Rigorous Elementary Mathematics

Volume 4: Geometry

Samer Seraj

Existsforall Academy

Copyright

ISBN 978-1-7389501-3-3

academy@existsforall.com

Acknowledgements

> "At the age of eleven, I began Euclid, with my brother as tutor. This was one of the great events of my life, as dazzling as first love. I had not imagined there was anything so delicious in the world. From that moment until I was thirty-eight, mathematics was my chief interest and my chief source of happiness."
>
> – *Bertrand Russell, Autobiography*

> "Mathematicians, like Proust and everyone else, are at their best when writing about their first love."
>
> – *Gian-Carlo Rota, Discrete Thoughts*

I express my gratitude to:

- The Almighty Creator, for providing me with this blessed and privileged life.
- My parents, for financing my mathematical education, and for supporting me during the time that this book series was written.
- My friends, for their companionship and for listening to me talk about mathematics.
- Euclid, for writing the *Elements*, which showed the world the meaning of eternal rigour.

Any mathematical errors or mistakes in the typesetting are my responsibility alone.

Contents

Preface vi

1 Lines 1
- 1.1 Equations of Lines . . . 1
- 1.2 Equipollence . . . 10

2 Angles 24
- 2.1 Basic Notions . . . 24
- 2.2 Parallel and Perpendicular . . . 33

3 Transformations 38
- 3.1 Formulas for Effects . . . 38
- 3.2 Preservation of Properties . . . 42

4 Dot Product 50
- 4.1 Algebraic Generalities . . . 50
- 4.2 Angles and Projections . . . 56

5 Polygons 61
- 5.1 Interior, Boundary, and Exterior . . . 61
- 5.2 Congruence and Similarity . . . 76

6 Circles I 80
- 6.1 Inscribed Angles . . . 80
- 6.2 Tangents, Secants, and Chords . . . 86

7 Classifications 95
- 7.1 Triangles . . . 95
- 7.2 Quadrilaterals . . . 106

8 Circles II 113
- 8.1 Cyclic Quadrilaterals . . . 113
- 8.2 Cyclic and Tangential Polygons . . . 124

9 Area 128
- 9.1 Using Lengths . . . 128
- 9.2 Using Coordinates . . . 136

10 Cevians 145
- 10.1 Computing Lengths . . . 145
- 10.2 Concurrency . . . 152

11 Triangle Centers 156
11.1 Examples . . . 156
11.2 Area Formulas . . . 168

12 Conics 172
12.1 Bivariate Quadratics . . . 172
12.2 Metric Properties . . . 190

13 Cross Product 196
13.1 Formulas . . . 196
13.2 Properties . . . 204

14 Three Dimensions 209
14.1 Prisms and Pyramids . . . 209
14.2 Curved Objects . . . 214

Appendices 219

A Solutions 220

List of Symbols 247

Bibliography 249

Index 250

Preface

> "In studying a philosopher, the right attitude is neither reverence nor contempt, but first a kind of hypothetical sympathy, until it is possible to know what it feels like to believe in his theories, and only then a revival of the critical attitude, which should resemble, as far as possible, the state of mind of a person abandoning opinions which he has hitherto held. Contempt interferes with the first part of this process, and reverence with the second. Two things are to be remembered: that a man whose opinions and theories are worth studying may be presumed to have had some intelligence, but that no man is likely to have arrived at complete and final truth on any subject whatever."
>
> – *Bertrand Russell, A History of Western Philosophy*

Mathematics is the study of ultimate regularity. Regularity entails order or predictability. Its antithesis is chaos. When there is regularity, there are discernible objects at play. In other words, there is structure. Wherever there is structure, there is symmetry. Symmetry means that, while one aspect of the object changes, another remains unchanged. The present series of books is an effort to rigorously systematize and provide an exposition of those aspects of elementary mathematics that appeal to the author. In the course of writing, it became evident that there are three recurring themes among the proof techniques used, all of which are forms of symmetry:

1. The discrete Fubini's principle instructs us to write the same thing in two different ways. For example, we have applied this principle in several ways:
   - A proof of the equality of vertical angles writes 180° in two ways (Theorem 2.11).
   - A proof of the Pythagorean theorem expresses the area of a square in two ways, equates them, and simplifies the result to get the famous $a^2 + b^2 = c^2$ (Theorem 9.12).
   - Writing the same vector as the sum of other vectors in multiple ways is helpful. This is used to get the barycentric coordinates of the incenter (Example 11.26).
   - The fourth height of a tetrahedron with three pairwise perpendicular edges that have a shared vertex is found by equating two ways of getting the volume (Problem 14.13).
2. Antisymmetry in a partial order is a powerful method of proof that lets us break down the strong notion of equality into the conjunction of two individually weaker statements. Examples that appear in this book are:
   - The proof that a certain type of equation represents lines and only lines in the plane invokes antisymmetry (Theorem 1.9).

- The proof of the standard triangle inequality (Theorem 3.10) uses antisymmetry.
- The set of points (usually in the plane) that satisfy a certain description is known as the locus of that criteria (Definition 6.4). Proving that a set of points is the locus of the description is a task for the two directions of the antisymmetry of sets. This is exemplified by the inscribed angle theorem (Theorem 6.6), perpendicular bisectors (Theorem 8.13), and angle bisectors (Theorem 8.18). Also, ellipses (Example 12.26) and hyperbolas (Problem 12.27) are found as certain loci.
- While studying conics (Chapter 12), we find that the minimum and maximum number of geometric constructions for a given conic are equal, so antisymmetry says that this is exactly the number of constructions.

3. Modding out by an equivalence relation allows us to focus on the essential properties of objects which are preserved under the relation.

   - To define vectors, we introduce an equivalence relation, called equipollence (Definition 1.20), on directed line segments. This allows us to work on certain preserved properties, such as magnitude and angles, with a convenient representative of the vector, called a position vector.
   - Ratios form important equivalence relations. We prove that two linear standard forms represent the same line if and only if the coefficients in one are all scaled by the same factor to produce the second, which means each line can be represented by a unique equivalence class or ratio of coefficients (Theorem 1.33). A similar result is proven for bivariate quadratics that represent constructible conics (Theorem 12.14). In the derivation of the cross product formula, we show that all normal vectors to a plane are scalar multiples of each other, with the cross product being one example, which means all coefficient quadruples of the standard equations of a plane are scalar multiples of each other (Theorem 13.14).

It is our hope that the reader will keep these proof techniques in mind while reading the book, and that the impression of the importance of symmetry will grow as the reader encounters the methods time and again.

The intended audience consists of students of math contests, competitions, and olympiads who want want to take a rigorous second look at the results they might be accustomed to taking for granted, and teachers, coaches, and trainers who want to reinforce their own understanding of what they teach.

Suggestions, comments, and error submissions would be greatly appreciated. These may include suggestions for strengthening or generalizing theorems, and additional material. Messages may be sent to

academy@existsforall.com

Samer Seraj
Mississauga, Ontario, Canada
June 25, 2024

Chapter 1

Lines

"As long as algebra and geometry have been separated, their progress have been slow and their uses limited, but when these two sciences have been united, they have lent each mutual forces, and have marched together towards perfection."

– *Joseph-Louis Lagrange*

We first define the Cartesian plane. Carving out subsets of the plane or space according to a function, equation, or a predicate is an immensely useful way of describing a geometric shape. This is a critical link between algebra and geometry. We will pay particular attention to the interchangeability of lines and linear functions. Afterwards, we will work with Euclidean vectors, which are often defined in elementary contexts as "arrows" such that arrows with the same length and direction are considered to be the same. This can be made rigorous using an equivalence relation called equipollence. Along the way, an accomplishment will be to prove the uniqueness of the standard form of a line in two dimensions, up to scaling.

1.1 Equations of Lines

We will model the Euclidean plane using the Cartesian plane. It is in this playground that we will be able to easily define and study geometric subsets of the plane, such as points, lines, line segments, rays, circles, and conics.

Definition 1.1. The Cartesian plane is the Cartesian product

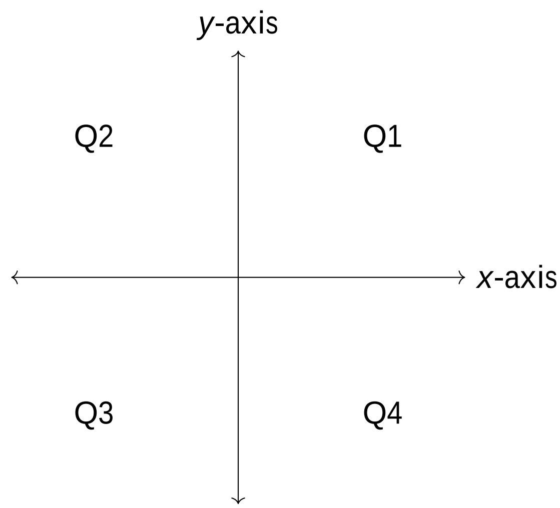
$$R^2 = R \quad R = f(x;y) : x \ 2 \ R; y \ 2 \ Rg;$$

the elements of which are called points and the components of each point $(x;y)$ are called its coordinates. Specifically, the x-coordinate is the left component and the y-coordinate is the right component. Visually, the Cartesian plane is a way of locating points on a flat plane. A point is represented by a single dot, since as it has no dimensions such as length or width, and a point is usually denoted by a capital letter like P. The x-coordinate of a point measures how far and in what direction the point lies horizontally, and the y-coordinate measures how far and in what direction the point lies vertically. Multiple points are said to coincide if they have the same pair of coordinates.

Definition 1.2. In the Cartesian plane, the origin is $(0;0)$, the y-axis is the vertical line through the origin consisting of all points $(x;y)$ such that $x = 0$, and the x-axis as the horizontal line through the origin consisting of all points $(x;y)$ such that $y = 0$. Intuitively, each axis can be considered to be a copy of the real number line.

Definition 1.3. The Cartesian plane has four quadrants, created by the axes. Starting from the top-right and going counterclockwise, these are:

1. Quadrant 1: on the top-right, where $x > 0$ and $y > 0$
2. Quadrant 2: on the top-left, where $x < 0$ and $y > 0$
3. Quadrant 3: on the bottom-left, where $x < 0$ and $y < 0$
4. Quadrant 4: on the bottom-right, where $x > 0$ and $y < 0$



Definition 1.4. The graph of a function $f : X \to Y$ is the set

$$\{(x, y) \in X \times Y : f(x) = y\},$$

though, formally speaking, this set of ordered pairs is actually the definition of the function itself. If X and Y are subsets of $\mathbb{R}$, then the graph of f can be represented visually by plotting all such pairs as points (x, y) on the Cartesian plane. As a reminder, recall that the definition of a function requires that every $x \in X$ has a corresponding $y \in Y$, but it is not necessarily true that every $y \in Y$ has a corresponding $x \in X$, since Y is a codomain and not necessarily the range.

Definition 1.5. If $f : X \to Y$ is a function where Y is a subset of the real numbers (or any set with an additive identity, like the field of complex numbers), then the preimage $f^{-1}(0)$ is called the zero set of f and its elements are called the roots or zeros of f. Geometrically, if X and Y are subsets of the real numbers, then the zero set is the set of points at which the x-axis intersects the function. The elements of $f^{-1}(0)$ are also sometimes called the solutions of the equation $f(x) = 0$.

The determination of the roots of functions is a fundamental problem in algebra. Part of what makes root-finding an even more frequent activity than one might imagine is that, for any fixed $y \in Y$, the determination of the preimage $f^{-1}(y)$ is equivalent to finding the roots of $g(x) = f(x) - y$ or the solutions of $f(x) = y$. In Volume 1, we developed methods of determining the roots of quadratic functions through factoring and the quadratic formula and, more generally, methods for polynomials.

Definition 1.6. If $k : X \times Y \to \mathbb{R}$ is a function, the **graph of the equation** $k(x, y) = 0$ is the set

$$\{(x, y) \in X \times Y : k(x, y) = 0\}.$$

This is also called the **zero set** of k. If X and Y are subsets of $\mathbb{R}$, then the graph can be plotted on the Cartesian plane. Sometimes, an equation is presented as

$$k_1(x, y) = k_2(x, y)$$

where k_1, k_2 are functions. In this case, its graph is defined as

$$\{(x, y) \in X \times Y : k_1(x, y) = k_2(x, y)\},$$

which is also the graph of $k(x, y)$ if we define $k = k_1 - k_2$.

Example. Graphs of equations are a generalization of the graphs of functions $f : X \to Y$ such that Y is a subset of $\mathbb{R}$: by defining $k(x, y) = f(x) - y$, we find that the graph of the equation $k(x, y) = 0$ is the graph of the function f. However, there are graphs of equations that are not the graph of any function, such as the circle

$$(x - a)^2 + (x - b)^2 = r^2$$

with center (a, b) and radius r (Corollary 2.3).

We will work on understanding the relationship between linear subsets of the plane (i.e. lines and its contiguous subsets) and linear functions.

Definition 1.7. Given two distinct points p, q in the plane, the **line through** p, q is defined as the set of points

$$\{p + t(q - p) : t \in \mathbb{R}\} = \{(1 - t)p + tq : t \in \mathbb{R}\}.$$

If the capital letters P, Q are used to denote the endpoints, then the line may be denoted, as an object, by $\overleftrightarrow{PQ}$. There are some contiguous subsets of lines:

- The **line segment with endpoints** p and q is the set

  $$\{p + t(q - p) : t \in [0, 1]\}.$$

  Note that we get p and q for $t = 0$ and $t = 1$ respectively. If the capital letters P, Q denote the endpoints p, q, then the segment is denoted as $\overline{PQ}$, though we will often drop the bar to simply write PQ. Moreover, PQ is also used to refer to the segment's **length** (this is the distance between the endpoints, as give by Theorem 2.1) or even the line through P, Q in informal settings. Rest assured, our terminology and notation will avoid ambiguity, even if it will not always be formal. The **midpoint** of the segment is the point in the set when the parameter is $t = \frac{1}{2}$.

- The **open line segment** corresponding to segment PQ is PQ without its endpoints P and Q, which means we are talking about the set

  $$\{p + t(q - p) : t \in (0, 1)\}.$$

  This is also called the **interior** of line segment PQ.

- The ray with origin p and containing the point q is the set

  $$\{p + t(q - p) : t \in \mathbb{R}_{\geq 0}\}.$$

  This is an object that is extended indefinitely on only one end instead of how a line is extended on both ends. If the capital letters P, Q are used to denote p, q respectively, then the ray is denoted as $\overrightarrow{PQ}$. Any point on the ray, except P may be chosen to be the direction-determining point Q.

Different regions of a line, relative to two points p and q on the line, are discussed in Theorem 3.12.

Lemma 1.8. A set ℓ is a line if and only if there exists a point $p = (p_1, p_2)$ and a non-zero point $v = (v_1, v_2)$ such that

$$\ell = \{p + tv : t \in \mathbb{R}\}.$$

Proof. In one direction, let

$$\ell = \{p + t(q - p) : t \in \mathbb{R}\}$$

be a line. We can let $v = q - p \neq 0$ to get that

$$\ell = \{p + tv : t \in \mathbb{R}\}.$$

In the other direction, let $\ell = \{p + tv : t \in \mathbb{R}\}$ for some non-zero point v in the plane. We can let $q = p + v$ so that $v = q - p$, and get the line

$$\ell = \{p + t(q - p) : t \in \mathbb{R}\}.$$

□

Theorem 1.9 (Equation of a general line). Let $p = (p_1, p_2)$ be a point in $\mathbb{R}^2$ and let $v = (v_1, v_2)$ be a point that is not the origin. Then

$$\{p + tv : t \in \mathbb{R}\} = \{(x, y) \in \mathbb{R}^2 : (y - p_2)v_1 = (x - p_1)v_2\}.$$

According to Lemma 1.8, a set is a line if and only if it can be represented in the form of the set on the left side, so the right side provides a second way of representing all lines (and only lines). A conversion process is described below.

Proof. We will prove this by showing that the two sets are subsets of each other, which allows us to invoke antisymmetry. If

$$(x, y) = (p_1, p_2) + t(v_1, v_2) = (p_1 + tv_1, p_2 + tv_2)$$

is an element of $\{p + tv : t \in \mathbb{R}\}$, then as long as $t \neq 0$,

$$(y - p_2)v_1 = (p_2 + tv_2 - p_2)\left(\frac{x - p_1}{t}\right) = (x - p_1)v_2$$

If $t = 0$, then we can immediately verify as a separate case that it holds that

$$(y - p_2)v_1 = 0 = (x - p_1)v_2$$

because $(x, y) = (p_1, p_2)$. Either way, this proves the inclusion.
Now suppose $(x, y) \in \mathbb{R}^2$ satisfies

$$(y - p_2)v_1 = (x - p_1)v_2.$$

If v_1 and v_2 are non-zero, then we can rewrite the equation as

$$\frac{y - p_2}{v_2} = \frac{x - p_1}{v_1}.$$

Let t be the common value that is equal to both sides of the equation. Then

$$y = tv_1 + p_1,$$
$$y = tv_2 + p_2,$$

which puts

$$(x, y) = (tv_1 + p_1, tv_2 + p_2) = (p_1, p_2) + t(v_1, v_2) = p + tv$$

in the desired form. Since (v_1, v_2) is assumed to not be the origin, we just need to take care of the possibility that $v_1 = 0$ and $v_2 \neq 0$, or $v_2 = 0$ and $v_1 \neq 0$. If $v_1 = 0$, then

$$(x - p_1)v_2 \implies x = p_1$$

and y can be anything, making a vertical line (horizontal and vertical lines are defined in Definition 1.12). Each point of the form (p_1, y) can be put in the form

$$(p_1 + t \cdot 0, p_2 + tv_2) = (p_1 + t \cdot v_1, p_2 + tv_2) = p + tv$$

for $t = \frac{y - p_2}{v_2}$. The case of $v_2 = 0$ involves a horizontal line and it may be handled symmetrically. This proves the inclusion.
Since the two representations can be obtained from each other from the shown processes, they are equivalent in representative power, in that they both capture all lines in $\mathbb{R}^2$ and only lines in $\mathbb{R}^2$. □

Corollary 1.10 (Point-point form of a line). A subset ℓ of the plane is a line if and only if there exist distinct points $p = (p_1, p_2)$ and $q = (q_1, q_2)$ in the plane such that

$$\ell = \{(x, y) \in \mathbb{R}^2 : (y - p_2)(q_1 - p_1) = (x - p_1)(q_2 - p_2)\}.$$

This is called the **point-point form** of a line because both points p, q satisfy the equation, so this is a way of finding the equation of a line that goes through two distinct points.

Proof. By the definition of a line, a set ℓ is a line if and only if there exist distinct points $p = (p_1, p_2)$ and $q = (q_1, q_2)$ such that

$$\ell = \{p + t(q - p) : t \in \mathbb{R}\}.$$

By Theorem 1.9,

$$\ell = \{(x, y) \in \mathbb{R}^2 : (y - p_2)(q_1 - p_1) = (x - p_1)(q_2 - p_2)\},$$

so we can go from one form to another, as desired. □

Corollary 1.11 (Standard form of a line). A subset ℓ of the plane is a line if and only if there exist real numbers A, B, C such that $A^2 + B^2 \neq 0$ (so at least one of A, B is non-zero) and

$$\ell = \{(x, y) \in \mathbb{R}^2 : Ax + By + C = 0\}.$$

This is called the **standard form** of a line.

Proof. Let ℓ be a line. By Theorem 1.9, there exists a point $p = (p_1, p_2)$ and a non-zero point $v = (v_1, v_2)$ in $\mathbb{R}^2$ such that

$$\ell = \{p + tv : t \in \mathbb{R}\} = \{(x, y) \in \mathbb{R}^2 : (y - p_2)v_1 = (x - p_1)v_2\}.$$

The form on the right side can be written as

$$\{(x, y) \in \mathbb{R}^2 : v_2x + (-v_1)y + (p_2v_1 - p_1v_2) = 0\}.$$

So we can take $A = v_2, B = -v_1, C = p_2v_1 - p_1v_2,$ where it holds that

$$A^2 + B^2 = v_2^2 + (-v_1)^2 = v_1^2 + v_2^2 \neq 0.$$

In the other direction, we want to show that any set

$$\ell = \{(x, y) \in \mathbb{R}^2 : Ax + By + C = 0\}$$

is a line if $A^2 + B^2 \neq 0.$ First we note that ℓ is non-empty: if one of A, B is zero, then we can isolate the non-vanishing variable to produce infinitely many points; if neither of A, B is zero, then we can substitute any value like 0 for one of the variables and isolate the other one to produce a point. This is helpful because we can find a point (p_1, p_2) and substitute it in to get

$$Ap_1 + Bp_2 + C = 0.$$

As a result,

$$\begin{aligned}\{(x, y) \in \mathbb{R}^2 : Ax + By + C = 0\} &= \{(x, y) \in \mathbb{R}^2 : Ax + By - (Ap_1 + Bp_2) = 0\}\\ &= \{(x, y) \in \mathbb{R}^2 : A(x - p_1) + B(y - p_2) = 0\}\\ &= \{(x, y) \in \mathbb{R}^2 : A(x - p_1) = (-B)(y - p_2)\},\end{aligned}$$

Since (A, B) is not the origin. this the equation of a line, according to Theorem 1.9. □

Standard forms are not unique since we can multiply through the coefficients by any non-zero factor without altering the underlying set of points. If A, B, C are rational, it is usually good practice to clear the denominators, divide out by the greatest common divisor of the new coefficients, and scale by -1, if needed, to avoid A from being negative.

Definition 1.12. A **horizontal line** is a set

$$\{(x, y) \in \mathbb{R}^2 : y = c\}$$

for some real constant $c.$ A **vertical line** is a set

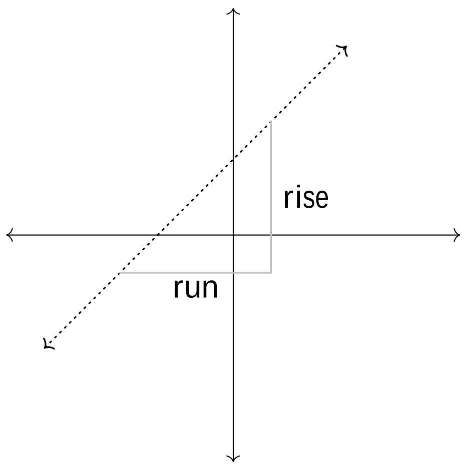
$$\{(x, y) \in \mathbb{R}^2 : x = c\}$$

for some real constant $c.$ It is easy to verify that horizontal lines and vertical lines are in fact lines, by taking $A = 0$ or $B = 0$ respectively in the standard form of a line $Ax + By + C = 0$.

Theorem 1.13. For any non-vertical line ℓ, there exists a constant m called the slope of the line, such that for any two distinct points (x_1, y_1) and (x_2, y_2) on the line,

$$m = \frac{y_2 - y_1}{x_2 - x_1}.$$

Note that this idea does not work for vertical lines because the fact that $x_1 = x_2$ would lead to division by 0.



Proof. Let ℓ be a non-vertical line and let a standard form of its equation be

$$Ax + By + C = 0,$$

where $A^2 + B^2 \neq 0$. Let (x_1, y_1) and (x_2, y_2) be two distinct points on the line (though it does not matter here, these must exist since the original definition of a line shows that every line has as many points as there are real numbers, meaning lines are in bijection with $\mathbb{R}$). Then

$$Ax_1 + By_1 + C = 0,$$
$$Ax_2 + By_2 + C = 0.$$

Subtracting the equations yields

$$A(x_1 - x_2) + B(y_1 - y_2) = 0$$

which implies that

$$\frac{y_1 - y_2}{x_1 - x_2} = -\frac{A}{B}$$

is a constant. Note that this computation is possible because $B \neq 0$ due to the line being non-vertical. □

Corollary 1.14 (Point-slope form of a line). A subset ℓ of the plane is a non-vertical line with slope m and contains the point (p_1, p_2) if and only if

$$\ell = \{(x, y) \in \mathbb{R}^2 : y - p_2 = m(x - p_1)\}$$

for real numbers m and a real pair of coordinates (p_1, p_2). This is called the **point-slope form** of a non-vertical line.

Proof. In one direction, suppose ℓ is a line. Let a standard form of the line ℓ be

$$Ax + By + C = 0.$$

By the proof of Corollary 1.11,

$$\begin{aligned} \ell &= \{(x, y) \in \mathbb{R}^2 : Ax + By + C = 0\} \\ &= \{(x, y) \in \mathbb{R}^2 : A(x - p_1) = (-B)(y - p_2) = 0\} \\ &= \{(x, y) \in \mathbb{R}^2 : y - p_2 = -\frac{A}{B}(x - p_1)\}. \end{aligned}$$

By the proof of Theorem 1.13, $-\dfrac{A}{B} = m$, so

$$\ell = \{(x, y) \in \mathbb{R}^2 : y - p_2 = m(x - p_1)\}.$$

In the other direction, let ℓ be a subset of the plane such that there exists a real number m and a point (p_1, p_2) satisfying

$$\ell = \{(x, y) \in \mathbb{R}^2 : y - p_2 = m(x - p_1)\}.$$

The defining equation is equivalent to

$$mx + (-1)y + (-mp_1 + p_2) = 0,$$

which is in standard form, making ℓ a line. Moreover, it is non-vertical because we can find two points that satisfy the equation with differing x-coordinates: (p_1, p_2) and $(p_1 + 1, p_2 + m)$. □

Definition 1.15. The **y-intercept** of a non-vertical line is the unique y-axis point $(0, y_0)$ that lies on the line. The **x-intercept** of a non-horizontal line is the unique y-axis point $(x_0, 0)$ that lies on the line. The existence and uniqueness of these points follow from plugging 0 into the appropriate variable in a standard form of a line and solving for the other variable.

Corollary 1.16 (Slope-intercept form of a line). A subset ℓ of the plane is a non-vertical line with slope m and y-intercept $(0, b)$ if and only if

$$\ell = \{(x, y) \in \mathbb{R}^2 : y = mx + b\}$$

for real numbers m and b. This is called the **slope-intercept form** of a non-vertical line.

Proof. Suppose ℓ is a line with slope m and y-intercept $(0, b)$. Plugging $(x, y) = (0, b)$ into the point-slope form of a line (Corollary 1.14) yields

$$\begin{aligned}\ell &= \{(x, y) \in \mathbb{R}^2 : y - b = m(x - 0)\} \\ &= \{(x, y) \in \mathbb{R}^2 : y = mx + b\}.\end{aligned}$$

Conversely, suppose

$$\ell = \{(x, y) \in \mathbb{R}^2 : y = mx + b\}$$

for some real numbers m and b. As above, since this is equivalent to

$$\ell = \{(x, y) \in \mathbb{R}^2 : y - b = m(x - 0)\},$$

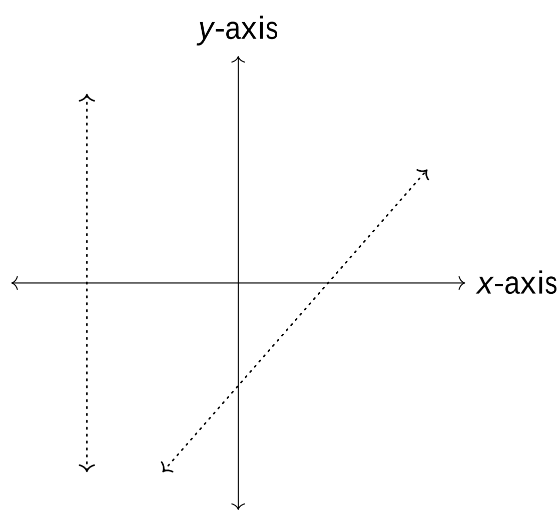
Corollary 1.14 tells us that ℓ is a non-vertical line with slope m and contains the point $(0, b)$, which must be the y-intercept since its x-coordinate is 0. □

Definition 1.17. A **univariate linear function** is any function $f : \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) = ax + b,$$

where a and b are real constants. Here, a is called the **linear coefficient** and b is the **constant term**. If $a = 0$ then f is a **constant function**, and if $a = 1$ and $b = 0$ then f is the **identity function** on the real numbers. It is easy to see that if $a \neq 0$, then f has exactly one root $-\frac{b}{a}$; if $a = 0$, then either f has no roots in the case that $b \neq 0$, or all real numbers are roots of f in the case that $b = 0$.

Example. A vertical line cannot be the graph of a function, but non-vertical lines are the graphs of linear functions $f(x) = mx + b$ for fixed real numbers m and b, and the graphs of all linear functions are non-vertical lines, thanks to the slope-intercept form (Corollary 1.16).



Problem 1.18. For two linear equations of the form

$$\begin{aligned}ax + by &= c, \\ \alpha x + \beta y &= \gamma,\end{aligned}$$

show that there must be exactly zero, one or infinite solutions by proving that the existence of two distinct solutions to such a system implies the existence of infinitely many solutions for the same system. Here, the variables are (x, y), and $a, b, c, \alpha, \beta, \gamma$ are constants.

1.2 Equipollence

While three-dimensional geometry will be covered in Chapter 14, let us briefly jump up to investigate n dimensions. Unfortunately, the third dimension is the highest one that we can visualize, and even there it is not feasible to visualize all cross sections simultaneously. So in a sense, we can draw and visualize up to only two-dimensional objects (and surfaces in three dimensions). This does not stop up from defining and studying higher dimensions, though our intuition for them will have to come from our direct experience with lower dimensions.

Definition 1.19. For each positive integer $n,$ the **n-dimensional Euclidean space** is defined as $\mathbb{R}^n,$ which is the set of n-tuples of real numbers. We may call this n-space, and each of its elements is called a **point**. For example, we extend the Cartesian coordinate system of pairs of real numbers (Definition 1.1) to three dimensional Cartesian space, which consists of triples of real numbers

$$\mathbb{R}^3 = \{(x, y, z) : x \in \mathbb{R}, y \in \mathbb{R}, z \in \mathbb{R}\}.$$

As formalized below, a Euclidean vector in n-space is described as a directed line segment, where "directed" means that we distinguish the starting "tail" endpoint of the segment and the ending "arrow" endpoint of the segment. There is an important caveat, which is that translating the segment does not change it, making vectors position-independent. This caveat indicates that there is an equivalence relation at play, and we will precisely define it momentarily. It is said that a Euclidean vector has magnitude (i.e. length) and direction (e.g. its counterclockwise angle with the positive x-axis), but no particular location.

Definition 1.20. A **directed line segment** in $\mathbb{R}^n$ is an ordered pair of elements $(p, q) \in \mathbb{R}^n \times \mathbb{R}^n.$ Visually, p is the starting point called the **tail** and q is the ending point called the **arrowhead**, and we draw the segment from p to $q.$ The relation

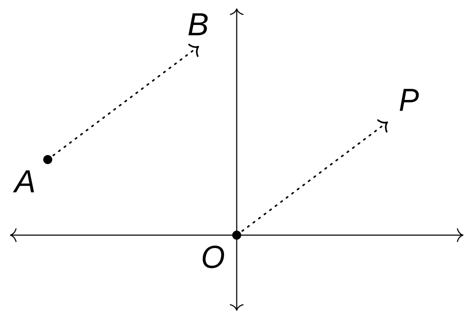
$$(p, q) \sim (x, y) \iff q - p = y - x$$

(where the subtraction is done component-wise in $\mathbb{R}^n$) is easily seen to be an equivalence relation because it is reflexive, symmetric and transitive, as defined in Volume 1. This relation is called **equipollence** and the resulting equivalence classes are called **Euclidean vectors** or simply **vectors**. If

$$(p, q) \sim (x, y),$$

then the two elements (p, q) and (x, y) are said to be **equipollent** to each other and the relation may be specifically denoted by the bump equality symbol $\Bumpeq$ instead of the generic $\sim.$

Definition 1.21. The element $(0, p)$ of an equipollence class where the first component is the zero vector is called the **canonical representative** or **position vector** of that class (make sure you see that a every class has a unique such canonical element). An arbitrary element (p, q) of an equipollence class is called a **displacement vector** from p to $q.$ A position vector $(0, p)$ may be denoted by $\vec{p}$ or simply p if there is no chance of confusion, say with Euclidean points or real numbers. A displacement vector (A, B) may be denoted by $\overrightarrow{AB}$: To denote the equipollence class of a displacement vector u; we use the notation $[u]$, which is the standard notation of equivalence classes. If $(0; p)$ is a position vector, we might also denote its equipollence class by $[p]$: Sometimes we use just a letter like v to denote an equipollence class, otherwise known as a vector.



Example. If $p = q$ are elements of $\mathbb{R}^m$, then the vector

$$[(p; q)] = [(p; p)] = [(0; 0)]$$

consists of all "points," so to speak, which are directed line segments with zero length. We call this the zero vector.

While equipollence classes are the formal route to Euclidean vectors, we will largely work with displacement vectors, that is directed line segments, and equate them with other equipollent segments, such as position vectors, as will be relevant or convenient.

Definition 1.22. Let $(0; v)$ and $(0; w)$ be position vectors in $\mathbb{R}^n$; and let r be a real number. There are two standard operations defined on vectors:

1. Vector addition: $[v] + [w] = [v + w]$ where the addition $v + w$ is done as the usual component-addition in $\mathbb{R}^n$: We will see a geometric interpretation of this momentarily in what is called the parallelogram law (Theorem 1.39).

2. Scalar multiplication: $r[v] = [rv]$; where rv is the usual component-wise scalar multiplication in $\mathbb{R}^n$: Intuitively, this scales (stretches or compresses) a vector, possibly combined with reflecting the direction to its opposite.

The definitions of these two operations rely on $(0; v)$ and $(0; w)$ being position vectors, and they cannot be replaced by arbitrary displacement vectors without adjusting the definitions. If $(p; q)$ and $(x; y)$ are displacement vectors, then we find their equipollent position vectors $(0; q \quad p)$ and $(0; y \quad x)$; and then apply addition or scalar multiplication and take the equipollence class of the result.

Theorem 1.23 (Vector space axioms). It may be verified without difficulty that the eight axioms of abstract vector spaces, as listed below, are fulfilled by the Euclidean vector space of $\mathbb{R}^n \quad \mathbb{R}^n$ under equipollence. Let $(0; u); (0; v); (0; w)$ be arbitrary position vectors in $\mathbb{R}^n$; and let $r; s$ be real numbers. Then:

1. After $\mathbb{R}^n \quad \mathbb{R}^n$ is "modded out" by l, it is an abelian group under vector addition $+$: This fancy language simply means that the following four conditions are met:

(a) Commutativity: $[v] + [w] = [w] + [v]$

(b) Associativity: $([u] + [v]) + [w] = [u] + ([v] + [w])$

(c) Existence of identity: $[0]$ satisfies $[0] + [u] = [u]$

(d) Existence of inverses: $[u] + [-u] = [0]$

A part of mathematical thinking is to correctly interpret the symbol $0,$ which appears as many objects from numbers to vectors, other algebraic structures, functions, and polynomials.

2. There are desirable regularities in scalar multiplication:

(a) $1[u] = [u]$

(b) $(rs)[u] = r(s[u])$

(c) $r([v] + [w]) = r[v] + r[w]$

(d) $(r + s)[v] = r[v] + s[v]$

We leave it to the reader to verify that these properties are satisfied by unwrapping the definitions of vector addition and scalar multiplication.

It maybe difficult to see at first why Euclidean vectors are a useful construct. One immediately upcoming benefit is that they will help us to prove a fact that may have already been suspected by the reader: the standard form of a line is unique up to multiplication by a non-zero scalar (Theorem 1.33). In the long run, vectors show their utility like any other equivalence relation, by allowing us to use more convenient representatives of an equipollence class (this representative is usually a position vector or has a particular tail), while preserving properties that are relevant in a given scenario.

Definition 1.24. A set of vectors $\{v_1, v_2, \ldots, v_m\}$ in $\mathbb{R}^n$ is said to be **linearly independent** when, for all $t_1, t_2, \ldots, t_m \in \mathbb{R},$ if

$$t_1v_1 + t_2v_2 + \cdots + t_mv_m = 0,$$

then

$$t_1 = t_2 = \cdots = t_m = 0.$$

(Do you see why linear independence implies that none of the v_i are the 0 vector?) In the negation, meaning there exist $t_1, t_2, \ldots, t_m \in \mathbb{R}$ (at least one of which is non-zero) such that

$$t_1v_1 + t_2v_2 + \cdots + t_mv_m = 0,$$

we say that the set of vectors $\{v_1, v_2, \ldots, v_m\}$ is **linearly dependent**. The expression

$$t_1v_1 + t_2v_2 + \cdots + t_mv_m$$

is called a **linear combination** of the set of vectors $\{v_i : i \in [m]\}$.

Lemma 1.25. The following conditions on two vectors v, w in $\mathbb{R}^n$ are equivalent:

1. $v = 0$ or $w = tv$ for some real t
2. v and w are linearly dependent
3. $w = tv$ or $v = tw$ for some real t

Note the similarity of the first condition with the equality condition for the Cauchy-Schwarz inequality from Volume 1.

Proof. We will prove that

$$(1) \implies (2) \implies (3) \implies (1).$$

- (1) $\implies$ (2): Suppose $v = 0$. Then, for $t_1 = 1$ and $t_2 = 0$, $t_1v + t_2w = 0$, proving linear dependence. Separately, suppose $w = tv$ for some real t. Then, for $t_1 = -t$ and $t_2 = 1$,
$$t_1v + t_2w = (-t)v + 1(tv) = (-t + t)v = 0v = 0;$$
again proving linear dependence.
- (2) $\implies$ (3): Suppose v and w are linearly dependent. Then there exist real t_1 and t_2, at least one of which is non-zero, such that $t_1v + t_2w = 0$. If $t_1 \neq 0$, then we can let $t = -\frac{t_1}{t_2}$ to get $w = tv$. If $t_2 \neq 0$, then we can let $t = -\frac{t_2}{t_1}$ to get $v = tw$.
- (3) $\implies$ (1): Supposing $w = tv$ for some real t immediately implies (1). So suppose $v = tw$ for some real t instead. If $v = 0$, then we are done again. So suppose instead that $v \neq 0$. Then it is not possible that $t = 0$ in $v = tw$. This allows us to divide by t in $v = tw$ to get $w = \frac{1}{t}v$, which completes the proof.

□

Definition 1.26. Recall from Volume 1 that an $m \times n$ matrix is a way of organizing mn entries, indexed by $[m] \times [n]$, in a block as shown in the 2×2 case below (here, $[m]$ refer to the set $\{1, 2, \ldots, m\}$, and similarly for $[n]$). The **matrix multiplication** of 2×2 matrices is defined by

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} p & q \\ r & s \end{bmatrix} = \begin{bmatrix} ap + br & aq + bs \\ cp + dr & cq + ds \end{bmatrix}$$

Similarly, a 2×2 matrix times a column vector is computed as

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} p \\ q \end{bmatrix} = \begin{bmatrix} ap + bq \\ cp + dq \end{bmatrix}.$$

The multiplication of larger matrices follows a similar definition. Matrix multiplication is associative, but not commutative in general. The multiplication of a matrix by a scalar (for us, a scalar is a real number but a more general definition involving field elements exists in abstract algebra) is defined as

$$\lambda \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} \lambda a & \lambda b \\ \lambda c & \lambda d \end{bmatrix}.$$

We call

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

the 2×2 **identity matrix** because $IM = MI$ for every 2×2 matrix M, as the reader should verify.

To foreshadow, we will need matrix multiplication in the proof of the barycentric shoelace formula (Theorem 9.28). The multiplication of a matrix with a column vector will be useful as well for succinctly expressing a system of equations in the proof of the uniqueness of bivariate quadratic representations of conics (Theorem 12.14).

Definition 1.27. We will not define the **determinant** of an $n \times n$ matrix in general but the following formulas show how to compute the determinants of 2×2 and 3×3 matrices:

- The determinant of a 2×2 matrix is the real number

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc.$$

- The determinant of a 3×3 matrix is the real number

$$\det \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} = aei + bfg + cdh - ceg - bdi - afh.$$

We will need the following properties that hold for the determinants of all square matrices, though we will need to apply them only to 2×2 and 3×3 matrices.

Theorem 1.28. The facts below about determinants may be verified using the algebraic definitions (Definition 1.27), albeit the direct proofs are are tedious:

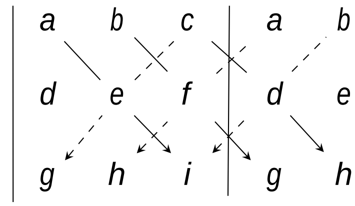
1. A scalar can be factored out of a single row or a single column of the determinant to outside the determinant without altering the value of the determinant. Equivalently, multiplying every element of one row or every element of one column by the same number results in multiplying the determinant by the same number.

2. Adding or subtracting (a scalar multiple) of a row from another row preserves the determinant. The same is true for the corresponding statement about columns.

3. Swapping two rows with each other or two columns with each other changes the sign of the determinant.

4. The determinant is preserved under taking the transpose of the matrix, meaning the element in each row i and column j is exchanged with the element in row j and column i. This operation is equivalent to reflecting the matrix across the diagonal that runs from the top-left to the bottom-right.

The following is a way of remembering how to compute the 3×3 determinant, called the **Rule of Sarrus**:



Lemma 1.29. In two dimensions, the position vectors (x_1, x_2) and (y_1, y_2) are linearly independent if and only if

$$\det \begin{bmatrix} x_1 & x_2 \\ y_1 & y_2 \end{bmatrix} = x_1 y_2 - x_2 y_1$$

is non-zero.

Proof. We will prove the contrapositive, that (x_1, x_2) and (y_1, y_2) are linearly dependent if and only if $x_1 y_2 = x_2 y_1$, since an equality is usually easier to manipulate than an inequation. For one direction, suppose they are linearly dependent. Then there exist constants $a, b,$ at least one of which is non-zero, such that

$$(0, 0) = a(x_1, y_1) + b(x_2, y_2) = (ax_1 + bx_2, ay_1 + by_2).$$

There are two scenarios, depending on whether $a \neq 0$ or $b \neq 0$ and they lead to the same conclusion (if both hold, then either argument applies):

$$a \neq 0 \implies a(x_1 y_2 - x_2 y_1) = (ax_1)y_2 - x_2(ay_1) = (-bx_2)y_2 - x_2(-by_2) = 0,$$
$$b \neq 0 \implies b(x_1 y_2 - x_2 y_1) = x_1(by_2) - (bx_2)y_1 = (-ay_1)x_1 - (-ax_1)y_1 = 0.$$

In either case, we can cancel $a \neq 0$ or $b \neq 0$ to get $x_1 y_2 - x_2 y_1 = 0.$
In the other direction, suppose $x_1 y_2 - x_2 y_1 = 0$. We will make use of Lemma 1.25 several times to prove linear dependence. Ideally, we would be able to rewrite the equation as

$$\frac{y_2}{x_2} = \frac{y_1}{x_1}.$$

In this case, we can set t equal to the common value so that $y_1 = tx_1$ and $y_2 = tx_2,$ resulting in the dependence relation

$$(y_1, y_2) = t \cdot (x_1, x_2).$$

The problem is that we might have $x_1 = 0$ or $x_2 = 0,$ in which case we cannot divide by one or both. If $x_1 = 0,$ then $x_2 = 0$ or $y_1 = 0.$ As a subcase, if $x_2 = 0$ then $(x_1, x_2) = 0$ and we have linear dependence. As another subcase, if $y_1 = 0$ then $(0, x_2)$ and $(0, y_2)$ are linearly dependent (work this out!). This takes care of the $x_1 = 0$ case. The analysis for the $x_2 = 0$ case is symmetrical, thereby completing the second direction of the proof. □

Corollary 1.30. If the position vectors (x_1, x_2) and (y_1, y_2) are linearly independent, then the matrix

$$M = \frac{1}{x_1 y_2 - x_2 y_1} \begin{bmatrix} y_2 & -x_2 \\ -y_1 & x_1 \end{bmatrix}$$

can be defined and it is the unique inverse of $N = \begin{bmatrix} x_1 & x_2 \\ y_1 & y_2 \end{bmatrix}$ in the sense that $M \cdot N = N \cdot M = I$. Conversely, if (x_1, x_2) and (y_1, y_2) are linearly dependent, then the above-defined matrix M has no inverse, meaning there does not exist a matrix N as described.

Proof. It is straightforward to show by the definition of matrix multiplication that

$$\left(\frac{1}{x_1y_2 - x_2y_1}\begin{bmatrix} y_2 & -x_2 \\ -y_1 & x_1 \end{bmatrix}\right)\begin{bmatrix} x_1 & x_2 \\ y_1 & y_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$
$$= \begin{bmatrix} x_1 & x_2 \\ y_1 & y_2 \end{bmatrix}\left(\frac{1}{x_1y_2 - x_2y_1}\begin{bmatrix} y_2 & -x_2 \\ -y_1 & x_1 \end{bmatrix}\right).$$

Recall from Volume 1 that inverses must be unique under an associative binary operation that has an identity. In this case, proving that the multiplication of 2×2 matrices is associative is not difficult (though it is algebraically tedious), and we find that I is the identity matrix. For the converse, suppose (x_1, x_2) and (y_1, y_2) are linearly dependent. By Lemma 1.25, one of them must be a scalar multiple of the other. Since the argument will be symmetrical in the two cases, suppose without loss of generality that there exists a real number t such that

$$(y_1, y_2) = t(x_1, x_2) = (tx_1, tx_2).$$

Now suppose, for the sake of contradiction, that there exists an inverse $\begin{bmatrix} p & q \\ r & s \end{bmatrix}$ so that

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} p & q \\ r & s \end{bmatrix}\begin{bmatrix} x_1 & x_2 \\ tx_1 & tx_2 \end{bmatrix}$$
$$= \begin{bmatrix} px_1 + tqx_1 & px_2 + tqx_2 \\ rx_1 + tsx_1 & rx_2 + tsx_2 \end{bmatrix}$$
$$= \begin{bmatrix} (p + tq)x_1 & (p + tq)x_2 \\ (r + ts)x_1 & (r + ts)x_2 \end{bmatrix}.$$

But the determinant of the far left side is 1 and the determinant of the far right side is 0, which is a contradiction. Therefore, an inverse does not exist for 2×2 matrices whose rows are linearly dependent. □

Problem 1.31. Prove that, in two dimensions, the set of arrowheads of the position vectors of linear combinations of two linearly independent position vectors generates the whole plane $\mathbb{R}^2$. For example, the arrowheads of the two position vectors might be the standard basis, $(1, 0)$ and $(0, 1)$, in which case

$$(a, b) = a(1, 0) + b(0, 1)$$

for every point $(a, b) \in \mathbb{R}^2$.

Problem 1.32. Let u, v, w be vectors in $\mathbb{R}^n$ such that v is non-zero. Prove that, if u, v are linearly dependent and v, w are linearly dependent, then u, w are linearly dependent. Intuitively, this means that linear dependence is partially transitive, specifically in the case where the transition vector is non-zero.

Theorem 1.33 (Uniqueness of standard form through two points)**.** If (x_1, y_1) and (x_2, y_2) are distinct points in $\mathbb{R}^2$ that both lie on both of the lines

$$\ell = \{(x, y) \in \mathbb{R}^2 : Ax + By + C = 0\},$$
$$\ell' = \{(x, y) \in \mathbb{R}^2 : A'x + B'y + C' = 0\},$$

where (A, B, C) and (A', B', C') are triples of real constants such that $A^2 + B^2 \neq 0$ and $A'^2 + B'^2 \neq 0$, then there exists a non-zero $t \in \mathbb{R}$ such that

$$(A, B, C) = t \cdot (A', B', C').$$

This implies that, if two lines ℓ and ℓ' in $\mathbb{R}^2$ have at least two distinct points in common, then the lines "coincide" in the sense that they are the same set $\ell = \ell'$. The analogous result for the uniqueness of bivariate quadratic representations of conics is Theorem 12.14.

Proof. By substitution, we obtain the equations

$$\begin{aligned} Ax_1 + By_1 + C &= 0, \\ Ax_2 + By_2 + C &= 0, \end{aligned}$$

and subtracting them yields

$$A(x_1 - x_2) - B(y_2 - y_1) = 0.$$

By Lemma 1.29, (A, B) and $(x_1 - x_2, y_2 - y_1)$ are linearly dependent. Similarly, by substitution, we obtain the analogous second set of equations

$$\begin{aligned} A'x_1 + B'y_1 + C' &= 0, \\ A'x_2 + B'y_2 + C' &= 0, \end{aligned}$$

and subtracting them yields

$$A'(x_1 - x_2) - B'(y_2 - y_1) = 0.$$

By Lemma 1.29, $(x_1 - x_2, y_2 - y_1)$ and (A', B') are linearly dependent. Since $(x_1, y_1), (x_2, y_2)$ are distinct and so $(x_1 - x_2, y_2 - y_1) \neq (0, 0)$, Problem 1.32 tells us that (A, B) and (A', B') are linearly dependent. Since $(A', B') \neq (0, 0)$, there exists $t \in \mathbb{R}$ such that $(A, B) = t \cdot (A', B')$ by Lemma 1.25. Since $(A, B) \neq (0, 0)$, we also obtain that the scale factor t is non-zero. Lastly,

$$\begin{aligned} C &= -Ax_1 - By_1 \\ &= -tA'x_1 - tB'x_1 \\ &= -t(A'x_1 + B'y_1) \\ &= -t(-C') \\ &= tC'. \end{aligned}$$

Therefore, there exists a non-zero real t such that

$$(A, B, C) = t \cdot (A', B', C').$$

□

Corollary 1.34. Two lines in $\mathbb{R}^2$ have only three possible options for their intersection: there is exactly one point in common, there are no points in common, or they have all points in common.

Proof. Suppose the first two options are untrue, meaning there exists a point in common, but there is not exactly one point in common. Then there exist at least two points in common. By Theorem 1.33, the two lines are identical, meaning they have all points in common. In other words, if the lines are ℓ and ℓ', then

$$\ell = \ell \cap \ell' = \ell'.$$

This strengthens the infinitude of common solutions in Problem 1.18 to completely overlapping. □

Definition 1.35. If two lines in $\mathbb{R}^2$ have exactly zero intersections then we call the lines **parallel**. If the two lines are the same set, then we call the lines **coincident**.

Theorem 1.36. *Let $p = (p_1, p_2)$ and $q = (q_1, q_2)$ be points in $\mathbb{R}^2$, and $v = (v_1, v_2)$ and $w = (w_1, w_2)$ be non-zero points in $\mathbb{R}^2$. Define the lines*

$$\ell_1 = \{(x, y) : (y - p_2)v_1 = (x - p_1)v_2\} = \{p + tv : t \in \mathbb{R}\},$$
$$\ell_2 = \{(x, y) : (y - q_2)w_1 = (x - q_1)w_2\} = \{q + sw : s \in \mathbb{R}\}.$$

Then the follow conditions are equivalent:

1. *The lines ℓ_1 and ℓ_2 are parallel or coincident.*
2. *It holds that*
$$\det \begin{bmatrix} w_1 & w_2 \\ v_1 & v_2 \end{bmatrix} = w_1v_2 - w_2v_1 = 0.$$
3. *There exists a non-zero constant $r \in \mathbb{R}$ such that $rv = w$.*

Proof. We will prove that

$$(1) \implies (2) \implies (3) \implies (1).$$

- $(1) \implies (2)$: First we will prove that, if the lines ℓ_1 and ℓ_2 are parallel or coincident, then $w_1v_2 = w_2v_1$ by proving the contrapositive: if $w_1v_2 - w_2v_1 \neq 0$, then ℓ_1 and ℓ_2 have a unique point of intersection, meaning we want there to exist a unique pair (t, s) of reals such that the follow equivalent equations hold:

$$\begin{aligned} p + tv &= q + sw \\ (p_1, p_2) + t(v_1, v_2) &= (q_1, q_2) + s(w_1, w_2) \\ t(v_1, v_2) - s(w_1, w_2) &= (q_1, q_2) - (p_1, p_2) \\ (tv_1 - sw_1, tv_2 - sw_2) &= (q_1 - p_1, q_2 - p_2) \\ \begin{bmatrix} v_1 & -w_1 \\ v_2 & -w_2 \end{bmatrix}\begin{bmatrix} t \\ s \end{bmatrix} &= \begin{bmatrix} q_1 - p_1 \\ q_2 - p_2 \end{bmatrix} \end{aligned}$$

  Under the assumption that

$$\det \begin{bmatrix} w_1 & w_2 \\ v_1 & v_2 \end{bmatrix} = w_1v_2 - w_2v_1 \neq 0,$$

the determinant of the matrix on the left satisfies

$$\det\begin{pmatrix} v_1 & -w_1 \\ v_2 & -w_2 \end{pmatrix} = v_1(-w_2) - (-w_1)v_2 = w_1v_2 - w_2v_1 \neq 0.$$

By Corollary 1.30, its inverse $\begin{pmatrix} v_1 & -w_1 \\ v_2 & -w_2 \end{pmatrix}^{-1}$ exists and we can multiply both sides of the equation by it to get the unique solution (t, s)

$$\begin{pmatrix} t \\ s \end{pmatrix} = \begin{pmatrix} v_1 & -w_1 \\ v_2 & -w_2 \end{pmatrix}^{-1} \begin{pmatrix} q_1 - p_1 \\ q_2 - p_2 \end{pmatrix}.$$

- (2) $\implies$ (3): Now we will show that, if $w_1v_2 - w_2v_1 = 0$, then there exists a real r such that $rv = w$. By the fact that neither v nor w is the origin,

$$v_1 = 0 \iff w_1 = 0,$$
$$v_2 = 0 \iff w_2 = 0,$$

and at most one of these two biconditional statements can hold. In the first case $r = \frac{w_2}{v_2}$ works, and in the second case $r = \frac{w_1}{v_1}$ works. If neither is true, then we may write

$$\frac{w_1}{v_1} = \frac{w_2}{v_2}.$$

Setting r equal to the common value, we get

$$rv = r(v_1, v_2) = (rv_1, rv_2) = (w_1, w_2) = w,$$

as desired. Note that r has to be non-zero, since w is non-zero.

- (3) $\implies$ (1): Finally, we will prove that if there exists a non-zero real r such that $rv = w$, then ℓ, ℓ' are parallel or coincident. Assuming such an r exists,

$$\begin{aligned} \ell_2 &= \{q + sw : s \in \mathbb{R}\} \\ &= \{q + srv : s \in \mathbb{R}\} \\ &= \{q + tv : t \in \mathbb{R}\}, \end{aligned}$$

since scaling the set of all reals by a non-zero constant r does not alter the set $\mathbb{R}$. Suppose ℓ_1, ℓ_2 are not parallel. We will prove that they are coincident. Since we are assuming they are not parallel, it means that there exists at least one point (x_0, y_0) of intersection. Then

$$(x_0, y_0) \in \ell_1 \implies \exists c_1 \in \mathbb{R} : (x_0, y_0) = p + c_1v,$$
$$(x_0, y_0) \in \ell_2 \implies \exists c_2 \in \mathbb{R} : (x_0, y_0) = q + c_2v.$$

Equating them, we get

$$\begin{aligned} p + c_1v &= q + c_2v \\ \implies p &= q + cv, \quad c = c_2 - c_1. \end{aligned}$$

Then

$$\begin{aligned}\ell_1 &= \{p + tv : t \in \mathbb{R}\}\\ &= \{q + cv + tv : t \in \mathbb{R}\}\\ &= \{q + (c+t)v : t \in \mathbb{R}\}\\ &= \{q + sv : s \in \mathbb{R}\} = \ell_2,\end{aligned}$$

since translating the set of all reals by a non-zero constant c does not alter the set $\mathbb{R}$. Therefore, ℓ_1, ℓ_2 are parallel or coincident.

□

Lemma 1.37 (Vector decomposition lemma). Let $\overrightarrow{AB}$ be a displacement vector in $\mathbb{R}^n$ and let $\vec{A}$ and $\vec{B}$ be position vectors (they represent $\overrightarrow{OA}$ and $\overrightarrow{OB}$, respectively, where O is the origin). Then

$$\overrightarrow{AB} \sim \vec{B} - \vec{A}.$$

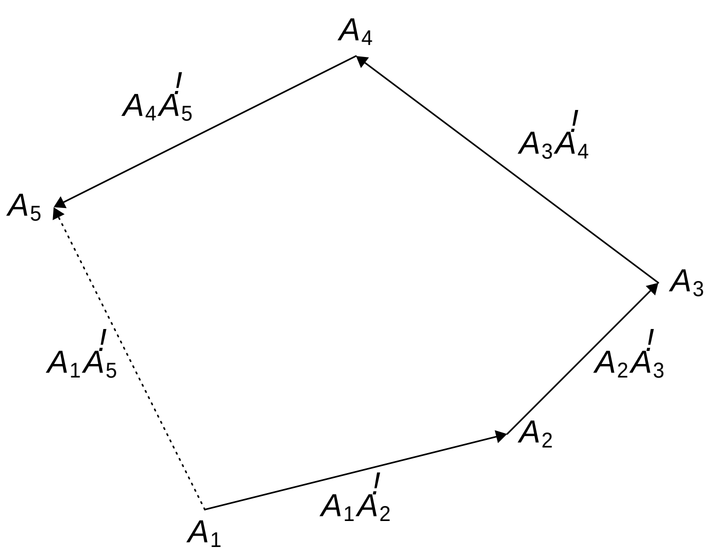
Subsequently,

$$\overrightarrow{AB} \sim -\overrightarrow{BA}.$$

A further pair of consequences is that if $A_1, A_2, \ldots, A_m$ are points in $\mathbb{R}^n$, then

$$\overrightarrow{A_1A_2} + \overrightarrow{A_2A_3} + \cdots + \overrightarrow{A_{m-1}A_m} \sim \overrightarrow{A_1A_m},$$
$$\overrightarrow{A_1A_2} + \overrightarrow{A_2A_3} + \cdots + \overrightarrow{A_{m-1}A_m} + \overrightarrow{A_mA_1} \sim 0.$$

Geometrically, this means that if we have finite sum of vectors, then we may place them in a sequence where the arrowhead of each vector is touching the tail of the next one (after translations under equipollence as necessary), and then the sum of the vectors is the same as the vector from the first tail to the last arrow.



Proof. Let $A = (a_1, a_2, \ldots, a_n)$ and $B = (b_1, b_2, \ldots, b_n)$. By the definition of vector addition and equipollence,

$$\begin{aligned}\vec{B} - \vec{A} &= (\underbrace{(0,0,\ldots,0)}_{n\text{-tuple of 0's}}, (b_1 - a_1, b_2 - a_2, \ldots, b_n - a_n)) \\ &\sim ((a_1, a_2, \ldots, a_n), (b_1, b_2, \ldots, b_n)) \\ &= (A, B) \\ &= \vec{AB}.\end{aligned}$$

This decomposition leads to

$$\vec{BA} = -(\vec{A} - \vec{B}) = \vec{B} - \vec{A} = \vec{AB}.$$

The remaining two corollaries are true by using the same decomposition and telescoping. □

Lemma 1.38. Let P_1, Q_1 and P_2, Q_2 be pairs of points such that $\vec{P_1Q_1} \sim \vec{P_2Q_2}$. Then the line

$$\ell_1 = \{\vec{P_1} + t\vec{P_1Q_1} : t \in \mathbb{R}\}$$

that runs through P_1, Q_1 is parallel to or coincident with the line

$$\ell_2 = \{\vec{P_2} + s\vec{P_2Q_2} : s \in \mathbb{R}\}$$

that runs through P_2, Q_2. Note that, since both lines are written in terms of vectors, our way of obtaining a concrete line (as in, a collection of points), is to interpret this notation as meaning the set of heads of the position representatives of all the included vectors.

Proof. Suppose the two lines are not parallel. We will show that the lines are equal as subsets of space. Since $\vec{P_1Q_1} \sim \vec{P_2Q_2}$, it means

$$P_1 - Q_1 = P_2 - Q_2.$$

In the case that $P_1 = P_2$, we would also get $Q_1 = Q_2$, which would make the lines coincide with ease. So suppose P_1 is not the same point as P_2. Under the initial assumption that the lines are not parallel, a point of intersection between the lines exist. So, there exist real numbers λ and μ such that

$$\begin{aligned}\vec{P_1} + \lambda\vec{P_1Q_1} &= \vec{P_2} + \mu\vec{P_2Q_2} \\ &= \vec{P_2} + \mu\vec{P_1Q_1} \\ \vec{P_2} - \vec{P_1} &= \lambda\vec{P_1Q_1} - \mu\vec{P_1Q_1} \\ \vec{P_1P_2} &= (\lambda - \mu)\vec{P_1Q_1} \sim (\lambda - \mu)\vec{P_2Q_2}.\end{aligned}$$

Note that $\quad \neq 0$, as, otherwise, P_1 would be the same point as P_2, which is a possibility that we have already excluded. This allows us to rewrite the lines as

$$\begin{aligned}
\ell_1 &= \{\vec{P_1} + t\vec{P_1Q_1} : t \in \mathbb{R}\} \\
&= \{\vec{P_1} + t(\quad)^{-1}\vec{P_1P_2} : t \in \mathbb{R}\} \\
&= \{\vec{P_1} + t\vec{P_1P_2} : t \in \mathbb{R}\} \\
&= \{\vec{P_1} + t(\vec{P_2} - \vec{P_1}) : t \in \mathbb{R}\} \\
&= \{(1-t)\vec{P_1} + t\vec{P_2} : t \in \mathbb{R}\}
\end{aligned}$$

and, similarly,

$$\begin{aligned}
\ell_2 &= \{\vec{P_2} + t\vec{P_2Q_2} : s \in \mathbb{R}\} \\
&= \{\vec{P_2} + s(\quad)^{-1}\vec{P_1P_2} : s \in \mathbb{R}\} \\
&= \{\vec{P_2} + s\vec{P_1P_2} : s \in \mathbb{R}\} \\
&= \{\vec{P_2} + s(\vec{P_2} - \vec{P_1}) : s \in \mathbb{R}\} \\
&= \{(-s)\vec{P_1} + (s+1)\vec{P_2} : s \in \mathbb{R}\}.
\end{aligned}$$

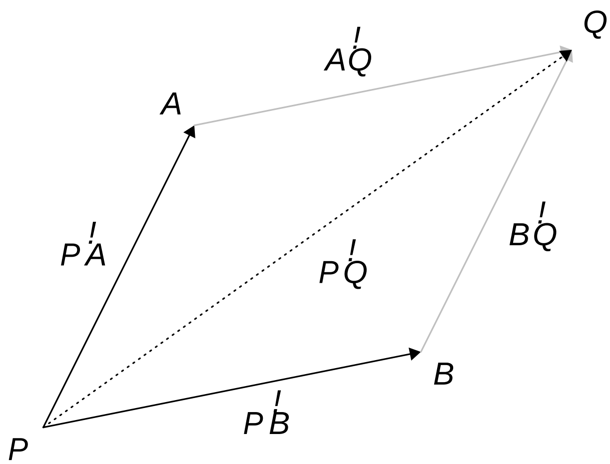
Note that the factor $(\quad)^{-1}$ was possible to make disappear because scaling the set of all real numbers t or s by a constant does not change the set $\mathbb{R}$. By the change of variables $t - s = 1$, we get $1 - t = -s$ and $t = s + 1$, so $\ell_1 = \ell_2$. □

Theorem 1.39 (Parallelogram law). *If P, A, B are distinct points in $\mathbb{R}^n$, then there is a representative of $\vec{PA} + \vec{PB}$ that is a directed diagonal, with its tail at P, of a parallelogram with adjacent sides PA and PB (a parallelogram is defined in Definition 7.19). Moreover, according to the vector decomposition lemma (Lemma 1.37), the other directed diagonal of the same parallelogram from A to B has a representative in*

$$\vec{AB} = \vec{PB} - \vec{PA}$$

and the directed diagonal from B to A has a representative in

$$\vec{BA} = -\vec{AB} = -(\vec{PB} - \vec{PA}) = \vec{PA} - \vec{PB}.$$



Proof. Let P, A, B be three distinct points in $\mathbb{R}^n$. Our two adjacent sides are representatives of the vectors

$$\overrightarrow{PA} = (P, A) \sim (0, A - P),$$
$$\overrightarrow{PB} = (P, B) \sim (0, B - P).$$

So their sum is

$$\overrightarrow{PA} + \overrightarrow{PB} \sim (0, A - P) + (0, B - P) = (0, A + B - 2P) \sim (P, A + B - P).$$

Let Q be the point $A + B - P$. We need to show that Q is the fourth vertex. Since we want a parallelogram forged by the adjacent sides PA and PB, it suffices to prove that the lines through $\overrightarrow{PA}$ and $\overrightarrow{BQ}$ are parallel or coincident, and likewise for $\overrightarrow{PB}$ and $\overrightarrow{AQ}$. Note that

$$\overrightarrow{PA} = (P, A) \sim (0, A - P) \sim (B, A + B - P) = (B, Q) = \overrightarrow{BQ},$$
$$\overrightarrow{PB} = (P, B) \sim (0, B - P) \sim (A, A + B - P) = (A, Q) = \overrightarrow{AQ}.$$

By Lemma 1.38, the lines through any two representatives of the same vector are parallel or coincident, so we are done. □

Chapter 2

Angles

> "You must not attempt this approach to parallels. I know this way to the very end. I have traversed this bottomless night, which extinguished all light and joy in my life. I entreat you, leave the science of parallels alone... Learn from my example."
>
> – *Farkas Bolyai*

We will describe angles, and their types and properties. This will allow us to prove some basic angle theorems about triangles. Afterwards, we will provide an exposition of parallel and perpendicular lines using slope-intercept form and the geometric interpretation of complex numbers. We conclude with an elementary proof of the formula for the perpendicular distance between a point and a line.

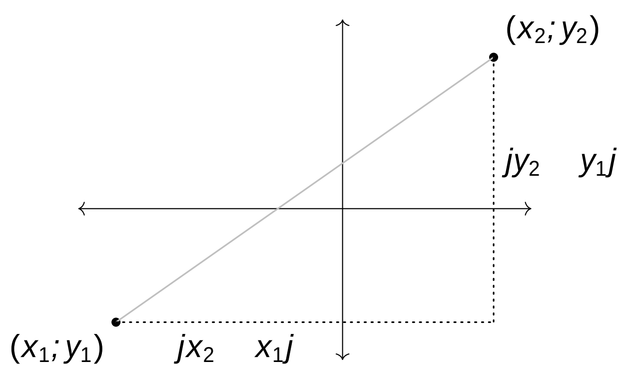
2.1 Basic Notions

Theorem 2.1 (Distance formula). The distance between two points (x_1, y_1) and (x_2, y_2) on the Cartesian plane is

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

Subsequently, this means the distance between two complex numbers z_1 and z_2 is $|z_1 - z_2|.$

Proof. If the two points are the same, then the distance between them is $0,$ which matches the formula. So we may assume that the two points are distinct. Then we can construct a right triangle with the segment between (x_1, y_1) and (x_2, y_2) as its hypotenuse and legs of length $|x_1 - x_2|$ and $|y_1 - y_2|$ that are parallel to the coordinate axes. This is done by drawing a line parallel to the x-axis through one point, and a line parallel to the y-axis through the other point.

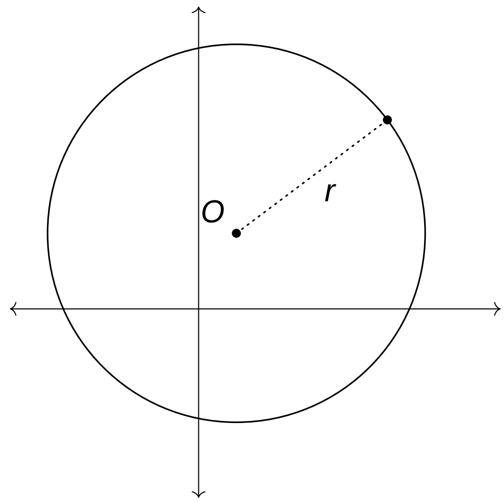


The formula then follows from the Pythagorean theorem. However, it is sensible to take this distance formula as a definition in analytic geometry, since we have not yet proven the Pythagorean theorem.
Regarding the relation to complex numbers, if we take $z_1 = (x_1; y_1)$ and $z_2 = (x_2; y_2)$ then the formula for the complex modulus yields

$$|z_1 - z_2| = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}.$$

□

Definition 2.2. A **circle** is the collection of all points in the plane at a fixed distance from a fixed point.



There are several ensuing concepts:

1. The fixed point is called the **center** of the circle.
2. The fixed distance is called the **radius** of the circle. A radius of a circle also refers to a line segment with one endpoint on the circle and one endpoint at the center of the circle. The plural of radius is radii. By the definition of a circle, all radii have the same length.
3. A **diameter** of a circle is a segment that goes through the center and has both endpoints on the circle, though the word can also refer to the length of such a segment.
4. The **circumference** of a circle is its perimeter. The term can also refer to the boundary of a closed disk (see Definition 5.1). We define the ratio of the circumference of a circle divided by its diameter to be the constant π, which has the same value for all circles.
5. The **unit circle** in the Cartesian plane is the circle of radius 1 and center at the origin.

Corollary 2.3. If a circle is the collection of all points at a fixed distance r from a fixed point $(a; b)$ in the Cartesian plane, then the equation of all points $(x; y)$ on the circle is given by

$$(x - a)^2 + (y - b)^2 = r^2.$$

Proof. By definition, (x, y) lies on the circle if and only if the distance from (x, y) to (a, b) is r. By the distance formula (Theorem 2.1), this means (x, y) lies on the circle if and only if

$$\sqrt{(x - a)^2 + (y - b)^2} = r,$$

which is equivalent to the proposed equation.

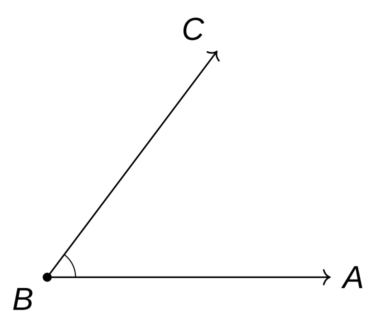
As a side note, the equation of a circle could be provided in the form

$$x^2 + Ax + y^2 + By + C = 0,$$

in which case we would have to complete the square in each variable (see the proof of the quadratic formula in Volume 1 for a general method of completing the square) in order to determine the radius and the coordinates of the center. Not all equations of this form lead to a circle, though. For example, upon completing the square in each variable, we might discover that the square of the "radius" is non-positive. Related comments for conics and bivariate quadratics are written under Definition 12.4. □

We need a way to measure how far we have rotated from one point to another points, relative to a central point that is equidistant from the origin and destination. Angles do this job.

Definition 2.4. Given two rays $\overrightarrow{BA}$ and $\overrightarrow{BC}$ with the common origin B, the angle between them refers to one of the two regions in between the rays, with the choice being made clearly at the time of writing. Whether the rays themselves are included depends on the context, but the interior of an angle is the same region excluding the rays. The common point of the rays is called the vertex of the angle and rays $\overrightarrow{BA}$ and $\overrightarrow{BC}$ are called the legs. The angle is denoted by $\angle ABC$.

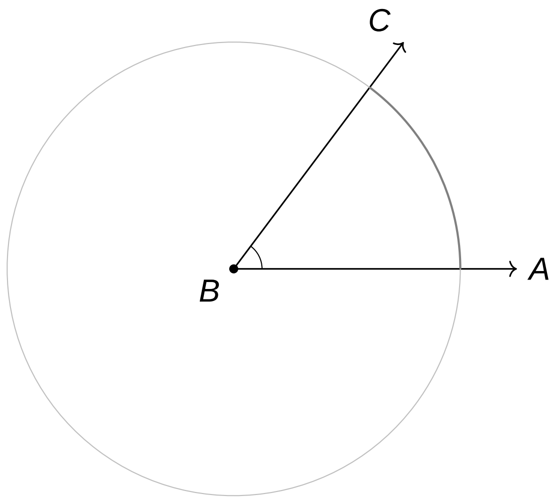


Note that we must clearly choose the angle from among the two regions that the plane is cut into by the two rays, otherwise the angle is not well-defined.

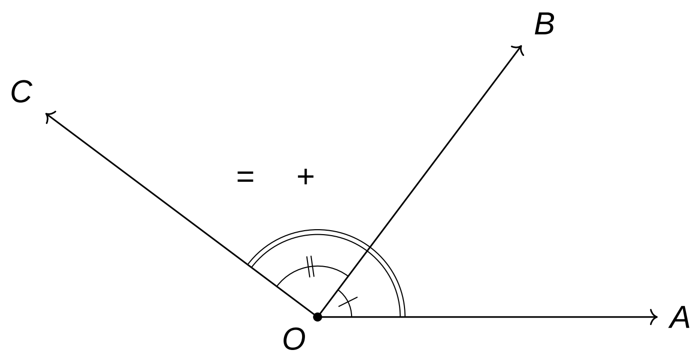
We will often conflate an angle, which is a region, with its measure, which is a non-negative real number. The measure of an angle is defined as follows.

Definition 2.5. Let $\angle ABC$ be an angle. After drawing a circle of radius 1 with the vertex of the angle as its center, the radian measure θ of the angle is the the arc length of the unit circle inside the angle. Note that π is defined to satisfy 2π being the circumference of the unit circle. Since radian measure can be uncomfortable for geometry, we can define degree measure to satisfy $2\pi = 360^\circ$. Then the degree measure φ of the same angle is given by

$$\frac{\varphi}{360} = \frac{\theta}{2\pi} \implies \varphi = \frac{180\theta}{\pi}.$$



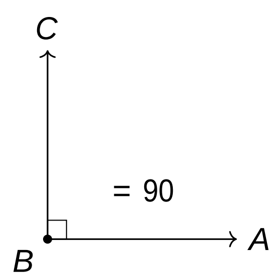
Theorem 2.6. Angles are additive in the sense that placing two angles beside each other with a common ray means the measure of the new angle is the sum of the measures of the original two angles.



Definition 2.7. There are several categories into which we will classify angles:

1. A **zero** angle represents no rotation and so has measure 0 :

2. A **right** angle is a fourth of a full rotation and so has measure 90 . One may use a square mark to indicate a right angle.



3. An **acute** angle is a positive angle that is smaller than a right angle.

4. A straight angle is half of a full rotation and so has measure 180 : The union of two rays that make a straight angle is a line. For example, a diameter of a circle splits into two radii that form two straight angles.

5. An obtuse angle is larger than a right angle but smaller than a straight angle.

6. A complete angle represents one full rotation and so has measure 360 :

7. A reflex angle is larger than a straight angle but smaller than a complete angle.

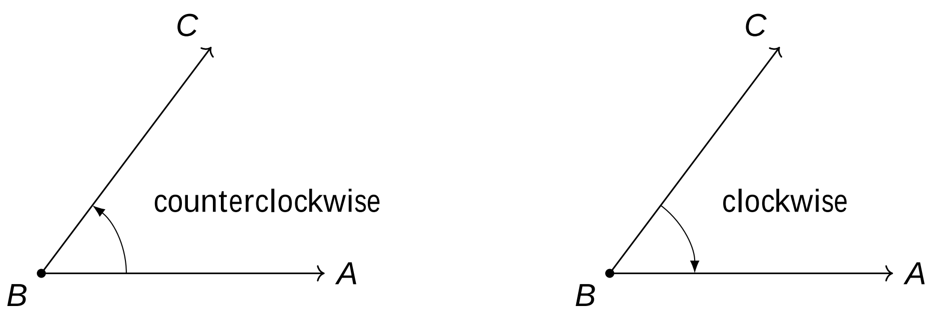
Definition 2.8. There are three important relationships between pairs of angles:

1. A pair of non-negative angles is called complementary if their measures sum to 90 .

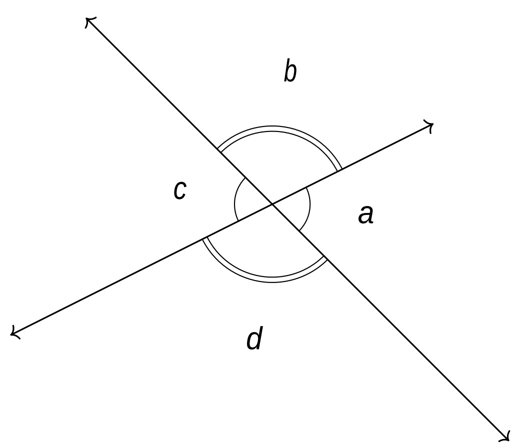
2. A pair of non-negative angles is called supplementary if their measures sum to 180 .

3. A pair of non-negative angles is called explementary if their measures sum to 360 .

Definition 2.9. We introduce negative angles by making counterclockwise angles the positive ones and clockwise angles the negative ones. In classical geometry, angles do not have such an orientation, but analytic or complex geometry can be imbued with oriented angles.



Definition 2.10. If two lines intersect, then they produces four angles. If the angles are $a; b; c; d$ in counterclockwise order, then $a; c$ are called vertical or opposite angles, as are $b; d$:



Theorem 2.11 (V-angle theorem)**.** If there are two intersecting lines, then the angles in each pair of vertical angles are equal.

Proof. Let the angles be a, b, c, d in clockwise order. Then

$$a + b = 180^\circ,$$
$$b + c = 180^\circ.$$

Equating them yields

$$a + b = b + c \implies a = c.$$

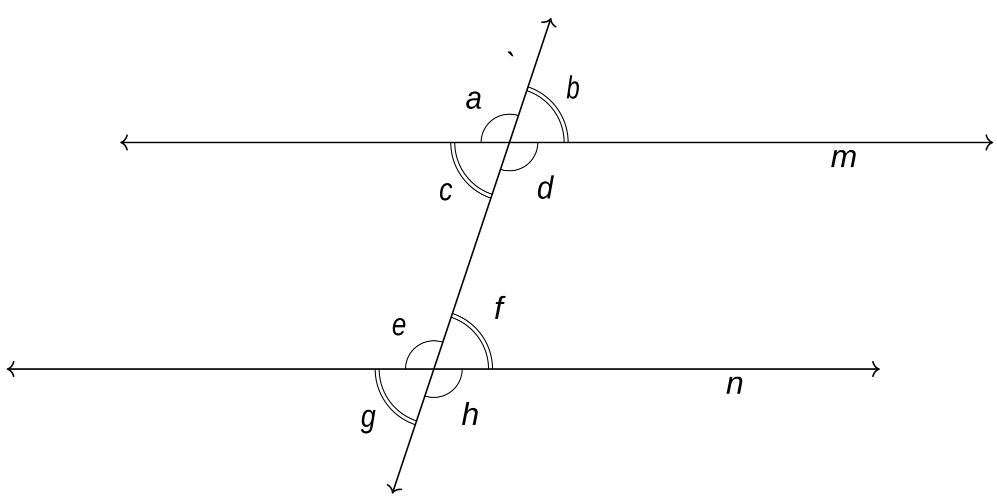
Symmetrically, $b = d$. □

Definition 2.12. Two lines are said to be **perpendicular** if the four angles created by their intersection are equal. Equivalently, the four angles all have measure $\frac{360^\circ}{4} = 90^\circ$. In fact, by vertical angles and supplementary angles we can show that if one angle of the intersection is 90°, so are the other three (try it!).

Definition 2.13. A **transversal** ℓ is a line that cuts through two lines m, n which are usually parallel for our purposes. This produces eight angles. Any two angles on the same side of the transversal are called **same-side** angles and any two angles on opposite sides of the transversal are called **alternate** angles. Angles in between the two parallel lines are called **interior** angles and the other four are called **exterior** angles. This terminology produces $2 \times 2 = 4$ types of pairs of angles in the figure:

- Alternate interior angles
- Alternate exterior angles
- Same-side interior angles
- Same-side exterior angles

Moreover, two same-side angles, one of which is an interior angle and one of which is an exterior angle, are called **corresponding** angles.



Theorem 2.14. Suppose there are two lines m, n and a transversal ℓ cutting through them. It is standard to assume that corresponding angles are equal if and only if m, n are parallel; this configuration is known as the F-angle theorem. Using supplementary angles, we get several other statements equivalent to m, n being parallel:

1. Alternate interior angles are equal (Z-angle theorem)
2. Alternate exterior angles are equal
3. Same-side interior angles are supplementary
4. Same-side exterior angles are supplementary

So a pair of parallel lines and a transversal creates two classes of four equal angles each, such that one angle from one class and one angle from another class are supplementary. In terms of the labels of the above diagram, the two classes are

$$a = d = e = h,$$
$$b = c = f = g.$$

Lemma 2.15. Let ℓ be a line, and m, n be two more lines.

1. If m, n are both perpendicular to $\ell,$ then m, n are parallel.
2. If m, ℓ are parallel and ℓ, n are parallel, then then m, n are parallel. This makes being parallel a transitive property.

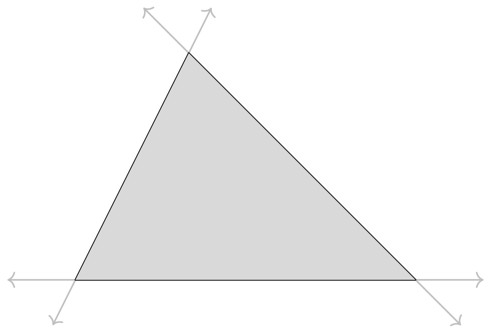
Proof. We will use the setup of two parallel lines and a transversal for both proofs:

1. Suppose m, n are both perpendicular to $\ell.$ Then eight right angles are produced at the two points of intersection. By the equality of corresponding angles, m, n must be parallel.
2. Suppose m, n are both parallel to $\ell.$ First we draw a line t through ℓ such that t and ℓ are perpendicular. Since ℓ, m are parallel, alternate interior angles tell us that t intersects m perpendicularly (there must be a point of intersection, otherwise m would be parallel to both ℓ, t which are perpendicular). Similarly, t intersects n perpendicularly. So t is a perpendicular to both of m, n and equal alternate interior angles tell us that m, n are parallel.

□

Definition 2.16. A **half-plane** is either side of a line and includes the line; this will be extended to 3D geometry in Definition 14.1. A **convex polygon** is a finite region formed by the intersection of half-planes, where we never include two half-planes whose union forms the whole plane. The **vertices** of the convex polygon are the corners where the boundaries of two such half-planes meet, and the **edges** or **sides** are the line segments between consecutive vertices. The **perimeter** of a convex polygon is the sum of the lengths of the edges. The **interior angles** of a convex polygon are the non-reflex angles created by pairs of edges at the vertices.

Example. A triangle (Definition 3.7) is a convex polygon, by taking the lines for the required half-planes to be the lines that run through the edges of the triangle.



Definition 2.17. Given a set of three or more points, if there exists a line that runs through all of the points, then the points are said to be **collinear**.

Theorem 2.18 (Interior angle theorem)**.** The sum of the interior angles of any triangle is 180°.

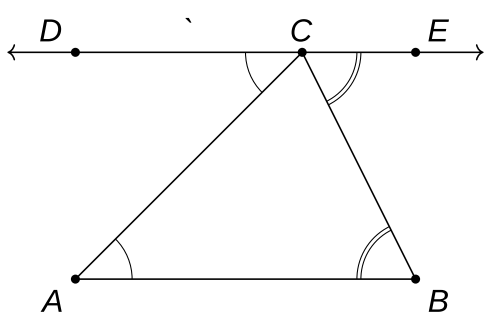
Proof. Let the triangle be $\triangle ABC$. First we draw a line ℓ through C that is parallel to the line through AB. Let D and E be points on ℓ, as shown. By alternate interior angles,

$$\angle BAC = \angle ACD,$$
$$\angle ABC = \angle BCE.$$

Then

$$\angle ABC + \angle BAC + \angle ACB = \angle BCE + \angle ACD + \angle ACB = 180^\circ,$$

due to D, C, E being collinear.



□

Corollary 2.19. If the interior angles of a triangle include a right or obtuse angle, then the other two angles are both acute.

Proof. Suppose a, b, c are the interior angles and that $a \geq 90^\circ$. Using that fact that

$$a + b + c = 180^\circ,$$

we find that

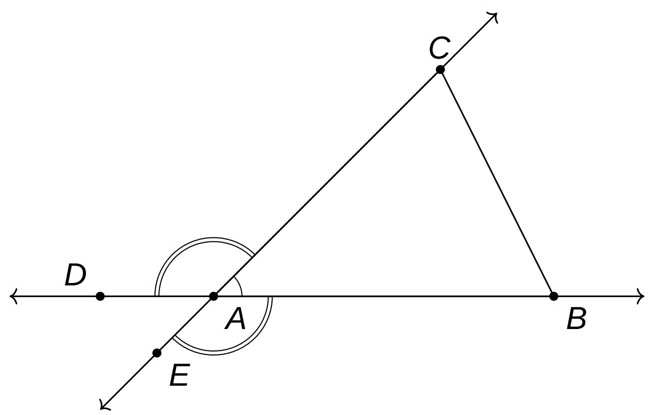
$$b + c = 180^\circ - a \leq 180^\circ - 90^\circ = 90^\circ.$$

Since b, c are both positive, they must both be acute. □

Definition 2.20. We can classify triangles according to the measures of their interior angles:

- **Acute triangle**: three acute interior angles
- **Right triangle**: one right interior angle, which causes the other two to be acute); the side opposite the right angle is called the **hypotenuse** and the other two sides are called **legs**
- **Obtuse triangle**: one obtuse interior angle, which causes the other two to be acute

Definition 2.21. The **exterior angle** corresponding to an interior angle of a convex polygon is the angle supplementary to the interior angle. Such an angle may be constructed by extending either of the two rays of the angle in the other direction, so there are two ways of doing this. By vertical angles, these two angles have the same measure, so the measure of an exterior angle corresponding to an interior angle is well-defined.



Corollary 2.22 (Triangle exterior angle theorem)**.** Let $4ABC$ be a triangle. Then the measure of the exterior angle at C is equal to the sum of the measures of the interior angles at A and B: As a consequence, the sum of the three exterior angles at $A; B; C$ is 360 . This generalized to convex polygons in Problem 5.29.

Proof. Let the measures of the interior angles at $A; B; C$ be ; ; ; and let the exterior angle at C measure ^{0}: We know that

$$+ \quad + \quad = 180 \; ;$$
$$+ \quad ^0 = 180 \; :$$

Equating the two yields

$$+ \quad + \quad = \quad + \quad ^0 =) \quad + \quad = \quad ^0;$$

which is the desired result, which is known as Euclid's first theorem. If the exterior angles at A and B are ^{0} and ^{0} respectively, then applying this result to them yields

$$^0 + \; ^0 + \; ^0 = (\; + \;) + (\; + \;) + (\; + \;)$$
$$= 2(\; + \; + \;)$$
$$= 2 \quad 180 \quad = 360 \; :$$

□

2.2 Parallel and Perpendicular

It is useful to be able to classify when two non-vertical lines are parallel or coincident according to the slopes of their slope-intercept equations. A criterion for perpendicular lines would also be helpful, for which we will need to invoke complex numbers, due to their friendliness with rotations.

Theorem 2.23. Suppose two lines are non-vertical.

1. The two lines are parallel or coincident if and only if their slopes are equal.

2. The two lines are perpendicular if and only if their slopes are negative reciprocals of each other.

Proof. Suppose there are two non-vertical lines. We prove the results in sequence, as the first part plays a role in the proof of the second part.

1. By Corollary 1.34, the negation of the proposition that two lines are parallel or coincident is that the two lines have exactly one intersection point. Instead of showing that if two lines are parallel or concident then their slopes are equal, we will show the contrapositive, which says that if two lines have unequal slopes then they have exactly one intersection point. Indeed, if the equations are

$$y = m_1 x + b_1,$$
$$y = m_2 x + b_2,$$

for $m_1 \neq m_2$, then solving for intersection points yields the single solution

$$(x_0, y_0) = \left(\frac{b_1 - b_2}{m_1 - m_2}, \frac{b_2 m_1 - b_1 m_2}{m_1 - m_2} \right).$$

In the other direction, we will show that if the lines have equal slopes, then they are parallel or concident. Suppose the lines have equal slope, so their equations are

$$y = mx + b_1,$$
$$y = mx + b_2.$$

Logically, the desired conclusion "they are parallel or concident" is equivalent to "if they are not parallel then they coincide." Since the negation of being parallel is that at least one point of intersection exists, it suffices to show that if one point of intersection exists then the lines coincide. Indeed, if (x_0, y_0) is a solution to both equations then setting the equations equal to each other yields

$$mx_1 + b_1 = y = mx_1 + b_2 \implies b_1 = b_2.$$

Thus, the equations are the same and represent the same line.

2. If two lines are perpendicular, then there must be a unique intersection point at which the angle is created. On the other hand, if the slopes are negative reciprocals, then the slopes cannot be equal and the previous part tells us that there exists a unique intersection point. So in either direction, we can assume the existence of a unique intersection point (x_0, y_0).

   Let the lines be ℓ_1 with slope m_1 and ℓ_2 with slope m_2, and let (x_1, y_1) be a point on ℓ_1 other than (x_0, y_0). Using the fact that multiplication by i causes a $\frac{\pi}{2}$ rotation counterclockwise around the origin, we can rotate $z_1 = x_1 + iy_1$ around $z_0 = x_0 + iy_0$ counterclockwise by $\frac{\pi}{2}$ radians to get the point

$$\begin{aligned} z_2 &= (z_1 - z_0)e^{\frac{\pi}{2}i} + z_0 \\ &= (x_1 + iy_1 - x_0 - iy_0)i + (x_0 + iy_0) \\ &= (ix_1 - y_1 - ix_0 + y_0) + (x_0 + iy_0) \\ &= (-y_1 + y_0 + x_0) + (x_1 - x_0 + y_0)i. \end{aligned}$$

   Let $z_2 = x_2 + iy_2$. Now we will tackle the two directions separately. If we assume that ℓ_1 and ℓ_2 are perpendicular, then we know that z_2 lies on ℓ_2, which tells us that

$$\begin{aligned} m_2 &= \frac{y_2 - y_0}{x_2 - x_0} \\ &= \frac{(x_1 - x_0 + y_0) - y_0}{(-y_1 + y_0 + x_0) - x_0} \\ &= \frac{x_1 - x_0}{-y_1 + y_0} = -\left(\frac{y_1 - y_0}{x_1 - x_0}\right)^{-1} = -\frac{1}{m_1}. \end{aligned}$$

   So the slopes are negative reciprocals. In the other direction, we suppose that $m_2 = -\frac{1}{m_1}$. Since z_1 was an arbitrary point on ℓ_1 other than z_0, it suffices to show that z_2 lies on ℓ_2. Indeed,

$$\begin{aligned} \frac{y_2 - y_0}{x_2 - x_0} &= \frac{(x_1 - x_0 + y_0) - y_0}{(-y_1 + y_0 + x_0) - x_0} \\ &= \frac{x_1 - x_0}{-y_1 + y_0} \\ &= -\left(\frac{y_1 - y_0}{x_1 - x_0}\right)^{-1} = -\frac{1}{m_1} = m_2, \\ y_1 - y_0 &= m_2(x_1 - x_0). \end{aligned}$$

   Thus, (x_2, y_2) lies on ℓ_2, since it satisfies the point-slope form of the equation of ℓ_2, which is

$$y - y_0 = m_2(x - x_0).$$

□

Now we will use the slope criteria in Theorem 2.23 to formulate similar criteria in terms of complex numbers. The new criteria will be universal in the sense that they will not require the lines to be non-vertical.

Theorem 2.24. Let a and b be complex numbers that are distinct from each other, and let c and d be complex numbers that are distinct from each other. Then:

1. The line through a, b is parallel to or coincident with the line through c, d if and only if $\frac{d-c}{b-a}$ is real. As a consequence, if a, b, c are distinct complex numbers, then they are collinear if and only if $\frac{c-a}{b-a}$ is real.

2. The line through a, b is perpendicular to the line through c, d if and only if $\frac{d-c}{b-a}$ is pure imaginary, meaning its real part is 0.

Proof. The propositions are easy to verify if either line is vertical, and we leave this to the reader. So we may assume that the lines are not vertical, which will allow us to use the slope criteria in Theorem 2.23. In order to calculate slopes in terms of the complex numbers, we will use the fact that, for any complex number z,

$$\begin{aligned} 2 \cdot \mathrm{Re}(z) &= z + \overline{z}, \\ 2i \cdot \mathrm{Im}(z) &= z - \overline{z}. \end{aligned}$$

1. The line through a, b is parallel to or coincident with with the line through c, d if and only if their slopes are equal. Using the Cartesian slope formula, we can take some reversible algebraic steps:

$$\begin{aligned} \frac{\mathrm{Im}(d) - \mathrm{Im}(c)}{\mathrm{Re}(d) - \mathrm{Re}(c)} &= \frac{\mathrm{Im}(b) - \mathrm{Im}(a)}{\mathrm{Re}(b) - \mathrm{Re}(a)} \\ \frac{\frac{1}{2i}(d - \overline{d}) - \frac{1}{2i}(c - \overline{c})}{\frac{1}{2}(d + \overline{d}) - \frac{1}{2}(c + \overline{c})} &= \frac{\frac{1}{2i}(b - \overline{b}) - \frac{1}{2i}(a - \overline{a})}{\frac{1}{2}(b + \overline{b}) - \frac{1}{2}(a + \overline{a})} \\ \frac{(d-c) - \overline{(d-c)}}{(d-c) + \overline{(d-c)}} &= \frac{(b-a) - \overline{(b-a)}}{(b-a) + \overline{(b-a)}} \\ (d-c)\overline{(b-a)} &= \overline{(d-c)}(b-a) \\ \frac{d-c}{b-a} &= \overline{\left(\frac{d-c}{b-a}\right)}. \end{aligned}$$

This is true if and only if $\frac{d-c}{b-a}$ is a real number.

If a, b, c are distinct complex numbers, then they are collinear if and only if the line through c, a is parallel to or coincides with the line through b, a (the loose criterion of being parallel or coincident implies that they will necessarily coincide due to the shared point a). By the criterion we just derived, this is true if and only if $\frac{c-a}{b-a}$ is a real number.

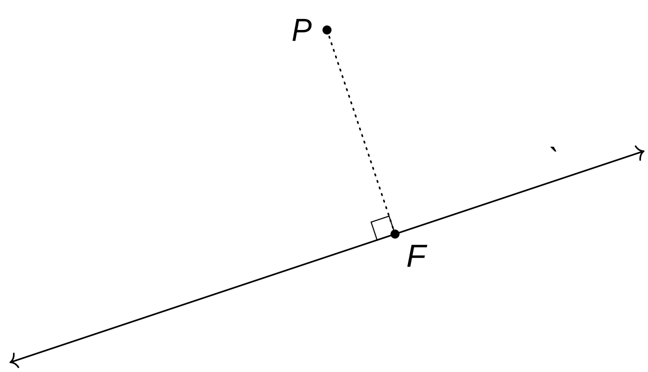
2. The line through a, b is perpendicular to the line through c, d if and only if their slopes are negative reciprocals. Using the Cartesian slope formula, we can take some reversible algebraic steps:

$$\begin{aligned}
\frac{\mathrm{Im}(d)-\mathrm{Im}(c)}{\mathrm{Re}(d)-\mathrm{Re}(c)} &= -\frac{\mathrm{Re}(b)-\mathrm{Re}(a)}{\mathrm{Im}(b)-\mathrm{Im}(a)} \\
\frac{\frac{1}{2i}(d-\overline{d})-\frac{1}{2i}(c-\overline{c})}{\frac{1}{2}(d+\overline{d})-\frac{1}{2}(c+\overline{c})} &= -\frac{\frac{1}{2}(b+\overline{b})-\frac{1}{2}(a+\overline{a})}{\frac{1}{2i}(b-\overline{b})-\frac{1}{2i}(a-\overline{a})} \\
\frac{(d-c)-\overline{(d-c)}}{(d-c)+\overline{(d-c)}} &= \frac{(b-a)+\overline{(b-a)}}{(b-a)-\overline{(b-a)}} \\
(d-c)\overline{(b-a)} &= -\overline{(d-c)}(b-a) \\
\frac{d-c}{b-a} &= -\overline{\left(\frac{d-c}{b-a}\right)}.
\end{aligned}$$

This is true if and only if $\frac{d-c}{b-a}$ is a pure imaginary number.

□

Definition 2.25. Given a line ℓ and a point $P = (x_0, y_0)$, it is clear that there exists a line m that is perpendicular to ℓ and runs through P. The reason is that, if ℓ is vertical, then m is horizontal, and otherwise m gets the slope that is the negative reciprocal of the slope of ℓ. The intersection of ℓ and m is called the **foot of the perpendicular** F from P to ℓ. The distance from P to F is called the **perpendicular distance** from P to ℓ.



Theorem 2.26 (Point-line distance formula)**.** Let ℓ be a line with equation $ax + by + c = 0$ and let P be a point in the plane with coordinates (x_0, y_0). Then the perpendicular distance from P to ℓ is

$$\frac{|ax_0 + by_0 + c|}{\sqrt{a^2 + b^2}}.$$

Proof. First we will deal with a few edge cases. If P lies on the ℓ, then $ax_0 + by_0 + c = 0$, which makes the given expression be equal to 0, as expected.
So now we may assume that P does not lie on ℓ. It is not possible for both a and b to be 0 because then ℓ would not be a line. If $a = 0$ and $b \neq 0$, then $y = -\frac{c}{b}$ and the line is

horizontal. The distance from P to ℓ is

$$\left|y_0 - \left(-\frac{c}{b}\right)\right| = \frac{|by_0 + c|}{|b|},$$

which agrees with the desired formula. Alternatively, if $a \neq 0$ and $b = 0,$ then $x = -\frac{c}{a}$ and the line is vertical. The distance from P to ℓ is

$$\left|x_0 - \left(-\frac{c}{a}\right)\right| = \frac{|ax_0 + c|}{|a|},$$

which again agrees with the desired formula. Now we may safely assume that neither a nor b is 0.

Let the foot of the perpendicular segment from (x_0, y_0) to ℓ be $(x_1, y_1).$ By the distance formula, we want to find an expression for

$$\sqrt{(x_0 - x_1)^2 + (y_0 - y_1)^2}$$

in terms of a, b, c and $x_0, y_0.$ The slope of ℓ is $-\frac{a}{b},$ so the slope of the perpendicular segment is its negative reciprocal, which is $\frac{b}{a}.$ Then $\frac{y_0 - y_1}{x_0 - x_1} = \frac{b}{a}$ or

$$a(y_0 - y_1) - b(x_0 - x_1) = 0.$$

Since we want to work with $(x_0 - x_1)^2$ and $(y_0 - y_1)^2,$ we square the equation and rearrange the result to get

$$a^2(y_0 - y_1)^2 + b^2(x_0 - x_1)^2 = 2ab(y_0 - y_1)(x_0 - x_1).$$

Inspired by the desired formula, we decide to work with the expression $(ax_0 + by_0 + c)^2,$ which miraculously yields

$$\begin{aligned}
(ax_0 + by_0 + c)^2 &= (ax_0 + by_0 - ax_1 - by_1)^2 \\
&= (a(x_0 - x_1) + b(y_0 - y_1))^2 \\
&= a^2(x_0 - x_1)^2 + b^2(y_0 - y_1)^2 + 2ab(x_0 - x_1)(y_0 - y_1) \\
&= a^2(x_0 - x_1)^2 + b^2(y_0 - y_1)^2 + a^2(y_0 - y_1)^2 + b^2(x_0 - x_1)^2 \\
&= (a^2 + b^2)((x_0 - x_1)^2 + (y_0 - y_1)^2).
\end{aligned}$$

This leads to

$$\sqrt{(x_0 - x_1)^2 + (y_0 - y_1)^2} = \frac{|ax_0 + by_0 + c|}{\sqrt{a^2 + b^2}}.$$

The point-line perpendicular distance formula will be proven more cleanly (and without reverse-engineering the expected formula) using vectors in Theorem 4.17. □

Problem 2.27. Let $f \in \mathbb{C}$ be the foot of the perpendicular from $z \in \mathbb{C}$ to the line through $a, b \in \mathbb{C}.$ Then

$$f = \frac{z(\overline{a} - \overline{b}) + \overline{z}(a - b) + \overline{a}b - a\overline{b}}{2(\overline{a} - \overline{b})}.$$

This formula works even if z lies on the line through a and $b.$

Chapter 3

Transformations

> "And that square root of minus one means that nature works with complex numbers and not with real numbers."
>
> – *Freeman Dyson, Birds and Frogs*

Complex numbers are essentially Cartesian coordinates with an extra multiplicative structure, which we will see has a geometric interpretation that involves rotation. This is what allows us, in some scenarios, to do two-dimensional geometry more conveniently using complex numbers than with other tools. First, we will describe all standard transformations using complex numbers. Secondly, we will look at preservation and alteration effects of these transformations. As a fruit of all this labour, we will use transformations to prove the cosine law and to prove the triangle inequality (for actual triangles). The triangle inequality will allow us to obtain "betweenness" results.

3.1 Formulas for Effects

Definition 3.1. There are four major transformations in plane geometry, which are defined in terms of complex numbers as follows. Each of these maps can be expressed in terms of Cartesian coordinates, but we have preferred to use complex numbers as it allows for more concise notation.

1. The **translation** by $w \in \mathbb{C}$ is the function $t : \mathbb{C} \to \mathbb{C}$, defined as
$$t(z) = z + w.$$

2. The **rotation** around $w \in \mathbb{C}$ by θ radians is the function $r : \mathbb{C} \to \mathbb{C}$, defined as
$$r(z) = (z - w)e^{i\theta} + w.$$
The rotation is counterclockwise for positive θ, and the rotation is clockwise for negative θ.

3. The **reflection** across the line $ax + by + c = 0$ is the function $f : \mathbb{C} \to \mathbb{C}$, defined as
$$f(z) = \frac{2ci + (b - ai)\overline{z}}{b + ai}.$$
If this function seems unnatural, see its natural decomposition using conjugation in Theorem 3.3.

4. The homothety from $w \in \mathbb{C}$ by a non-zero factor of $k \in \mathbb{R}$ is the function $h : \mathbb{C} \to \mathbb{C},$ defined as
$$h(z) = (z - w)k + w.$$
This is also called a dilation. If $k > 0$ we call it a positive homothety or dilation, and if $k < 0$ we call it a negative homothety or dilation. Either way, k is the dilation factor.

Translations, rotations, reflections, and compositions of any or all of them are called Euclidean isometries. Compositions of any or all of the four maps are called similarity transformations.

Example. A homothety from w by a factor of -1 is equivalent to a rotation around w by π radians because $e^{i\pi} = -1.$ This transformation is also called a point reflection across w. Keep in mind that there is a difference between a reflection across a line and a reflection across a point.

Theorem 3.2. Translations, rotations, reflections and homotheties are bijective from $\mathbb{C}$ to $\mathbb{C},$ and the inverse map of any function of each type is a transformation of the same type.

Proof. Recall from Volume 1 that a function is bijective if and only if it has an inverse. Using this fact, the existence of an inverse implies bijectivity, so it suffices to prove that any function of each type has an inverse of the same type. In the notation of Definition 3.1, it can be readily verified that the functions

$$\begin{aligned} t^{-1}(z) &= z - w, \\ r^{-1}(z) &= (z - w)e^{i(-\theta)} + w, \\ f^{-1}(z) &= \frac{2ci + (b - ai)\overline{z}}{b + ai}, \\ h^{-1}(z) &= (z - w)\frac{1}{k} + w. \end{aligned}$$

are the respective inverses. In particular, note that every reflection is its own inverse, which makes each reflection an involution. □

Theorem 3.3. The transformations in Definition 3.1 may be decomposed as follows:

1. Every rotation is a translation, followed by a rotation around the origin, followed by the inverse translation.

2. Every homothety by a factor of k is a translation, followed by a homothety by a factor of k from the origin, followed by the inverse translation.

3. A conjugation is defined as a reflection across the x-axis, which matches the definition of the conjugate of a complex number. Then every reflection is a translation, followed by a rotation, followed by a conjugation, and then the inverse rotation, and finally the inverse translation.

Since Euclidean isometries are compositions of translations, rotations, and reflections, this result implies that Euclidean isometries are compositions of translations, rotations around the origin, and conjugations. In the same way, similarity transformations are compositions of translations, rotations around the origin, conjugations, and homotheties from the origin.

Proof. The first two properties are immediately found to be true using the definitions of rotation and homothety, so we will only prove the last one. Let the line of reflection be $ax + by + c = 0$ and let the point being reflected be the complex number $z.$ The formula

$$\frac{2ci + (b - ai)\overline{z}}{b + ai} = -\frac{2ci}{b - ai} + \frac{b - ai}{b + ai}\,\overline{z},$$

tells us that we apply a conjugation, followed by multiplying by $\frac{b-ai}{b+ai}$ which is a rotation because its modulus is $1,$ followed by adding $-\frac{2ci}{b-ai}$ which is a translation. It technically satisfies the description of the sought decomposition, but it is not quite what we want morally. It is also not entirely satisfying because the formula for reflection across a line is unnatural and unjustified compared to those for translation, rotation, and homothety. We will intuitively derive the formula, and we will prove the desired result along the way.

The x-intercept of the line is at $y = 0$ which gives $x = -\frac{c}{a}.$ So we subtract $-\frac{c}{a}$ from the x-coordinates of the line and the point to produce the line $ax + by = 0,$ which goes through the origin, and get the point $z + \frac{c}{a}.$ It is not possible for both a and b to be 0 because then $ax+by+c = 0$ would not be a line, so $(b, -a)$ is a non-origin point on the new line. Applying a clockwise rotation by $\arg(b - ai)$ around the origin sends the line to the x-axis and $z + \frac{c}{a}$ to $\frac{z+\frac{c}{a}}{e^{i\arg(b-ai)}}.$ Now we apply conjugation as it is equivalent to reflection across the x-axis, and apply the inverse rotation and then the inverse translation to get the final reflected point

$$\begin{aligned}
e^{i\arg(b-ai)}\,\overline{\left(\frac{z+\frac{c}{a}}{e^{i\arg(b-ai)}}\right)} - \frac{c}{a} &= \frac{e^{i\arg(b-ai)}}{e^{i(-\arg(b-ai))}}\left(\overline{z} + \frac{c}{a}\right) - \frac{c}{a}\\
&= \frac{e^{i\arg(b-ai)}}{e^{i\arg(b+ai)}}\left(\overline{z} + \frac{c}{a}\right) - \frac{c}{a}\\
&= \frac{|b-ai|\,e^{i\arg(b-ai)}}{|b+ai|\,e^{i\arg(b+ai)}}\left(\overline{z} + \frac{c}{a}\right) - \frac{c}{a}\\
&= \frac{b-ai}{b+ai}\left(\overline{z} + \frac{c}{a}\right) - \frac{c}{a},
\end{aligned}$$

which is equivalent to the reflection formula. The top-left expression in the above array of expressions shows that a reflection is a conjugation that is nested inside a pair of inverse rotations and, outside that, a pair of inverse translations. □

Problem 3.4. If $f : X \to Y$ is a bijective function where X and Y are subsets of $\mathbb{R},$ show that the graph of the inverse f^{-1} is the reflection of the graph of f across the line $x = y$.

Lemma 3.5. Let $k : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be a function.

1. If the translation $(x, y) \mapsto (x+u, y+v)$ is applied to the graph of $k(x, y) = 0,$ then the resulting set is the graph of $k(x-u, y-v) = 0.$
2. If the counterclockwise rotation $(x, y) \mapsto (x\cos\theta - y\sin\theta, x\sin\theta + y\cos\theta)$ by θ around the origin is applied to the graph of $k(x, y) = 0,$ then the resulting set is the graph of
$$k(x\cos\theta + y\sin\theta, y\cos\theta - x\sin\theta) = 0.$$
3. If the conjugation $(x, y) \mapsto (x, -y)$ is applied to the graph of $k(x, y) = 0,$ then the resulting set is the graph of $k(x, -y).$

Proof. We will prove the result about rotation and leave the proofs for translation and conjugation to the reader. The latter two can be done using the same technique as the former and are easier, so they should be doable. As a preliminary note, we will justify the assumption in the assertion that the map

$$(x, y) \mapsto (x\cos\theta - y\sin\theta, x\sin\theta + y\cos\theta)$$

applies a counterclockwise rotation to (x, y) by θ around the origin. We can do this using complex numbers by letting $z = x + iy$ and expanding its product with $e^{i\theta} = \cos\theta + i\sin\theta$:

$$\begin{aligned} z \cdot e^{i\theta} &= (x+iy)(\cos\theta + i\sin\theta) \\ &= x\cos\theta + ix\sin\theta + iy\cos\theta + i^2 y\sin\theta \\ &= x\cos\theta + ix\sin\theta + iy\cos\theta - y\sin\theta \\ &= (x\cos\theta - y\sin\theta) + +i(x\sin\theta + y\cos\theta). \end{aligned}$$

So, we are looking at the set

$$S = \left\{(x\cos\theta - y\sin\theta, x\sin\theta + y\cos\theta) \in \mathbb{R}^2 : k(x, y) = 0\right\}.$$

To cause a change of variables, we let

$$\begin{aligned} x' &= x\cos\theta - y\sin\theta, \\ y' &= x\sin\theta + y\cos\theta. \end{aligned}$$

By multiplying these equations by $\cos\theta$ and $\sin\theta$ and then using elimination, along with the Pythagorean identity, it can be shown that

$$\begin{aligned} x &= x'\cos\theta + y'\sin\theta, \\ y &= y'\cos\theta - x'\sin\theta. \end{aligned}$$

The reader is encouraged to write out the details. By a similar process, this step can be reversed. Thus, we can apply a change of variables to show that

$$S = \left\{(x', y') \in \mathbb{R}^2 : k(x'\cos\theta + y'\sin\theta, y'\cos\theta - x'\sin\theta)\right\}.$$

Finally, we can replace the symbols x' and y' with x and y, respectively, in this definition of S, as the choice of symbols is irrelevant. □

3.2 Preservation of Properties

Theorem 3.6. Translations, rotations, reflections and homotheties have certain effects on the lengths of line segments and on the measures and orientations of counterclockwise angles:

1. A translation t maps a line segment from z_1 to z_2 to a line segment from $t(z_1)$ to $t(z_2)$ of the same length, and preserves the measure and orientation of counterclockwise angles in $[0, 2\pi).$

2. A rotation r maps a line segment from z_1 to z_2 to a line segment from $r(z_1)$ to $r(z_2)$ of the same length, and preserves the measure and orientation of counterclockwise angles in $[0, 2\pi).$

3. A reflection f across a line maps a line segment from z_1 to z_2 to a line segment from $f(z_1)$ to $f(z_2)$ of the same length, and maps a counterclockwise angle in $[0, 2\pi)$ to a clockwise angle with the same measure.

4. A homothety h by a factor of k maps a line segment of length l from z_1 to z_2 to a line segment of length $|k| \cdot l$ from $h(z_1)$ to $h(z_2),$ and preserves the measure and orientation of counterclockwise angles $[0, 2\pi).$

Proof. Certainly, the distance between two same points is $0,$ which is the same as the distance between the two same points to which they are mapped by a function. So we can assume that z_1 and z_2 are distinct. By Definition 1.7, the line segment between two complex numbers z_1 and z_2 is given by

$$\ell(x) = (1 - x)z_1 + xz_2$$

for $x \in [0, 1].$ When handling angles, there will be a third point z_3 distinct from z_2 so that we can measure the counterclockwise rotation of z_1 around z_2 that, along with a positive dilation from $z_2,$ causes z_1 to coincide with $z_3.$ Let $\arg\left(\frac{z_3 - z_2}{z_1 - z_2}\right) = \theta$ and let $\left|\frac{z_3 - z_2}{z_1 - z_2}\right| = s.$

1. Let $t(z) = z + w$ be a translation. Then

$$\begin{aligned} t((1 - x)z_1 + xz_2) &= (1 - x)z_1 + xz_2 + w \\ &= (1 - x)(z_1 + w) + x(z_2 + w) \\ &= (1 - x) \cdot t(z_1) + x \cdot t(z_2). \end{aligned}$$

   So the line segment from z_1 to z_2 gets mapped to the line segment from $t(z_1)$ to $t(z_2).$ The length of the new segment is

$$|t(z_2) - t(z_1)| = |(z_2 + w) - (z_1 + w)| = |z_2 - z_1|,$$

   so distances are preserved. Counterclockwise angles in $[0, 2\pi)$ are preserved as well because

$$\frac{t(z_3) - t(z_2)}{t(z_1) - t(z_2)} = \frac{z_3 + w - z_2 - w}{z_1 + w - z_2 - w} = \frac{z_3 - z_2}{z_1 - z_2} = se^{i\theta}.$$

2. In Theorem 3.3, we found that a rotation is a translation, followed by a rotation around the origin, followed by the inverse translation. Since a translation has the preservation properties previously proven, it suffices to prove the results for only rotations around the origin. Let $r(z) = ze^{i\theta}$ be a rotation around the origin. Then

$$\begin{aligned} r((1-x)z_1 + xz_2) &= ((1-x)z_1 + xz_2)e^{i\theta} \\ &= (1-x)z_1e^{i\theta} + xz_2e^{i\theta} \\ &= (1-x)\cdot r(z_1) + x\cdot r(z_2). \end{aligned}$$

So the line segment from z_1 to z_2 gets mapped to the line segment from $r(z_1)$ to $r(z_2)$. The length of the new segment is

$$|r(z_2) - r(z_1)| = |z_2e^{i\theta} - z_1e^{i\theta}| = |z_2 - z_1|\,|e^{i\theta}| = |z_2 - z_1|,$$

so distances are preserved. Counterclockwise angles in $[0, 2\pi)$ are preserved as well because

$$\frac{r(z_3) - r(z_2)}{r(z_1) - r(z_2)} = \frac{z_3e^{i\theta} - z_2e^{i\theta}}{z_1e^{i\theta} - z_2e^{i\theta}} = \frac{z_3 - z_2}{z_1 - z_2} = se^{i\theta}.$$

3. According to Theorem 3.3, a reflection is a conjugation enveloped by a rotation and its inverse, followed by being sandwiched between a translation and its inverse. Since rotations and translations have the preservation properties proven in the previous parts, it suffices to prove the result for only conjugation. Indeed, using properties of complex conjugates from Volume 1,

$$\overline{(1-x)z_1 + xz_2} = (1-x)\cdot\overline{z_1} + x\cdot\overline{z_2}.$$

So the line segment from z_1 to z_2 gets mapped to the line segment from $\overline{z_1}$ to $\overline{z_2}$. The length of the new segment is

$$|\overline{z_2} - \overline{z_1}| = |\overline{z_2 - z_1}| = |z_2 - z_1|,$$

so distances are preserved. Each counterclockwise angle in $[0, 2\pi)$ is turned into the explementary counterclockwise angle because

$$\frac{\overline{z_3} - \overline{z_2}}{\overline{z_1} - \overline{z_2}} = \overline{\left(\frac{z_3 - z_2}{z_1 - z_2}\right)} = \overline{se^{i\theta}} = se^{i(2\pi-\theta)},$$

which is a clockwise rotation with the same absolute measure.

4. By Theorem 3.3. a homothety is a translation, followed by a homothety around the origin, followed by the inverse translation. Since a translation has the preservation properties that were previously proven, it suffices to work with only homotheties from the origin. Let $h(z) = zk$ be a homothety from the origin by a factor of k. Then

$$\begin{aligned} h((1-x)z_1 + xz_2) &= ((1-x)z_1 + xz_2)k \\ &= (1-x)z_1k + xz_2k \\ &= (1-x)\cdot h(z_1) + x\cdot h(z_2). \end{aligned}$$

So the line segment from z_1 to z_2 gets mapped to the line segment from $h(z_1)$ to $h(z_2)$. The length of the new segment is

$$|h(z_2) - h(z_1)| = |z_2 k - z_1 k| = |k|\,|z_2 - z_1|,$$

so lengths are multiplied by a factor of $|k|$. Counterclockwise angles in $[0, 2\pi)$ are preserved because

$$\frac{h(z_3) - h(z_2)}{h(z_1) - h(z_2)} = \frac{z_3 k - z_2 k}{z_1 k - z_2 k} = \frac{z_3 - z_2}{z_1 - z_2} = se^{i\theta}.$$

□

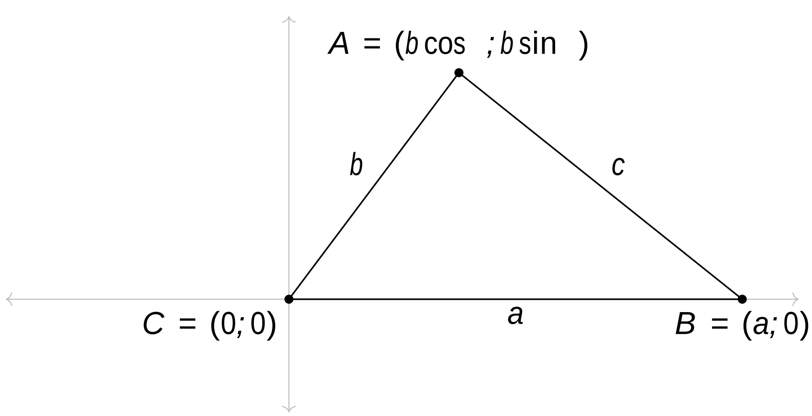
Definition 3.7. A **triangle** is produced by pairwise connecting three non-collinear distinct points, called **vertices**, with three line segments, called **edges**. A **degenerate triangle** is created by the three segments between three points that are collinear. For our purposes, a degenerate triangle is technically not a triangle, though we might clarify occasionally that a particular triangle is non-degenerate.

Theorem 3.8 (Law of cosines)**.** If $\triangle ABC$ has sides a, b, c opposite to vertices A, B, C, respectively, and the interior angle at C measures θ, then

$$c^2 = a^2 + b^2 - 2ab\cos\theta.$$

This contains the Pythagorean theorem as a special case (Theorem 9.12).

Proof. We will use Cartesian coordinates. We translate the triangle so that C lies at the origin, then rotate the triangle around the origin until B lies on the positive x-axis, and then (if necessary) reflect the triangle across the x-axis so that the y-coordinate of A is positive. According to Theorem 3.6, all side lengths and absolute measures of interior angles of the triangle are preserved. This setup provides a concrete configuration on which we can perform computations.



We already know that the coordinates of C are $(0, 0)$. Since $CB = a$ and B lies on the positive x-axis, the coordinates of B are $(a, 0)$. Finally, $CA = b$ and $\angle BCA = \theta$, so the polar coordinates of A are (b, θ), which can be converted to the Cartesian coordinates $(b\cos\theta, b\sin\theta)$.

This technique of placing points on the origin or axes is a common simplifying reduction in analytic geometry.

By applying the distance formula to A and B, we derive

$$\begin{aligned} c &= \sqrt{(a - b\cos\theta)^2 + (0 - b\sin\theta)^2}, \\ c^2 &= (a - b\cos\theta)^2 + (0 - b\sin\theta)^2 \\ &= a^2 + b^2(\cos^2\theta + \sin^2\theta) - 2ab\cos\theta \\ &= a^2 + b^2 - 2ab\cos\theta. \end{aligned}$$

Note that the Pythagorean theorem is the special case of $\theta = 90^\circ$ since $\cos 90^\circ = 0$ annihilates the $2ab\cos\theta$ term, leaving

$$a^2 + b^2 = c^2.$$

□

Corollary 3.9. In $\triangle ABC$, let the sides opposite to A, B, C be a, b, c respectively, and let the interior angle at C measure θ. Then

$$\operatorname{sgn}(a^2 + b^2 - c^2) = \begin{cases} 1 & \text{if } \theta \text{ is acute} \\ 0 & \text{if } \theta \text{ is right} \\ -1 & \text{if } \theta \text{ is obtuse} \end{cases}.$$

Proof. By the cosine law, $a^2 + b^2 - c^2 = 2ab\cos\theta$. Since $2ab$ is positive,

$$\operatorname{sgn}(a^2 + b^2 - c^2) = \operatorname{sgn}(\cos\theta).$$

By observing the unit circle, we can see that $\cos\theta$ has the signs prescribed in each case.

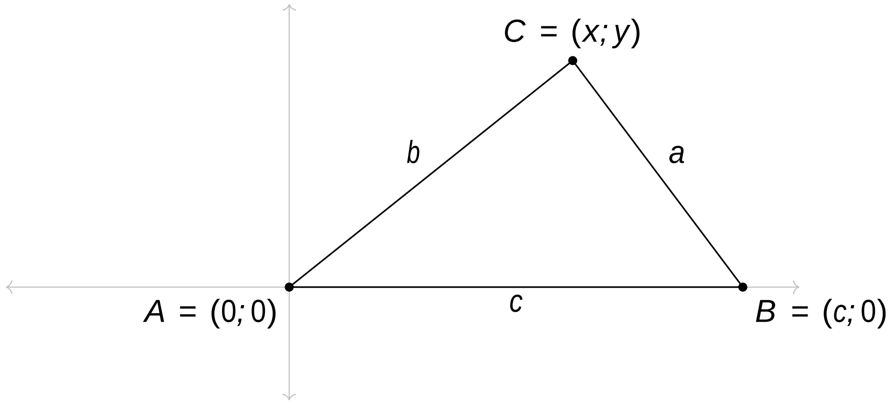
As the reader may have noticed, this contains the Pythagorean theorem as a special case. □

Theorem 3.10 (Triangle inequality). If three positive real numbers a, b, c satisfy the triangle inequalities

$$\begin{aligned} a + b &> c, \\ b + c &> a, \\ c + a &> b, \end{aligned}$$

then they are the sides of a non-degenerate triangle. Conversely (or rather, the contrapositive of the converse), if at least one of the inequalities is broken, then a degenerate triangle is produced. In this case, one of these strict inequalities is replaced by the corresponding equality and the other two inequalities remain strict.

Proof. Let the triangle be $\triangle ABC$ with sides $a = BC, b = CA, c = AB$. Suppose the three triangle inequalities hold. We wish to show that the three vertices are non-collinear, which will prove that $\triangle ABC$ is non-degenerate. By utilizing the same Euclidean isometries as those described in the proof of the cosine law (Theorem 3.8), we let $A = (0, 0)$ be the origin, $B = (c, 0)$ be on the positive x-axis (so, $c > 0$), and $C = (x, y)$ with $y \geq 0$. It suffices to prove that $y > 0$ so that is does not lie on the x-axis, which is the line that runs through AB.



We first aim to find explicit coordinates for $C = (x, y)$. By the distance formula, we get the two equations

$$\begin{aligned} x^2 + y^2 &= b^2, \\ (x - c)^2 + y^2 &= a^2. \end{aligned}$$

Subtracting the bottom from the top and rearranging yields

$$\begin{aligned} 2cx - c^2 &= b^2 - a^2 \\ x &= \frac{b^2 + c^2 - a^2}{2c}. \end{aligned}$$

Since $y \geq 0$, we can isolate it as

$$\begin{aligned} y &= \sqrt{b^2 - x^2} \\ &= \sqrt{(b - x)(b + x)} \\ &= \sqrt{\left(b - \frac{b^2 + c^2 - a^2}{2c}\right)\left(b + \frac{b^2 + c^2 - a^2}{2c}\right)} \\ &= \sqrt{\left(\frac{c^2 - (a - b)^2}{2a}\right)\left(\frac{(a + b)^2 - c^2}{2a}\right)} \\ &= \frac{1}{2a}\sqrt{(b + c - a)(c + a - b)(a + b - c)(a + b + c)}. \end{aligned}$$

The argument inside the square root exists and is positive due to the triangle inequalities, so $y > 0$. Therefore, $\triangle ABC$ exists as a non-degenerate triangle, as desired.

Conversely, suppose at least one of the following "broken" triangle inequalities is true:

$$\begin{aligned} a + b &\leq c, \\ b + c &\leq a, \\ c + a &\leq b. \end{aligned}$$

Without loss of generality, pick $c + a \leq b$. By the triangle inequality for Euclidean vectors (Corollary 4.9),

$$|\overrightarrow{AB}| + |\overrightarrow{BC}| \geq |\overrightarrow{AB} + \overrightarrow{BC}| = |\overrightarrow{AC}|,$$

which is equivalent to $c + a \leq b$. By antisymmetry, equality holds. By the equality condition for the Euclidean vector triangle inequality, this means $\vec{AB}$ and $\vec{BC}$ are scalar multiples of each other, which means the points $A; B; C$ are collinear, implying degeneracy of the triangle. Further, for the sake of contradiction, suppose $a + b \leq c$ or $b + c \leq a$. Using $c + a = b$, we get $a \leq 0$ or $c \leq 0$, respectively, which are contradictions. So, exactly one of the triangle inequalities is broken in this case and it becomes an equality. □

Problem 3.11. Given distinct collinear points $A; B; C$, Euclidean geometry says that B lies strictly between A and C if and only if $AB + BC = AC$. This will be our definition of B being strictly between A and C. Prove that it suffices to show that $AB < AC$ and $BC < AC$ in order to prove that B lies strictly between A and C.

Theorem 3.12. Let $z; w$ be distinct complex numbers. A complex number p lies on the line through z and w if and only if there exists some $x \in \mathbb{R}$ such that

$$p = (1 - x)z + xw.$$

More specifically,

$$\begin{cases} z \text{ lies strictly between } p \text{ and } w & \text{if } x < 0 \\ p \text{ coincides with } z & \text{if } x = 0 \\ p \text{ lies strictly between } z \text{ and } w & \text{if } 0 < x < 1 \\ p \text{ coincides with } w & \text{if } x = 1 \\ w \text{ lies strictly between } z \text{ and } p & \text{if } x > 1 \end{cases}.$$

Proof. Suppose

$$p = (1 - x)z + xw$$

for some real $x.$ Isolating x yields the equivalent equation

$$x = \frac{p - z}{w - z};$$

which is true if and only if $z; p; w$ are collinear in some order by the preceding theorem. Now we need to classify when they are in what order. If $x = 0$ then $p = z;$ and if $x = 1$ then $p = w.$ These are the easy cases. For the three "between" cases, we will apply Problem 3.11:

- Suppose $x < 0.$ Then $1 < 1 - x$, so

  $$1 < |1 - x| = \left|\frac{w - p}{w - z}\right| \implies |w - z| < |w - p|.$$

  Moreover, $x < 0$ implies that $\frac{1}{x} < 0$. Then $1 < 1 - \frac{1}{x}$. so

  $$1 < \left|1 - \frac{1}{x}\right| = \left|\frac{p - w}{p - z}\right| \implies |p - z| < |p - w|.$$

  Thus, z must lie strictly in between p and $w.$

- Suppose $0 < x < 1.$ Then
$$1 > x = |x| = \left|\frac{p-z}{w-z}\right| \implies |p-z| < |w-z|.$$
Moreover, $1 > x > 0$ leads to $1 > 1-x > 0,$ which means
$$1 > |1-x| = \left|\frac{w-p}{w-z}\right| \implies |w-p| < |w-z|$$
as well, so p must be strictly in between z and $w.$

- Suppose $x > 1.$ Then $1 < |x| = \left|\frac{p-z}{w-z}\right|$ or $|w-z| < |p-z|.$ Moreover,
$$x > 1 \implies 0 < \frac{1}{x} < 1 \implies 0 < 1-\frac{1}{x} < 1$$
$$\implies 1 < \frac{1}{1-\frac{1}{x}} = \left|\frac{1}{1-\frac{1}{x}}\right| = \left|\frac{p-z}{p-w}\right|.$$
Then $|p-w| < |p-z|.$ Thus, w must lie strictly in between z and $p.$

□

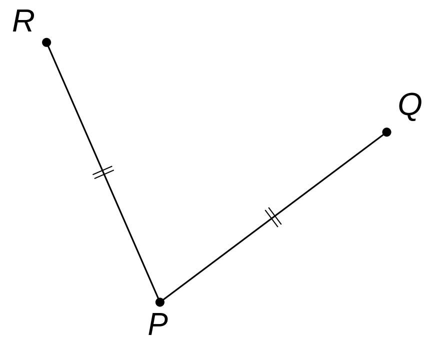
Theorem 3.13. If k is a Euclidean isometry, then k maps the line through $a, b \in \mathbb{C}$ to the line through $k(a), k(b).$ Moreover, if z is some point in the plane, then the perpendicular distance from z to the line through a, b is equal to the distance from $k(z)$ to the line through $k(a), k(b).$

Proof. By Theorem 3.3, every Euclidean isometry may be decomposed into translations, rotations around the origin, and conjugations. So, it suffices to prove the two assertions for each of these three specific types of transformations. Recall that the translation of a point amounts to adding some $w \in \mathbb{C}$ to the point, rotation of a point around the origin is multiplication of the point by $e^{i\theta}$ for some real $0 \le \theta < 2\pi,$ and conjugation applies complex conjugation to the point.

We know that the line through a, b is the collection of all points $(1-x)a + xb$ for $x \in \mathbb{R}.$ It is easy to show that, for any of the three types of transformations k stated,
$$k((1-x)a + xb) = (1-x)k(a) + xk(b),$$
which shows that k maps the line through a, b to the line through $k(a), k(b).$

Let z be a point in the point. By the complex foot formula (Problem 2.27), we want to show that the distance
$$\left|z - \frac{z(\overline{a}-\overline{b}) + \overline{z}(a-b) + \overline{a}b - a\overline{b}}{2(\overline{a}-\overline{b})}\right| = \left|\frac{z(\overline{a}-\overline{b}) - \overline{z}(a-b) + a\overline{b} - \overline{a}b}{2(\overline{a}-\overline{b})}\right|$$
is invariant under translations, rotations around the origin, and conjugations. This is a computational exercise that we leave to the reader; it is not as intensive as it looks. □

Definition 3.14. A point P is said to be **equidistant** from two points Q and R if $PQ = PR:$ We denote that two line segments have equal length using the same number of tick marks.



Problem 3.15. Prove that the midpoint of the segment between two complex numbers z_1 and z_2 is

$$z_3 = \frac{z_1 + z_2}{2}:$$

Problem 3.16. It can be shown that the reflection of a point P across a line \` is the unique point $Q$ such that \` is the perpendicular bisector of $PQ;$ assuming P does not lie on \`: Taking this property for granted, find a formula for the reflection of z 2 C across the line through $a; b$ 2 C; in terms of $z; a; b$.

Chapter 4

Dot Product

> "It is difficult to give an idea of the vast extent of modern mathematics. The word "extent" is not the right one: I mean extent crowded with beautiful detail - not an extent of mere uniformity such as an objectless plain, but of a tract of beautiful country seen at first in the distance, but which will bear to be rambled through and studied in every detail of hillside and valley, stream, rock, wood, and flower."
>
> – *Arthur Cayley*

Having been introduced to Euclidean vectors through equipollence, we now dive deeper into geometric properties of vectors, specifically angles between vectors. To that end, the dot product of Euclidean vectors that we will study will be helpful for handling angles between vectors. Throughout, we will see the Cauchy-Schwarz inequality, orthogonal projections, and a more sophisticated proof of the point-line perpendicular distance formula.

4.1 Algebraic Generalities

We begin with a question that motivates the definition of the dot product: In two dimensions, given two vectors with a common tail, how does one obtain the angle between the tails? The answer is to use the cosine law, as done below.

Definition 4.1. The Euclidean **norm** or **magnitude** of a position vector

$$v = (v_1, v_2, \ldots, v_n)$$

in $\mathbb{R}^n$ is denoted by and defined as

$$\|v\| = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}.$$

As the Pythagorean theorem (Theorem 9.12) and its 3D box analogue for space diagonals (Problem 14.9) inspire, this is the "length" of the position vector, which is the same as the distance from the origin to the point $(v_1, v_2, \ldots, v_n)$. A **unit vector** is a vector whose norm is 1. To divide a non-zero vector v by the scalar $\|v\|$ in order to produce a unit vector in the same direction is called **normalization**; the reader should verify that a normalized vector is indeed a unit vector.

Theorem 4.2. If $v = (x_1, y_1), w = (x_2, y_2)$ are two-dimensional vectors, then the angle θ between their tails is given by

$$\cos\theta = \frac{x_1x_2 + y_1y_2}{\sqrt{x_1^2 + y_1^2}\,\sqrt{x_2^2 + y_2^2}}.$$

Note that the explementary angles θ and $2\pi - \theta$ have the same cosine, so we can choose to solve for the non-reflex angle.

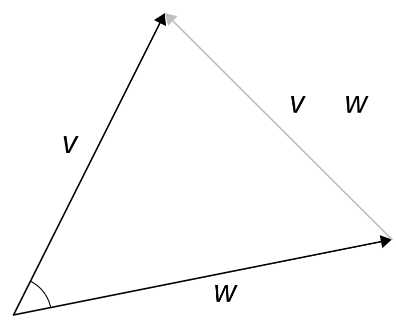
Proof. The two vectors v and w are like the sides of a triangle that has the interior angle θ between them. By the parallelogram law (Theorem 1.39), the length of the third side is the magnitude of the vector

$$v - w = (x_1 - x_2, y_1 - y_2).$$

By the cosine law (Theorem 3.8),

$$\begin{aligned}
2\,\|v\|\,\|w\|\cos\theta &= \|v\|^2 + \|w\|^2 - \|v - w\|^2 \\
&= (x_1^2 + y_1^2) + (x_2^2 + y_2^2) - [(x_1 - x_2)^2 + (y_1 - y_2^2)] \\
&= x_1^2 + y_1^2 + x_2^2 + y_2^2 - (x_1^2 + x_2^2 - 2x_1x_2 + y_1^2 + y_2^2 - 2y_1y_2) \\
&= 2x_1x_2 + 2y_1y_2 \\
\|v\|\,\|w\|\cos\theta &= x_1x_2 + y_1y_2,
\end{aligned}$$

which is equivalent to what we wished to prove.



□

This leads to the more general concept in n-dimensional Euclidean space below.

Definition 4.3. The **Euclidean dot product** of two position vectors

$$v = (v_1, v_2, \dots, v_n) \text{ and } w = (w_1, w_2, \dots, w_n)$$

in $\mathbb{R}^n$ is a real number (in particular, not a vector) that is denoted by and defined as

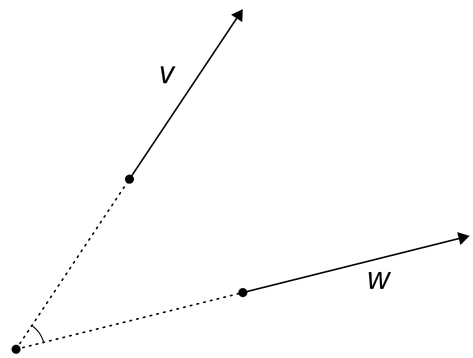
$$v \cdot w = v_1w_1 + v_2w_2 + \cdots + v_nw_n.$$

In more abstract scenarios, the dot product is called an **inner product** of vectors and it is denoted by $\langle v, w\rangle.$ We will occasionally prefer this notation to distinguish the dot product from real multiplication or scalar multiplication of vectors.

To find dot products or norms of displacement vectors instead of position vectors, we can replace the displacement vectors with their equipollent position vectors, and then apply the dot product or norm. So, the norms of vectors in the same equipollence class are equal. Intuitively, this makes sense because, thanks to the usual Euclidean distance formula (Theorem 2.1)

$$\sqrt{(v_1 - w_1)^2 + (v_2 - w_2)^2 + \cdots + (v_n - w_n)^2}$$

defined on $\mathbb{R}^n$, the tail-to-arrowhead distances of equipollent vectors are the same. On the other hand, the dot product is close to representing "the angle" between two vectors, independent of the chosen displacement vectors. Informally, such an angle always exists because we can always extend the tail ends of linearly independent vectors until they meet to form an angle. See if you can justify that the angle is well-defined on vectors in the sense that its measure is independent of the choice of displacement representatives of the two vectors.



Theorem 4.4. The dot product $\langle \cdot, \cdot \rangle$ is a "symmetric bilinear form," which is fancy language for meaning that the following properties hold for all vectors u, v, w in $\mathbb{R}^n$ and all real λ:

1. Commutative: $\langle v, w \rangle = \langle w, v \rangle$
2. Distributive: $\langle u + v, w \rangle = \langle u, w \rangle + \langle v, w \rangle$
3. Homogeneous: $\langle \lambda v, w \rangle = \lambda \langle v, w \rangle$

As a consequence of commutativity, we can deduce the further properties corresponding to the second and third properties:

$$\langle w, u + v \rangle = \langle w, u \rangle + \langle w, v \rangle$$
$$\langle w, \lambda v \rangle = \lambda \langle w, v \rangle.$$

Moreover, $\langle v, 0 \rangle = 0$.

We leave the proofs of these properties to the reader as they are involve simple algebraic manipulations. As a side note, associativity is impossible for the dot product because its domain and range are in conflict. Do you see how?

Theorem 4.5 (Weak cancellation rule). If u, v, w are vectors such that $u \cdot v = u \cdot w$, then it is not necessarily true that $v = w$, so we cannot generally cancel u from the left sides. For example, it might be true that u is the zero vector in $\mathbb{R}^n$. However, if v, w are vectors in $\mathbb{R}^n$ such that $\langle u, v \rangle = \langle u, w \rangle$ for *all* vectors u in $\mathbb{R}^n$, then it holds that $v = w$.

Proof. Note that $\langle u, v\rangle = \langle u, w\rangle$ if and only if $\langle u, v - w\rangle = 0$; and that $v = w$ if and only if $v - w = 0$. So it suffices to prove that if $\langle u, z\rangle = 0$ for all vectors u, then $z = 0$. Let

$$z = (z_1, z_2, \ldots, z_n).$$

The idea is to take u to be each of the n vectors such that the arrowhead has exactly one component equal to 1 and the rest of the components are 0; these are called the **standard basis vectors** of $\mathbb{R}^n$. So by taking

$$\begin{aligned} u &= (1, 0, 0, \ldots, 0, 0), \\ u &= (0, 1, 0, \ldots, 0, 0), \\ u &= (0, 0, 1, \ldots, 0, 0), \\ &\vdots \\ u &= (0, 0, 0, \ldots, 1, 0), \\ u &= (0, 0, 0, \ldots, 0, 1), \end{aligned}$$

we successively find that

$$0 = z_1 = z_2 = z_3 = \cdots = z_{n-1} = z_n.$$

Therefore, z is the 0 vector. □

Theorem 4.6 (Algebraic properties of Euclidean norm)**.** Let v be a vector in $\mathbb{R}^n$ and let λ be a real number. Then:

1. Relation between norm and dot product: $\|v\| = \sqrt{\langle v, v\rangle}$
2. Trivial inequality: $\|v\| \geq 0$ and equality holds if and only if v is the zero vector
3. $\|\lambda v\| = |\lambda| \, \|v\|$

These are all easy and instructive to verify. We encourage the reader to do so.

Problem 4.7. Prove that the sum of the squares of the lengths of the diagonals of any parallelogram is equal to the sum of the squares of the lengths of the four sides. Complete this proof in a way that it holds regardless of the dimension of the Euclidean space in which we are working (so, the two-dimensional definition of a parallelogram should be irrelevant).

Theorem 4.8 (Euclidean Cauchy-Schwarz inequality)**.** For any two position vectors v and w in $\mathbb{R}^n$, it holds that

$$\|v\| \, \|w\| \geq |\langle v, w\rangle|.$$

Equality holds if and only if v and w are linearly dependent.

Proof. First, we provide the abstract proof, which is a bit out of the blue. If $w = 0$, then it may immediately be checked that the inequality holds, and, in fact, equality holds with both sides being 0. So suppose $w \neq 0$. For any $c \in \mathbb{R}$, we can manipulate

$$\begin{aligned} 0 &\leq \|v - cw\|^2 \\ &= \langle v - cw, v - cw\rangle \\ &= \langle v, v\rangle + \langle v, -cw\rangle + \langle -cw, v\rangle + \langle -cw, -cw\rangle \\ &= \|v\|^2 - 2c\langle v, w\rangle + c^2\langle w, w\rangle. \end{aligned}$$

The step that is difficult to motivate is that we let $c = \frac{\langle v, w\rangle}{\langle w, w\rangle}$. This leads to

$$\begin{aligned} \|v\|^2 - 2c\langle v, w\rangle + c^2\langle w, w\rangle &= \|v\|^2 - 2\frac{\langle v, w\rangle}{\langle w, w\rangle}\langle v, w\rangle + \frac{\langle v, w\rangle^2}{\langle w, w\rangle^2}\langle w, w\rangle \\ &= \|v\|^2 - \frac{\langle v, w\rangle^2}{\langle w, w\rangle} \\ &= \|v\|^2 - \frac{|\langle v, w\rangle|^2}{\|w\|^2}. \end{aligned}$$

Since we started by saying $0 \leq \|v - cw\|^2$, this means

$$\begin{aligned} \|v\|^2 - \frac{|\langle v, w\rangle|^2}{\|w\|^2} &\geq 0, \\ \|v\|^2\|w\|^2 &\geq |\langle v, w\rangle|^2 \\ \|v\|\|w\| &\geq |\langle v, w\rangle|. \end{aligned}$$

Now we will prove the equality criterion using Lemma 1.25. If equality holds, then $v = cw$, which proves linear dependence. On the other hand, if we are assuming linear dependence, and since we are working in the $w \neq 0$ case, there must exist $r \in \mathbb{R}$ such that $v = rw$. It is straightforward to then check that equality holds with

$$\|v\|\|w\| = |r|\|w\|^2 = |\langle v, w\rangle|.$$

As an alternate justification that is more concrete, both sides of the inequality are non-negative, so we may square it to product the equivalent inequality

$$\|v\|^2\|w\|^2 \geq \langle v, w\rangle^2.$$

Letting $v = (v_1, v_2, \dots, v_n)$ and $w = (w_1, w_2, \dots, w_n)$ be position vectors, the inequality may be written as

$$(v_1^2 + v_2^2 + \cdots + v_n^2)(w_1^2 + w_2^2 + \cdots + w_n^2) \geq (v_1w_1 + v_2w_2 + \cdots + v_nw_n)^2.$$

This is the standard Cauchy-Schwarz inequality on real numbers that we proved using the quadratic discriminant in Volume 1. The equality condition that we derived there is that $v = 0$ or $w = tv$ for some real t, which is equivalent to linear dependence of v and w, according to Lemma 1.25. □

Corollary 4.9 (Euclidean triangle inequality). Let $v_1, v_2, \dots, v_m$ be vectors in $\mathbb{R}^n$. Then

$$\|v_1\| + \|v_2\| + \cdots + \|v_m\| \geq \|v_1 + v_2 + \cdots + v_m\|.$$

Equality holds if and only if all of the v_i are 0, or they are all equal to non-negative real numbers times one particular non-zero v_i; in other words, they all point in the same direction (not even opposite directions suffice).

Proof. We will start with $m = 2$ as the base case and proceed by induction on integers $m \geq 2$. By the Cauchy-Schwarz inequality,

$$\begin{aligned}\|v + w\|^2 &= \langle v + w, v + w\rangle \\ &= \|v\|^2 + \|w\|^2 + 2\langle v, w\rangle \\ &\leq \|v\|^2 + \|w\|^2 + 2|\langle v, w\rangle| \\ &\leq \|v\|^2 + \|w\|^2 + 2\|v\|\,\|w\| \\ &= (\|v\| + \|w\|)^2,\end{aligned}$$

so taking square roots yields

$$\|v + w\| \leq \|v\| + \|w\|.$$

Equality holds if and only if

$$\langle v, w\rangle = |\langle v, w\rangle| = \|v\|\,\|w\|.$$

We want to show this equality condition is equivalent to: $v = w = 0$, or $v = tw$ for some non-negative real t if $w \neq 0$, or $w = tv$ for some non-negative real t if $v \neq 0$. For one direction, suppose it holds that

$$\langle v, w\rangle = |\langle v, w\rangle| = \|v\|\,\|w\|.$$

By the Cauchy-Schwarz equality criterion, v, w are linearly dependent. By Lemma 1.25, this is equivalent to $v = tw$ or $w = tv$ for some real t. If $v = 0$ and $w = 0$, then we are done. If $v = 0$ and $w \neq 0$, then we may take $t = 0$ in $v = tw$. Similarly, if $v \neq 0$ and $w = 0$, then we may take $t = 0$ in $w = tv$. So now we may assume that $v \neq 0$ and $w \neq 0$. Substituting either of $v = tw$ or $w = tv$ (whichever holds) into $\langle v, w\rangle = |\langle v, w\rangle|$ leads to $t = |t|$, meaning t is non-negative. This proves one direction of the equivalence.
Conversely, suppose: $v = w = 0$, or $v = tw$ for some non-negative real t if $w \neq 0$, or $w = tv$ for some non-negative real t if $v \neq 0$. We get what we want if $v = w = 0$, so that case is easy. Otherwise, the fact that $v = tw$ or $w = tv$ implies that $|\langle v, w\rangle| = \|v\|\,\|w\|$. The fact that t is non-negative implies that $\langle v, w\rangle = |\langle v, w\rangle|$. This establishes the base case.
Suppose the result holds for some integer $m \geq 2$. Let $v_1, v_2, \dots, v_m, v_{m+1}$ be vectors in $\mathbb{R}^n$, which we may assume are all non-zero vectors because otherwise the inequality reduces to a lower induction case (we can use strong induction). By the induction hypothesis,

$$\|v_1\| + \|v_2\| + \cdots + \|v_{m-1}\| + \|v_m + v_{m+1}\| \geq \|v_1 + v_2 + \cdots + v_m + v_{m+1}\|,$$

with equality holding if and only if $v_1, v_2, \dots, v_{m-1}$, and $v_m + v_{m+1}$ all point in the same direction. Since they are all assumed to be non-zero, we may say that $v_2, \dots, v_{m-1}$, and

$v_m + v_{m+1}$ are all positive multiples of v_1. The desired inequality follows from transitivity because the base case tells us that

$$\|v_m\| + \|v_{m+1}\| \geq \|v_m + v_{m+1}\|,$$

with equality holding if and only if $tv_m = v_{m+1}$ for some positive real t. Recall that there exists a positive real s such that

$$sv_1 = v_m + v_{m+1} = v_m + tv_m = (t+1)v_m.$$

This implies that

$$\begin{aligned} v_m &= \frac{s}{t+1}v_1, \\ v_{m+1} = tv_m &= \frac{st}{t+1}v_1, \end{aligned}$$

which makes all of the v_i positive multiples of v_1. The converse holds as well because for positive t_i,

$$\begin{aligned} \|v_1\| + \|v_2\| + \cdots + \|v_m\| + \|v_{m+1}\| &= \|v_1\| + \|t_2 v_1\| + \cdots + \|t_m v_1\| + \|t_{m+1} v_1\| \\ &= (1 + |t_2| + \cdots + |t_m| + |t_{m+1}|)\|v_1\| \\ &= (1 + t_2 + \cdots + t_m + t_{m+1})\|v_1\| \\ &= |1 + t_2 + \cdots + t_m + t_{m+1}| \, \|v_1\| \\ &= \|v_1 + t_2 v_1 + \cdots + t_m v_1 + t_{m+1} v_1\| \\ &= \|v_1 + v_2 + \cdots + v_m + v_{m+1}\|, \end{aligned}$$

so equality holds under this condition. □

4.2 Angles and Projections

Theorem 4.10 (Trigonometric dot product). In two or three dimensions, there is a well-defined non-reflex angle θ and its explementary angle $2\pi - \theta$ between two given non-zero vectors v, w. We claim that

$$v \cdot w = \|v\| \, \|w\| \cos\theta.$$

Note that

$$\cos\theta = \cos(2\pi - \theta),$$

so θ can refer to either angle here.

Proof. Let v, w be non-zero vectors in two dimensions or in three dimensions. By the cosine law (Theorem 3.8),

$$\begin{aligned} \|v\|^2 + \|w\|^2 - 2\|v\| \, \|w\| \cos\theta &= \|v - w\|^2 \\ &= \langle v - w, v - w\rangle \\ &= \langle v, v\rangle + \langle v, -w\rangle + \langle -w, v\rangle + \langle w, w\rangle \\ &= \|v\|^2 + \|w\|^2 - 2\langle v, w\rangle, \end{aligned}$$

which simplifies to

$$\langle v, w\rangle = \|v\|\,\|w\|\cos\theta,$$

as desired. This argument works as long as we have the cosine law, which we proved in two dimensions. It also holds in three dimensions because every pair of linearly independent displacement vectors with a shared tail in 3D has a copy of the 2D plane running through it (see the definition of the plane in Definition 13.3). Even though we have not proven it, what this means is that 3D rigid motions can be used to rotate the (x, y)-plane into any plane, and rigid motions preserve angles and lengths. □

This proves the relationship between the dot product and angles. In dimensions higher than three, we have no geometric intuition, but we can use this θ as the definition of the "angle" between two non-zero vectors.

Definition 4.11. According to Theorem 4.10, two non-zero vectors in two dimensions or three dimensions are perpendicular to each other if and only if their dot product is 0, since

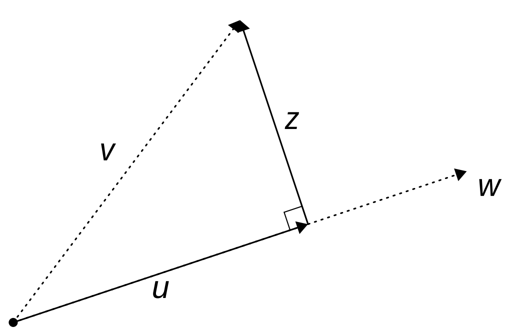
$$\cos\theta = 0 \iff \theta \equiv \frac{\pi}{2}, \frac{3\pi}{2} \pmod{2\pi}.$$

Thus, we define that two vectors in $\mathbb{R}^n$ are **orthogonal** (meaning perpendicular) if their dot product is 0. Note that this means that if one of the two vectors is 0, then the two vectors are automatically orthogonal, even though there is no angle between them of which we can speak.

Problem 4.12. Prove the following variation of the difference of squares factorization for the dot product. If v and w are vectors, then

$$(v - w)\cdot(v + w) = \|v\|^2 - \|w\|^2.$$

Theorem 4.13 (Orthogonal projection). Let v and w be vectors in $\mathbb{R}^n$ where w is non-zero. Then there exist unique vectors u and z such that u is parallel to (as in, a scalar multiple of) w, and z is orthogonal to w and $v = u + z$. Here, u is called the **projection of** v *to* w and we denote it by $u = \mathrm{proj}_w v$, and z is called the **rejection of** v *from* w and we denote it by $z = \mathrm{oproj}_w v$. In a sense, we are uniquely decomposing v into the shadow u that v casts on w and what is left out, z.



Proof. Suppose such vectors u and x exist. Then there exists a real number t such that $u = tw.$ This allows us to compute

$$0 = \langle z, w\rangle = \langle v - u, w\rangle = \langle v - tw, w\rangle = \langle v, w\rangle - t\langle w, w\rangle,$$
$$t = \frac{\langle v, w\rangle}{\langle w, w\rangle},$$

where the division by $\langle w, w\rangle$ is permissible because $w \neq 0.$ It is automatically true that $u = tw$ is parallel to w, and trying out this particular t shows that

$$z = v - u = v - tw = v - \frac{\langle v, w\rangle}{\langle w, w\rangle} w$$

satisfies $z \perp w.$ We leave this last computation to the reader. Thus, u and z exist, and they are unique because we initially uniquely isolated t and uniquely isolated $z = v - u.$ □

Theorem 4.14 (Vector Pythagorean theorem). Let v and w vector with w non-zero. Let the projection of v to w be $u = \mathrm{proj}_w v$ and let the rejection of v from w be $z = \mathrm{oproj}_w v$. Then

$$\|u\|^2 + \|z\|^2 = \|v\|^2.$$

Proof. Based on the definitions of u and z, we know that

$$v = u + z,$$
$$u \cdot z = 0.$$

Dotting each side of the first equation with itself yields

$$\begin{aligned} \|v\|^2 &= v \cdot v \\ &= (u + z) \cdot (u + z) \\ &= u \cdot u + u \cdot z + z \cdot u + z \cdot z \\ &= \|u\|^2 + 2 \cdot 0 + \|z\|^2. \end{aligned}$$

□

Lemma 4.15. In two dimensions, if a vector v is orthogonal to a vector w, then v is orthogonal to tw for every real $t.$ In this case, if v is also orthogonal to another vector that cannot be written in the form tw (meaning a vector that is linearly independent with w), then v is the 0 vector. Note that this means that if v is non-zero and orthogonal to $w,$ then v orthogonal to each tw and nothing else.

Proof. If $\langle v, w\rangle = 0,$ then it is clear that

$$0 = \langle v, tw\rangle = t\langle v, w\rangle = t \cdot 0 = 0.$$

Now suppose that u is a vector that is linearly independent with w and such that $\langle u, v\rangle = 0.$ Then $\langle v, au + bw\rangle = 0$ for all real a, b. But u, w are linearly independent, so they generate the whole plane by Problem 1.31. Thus, we may use the weak cancellation rule (Theorem 4.5) to get $v = 0.$ □

Definition 4.16. A vector v is **orthogonal** or a **normal vector** to a line $\ell = \{p+tw : t \in \mathbb{R}\}$ if v is orthogonal to w, since being orthogonal to one direction vector of the line means being orthogonal to all direction vectors of the line, according to Lemma 4.15.

Theorem 4.17. In two dimensions, let ℓ be a line defined in standard form by $Ax+By+C = 0,$ where at least one of A, B is non-zero. Let (x_1, y_1) and (x_2, y_2) be distinct points on ℓ, and (x_0, y_0) be an arbitrary point in the plane. Let

$$v = ((x_1, x_1), (x_0, y_0))$$

be a displacement vector so that its tail is on $\ell,$ and let $w = ((x_1, y_1), (x_2, y_2))$ be a vector that lies on $\ell.$ Then the perpendicular distance from (x_0, y_0) to ℓ is

$$\|\text{oproj}_w v\| = \|\text{proj}_z v\| = \frac{|Ax_0 + By_0 + C|}{\sqrt{A^2 + B^2}},$$

where $z = ((0, 0), (A, B))$, which turns out to be a normal vector to $\ell.$

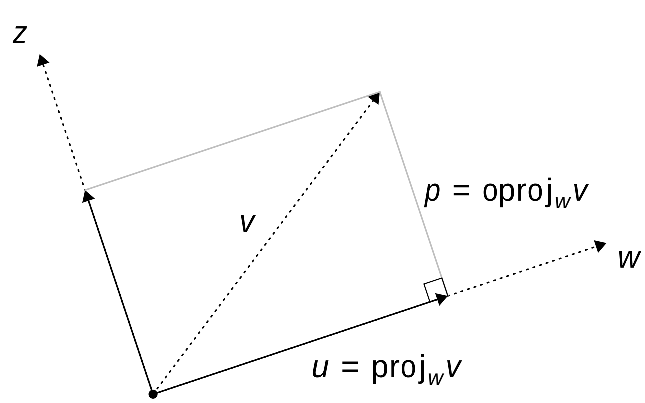
Proof. Firstly, letting $p = \text{oproj}_w v$ we are seeking $\|p\|$ since the enclosed vector is perpendicular to every direction vector of the line. As an intermediate step, we will show that (A, B) is a normal vector of the line. Since (x_1, y_1) and (x_2, y_2) are on ℓ they satisfy

$$\left\{\begin{aligned} Ax_1 + By_1 + C &= 0, \\ Ax_2 + By_2 + C &= 0 \end{aligned}\right. \implies A(x_2 - x_1) + B(y_2 - y_1) = 0.$$

Equivalently, the dot product of the position vector (A, B) and the displacement vector w is $0,$ so (A, B) is indeed normal to ℓ.

Due to the following diagram of the parallelogram law in the rectangular case, we see that that

$$\|\text{oproj}_w v\| = \|v - \text{proj}_w v\| = \|\text{proj}_z v\|.$$



By Theorem 4.13,

$$\begin{aligned}
\|\operatorname{proj}_z v\| &= \left\| \frac{\langle (x_0 - x_1, y_0 - y_1), (A, B) \rangle}{\langle (A, B), (A, B) \rangle} (A, B) \right\| \\
&= \frac{|\langle (x_0 - x_1, y_0 - y_1), (A, B) \rangle|}{\|(A, B)\|} \\
&= \frac{|Ax_0 - Ax_1 + By_0 - By_1|}{\sqrt{A^2 + B^2}} \\
&= \frac{|Ax_0 + By_0 + C|}{\sqrt{A^2 + B^2}},
\end{aligned}$$

where we have used to the fact that $Ax_1 + By_1 + C = 0$ in the final step. □

Chapter 5

Polygons

> "Emptiness is everywhere and it can be calculated, which gives us a great opportunity. I know how to control the universe. So tell me, why should I run for a million [dollars]?"
>
> – *Grigori Perelman (apocryphal)*

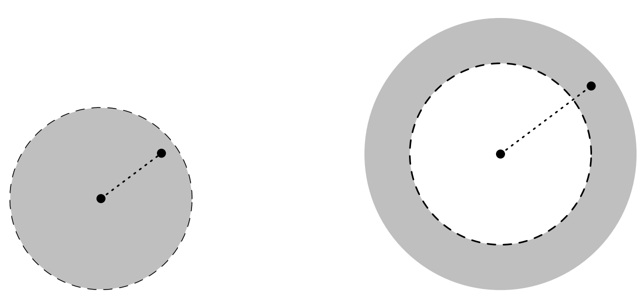
Our preliminary material on polygons will be technical, but necessary. We will state and use the Jordan curve theorem from topology. We will also show that the two-ears theorem is a useful tool for dealing with polygons that are not necessarily convex because it allows for the technique of ear-clipping. We will end with general definitions and results about congruent and similar polygons.

5.1 Interior, Boundary, and Exterior

In order to handle polygons in general, there are a few definitions that we will need from an area of higher mathematics, called topology, but we will keep these concepts at a bare minimum.

Definition 5.1. There are several concepts related to circles:

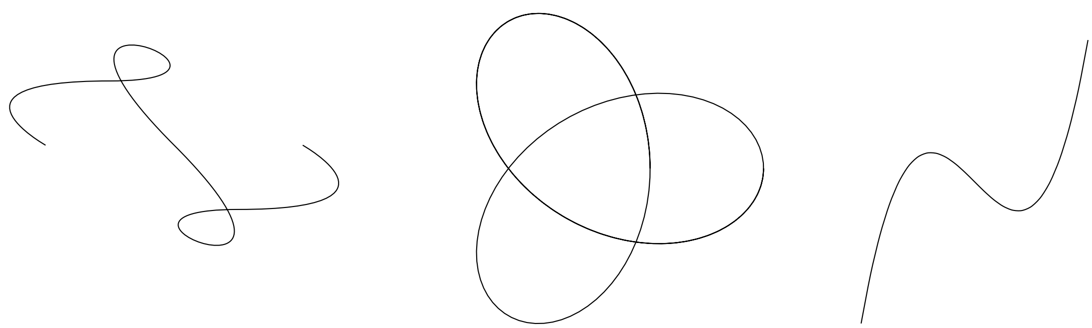
1. The interior or inside of a circle is the collection of points whose distance from the center of the circle is less than the radius.
2. The exterior or outside of a circle is the collection of points whose distance from the center of the circle is greater than the radius.



3. An open disk is only the interior of a circle but without the circle itself, and a closed disk is the union of a circle and its interior.

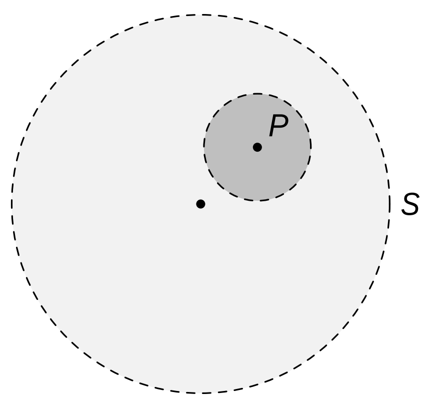
Definition 5.2. A path in the plane is a continuous curve, where continuous informally means that it can be drawn without lifting the writing instrument from the surface. For our purposes, there are two important classes of curves:

- A closed path is a path such that, when it is drawn, the two endpoints are equal, so it forms a loop.
- A simple path is a path that does not intersect itself, with the exception of the two endpoints being equal in the case of a closed path.



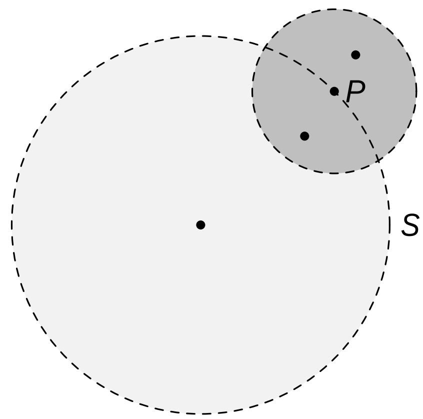
A simple, closed path is called a Jordan curve.

Definition 5.3. A set of points S in the plane is said to be open if for every point P in S, there exists an open disk around P that is contained in S.



Definition 5.4. A set of points S in the plane is said to be connected if there is a path between any two points in the set such that every point in the path is contained in S.

Definition 5.5. Given a set S of points in the plane, its boundary consists of all points P in the plane such that, for every open disk centered at P, the open disk contains a point that is in S and a point that is not in S.



Theorem 5.6 (Jordan curve theorem). For any Jordan curve in the plane, its complement is the union of two non-intersecting regions that are each a connected set: a bounded interior and an unbounded exterior. Moreover, the boundary of each of the two regions is precisely the Jordan curve. There are two consequences of the theorem that we will use:

1. The interior is an open set, and the exterior is an open set.

2. Any path with one endpoint in the interior and one endpoint in the exterior must intersect with the Jordan curve.

Proof. Elementary proofs of this simple-sounding theorem are technical enough that the mathematician Tverberg wrote "*there are many, even among professional mathematicians, who have never read a proof of it.*" As such, we do not provide a proof, but we encourage the reader to draw some simple, closed paths that are a little convoluted and verify the theorem for them. □

Lemma 5.7. An open disk centered at a point on a Jordan curve must contain both an exterior point and an interior point of the Jordan curve.

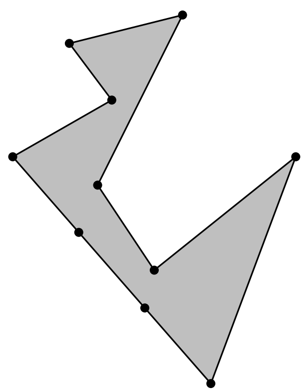
Proof. According to the Jordan curve theorem, a Jordan curve is the boundary of its interior and the boundary of its exterior. By the definition of the boundary of the interior, every open disk centered at every boundary point must contain a point in the interior and a point not in the interior. Similarly, by the definition of the boundary of the exterior, every open disk centered at every boundary point must contain a point in the exterior and a point not in the interior. Thus, every open disk centered at a point on a Jordan curve contains an interior point and an exterior point. □

The following is a technique that we will use several times.

Lemma 5.8. Let J be a Jordan curve and let C be a connected subset of the plane such that C does not contain any points of J: Then C consists of only interior points of J or only exterior points of J. Thus, knowing that C contains one point of the interior of J is enough to conclude that all points of C are in the interior of J; similarly, knowing that C contains one point of the exterior of J is enough to conclude that all points of C are in the exterior of J.

Proof. Suppose C does not contain any points of J. Suppose, for contradiction, that C contains both an interior point and an exterior point of J. Since C is connected, there exists a path with these two points as its endpoints such that the path is consists of only points in C. By the Jordan curve theorem, there should be a point of J on this path, which contradicts the fact that C contains no points of J. Thus, C cannot contain both an interior point and an exterior points of J. The other two conclusions follow immediately. □

Definition 5.9. For our purposes, a **generalized polygon** is a Jordan curve consisting of line segments, along with the interior region of the Jordan curve. To be precise, the boundary of the generalized polygon $A_0A_1 \ldots A_{n-1}$ consists of n distinct points $\{A_0, A_1, \ldots, A_{n-1}\}$ called **vertices** (the singular form is **vertex**), and n line segments $\{A_0A_1, A_1A_2, \ldots, A_{n-1}A_0\}$ called **edges** or **sides**. A generalized polygon has at least 3 vertices and 3 edges; an n-sided generalized polygon is called a generalized n-gon. The **neighbouring edges** of a particular vertex in a generalized polygon are the two edges that emanate from the vertex, whereas its **neighbouring vertices** are the two vertices to which the vertex is attached by an edge.



If we do not allow consecutive edges in a generalized polygon to lie on the same line, so that no three consecutive vertices can be collinear (and so that it is not possible to add a vertex by splitting an edge into two pieces), then the generalized polygon is simply called a **polygon**. A polygon with n edges is called an n-gon. Note that the set of polygons are a subset of the set of generalized polygons.

Example. In order from the least number of sides to the most number of sides, the names of n-gons for $3 \le n \le 12$ are: triangle, quadrilateral, pentagon, hexagon, heptagon, octagon, nonagon, decagon, hendecagon, dodecagon. Non-examples of generalized polygons are those that fail to be simple, or are not closed, or do not consist of line segments. Another non-example of a generalized polygon are degenerate triangles, since that the vertices of a generalized polygon cannot all be collinear due to the "simple" criterion of a Jordan curve.

Theorem 5.10 (Polygon inequality). If $P = A_1A_2 \ldots A_{n-1}A_n$ is a path consisting of line segments (some of which possibly intersect others), then

$$A_1A_2 + A_2A_3 + \cdots + A_{n-2}A_{n-1} + A_{n-1}A_n \geq A_1A_n.$$

Equality holds if and only if the points $A_1, A_2, \ldots, A_n$ are collinear in that order. In particular, in the case of generalized n-gons $A_1A_2 \ldots A_{n-1}A_n$ (note that A_n connects to A_1), the inequality is strict.

Proof. We proceed by induction on the number of sides $n \geq 3$ of the generalized polygon. In the base case $n = 3$, Theorem 3.10 says that $A_1A_2 + A_2A_3 \geq A_1A_3$ for any three points A_1, A_2, A_3 with equality holding if and only if A_1, A_2, A_2 are collinear in that order (since any other order would cause the inequality to be strict). Now suppose the result holds for all generalized n-gons some $n \geq 3$. Let $P = A_1A_2 \ldots A_{n-1}A_nA_{n+1}$ be a path consisting of $n + 1$ vertices. By the induction hypothesis,

$$A_1A_2 + A_2A_3 + \cdots + A_{n-2}A_{n-1} + A_{n-1}A_n \geq A_1A_n,$$

with equality holding if and only if $A_1, A_2, \ldots, A_n$ are collinear in that order. Adding A_nA_{n+1} to both sides yields

$$A_1A_2 + A_2A_3 + \cdots + A_{n-2}A_{n-1} + A_{n-1}A_n + A_nA_{n+1} \geq A_1A_n + A_nA_{n+1} \geq A_1A_{n+1},$$

where we reused the base base in the final step. Equality holds in

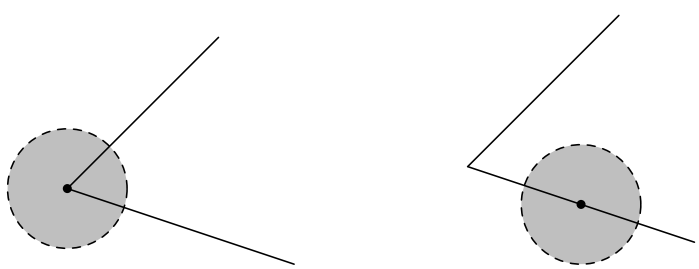
$$A_1A_n + A_nA_{n+1} \geq A_1A_{n+1}$$

if and only if A_1, A_n, A_{n+1} are collinear in that order, implying the overall desired equality criterion.
In the case of a generalized polygon, the base case is a non-degenerate triangle, so the inequality is strict from the beginning onward due to the triangle inequalities being strict for non-degenerate triangles in the plane (Theorem 3.10). □

Definition 5.11. It is clear what we mean by saying that a point (or a set of points, like a line segment) lies on the boundary of a generalized polygon. When we say a point lies on a generalized polygon, we will mean that it lies either on its boundary in the interior of the path formed by the boundary. However, when we say a point lies in the interior of a generalized polygon, we will mean that it lies in the interior of the path formed by the boundary but not on the boundary itself.

Lemma 5.12. For every point P on the boundary of a generalized polygon, there exists an open disk centered around P such that excluding the one or two edges on which P lies from the disk results in two sectors: one sector that lies entirely in the interior and one sector that lies entirely in the exterior of the generalized polygon. Of course, once we find such an open disk, all open disks of smaller radius satisfy the same property. (See Definition 6.1 for the definition of a sector.)

Proof. Let P be a point on the boundary of the generalized polygon. Note that if P is a vertex then it has two neighbouring edges, and otherwise P lies in the interior of an edge.

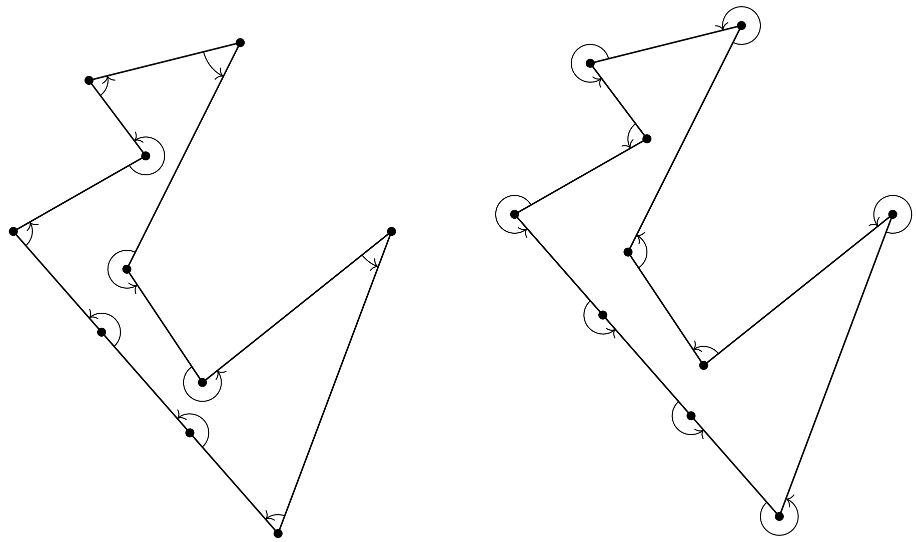


First we show that we can find an open disk centered at P which does not include any boundary points other than parts of the one or two edges on which P lies. We just need a small enough open disk that additional edges do not intersect the disk; as a result, no additional vertices will intersect the disk either because each vertex is includes in an edge. There are finitely many edges on which P does not already lie, so we find the shortest distances from P to the line though each of those other edges, and let the radius of the disk be a positive real number smaller than all of those distances. Thus, there exists a disk that is centered at P and does not contain any new boundary points. So each sector (where the two sectors exclude the edges through P) can contain only interior and exterior points of the generalized polygon. It is easy to see that each sector is a connected set. By Lemma 5.8, each sector consists of only interior points or only exterior points. By Lemma 5.7, an open disk centered at a boundary point must contain both an exterior point and interior point. Therefore, one sector consists of only interior points and the other sector consists of only exterior points. □

Definition 5.13. In a generalized polygon, each vertex has two neighbouring edges. This creates two explementary angles. The **interior angle** at that vertex is chosen out of those two angles such that the intersection of the interior of the angle (recall that we do not include the rays of the angle in its interior) with a sufficiently small open disk around the vertex produces a sector that lies in the interior of the generalized polygon. Lemma 5.12 guarantees that exactly one of the two explementary angles allows for this.

Theorem 5.14. Given a generalized n-gon $A_0A_1 \ldots A_{n-1}$, it is clear that for each index $0 \le i \le n-1$, there exists a counterclockwise rotation around A_i, along with a positive dilation from A_i, that sends $A_{i-1}A_i$ to $A_{i+1}A_i$, where indices are reduced modulo n. Each of these counterclockwise rotations is either on the side of the interior of the generalized polygon (meaning it coincides with the interior angle at A_i) or is on the side of the exterior of the generalized polygon (meaning it coincides with the explementary angle of the interior angle at A_i). We assert that either all or none of the n counterclockwise rotations are on the side of the interior.

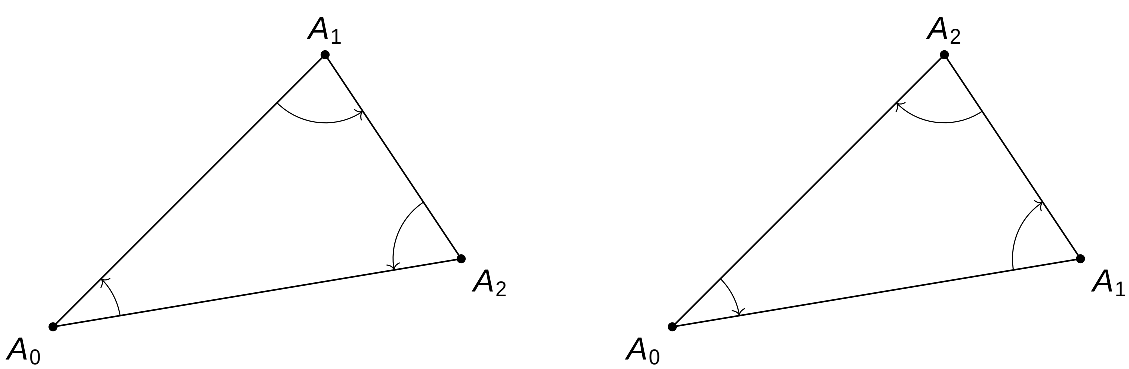
Proof. If none of the counterclockwise rotations lie on the side of the interior, then we are done. So suppose at least one of the counterclockwise rotations lies on the side of the interior. It suffices to show that all of the counterclockwise rotations lie on the side of the interior. By the principle of induction, it is enough to show that if this property is satisfied by the rotation around A_i then this property is satisfied by the rotation around A_{i+1}. Suppose for contradiction that the counterclockwise rotation around A_i lies on the side of the interior, but the counterclockwise rotation around A_{i+1} lies on the side of the exterior. We know that there exists an open disk centered at each of A_i and A_{i+1} such that excluding the neighbouring edges of these vertices from their respective disks results in two sectors for each disk: one sector that lies in the interior and one sector that lies in the exterior. Then the counterclockwise rotation around A_i runs through the interior sector at A_i, and the counterclockwise rotation around A_{i+1} runs through the exterior sector at A_{i+1}. We will show that there exists a path from inside the interior sector at A_i to inside the exterior sector at A_{i+1} without crossing any boundary points, which contradicts the Jordan curve theorem.



Let $P \notin A_i$ be a point on A_iA_{i+1} that is inside the disk at A_i and let $Q \notin A_{i+1}$ be a point on A_iA_{i+1} that is inside the disk at A_{i+1}. We will show that there exists an $\varepsilon > 0$ such that, for each point X on the line segment PQ, excluding A_iA_{i+1} from the open disk of radius ε centered at X yields two half-disks that do not contain any boundary points. Finding ε is a matter of finding the shortest distance from PQ to each edge that is not A_iA_{i+1}, and then choosing ε to be smaller than all of those distances. This allows us to travel from inside the interior sector at A_i to inside the exterior sector at A_{i+1} through half-disks that do not contain any boundary points, producing the desired contradiction. Specifically, we go from A_i to P to Q to A_{i+1}. □

Definition 5.15. Given a generalized n-gon $A_0A_1 \ldots A_{n-1}$, for each index $0 \le i \le n-1$, there is a clockwise or counterclockwise rotation of A_{i-1} around A_i, along with a positive dilation from A_i, that causes $A_{i-1}A_i$ to coincide with $A_{i+1}A_i$, where the angle of the rotation is on the side of the interior angle $\angle A_{i-1}A_iA_{i+1}$. By Theorem 5.14, all of these rotations are in the same direction, clockwise or counterclockwise. Contrary to intuition, we say that the ordering $A_0, A_1, \ldots, A_{n-1}$ of consecutive vertices is a clockwise orientation if the rotations are counterclockwise, and a counterclockwise orientation if the rotations are clockwise.

Example. By observing the direction in which vertices are labelled in convex polygon, like the non-degenerate triangles below, it becomes clear why the direction in which the vertices are ordered is the opposite of the direction of the aforementioned rotations.



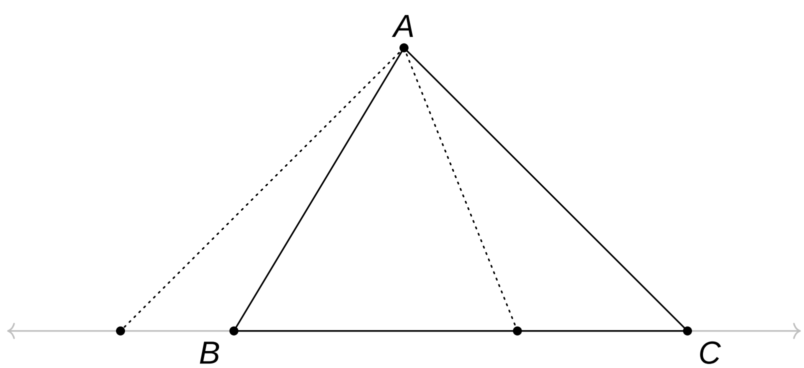
Definition 5.16. A diagonal of a generalized polygon is a line segment that connects two non-consecutive vertices.

Theorem 5.17. Recall that a convex polygon was defined in Definition 2.16. Equivalently, a convex polygon is a polygon that fulfils any one of the following conditions:

- The polygon forms a convex set, meaning the line segment between any two points on the polygon lies on the polygon.
- All interior angles of the polygon are strictly less than 180°. (Note that this implies that all non-degenerate triangles are convex polygons.)
- The interiors of all diagonals of the polygon lie in the interior polygon.
- For each edge, the polygon lies entirely on a half-plane defined by the line running through the edge. Moreover, a convex polygon is the region formed by the intersection of half-planes defined by the lines running through its edges.
- The polygon lies on the rays and interior of each of its interior angles (where we extend the rays of the angle infinitely). More specifically, a convex polygon is the region formed by the intersection of its (indefinitely extended) interior angles.

As a non-example, a generalized polygon with a straight interior angle is not convex. A polygon that is not convex is called concave.

Definition 5.18. A cevian of a triangle is a line segment that has one endpoint on a vertex of the triangle, and the other endpoint, called the foot of the cevian, on the interior of the edge opposite to that vertex. A generalized cevian of a triangle is a line segment that still has one endpoint on a vertex of a triangle, but the other endpoint can be anywhere on the line running through the opposite edge. If A is the vertex from which the cevian emanates, then we may call the cevian an A-cevian.



Theorem 5.19. Suppose we have a triangle $\triangle ABC$ and a generalized cevian emanating from vertex A with foot F. Then the interior of AF lies entirely in:

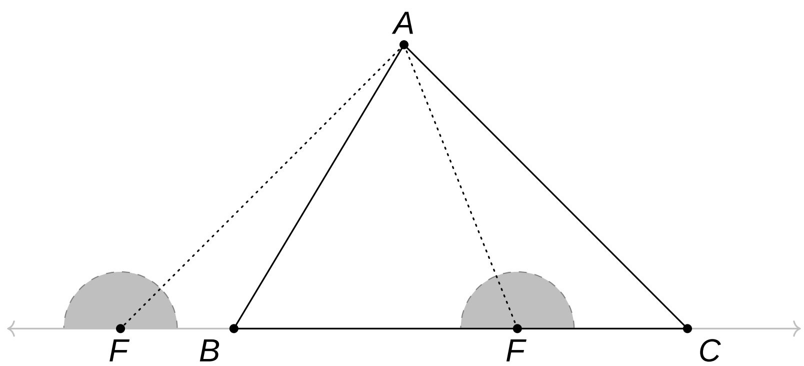
1. The boundary of $\triangle ABC$ if F coincides with B or C. In this case, the interior of AF is the interior of AB or the interior of AC.

2. The interior of $\triangle ABC$ if F lies in the interior of $BC:$

3. The exterior of $\triangle ABC$ if F lies outside $BC:$

As a consequence, the interior of every cevian lies in the interior of the triangle.

Proof. It is clear that if F coincides with B or C, then the interior of AF lies is the interior of AB or $AC;$ respectively. So suppose F lies on the line through BC but not at the endpoints of $BC:$
Note that AF intersects the line through each edge once, in particular at A and $F:$ So removing the endpoints of AF produces the interior of a line segment that does not contain any boundary points of the triangle, as two distinct lines cannot intersect at more than one point. By Lemma 5.8, the interior of AF can contain only interior points or only exterior points of the triangle since the interior of AF is connected. Now we will specialize this result to each of the two remaining cases.



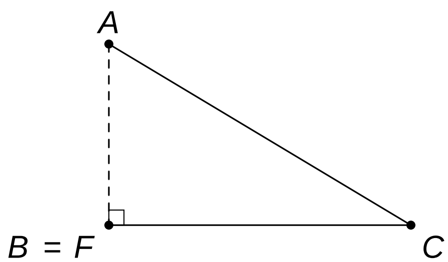
If F lies in the interior of $BC;$ then there is an open disk around F such that excluding BC from the disk results in two half-disks, one of which is contained in the interior of the triangle and intersects with $AF:$ So there is an interior point of the triangle in the interior of $AF;$ causing the interior of AF to be fully composed of interior points of the triangle. If F does not lie in the interior of $BC;$ then it lies in the exterior of the triangle, and so there is an open disk around F that contains only exterior points. So there is an exterior point of the triangle in the interior of $AF;$ causing the interior of AF to be fully composed of exterior points of the triangle. □

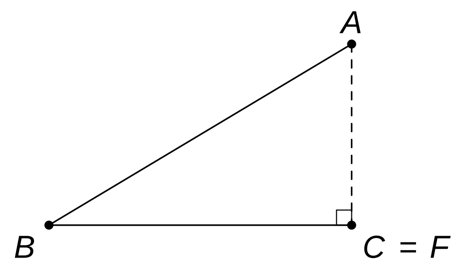
Corollary 5.20 (Crossbar theorem). For each vertex of a triangle and each point in the interior of the triangle, there is a cevian emanating from the vertex that includes the point.

Proof. Suppose we have a triangle $\triangle ABC$ and that P is a point in its interior. Without loss of generality, it suffices to show that there exists a cevian emanating from A that includes $P:$ We can certainly draw a generalized cevian emanating from A through the point $P;$ so it is a matter of showing that the foot F of this cevian lies in the interior of $BC:$ First, suppose for contradiction that F lies outside segment BC and so in the exterior of $\triangle ABC:$ Since P is an interior point, the interior of PF must contain a boundary point of the triangle, which also falls in the interior of $AF:$ By Theorem 5.19, this contradicts the fact that the interior of a generalized cevian cannot contain both a boundary point and an interior point P of the triangle. Thus, F lies on BC. Moreover, we can say that F does not coincide with B or C, because that would cause P to lie on AB or AC instead of the interior of the triangle. So F lies in the interior of BC. □

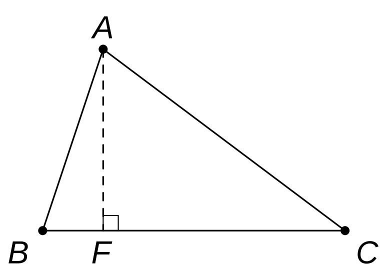
Theorem 5.21. Let A, B, C be distinct points in the plane that are not all collinear. Then the foot F of the perpendicular segment from A to the line through BC lies on:

1. B or C if one of the interior angles $\angle ABC$ or $\angle ACB$ of $\triangle ABC$ is right, respectively.

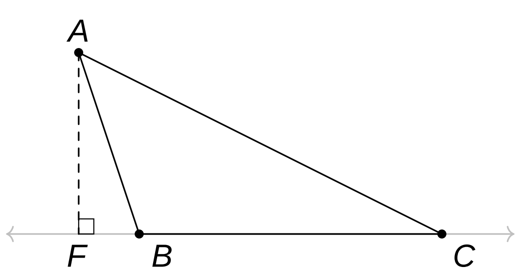


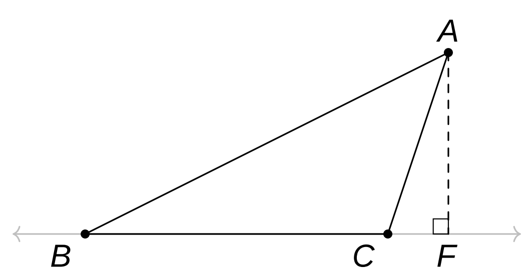


2. The interior of the segment BC if both of the interior angles $\angle ABC$ and $\angle ACB$ of $\triangle ABC$ are acute.



3. Outside the segment BC if one of the interior angles $\angle ABC$ and $\angle ACB$ of $\triangle ABC$ is obtuse.





Proof. We treat the cases separately:

1. If the interior angle $\angle ABC$ of $\triangle ABC$ is right, then AF and AB are both perpendicular to BC, which means they are parallel or they coincide. Since they share the point A, they lie on the same line and $F = B$. If the interior angle $\angle ACB$ of $\triangle ABC$ is right, then the proof of $F = C$ is analogous.

2. Suppose the interior angles $\angle ABC$ and $\angle ACB$ of $\triangle ABC$ are acute. Suppose for contradiction that F does not lie in the interior of the segment BC. Clearly, F cannot lie on B or C because then AF would not form a right angle with the line through BC. So F must lie outside the segment BC. Then the interior of AF lies in the exterior of $\triangle ABC$. Let D be the vertex out of B and C to which F is closer and let E be the other vertex out of B and C. Then $\triangle AFD$ is a right triangle, so its interior angle $\angle ADF$ is acute. Subsequently, the supplementary angle $\angle ADE$ is an obtuse interior angle of $\triangle ABC$, which is a contradiction. Thus, F lies in the interior of BC.

3. Suppose one of the interior angles $\angle ABC$ and $\angle ACB$ of $\triangle ABC$ is obtuse. Suppose for contradiction that F does not lie outside the segment BC. As in the last part, F cannot lie on B or C because then AF would not form a right angle with the line through BC. So F must lie in the interior of the segment BC. Then the interior of AF lies in the interior of $\triangle ABC$. This causes $\triangle ABC$ to split into right triangles $\triangle AFB$ and $\triangle AFC$, implying that that interior angles $\angle ABC = \angle ABF$ and $\angle ACB = \angle ACF$ are both acute. This contradicts the hypothesis, so F lies outside segment BC.

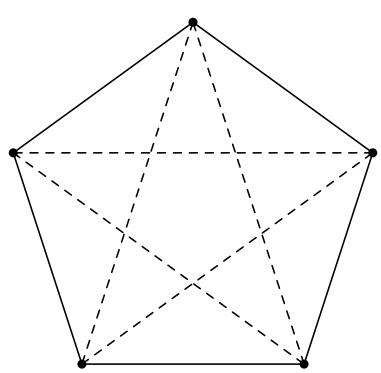
□

Problem 5.22. Let $\triangle ABC$ be isosceles with $CA = CB$. Show that the foot F of the perpendicular from C to the line through AB lies in the interior of AB.

To generalize several results about triangles to generalized polygons, we will use the concept of polygonal ears, as defined below.

Definition 5.23. An ear of a generalized n-gon for $n \geq 4$ is a vertex A such that the interior of the diagonal connecting its neighbouring vertices lies in the interior of the generalized polygon.

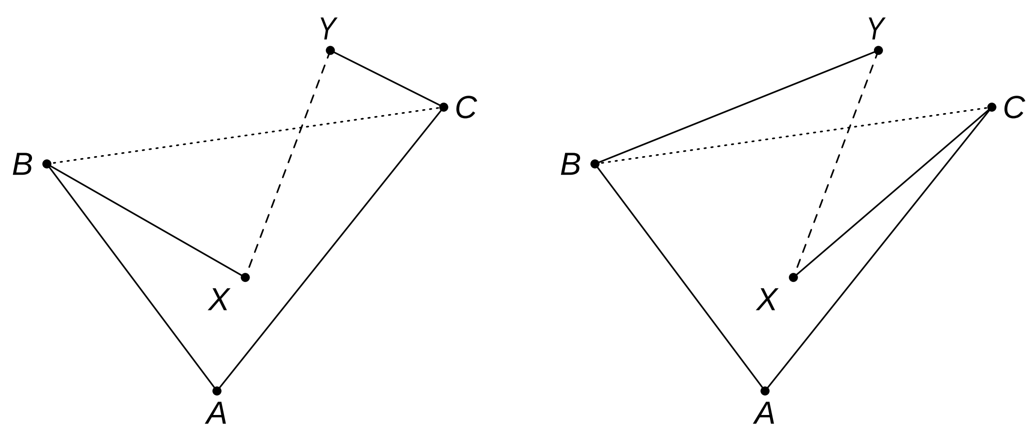
Example. Every vertex of a convex polygon is an ear, since the interior of every diagonal lies in the interior of the polygon.



Lemma 5.24. If a vertex A is an ear of a generalized polygon P then the interior of the triangle that it forms with its neighbouring vertices B and C lies in the interior of P. We call $\triangle ABC$ the triangle induced by A.

Proof. Suppose A is an ear of a generalized polygon P with neighbouring vertices B and $C.$ Then the open segment BC lies in the interior of $P.$ We want to show that every point Z in the interior of $\triangle ABC$ lies in the interior of $P.$ The two snags that we might run into are that the interior of $\triangle ABC$ contains a point on the boundary or exterior of $P.$ We will show that both are impossible, in that order.

Suppose, for contradiction, that the interior of $\triangle ABC$ contains a boundary point $X.$ If all boundary points of P other than those on AB and AC are in the interior of $\triangle ABC$, then P is bounded by $\triangle ABC$ and so none of the interior points of P would be on the side of the line through BC not containing $A.$ This would contradict the fact that the interior of segment BC consists of interior points of P, as every interior point of P has a sufficiently small open disk around it that is contained in the interior of P, which would produce interior points on the side of BC not containing A. So there must exist a boundary point Y of P that is in the exterior of $\triangle ABC.$ There are two simple paths that lie on the boundary of P between the distinct points X and Y, where the two paths are disjoint except at their endpoints. We want to show that one of these paths must cross the interior of $BC,$ which will contradict the fact that the interior of BC cannot include boundary points. Since X is in the interior of $\triangle ABC$ and Y is in the exterior of $\triangle ABC$, each path must cross the boundary of $\triangle ABC$ at some point Q. Since the boundary of P cannot intersect itself and since the interior of BC cannot contain any boundary points, the only possibilities for Q are B and C. If we travel from X to B without going to C first then the same path must continue to A and then C; similarly, if we travel from X to C without going to B first then the same path must continue to A and then $B.$ Thus, one of the two aforementioned paths between X and Y uses up both B and C, forcing the other to pass through the interior of BC. Thus, the interior of $\triangle ABC$ cannot include boundary points.

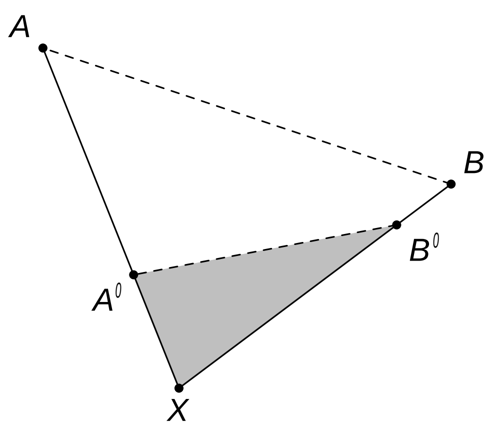


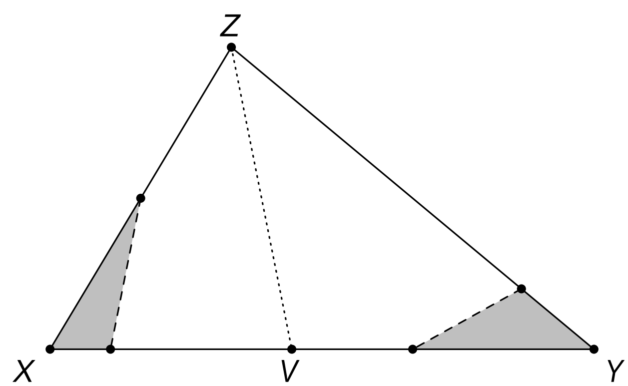
However, this means that the interior of $\triangle ABC$ consists of only interior and exterior points of $P.$ By the crossbar theorem (Corollary 5.20), for every point Z in the interior of $\triangle ABC,$ there is a cevian AF emanating from A with its foot F in in the interior of BC such that the cevian includes $Z.$ This shows that every such point Z can be attached to a point F in the interior of BC by a line segment ZF that does not intersect AB or $AC.$ As the only boundary points of P on $\triangle ABC$ lie on AB and $AC,$ the segment ZF has only interior and exterior points of $P.$ If it has both an interior and an exterior point, that would force there to be a boundary point in the interior of $ZF,$ which is a contradiction. So ZF consists of only interior points or only exterior points of P; as F is an interior point, Z must be an interior point as well, which completes the proof. □

Definition 5.25. Two ears of a generalized polygon are said to be **non-overlapping** if the interiors of the two induced triangles do not overlap. Note that it is acceptable for the induced triangles to overlap at their vertices or edges.

Theorem 5.26 (Two ears theorem). Every polygon with more than three vertices has at least two non-overlapping ears. This implies the same result for generalized polygons.

Proof. Proving the two ears theorem for polygons would be too much of a detour, but we will show that it implies the same result for generalized polygons; for a true proof, see [7].
Instead, we will prove that the result for polygons implies the result for generalized polygons. Suppose we have a generalized polygon $P.$ For each maximal sequence of consecutive edges that lie on the same line, we collapse it into one line segment so that we can interpret P as a (non-generalized) polygon Q. We now treat the case where Q is a triangle separately from when it is not.
If Q is not a triangle, then the two ears theorem tells us Q has two non-overlapping ears X and Y. This means the interiors of the two triangles induced by the ears are non-overlapping, and moreover, both interiors are contained in the interior of Q (which is the same as the interior of P). Then X and Y are also ears of P by the following line of reasoning. Let A, B be the neighbouring vertices of X in $Q,$ and let A', B' be the neighbouring vertices of X in $P.$ Then $XA' \subseteq XA$ and $XB' \subseteq XB.$ So the interior of $A'B'$ lies in the interior of $\triangle AXB$ which lies in the interior of $P.$ The argument for Y is analogous.





Now suppose Q is a triangle. At least one of its edges has a vertex V of P in the interior of the edge, so label the vertices of Q as X, Y, Z in order that V lies in the interior of $XY.$ We claim that X and Y are ears of $P.$ Certainly, connecting the neighbouring vertices of X in P creates a diagonal whose interior lies in the interior of Q and so the interior of $P,$ since a triangle is a convex set; the analogous result holds for $Y.$ So X and Y are ears and it remains to be shown that they are non-overlapping. This is true because $\triangle XVZ$ and $\triangle YVZ$ have non-overlapping interiors, and the interior of triangle created by X and its neighbouring vertices in P lies in the interior of $\triangle XVZ,$ and the interior of the triangle created by Y and its neighbouring vertices in P lies in the interior of $\triangle YVZ.$ Thus, X and Y are non-overlapping ears. □

Definition 5.27. A technique that we will use several times is the **ear-clipping** of a generalized $(n+1)$-gon, where the triangle induced by an ear is removed and the diagonal between the neighbouring vertices of the former ear is replaced by a boundary line segment to produce a new generalized n-gon.

As can be imagined, concave polygons can appear quite strange and it can be di□cult to establish theorems about them, so people usually focus on convex polygons. In our experience, it it can even be assumed by laypeople that a polygon refers to a convex polygon. The upcoming results are easier to establish for convex polygons because all of the vertices of a convex polygon are ears. We encourage the reader to develop those proofs independently.

Theorem 5.28 (Sum of interior angles)**.** *The sum of the interior angles of a generalized n-gon is $180\ (n\quad 2).$ In an argument within the proof, we will prove that if an ear V_n of a generalized $(n+1)$-gon $P = V_0V_1 ::: V_n$ is clipped, then the induced triangle $T = V_0V_{n\ \ 1}V_n$ and the generalized n-gon $Q = V_0V_1 ::: V_{n\ \ 1}$ have the same orientation as $P.$*

Proof. We proceed by induction on the number of sides of the generalized polygon. In the base case $n = 3;$ we already know that the result holds for triangles (Theorem 2.18). Suppose the result holds for all generalized n-gons for some $n\quad 3.$ Let P be a generalized $(n+1)$-gon. Then clipping an ear of P results in a generalized n-gon $Q;$ the interior angles of which we know to sum to $180\ (n\quad 2)$ by the induction hypothesis.
Label P as $V_0V_1 ::: V_n$ so that the ear that we clipped is $V_n.$ First we will show that the induced triangle $T = V_0V_{n\ \ 1}V_n$ and $Q = V_0V_1 ::: V_{n\ \ 1}$ have the same orientation as $P.$ In order to have clear notation, we need to label all of the angles involved. Let $\ _0;\ _{n\ \ 1};\ _n$ be the interior angles of T at vertices $V_0; V_{n\ \ 1}; V_n$, respectively. Let $\ _0;\ _1; ::: ;\ _{n\ \ 1}$ be the interior angles of Q at vertices $V_0; V_1; ::: ; V_{n\ \ 1}$, respectively. Let $\ _0;\ _1; ::: ;\ _n$ be the interior angles of P at vertices $V_0; V_1; ::: ; V_n$, respectively. Since V_n is an ear, the interior of T is contained in the interior of $P.$ This means the interiors of angles $\ _n =\ _n$ coincide, and so T has the same orientation as $P.$ Now suppose for contradiction that Q and P have opposite orientations. Then

$$
\begin{aligned}
{}_i \quad {}_i &= 360 \quad {}_i \text{ for } i = 0; n \quad 1;\\
{}_i &= 360 \quad {}_i \text{ for } 1 \quad i \quad n \quad 2.
\end{aligned}
$$

In this case, the "exterior" of Q is a subset of the interior of $P.$ This is a contradiction because the exterior is supposed to be an unbounded region whereas the interior of P is bounded. Thus, Q and P have the same orientation. Then

$$
\begin{aligned}
{}_i + {}_i &= {}_i \text{ for } i = 0; n \quad 1;\\
{}_i &= {}_i \text{ for } 1 \quad i \quad n \quad 2.
\end{aligned}
$$

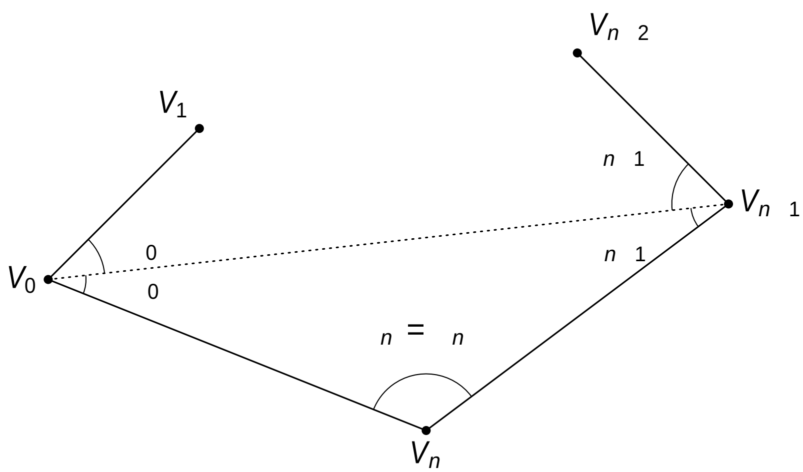
By the induction hypothesis,

$$
\begin{aligned}
180\ (n \quad 2) &= {}_0 + {}_1 + \quad + {}_{n\ \ 2} + {}_{n\ \ 1}\\
&= ({}_0 \quad {}_0) + {}_1 + \quad + {}_{n\ \ 2} + ({}_{n\ \ 1} \quad {}_{n\ \ 1})\\
&= ({}_0 + \quad + {}_{n\ \ 1}) \quad {}_0 \quad {}_{n\ \ 1}\\
&= ({}_0 + \quad + {}_{n\ \ 1}) + {}_n \quad 180\\
&= ({}_0 + \quad + {}_{n\ \ 1} + {}_n) \quad 180 .
\end{aligned}
$$

Therefore, the sum of the interior angles of P is

$${}_0 + \cdots + {}_{n-1} + {}_n = 180^\circ(n-2) + 180^\circ = 180^\circ(n-1).$$

This completes the induction.



□

Problem 5.29. Recall from Definition 2.21 that an **exterior angle** in a convex polygon is an angle supplementary to an interior angle. Find the sum of the exterior angles of any convex polygon is 360°, where we take the sum of one exterior angle corresponding to each interior angle.

Definition 5.30. A polygon is called **equilateral** if its side lengths are all equal. A polygon is called **equiangular** if all of its interior angles are equal. A **regular** polygon is one that is both equilateral and equiangular.

Problem 5.31. Find the measure of an individual interior angle and an individual exterior angle in an equiangular n-gon, where $n \geq 3$ is an integer. What does each approach as $n \to \infty$?

Problem 5.32 (Pick's theorem). A **lattice point** on the Cartesian plane is a point whose coordinates are both integers. A **generalized lattice polygon** is a generalized polygon on the Cartesian plane whose vertices are all lattice points. Let P be a generalized lattice polygon. If B is the number of lattice points on the boundary of P, including its vertices, and I is the number of lattice points in the interior of P, then area of P is given by

$$[P] = I + \frac{B}{2} - 1.$$

Assuming that this works for triangles, prove it for P with more than 3 sides.

Definition 5.33. A **triangulation** of a generalized polygon P is a collection of triangles such that their interiors are non-intersecting and the union of the triangles is P. A triangulation of P is said to have **no extra vertices** if the vertices of the triangles in the triangulation are chosen from only the vertices of P.

Proofs based on clipping an ear and then invoking an induction hypothesis are essentially using a triangulation of a generalized polygon.

Theorem 5.34. Every generalized polygon has a triangulation that has no extra vertices.

Proof. The proof is straightforward using induction and ear-clipping, as follows. In the base case, every triangle is a triangulation of itself. For the induction hypothesis, we assume that every generalized n-gon has a triangulation. Suppose we have a generalized $(n+1)$-gon. Then we clip an ear to produce a generalized n-gon. A triangulation of the generalized $(n+1)$-gon is the combination of the triangulation of the generalized n-gon, given by the induction hypothesis, and the triangle induced by the ear. □

In Volume 2, we determined the number of triangulations of a convex n-gon in terms of n when studying the Catalan numbers.

5.2 Congruence and Similarity

We can extend the notion of congruence from triangles to generalized polygons as follows.

Definition 5.35. Two generalized n-gons $A_0A_1 \ldots A_{n-1}$ and $B_0B_1 \ldots B_{n-1}$ are said to be congruent if corresponding sides A_iA_{i+1} and B_iB_{i+1} are equal for $0 \le i \le n-1$, and corresponding interior angles $A_{i-1}A_iA_{i+1}$ and $B_{i-1}B_iB_{i+1}$ are equal for $0 \le i \le n-1$, where indices are reduced modulo $n.$

Theorem 5.36. Applying any Euclidean isometry to a generalized polygon produces a congruent generalized polygon; translations and rotations preserve the orientation of a polygon, whereas a reflection alters it. Conversely, if there are two congruent generalized polygons in the plane, then one can be transformed into the other by a Euclidean isometry.

Proof. Let $P = V_0V_1 \ldots V_{n-1}$ be a generalized polygon labeled in clockwise orientation, which is without loss of generality because congruence does not depend on the orientation in which vertices are labelled. By Theorem 3.6, we know that if g is a translation, rotation or reflection, then g maps a line segment from z_1 to z_2 to a line segment from $g(z_1)$ to $g(z_2)$ of the same length. Thus, the same is true for all Euclidean isometries, which are their compositions. Moreover, Euclidean isometries are bijections, so the n distinct vertices are mapped to n distinct points, and since the interiors of the n edges of P are pairwise disjoint, the same holds for the n line segments in the image. So every Euclidean isometry does take the generalized polygon $P,$ which is a simple, closed path consisting of line segments along with the interior of the path, to a generalized polygon $P' = V'_0V'_1 \ldots V'_{n-1}$ with the same number of vertices.

Now we need to show that P and P' are congruent. We already know that Euclidean isometries map line segments to line segments of equal length, so we only need to establish that the interior angles at corresponding vertices are of equal measure. For any generalized polygon $A_0A_1 \ldots A_{n-1},$ and for each index $0 \le i \le n-1,$ there exists a rotation around A_i that, along with a positive dilation from $A_i,$ causes $A_{i-1}A_i$ to coincide with $A_{i+1}A_i,$ where indices are reduced modulo $n,$ and where the rotation has the measure of the interior angle

$\angle A_{i-1}A_iA_{i+1}.$ We have previously established that all of these rotations can be chosen to be in the same direction, clockwise or counterclockwise. Since P is oriented clockwise, these rotations are all counterclockwise in it. By Theorem 3.6, translations and rotations map each counterclockwise angle in $[0, 360^\circ)$ to a counterclockwise angle of the same measure, and reflections map each counterclockwise angle in $[0, 360^\circ)$ to a counterclockwise angle of explementary measure. Then, to prove that the interior angle at V_i' has the same measure as the interior angle at V_i, it suffices to show that any translation or rotation of P is oriented clockwise and that any reflection of P is oriented counterclockwise. Let the measures of the interior angles at $V_0, V_1, \ldots, V_{n-1}$ be $\theta_0, \theta_1, \ldots, \theta_{n-1}$ respectively, so that

$$\theta_0 + \theta_1 + \cdots + \theta_{n-1} = 180^\circ(n-2),$$

by Theorem 5.28.

Suppose, for contradiction, that a translation or rotation P' of P is oriented counterclockwise, or a reflection P' of P is oriented clockwise. In any of these cases, the sum of the interior angles of P' would be the sum of the explementary angles of the interior angles of P, which yields the contradictory sum

$$\begin{aligned}(360^\circ - \theta_0) + (360^\circ - \theta_1) + \cdots + (360^\circ - \theta_{n-1}) &= 360^\circ n - 180^\circ(n-2)\\ &= 180^\circ(n+2)\\ &\neq 180^\circ(n-2).\end{aligned}$$

Conversely, suppose there are two congruent generalized polygons $P = V_0V_1 \ldots V_{n-1}$ and $P' = V_0'V_1' \ldots V_{n-1}'$ in the plane. We want to show that a Euclidean isometry maps P to P'. Since we have established that Euclidean isometries map any generalized polygon to a congruent generalized polygon, and that each Euclidean isometry has an inverse that is a Euclidean isometry, it suffices to apply Euclidean isometries to both P and P' until their corresponding vertices coincide. First we translate P and P' so that V_0 and V_0' are at the origin, and then rotate each generalized polygon around the origin until V_1 and V_1' are at the origin. By the previous part of the proof, the reflection of a generalized polygon with clockwise orientation is a congruent generalized polygon with counterclockwise orientation; by the same proof technique of summing the interior angles to derive a contradiction, we can show that reflections turn counterclockwise orientations into clockwise orientations (and that translations and rotations preserve counterclockwise orientations). So, by reflecting across the x-axis if necessary, we can also assume that P and P' are both oriented clockwise. Note that, despite having applied transformations to P and P', we have continued to use the names P and P' along with their vertices V_i and V_i' for the sake of not introducing excessive new notation.

For each index $0 \le i \le n-1$, let the complex number corresponding to V_i be v_i, and let the complex number corresponding to V_i' be v_i'. Now we will show by strong induction that $v_i' = v_i$ for each i. In the base case, $v_0' = 0 = v_0$. Then, since v_1' and v_1 both lie on the positive x-axis and congruence implies $V_0'V_1' = V_0V_1$, we also know that $v_1' = v_1$. Now suppose $v_j' = v_j$ for all indices $0 \le j \le i$ for some index $1 \le i$. We will show that $v_{i+1}' = v_{i+1}$, where indices are reduced modulo n. Let the interior angles at each of V_i and V_i' be θ_i. Since P and P' are oriented clockwise, we compute

$$\begin{aligned}\frac{v_{i+1}-v_i}{v_{i-1}-v_i} &= \frac{|v_{i+1}-v_i|}{|v_{i-1}-v_i|}e^{i\theta_i} = \frac{V_{i+1}V_i}{V_{i-1}V_i}e^{i\theta_i};\\ \frac{v'_{i+1}-v'_i}{v'_{i-1}-v'_i} &= \frac{|v'_{i+1}-v'_i|}{|v'_{i-1}-v'_i|}e^{i\theta_i} = \frac{V'_{i+1}V'_i}{V'_{i-1}V'_i}e^{i\theta_i};\end{aligned}$$

where there is a distinction between the imaginary number i and the index $i.$ By congruence, $V_{i+1}V_i = V'_{i+1}V'_i$ and $V_{i-1}V_i = V'_{i-1}V'_i;$ so

$$\frac{v'_{i+1}-v'_i}{v'_{i-1}-v'_i} = \frac{v_{i+1}-v_i}{v_{i-1}-v_i}.$$

By invoking the induction hypothesis on the previous two instances,

$$\begin{aligned}v'_{i+1} &= \frac{v_{i+1}-v_i}{v_{i-1}-v_i}(v'_{i-1}-v'_i)+v'_i\\ &= \frac{v_{i+1}-v_i}{v_{i-1}-v_i}(v_{i-1}-v_i)+v_i\\ &= v_{i+1}.\end{aligned}$$

This completes the induction, showing that all corresponding vertices can be made to coincide. □

Like congruence, we can extend the notion of similarity from triangles to generalized polygons.

Definition 5.37. Two generalized n-gons are said to be **similar** if the ratio of any pair of corresponding sides $\frac{A_iA_{i+1}}{B_iB_{i+1}}$ equal to the same **similarity ratio** k for $0 \le i \le n-1$, and corresponding interior angles $A_{i-1}A_iA_{i+1}$ and $B_{i-1}B_iB_{i+1}$ are equal for $0 \le i \le n-1$, where indices are reduced modulo $n.$ Note that congruence is a special case of similarity when we take $k = 1$ as the similarity ratio.

Theorem 5.38. Applying any similarity transformation to a generalized polygon P produces a similar generalized polygon $P'.$ In particular, applying only a homothety of factor k results in the similarity ratio $|k|$ of the lengths of P' to the lengths of $P.$ Moreover, a homothety (positive or negative) preserves the orientation of the generalized polygon. Conversely, if there are two similar generalized polygons P and P' in the plane, then P can be transformed into P' by a similarity transformation. Specifically, if the similarity ratio of the lengths of P' to the lengths of P is $k;$ then we can choose a similarity transformation of P to P' that consists of exactly one homothety of factor k from the origin, followed by a Euclidean isometry. Thus, the term "similarity transformation" is justified.

Proof. For the first direction, since we already know that Euclidean isometries map each generalized polygons to a congruent generalized polygon, it suffices to prove the assertion for only homotheties. Moreover, every homothety is a homothety from the origin sandwiched between inverse translations, so it suffices to work with only homotheties from the origin.

Let $P = V_0V_1 \dots V_{n-1}$ be a generalized polygon labeled in clockwise orientation without loss of generality, and let h be a homothety of factor $k.$ We know that h maps a line segment of length l from z_1 to z_2 to a line segment of length $|k| \cdot l$ from $h(z_1)$ to $h(z_2).$ Moreover, homotheties are bijections, so the n distinct vertices are mapped to n distinct points, and since the interiors of the n edges of P are pairwise disjoint, the same holds for the n line segments in the image. So h does take the generalized polygon P to a generalized polygon $P' = V_0'V_1' \dots V_{n-1}'$ with the same number of vertices.
Now we need to show that P and P' are similar with a similarity ratio $|k|$ of the lengths of P' to the lengths of $P.$ We have already stated that h maps a line segment of length l to a line segment of length $|k| \cdot l,$ so we only need to show that the interior angles at corresponding vertices are of equal measure. By Theorem 3.6, homotheties map each counterclockwise angle in $[0, 360^\circ)$ to a counterclockwise angle of the same measure. Then, to prove that the interior angle at V_i' has the same measure as the interior angle at $V_i,$ it suffices to show that P' is oriented clockwise. Supposing for contradiction that P' is oriented counterclockwise, the proof is then the same as for Euclidean isometries, where this assumption leads to showing that the sum of the interior angles of P' is $180^\circ(n+2),$ which is not $180^\circ(n-2).$ Likewise, we can prove that a homothety takes a counterclockwise oriented generalized polygon to the same.
Conversely, suppose there are two similar generalized polygon

$$P = V_0V_1 \dots V_{n-1},$$
$$P' = V_0'V_1' \dots V_{n-1}'$$

in the plane such that the similarity ratio of the lengths of P' to the lengths of P is $k.$ First we apply a homothety to P from the origin by a factor of k to produce a generalized polygon P^* that is congruent to $P'.$ Then we know that there exists a Euclidean isometry that maps P^* to P' by the preceding theorem. □

Example 5.39. For any non-zero complex number $w,$ the function $s_w : \mathbb{C} \to \mathbb{C},$ defined by $s_w(z) = zw,$ is called a **spiral similarity**. Show that this name is sensible.

Solution. Let $\arg(w) = \theta.$ Then

$$s_w(z) = zw = z \cdot |w| \cdot e^{i\theta}.$$

This means s_w is the composition of a homothety from the origin by a factor of $|w|$ and a counterclockwise rotation around the origin by $\theta.$ We call this a "spiral similarity" because the rotation is a *spiral* that maps a generalized polygon to a congruent generalized polygon and the homothety maps a generalized polygon to a *similar* generalized polygon. □

Chapter 6

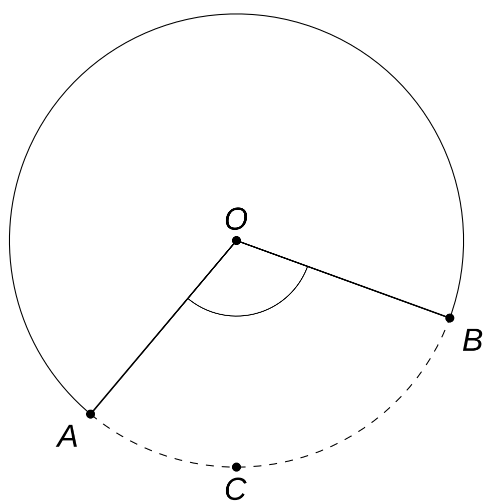
Circles I

> "Noli turbare circulos meos! (Do not disturb my circles!)"
> (in response to invading soldiers)
>
> – *Archimedes*

Circles are easy to define, yet the geometry involved with circles can easily become complicated. We will first study the inscribed angle theorem. Then, we will study theorems about tangents, secants, and chords in circles, specifically their interactions with angles.

6.1 Inscribed Angles

Definition 6.1. If an angle $\angle AOB$ is drawn such that O is the center of a circle and AO and BO are radii of the circle, then $\angle AOB$ is called a **central angle**. Several further pieces of terminology come into play:



- The part of the circle that lies inside the central angle, including endpoints, is called an **arc**, and is denoted by $\overset{\frown}{AB}$.
- Every pair of points A and B on the circle defines two arcs between them, which we call **opposite arcs**. If it is clear to which of the two arcs we are referring, we call the arc **well-defined**. To make it clear to which arc we are referring, we can pick a point C on the arc and denote the arc as $\overset{\frown}{ACB}$.

- If it is ambiguous whether we are working with a non-flat central angle or its explementary angle, the **minor arc** is the arc induced by the non-reflex central angle and the **major arc** is the arc induced by the reflex central angle.

- A **sector** AOB is the part of the closed disk that lies inside $\angle AOB$ and includes the two boundary radii OA and OB. The corresponding **circular segment** is the sector AOB without $\triangle AOB$ but including the segment AB.

- The **measure** of the arc $\overset{\frown}{AB}$ or the sector AOB is the measure of the central angle $\angle AOB$ within which the arc or sector lies. If the measure of a central angle is m in radians, then the fraction of the circle it encompasses is $\frac{m}{2\pi}$ in radians. So the length a of the corresponding arc is found by the proportion

$$\frac{a}{2\pi r} = \frac{m}{2\pi} \Longrightarrow a = rm$$

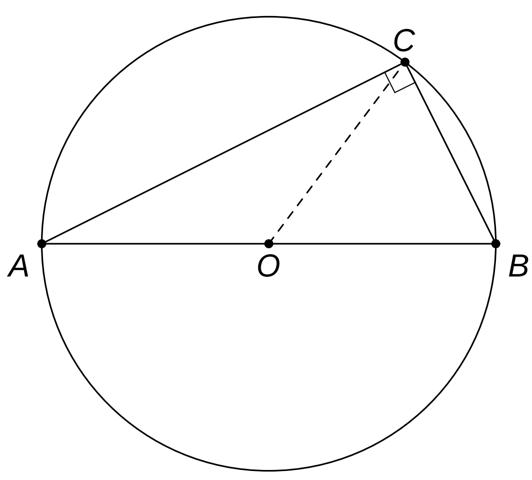
  and the area A of the corresponding sector is found by the proportion

$$\frac{A}{\pi r^2} = \frac{m}{2\pi} \Longrightarrow A = \frac{r^2 m}{2}.$$

  Both of these formulas are for angles m in radian measure, which can be converted to degrees, as necessary.

Theorem 6.2 (Thales's theorem). *Let A, B, C be distinct points on a circle. Then $\angle ACB$ is a right angle if and only if AB is a diameter.*

Proof. First, we make some preliminary remarks. Let O be the center of the circle. Then $AO = BO = CO$, which implies that $\triangle AOC$ and $\triangle BOC$ are isosceles with the vertex angles of both triangles at O. Below, all angles mentioned are interior angles of their respective triangles.



For one direction, suppose AB is a diameter. Then the midpoint of AB is O, so

$$\angle AOC + \angle BOC = 180^\circ.$$

This allows us to calculate

$$\begin{aligned}\angle ACB &= \angle ACO + \angle BCO \\ &= \frac{180^\circ - \angle AOC}{2} + \frac{180^\circ - \angle BOC}{2} \\ &= 180^\circ - \frac{\angle AOC + \angle BOC}{2} \\ &= 90^\circ .\end{aligned}$$

Conversely, suppose $\angle ACB = 90^\circ$. Then

$$\begin{aligned}\angle ACO + \angle BCO &= \angle ACB \\ &= 90^\circ .\end{aligned}$$

This allows us to compute that

$$\begin{aligned}\angle AOC + \angle BOC &= (180^\circ - 2\angle ACO) + (180^\circ - 2\angle BCO) \\ &= 360^\circ - 2(\angle ACO + \angle BCO) \\ &= 180^\circ .\end{aligned}$$

Thus, A, O, B are collinear, making AB a diameter. □

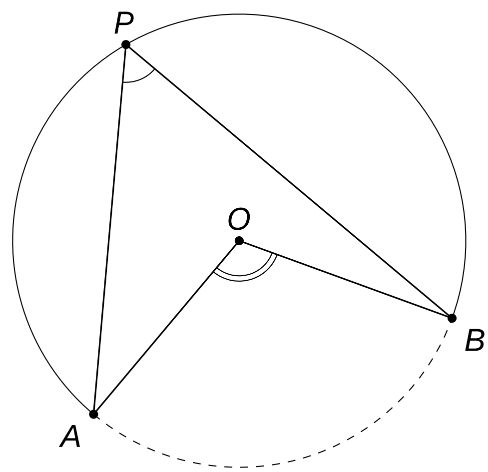
Problem 6.3. Prove the following two results:

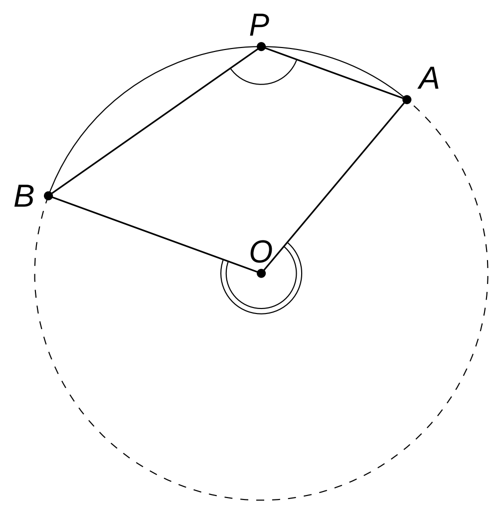
1. Suppose ABC is a triangle. Let M be the midpoint of AB. Then $\angle ACB = 90^\circ$ if and only if $MC = \frac{AB}{2}$.
2. Suppose UV is a diameter of a circle and W is a point on the plane that is distinct from U and V. Then $\angle UWV = 90^\circ$ if and only if W lies on .

Definition 6.4. Given some criteria to be satisfied by points in the plane, the set of all points that fulfil the criteria is called their **locus**. Proving that a set is the desired locus typically involves proving two set inclusions: showing that every point in the set fulfils the criteria, and showing that every point that fulfils the criteria lives in the set.

The second part of the preceding problem is generalized by the next result, which is called the inscribed angle theorem. In one direction, the inscribed angle theorem establishes an equality of angles, and the converse shows that a point lies a circle. Together, they form a locus result.

Definition 6.5. If the points A, P, B lie on a circle, then the **inscribed angle** $\angle APB$ is the one chosen between the two explementary angles at P so that the $\angle APB$ lies on the side of the interior of the circle. Letting $\overset{\frown}{AB}$ be the arc between A and B on which P does not lie, we say that $\overset{\frown}{AB}$ **subtends** $\angle APB$, or that $\angle APB$ is subtended by $\overset{\frown}{AB}$; in the other direction, we say that $\angle APB$ **intercepts** $\overset{\frown}{AB}$, or that $\overset{\frown}{AB}$ is intercepted by $\angle APB$. In other words, arcs subtend angles and angles intercept arcs.

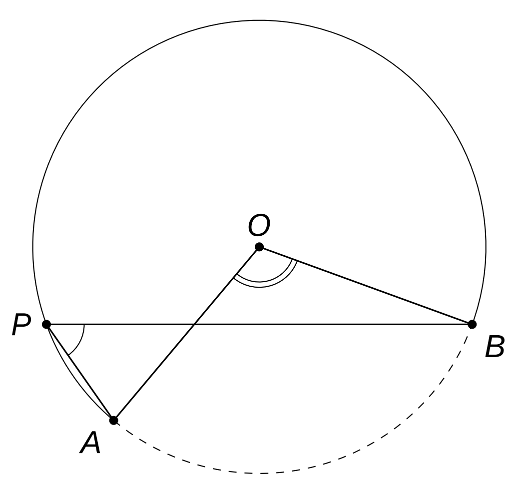




Note that all inscribed angles are non-reflex as the angle lies entirely on one side of the tangent through its vertex.

Theorem 6.6 (Inscribed angle theorem). Let A and B be distinct points on a circle, and P be a point in the plane distinct from A and B; and let $\overset{\frown}{AB}$ be a well-defined arc between A and B. Then P lies on the opposite arc between A and B if and only if the non-reflex $\angle APB$ has half the measure of $\overset{\frown}{AB}$ and P lies on the side of the line through AB that does not include $\overset{\frown}{AB}$: This holds regardless of whether $\overset{\frown}{AB}$ is a minor arc, a semicircle, or a major arc.

Proof. The usual synthetic proof breaks the argument up, based on the configurations and uses triangle geometry to tackle each configuration of each direction. The thorniest configuration is drawn below for illustrative purposes. We provide a unified proof of all cases of the inscribed angle theorem and its converse using complex numbers. The reader is encouraged to develop the standard proofs independently.



First we will set up an equation that will be used for proving both directions. In both the theorem and its converse, P is a point that lies on the side of the line through AB that does not include $\overset{\frown}{AB};$ as this is explicitly assumed in one direction and is implicit in the other direction due to P lying on the opposite arc. Without loss of generality, we assume that the center of the circle is the origin 0 and that $A; B; P$ lie in counterclockwise, in that order (the clockwise order leads to the same result). Let the complex numbers corresponding to $A; B; P$ be $a; b; p$ respectively. If the radius of the circle is r then $a = re^{i}$ and $b = re^{i}$ where the angles and are the arguments of a and b respectively. Let $p = se^{i}$ where s is the distance of p from the origin and is its argument. We want to investigate the non-reflex angle $\backslash APB;$ which is the argument of $\frac{b \quad p}{a \quad p}:$ Letting t be the modulus of $\frac{b \quad p}{a \quad p};$ we get the equation

$$te^{i} = \frac{b \quad p}{a \quad p} = \frac{re^{i} \quad se^{i}}{re^{i} \quad se^{i}}:$$

Taking the conjugate of both sides of this equation yields

$$te^{i(\quad)} = \frac{re^{i(\quad)} \quad se^{i(\quad)}}{re^{i(\quad)} \quad se^{i(\quad)}} = \frac{e^{i} \ (se^{i} \quad re^{i} \ )}{e^{i} \ (se^{i} \quad re^{i} \ )}:$$

To get a grasp on 2 , we divide the original equation by the conjugate equation to get

$$e^{i(2 \ )} = e^{i(\quad)} \ \frac{(se^{i} \quad re^{i} \ )(re^{i} \quad se^{i} \ )}{(se^{i} \quad re^{i} \ )(re^{i} \quad se^{i} \ )}:$$

Now we approach each direction separately.

1. In proving the inscribed angle theorem, we are assuming that P lies on the circle, so $s = r;$ which reduces the above equation to $e^{i(2 \ )} = e^{i(\quad)}:$ This implies 2 $:$ Since is the measure of $\overset{\frown}{AB};$ all we need is for this congruence to be an actual equality. Since is an inscribed angle, it is non-reflex, and we also know that is a central angle. Thus, 2 and both lie in $(0; 2 \ )$, which makes the congruence an equation.

2. For the converse, we are assuming that 2 $=$ $:$ As a result, $e^{i(2 \ )} = e^{i(\quad)};$ which reduces our equation in the preamble to

$$(se^{i} \quad re^{i} \ )(re^{i} \quad se^{i} \ ) = (se^{i} \quad re^{i} \ )(re^{i} \quad se^{i} \ ):$$

   Taking everything to one side, expanding and factoring yields

$$\begin{aligned} 0 &= (se^{i} \quad re^{i} \ )(re^{i} \quad se^{i} \ ) \quad (se^{i} \quad re^{i} \ )(re^{i} \quad se^{i} \ ) \\ &= (r^2 \quad s^2)e^{i} \ (e^{i} \quad e^{i} \ ): \end{aligned}$$

   It is not possible for e^{i} to be $0;$ and since $a \neq b$ we also know that $e^{i} \neq e^{i}$ $:$ Thus, the only possibility is that $r = s;$ which puts P on the circle. Moreover, P lies on the arc opposite to $\overset{\frown}{AB}$ because we have assumed that P lies on the side of the line through AB that does not include $\overset{\frown}{AB}:$

In practical usage, it is common to come across two inscribed angles that intercept the same arc. As a result, they have a shared central angle, and therefore the inscribed angles have the same measure. □

Theorem 6.7 (Extended law of sines)**.** Given $\triangle ABC$ with edges a, b, c opposite to vertices A, B, C respectively and circumradius $R,$

$$\frac{a}{\sin A} = \frac{b}{\sin B} = \frac{c}{\sin C} = 2R.$$

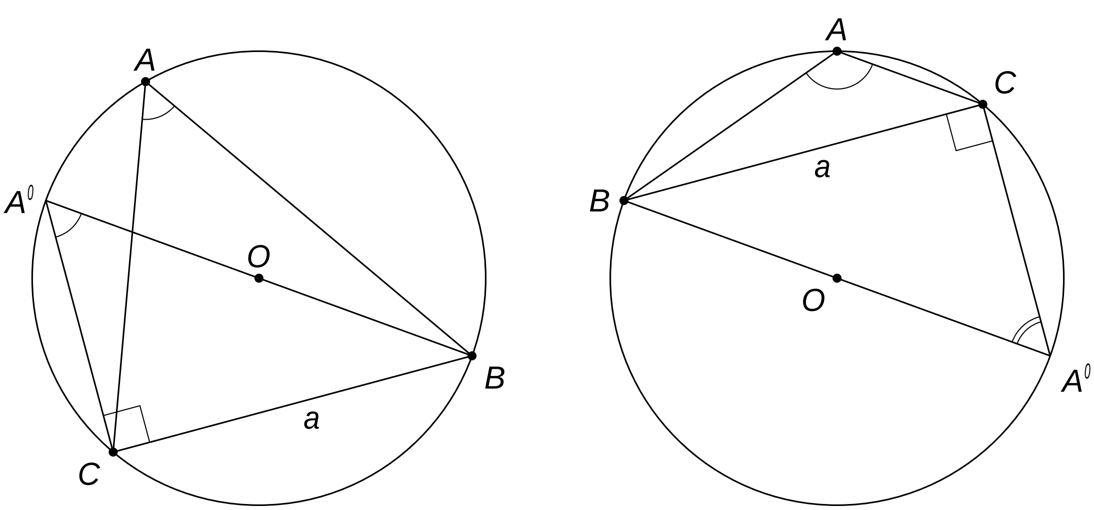
This is called the "extended" law of lines because the ordinary version does not state the final equality with $2R.$ For ease of notation, we have dropped the angle signs and instead referred to each interior angle by its vertex.

Proof. We will show that $\sin A = \frac{a}{2R}$ and the other equations will follow because the argument will not rest on any special property of A or $a.$ Let A' be the other endpoint of the diameter through B of the circumcircle. Then there are three possible configurations:

1. If A' does not coincide with C and lies on the arc $\overset{\frown}{BAC},$ then $\triangle BA'C$ is right with its right angle at C since BA' is a diameter, and $\angle BAC = \angle BA'C$ by the inscribed angle theorem. Then
$$\sin A = \sin \angle BA'C = \frac{BC}{BA'} = \frac{a}{2R}.$$

2. If A' coincides with $C,$ then BC is a diameter and $\triangle BAC$ is a right triangle with its right angle at $A.$ Then
$$\sin A = \sin \frac{\pi}{2} = 1 = \frac{BC}{BC} = \frac{a}{2R}.$$

3. If A' does not coincide with C and lies on the arc $\overset{\frown}{BC}$ that does not contain $A,$ then $\triangle BA'C$ is right with its right angle at C since BA' is a diameter, and $\angle BAC = \pi - \angle BA'C$ by the inscribed angle theorem. Then
$$\sin A = \sin(\pi - \angle BA'C) = \sin \angle BA'C = \frac{BC}{BA'} = \frac{a}{2R}.$$



□

Problem 6.8 (Law of tangents). Given $\triangle ABC$ with edges a, b opposite to vertices A, B respectively, show that

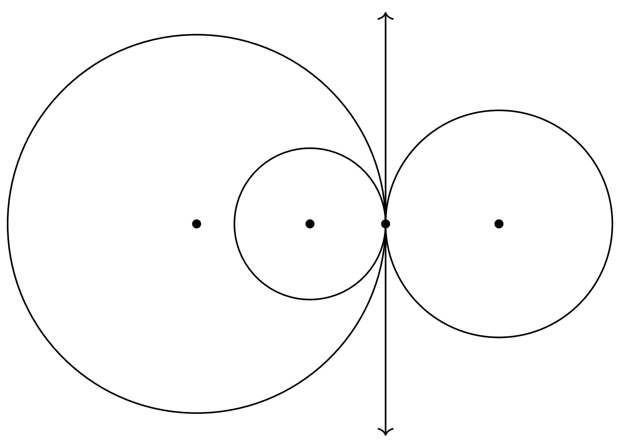
$$\frac{\tan\left(\frac{A-B}{2}\right)}{\tan\left(\frac{A+B}{2}\right)} = \frac{a-b}{a+b}.$$

Although this "law" is not as useful as the cosine law or sine law, it is a fun problem.

6.2 Tangents, Secants, and Chords

Definition 6.9. There are two notions of tangency that we will need:

- A **tangent line**, which is also just called a **tangent**, to a circle is a line in the plane that intersects the circle at exactly one point.
- Two circles are said to be **tangent** if they intersect at precisely one point. Two circles are **externally** tangent if the interior of each circle is in the exterior of the other, or they are **internally** tangent if the interior of one circle is contained in the interior of the other.



Theorem 6.10. If ℓ is a tangent to a circle then, other than the point of tangency, it lies in the exterior of the circle.

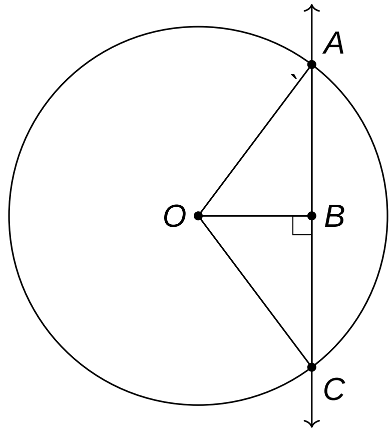
Proof. Let the point of tangency be A. Suppose, for contradiction, that P is a point on ℓ that is inside the circle. If ℓ runs through the center O of the circle, then A is a part of a diameter, and the existence of the other endpoint of the diameter would imply that ℓ intersects the circle at two distinct points. So we can assume that ℓ does not run through O. Let B be the foot of the perpendicular from O to ℓ. If $B = P$, then relabel P as B for an upcoming step in the argument. Otherwise, $\triangle OPB$ is a right triangle that is not degenerate with a right angle at B and hypotenuse OP; then

$$OP = \sqrt{OB^2 + BP^2} > \sqrt{OB^2} = OB$$

and so OB also lies inside the circle. In either case, extend AB from A through B to C, where C is defined to satisfy $AB = BC$. By the Pythagorean theorem,

$$OC = \sqrt{OB^2 + BC^2} = \sqrt{OB^2 + AB^2} = OA.$$

So $OC = OA$ is a radius, showing that C is a point on ℓ that is distinct from A yet lies on the circle. This contradicts the fact that ℓ touches the circle only once.



□

Theorem 6.11. If a line is tangent to a circle, then the line is perpendicular to the radius that touches the point of tangency. Conversely, if a line intersects a circle at some point and the radius that touches that point of intersection is perpendicular to the line, then the line is tangent to the circle.

Proof. Suppose ℓ is a line that is tangent to a circle with center O. Let A be the point of tangency. The radius OA has the endpoint A on ℓ. Now suppose, for the sake of contradiction, that OA is not perpendicular to ℓ. Then let the foot of the perpendicular from O to ℓ be B. Then B is on the tangent line, so it is outside the circle, by Theorem 6.10. So $\triangle OBA$ is a right triangle with a right angle at B. Then the hypotenuse OA is greater than OB, which contradicts the fact that OB should be larger than the radius OA, since B is outside the circle. Thus, ℓ is perpendicular to OA.
In the other direction, suppose ℓ is a line that intersects a circle at A and that the radius OA is perpendicular to ℓ. Let P be any point on ℓ other than A. Then OPA is a right triangle with a right angle at A. By the Pythagorean theorem,

$$OP = \sqrt{OA^2 + AP^2} > \sqrt{OA^2} = OA,$$

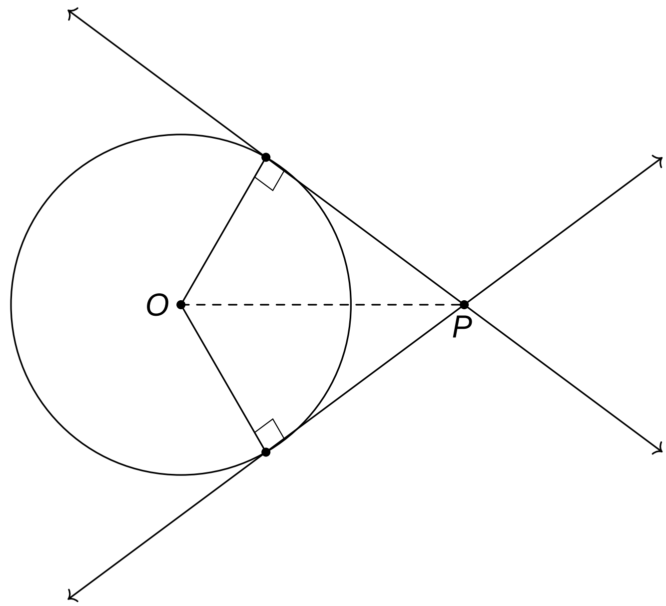
meaning P lies outside the circle. Thus, ℓ intersects the circle only at A, and so ℓ is a tangent. □

Theorem 6.12. Suppose we are given a circle and a point P in the plane outside . For each tangent from P to , the length of the segment from P to the point of tangency is equal to the others. Subsequently, there are exactly two lines that can be drawn through P that are tangent to .

Proof. Let P be any point outside , let O be the center of , and let A be a point on such that the line running through PA is tangent to . Then OA is perpendicular to PA. Since PO has a fixed length and AO is a radius which also has a constant length across all radii, the Pythagorean theorem tells us that

$$PA = \sqrt{PO^2 - OA^2}$$

has a fixed length, which proves the first claim.



For the second claim, a process similar to the one in Theorem 6.14 may be used to solve for precisely two tangents. We encourage the reader to follow through with a derivation. □

Problem 6.13. Prove the following:

1. For two internally or externally tangent circles, the point of tangency of the circles and the centers of the two circles are collinear.

2. For two internally or externally tangent circles, the common tangent line through the point of tangency of the circles is perpendicular to the line through the centers of the circles.

Theorem 6.14. Two circles are said to be in general position if they are external to each other and do not intersect. Given any two circles in general position, there exist four tangents that are common to both circles. These are called bitangents.

Proof. Let the circle with radius r_1 have center (x_1, y_1) and the circle with radius r_2 have center (x_2, y_2). Let a bitangent ℓ be given by

$$ax + by + c = 0.$$

Since a and b cannot simultaneously be 0, we can divide both sides by the non-zero number $\sqrt{a^2 + b^2}$, which allows us to assume that $a^2 + b^2 = 1$ without loss of generality. The perpendicular distance between ℓ and (x_1, y_1) is r_1 and the perpendicular distance between ℓ and (x_2, y_2) is r_2. By the point-line distance formula (Theorem 2.26),

$$|ax_1 + by_1 + c| = r_1,$$
$$|ax_2 + by_2 + c| = r_2,$$

since $a^2+b^2=1$. Opening up the absolute values result in $\pm$ signs on the r_1 and r_2, where the two $\pm$ signs are independent of each other. Given a triple (a, b, c) such that $ax+by+c=0$ represents ℓ and $a^2+b^2=1$, the triple $(-a, -b, -c)$ works as well. So we engineer (a, b, c) to ensure that

$$ax_2+by_2+c=r_2.$$

The other equation is

$$ax_1+by_1+c=\pm r_1.$$

We wish to show that exactly four distinct such triples (a, b, c) exist. Subtracting the second equation from the first yields

$$a(x_2-x_1)+b(y_2-y_1)=r_2\mp r_1.$$

Letting

$$\begin{aligned} d &= \sqrt{(x_2-x_1)^2+(y_2-y_1)^2}, \\ (x_0, y+0, r_0) &= \left(\frac{x_2-x_1}{d}, \frac{y_2-y_1}{d}, \frac{r_2\mp r_1}{d}\right), \end{aligned}$$

and dividing both sides of the equation by d, we get

$$ax_0+by_0=r_0,$$

where it is true that

$$\begin{aligned} a^2+b^2 &= 1, \\ x_0^2+y_0^2 &= 1. \end{aligned}$$

Our job is to solve for a and b in terms of x_0, y_0, and r_0. We do this as follows:

$$\begin{aligned} ax_0+by_0 &= r_0 \\ by_0 &= r_0-ax_0 \\ b^2y_0^2 &= (r_0-ax_0)^2 \\ &= r_0^2-2r_0ax_0+a^2x_0^2 \\ (1-a^2)y_0^2 &= r_0^2-2r_0ax_0+a^2x_0^2 \\ a^2(x_0^2+y_0^2)-2r_0x_0a+(r_0^2-y_0^2) &= 0 \\ a^2-2r_0x_0a+(r_0^2-y_0^2) &= 0. \end{aligned}$$

By the quadratic formula,

$$\begin{aligned} a &= \frac{2r_0x_0\pm\sqrt{(2r_0x_0)^2-4(r_0^2-y_0^2)}}{2} \\ &= r_0x_0\pm\sqrt{r_0^2x_0^2-r_0^2+y_0^2} \\ &= r_0x_0\pm\sqrt{r_0^2(x_0^2-1)+y_0^2} \\ &= r_0x_0\pm\sqrt{-r_0^2y_0^2+y_0^2} \\ &= r_0x_0\pm y_0\sqrt{1-r_0^2}. \end{aligned}$$

Then

$$
\begin{aligned}
b &= \frac{r_0 - ax_0}{y_0} \\
&= \frac{r_0 - (r_0x_0 \pm y_0\sqrt{1 - r_0^2})x_0}{y_0} \\
&= \frac{r_0 - r_0x_0^2 \mp x_0y_0\sqrt{1 - r_0^2}}{y_0} \\
&= \frac{r_0(1 - x_0^2) \mp x_0y_0\sqrt{1 - r_0^2}}{y_0} \\
&= \frac{r_0y_0^2 \mp x_0y_0\sqrt{1 - r_0^2}}{y_0} \\
&= r_0y_0 \mp x_0\sqrt{1 - r_0^2}.
\end{aligned}
$$

Therefore, we obtain

$$
\begin{aligned}
a &= r_0x_0 \pm y_0\sqrt{1 - r_0^2}, \\
b &= r_0y_0 \mp x_0\sqrt{1 - r_0^2}, \\
c &= r_2 - (ax_2 + by_2),
\end{aligned}
$$

where there $\pm$ and $\mp$ signs are opposites, and the $\pm$ sign in $r_0 = \frac{r_2 \pm r_1}{d}$ is independent of the other two signs. Thus, there are four 2×2 solutions, due to there being two independent signs. Finally, these four solutions all exist, regardless of which of the two possibilities for r_0 is taken, since $\sqrt{1 - r_0^2}$ exists if and only if

$$
\begin{aligned}
1 - r_0^2 &> 0 \\
1 &> \left(\frac{r_2 \pm r_1}{d}\right)^2 \\
d &> |r_2 \pm r_1|.
\end{aligned}
$$

This is true because the circles are purely external to each other, leading to

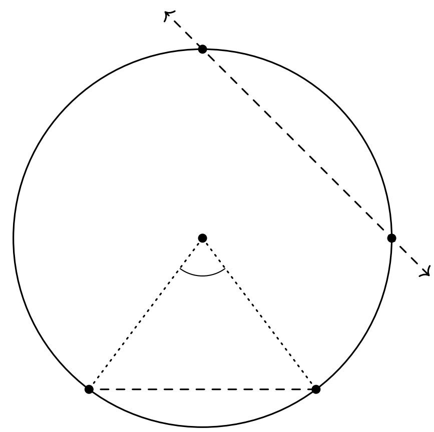
$$d > r_2 + r_1 > |r_2 - r_1|.$$

□

Definition 6.15. A **chord** of a circle is a line segment with both endpoints on the circle. In particular, every diameter is a chord. Note that the interior of a chord lies entirely in the interior of the circle. This leads to more definitions:

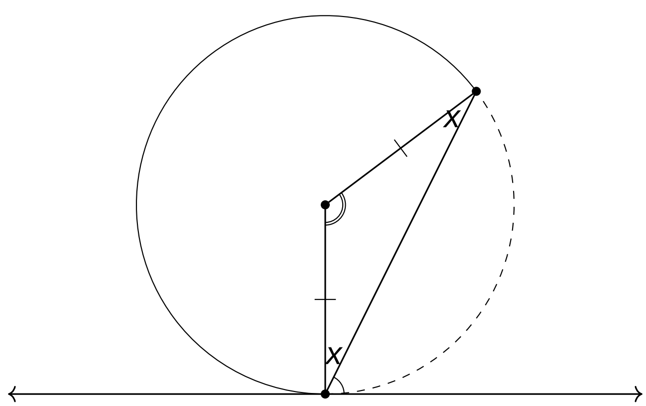
- Since two radii can be drawn to the endpoints of a chord, we call the arc intercepted by the resulting central angle that contains the chord (which is necessarily non-reflex) the **intercepted arc** of the chord. In the case that the chord is a diameter, the intercepted arc could be either semicircle.

- A secant of a circle is a line in the plane that intersects the circle at exactly two points. In other words, a secant is like a chord that extends infinitely in both directions.



We need the following lemma as a precursor to the multi-part result that is Theorem 6.18, which expand the inscribed angle theorem, though the lemma could be considered to be a part of that set of results.

Lemma 6.16 (Chord-tangent angle theorem). Suppose a chord of a circle meets a tangent line to the circle at a non-right angle at one of the endpoints of the chord. Then the acute angle between them is equal to half the measure of the intercepted arc of the chord. Consequently, the obtuse angle between the chord and the tangent line is equal to half the measure of the arc opposite to the intercepted arc of the chord. Alternatively, if the chord and the tangent line meet at a right angle, then the chord is a diameter and both arcs that are cut off by the chord are semicircles. In general, either angle created by the chord and tangent is equal to half the measure of the arc on its side.



Proof. Let the acute angle between the chord and the tangent line passing through the chord measure *:* Let the measure of the intercepted arc of the chord be *:* Drawing the radii to the endpoints of the chord creates an isosceles triangle with base angles x*:* Then

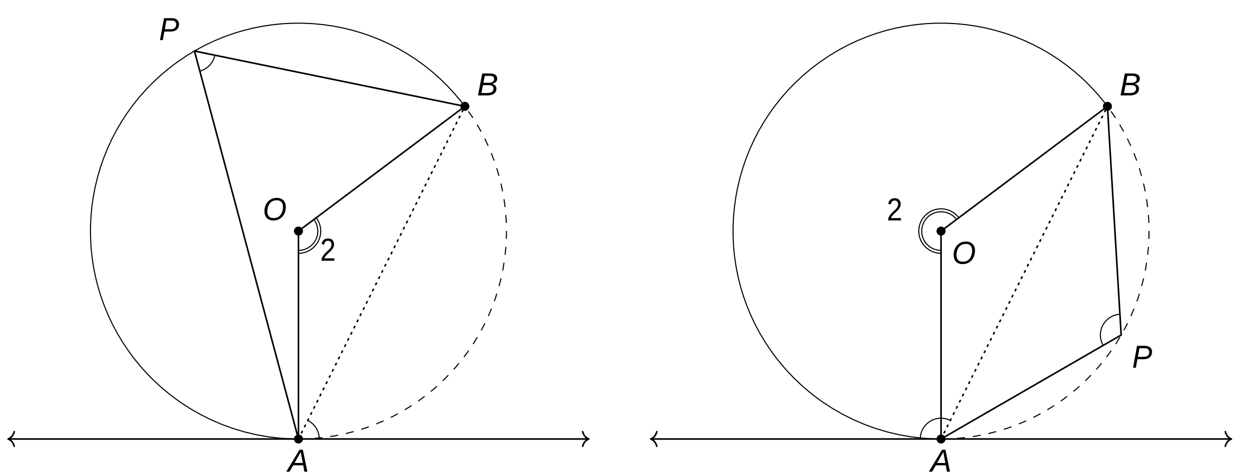
$$+ 2x = 180 \quad = 2 \quad 90 \quad = 2 \quad (x + \quad):$$

It follows that = 2 : The second part of the lemma is true because = 2 implies

$$360 \quad = 2 \ (180 \quad):$$

In the other case, suppose the chord and tangent meet at a right angle. Since a radius goes through the point of tangency at a right angle as well, the same line runs through the chord and the radius. Thus, the chord is the diameter resulting from extending the radius. Since a diameter cuts o two arcs that are both semicircles, we are done. □

Problem 6.17. Suppose a chord AB of a circle meets a tangent line to the circle at A at a non-right angle. Show that if an inscribed angle \ APB has P lying on the major arc $\overset{\frown}{AB}$ then the measure of \ APB is the same as the measure of the acute angle between the chord and the tangent. Analogously, show that if an inscribed angle \ APB has P lying on the minor arc $\overset{\frown}{AB}$ then the measure of \ APB is the same as the measure of the obtuse angle between the chord and the tangent.



The inscribed angle theorem can be thought of as providing a relationship between an angle between two lines emanating from a point P on a circle and the measure of the arc inside the angle. The inscribed angle theorem can be used to extend itself to points P inside and outside the circle as follows.

Theorem 6.18. In addition Lemma 6.16, which is a chord-tangent arc theorem, there are four more angle theorems about intersecting chords, secants and tangents:

1. Chord-chord: Two chords that intersect inside the circle create two pairs of vertical angles and cut o four arcs. Picking one pair of vertical angles, the measure of either of those angles is the average of the measures of the two arcs that they contain.

2. Secant-secant: Two secants that intersect outside the circle form an angle that cuts o four arcs, two of which are contained in the angle. The measure of the angle created by the two secants is half the positive di erence of the measures of the two arcs that lie inside this angle.

3. Secant-tangent: A secant and a tangent that intersect outside the circle form an angle that cuts off three arcs, two of which are contained in the angle. The measure of the angle created by the secant and the tangent is half the positive difference of the measures of the two arcs that lie inside this angle.

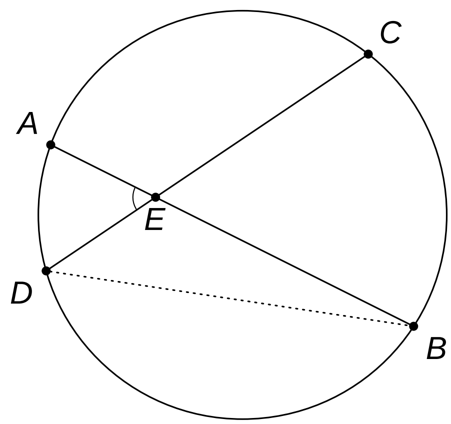
4. Tangent-tangent: Two tangents that intersect outside the circle form an angle that cuts off two arcs, both of which are contained in the angle. The measure of the angle created by the two tangents is half the positive difference of the measures of the two arcs.

Note that there is no similar chord-secant theorem that is meaningfully different from the scenarios that have already been mentioned.

Proof. Let the angles and arc measures be labelled as above.

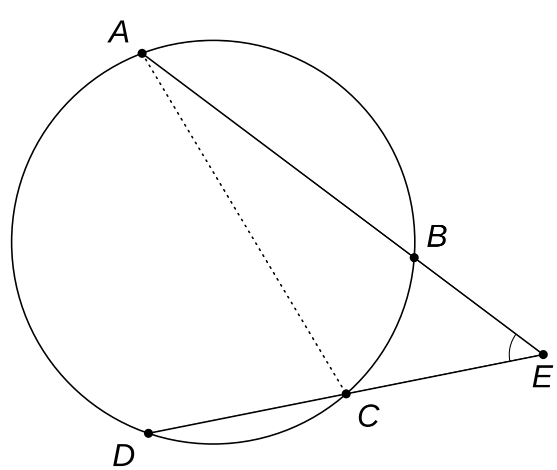
1. By the inscribed angle theorem, $= 2$ and $= 2$. Since is an exterior angle to the triangle drawn, $= +$. As a result,

$$= + = \frac{+}{2}:$$



2. Since is an exterior angle to the triangle drawn, $= +$. Then

$$= = \frac{}{2}:$$

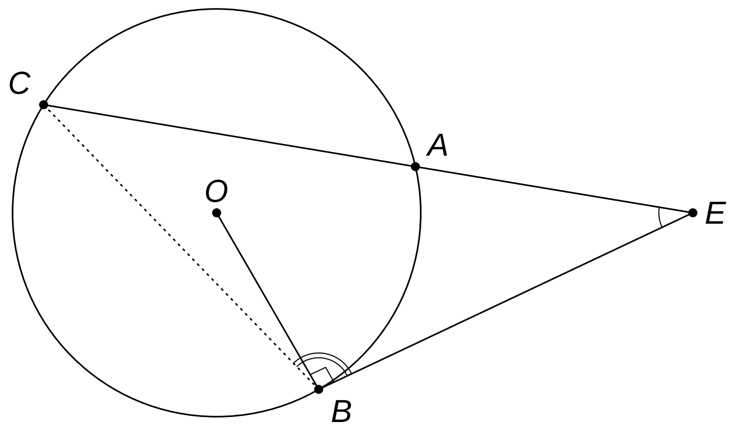


3. By Lemma 6.16,

$$= \frac{360}{2}:$$

By the inscribed angle theorem, $= \frac{}{2}$. Then

$$= 180 \quad = 180 \quad \frac{}{2} \quad \frac{360}{2} = \frac{}{2}:$$

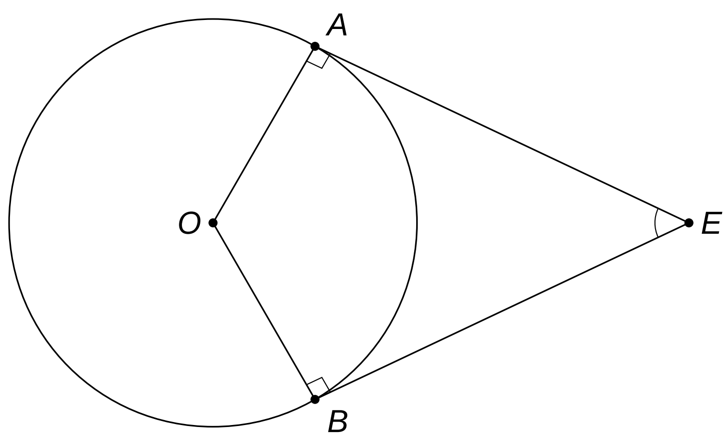


4. In quadrilateral $OAEB$, the angles add up 360 , so

$$= 360 \quad 2 \quad 90 \quad = 180 \quad :$$

The arcs and make a partition of the circle, so

$$= 180 \quad = \frac{(360 \quad)}{2} = \frac{}{2}:$$



□

Problem 6.19. Suppose two parallel lines intersect a circle, each at one or two points. So they are both secants, both tangents, or one is a secant while the other is a tangent. Show that the arcs cut o in between them have equal measure.

Problem 6.20. Show that two chords of the same circle have equal length if and only if their intercepted arcs have equal measure.

Chapter 7

Classifications

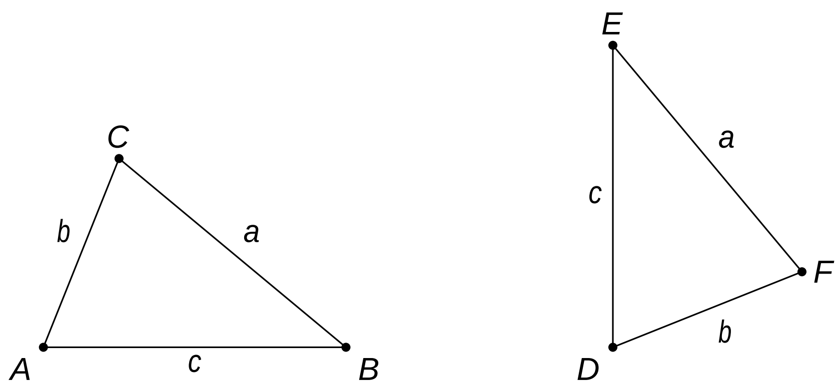
> "Give him a coin if he must profit from what he learns."
> (in response to a slave who asked what he would gain from studying geometry)
>
> – *Euclid*

Trigonometry is the study of turning incomplete information about a triangle's side lengths and angles into complete information. First we develop congruence criteria, which tell us when this task of filling in the blank is feasible. Then we deal with the "ambiguous case," which showcases a particular scenario where the blanks are not always possible to fill in uniquely. We also develop criteria for proving that two triangles are similar. In the second section, we define and prove classification criteria for special classes of convex quadrilaterals, specifically trapezoids, parallelograms, rhombi, rectangles, and kites.

7.1 Triangles

Intuitively, congruent shapes are shapes that have, well, the same shape.

Definition 7.1. Following the definition of congruent polygons (Definition 5.35), two triangles are congruent, denoted by $\triangle ABC = \triangle DEF$ (the order of the vertices matters in this notation), if corresponding sides are equal and corresponding angles are equal. More precisely, $AB = DE; BC = EF; CA = FD$ and $\angle ABC = \angle DEF; \angle BCA = \angle EFD; \angle CAB = \angle FDE$. So congruence of triangles preserves lengths and interior angles.

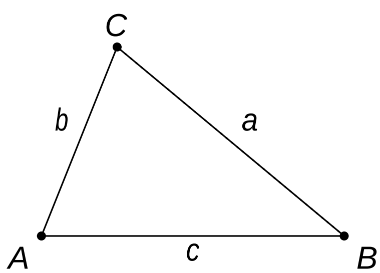


Theorem 7.2. To establish congruence, it is not necessary to show that all corresponding sides are equal and corresponding angles are equal. There are criteria which allow us to get away with less. Let S stand for "side" and A stand for "angle." Then the following sets of data each allow for one unique triangle to satisfy the data, proving four congruence theorems:

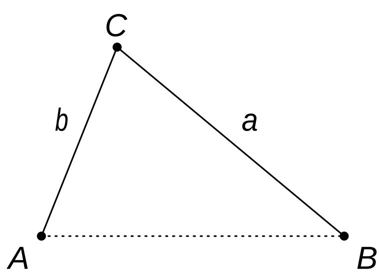
1. SSS: the three side lengths are known
2. SAS: two side lengths and the angle between those two sides are known
3. ASA: two angle measures and the length of the side between those two angles is known
4. AAS: two angle measures and the length of a side not between those two angles is known

Proof. We will use the cosine law and sine law to uniquely determine all missing angles and side lengths, as denoted for $4ABC$ in Definition 7.1.

1. The cosine law can be used to determine cos , cos , and cos . This leads to unique angles ; ; because the cosine function is injective on the domain of interior angles (0 ; 180).



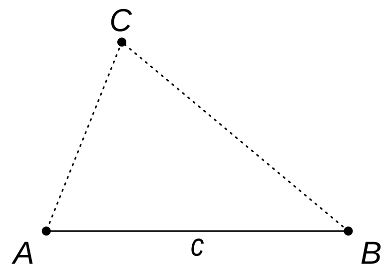
2. The cosine law can be used to obtain c, from which the cosine law can again be used as it was in SSS to obtain and .



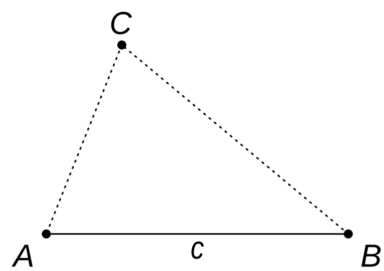
3. First we obtain = 180 . Then the sine law equations

$$\frac{a}{\sin} = \frac{c}{\sin};$$
$$\frac{b}{\sin} = \frac{c}{\sin}$$

can be used to isolate and solve for a and b.



4. This reduces to ASA because the third angle can be deduced by computing $= 180$.



We deliberately avoided using the sine law to determine any angles because the fact that

$$\sin \quad = \sin(180 \quad)$$

makes sine non-injective on $(0, 180)$. Using sine instead of cosine would thus increase our verification work. This unfortunate reality is a reflection of the "ambiguous case," which will be studied momentarily. □

Corollary 7.3. If two triangles are known to be right, then they are congruent if either of the following criteria are fulfilled:

1. LL congruence: each pair of corresponding legs are (separately) of equal length
2. HL congruence: one pair of corresponding legs are of equal length and the hypotenuses are of equal length

Proof. LL congruence follows from SAS congruence since the right angle is between the legs. For HL congruence, we may obtain the remaining leg using the Pythagorean theorem (Theorem 9.12), which reduces the problem to LL congruence. □

Corollary 7.4. The following two criteria each separately establish congruence of two triangles:

1. Two right triangles have an equal non-right angle and an equal leg.
2. Two right triangles have an equal non-right angle and equal hypotenuses.

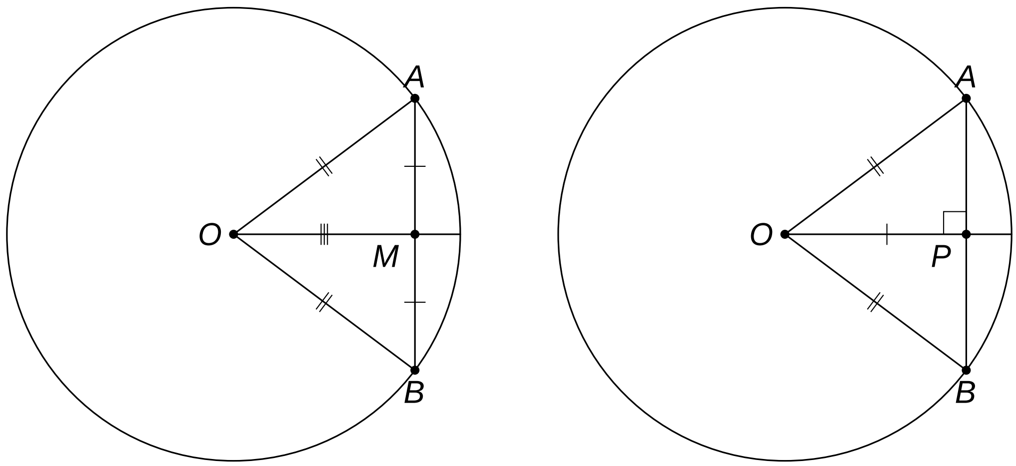
Proof. We will not need the Pythagorean theorem for these results. In both cases, we can deduce the third angle which is equal in both triangles using the sum of interior angles result for triangles, and then use ASA congruence. □

Theorem 7.5. Given a radius and a chord of a circle that is not a diameter, the radius bisects the chord if and only if the radius is perpendicular to the chord.

Proof. Let the center of the circle be O and let the chord be AB. Supposing the radius bisects the chord, let the midpoint of the chord be M. Then $AM = BM$ and $AO = BO$. By SSS congruence, $\triangle AMO \cong \triangle BMO$, and so

$$\angle AMO + \angle BMO = 180^\circ \implies \angle AMO = \angle BMO = 90^\circ.$$

In the other direction, suppose the radius is perpendicular to the chord. Let the intersection of the chord with the radius be P. By HL congruence, $\triangle APO \cong \triangle BPO$, and so $AP = BP$.



□

Problem 7.6. Show that, if two circles intersect at two distinct points, then the line through the centers of the circles is the perpendicular bisector of the common chord drawn by connecting the two points of intersection of the circles.

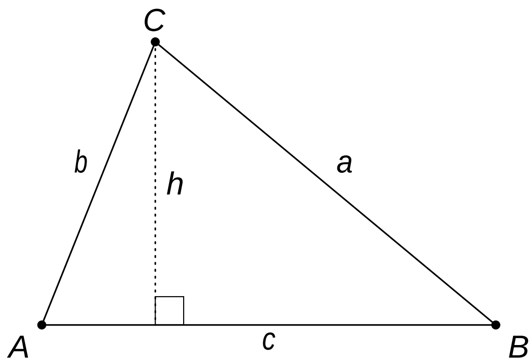
Problem 7.7. Given $\triangle ABC$ with edges a, b, c opposite to vertices A, B, C respectively, show that $\angle A > \angle B$ if and only if $a > b$. As a consequence, the longest side of a triangle is always opposite to the largest angle, and the shortest side of a triangle is always opposite to the smallest angle.

Theorem 7.8 (Ambiguous case). SSA is not a congruence theorem in its most general form. The number of solutions can be classified as follows. In $\triangle ABC$, sides a, b, c lie opposite to angles $\angle A = \alpha, \angle B = \beta, \angle C = \gamma$, respectively. Suppose we have an SSA scenario, where side lengths a and b are known and α is known. Since $\sin\alpha = \frac{h}{b}$, where h is the altitude emanating from C, let $h = b\sin\alpha$, which is always less than or equal to b. Based on considering the length a relative to h and h, the number of possible triples (c, β, γ) that produce a non-degenerate triangle is:

1. If α is acute and $a < h$, then there are exactly 0 triangles.

2. If α is acute and $a = h$, then there is exactly 1 triangle.

3. If α is acute and $h < a < b$, then there are exactly 2 triangles.
4. If α is acute and $a \geq b$, there is exactly 1 triangle.
5. If α is right or obtuse and $a \leq b$, then there are exactly 0 triangles.
6. If α is right or obtuse and $a > b$, then there is exactly 1 triangle.



Proof. We will find triples $(c; \beta; \gamma)$ that satisfy the sine law

$$\frac{a}{\sin \alpha} = \frac{b}{\sin \beta} = \frac{c}{\sin \gamma} = k;$$

where $k > 0$ just represents the common value, and $\alpha + \beta + \gamma = 180^\circ$. We claim that this is sufficient to construct a triangle by proving that the triangle inequalities (Theorem 3.10) hold under these conditions. For example, we can use reversible steps to go backwards and prove, using a compound angle identity, that

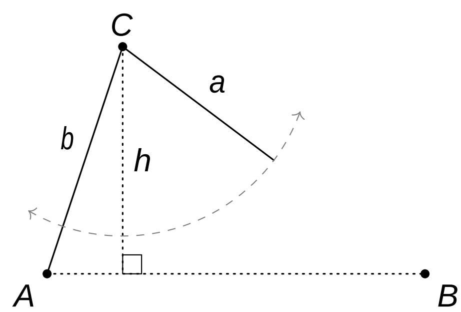
$$\begin{aligned} a + b &> c \\ k\sin\alpha + k\sin\beta &> k\sin\gamma \\ \sin\alpha + \sin\beta &> \sin\gamma \\ &= \sin(180^\circ - \alpha - \beta) \\ &= \sin(\alpha + \beta) \\ &= \sin\alpha\cos\beta + \sin\beta\cos\alpha \\ \iff \sin\alpha(1 - \cos\beta) + \sin\beta(1 - \cos\alpha) &> 0. \end{aligned}$$

The inequality in the last line is true because $\sin\alpha$ and $\sin\beta$ are positive and $\cos\alpha$ and $\cos\beta$ are strictly bounded above by 1. This proves that $a + b > c$, and the other two inequalities $b + c > a$ and $c + a > b$ are similarly proven.

1. Suppose α is acute and $a < h$. If $\triangle ABC$ exists, then the sine law gives

$$\sin\beta = \frac{b}{a}\sin\alpha = \frac{b}{a}\cdot\frac{h}{b} = \frac{h}{a} > 1.$$

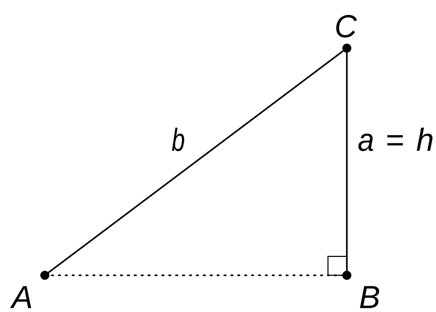
Since it is impossible for $\sin\beta > 1$ to have a solution, we get a contradiction and no triangle exists.



2. Suppose is acute and $a = h$. By the same sequence of equalities as in the first case, sin = 1, so = 90 . Then we obtain = 180 . With all the angles obtained, we can use the sine law to get

$$c = \sin \frac{a}{\sin}:$$

So, exactly, one triangle exists.



3. Suppose is acute and $h < a < b$. Then

$$\sin = \frac{h}{a} < 1$$

has two supplementary solutions for (one acute $_1$ and one obtuse $_2$), due to the reflection identity sin = sin(180) from Volume 1. The acute $_1$ leads to

$$_1 = 180 \qquad _1 > 0$$

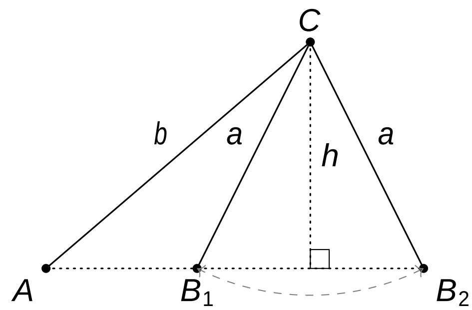
and c as in the second case above. However, with the obtuse $_2$, we need to check that

$$_2 = 180 \qquad _2$$

will be positive, which is equivalent to

$$< 180 \qquad _2 = \quad _1:$$

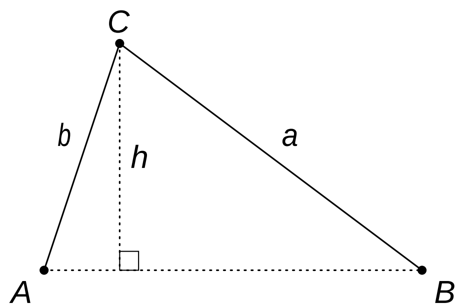
By Problem 7.7, the requirement $< \; _1$ is equivalent to $a > b$, which is inherent to this case. Thus, exactly two triangles exist.



4. Suppose is acute and $a \;\; b$. Since sin < 1, we get $b = \dfrac{h}{\sin \;}$ $> h$, so $a \;\; h$. As in the third case above, two supplementary $_1$; $_2$ are found, with $_1$ acute and $_2$ obtuse. From $_1$, we can obtain

$$= 180 \qquad {}_1 > 0.$$

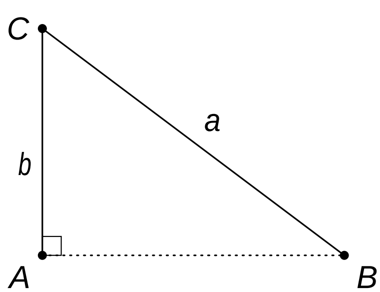
However, the obtuse $_2$ leads to $_2 >$, which is equivalent to $b > a$ by Problem 7.7, contradicting $a \;\; b$. So $_2$ is extraneous and there is exactly one triangle.

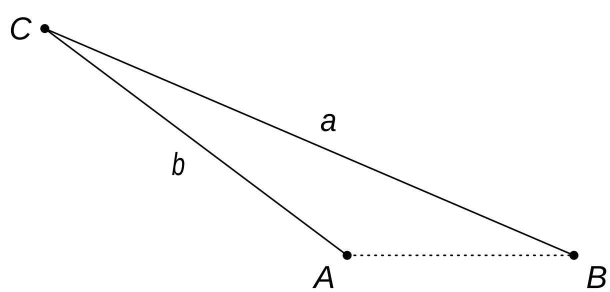


5. Suppose is right or obtuse, and $a > b$. The first condition gives $0 < \sin \qquad 1$ and the second condition gives $\dfrac{b}{a} < 1$. Combining them with the sine law yields

$$\sin \; = \frac{b}{a} \sin \qquad \frac{b}{a} < 1;$$

so a solution to that satisfies the sine law exists. In fact, two supplementary solutions exist, which are an acute $_1$ and an obtuse $_2$. However, we can only pick $_1$ because is already obtuse, and we do not want two obtuse angles since they would add up to more than 180 . Picking $_1$, we can find and c as in previous cases.





6. Suppose is right or obtuse, and a b. The first condition gives $0 <$ sin 1 and the second condition gives $\frac{b}{a}$ 1 Suppose exists. Then the sine law gives

$$\sin \quad = \frac{b}{a} \sin \quad \sin \quad = \sin(180 \quad);$$

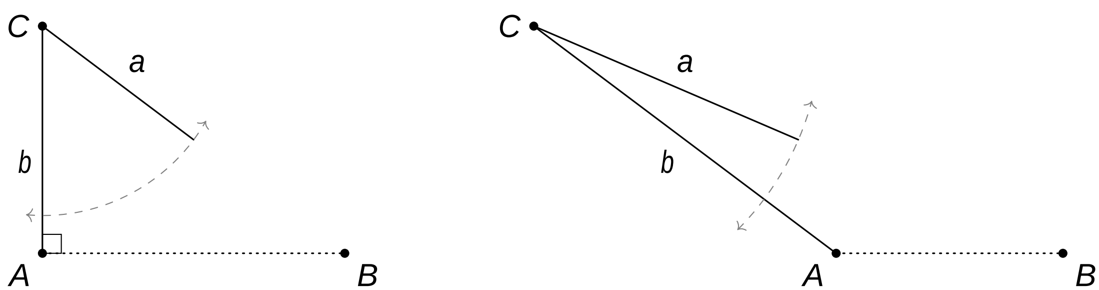
where 180 is acute, due to being obtuse. So

$$\sin \quad \sin(180 \quad) =) \quad 180 \quad ;$$

since the sine function is increasing in the first quadrant that consists of all acute angles. Then we get the contradiction

$$= 180 \quad 0:$$

Therefore, there is no triangle in this case.



This covers all cases of the SSA scenario. □

Definition 7.9. Suppose $4ABC$ has edges $a; b$ opposite to vertices $A; B$ respectively, such that $\backslash A$ is known and acute, and $a; b$ are known and they satisfy $b \sin A < a < b:$ Note that $b \sin A$ is the length of the height emanating from $C:$ By Theorem 7.8, there are two configurations for $4ABC$ that satisfy these conditions. This is called the ambiguous case.

Theorem 7.10. Two sides of a triangle are equal if and only if the angles opposite them are equal.

Proof. Let the triangle be $4ABC:$ For one direction, suppose $AB = AC:$ The amazing trick here is that $4ABC = 4ACB$ because $\backslash BAC = \backslash CAB$ so we can use SAS congruence. Then $\backslash B = \backslash C$ by the definition of congruence.
In the other direction, suppose $\backslash B = \backslash C:$ A similar trick that we can pull out is that $BC = CB;$ so $4ABC = 4ACB$ by ASA congruence. Then $AB = AC$ by the definition of congruence. □

Definition 7.11. Using Theorem 7.10, we can classify triangles into three types, each of which can be given two equivalent formulations, one using lengths of sides and one using measures of interior angles:

- Equilateral triangle: three equal side lengths or three equal interior angles

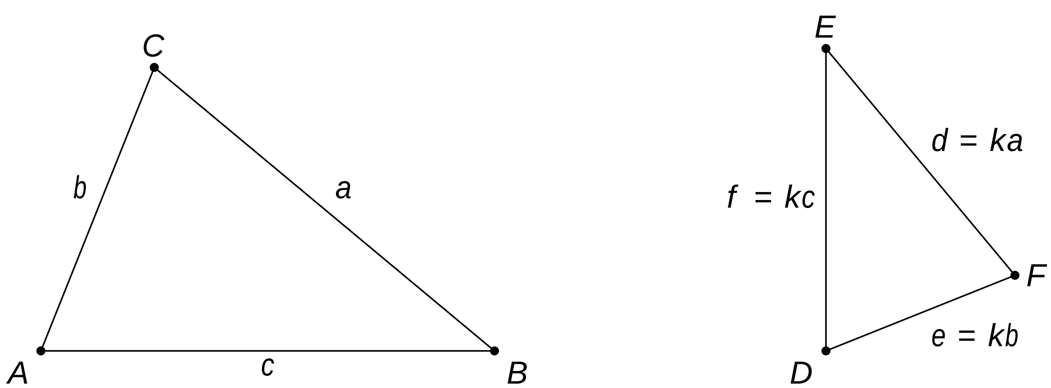
- Isosceles triangle: at least two equal side lengths or at least two equal interior angles. In an isosceles triangle, the legs are the equal sides and the base is the remaining side. The equal angles are called the base angles and the vertex angle is the remaining angle. The vertex from which the vertex angle emanates is called the apex. Note that equilateral triangles qualify as isosceles triangles.
- Scalene triangle: three different side lengths or three different interior angles

Intuitively, two shapes are similar if they are scaled versions of one another.

Definition 7.12. Following the definition of similar polygons (Definition 5.37), two triangles are similar, denoted by $\triangle ABC \sim \triangle DEF$ (as with congruence, correctly ordering the vertices in this notation is important), if corresponding angles are equal and the ratios between corresponding sides are equal. More precisely, $\angle ABC = \angle DEF; \angle BCA = \angle EFD; \angle CAB = \angle FDE$ and

$$\frac{DE}{AB} = \frac{EF}{BC} = \frac{FD}{CA}$$

where the common value k of the proportions is called the similarity ratio. So, similarity preserves interior angles, but not lengths. Similar triangles are said to have the same orientation if starting with matching angles and going through the other two angles counterclockwise produces the same angles in the same order; otherwise, the similar triangles are called oppositely oriented.



Theorem 7.13. As with congruence, there are similarity criteria which allow us to establish similarity without showing that all of the conditions in the definition of similarity hold:

1. AA: two pairs of corresponding angles are separately equal (and equality of the third angles can be deduced)
2. SSS: the three ratios of corresponding sides are equal
3. SAS: the ratios of two corresponding sides are equal, and the angles between them are equal

Proof. As with the four congruence criteria, we will use the sine law and cosine law to prove these three similarity criteria. Let the triangles be $\triangle ABC$ with sides $a; b; c$ opposite angles $A; B; C$, respectively, and $\triangle DEF$ with sides $d; e; f$ opposite angles $D; E; F$, respectively.

1. Suppose $\angle A$ and $\angle D$ have a common value of , and that $\angle B$ and $\angle E$ have a common value of . Then $\angle C$ and $\angle F$ have a common value of

$$= 180 \quad :$$

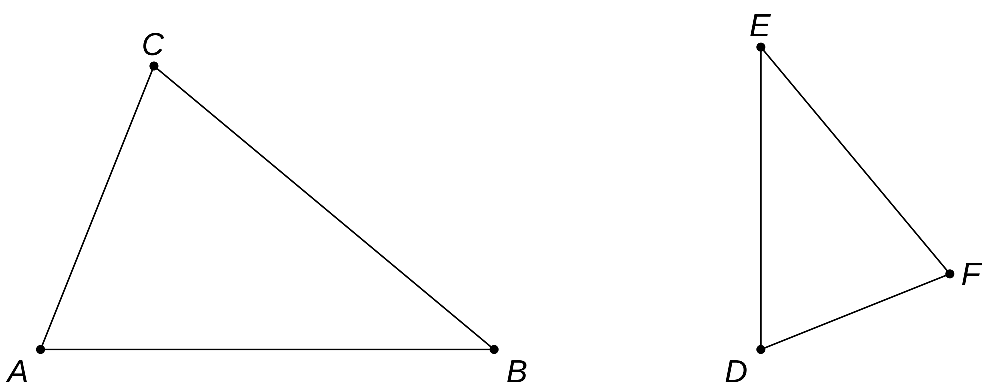
By the sine law,

$$\frac{a}{\sin} = \frac{b}{\sin} = \frac{c}{\sin};$$
$$\frac{d}{\sin} = \frac{e}{\sin} = \frac{f}{\sin}:$$

Dividing the top equation by the bottom one causes cancellations, yielding

$$\frac{a}{d} = \frac{b}{e} = \frac{c}{f};$$

which are the similarity proportions that we seek.



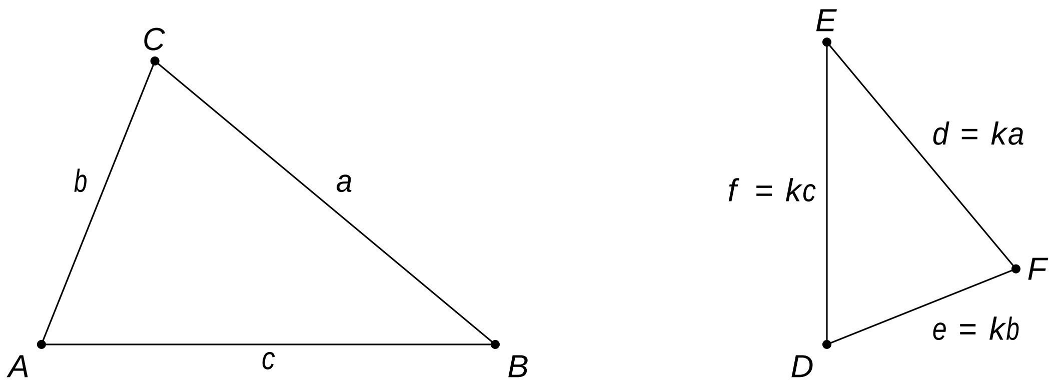
2. Suppose

$$\frac{d}{a} = \frac{e}{b} = \frac{f}{c} = k$$

for some common value k. Then $d = ka; e = kb;$ and $f = kc$. Then

$$\begin{aligned}\cos B &= \frac{c^2 + a^2 \quad b^2}{2ca} \\ &= \frac{(kc)^2 + (ka)^2 \quad (kb)^2}{2(kc)(ka)} \\ &= \frac{f^2 + d^2 \quad e^2}{2fd} \\ &= \cos E;\end{aligned}$$

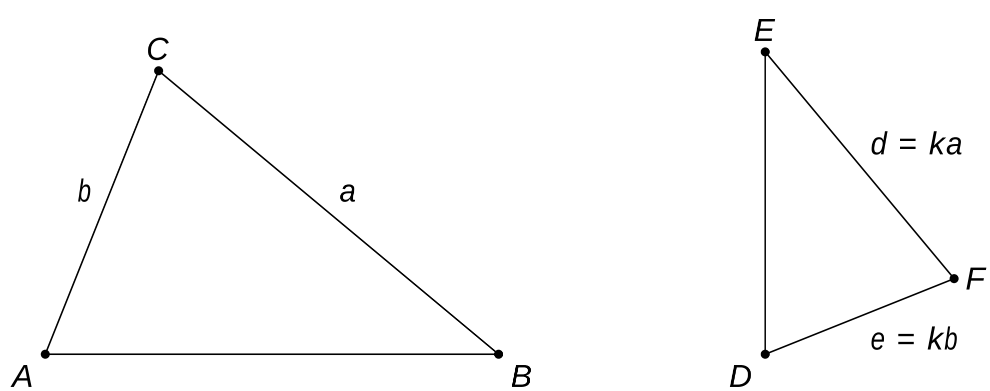
which gives $\angle B = \angle E$, due to the injectivity of the cosine function on $(0 ;180)$. Similarly, $\angle A = \angle D$ and $\angle C = \angle F$.



3. Suppose $\frac{d}{a} = \frac{e}{b} = k$ for some common value k and that $\angle C = \angle F$ have a common value of θ. Then $d = ak$ and $e = bk$. So,

$$\begin{aligned}\frac{f}{c} &= \frac{\sqrt{d^2 + e^2 - 2de\cos\theta}}{\sqrt{a^2 + b^2 - 2ab\cos\theta}} \\ &= \frac{\sqrt{(ka)^2 + (kb)^2 - 2(ka)(kb)\cos\theta}}{\sqrt{a^2 + b^2 - 2ab\cos\theta}} \\ &= k = \frac{d}{a} = \frac{e}{b},\end{aligned}$$

which proves the similarity proportions. This reduces the problem to SSS similarity, which was addressed above.



□

Problem 7.14. Prove that $\triangle ABC$ is equilateral if and only if $\triangle ABC \sim \triangle BCA$.

Theorem 7.15. A common setup is when there are two triangles with an equal angle and one triangle is wedged inside the other, which we will call **nested triangles**. The triangles are similar and have the same orientation if and only if the sides opposite the nested angles are parallel. (Note that, in a different scenario, the nested triangles could be similar yet oppositely oriented.)

Proof. Let the triangles be $\triangle ABC$ and $\triangle A'B'C'$ with $\angle A = \angle A'$, and A, A' coinciding. By corresponding angles, BC and $B'C'$ are parallel if and only if $\angle ABC = \angle A'B'C'$ and $\angle ACB = \angle A'C'B'$. By AA similarity, this is true if and only if $\triangle ABC \sim \triangle A'B'C'$ and the two triangles have the same orientation. □

Problem 7.16. Show that connecting the midpoints of two sides of a triangle creates a line segment that has half the length of the third side and that is parallel to the third side. Use this to prove that connecting the midpoints of all three sides to each other creates four congruent triangles. The central one is called the **medial triangle**.

7.2 Quadrilaterals

Lemma 7.17. "The" **shortest distance** between two parallel lines in the plane is a well-defined concept because:

1. Suppose we are given a line and a point in the plane that does not lie on the line. Then there exists a unique line segment of shortest length with one endpoint on the point and the other endpoint on the line. The segment is perpendicular to any direction vectors on the line.

2. The distance from a point on one line to the closest point on a parallel plane is a constant, regardless of the first point.

3. The shortest distance from any point on the first line to the second line is the same as the shortest distance from any point on the second line to the first line.

Proof. In essence, the concepts of perpendicular distance and shortest distance converge.

1. Let the point be P and the line be ℓ. Following Theorem 4.17, let v be a vector with its tail on the line and its head on P, and let w be a direction vector of the line. Let $u = \text{proj}_w v$ and $z = \text{oproj}_w v$. Then $\|z\|$ is the perpendicular distance from P to ℓ. We claim that this is also the shortest possible distance from P to anywhere on ℓ. Indeed, by the vector Pythagorean theorem (Theorem 4.14),

$$\|v\|^2 = \|u\|^2 + \|z\|^2 \geq \|z\|^2 \implies \|v\| \geq \|z\|,$$

with equality holding if and only if $u = 0$. So, the point on ℓ at which the minimum distance $\|z\|$ from P occurs is unique, and this is also where the perpendicular distance occurs.

2. Suppose parallel lines are given by

$$ax + by + c = 0,$$
$$Ax + By + C = 0.$$

Let (x_0, y_0) and (x_1, y_1) be distinct points on the first line. Using the formula for the perpendicular distance between a point and a line in the plane (Theorem 4.17), we work backwards:

$$\begin{aligned} \frac{|Ax_0 + By_0 + C|}{\sqrt{A^2 + B^2}} &= \frac{|Ax_1 + By_1 + C|}{\sqrt{A^2 + B^2}} \\ (Ax_0 + By_0 + C)^2 &= (Ax_1 + By_1 + C)^2 \\ [A(x_0 - x_1) + B(y_0 - y_1)] \cdot [A(x_0 + x_1) + B(y_0 + y_1) + 2C] &= 0, \end{aligned}$$

where we used the difference of squares factorization in the last step. This is true because the first factor in the last line is equal to

$$(A, B) \cdot (x_0 - x_1, y_0 - y_1),$$

which is zero because $(x_0 - x_1, y_0 - y_1)$ is a direction vector of the line and (A, B) is a normal vector to the line.

3. Since we have shown that the concepts of perpendicular distance and shortest distance are the same, it suffices to prove that if P is a point on the first line and Q is a point on the second line, and if w is a direction vector common to both lines (which exists because they are parallel), then, for $v = \overrightarrow{PQ}$,

$$\|\mathrm{oproj}_w v\| = \|\mathrm{oproj}_w(-v)\|.$$

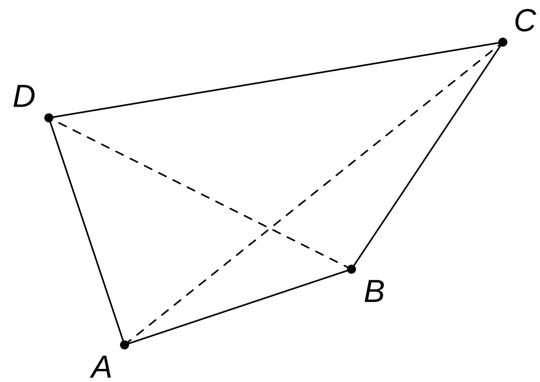
Indeed, by the formula in Theorem 4.13, we get

$$\begin{aligned} \|\mathrm{oproj}_w(-v)\| &= \|-v - \mathrm{proj}_w(-v)\| \\ &= \left\|-v - \frac{\langle -v, w\rangle}{\langle w, w\rangle} w\right\| \\ &= \left\|-v + \frac{\langle v, w\rangle}{\langle w, w\rangle} w\right\| \\ &= \left\|v - \frac{\langle v, w\rangle}{\langle w, w\rangle} w\right\| \\ &= \|v - \mathrm{proj}_w v\| \\ &= \|\mathrm{oproj}_w v\|, \end{aligned}$$

as desired.

This will allow us to refer to a well-defined concept of the shortest and perpendicular "distance" between parallel lines, particularly when working with special classes of quadrilaterals. □

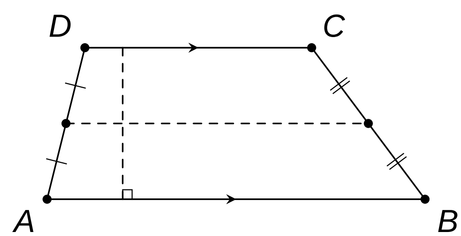
Definition 7.18. A **convex quadrilateral** is a convex polygon with four vertices and four edges. The **diagonals** of a convex quadrilateral $ABCD$ are the line segments AC and BD that connect non-adjacent vertices.



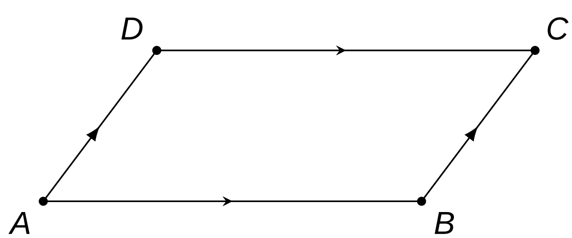
There also exist non-convex (i.e. concave) quadrilaterals, such as darts, but we will not study them.

Definition 7.19. There are several classes of convex quadrilaterals, defined by their important regularities:

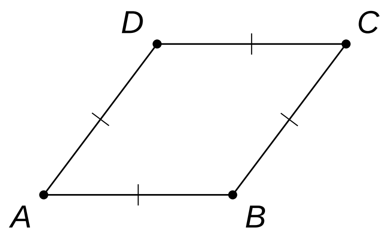
1. A trapezoid is a convex quadrilateral with at least one pair of parallel sides. If one pair of parallel sides is distinguished then they are called bases and the other two sides are called legs. The height of a trapezoid is the distance between the lines through the bases, which is well-defined according to Lemma 7.17. The median of a trapezoid is the line segment connecting the midpoints of the two legs.



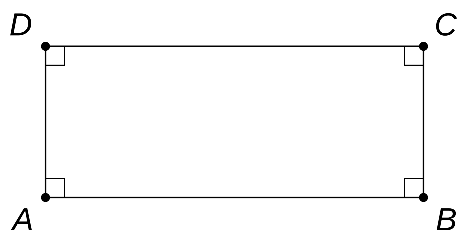
2. A parallelogram is a convex quadrilateral such that each pair of opposite sides is parallel. Note that every parallelogram is a trapezoid.



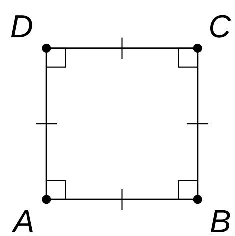
3. A rhombus is a convex quadrilateral such that all four edges have equal length. By drawing a diagonal and using SSS congruence, we find that the two triangles are congruent isosceles triangles; this causes the existence of alternate interior angles angles twice, which shows that opposite sides of a rhombus are parallel. So, every rhombus a parallelogram.



4. A **rectangle** is a convex quadrilateral with all four interior angles having equal measure. By drawing a diagonal, we can split any convex quadrilateral into two triangles and so the sum of the interior angles of a convex quadrilateral is $2 \cdot 180^\circ = 360^\circ.$ Then, the interior angles of a rectangle all measure 90°; this causes the existence of alternate interior angles, which shows that each pair of opposite sides is parallel. So, every rectangle a parallelogram.



5. A **square** is a convex quadrilateral with all four edges having equal length and all four interior angles having equal measure, making it both a rhombus and a rectangle.



Therefore, we have the following inclusion hierarchy of convex quadrilaterals:

$$\begin{aligned} \text{Squares} &\subset \text{Rectangles and Rhombi} \\ &\subset \text{Parallelograms} \\ &\subset \text{Trapezoids} \\ &\subset \text{Convex Quadrilaterals}. \end{aligned}$$

All of the inclusions are actually proper but we will not belabour this obvious point.

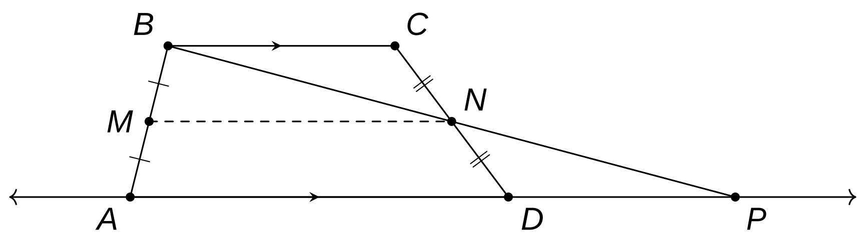
Problem 7.20. Prove that, in any convex quadrilateral, a parallelogram is created by connecting the midpoints of consecutive sides.

Theorem 7.21. Given any trapezoid, the median is parallel to the bases and the length of the median is the average of the lengths of the bases.

Proof. Let trapezoid $ABCD$ have bases AD and BC. Let M be the midpoint of AB and N be the midpoint of CD. First we extend BN until it meets the line through AD at P. Since $\angle CBP = \angle DPB$ and $\angle BCD = \angle PDC$ and $DN = CN$, we get that $\triangle BCN \cong \triangle PDN$ by AAS congruence. Looking at MN as connecting the midpoints of two sides of $\triangle ABP$, Problem 7.16 implies that MN is parallel to AD and

$$MN = \frac{AP}{2} = \frac{AD + DP}{2} = \frac{AD + BC}{2},$$

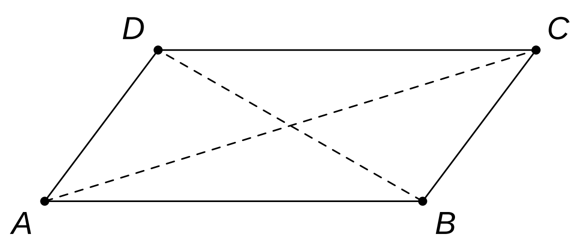
since $DP = BC$. It is also true that MN is parallel to BC because MN is parallel to AP, and AD is parallel to BC.



□

Theorem 7.22. A convex quadrilateral is a parallelogram if and only if any one of the following conditions hold for it:

1. Each pair of opposite sides consists of two sides of equal length.
2. Each pair of opposite interior angles consists of two equal angles.
3. The diagonals bisect each other.
4. There are two opposite sides that are parallel and equal in length.



Proof. Letting the original definition of a parallelogram be (0), we will prove that

$$(0) \implies (1) \implies (2) \implies (3) \implies (4) \implies (0).$$

1. Suppose $ABCD$ is a parallelogram, meaning each pair of opposite sides is parallel. By drawing the diagonal BD and using alternate interior angles to invoke ASA congruence, we find that $\triangle ABD \cong \triangle CDB$. Thus, $AB = CD$ and $AD = CB$.

2. Suppose $AB = CD$ and $AD = CB$. Again, we draw the diagonal BD. This time, we use SSS congruence to get that $\triangle ABD = \triangle CDB$. Then $\angle ABD = \angle CDB$ and $\angle ADB = \angle CBD$, so

$$\angle ABC = \angle ABD + \angle DBC = \angle CDB + \angle BDA = \angle CDA.$$

By the same SSS congruence, we also get $\angle BAD = \angle DCB$.

3. Suppose $\angle ABC = \angle CDA$ and $\angle BAD = \angle DCB$. Since the sum of the interior angles of a quadrilateral is 360, we find that $\angle DCB$ and $\angle CDA$ and supplementary. Thus, they are same-side interior angles in a setup where AD and BC have parallel lines running through them and CD has a transversal running through it. Similarly, AB and CD are parallel too. As in the first implication, $AB = CD$ and $AD = CB$. Now we draw the diagonals. By using the equality of alternate interior angles and ASA congruence, we find that the diagonals bisect each other. The detail are left to the reader.

4. Suppose the diagonals of $ABCD$ bisect each other with the point of intersection being E. By the equality of vertical angles and by SAS congruence, $AD = BC$. Then alternate interior angles imply that AD and BC are parallel.

5. Suppose $AD = BC$ and that AD and BC are parallel. First we draw the diagonal BD and use alternate interior angles to get $\angle ADB = \angle CBD$. By SAS congruence, $\triangle ADB = \triangle CBD$. Then $\angle ABD = \angle CDB$. By alternate interior angles, AB and CD are parallel too. Thus, the two pairs of opposite sides of $ABCD$ each consists of parallel sides, making it a parallelogram.

Therefore, we have five equivalent criteria for identifying a parallelogram. □

Note that we drew auxiliary lines like diagonals to facilitate the proof of Theorem 7.22. One way of thinking of drawing auxiliary lines is that we are not drawing a line, but rather the line is already there as a subset of the plane. We have only noticed that it is there. The geometer wishes to become competent at observing what is already there.

Problem 7.23. Prove that a convex quadrilateral is a rhombus if and only if diagonals are perpendicular and bisect each other.

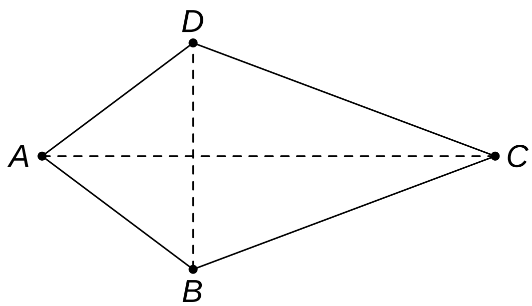
Problem 7.24. Prove that a convex quadrilateral is a rectangle if and only if the diagonals are equal in length and bisect each other.

Since a convex quadrilateral is a square if and only if it is both a rectangle and a rhombus, combining Problem 7.23 and Problem 7.24 yields that a convex quadrilateral is a square if and only if the diagonals are perpendicular, equal, and bisect each other.

Definition 7.25. A **kite** is a convex quadrilateral with two disjoint pairs of sides such that each pair consists of adjacent sides that are equal. For example, rhombuses are kites, as are squares.

Theorem 7.26. A convex quadrilateral is a kite if and only if any one of the following conditions hold.

1. A diagonal cuts the quadrilateral into two congruent triangles (so, there is a line of symmetry).
2. A diagonal bisects a pair of opposite interior angles.
3. A diagonal bisects the other diagonal, and the two diagonals are perpendicular to each other.



Proof. Letting the original definition of a kite be (0), we will prove that

$$(0) \;=)\; (1) \;=)\; (2) \;=)\; (3) \;=)\; (0):$$

1. Suppose we have a kite $ABCD$ such that $AB = AD$ and $CB = CD:$ By SSS congruence $4ABC = 4ADC:$ So the diagonal AC is the one we seek.
2. Suppose $ABCD$ is a convex quadrilateral such that $4ABC = 4ADC:$ Then $\backslash BAC = \backslash DAC$ and $\backslash BCA = \backslash DCA;$ so the diagonal AC bisects each of the interior angles $\backslash BAD$ and $\backslash BCD:$
3. Suppose $ABCD$ is a convex quadrilateral such that the diagonal AC bisects each of the interior angles $\backslash BAD$ and $\backslash BCD:$ Let E be the foot of the height of $4ABC$ from B to $AC;$ and F be the foot of the height of $4ADC$ from D to $AC:$ By AAS congruence, $4ABE = 4ADF;$ so $AE = AF:$ So the diagonal BD intersects the diagonal AC at $E = F:$ Since $E = F$ is the feet of two altitudes, the intersection forms right angles. Moreover, the aforementioned congruence yields $BE = DF;$ so BD is bisected by $AC:$
4. Suppose $ABCD$ is a convex quadrilateral such that AC perpendicularly bisects BD at $E:$ By SAS congruence $4AEB = 4AED$ and $4CEB = 4CED:$ Thus, $AB = AD$ and $CB = DC;$ making $ABCD$ a kite.

Therefore, we have four equivalent criteria for identifying a kite. □

Chapter 8

Circles II

"Let no one untrained in geometry enter."

– *written over the entrance to Plato's academy*

We will start by studying the power of a point theorem and its converse. Then we will look at several useful and interesting criteria for knowing when four points all lie on a circle, since cyclic quadrilaterals are rich in structure. Afterwards, we will look into cyclic and tangential polygons.

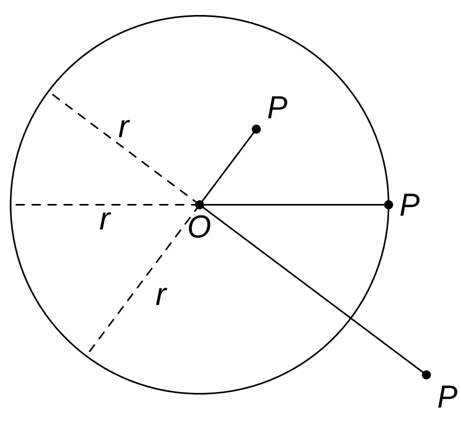
8.1 Cyclic Quadrilaterals

Definition 8.1. The power of a point P with respect to a circle with center O and radius r is

$$\text{Pow}\ (P) = OP^2 \quad r^2:$$

This is the same as $(OP + r)(OP \quad r);$ so

$$\text{sgn}\,[\text{Pow}\ (P)] = \begin{cases} 1 & \text{if } P \text{ is in the exterior of} \\ 0 & \text{if } P \text{ is on} \\ 1 & \text{if } P \text{ is in the interior of} \end{cases} \quad :$$



Theorem 8.2 (Power of a point theorem). Let P be a point, let be a circle, and let \` be a line through $P:$ Then:

1. If P is in the exterior of and \` intersects at two distinct points X and $Y;$ then

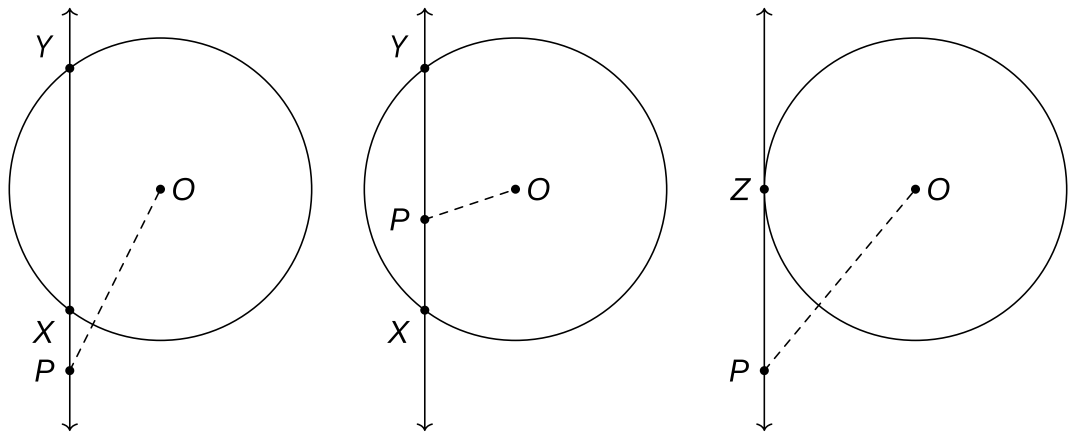
$$PX \quad PY = \text{Pow}\ (P):$$

2. If P is in the interior of ω and ℓ intersects ω at two distinct points X and Y, then

$$PX \cdot PY = -\operatorname{Pow}_\omega(P).$$

3. If P is in the exterior of ω and ℓ is tangent to ω at Z then

$$PZ^2 = \operatorname{Pow}_\omega(P).$$



Proof. Authors often prove the power of a point theorem for each configuration separately using triangle geometry. Some go straight to proving the consequential configurations in Corollary 8.3. We will tackle all three cases at once using complex numbers.
Let the center of ω be the origin, without loss of generality, and let its radius be r, Then a general point on ω is $z = re^{i\theta}$. Let the point P be located at the complex number p, and let the counterclockwise angle from the x-axis to ℓ be φ. Then a general point on ℓ is

$$z = p + se^{i\varphi},$$

where s is a directed length from p to z, where the sign of s depends on the of side p on ℓ on which z lies. Letting z represent the points where ℓ intersects ω, we can equate

$$p + se^{i\varphi} = z = re^{i\theta}.$$

The conjugate of this equation is

$$\overline{p} + se^{i(-\varphi)} = re^{i(-\theta)}.$$

Multiplying the original equation by the conjugate equation yields

$$\begin{aligned} r^2 &= (re^{i\theta}) \cdot (re^{i(-\theta)}) \\ &= (p + se^{i\varphi})(\overline{p} + se^{i(-\varphi)}) \\ &= |p|^2 + 2\left(\frac{pe^{i(-\varphi)} + \overline{p}e^{i\varphi}}{2}\right)s + s^2 \\ &= |p|^2 + \left[2\operatorname{Re}(pe^{i(-\varphi)})\right]s + s^2. \end{aligned}$$

This provides a quadratic equation in the variable s with real coefficients, namely

$$s^2 + \left[2\,\mathrm{Re}(pe^{i(\ \ )})\right] s + (|p|^2 - r^2) = 0.$$

If the two solutions to this quadratic are s_1 and s_2 then we are seeking $|s_1| \cdot |s_2|$. By Vieta's formulas,

$$|s_1| \cdot |s_2| = |s_1 \cdot s_2| = \left||p|^2 - r^2\right| = |\operatorname{Pow}_{\omega}(P)|.$$

Now we can specialize the formula to each case:

1. If P is in the exterior of ω and ℓ intersects ω at two distinct points X and Y, then

$$PX \cdot PY = |s_1| \cdot |s_2| = |\operatorname{Pow}_{\omega}(P)| = \operatorname{Pow}_{\omega}(P).$$

2. If P is in the interior of ω and ℓ intersects ω at two distinct points X and Y, then

$$PX \cdot PY = |s_1| \cdot |s_2| = |\operatorname{Pow}_{\omega}(P)| = -\operatorname{Pow}_{\omega}(P).$$

3. If P is in the exterior of ω and ℓ is tangent to ω at Z, then there is one distinct solution to the quadratic and $s_1 = s_2$, which leads to

$$PZ^2 = |s_1|^2 = |s_1| \cdot |s_2| = |\operatorname{Pow}_{\omega}(P)| = \operatorname{Pow}_{\omega}(P).$$

□

Corollary 8.3. In practice, the power of a point theorem comes in three configurations that follow directly from Theorem 8.2. Let P be a point and let ω be a circle.

1. Chord-chord: If P is in the interior of ω, and AB and CD are two chords of the circle that intersect at P, then
$$PA \cdot PB = PC \cdot PD.$$

2. Secant-secant: If P is in the exterior of ω, and one secant through P intersects the circle at two distinct points A and B, and another secant through P intersects the circle at two distinct points C and D, then
$$PA \cdot PB = PC \cdot PD.$$

3. Secant-tangent: If P is in the exterior of ω, and a secant through P intersects the circle at two distinct points A and B, and a tangent through P intersects the circle at C, then
$$PA \cdot PB = PC^2.$$

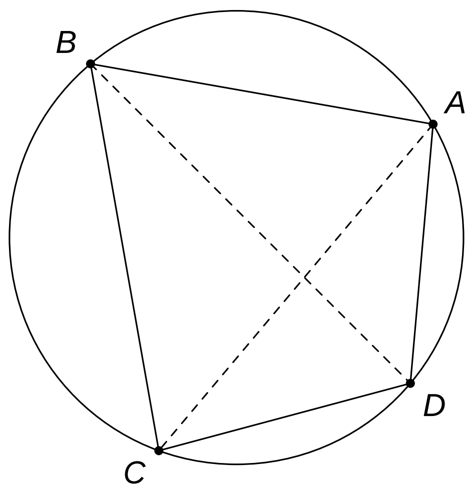
We could also develop a tangent-tangent case, but this would be equivalent to saying that the two tangent segments from an exterior point have the same length, which was proven in Theorem 6.12.

Definition 8.4. If four points lie on a circle then they form a **cyclic quadrilateral**. To be clear, to name it we would still write the vertices in clockwise or counterclockwise fashion, starting with any vertex, as usual.

Theorem 8.5 (Cyclic quadrilateral criteria)**.** Let $ABCD$ be a convex quadrilateral. Then the following conditions are equivalent:

1. $ABCD$ is a cyclic quadrilateral
2. $\angle ADB = \angle ACB$
3. $\angle ABC + \angle ADC = 180$

We have assumed convexity in order for the angles in question to be well-defined angles whose interiors lie in the interior of the quadrilateral. Otherwise, it would not be clear to which of the two explementary angles we are referring in some cases where we refer to an angle.



Proof. We will show that

$$(2) \iff (1) \iff (3)$$

follows from the inscribed angle theorem and its converse (Theorem 6.6). By naming conventions, since $ABCD$ is a convex quadrilateral, C and D lie on the same side of the line through AB; and B and D lie on opposite sides of the line through AC. Thus:

- (1) $\implies$ (2): We know that C and D lie on the same arc between A and B. By the inscribed angle theorem, $\angle ADB$ and $\angle ACB$ are both equal to half the measure of their intercepted arc $\overset{\frown}{AB}$.

- (2) $\implies$ (1): First we draw the circumcircle of $\triangle ADB$. Letting $\overset{\frown}{AB}$ be the arc not containing D, $\angle ADB$ is equal to half the measure of $\overset{\frown}{AB}$. Since $\angle ACB = \angle ADB$ and C lies on the side of the line through AB not containing $\overset{\frown}{AB}$, this forces C to lie on arc $\overset{\frown}{ADB}$ as well by the converse of the inscribed angle theorem. Thus, $ABCD$ is cyclic.

- (1) $\Rightarrow$ (3): We know that B and D lie on opposite sides of the line through AC, so the inscribed angle theorem tells us that $\angle ABC$ and $\angle ADC$ are equal to half the measures of $\overset{\frown}{ADC}$ and $\overset{\frown}{ABC}$, respectively. Since the two arcs sum to a full rotation, the sum of the inscribed angles is

$$\angle ABC + \angle ADC = \frac{\overset{\frown}{ADC}}{2} + \frac{\overset{\frown}{ABC}}{2} = \frac{360^\circ}{2} = 180^\circ.$$

- (3) $\Rightarrow$ (1): First we draw the circumcircle of $\triangle ABC$. Letting $\overset{\frown}{AC}$ be the arc not containing B, $\angle ABC$ is equal to half the measure of $\overset{\frown}{AC}$. By assumption and the inscribed angle theorem,

$$\angle ADC = 180^\circ - \angle ABC = 180^\circ - \frac{\overset{\frown}{AC}}{2} = \frac{360^\circ - \overset{\frown}{ABC}}{2} = \frac{\overset{\frown}{ABC}}{2}.$$

  Since D and B lie on the opposite sides of the line through AC, D lies on the side of the line through AC not including $\overset{\frown}{ABC}$. By the converse of the inscribed angle theorem, this forces D to lie on $\overset{\frown}{AC}$. Thus, $ABCD$ is cyclic.

□

Problem 8.6. Find a formula for each diagonal of a cyclic quadrilateral purely in terms of the side lengths of the quadrilateral.

Theorem 8.7. A trapezoid $ABCD$ with parallel sides AB and CD is said to be isosceles if it satisfies any one of the following conditions. Then these conditions are equivalent:

1. The diagonals AC and BD are equal in length.

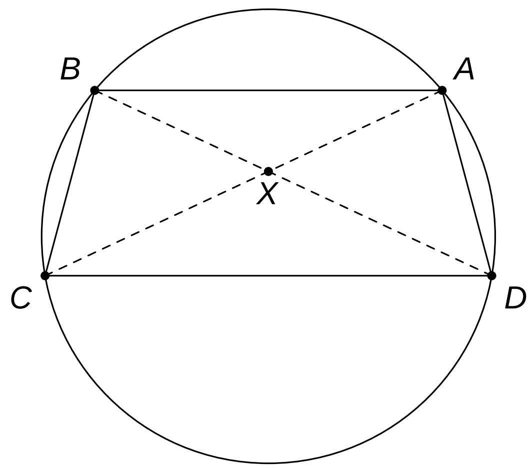
2. If the point of intersection of the diagonals is X, then $AX = BX$ and $CX = DX$.

3. A pair of the base angles are equal, where $\angle ADC$ and $\angle BCD$ is one pair of base angles, and $\angle DAB$ and $\angle CBA$ is another pair of base angles. Note that if one pair of base angles are equal then the other pair is also equal since same-side interior angles of a transversal are supplementary, which implies that

$$\angle ADC + \angle DAB = 180^\circ = \angle BCD + \angle CBA.$$

4. A pair of opposite interior angles are supplementary. Note that one pair of opposite interior angles being supplementary implies cyclicity, which implies that the other pair of opposite interior angles is also supplementary.

Contrary to intuition, the legs (i.e. the two opposite sides that are not necessarily parallel) being equal in length is not sufficient for the trapezoid to be isosceles. Otherwise, a non-square rhombus would qualify,even though its diagonals are not equal, for example.



Proof. We will cycle through the conditions:

- (1) $\implies$ (2): If $AC = BD$, then alternate interior angles of a transversal tells us that $\triangle ABX \sim \triangle CDX$. By similarity ratios, $\frac{AX}{CX} = \frac{BX}{DX}$, and the hypothesis is that $AX + CX = BX + DX$. By an algebraic ratio trick,
$$\frac{AX}{BX} = \frac{CX}{DX} = \frac{AX + CX}{BX + DX} = 1.$$

- (2) $\implies$ (3): If $AX = BX$ and $CX = DX$, then isosceles triangles and alternate interior angles of a transversal tell us that $\angle ABD = \angle BDC = \angle ACD$. Since trapezoids are always convex, this means $ABCD$ is cyclic. Then $\angle ADB = \angle ACB$ as well, which means
$$\angle ADC = \angle ADB + \angle BDC = \angle ACB + \angle ACD = \angle BCD,$$
so two base angles are equal.

- (3) $\implies$ (4): If a pair of base angles are equal then the other pair of base angles are also equal, as argued in the statement of the theorem. Since same-side interior angles of a transversal are supplementary,
$$\angle ABC + \angle ADC = \angle ABC + \angle BCD = 180^\circ.$$

- (4) $\implies$ (1): If opposite interior angles are supplementary, then using the fact that same-side interior angles of a transversal are supplementary,
$$\angle ADC = 180^\circ - \angle ABC = 180^\circ - (180^\circ - \angle BCD) = \angle BCD.$$
Moreover, opposite interior angles being supplementary in a convex quadrilateral implies cyclicity, which implies $\angle DAC = \angle CBD$. Since $\triangle ADC$ and $\triangle BCD$ share DC, it follows from AAS congruence that they are congruent, which allows us to conclude that $AC = BD$.

□

Problem 8.8. Show that a trapezoid is isosceles if and only if connecting the midpoints of the bases creates a segment perpendicular to both bases.

Now we will develop several concyclicity criteria for four points in the plane. The first is based on the complex number "cross-ratio," the second is a converse to the power of a point theorem, and the third involves all $\binom{4}{2} = 6$ pairwise lengths between the four points.

Theorem 8.9 (Complex concyclicity). Let $A; B; C; D$ be distinct points in the plane such that they are not all collinear. Let the corresponding complex numbers in the complex plane be $a; b; c; d$, respectively. Then $A; B; C; D$ are concyclic if and only if

$$(a; b; c; d) = \frac{b - c}{a - c} \cdot \frac{a - d}{b - d}$$

is a non-zero real number. We will prove the following more refined version:

1. $A; B; C; D$ are concyclic such that $C; D$ lie on the same side of the line through AB if and only if $(a; b; c; d)$ is a positive real number.

2. $A; B; C; D$ are concyclic such that $C; D$ lie on opposite sides of the line through AB if and only if $(a; b; c; d)$ is a negative real number.

Proof. If $A; B; C; D$ were concyclic in some order then the resulting quadrilateral would be cyclic. So the quadrilateral would be convex, which would mean that no three of the vertices are collinear and none of the vertices lie in the interior of the triangle formed by the other three vertices. Before we delve into the heart of the proof, we will briefly argue in the other direction that, if $(a; b; c; d)$ is a non-zero real number, then no one of $A; B; C; D$ lies in the interior or boundary of the triangle formed by the other three vertices. Suppose otherwise for contradiction. Then the fourth vertex lies in the interior of the triangle or the interior of one of the three edges. In essence, $(a; b; c; d)$ is a non-zero real number if and only if

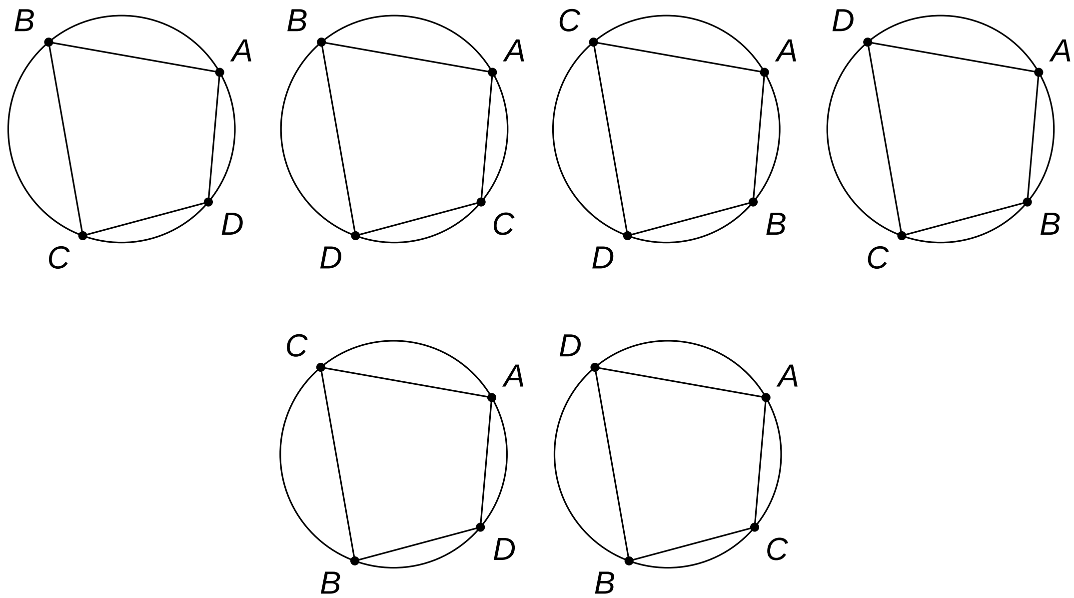
$$\arg\left(\frac{b - c}{a - c}\right) + \arg\left(\frac{a - d}{b - d}\right) \equiv 0; \pi.$$

Geometrically, this is equivalent to meaning the counterclockwise angle around C that makes AC coincide with BC plus the counterclockwise angle around D that makes BD coincide with AD is congruent to 0 or π modulo 2π. Due to size considerations of the two counterclockwise angles in question, this never happens whenever one vertex lies in the interior of or on the boundary of the triangle formed by the other three vertices. We encourage the reader to fill in the details.

We have mentioned this so that, when we assume that $(a; b; c; d)$ is a positive or negative real number, we can safely work on only the cases where $A; B; C; D$ form a convex quadrilateral in some order. In those cases, there are $4! = 24$ ways in which the vertices can be ordered; due to the symmetry of rotations, we can assume the "first" vertex is A; which thankfully gives 6 cases instead of 24 where we go counterclockwise from A:

$$ABCD; ABDC; ACDB; ADCB \text{ and } ACBD; ADBC.$$

In the first four cases $C; D$ lie on the same side of the line through AB; and in the last two cases $C; D$ lie on opposite sides of the line through AB.



Thanks to reflections, we can actually cut the cases down to three, as in every alternate one shown. We are finally ready to prove the two parts of the theorem. In each step of the argument, the reader should verify that it make sense in the relevant diagrams.

1. Suppose A, B, C, D are concyclic and C, D lie on the same side of AB. By a criterion for cyclic quadrilaterals, $\angle ACB = \angle ADB$ and so

$$\arg\left(\frac{b-c}{a-c}\right) = \arg\left(\frac{b-d}{a-d}\right),$$

as this arg equation is equivalent to $\angle ACB = \angle ADB$ or $2\pi - \angle ACB = 2\pi - \angle ADB$, where we are referring to the non-reflex versions of $\angle ACB$ and $\angle ADB$.

Conversely, if this arg equation holds, it means the counterclockwise angle resulting from rotating AC around C to coincide with BC is equal to the counterclockwise angle resulting from rotating AD around D to coincide with BD. Suppose for contradiction that C, D lie on opposite sides of AB. Then one of the aforementioned counterclockwise angles would be less than π and the other would be greater than π, meaning they cannot be equal. So C, D lie on the same side of the line through AB. By the arg equation, the non-reflex $\angle ACB$ and $\angle ADB$ are equal, as $\angle ACB$ and $\angle ADB$ are respectively $\arg\left(\frac{b-c}{a-c}\right)$ and $\arg\left(\frac{b-d}{a-d}\right)$, or respectively $2\pi - \arg\left(\frac{b-c}{a-c}\right)$ and $2\pi - \arg\left(\frac{b-d}{a-d}\right)$. Then a criterion for cyclic quadrilaterals tells us that A, B, C, D are concyclic.

So the initial conditions are equivalent to

$$0 = \arg\left(\frac{b-c}{a-c}\right) - \arg\left(\frac{b-d}{a-d}\right) = \arg\left(\frac{b-c}{a-c} \Big/ \frac{b-d}{a-d}\right),$$

which is equivalent to $(a, b; c, d)$ being a positive real number.

2. Suppose A, B, C, D are concyclic and C, D lie on opposite sides of $AB.$ By a criterion for cyclic quadrilaterals, $\angle ACB + \angle ADB = \pi$ and so

$$\arg\left(\frac{b-c}{a-c}\right) + \arg\left(\frac{a-d}{b-d}\right) \equiv \pi,$$

as this congruence is equivalent to $\angle ACB + \angle ADB \equiv \pi$ or $(2\pi - \angle ACB) + (2\pi - \angle ADB) \equiv \pi,$ where we are referring to the non-reflex version of $\angle ACB$ and $\angle ADB.$

Conversely, if this arg congruence holds, it means the counterclockwise angle resulting from rotating AC around C to coincide with BC plus the counterclockwise angle resulting from rotating BD around D to coincide with AD is congruent to a flat angle. Suppose for the sake of contradiction that C, D lie on the same side of the line through AB. Then one of the aforementioned counterclockwise angles will lie in $(0, \pi)$ and the other will lie in $(\pi, 2\pi).$ Then their sum minus π will lie in $(0, 2\pi),$ contradicting the fact that it should be congruent to 0 modulo $2\pi.$ So C, D lie on opposite sides of $AB.$ By the arg congruence, the non-reflex $\angle ACB$ and $\angle ADB$ satisfy $\angle ACB + \angle ADB \equiv \pi,$ as $\angle ACB$ and $\angle ADB$ are respectively $\arg\left(\frac{b-c}{a-c}\right)$ and $\arg\left(\frac{a-d}{b-d}\right),$ or respectively $2\pi - \arg\left(\frac{b-c}{a-c}\right)$ and $2\pi - \arg\left(\frac{a-d}{b-d}\right).$ This congruence is actually the equality $\angle ACB + \angle ADB = \pi$ because both of the angles on the left lie in $(0, \pi),$ which makes their sum lie in $(0, 2\pi).$ Then a criterion for cyclic quadrilaterals tells us that A, B, C, D are concyclic.

So the initial conditions are equivalent to

$$\arg\left(\frac{b-c}{a-c}\right) + \arg\left(\frac{a-d}{b-d}\right) \equiv \arg\left(\frac{b-c}{a-c} \cdot \frac{a-d}{b-d}\right) \equiv \pi,$$

which is equivalent to (a, b, c, d) being a negative real number.

□

Theorem 8.10 (Converse of power of a point). Let A, B, C, D be distinct points in the plane that are not all collinear. Let P be the intersection of the lines through AB and $CD.$ Suppose P either lies on both of the segments AB and CD or neither. If

$$PA \cdot PB = PC \cdot PD,$$

then A, B, C, D are concyclic, not necessarily in that order.

Proof. We will use complex numbers. Let the four points be a, b, c, d, respectively. Since the setup can be translated without issue, let P be the complex origin 0. Knowing that $a, b, 0$ are collinear and that $c, d, 0$ are collinear, there exist real numbers x, y such that

$$a = bx,$$
$$c = dy.$$

Due to the fact that

$$PA \cdot PB = PC \cdot PD;$$

we know that P cannot equal any of the $A;B;C;D$ because it would prevent at least two of these four points from being distinct. So P is either in the interior of both segments $AB;CD$ or in the exterior of both segments. If 0 (as in, P) lies on the segment AB and on the segment CD then $x;y$ are both negative since a negative dilation is required to send b to a and, similarly, a negative dilation is required to to send d to c. On the other hand, if 0 lies on neither the segment AB nor the segment CD, then the negative dilations are replaced by positive dilations. Either way, xy is a positive real number, since the signs of $x;y$ are the same. The last bit of preliminary observation is that the equation

$$PA \cdot PB = PC \cdot PD$$

is equivalent to

$$a\overline{a}b\overline{b} = c\overline{c}d\overline{d}$$

or

$$\frac{ab}{cd} = \overline{\left(\frac{cd}{ab}\right)}:$$

By the complex concyclicity criterion, it suffices to prove that $(a;c;b;d) \in \mathbb{R}$. To that end, we compute

$$\begin{aligned}
(a;c;b;d) &= \frac{c-b}{a-b} \cdot \frac{a-d}{c-d} \\
&= \frac{c-b}{bx-b} \cdot \frac{a-d}{dy-d} \\
&= \frac{c-b}{b(x-1)} \cdot \frac{a-d}{d(y-1)} \\
&= \frac{c-b}{b} \cdot \frac{a-d}{d} \cdot \frac{1}{x-1} \cdot \frac{1}{y-1} \\
&= \left(\frac{c}{b}-1\right)\left(\frac{a}{d}-1\right) \cdot \frac{1}{x-1} \cdot \frac{1}{y-1} \\
&= \left(\frac{ac}{bd}-\frac{a}{d}-\frac{c}{b}+1\right) \cdot \frac{1}{x-1} \cdot \frac{1}{y-1}:
\end{aligned}$$

Since $x;y$ are defined to be real and

$$\frac{ac}{bd} = \frac{a}{b} \cdot \frac{c}{d} = xy$$

is real, it suffices to prove that $\frac{a}{d}+\frac{c}{b}$ is real. At this point, our tactic will be to prove that $\left(\frac{a}{d}+\frac{c}{b}\right)^2$ is real and non-negative, which will imply $\frac{a}{d}+\frac{c}{b}$ (without the square) is real. First we note that

$$\frac{a}{d}+\frac{c}{b} = \frac{bx}{\left(\frac{c}{y}\right)} + \frac{dy}{\left(\frac{a}{x}\right)} = \left(\frac{d}{a}+\frac{b}{c}\right) xy:$$

Since xy is already real and positive, it suffices to prove that

$$\left(\frac{a}{d}+\frac{c}{b}\right)\left(\frac{d}{a}+\frac{b}{c}\right)=\frac{ab}{cd}+\frac{cd}{ab}+2$$

is real and non-negative. Since

$$|a|\cdot|b|=|c|\cdot|d| \implies \left|\frac{ab}{cd}\right|=1,$$

we know that $\frac{ab}{cd}=e^{i\theta}$ lies on the complex unit circle. Therefore,

$$\begin{aligned}\frac{ab}{cd}+\frac{cd}{ab} &= e^{i\theta}+\frac{1}{e^{i\theta}}\\ &= e^{i\theta}+e^{i(-\theta)}\\ &= (\cos\theta+i\sin\theta)+(\cos(-\theta)+i\sin(-\theta))\\ &= (\cos\theta+i\sin\theta)+(\cos\theta-i\sin\theta)\\ &= 2\cos\theta \ge -2.\end{aligned}$$

This completes the proof. □

Theorem 8.11 (Ptolemy's inequality). If A, B, C, D are four distinct points in the plane, then

$$AB\cdot CD+BC\cdot DA \ge AC\cdot BD.$$

If they are not all collinear, then equality holds if and only if A, B, C, D are concyclic in that clockwise or counterclockwise order. The equality criterion is called **Ptolemy's theorem**.

Proof. We place A, B, C, D on the complex plane by letting them be represented by the distinct points $0, z_1, z_2, z_3$ respectively, without loss of generality. Then we wish to show that

$$|z_1|\cdot|z_2-z_3|+|z_1-z_2|\cdot|z_3| \ge |z_2|\cdot|z_1-z_3|.$$

Dividing both sides by $|z_1z_2z_3|$ yields the equivalent inequality

$$\left|\frac{1}{z_3}-\frac{1}{z_2}\right|+\left|\frac{1}{z_2}-\frac{1}{z_1}\right| \ge \left|\frac{1}{z_3}-\frac{1}{z_1}\right|.$$

This is true by the complex triangle inequality. Since neither term on the left side is 0 due to the distinctness of the points, equality holds if and only if

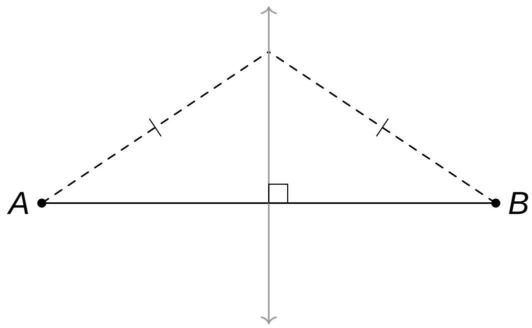
$$\frac{\frac{1}{z_3}-\frac{1}{z_2}}{\frac{1}{z_2}-\frac{1}{z_1}}=\frac{z_2-z_3}{0-z_3}\cdot\frac{0-z_1}{z_2-z_1}=-(0, z_2, z_3, z_1)$$

is a positive real number, by the equality criterion of the complex triangle inequality (see Volume 1). Equivalently, $(0, z_2, z_3, z_1)$ is a negative real number. Assuming A, B, C, D are not all collinear, the complex concyclicity criterion tells us that this is true if and only if $0, z_1, z_2, z_3$ are concyclic such that z_1, z_3 lie on opposite sides of the line through 0 and z_2. This is the result we seek. □

8.2 Cyclic and Tangential Polygons

Definition 8.12. The **perpendicular bisector** of a line segment is the line that runs through the midpoint of the segment and is perpendicular to the segment.

Theorem 8.13. Given two distinct points A and B, the locus of points that are equidistant from A and B is the perpendicular bisector of AB.



Proof. Suppose P is a point that is equidistant from A and B. Then $\triangle APB$ is isosceles with $PA = PB$. Let F be the foot of the perpendicular from P to AB. By HL congruence, $\triangle APF \cong \triangle BPF$, which implies $AF = BF$. Thus, P lies on the perpendicular bisector of AB.

For the other inclusion, suppose Q lies on the perpendicular bisector of AB. Let the midpoint of AB be M, so that $AM = BM$. By SAS congruence, $\triangle AMQ \cong \triangle BMQ$, which implies $QA = QB$. Thus, Q is equidistant from A and B. □

Definition 8.14. A set of points is said to be **concyclic** if there exists a circle on which they all lie. A **cyclic polygon** is one whose vertices are concyclic, where the circle on which all the vertices can be placed is called the **circumcircle**. In other words, there is a point called the **circumcenter** from which all the vertices are at a fixed distance called the **circumradius**. Every cyclic polygon is convex because each interior angle is an inscribed angle, which must be half of the corresponding central angle; since the central angle must be less than 360°, each interior angle is less than 180°.

Definition 8.15. A collection of lines, rays or line segments is said to **concur** or be **concurrent** if there is a point that lies on all of them. This point is called a **point of concurrency**.

Theorem 8.16. A polygon is cyclic if and only if the perpendicular bisectors of all of its edges concur at some point. If this point exists, it is the unique circumcenter.

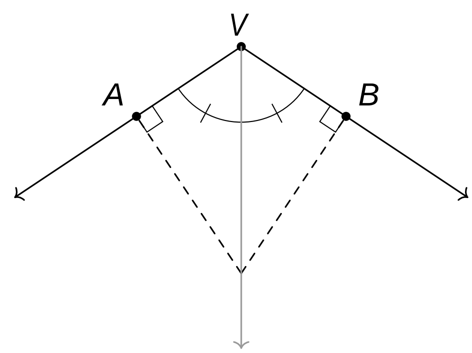
Proof. If the circumcenter exists, then it is a point P that is equidistant from every vertex, which means P is equidistant from the endpoints of every edge, and so P lies on the perpendicular bisector of every edge, by Theorem 8.13.

Conversely, if the perpendicular bisectors of all of the edges concur at Q, then Q must be equidistant from the endpoints of every edge, by Theorem 8.13. This means Q is equidistant from every vertex, which means it is a circumcenter. Thus, P and Q are the same point if either exists.

Suppose for contradiction that there are two distinct circumcenters. Then the perpendicular bisectors of the edges have more than one intersection point, causing them to all lie on the same line. This would mean that each pair of edges is parallel or they lie on the same line, both of which are impossible since consecutive edges share a vertex and create an interior angle less than 180° in a convex polygon. Thus, the circumcenter is unique, and so is the circumcircle since the uniqueness of the circumcenter implies the uniqueness of the circumradius. □

Definition 8.17. The **angle bisector** of an angle is the ray with the same vertex as the angle and, with the exception of that vertex, the ray lies in the interior of the angle such that it splits the angle into two equal angles.

Theorem 8.18. For a non-reflex angle, the locus of all points in the interior of the angle, the feet of whose perpendicular segments to the line through each ray lies on the ray, and that are equidistant from the rays of the angle is the angle bisector of the angle without its endpoint.



Proof. Let the vertex of the non-reflex angle be V. Suppose P is a point such that the feet of its perpendicular segments to the line through each ray lies on the ray, and P is equidistant from the rays of the angle. Let A be one foot and let B be the other foot. By HL congruence, $\triangle PAV \cong \triangle PBV$, which implies $\angle PVA = \angle PVB$. Thus, P is on the angle bisector of $\angle AVP$.

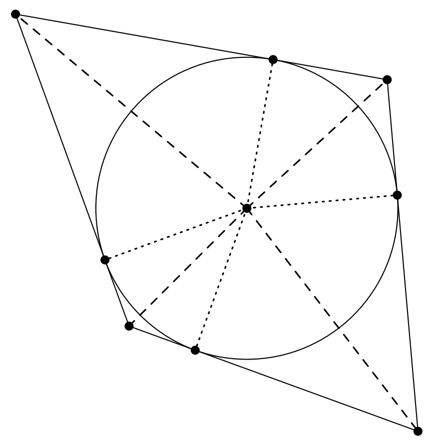
For the other inclusion, suppose Q lies on the angle bisector of the angle without its endpoint. Then Q lies in the interior of the angle. Let ℓ be a line through Q that intersects the rays at C and D such that CD is perpendicular to VQ. The angles $\angle CVQ$ and $\angle DVQ$ are each half of the original non-reflex angle, so they are both acute; then their complementary angles $\angle VCQ$ and $\angle VDQ$ are also acute. This means the perpendicular segments from Q to the lines through VC and VD lie on the respective segments, which means the perpendicular segments from Q to the line through each ray lies on the ray. Let the foot of the perpendicular segment from Q to VC be X and the foot of the perpendicular segment from Q to VD be Y. By AAS congruence, $\triangle VXQ \cong \triangle VYQ$ and so $QX = QY$. Thus, Q is equidistant from the rays of the angle. □

Definition 8.19. We already know what it means for a line to be tangent to a circle. On the other hand, we say that a line segment is **tangent** to a circle if the line through the segment is tangent to the circle and the point of tangency lies on the segment.

Definition 8.20. A **tangential** polygon is one whose edges are all tangent to the same circle, called the **incircle**. In other words, there is a point called the **incenter** which is at a fixed perpendicular distance called the **inradius** from all the edges. The interior of the incircle lies in the interior of the polygon. Also, every tangential polygon is convex because each interior angle is supplementary with the opposite central angle, and so the interior angle cannot exceed 180 .

Notice that the point at which the incircle is tangent to an edge must lie in the interior of the edge, because, if it lay on a vertex, then the two edges emanating from the vertex would form a straight angle.

Theorem 8.21. A polygon is tangential if and only if it is convex and the angle bisectors of its interior angles concur at some point. If this point exists, it is the unique incenter.



Proof. Recall from Theorem 5.17 that a convex polygon is equal to the region formed by the intersection of its interior angles, from which we can derive that the intersection of the interiors of the interior angles is equal to the interior of the polygon. If the polygon is tangential, then convexity is guaranteed; moreover, there is an incenter P in the interior of the polygon, the foot of whose perpendicular segment to the line through each edge lies on the edge, and P is equidistant from every edge. By convexity, P is in the interior of every interior angle, and by the preceding statement, P is equidistant from every pair of consecutive edges. So P lies on the angle bisector of every interior angle.
Conversely, suppose the polygon is convex and the angle bisectors of the interior angles concur at Q. By convexity, a point, such as Q; that is in the interior of all the interior angles is in the interior of the polygon. Since the angle bisectors of the interior angles concur at Q, the perpendicular segment from Q to the line running through each edge falls on the edge (this is because half of each interior angle of a convex polygon is necessarily acute) and Q is equidistant from every pair of consecutive edges. So Q must be equidistant from every edge, which means it is an incenter. Thus, P and Q are the same point if either exists.
Suppose for contradiction that there are two distinct incenters. Then the angle bisectors of the interior angles have more than one intersection point, causing them to all lie on the same line. This would mean all the vertices of the polygon are collinear, which is a contradiction. Thus, the incenter is unique, and so is the incircle because the uniqueness of the incenter implies uniqueness of the inradius. □

Problem 8.22. If the sides of a tangential polygon P are $s_0, s_1, \ldots, s_{n-1}$ in clockwise or counterclockwise order, then show that there exists a list of n positive real numbers $(t_0, t_1, \ldots, t_{n-1})$ such that $s_i = t_i + t_{i+1}$ for each $0 \le i \le n-1$, where indices are reduced modulo n.

Problem 8.23. Show that the area of a tangential polygon is rs, where r is the inradius and s is the semiperimeter (half the perimeter).

Theorem 8.24 (Pitot's theorem). If a quadrilateral $ABCD$ is tangential then

$$AB + CD = AD + BC.$$

Conversely, if the sum of the lengths of two opposite edges of a convex quadrilateral equals the sum of the lengths of the other two opposite edges, then the quadrilateral is tangential.

Proof. Suppose $ABCD$ is tangential and let $AB, BC, CD,$ and DA be tangent to the incircle at $A', B', C',$ and D' respectively. Using the fact that both tangent segments from the same external point have the same length,

$$\begin{aligned} AB + CD &= (AA' + A'B) + (CC' + C'D) \\ &= (AD' + BB') + (CB' + DD') \\ &= (AD' + D'D) + (BB' + B'C) \\ &= AD + BC. \end{aligned}$$

There exists a natural "proof" of the converse that is shown in several well-known source, such as [1], but it contains a hidden mistake that was explored by Alexander Bogomolny in [3]. A rigorous proof, which requires carefully considering the implications of the convexity assumption, is explained in detail in [6]. □

Problem 8.25. According to Definition 7.25, a kite is a convex quadrilateral with two disjoint pairs of sides such that each pair consists of adjacent sides that are equal. Without using the converse of Pitot's theorem, show that kites, and therefore rhombuses, are tangential.

Definition 8.26. A polygon is bicentric if it is both cyclic and tangential. Two circles are said to be concentric if their centers coincide.

Problem 8.27. According to Definition 5.30, a regular polygon is a polygon whose sides are all equal and all of its interior angles are equal. Show that all regular polygons are bicentric, with the circumcircle and incircle being concentric. However, the circumcircle and incircle are not necessarily concentric in a bicentric polygon, as will be evident when we study triangle centers in Chapter 11.

Chapter 9

Area

> "And purchased a site, which was named 'Bull's Hide' after the bargain by which they should get as much land as they could enclose with a bull's hide."
>
> – *Virgil's Aeneid*

We introduce the concept of area as a measure of subsets of two-dimensional space. In the first section, area formulas are developed for specific polygons in the plane using their side lengths. In the second section, general "shoelace" formulas are developed for computing areas using Cartesian, complex, and barycentric coordinates.

9.1 Using Lengths

Definition 9.1. Intuitively, the area of a polygon is a numerical measure of the set of points that it occupies in the plane. This is the number of 1×1 unit squares that would fill the polygon, including partial unit squares. We denote the area of a region P in the plane using square brackets like $[P]$. We will not define area more precisely, nor will we make it clear which subsets of the plane can actually be assigned area, as both tasks would require measure theory.

Theorem 9.2. The notion of area satisfies the following properties:

1. Partitioning regions: If we partition a region into finitely many regions (i.e. split into non-overlapping regions, except perhaps at boundaries) whose areas can be calculated, like triangles and special convex quadrilaterals, then the area of the original region is the sum of the areas of the new smaller regions.

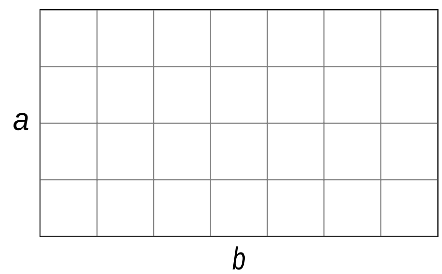
2. Complementary regions: If we place a region R inside a larger region S such that each of S and the excess parts T (i.e. the region outside R but inside S) have areas that can be calculated, then $[R] = [S] - [T]$.

3. Overlapping regions: If we express a region as the finite union of potentially overlapping regions, where the areas of the new regions and their intersections can be calculated, then the area of the original region is the sum of the new smaller regions, with subtracting and adding overlaps, as needed. The counting becomes more elaborate as an increasing number of regions overlap in places.

4. Tiling regions: If we copy and paste n congruent copies of a region R together in a non-overlapping way (except possibly at boundaries) to produce a region S whose area can be calculated, then $[R] = \frac{[S]}{n}$.

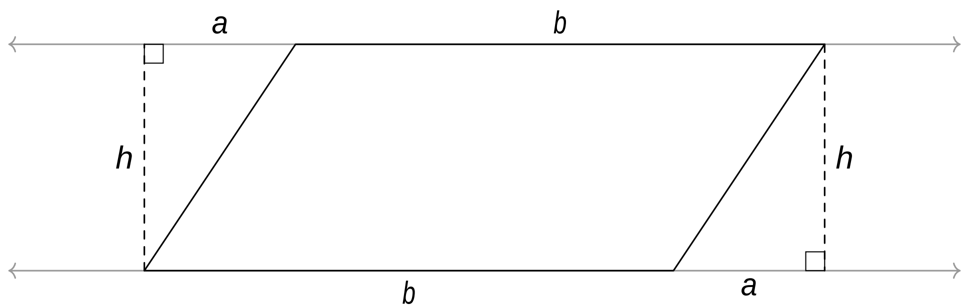
These principles pertaining to areas can be especially helpful for finding the areas of non-standard regions. The same techniques can be applied to surface area and volume calculations in 3D geometry. The above methods are analogous to combinatorial principles from Volume 2, specifically the addition principle, subtraction principle, the principle of inclusion-exclusion, and the division principle, respectively.

Definition 9.3. We define the **area of a rectangle** with side lengths a and b to be $ab.$ This is consistent with the fact that if a and b are integers, then there would fit exactly ab unit squares inside the rectangle. As a consequence, the area of a square with side length s is $s^2.$



Theorem 9.4. If two opposite bases of a parallelogram have length b and corresponding height $h,$ then its area is $bh.$

Proof. Suppose we have a parallelogram with parallel bases that have length b and a corresponding height $h.$ The parallelogram can be placed inside a rectangle by drawing two heights perpendicular to the bases and emanating from the outermost two opposite vertices. The two resulting right triangles are congruent by SAS congruence with legs of length a and $h,$ and they can be placed together to form an $a \times h$ rectangle.



Subtracting the area of the smaller constructed rectangle from the larger rectangle tells us that the area of the parallelogram is

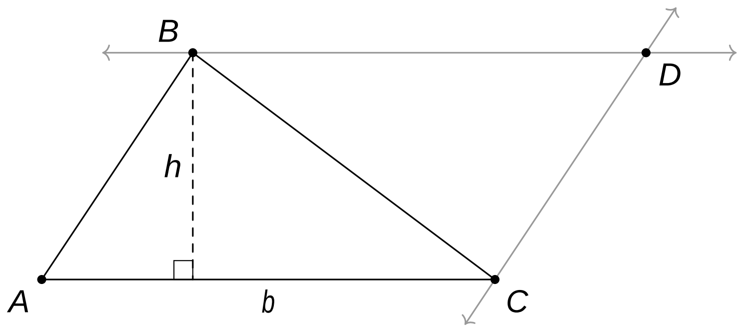
$$(a+b)h - ah = bh.$$

□

Theorem 9.5. If $\triangle ABC$ is a triangle whose height corresponding to base $AC = b$ is $h,$ then

$$[ABC] = \frac{bh}{2}.$$

Proof. Intuitively, we will draw a second copy of the triangle and put it together with the first to produce a parallelogram whose area we can calculate. More rigorously, we draw the line through B that is parallel to $AC;$ and then we draw the line through C that is parallel to $AB:$ Let the intersection of these two lines be $D:$ Then $ABDC$ is a parallelogram with bases $AC; BD$ of length b and a corresponding height $h:$



Moreover, by SSS congruence, $4ABC = 4DCB;$ so

$$[ABC] = \frac{[ABC] + [DCB]}{2} = \frac{[ABDC]}{2} = \frac{bh}{2}:$$

□

Corollary 9.6. The area of a right triangle with legs of length a and b is $\frac{ab}{2}:$

Corollary 9.7. The area of any two triangles with the same base and same height have the same area. In particular, if we fix a triangle ABC with base $AB;$ then for any C^0 on the line ` through C and parallel to $AB;$

$$[4ABC^0] = [4ABC]:$$

Corollary 9.8. If two triangles each have a base that has a certain line running through both of these cases and there is a common height corresponding to the bases, then the ratio of their areas is the ratio of the lengths of the bases, as long as the order of the triangles is preserved in both ratios.

Proof. Let the triangles be $4ABC$ and $4DEC$ with bases AB and DE that lie on the same line, and a common height h emanating from $C:$ Then

$$\frac{[ABC]}{[DEC]} = \frac{\left(\frac{AB\ h}{2}\right)}{\left(\frac{DE\ h}{2}\right)} = \frac{AB}{DE}:$$

□

Problem 9.9. An orthodiagonal convex quadrilateral is a convex quadrilateral whose diagonals are perpendicular. An example is a kite. Prove that the area of a convex orthodiagonal quadrilateral with diagonals of length c and d is $\frac{cd}{2}:$

Example 9.10. Show that if a triangle has heights f, g, h then

$$\frac{1}{f} + \frac{1}{g} > \frac{1}{h}.$$

Proof. Let the bases corresponding to the heights f, g, h be a, b, c respectively. Then three ways of expressing the area Δ of the triangle are

$$\Delta = \frac{af}{2} = \frac{bg}{2} = \frac{ch}{2}.$$

By the triangle inequality,

$$a + b > c,$$

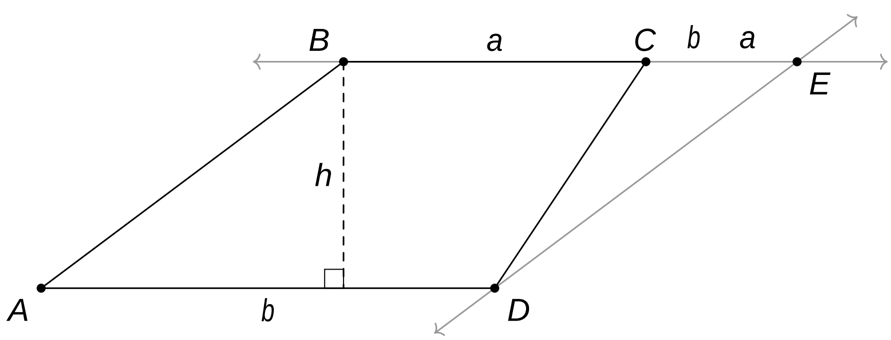
which we can rewrite as

$$\frac{2\Delta}{f} + \frac{2\Delta}{g} > \frac{2\Delta}{h}.$$

This is equivalent to what we want to prove because we can cancel the numerators. In general, it can be fruitful to find the area of a region in more than one way in order to gain information about variables involved. For example, a triangle has three pairs of corresponding bases and heights as we just showed, and a parallelogram has two pairs of two equal bases and a corresponding height. □

Theorem 9.11. If a trapezoid has bases with length a and b and a corresponding height h, then the area of the trapezoid is $\frac{(a+b)h}{2}$.

Proof. Let the trapezoid be $ABCD$ with bases $AD = b$ and $BC = a$. If $a = b$, then the trapezoid has two opposite sides that are parallel and equal in length, so it is parallelogram and the formula is easy to verify in this case. Now suppose $b > a$. We draw the line through D that is parallel to AB and the line through BC. Letting the intersection of the two lines be E, we get a parallelogram $ABED$. Conveniently, $\triangle CED$ has base $CE = BE - BC = b - a$ with the corresponding height h.



Using complementary regions, the area of the trapezoid is

$$[ABCD] = [ABED] - [CED] = bh - \frac{(b-a)h}{2} = \frac{(a+b)h}{2}.$$

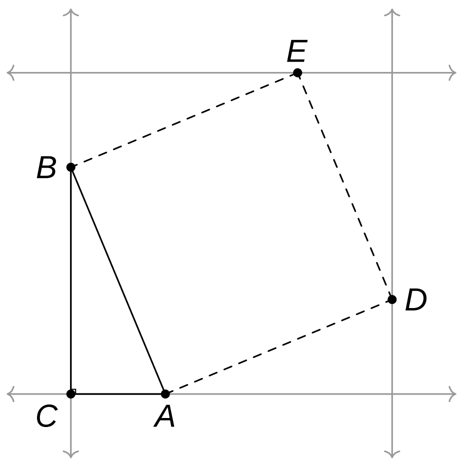
□

Theorem 9.12 (Pythagorean theorem). If a right triangle has legs of length a and b and hypotenuse c, then

$$a^2 + b^2 = c^2.$$

Conversely, if a, b, c are the side lengths of a triangle such that $a^2 + b^2 = c^2$, the triangle is a right triangle with a right angle opposite the side with lengths c.
Note that we took the Pythagorean theorem for triangles whose legs are parallel to the coordinate axes as the definition of Euclidean distance (Theorem 2.1), whereas the present result is for any right triangle.

Proof. The result was already known as a special case of the cosine law (Theorem 3.8), but it is worth showing the marvelous proof below. Let $\triangle ABC$ be a triangle with a right angle at vertex C, and side lengths $BC = a, CA = b, AB = c$. First we construct a square $ADEB$ on the side of the line through AB that does not contain C. Now we extend CA into a ray through A, and CB into a ray through B. Then we draw a line through D that is parallel to CB, and a line through E that is parallel to CA. This produces three more right triangles that are similar to $\triangle ABC$ by AA similarity. In fact, all four triangles are congruent by ASA congruence since they all have hypotenuses of equal length c. The final observation is that the legs of the four triangles produce a square, which we leave to the reader to easily verify by checking that the sides are all a result of 180° angles and that they have equal length $a + b$.



This allows us to write the area of the larger square is two ways:

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

Conversely, suppose $\triangle ABC$ is a triangle with side lengths $BC = a, CA = b, AB = c$ such that $a^2 + b^2 = c^2$. We construct a right triangle $\triangle PQR$ with a right angle at vertex R and side lengths $RP = b$ and $RQ = a$. Then the Pythagorean theorem tells us that

$$PQ = \sqrt{RP^2 + RQ^2} = \sqrt{a^2 + b^2} = c.$$

By SSS congruence, $\triangle ABC \cong \triangle PQR$, which means $\triangle ABC$ is a right triangle too with a right angle at vertex C. □

Corollary 9.13. In a right triangle, the hypotenuse is strictly longer than either leg.

Proof. Let the legs have length a and b and the hypotenuse have length c. By the Pythagorean theorem,

$$c = \sqrt{a^2 + b^2} > \sqrt{a^2} = a,$$
$$c = \sqrt{a^2 + b^2} > \sqrt{b^2} = b.$$

□

Theorem 9.14 (HL similarity). Two right triangles with hypotenuses and one pair of legs in the same ratio are similar.

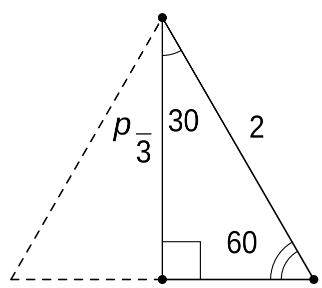
Proof. Let the length of the hypotenuses be h and kh for some positive constant k, and the legs in question have lengths l and kl. Then the remaining legs m and m' are related by

$$m' = \sqrt{(kh)^2 - (kl)^2} = k\sqrt{h^2 - l^2} = km,$$

where we have used the Pythagorean theorem. Now SSS similarity can be applied. □

Finally, we look at some special triangles and related lengths that, for example, help us in the evaluation of trigonometric functions at particular values (see Volume 1).

Theorem 9.15 ($30^\circ - 60^\circ - 90^\circ$ triangle). The height of an equilateral triangle with sides of length s is $\frac{\sqrt{3}s}{2}$. Thus, the area of such an equilateral triangle is $\frac{\sqrt{3}s^2}{4}$. Moreover, if a triangle has angles $30^\circ - 60^\circ - 90^\circ$, then the sides opposite those angles in that order are in the ratio $1 : \sqrt{3} : 2$.



Proof. By dropping a height, we can use AAS congruence to establish that the two new triangles are congruent right triangles. This means the height bisects the corresponding base. If the side length of the equilateral triangle is s, then, by the Pythagorean theorem, the height is $\sqrt{s^2 - \left(\frac{s}{2}\right)^2} = \frac{\sqrt{3}s}{2}$. (Note that the same technique can be used to calculate the height of an isosceles triangle that emanates from the apex.) As a consequence, the area of the equilateral triangle is

$$\frac{1}{2} \cdot \frac{\sqrt{3}s}{2} \cdot s = \frac{\sqrt{3}s^2}{4}.$$

As a second consequence, we can observe from either of the two new right triangles that the sides opposite the $30^\circ, 60^\circ, 90^\circ$ angles are in the ratio

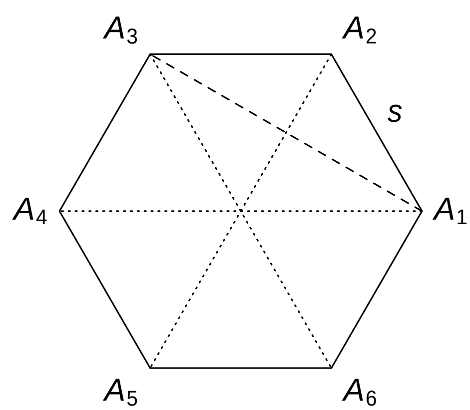
$$\frac{s}{2} : \frac{\sqrt{3}s}{2} : s = 1 : \sqrt{3} : 2.$$

□

Example 9.16. A regular hexagon is a six-sided convex polygon such that all of the side lengths are equal and all of the interior angles are equal. Determine the area of a regular hexagon with sides of length $s.$ Moreover, if the regular hexagon is $A_1A_2A_3A_4A_5A_6,$ then determine the length of A_1A_3 in terms of $s.$

Solution. The key is to notice that we can produce such a hexagon by gluing together six equilateral triangles that have sides of length $s.$ Indeed, since equilateral triangles have interior angles measuring $\frac{180^\circ}{3} = 60^\circ,$ the central angle in such a construction will be $60^\circ \times 6 = 360^\circ,$ so there will be no gaps. Moreover, all of the interior angles of the resulting six-sided convex polygon will be $2 \times 60^\circ = 120^\circ,$ and it will have six sides of length $s.$ By Theorem 9.15, the area of the hexagon is

$$6\left(\frac{\sqrt{3}s^2}{4}\right) = \frac{3\sqrt{3}}{2}s^2.$$



For the length of $A_1A_3,$ we notice that $A_1A_2 = A_3A_2,$ so we drop an altitude from A_2 to $A_1A_3.$ By HL congruence, the resulting two triangles are congruent. The two angles at A_2 measure $\frac{120^\circ}{2} = 60^\circ,$ so

$$A_1A_3 = 2\left(\frac{\sqrt{3}s}{2}\right) = \sqrt{3}s.$$

□

Problem 9.17. Let $A_1A_2A_3 \ldots A_{12}$ be a regular dodecagon, which is a twelve-sided convex polygon such that all side lengths are equal to s and all interior angles are equal. Determine the length of A_1A_4 in terms of s.

Theorem 9.18 ($45^\circ - 45^\circ - 90^\circ$ triangle). The sides opposite the $45^\circ, 45^\circ, 90^\circ$ angles respectively in an isosceles right triangle are in the ratio $1 : 1 : \sqrt{2}.$

Proof. A square with side length s can be split along a diagonal to construct a $45^\circ-45^\circ-90^\circ$ triangle, whose diagonal has length

$$\sqrt{s^2+s^2} = s\sqrt{2}$$

by the Pythagorean theorem. Thus, the sides opposite the $45^\circ, 45^\circ, 90^\circ$ angles are in the ratio

$$s : s : \sqrt{2}s = 1 : 1 : \sqrt{2}.$$

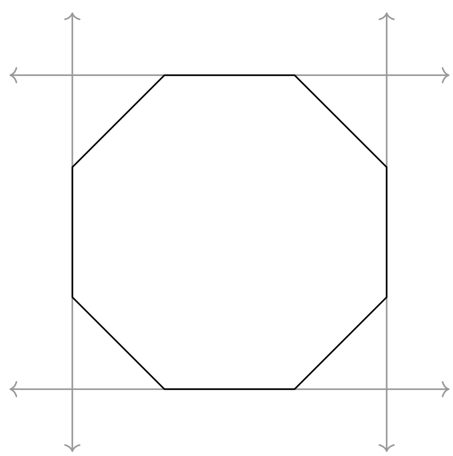
□

Example 9.19. A **regular octagon** is an eight-sided convex polygon such that all of the side lengths are equal and all of the interior angles are equal. Determine the area of a regular octagon with sides of length s.

Solution. Each interior angle of a regular octagon is be calculated to be

$$\frac{180^\circ(8-2)}{8} = 135^\circ.$$

The complement of this angle is 45°, so extending every alternate edge into a line produces four isosceles right triangles and an overarching square.



The triangles have area

$$\frac{1}{2} \cdot \frac{s}{\sqrt{2}} \cdot \frac{s}{\sqrt{2}} = \frac{s^2}{4}$$

and the overarching square has area

$$\left(s + 2\cdot\frac{s}{\sqrt{2}}\right)^2 = (s+\sqrt{2}s)^2 = s^2(3+2\sqrt{2}).$$

By complementary regions, the area of the regular octagon is

$$s^2(3+2\sqrt{2}) - 4\cdot\frac{s^2}{4} = 2s^2\cdot(1+\sqrt{2}).$$

□

9.2 Using Coordinates

Although determinants will feature prominently in the presentation, we will use little to no machinery from linear algebra. The main purpose of determinants in the material here is to present certain algebraic expressions in an organized fashion.

Definition 9.20. For a generalized polygon P, we define its sign as

$$\operatorname{sgn}(P) = \begin{cases} 1 & \text{if } P \text{ is oriented counterclockwise} \\ -1 & \text{if } P \text{ is oriented clockwise} \end{cases}.$$

This is well-defined concept, by Definition 5.15.

Lemma 9.21. Let $A = (0,0), B = (x_b, y_b), C = (x_c, y_c)$, which are allowed to be collinear. Then

$$[ABC] = \operatorname{sgn}(ABC) \cdot \frac{1}{2} \cdot \det \begin{bmatrix} x_b & x_c \\ y_b & y_c \end{bmatrix}.$$

Proof. If A, B, C are collinear, then we have a degenerate triangle which should have area 0. Indeed, since collinearity implies the equality of the slopes

$$\frac{y_b - 0}{x_b - 0} = \frac{y_c - 0}{x_c - 0},$$

we get $x_b y_c - x_c y_b = 0$, as desired. The case where B and C are both on the y-axis can be handled separately with ease.

Now suppose A, B, C are not collinear. Let θ be the measure of the counterclockwise rotation around A that, along with a positive dilation from A, causes B to coincide with C. Since interior angles of triangles measure less than π, the interior angle of $\triangle ABC$ at A is

$$\angle BAC = \begin{cases} \theta & \text{if } \operatorname{sgn}(ABC) = 1 \\ 2\pi - \theta & \text{if } \operatorname{sgn}(ABC) = -1 \end{cases}.$$

As a result, letting $h = AB \cdot \sin \angle BAC$ be the height emanating from B and with foot on AC,

$$\begin{aligned} [ABC] &= \frac{AC \cdot h}{2} \\ &= \frac{1}{2} \cdot AB \cdot AC \cdot \sin \angle BAC \\ &= \operatorname{sgn}(ABC) \cdot \frac{1}{2} \cdot AB \cdot AC \cdot \sin \theta. \end{aligned}$$

So now it suffices to show that

$$AB \cdot AC \cdot \sin \theta = \det \begin{bmatrix} x_b & x_c \\ y_b & y_c \end{bmatrix}.$$

Let $b = x_b + iy_b$ and $c = x_c + iy_c$. By the definition of θ, we know that $\frac{c}{b} = \frac{|c|}{|b|}e^{i\theta}$. By writing out the components of each complex number, we get the equation

$$\begin{aligned} |b|(x_c + iy_c) &= |c|(x_b + iy_b)(\cos\theta + i\sin\theta) \\ &= |c|((x_b\cos\theta - y_b\sin\theta) + i(x_b\sin\theta + y_b\cos\theta)). \end{aligned}$$

Equating real parts and imaginary parts yields the system of equations

$$\begin{aligned} |b|x_c &= |c|x_b\cos\theta - |c|y_b\sin\theta, \\ |b|y_c &= |c|x_b\sin\theta + |c|y_b\cos\theta. \end{aligned}$$

Subtracting y_b times the first equation from x_b times the second equation proves that

$$|b|(x_b y_c - x_c y_b) = |c|(x_b^2 + y_b^2)\sin\theta,$$

which finally implies

$$\det\begin{bmatrix} x_b & x_c \\ y_b & y_c \end{bmatrix} = |c| \cdot \frac{|b|^2}{|b|}\sin\theta = |b| \cdot |c| \cdot \sin\theta.$$

□

Theorem 9.22 (Shoelace formula for triangles). Let $A = (x_0, y_0), B = (x_1, y_1), C = (x_2, y_2)$, which are allowed to be collinear. Then

$$\begin{aligned} 2 \cdot \operatorname{sgn}(ABC) \cdot [ABC] &= \det\begin{bmatrix} x_1 - x_0 & x_2 - x_0 \\ y_1 - y_0 & y_2 - y_0 \end{bmatrix} \\ &= \det\begin{bmatrix} 1 & 1 & 1 \\ x_0 & x_1 & x_2 \\ y_0 & y_1 & y_2 \end{bmatrix} \\ &= \det\begin{bmatrix} x_0 & x_1 \\ y_0 & y_1 \end{bmatrix} + \det\begin{bmatrix} x_1 & x_2 \\ y_1 & y_2 \end{bmatrix} + \det\begin{bmatrix} x_2 & x_0 \\ y_2 & y_0 \end{bmatrix}. \end{aligned}$$

Moreover, A, B, C are collinear in some order if and only if this quantity is 0.

Proof. Any translation sends a triangle to a congruent triangle with the same orientation, and congruent triangles have the same area, due to to SAS congruence and the sine area formula derived in Lemma 9.21. So we apply the translation $(x, y) \mapsto (x, y) - (x_0, y_0)$ to get the three new points $A' = (0, 0), B' = (x_1 - x_0, y_1 - y_0), C' = (x_2 - x_0, y_2 - y_0)$. By Lemma 9.21,

$$2 \cdot \operatorname{sgn}(ABC) \cdot [ABC] = 2 \cdot \operatorname{sgn}(A'B'C') \cdot [A'B'C'] = \det\begin{bmatrix} x_1 - x_0 & x_2 - x_0 \\ y_1 - y_0 & y_2 - y_0 \end{bmatrix}.$$

It is then immediately true that the other two forms are equal to this quantity by applying the formulas for computing determinants.

Note that A, B, C are collinear if and only if $[ABC] = 0$, which gives us the nice Cartesian collinearity criterion

$$\det \begin{pmatrix} 1 & 1 & 1 \\ x_0 & x_1 & x_2 \\ y_0 & y_1 & y_2 \end{pmatrix} = 0.$$

This is a symmetric criterion, as opposed to the criterion stating that the lines between CA and BA have equal slopes. □

Corollary 9.23 (Polygon shoelace formula). Let $P = V_0V_1 \cdots V_{n-1}$ be a generalized n-gon with each vertex V_k having coordinates (x_k, y_k). Then

$$2 \cdot \operatorname{sgn}(P) \cdot [P] = \sum_{k=0}^{n-1} \det \begin{pmatrix} x_k & x_{k+1} \\ y_k & y_{k+1} \end{pmatrix},$$

where indices are reduced modulo n. The name of this result is due to the crisscross nature of this series of 2×2 determinants.

Proof. The proof is by induction on n. We already know the result for triangles, that is $n = 3$ from Theorem 9.22. Now suppose the result holds for generalized n-gons for some $n \geq 3$ and let P be a generalized $(n+1)$-gon. Then P has an ear, so label P as $V_0V_1 \ldots V_n$ in either counterclockwise or clockwise orientation such that V_n is an ear of P. Let $V_k = (x_k, y_k)$ for each index $0 \leq k \leq n$. The ear V_n induces the triangle $T = V_0V_{n-1}V_n$ and clipping this ear yields the generalized n-gon $Q = V_0V_1 \ldots V_{n-1}$. Recall that T and Q both have the same orientation as P. By the induction hypothesis for n-gons and the base case for triangles,

$$\begin{aligned} 2 \cdot \operatorname{sgn}(P) \cdot [P] &= 2 \cdot \operatorname{sgn}(P)([Q] + [T]) \\ &= 2 \cdot \operatorname{sgn}(Q) \cdot [Q] + 2 \cdot \operatorname{sgn}(T) \cdot [T] \\ &= \sum_{k=0}^{n-1} \det \begin{pmatrix} x_k & x_{k+1} \\ y_k & y_{k+1} \end{pmatrix} + \det \begin{pmatrix} x_0 & x_{n-1} \\ y_0 & y_{n-1} \end{pmatrix} + \det \begin{pmatrix} x_{n-1} & x_n \\ y_{n-1} & y_n \end{pmatrix} + \det \begin{pmatrix} x_n & x_0 \\ y_n & y_0 \end{pmatrix} \\ &= \sum_{k=0}^{n-2} \det \begin{pmatrix} x_k & x_{k+1} \\ y_k & y_{k+1} \end{pmatrix} + \det \begin{pmatrix} x_{n-1} & x_n \\ y_{n-1} & y_n \end{pmatrix} + \det \begin{pmatrix} x_n & x_0 \\ y_n & y_0 \end{pmatrix} \\ &= \sum_{k=0}^{n} \det \begin{pmatrix} x_k & x_{k+1} \\ y_k & y_{k+1} \end{pmatrix}. \end{aligned}$$

Note that we sneakily cancelled the two terms

$$\det \begin{pmatrix} x_{n-1} & x_0 \\ y_{n-1} & y_0 \end{pmatrix} + \det \begin{pmatrix} x_0 & x_{n-1} \\ y_0 & y_{n-1} \end{pmatrix}$$

in the middle of the computation. The reader should algebraically verify that their sum is in fact zero, though it is immediate from a property of determinants that states that swapping columns in a matrix negates the determinant.

One issue is that our derivation of this formula seems to rely on labelling $P = V_0V_1 \cdots V_n$ in a way that V_n is an ear. This is not the case because the formula is cyclic in the sense that, for any integer i,

$$\sum_{k=0}^{n} \det \begin{pmatrix} x_k & x_{k+1} \\ y_k & y_{k+1} \end{pmatrix} = \sum_{k=0}^{n} \det \begin{pmatrix} x_{k+i} & x_{k+i+1} \\ y_{k+i} & y_{k+i+1} \end{pmatrix}.$$

So it acceptable to relabel $V_0, V_1, \ldots, V_n$ as $V_i, V_{1+i}, \ldots, V_{n+i}$, where indices are reduced modulo n. Moreover, the derivation works for both types of orientations for P. Thus, the formula holds for any labelling of P and the induction is complete. □

Corollary 9.24. Suppose P and Q are similar generalized polygons in the plane such that the ratio of the lengths of Q to the lengths of P is $k > 0$. Then

$$[Q] = k^2 \cdot [P].$$

In particular, if P and Q are congruent, then $k = 1$ and they have the same area.

Proof. Suppose P and Q are as stated, each with n vertices. By Theorem 3.3, there exists a similarity transformation of P to Q that consists of exactly one homothety of factor k from the origin, followed by a Euclidean isometry. By the quoted result, every Euclidean isometry can be expressed as a composition of translations, rotations around the origin and conjugations. Thus, it suffices to show that applying a homothety from the origin multiplies the area by k^2, and that area is preserved under translations, rotations around the origin and conjugations. We will prove each of these facts using the shoelace formula. Starting from any vertex of P and going counterclockwise, let the coordinates of the vertices be $(x_0, y_0), (x_1, y_1), \ldots, (x_{n-1}, y_{n-1})$.

1. If Q is the image of P under a homothety of factor k from the origin, then $\operatorname{sgn}(Q) = \operatorname{sgn}(P) = 1$. We compute

$$[Q] = \frac{1}{2} \cdot \sum_{i=0}^{n-1} \det \begin{pmatrix} kx_i & kx_{i+1} \\ ky_i & ky_{i+1} \end{pmatrix} = \frac{k^2}{2} \cdot \sum_{i=0}^{n-1} \det \begin{pmatrix} x_i & x_{i+1} \\ y_i & y_{i+1} \end{pmatrix} = k^2 \cdot [P].$$

2. If Q is the image of P under a translation by $z = x + iy$, then $\operatorname{sgn}(Q) = \operatorname{sgn}(P) = 1$. We compute

$$\begin{aligned} [Q] &= \frac{1}{2} \cdot \sum_{i=0}^{n-1} \det \begin{pmatrix} x_i + x & x_{i+1} + x \\ y_i + y & y_{i+1} + y \end{pmatrix} \\ &= \frac{1}{2} \cdot \sum_{i=0}^{n-1} ((x_i y_{i+1} - x_{i+1} y_i) + x(y_{i+1} - y_i) + y(x_i - x_{i+1})) \\ &= \frac{1}{2} \cdot \sum_{i=0}^{n-1} \det \begin{pmatrix} x_i & x_{i+1} \\ y_i & y_{i+1} \end{pmatrix} + \frac{x}{2} \cdot \sum_{i=0}^{n-1} (y_{i+1} - y_i) + \frac{y}{2} \cdot \sum_{i=0}^{n-1} (x_i - x_{i+1}) \\ &= \frac{1}{2} \cdot \sum_{i=0}^{n-1} \det \begin{pmatrix} x_i & x_{i+1} \\ y_i & y_{i+1} \end{pmatrix} + 0 + 0 = [P], \end{aligned}$$

where indices are reduced modulo n. The second and third sums disappeared in the end due to telescoping.

3. If Q is the image of P under a counterclockwise rotation by θ radians around the origin, then $\operatorname{sgn}(Q) = \operatorname{sgn}(P) = 1.$ Since

$$\begin{aligned}(x+iy)e^{i\theta} &= (x+iy)(\cos\theta + i\sin\theta)\\ &= (x\cos\theta - y\sin\theta) + i(x\sin\theta + y\cos\theta);\end{aligned}$$

we want to compute

$$[Q] = \frac{1}{2}\sum_{i=0}^{n-1}\det\begin{pmatrix} x_i\cos\theta - y_i\sin\theta & x_{i+1}\cos\theta - y_{i+1}\sin\theta \\ x_i\sin\theta + y_i\cos\theta & x_{i+1}\sin\theta + y_{i+1}\cos\theta\end{pmatrix}.$$

We could expand this via the 2×2 determinant formula and use the Pythagorean identity to simplify it, but there is a more sophisticated method available. For those familiar with matrix multiplication and the multiplicative property of the determinant, each term in the sum is actually

$$\begin{aligned}\det\left(\begin{pmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{pmatrix}\begin{pmatrix}x_i & x_{i+1}\\ y_i & y_{i+1}\end{pmatrix}\right) &= \det\begin{pmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{pmatrix}\det\begin{pmatrix}x_i & x_{i+1}\\ y_i & y_{i+1}\end{pmatrix}\\ &= (\cos^2\theta + \sin^2\theta)\det\begin{pmatrix}x_i & x_{i+1}\\ y_i & y_{i+1}\end{pmatrix}\\ &= \det\begin{pmatrix}x_i & x_{i+1}\\ y_i & y_{i+1}\end{pmatrix},\end{aligned}$$

which means $[Q] = [P]$ by the shoelace formula. Readers who are not familiar with matrix multiplication should expand each term in the sum manually and watch the terms cancel.

4. If Q is the image of P under a conjugation, $\operatorname{sgn}(Q) = -\operatorname{sgn}(P) = -1.$ We compute

$$[Q] = -\frac{1}{2}\sum_{i=0}^{n-1}\det\begin{pmatrix}x_i & x_{i+1}\\ -y_i & -y_{i+1}\end{pmatrix} = \frac{1}{2}\sum_{i=0}^{n-1}\det\begin{pmatrix}x_i & x_{i+1}\\ y_i & y_{i+1}\end{pmatrix} = [P].$$

Thus, the proposition is proven. □

We will now see variants of the shoelace formula for triangles, using complex coordinates and barycentric coordinates.

Corollary 9.25 (Complex shoelace for triangle). If the vertices of $\triangle ABC$ are placed on the complex coordinates $A = a, B = c, B = c$, then the signed area of $\triangle ABC$ is

$$\operatorname{sgn}(ABC)\cdot[ABC] = \frac{i}{4}\det\begin{pmatrix}a & \overline{a} & 1\\ b & \overline{b} & 1\\ c & \overline{c} & 1\end{pmatrix}.$$

Proof. Using Theorem 1.28 and setting $A = (x_0, y_0), B = (x_1, y_1), C = (x_2, y_2)$, we compute that twice the signed area of $\triangle ABC$ is

$$
\begin{aligned}
2 \cdot \operatorname{sgn}(ABC) \cdot [ABC] &= \det \begin{pmatrix} 1 & 1 & 1 \\ x_0 & x_1 & x_2 \\ y_0 & y_1 & y_2 \end{pmatrix} = \det \begin{pmatrix} 1 & x_0 & y_0 \\ 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \end{pmatrix} = \det \begin{pmatrix} x_0 & y_0 & 1 \\ x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \end{pmatrix} \\
&= \det \begin{pmatrix} x_0 + iy_0 & y_0 & 1 \\ x_1 + iy_1 & y_1 & 1 \\ x_2 + iy_2 & y_2 & 1 \end{pmatrix} = \frac{1}{2i} \det \begin{pmatrix} x_0 + iy_0 & 2iy_0 & 1 \\ x_1 + iy_1 & 2iy_1 & 1 \\ x_2 + iy_2 & 2iy_2 & 1 \end{pmatrix} \\
&= -\frac{1}{2i} \det \begin{pmatrix} x_0 + iy_0 & x_0 - iy_0 & 1 \\ x_1 + iy_1 & x_1 - iy_1 & 1 \\ x_2 + iy_2 & x_2 - iy_2 & 1 \end{pmatrix} = \frac{i}{2} \det \begin{pmatrix} a & \overline{a} & 1 \\ b & \overline{b} & 1 \\ c & \overline{c} & 1 \end{pmatrix},
\end{aligned}
$$

where we used $i = -\frac{1}{i}$ at the end. □

Next, we will develop the underpinnings of barycentric coordinates. This is a powerful technique that can be used to solve many geometry problems in mathematical olympiads.

Lemma 9.26. Let A, B, C be distinct points in $\mathbb{R}^n$. The three points are non-collinear if and only if $\overrightarrow{AC}$ and $\overrightarrow{BC}$ are linearly independent. Equivalently, A, B, C are collinear if and only if $\overrightarrow{AB}$ and $\overrightarrow{BC}$ are linearly dependent,

Proof. We will prove the contrapositive of each direction, that A, B, C are collinear if and only if $\overrightarrow{AC}$ and $\overrightarrow{BC}$ are linearly dependent. If A, B, C are collinear, then a line

$$\ell = \{p + tv : t \in \mathbb{R}\}$$

runs through all three points. So there exist distinct real constants t_1, t_2, t_3 such that

$$
\begin{aligned}
p + t_1 v &= A, \\
p + t_2 v &= B, \\
p + t_3 v &= C.
\end{aligned}
$$

Then $(t_3 - t_1)v = C - A$ and $(t_3 - t_2)v = C - B$, so the position vector $\overrightarrow{v}$ satisfies

$$\overrightarrow{v} = \frac{1}{t_3 - t_1}\overrightarrow{AC} = \frac{1}{t_3 - t_2}\overrightarrow{BC},$$

making $\overrightarrow{AC}$ and $\overrightarrow{BC}$ linearly dependent.

Conversely, suppose $\overrightarrow{AC}$ and $\overrightarrow{BC}$ are linearly dependent. Then there exist real constants λ and μ such that at least one of them is non-zero and

$$\lambda \overrightarrow{AC} + \mu \overrightarrow{BC} = 0.$$

In fact, neither of λ, μ can be zero because one of them being zero would cause B, C to coincide or A, C to coincide, contradicting that the three points are distinct. Let $\overrightarrow{v}$ be the position vector that is common to both sides of

$$\lambda(\overrightarrow{C} - \overrightarrow{A}) = -\mu(\overrightarrow{C} - \overrightarrow{B}).$$

Then

$$\begin{aligned}\vec{A} &= \vec{C} - \frac{1}{\ }\vec{v};\\ \vec{B} &= \vec{C} + \frac{1}{\ }\vec{v};\\ \vec{C} &= \vec{C} + 0\vec{v}.\end{aligned}$$

By taking the arrowheads of all of these position vectors, we find that A, B, C all lie on the same line $\{C + tv : t \in \mathbb{R}\}$, making these three points collinear. □

Lemma 9.27. Let $\triangle ABC$ be a non-degenerate triangle and P be any point in the plane. Then there exists a unique triple of real numbers (x, y, z) such that

$$\vec{P} = x\vec{A} + y\vec{B} + z\vec{C}$$

and $x + y + z = 1$. These entries (x, y, z) are called the **barycentric coordinates** of point P with respect to the **reference triangle** $\triangle ABC$.

Proof. Using $z = 1 - x - y$, we work backwards to get

$$\begin{aligned}\vec{P} &= x\vec{A} + y\vec{B} + z\vec{C}\\ &= x\vec{A} + y\vec{B} + (1 - x - y)\vec{C}\\ &= x(\vec{A} - \vec{C}) + y(\vec{B} - \vec{C}) + \vec{C}\\ \vec{P} - \vec{C} &= x(\vec{A} - \vec{C}) + y(\vec{B} - \vec{C})\\ \vec{CP} &= x\vec{CA} + y\vec{CB}.\end{aligned}$$

So we need to show that real numbers x and y exist such that they satisfy the equation in the last line. Since $\triangle ABC$ is non-degenerate, it means A, B, C are non-collinear. By Lemma 9.26, $\vec{CA}$ and $\vec{CB}$ are linearly independent, allowing for a unique pair of coefficients x and y to exist. Letting $z = 1 - x - y$, we are done. □

Theorem 9.28 (Barycentric shoelace for triangles). Let $\triangle ABC$ be the reference triangle (necessarily non-degenerate) in a barycentric system. Let $\triangle PQR$ be an arbitrary triangle (possibly degenerate) whose vertices have barycentric coordinates given by

$$\begin{aligned}P &= (x_1, y_1, z_1),\\ Q &= (x_2, y_2, z_2),\\ R &= (x_3, y_3, z_3).\end{aligned}$$

Then,

$$\frac{\operatorname{sgn}(PQR) \cdot [PQR]}{\operatorname{sgn}(ABC) \cdot [ABC]} = \det\begin{pmatrix} x_1 & y_1 & z_1\\ x_2 & y_2 & z_2\\ x_3 & y_3 & z_3\end{pmatrix}$$

Proof. In Cartesian coordinates, let $A = (a_1; a_2); B = (b_1; b_2); C = (c_1; c_2)$. We are seeking the area of the triangle between the vectors $\overrightarrow{RP}$ and $\overrightarrow{RQ}$. These two can be decomposed as

$$\begin{aligned}
\overrightarrow{RP} &= \overrightarrow{CP} - \overrightarrow{CR} \\
&= (x_1\overrightarrow{CA} + y_1\overrightarrow{CB}) - (x_3\overrightarrow{CA} + y_3\overrightarrow{CB}) \\
&= (x_1 - x_3)\overrightarrow{CA} + (y_1 - y_3)\overrightarrow{CB} \\
&= (x_1 - x_3)(a_1 - c_1; a_2 - c_2) + (y_1 - y_3)(b_1 - c_1; b_2 - c_2) \\
&= ((x_1 - x_3)(a_1 - c_1); (x_1 - x_3)(a_2 - c_2)) + ((y_1 - y_3)(b_1 - c_1); (y_1 - y_3)(b_2 - c_2)) \\
&= ((x_1 - x_3)(a_1 - c_1) + (y_1 - y_3)(b_1 - c_1); (x_1 - x_3)(a_2 - c_2) + (y_1 - y_3)(b_2 - c_2))
\end{aligned}$$

and

$$\begin{aligned}
\overrightarrow{RQ} &= \overrightarrow{CQ} - \overrightarrow{CR} \\
&= (x_2\overrightarrow{CA} + y_2\overrightarrow{CB}) - (x_3\overrightarrow{CA} + y_3\overrightarrow{CB}) \\
&= (x_2 - x_3)\overrightarrow{CA} + (y_2 - y_3)\overrightarrow{CB} \\
&= (x_2 - x_3)(a_1 - c_1; a_2 - c_2) + (y_2 - y_3)(b_1 - c_1; b_2 - c_2) \\
&= ((x_2 - x_3)(a_1 - c_1); (x_2 - x_3)(a_2 - c_2)) + ((y_2 - y_3)(b_1 - c_1); (y_2 - y_3)(b_2 - c_2)) \\
&= ((x_2 - x_3)(a_1 - c_1) + (y_2 - y_3)(b_1 - c_1); (x_2 - x_3)(a_2 - c_2) + (y_2 - y_3)(b_2 - c_2)).
\end{aligned}$$

By the first line of the shoelace formula for triangles (Theorem 9.22), since $\triangle RPQ$ has the same orientation and area as $\triangle PQR$, we find that

$$\begin{aligned}
&2 \cdot \operatorname{sgn}(RPQ) \cdot [RPQ] \\
&= 2 \cdot \operatorname{sgn}(PQR) \cdot [PQR] \\
&= \det \begin{pmatrix} (x_1 - x_3)(a_1 - c_1) + (y_1 - y_3)(b_1 - c_1) & (x_2 - x_3)(a_1 - c_1) + (y_2 - y_3)(b_1 - c_1) \\ (x_1 - x_3)(a_2 - c_2) + (y_1 - y_3)(b_2 - c_2) & (x_2 - x_3)(a_2 - c_2) + (y_2 - y_3)(b_2 - c_2) \end{pmatrix}.
\end{aligned}$$

By matrix multiplication and the multiplicative property of determinants, this is equivalent to

$$\begin{aligned}
&\det\left(\begin{pmatrix} a_1 - c_1 & b_1 - c_1 \\ a_2 - c_2 & b_2 - c_2 \end{pmatrix}\begin{pmatrix} x_1 - x_3 & x_2 - x_3 \\ y_1 - y_3 & y_2 - y_3 \end{pmatrix}\right) \\
&= \det\begin{pmatrix} a_1 - c_1 & b_1 - c_1 \\ a_2 - c_2 & b_2 - c_2 \end{pmatrix} \cdot \det\begin{pmatrix} x_1 - x_3 & x_2 - x_3 \\ y_1 - y_3 & y_2 - y_3 \end{pmatrix} \\
&= 2 \cdot \operatorname{sgn}(ABC) \cdot [ABC] \cdot \det\begin{bmatrix} 1 & 1 & 1 \\ x_3 & x_1 & x_2 \\ y_3 & y_1 & y_2 \end{bmatrix}.
\end{aligned}$$

As a result, we obtain that

$$\begin{aligned}
\frac{2\cdot\operatorname{sgn}(RPQ)\cdot[RPQ]}{2\cdot\operatorname{sgn}(ABC)\cdot[ABC]} &= \det\begin{pmatrix}1 & 1 & 1\\ x_3 & x_1 & x_2\\ y_3 & y_1 & y_2\end{pmatrix}\\
&= \det\begin{pmatrix}x_3+y_3+z_3 & x_1+y_1+z_1 & x_2+y_2+z_2\\ x_3 & x_1 & x_2\\ y_3 & y_1 & y_2\end{pmatrix}\\
&= \det\begin{pmatrix}z_3 & z_1 & z_2\\ x_3 & x_1 & x_2\\ y_3 & y_1 & y_2\end{pmatrix} = \det\begin{pmatrix}x_3 & x_1 & x_2\\ y_3 & y_1 & y_2\\ z_3 & z_1 & z_2\end{pmatrix}\\
&= \det\begin{pmatrix}x_1 & x_2 & x_3\\ y_1 & y_2 & y_3\\ z_1 & z_2 & z_3\end{pmatrix}.
\end{aligned}$$

Here, we used several of the preservation properties of determinants, which were listed in Theorem 1.28. □

Corollary 9.29. The barycentric coordinates of a point P with respect to the reference triangle $\triangle ABC$ are

$$(x, y, z) = \left(\frac{\operatorname{sgn}(PBC)\cdot[PBC]}{\operatorname{sgn}(ABC)\cdot[ABC]}, \frac{\operatorname{sgn}(PCA)\cdot[PCA]}{\operatorname{sgn}(ABC)\cdot[ABC]}, \frac{\operatorname{sgn}(PAB)\cdot[PAB]}{\operatorname{sgn}(ABC)\cdot[ABC]}\right).$$

Proof. First we observe that the barycentric coordinates of A, B, C themselves are given by the coefficients in

$$\begin{aligned}
\vec{A} &= 1\cdot\vec{A} + 0\cdot\vec{B} + 0\cdot\vec{C}\\
\vec{B} &= 0\cdot\vec{A} + 1\cdot\vec{B} + 0\cdot\vec{C}\\
\vec{C} &= 0\cdot\vec{A} + 0\cdot\vec{B} + 1\cdot\vec{C}.
\end{aligned}$$

By Theorem 9.28,

$$\begin{aligned}
\frac{\operatorname{sgn}(PBC)\cdot[PBC]}{\operatorname{sgn}(ABC)\cdot[ABC]} &= \det\begin{pmatrix}x & y & z\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix} = x,\\
\frac{\operatorname{sgn}(PBC)\cdot[PBC]}{\operatorname{sgn}(ABC)\cdot[ABC]} &= \det\begin{pmatrix}x & y & z\\ 0 & 0 & 1\\ 1 & 0 & 0\end{pmatrix} = y,\\
\frac{\operatorname{sgn}(PBC)\cdot[PBC]}{\operatorname{sgn}(ABC)\cdot[ABC]} &= \det\begin{pmatrix}x & y & z\\ 1 & 0 & 0\\ 0 & 1 & 0\end{pmatrix} = z,
\end{aligned}$$

which completes the proof. □

Chapter 10

Cevians

> "I must ask you to forgive me if I concentrate on my own favourite branches, and I must take the risk of offending various geometers who will ask why I have not dealt with algebraic geometry, differential geometry, symplectic geometry, continuous geometry, metric spaces, Banach spaces, linear programming, and so on."
>
> – *H. S. M. Coxeter*

For computing lengths of cevians, Stewart's theorem is indispensable. In the study of concurrent cevians, the key results are Ceva's theorem, its converse, and its trigonometric variant. We will study all of these results plus applications to interesting problems.

10.1 Computing Lengths

Definition 10.1. Recall that cevians were defined in Definition 5.18. The three most common cevians of triangles are as follows.

- A median is a cevian whose foot is the midpoint of the edge on which it lies, so the median bisects its edge.
- An angle bisector is a cevian that cuts in half the interior angle inside in which it lies, so it bisects the interior angle.
- An altitude is a generalized cevian that makes a right angle with the line through the edge on which its foot lies. An altitude is also called a height, though the latter term is often reserved for the length of the altitude.

Problem 10.2. In a right triangle, show that dropping the altitude from the right angle's vertex creates two triangles that are similar to the original triangle.

Theorem 10.3. Suppose $\triangle ABC$ and $\triangle A'B'C'$ are triangles such that $\triangle ABC \sim \triangle A'B'C'.$ Let D, E, F be the feet of the median, angle bisector, height respectively of $\triangle ABC$ from A to the line through $BC,$ and define D', E', F' similarly for $\triangle A'B'C'.$ If the similarity ratio of the side lengths of $\triangle ABC$ to the corresponding side lengths of $\triangle A'B'C'$ is $k,$ then

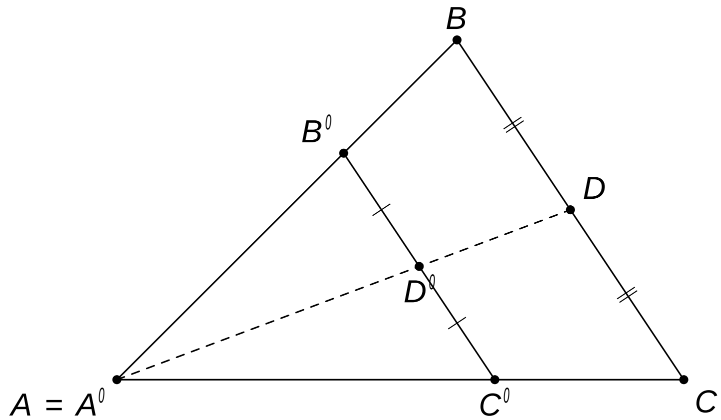
$$\frac{AD}{A'D'} = \frac{AE}{A'E'} = \frac{AF}{A'F'} = k.$$

Proof. It will be helpful to conceptualize the following proofs by nesting one of the triangles inside the other with $\angle A = \angle A'$ being the shared angle.

1. Medians: Note that $\dfrac{BD}{B'D'} = \dfrac{BC/2}{B'C'/2} = \dfrac{BC}{B'C'} = k.$ Since $\dfrac{BA}{B'A'} = k$ and $\angle B = \angle B',$ we find using SAS similarity that

$$\triangle BAD \sim \triangle B'A'D'$$

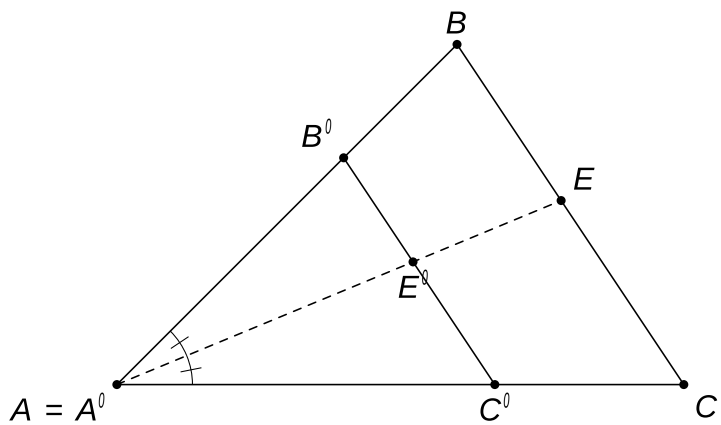
with similarity ratio $k,$ which allows us to conclude that $\dfrac{AD}{A'D'} = k.$



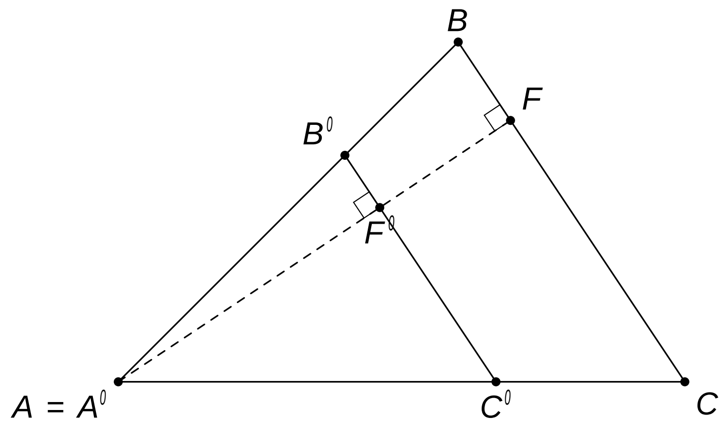
2. Angle bisectors: Note that $\angle B = \angle B'$ and that

$$\angle BAE = \frac{\angle BAC}{2} = \frac{\angle B'A'C'}{2} = \angle B'A'E'.$$

By AA similarity, $\triangle BAE \sim \triangle B'A'E'.$ Since $\dfrac{BA}{B'A'} = k,$ this is also the similarity ratio of $\triangle BAE$ to $\triangle B'A'E',$ which in turn gives us $\dfrac{AE}{A'E'} = k.$



3. Altitudes: Since $\angle B = \angle B'$ and $\angle BFA = \angle B'F'A' = 90^\circ,$ we get that $\triangle BAF \sim \triangle B'A'F'$ by AA similarity. Once again, since $\dfrac{BA}{B'A'} = k,$ this is also the similarity ratio of $\triangle BAF$ to $\triangle B'A'F'.$ Thus, $\dfrac{AF}{A'F'} = k$ as well.

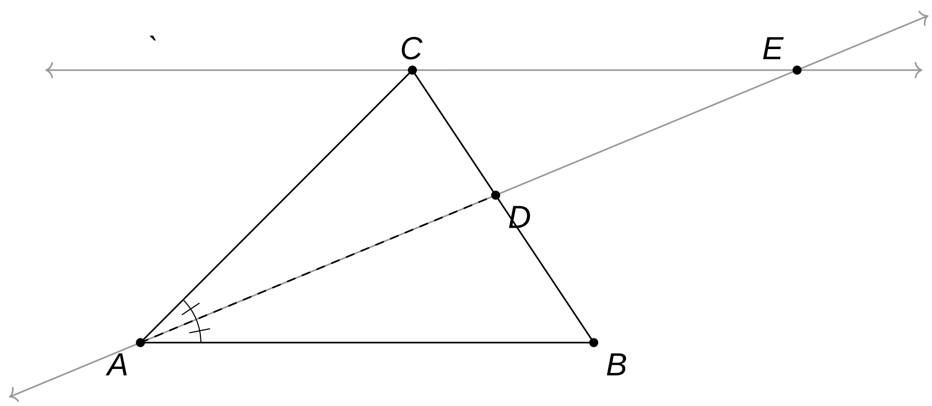


∎

Theorem 10.4 (Angle bisector theorem). Let $\triangle ABC$ have an angle bisector AD where D lies on BC. Then

$$AB \cdot CD = AC \cdot BD.$$

Proof. First we draw the line ℓ through C that is parallel to the line through AB. Then we extend AD as a ray through D so that it meets ℓ at E.



Since AB and CE are parallel, we get the equal alternate interior angles $\angle BAE = \angle AEC$. By alternate interior angles,

$$\angle CEA = \angle BAE = \angle CAE,$$

so $\triangle ACE$ is isosceles with $AC = EC$. Moreover, we can observe by the same alternate interior angles and $\angle ABC = \angle BCE$ that $\triangle BAD \sim \triangle CED$. By similarity ratios and the isosceles triangle,

$$\frac{BD}{CD} = \frac{AB}{EC} = \frac{AB}{AC},$$

which implies $AB \cdot CD = AC \cdot BD$ as desired. ∎

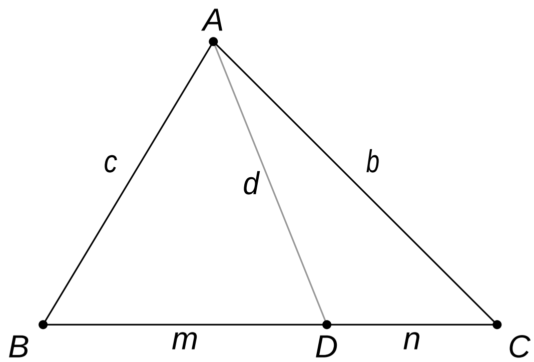
Theorem 10.5 (Stewart's theorem). Let AD be a cevian of $\triangle ABC$. Let $BC = a, CA = b, AB = c, BD = m, CD = n, AD = d$. Then all the lengths are related by

$$(d^2 + mn)a = b^2m + c^2n.$$

A common mnemonic for remembering this formula is to write it as

$$dad + man = bmb + cnc;$$

though it would also be necessary to memorize the configuration with its labels.



Proof. First we label the non-reflex angles \ ADB = and \ ADC = : By the cosine law,

$$\cos \quad = \frac{d^2 + m^2 \quad c^2}{2dm};$$
$$\cos \quad = \frac{d^2 + n^2 \quad b^2}{2dn}:$$

Since + = 180 ;

$$\cos \quad = \cos(180 \qquad) = \quad \cos \quad :$$

Then our two expressions from the cosine law are negations of each other, meaning

$$\frac{d^2 + m^2 \quad c^2}{2dm} = \quad \frac{d^2 + n^2 \quad b^2}{2dn}:$$

Clearing the denominators and rearranging the terms yields

$$d^2(m + n) + mn(m + n) = b^2m + c^2n:$$

Since $m + n = a;$ we are done. □

Problem 10.6 (Apollonius's theorem). For 4ABC, find an expression in terms of $a = BC; b = CA; c = AB$ for the median emanating from A.

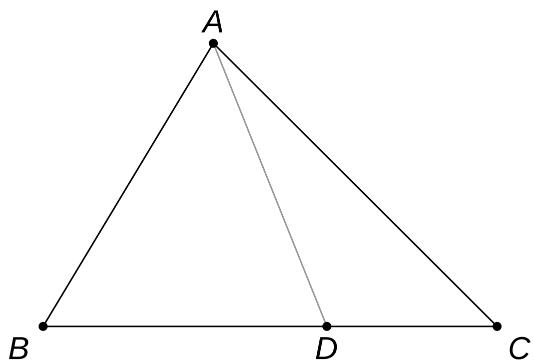
Problem 10.7. Given 4ABC, let the medians with feet on $BC = a; CA = b; AB = c$ be $m_a; m_b; m_c$, respectively. Suppose $a \quad b \quad c$. Prove that $a; b; c$ can each be scaled by the same factor to produce $m_c; m_b; m_a$ respectively if and only if $a^2; b^2; c^2$ form an arithmetic sequence in that order. In this case, the triangle with side lengths $m_a; m_b; m_c$ is known as an automedian triangle.

Problem 10.8. Given 4ABC, find an expression in terms of $a = BC; b = CA; c = AB$ for the cevian that is the angle bisector emanating from A.

Formulas for the lengths of specific cevians are not worth memorizing as they can be derived from Stewart's theorem when necessary.

Theorem 10.9 (Ratio lemma). Let D be a point on BC in $\triangle ABC$. Then

$$\frac{BD}{CD} = \frac{BA}{CA} \cdot \frac{\sin BAD}{\sin CAD}.$$



Proof. By the applying the law of sines (Theorem 6.7) to $\triangle ABD$ and $\triangle ACD$,

$$\frac{AD}{\sin ABD} = \frac{BD}{\sin BAD} = \frac{AB}{\sin ADB},$$
$$\frac{AD}{\sin ACD} = \frac{CD}{\sin CAD} = \frac{AC}{\sin ADC}.$$

Then

$$BD = \frac{AD \sin BAD}{\sin ABD},$$
$$CD = \frac{AD \sin CAD}{\sin ACD},$$

and dividing the first by the second yields

$$\frac{BD}{CD} = \frac{\sin BAD \sin ACD}{\sin CAD \sin ABD}.$$

Now note that

$$\sin ADB = \sin(180 - ADC) = \sin ADC$$

by a reflection trigonometric identity. Then, the equations from the law of sines also give equations

$$\sin ADB = \frac{AB \sin ABD}{AD},$$
$$\sin ADC = \frac{AC \sin ACD}{AD}$$

that can be set equal to each other to get

$$\frac{AB \sin ABD}{AD} = \frac{AC \sin ACD}{AD} \implies \frac{\sin ACD}{\sin ABD} = \frac{AB}{AC}.$$

Therefore,

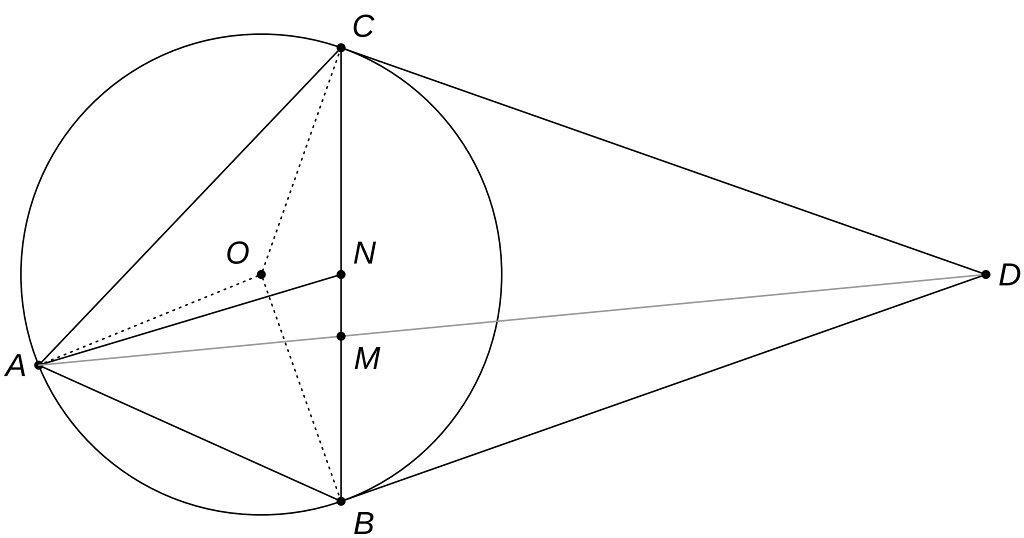
$$\frac{BD}{CD} = \frac{\sin BAD \sin ACD}{\sin CAD \sin ABD} = \frac{\sin BAD}{\sin CAD} \cdot \frac{AB}{AC},$$

as desired. Do you see how this contains the angle bisector theorem (Theorem 10.4) as a special case? □

Definition 10.10. A symmedian through a vertex V of a triangle is a cevian that is produced by reflecting the line through the median emanating from V over the line through the angle bisector emanating from V.

Theorem 10.11 (Symmedian lemma)**.** Given $\triangle ABC$, let the tangents to its circumcircle at B and C intersect at D. Prove that AD runs through the symmedian emanating from A.

Proof. Let the intersection of AD with BC be M. Let the reflection of AD across the angle bisector emanating from A intersect with BC at N. We wish to prove that N is the midpoint of BC, so it suffices to prove that $BN = NC$. Let O be the circumcenter so that $OC \perp CD$ and $OB \perp BD$.



By the sine law,

$$BN = AN \cdot \frac{\sin BAN}{\sin ABN},$$

$$NC = AN \cdot \frac{\sin CAN}{\sin ACN}.$$

Taking the quotient of the top equation divided by the bottom one, we get

$$\frac{BN}{NC} = \frac{\sin BAN}{\sin ABN} \cdot \frac{\sin ACN}{\sin CAN}.$$

Now we work to find interrelations between the angles that characterize this configuration. Using the three isosceles triangles centred at O, let

$$x = \angle OBC = \angle OCB,$$

$$y = \angle OCA = \angle OAC,$$

$$z = \angle OAB = \angle OBA.$$

Using the shorthand notation A, B, C for the interior angles of $\triangle ABC$ at these respective vertices, we get the system

$$\begin{aligned} A &= \angle BAC = y + z, \\ B &= \angle CBA = z + x, \\ C &= \angle ACB = x + y, \end{aligned}$$

which we solve for x, y, z to get

$$\begin{aligned} 2x &= B + C - A, \\ 2y &= C + A - B, \\ 2z &= A + B - C. \end{aligned}$$

Some exploration and angle-chasing leads to

$$\begin{aligned} \angle ABD &= z + 90^\circ \\ &= \frac{A + B - C}{2} + 90^\circ \\ &= \frac{A + B - C}{2} + \frac{A + B + C}{2} \\ &= A + B = 180^\circ - C \\ \angle ACD &= y + 90^\circ \\ &= \frac{C + A - B}{2} + 90^\circ \\ &= \frac{C + A - B}{2} + \frac{A + B + C}{2} \\ &= C + A = 180^\circ - B. \end{aligned}$$

Continuing from earlier, we use the fact that the line through AN is a reflection of the line through AM across the angle bisector of A to get

$$\begin{aligned} \frac{BN}{NC} &= \frac{\sin BAN}{\sin ABN} \cdot \frac{\sin ACN}{\sin CAN} \\ &= \frac{\sin BAN}{\sin B} \cdot \frac{\sin C}{\sin CAN} \\ &= \frac{\sin CAM}{\sin(180^\circ - ACD)} \cdot \frac{\sin(180^\circ - ABD)}{\sin BAM} \\ &= \frac{\sin CAM}{\sin ACD} \cdot \frac{\sin ABD}{\sin BAM} \\ &= \frac{CD}{AD} \cdot \frac{AD}{BD} = \frac{CD}{BD} = 1, \end{aligned}$$

since the tangent lengths $BD = CD$ from D are equal in length. We also used the sine law on $\triangle ACD$ and $\triangle ABD$ near the end. □

10.2 Concurrency

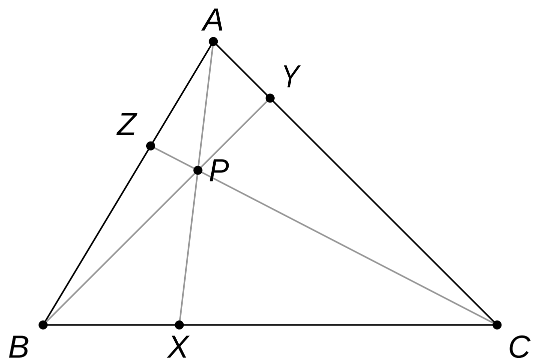
In what follows, we will repeatedly use the algebraic ratio trick that

$$\frac{a}{b} = \frac{c}{d} = k \implies \frac{a+c}{b+d} = \frac{a-c}{b-d} = k,$$

assuming these denominators are not 0. This trick was proven in Volume 1.

Theorem 10.12 (Ceva's theorem). In $\triangle ABC,$ the three cevians AX, BY, CZ concur at a point P if and only if

$$\frac{BX}{XC} \cdot \frac{CY}{YA} \cdot \frac{AZ}{ZB} = 1.$$



Proof. Suppose the three cevians concur. Recall that if two triangles each have a base running through the same line and the corresponding two heights are the same line segment, then the ratio of these bases is the ratio of the areas of the corresponding triangles. Then

$$\frac{BX}{XC} = \frac{[ABX]}{[ACX]} = \frac{[PBX]}{[PCX]} = \frac{[ABX]-[PBX]}{[ACX]-[PCX]} = \frac{[ABP]}{[ACP]}.$$

By similar derivations,

$$\frac{CY}{YA} = \frac{[BCP]}{[BAP]},$$
$$\frac{AZ}{ZB} = \frac{[CAP]}{[CBP]}.$$

Multiplying the three equations yields the telescoping product

$$\frac{BX}{XC} \cdot \frac{CY}{YA} \cdot \frac{AZ}{ZB} = \frac{[ABP]}{[ACP]} \cdot \frac{[BCP]}{[BAP]} \cdot \frac{[CAP]}{[CBP]} = 1.$$

Conversely, suppose $\frac{BX}{XC} \cdot \frac{CY}{YA} \cdot \frac{AZ}{ZB} = 1.$ Let P be the interior point of the triangle at which BY and CZ intersect. Let X' be the foot of the cevian emanating from A that goes through $P.$ By Ceva's theorem,

$$\frac{BX'}{X'C} \cdot \frac{CY}{YA} \cdot \frac{AZ}{ZB} = 1.$$

Equating this with the equation in the hypothesis, we get

$$\frac{BX}{XC} = \frac{BX'}{X'C}.$$

Using

$$BC = BX + XC = BX' + X'C,$$

we get

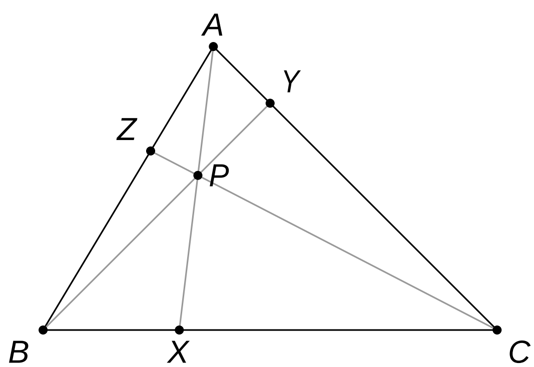
$$\frac{BC - XC}{XC} = \frac{BC - X'C}{X'C}.$$

So $XC = X'C$ and $BX = BX'$, and therefore $X = X'$.
Here, we defined a point that conveniently fulfils a desired property and then showed that a given point is the same point. This technique is called using "phantom points" and is frequently used to prove the converse of a theorem. □

There is a more general version of Ceva's theorem that holds for generalized cevians, but we will not foray into it because it involves the ratios of directed lengths. For the same reason, we will avoid the related collinearity result called Menelaus's theorem and its converse.

Theorem 10.13 (Van Aubel's theorem). If the cevians AX, BY, CZ of $\triangle ABC$ concur at P, then

$$\frac{AP}{PX} = \frac{AY}{YC} + \frac{AZ}{ZB}.$$



Proof. Similar to our proof of Ceva's theorem, we find that

$$\frac{AP}{PX} = \frac{[ABP]}{[PBX]} = \frac{[ACP]}{[PCX]} = \frac{[ABP] + [ACP]}{[PBX] + [PCX]} = \frac{[ABP] + [ACP]}{[PBC]}.$$

In the proof of Ceva's theorem, we found that

$$\frac{AY}{YC} = \frac{[ABP]}{[CBP]},$$
$$\frac{AZ}{ZB} = \frac{[ACP]}{[BCP]}.$$

Then van Aubel's theorem follows immediately. □

Problem 10.14 (Gergonne's theorem). Prove that if $\triangle ABC$ has cevians AX, BY, CZ that concur at $P,$ then

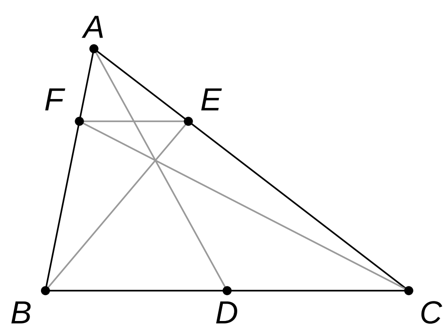
$$\frac{PX}{AX} + \frac{PY}{BY} + \frac{PZ}{CZ} = 1.$$

Subtracting each side from 3 yields the equivalent identity

$$\frac{AP}{AX} + \frac{BP}{BY} + \frac{CP}{CZ} = 2.$$

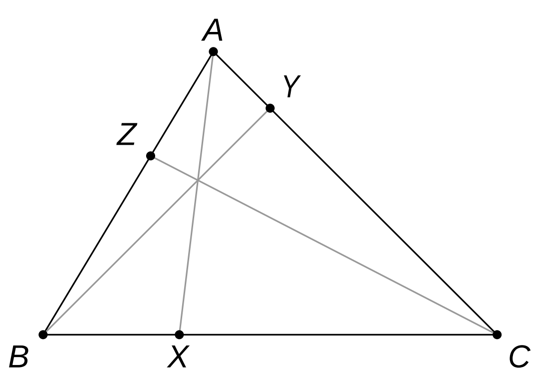
The preceding two results attributed, to Van Aubel and Gergonne, are not well-known, but they appear with proofs in [2].

Problem 10.15. Let D be the midpoint of BC in $\triangle ABC$, and let E be a point on AC, and F be a point on AB. Prove that the cevians AD, BE, CF concur if and only if EF is parallel to BC.



Theorem 10.16 (Trigonometric Ceva's theorem). The cevians AX, BY, CZ of $\triangle ABC$ are concurrent if and only if

$$\frac{\sin BAX}{\sin XAC} \frac{\sin CBY}{\sin YBA} \frac{\sin ACZ}{\sin ZCB} = 1.$$



Proof. Using the sine law and the fact that the sines of supplementary angles are equal, we get

$$\frac{\sin BAX}{\sin XAC} = \frac{\frac{BX}{BA} \sin BXA}{\frac{XC}{CA} \sin CXA} = \frac{BX}{XC} \frac{CA}{AB}.$$

Similar derivations yield

$$\frac{\sin CBY}{\sin YBA} = \frac{CY}{YA} \cdot \frac{AB}{BC},$$
$$\frac{\sin ACZ}{\sin ZCB} = \frac{AZ}{ZB} \cdot \frac{BC}{CA}.$$

Now we get the partially telescoping product

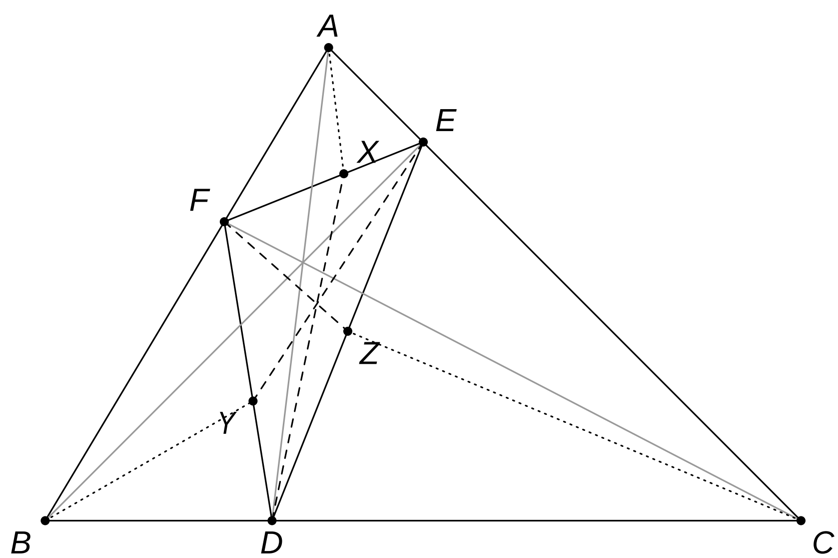
$$\begin{aligned}\frac{\sin BAX}{\sin XAC} \cdot \frac{\sin CBY}{\sin YBA} \cdot \frac{\sin ACZ}{\sin ZCB} &= \left(\frac{BX}{XC} \cdot \frac{CA}{AB}\right) \cdot \left(\frac{CY}{YA} \cdot \frac{AB}{CB}\right) \cdot \left(\frac{AZ}{ZB} \cdot \frac{BC}{AC}\right)\\ &= \left(\frac{BX}{XC} \cdot \frac{CY}{YA} \cdot \frac{AZ}{ZB}\right) \cdot \left(\frac{CA}{AB} \cdot \frac{AB}{CB} \cdot \frac{BC}{AC}\right)\\ &= \frac{BX}{XC} \cdot \frac{CY}{YA} \cdot \frac{AZ}{ZB}.\end{aligned}$$

Ceva's theorem says that AX, BY, CZ are concurrent if and only if

$$\frac{BX}{XC} \cdot \frac{CY}{YA} \cdot \frac{AZ}{ZB} = 1,$$

so we are done. □

Problem 10.17 (Cevian nest). In $\triangle ABC$, suppose the cevians AD, BE, CF concur. Suppose also that in $\triangle DEF$, cevians DX, EY, CZ concur. Prove that the lines through AX, BY, CZ concur. Hint: the ratio lemma, Ceva's theorem, and the trigonometric form of Ceva's theorem will be helpful.



Chapter 11

Triangle Centers

"There is no royal road to geometry." (in response to Ptolemy I Soter)

– *Euclid*

We will take a look at the most important triangle centers and their properties. Our list will be focused, in comparison to the thousands of triangle centers listed by Clark Kimberling. Specific triangle centers we will find and study include the centroid, incenter, excenters, orthocenter and circumcenter. Along the way, we will come across additional concepts, such as the Nagel point, Gergonne point, and the Euler line.

11.1 Examples

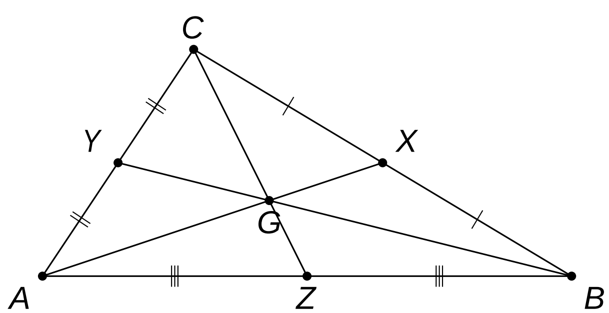
Theorem 11.1. Recall that median is a cevian whose foot lies on the midpoint of an edge. The three medians of any triangle are concurrent.

Proof. Let the medians be $AX; BY; CZ:$ By the definition of a median,

$$\frac{BX}{XC} \frac{CY}{YA} \frac{AZ}{ZB} = 1$$

because in each of the three fractions, the numerator is equal to the denominator. By the converse of Ceva's theorem, the medians are concurrent. □

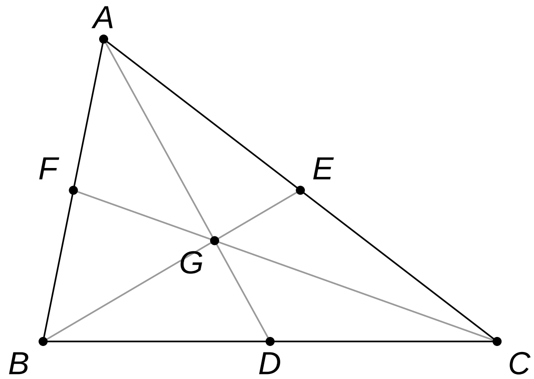
Definition 11.2. The point of concurrency of the medians of a triangle is called its **centroid**, which is often denoted by the letter G.



Problem 11.3. Show that, if the medians $AX; BY; CZ$ of $\triangle ABC$ concur at $P;$ then

$$\frac{AP}{PX} = \frac{BP}{PY} = \frac{CP}{PZ} = 2:$$

Use this to show that the six regions into which the medians of $\triangle ABC$ split the triangle are of equal area.



Definition 11.4. Recall that an interior angle bisector is the ray that splits an interior angle in half. An interior angle bisector can also refer to the part of such a ray that lies on the triangle. When we refer to an angle bisector, it will be clear from the context whether we are referring to the ray or the cevian.

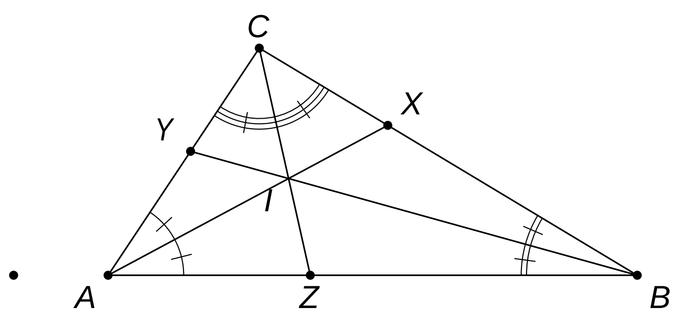
Theorem 11.5. In any triangle, the three cevians that are the interior angle bisectors are concurrent. This means that every triangle has a unique incenter, which we often denote by the letter I.

Proof. The angle bisector theorem tells us that

$$\frac{BX}{XC} \cdot \frac{CY}{YA} \cdot \frac{AZ}{ZB} = \frac{AB}{AC} \cdot \frac{BC}{BA} \cdot \frac{CA}{CB} = 1.$$

By the converse of Ceva's theorem, the angle bisectors are concurrent. Alternatively, the trigonometric form of Ceva's theorem proves the result immediately. The incenter is unique by Theorem 8.21. □

Definition 11.6. The incenter of a triangle is defined as the concurrency point of the three angle interior angle bisectors.



Definition 11.7. The perimeter of a triangle is the sum of its three side lengths. The semiperimeter of a triangle is half the perimeter; this might seem like a strange concept, but it comes up in Heron's formula in Theorem 11.27, for example.

The following is a useful substitution that allows us to turn computational problems about triangles, such as geometric inequalities, into problems about positive real numbers, or vice versa.

Theorem 11.8 (Ravi substitution). The positive real numbers $a; b; c$ are the three side lengths of some triangle if and only if there exist positive real numbers $x; y; z$ such that

$$\begin{aligned} a &= y + z; \\ b &= z + x; \\ c &= x + y: \end{aligned}$$

In the case that these $x; y; z$ exist, they are unique.

Proof. Let $a; b; c$ be positive numbers. Recall that the triangle inequality for triangles in the plane (Theorem 3.10) tells us that $a; b; c$ are the three side lengths of some triangle if and only if the following three inequalities hold:

$$\begin{aligned} b + c &> a; \\ c + a &> b; \\ a + b &> c: \end{aligned}$$

Suppose there exist positive real numbers $x; y; z$ that satisfy the three equations in the statement of the theorem. Then

$$\begin{aligned} b + c \quad a &= 2x > 0; \\ c + a \quad b &= 2y > 0; \\ a + b \quad c &= 2z > 0: \end{aligned}$$

In the other direction, suppose $a; b; c$ are the three side lengths of some triangle. Then the incenter exists. Suppose the incircle touches $BC; CA; AB$ at $X; Y; Z$ respectively. Thanks to the equality of both tangent lengths from the same point, we can let

$$\begin{aligned} x &= AY = AZ; \\ y &= BZ = BX; \\ z &= CX = CY: \end{aligned}$$

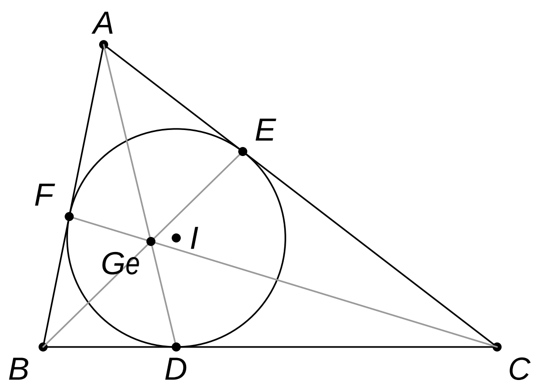
In fact, we already proved the existence of the positive $x; y; z$ in the same way when working with tangential polygons in Problem 8.22.

In the case that $x; y; z$ exist, they are unique because we can isolate them in terms of $a; b; c$ as

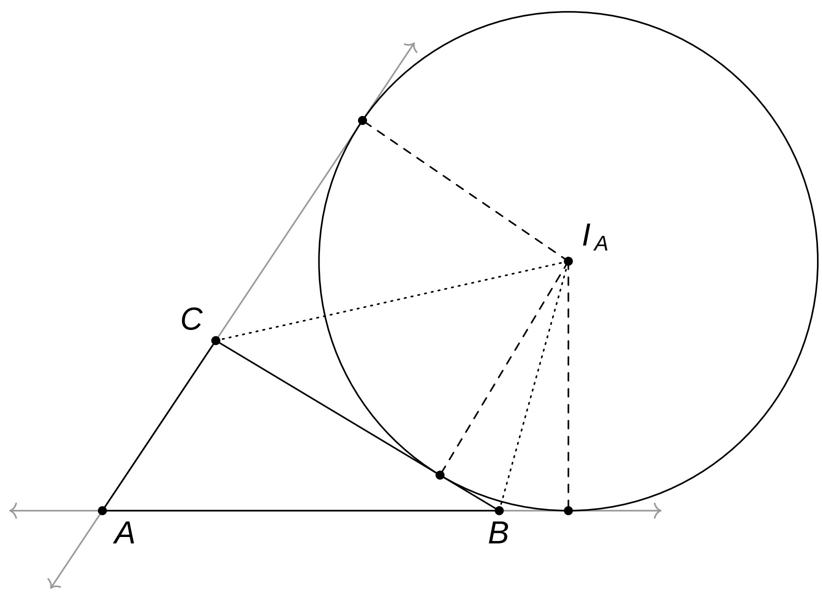
$$\begin{aligned} x &= s \quad a; \\ y &= s \quad b; \\ z &= s \quad c; \end{aligned}$$

where $s = \dfrac{a + b + c}{2}$ is the semiperimeter of the triangle. □

Problem 11.9. Suppose the incircle of $\triangle ABC$ touches BC, CA, AB at X, Y, Z respectively. Show that the cevians AX, BY, CZ concur. The point of concurrency is called the Gergonne point Ge of the triangle. Note: this is different from the incenter I, as shown in the diagram.



Lemma 11.10. There is a unique point I_A in the exterior of $\triangle ABC$ that is equidistant from the segment BC, the ray emanating from A through B, and the ray emanating from A through C. It is implicit in this equidistant condition that the foot of the perpendicular segment from I_A to the line through BC lies on BC, to the line through ray AB lies on ray AB, and to the line through ray AC lies on ray AC.



Proof. First we construct such a point I_A. Each interior angle has two equal exterior angles that are supplementary to it. Choosing the exterior angle of $\angle ABC$ that is on the same side of the line through AB as C, and choosing the exterior angle of $\angle ACB$ that is on the same side of the line through AC as B, let I_A be the intersection of the bisectors of these two exterior angles. Because I_A lies on angle bisectors, it is equidistant from the rays of each of these angles, which is what we wanted (the lines through two of these rays coincide and intersection of two of these two rays is BC), due to Theorem 8.18.

Now suppose I_A is any point that satisfies the definition in the statement of the lemma. We want to show that I_A is unique. Since the intersection of the aforementioned two angle bisectors is unique, it suffices to prove that I_A lies on both. Since I_A is equidistant from ray AB and $BC,$ it lies on the angle bisector of the exterior angle of $\angle ABC$ that is on the same side of the line through AB as $C.$ Similarly, since I_A is equidistant from ray AC and $BC,$ it lies on the angle bisector of the exterior angle of $\angle ABC$ that is on the same side of the line through AC as $B.$ □

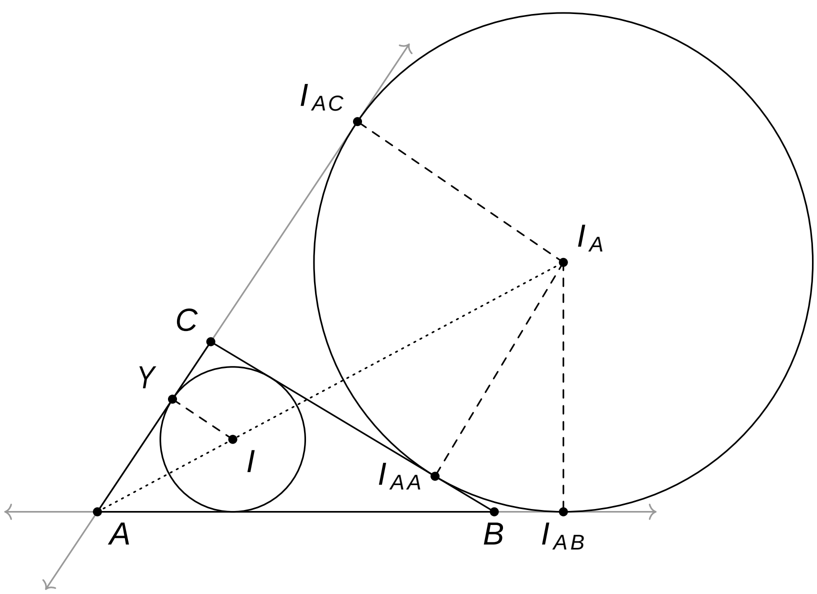
Definition 11.11. In the notation of Lemma 11.10, I_A an excenter of $\triangle ABC$. The circle with center I_A that is tangent to $BC,$ the ray AB and the ray AC is an excircle. Its radius is called an exradius. We can similarly define an excircle that is tangent to the segment CA or the segment $AB.$ To distinguish between the three excircles, we may refer to the one in our argument as the A-excircle with A-excenter I_A and A-exradius $r_A,$, with symmetric definitions for B and C.

Example 11.12. Given $\triangle ABC$, show that A, the incenter I, and the A-excenter I_A are collinear. Use this to show that the A-exradius is equal to

$$r_A = \frac{rs}{s-a},$$

where r is the inradius, s is the semiperimeter and $a = BC.$

Solution. Since I_A is in the interior of the interior angle $\angle BAC$ with AB and AC extended as rays past B and C respectively, and since I_A is equidistant from the rays AB and $AC,$ I_A lies on the ray that is the angle bisector of the interior angle $\angle BAC.$ As A and I also lie on this angle bisector, A, I, I_A are collinear.
Now let the foot of the perpendicular segment from I to AC be $Y.$ Let the A-excircle touch BC at $I_{AA},$ the ray AB at $I_{AB},$ and the ray AC at $I_{AC}.$



The critical observation is that, by tangency,

$$AB + BI_{AA} = AB + BI_{AB} = AI_{AB} = AI_{AC} = AC + CI_{AC} = AC + CI_{AA}.$$

At the same time,

$$AB + BI_{AA} + CI_{AA} + AC = AB + BC + CA = 2s,$$

so $AI_{AB} = AI_{AC} = s$. By nested angles, $\triangle AIY$ and $\triangle AI_AI_{AC}$ are similar right triangles. Then we get the similarity ratio

$$\frac{r_A}{s} = \frac{I_AI_{AC}}{AI_{AC}} = \frac{IY}{AY} = \frac{r}{s-a}.$$

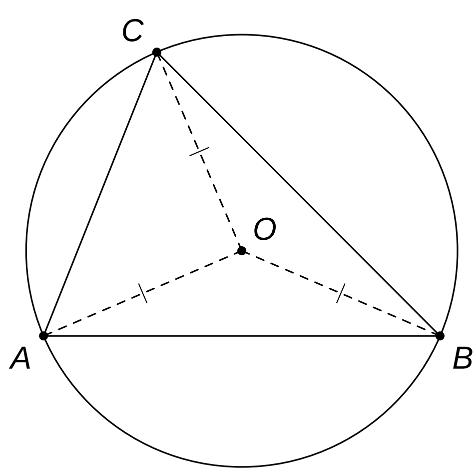
Multiplying both sides is s produces the desired formula. □

Problem 11.13. Given $\triangle ABC$, let the A-excircle touch BC at I_{AA}, the B-excircle touch CA at I_{BB}, and the C-excircle touch AB at I_{CC}. Show that the cevians AI_{AA}, BI_{BB} and CI_{CC} are concurrent. The point of concurrency is called the Nagel point of the triangle.

Theorem 11.14. Every triangle has a unique circumcenter, which we often denote by the letter O. As a result, the perpendicular bisectors of the edges of the triangle are concurrent at the same point.

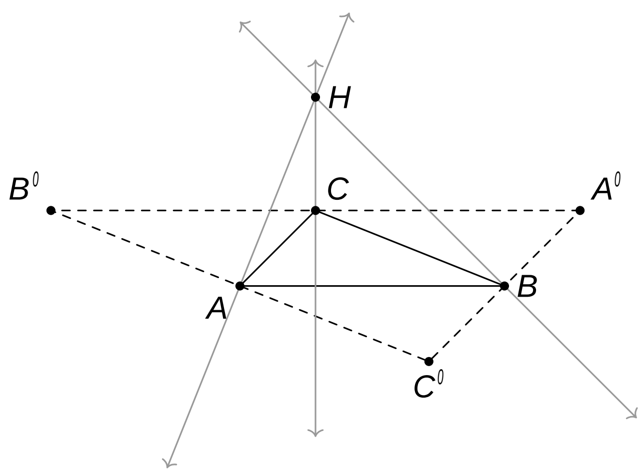
Proof. Let the triangle be $\triangle ABC$. The edges of a triangle are not parallel, so the perpendicular bisectors of BC and CA must intersect at a point O. Since O is on the perpendicular bisector of BC, O is equidistant from B and C. Similarly, since O is on the perpendicular bisector of CA, O is equidistant from C and A. Since O equidistant from all of A, B, C, it is a circumcenter, which we know to be unique when it exists by Theorem 8.16. As a consequence, the perpendicular bisectors of the three edges are concurrent at O. □

Definition 11.15. The circumcenter of a triangle is the unique point that is equidistant from the three vertices of the triangle. Note that the circumcenter does not necessarily lie in the interior of the triangle.



Theorem 11.16. Recall that an altitude or height of a triangle is a generalized cevian that is the perpendicular segment from a vertex to the line running through the opposite edge. Since the foot of an altitude may lie in the interior, endpoints, or outside of an edge, it is not always the case that an altitude is a cevian. Nonetheless, the lines through the three altitudes of any triangle are concurrent.

Proof. As with perpendicular bisectors and the circumcenter, the supposed point of concurrency might not lie in the interior of the triangle because an altitude is not necessarily a cevian. Since both altitudes and perpendicular bisectors involve perpendicularity, perhaps we can use the concurrency of the perpendicular bisectors of the edges of a triangle. We would be done if the lines running through the altitudes of our triangle were the perpendicular bisectors of some other triangle. To this end, we recall that if a triangle's vertices are the midpoints of the edges of another triangle, then the former is called the medial triangle of the latter. We will show that the lines through the altitude of the medial triangle are the perpendicular bisectors of the larger triangle, and then we will show that every triangle is the medial triangle of some other triangle.



Let $A'B'C'$ be a triangle. Let the midpoints of $B'C', C'A', A'B'$ be A, B, C respectively. By SSS congruence,

$$\triangle A'CB,\ \triangle BAC',\ \triangle ABC,\ \triangle CB'A$$

are all congruent. Then equal alternate interior angles of a transversal tell us that the line through BC is parallel to the line through $B'C'.$ Since the line through the altitude h_A of $\triangle ABC$ emanating from A is perpendicular to the line through $BC,$ it is also perpendicular to the line through $B'C'.$ Since h_A intersects $B'C'$ at A and A is the midpoint of $B'C',$ the line running through h_A is the perpendicular bisector of $B'C'.$ Symmetric arguments hold for the heights emanating from B and $C.$

Given $\triangle ABC,$ we can see that it is the medial triangle of some triangle as follows. Let ℓ_A be the line through A that is parallel to $BC;$ let ℓ_B be the line through B that is parallel to $CA;$ and let ℓ_C be the line through C that is parallel to $AB.$ We leave it to the reader to use triangle congruence theorems to show that it suffices for our purposes to use the triangle whose vertices are the pairwise intersections of $\ell_A, \ell_B, \ell_C.$ □

Definition 11.17. The point of concurrency of the lines through the altitudes of a triangle is called its **orthocenter**, which we often denoted by the letter H. Note that the orthocenter does not necessarily lie in the interior of the triangle.

Problem 11.18. Given $\triangle ABC,$ let let r be the inradius, let h_a, h_b, h_c be the altitudes emanating from A, B, C respectively, and let r_a, r_b, r_c be the exradii of the excircles tangent to BC, CA, AB respectively. Show that

$$\frac{1}{h_a} + \frac{1}{h_b} + \frac{1}{h_c} = \frac{1}{r} = \frac{1}{r_a} + \frac{1}{r_b} + \frac{1}{r_c}.$$

Example 11.19. Let $\triangle ABC$ be isosceles with apex A. Then the altitude emanating from A is contained in the ray that is the interior angle bisector emanating from A, which is contained in the perpendicular bisector of BC.

Solution. Let F be the foot of the altitude emanating from A, which lies in the interior of BC since $\triangle ABC$ is isosceles with apex A. By HL congruence, $\triangle ABF \cong \triangle ACF$. So $\angle BAF \cong \angle CAF$, which shows that that the altitude emanating from A lies on the interior angle bisector emanating from A.
Now let E be the point in the interior of BC at which which the interior angle bisector emanating from A intersects BC. By SAS congruence, $\triangle BAE \cong \triangle CAE$. So $\angle BEA = \angle CEA = 90^\circ$ and $BE = CE$, which shows that the interior angle bisector emanating from A is contained in the perpendicular bisector of BC. □

Problem 11.20. Show that, in any equilateral triangle, the centroid, incenter, circumcenter, and orthocenter are all the same point. This is called the **center** of the equilateral triangle.

Problem 11.21. Let A, B, C be three distinct non-collinear points in the plane. If O is the circumcenter and H is the orthocenter of $\triangle ABC$, verify that

$$\overrightarrow{OH} = \overrightarrow{OA} + \overrightarrow{OB} + \overrightarrow{OC}.$$

Using the dot product, deduce that

$$OH^2 = 9R^2 - a^2 - b^2 - c^2,$$

where $a = BC, b = CA, c = AB$ are lengths, and R is the circumradius.

Theorem 11.22 (Euler line)**.** For any $\triangle ABC$, the circumcenter O, centroid G and orthocenter H are collinear in that order such that $OG : GH = 1 : 2$. The line that runs through them is called the **Euler line**. This is true as a result of the following computations. Let a, b, c be the complex numbers corresponding to A, B, C respectively.

1. The centroid is $g = \dfrac{a+b+c}{3}$.
2. If the circumcenter is at the origin, then the orthocenter is $h = a + b + c$.

Proof. We will first compute g and h, and then we will prove their collinearity with the circumcenter.

1. Recall that the line segment from a complex number z to a complex number w is given by $(1-t)z + tw$ for $t \in [0, 1]$. Moreover, assuming z and w are distinct, the ratio between the distance from z to $(1-t)z + tw$ and the distance from z to w is

$$\frac{|(1-t)z + tw - z|}{|w - z|} = t.$$

The foot of the median emanating from a is the midpoint of b and c, which is $\frac{b+c}{2}$. Since the distance from a to the centroid g should be two-thirds of the distance from a to $\frac{b+c}{2}$, we substitute $t = \frac{2}{3}$ into

$$(1-t)a + t\left(\frac{b+c}{2}\right)$$

to get the centroid

$$\begin{aligned} g &= \left(1 - \frac{2}{3}\right)a + \frac{2}{3}\left(\frac{b+c}{2}\right) \\ &= \frac{a+b+c}{3}. \end{aligned}$$

2. Suppose the circumcenter is at the origin 0 and let the circumradius be R. In many arguments involving the geometry of complex numbers, the barrier to computing the location of a point is the removal of its conjugate from an expression. To lay the tracks, we note that since a, b, c lie on the circumcircle,

$$\begin{aligned} a\overline{a} &= |a|^2 = R^2, \\ b\overline{b} &= |b|^2 = R^2, \\ c\overline{c} &= |c|^2 = R^2. \end{aligned}$$

Let h be the orthocenter. Then the line through a and h is perpendicular to the line through b and c. By the complex perpendicularity criterion,

$$\begin{aligned} \frac{h-a}{b-c} &= -\overline{\left(\frac{h-a}{b-c}\right)} \\ &= -\left(\frac{\overline{h} - \frac{R^2}{a}}{\frac{R^2}{b} - \frac{R^2}{c}}\right) \\ &= \frac{bc}{aR^2} \cdot \frac{a\overline{h} - R^2}{b-c}. \end{aligned}$$

Clearing the denominators and rearranging the equation yields

$$haR^2 - a^2R^2 + bcR^2 = abc\overline{h}.$$

Similarly, since the line through b and h is perpendicular to the line through c and a, a symmetric derivation yields

$$hbR^2 - b^2R^2 + caR^2 = abc\overline{h}.$$

Since the right sides of the two equations are equal, we can equate the left sides to compute that

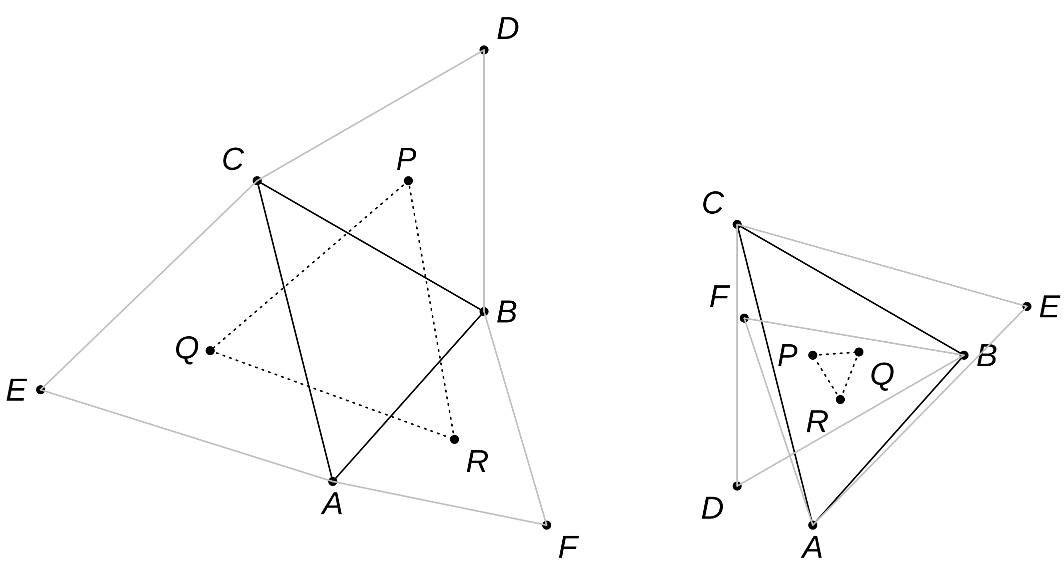
$$\begin{aligned} haR^2 - a^2R^2 + bcR^2 &= hbR^2 - b^2R^2 + caR^2 \\ ha - a^2 + bc &= hb - b^2 + ca \\ h(a-b) - (a-b)(a+b) - c(a-b) &= 0 \\ h &= a+b+c. \end{aligned}$$

Since translation preserves distances and counterclockwise angles, it preserves three relevant properties: collinearity, the order in which points are collinear, and the pairwise distances between the collinear points. So we may assume without loss of generality that the circumcenter O of $\triangle ABC$ is the origin $0.$ Then the centroid G is $g = \frac{a+b+c}{3}$ and the orthocenter H is $h = a + b + c.$ The segment from 0 to h is given by

$$(1 - t)0 + th = t(a + b + c).$$

Taking $t = \frac{1}{3}$ show that g lies $\frac{1}{3}$ of the way through the segment from 0 to $h,$ which simultaneously proves the collinearity of O, G, H in that order and the ratio $OG : GH = 1 : 2.$ As a side note, the Euler line contains other important triangle centers, such as the center of the nine-point circle and the Exeter point, but we will not discuss them. □

Problem 11.23 (Napoleon's theorem). Let $\triangle ABC$ be any (non-degenerate) triangle. If an equilateral triangle is erected on each side, with each one exterior to the triangle, prove that the centroids of those three equilateral triangles are the vertices of another equilateral triangle. Also prove that the result still holds if all the three equilateral triangles are drawn on the half-planes that cause their interiors to overlap with the interior of $\triangle ABC$.



Lemma 11.24. If B and C are points in $\mathbb{R}^2,$ and D is a point on segment BC such that we have the lengths $BD = x$ and $CD = y,$ then

$$\vec{D} = \frac{y\vec{B} + x\vec{C}}{x + y}.$$

Proof. By scaling down $\overrightarrow{BC},$ we find that

$$\frac{x}{x+y}\overrightarrow{BC} = \overrightarrow{BD},$$

$$\frac{x}{x+y}\vec{C} - \frac{x}{x+y}\vec{B} = \vec{D} - \vec{B}.$$

Then we can isolate

$$\vec{D} = \frac{y}{x+y}\vec{B} + \frac{x}{x+y}\vec{C}.$$

□

Example 11.25. Let A, B, C be three distinct non-collinear points in the plane. Prove that the position vector of the centroid of $\triangle ABC$ is

$$\vec{G} = \frac{\vec{A} + \vec{B} + \vec{C}}{3},$$

so that the barycentric coordinates of the centroid are $\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$.

Solution. By Lemma 11.24, the position vector of the midpoint M of AB is $\vec{M} = \frac{\vec{A} + \vec{B}}{2}$. By the geometric fact that the medians cut each other in a $2:1$ ratio with the larger segment being closer to the emanating vertex, we can compute that

$$\vec{G} - \vec{C} = \vec{CG} = \frac{2}{3}\left(\frac{\vec{A} + \vec{B}}{2} - \vec{C}\right) = \frac{\vec{A} + \vec{B}}{3} - \frac{2}{3}\vec{C}$$

$$\vec{G} = \frac{\vec{A} + \vec{B}}{3} - \frac{2}{3}\vec{C} + \vec{C} = \frac{\vec{A} + \vec{B} + \vec{C}}{3}.$$

There is also a way to get this result without using the $2:1$ intersection property of the centroid, and it uses the comparison of coefficients of linearly independent vectors. We will use this other method in Example 11.26 in order to find the position vector of the incenter. □

Example 11.26. Let A, B, C be three distinct non-collinear points on the plane. If I is the incenter of $\triangle ABC$, then the position vector of I is

$$\vec{I} = \frac{a\vec{A} + b\vec{B} + c\vec{C}}{a+b+c},$$

where $a = BC, b = CA, c = AB$ are lengths. Equivalently, a triple of unhomogenized barycentric coordinates for the incenter is (a, b, c).

Solution. First we will zone into the case of Lemma 11.24 where D is the foot of an angle bisector. If D is the foot of the angle bisector from A to BC, then $bx = cy$ by the angle bisector theorem, where $x = DB$ and $y = DC$. Then

$$\vec{D} = \frac{y\vec{B} + x\vec{C}}{x+y} = \frac{\frac{bx}{c}\vec{B} + x\vec{C}}{x + \frac{bx}{c}} = \frac{b\vec{B} + c\vec{C}}{b+c}.$$

Analogously, if E is the foot of the angle bisector from B to AC, then

$$\vec{E} = \frac{c\vec{C} + a\vec{A}}{c+a}.$$

Since I lies at the intersection of AD and BE, some scaling of $\vec{AD}$ will equal $\vec{AI} = \vec{I} - \vec{A}$, and some scaling of $\vec{BE}$ will equal $\vec{BI} = \vec{I} - \vec{B}$. So there exist constants λ and μ such that

$$\lambda\left(\frac{b\vec{B} + c\vec{C}}{b+c} - \vec{A}\right) = \vec{I} - \vec{A},$$
$$\mu\left(\frac{c\vec{C} + a\vec{A}}{c+a} - \vec{B}\right) = \vec{I} - \vec{B}.$$

As a result,

$$\vec{I} = \lambda\left(\frac{b\vec{B} + c\vec{C}}{b+c} - \vec{A}\right) + \vec{A} = \mu\left(\frac{c\vec{C} + a\vec{A}}{c+a} - \vec{B}\right) + \vec{B},$$

which we can rewrite as

$$\lambda\left(\frac{b(\vec{B} - \vec{A}) + c(\vec{C} - \vec{A})}{b+c}\right) + (\vec{A} - \vec{B}) = \mu\left(\frac{c(\vec{C} - \vec{B}) + a(\vec{A} - \vec{B})}{c+a}\right)$$
$$\frac{\lambda b}{b+c}\vec{AB} + \frac{\lambda c}{b+c}\vec{AC} - \vec{AB} + \frac{\mu c}{c+a}\vec{CB} + \frac{\mu a}{c+a}\vec{AB} = 0.$$

By substituting $\vec{AB} = -\vec{BC} - \vec{CA}$ into this, we get

$$0 = -\frac{\lambda b}{b+c}\vec{BC} - \frac{\lambda b}{b+c}\vec{CA} + \frac{\lambda c}{b+c}\vec{AC} + \vec{BC} + \vec{CA} + \frac{\mu c}{c+a}\vec{CB} - \frac{\mu a}{c+a}\vec{BC} - \frac{\mu a}{c+a}\vec{CA};$$
$$0 = \left(\frac{\lambda b}{b+c} + \mu - 1\right)\vec{BC} + \left(\frac{\mu a}{c+a} + \lambda - 1\right)\vec{CA}.$$

Since A, B, C are distinct and non-collinear, $\vec{BC}$ and $\vec{CA}$ are linearly independent, so

$$\frac{\lambda b}{b+c} + \mu = \frac{\mu a}{c+a} + \lambda = 1.$$

Solving for λ and μ yields

$$\lambda = \frac{b+c}{a+b+c} \text{ and } \mu = \frac{c+a}{a+b+c}.$$

Therefore, we compute

$$\vec{I} = \lambda\left(\frac{b\vec{B} + c\vec{C}}{b+c} - \vec{A}\right) + \vec{A} = \frac{b+c}{a+b+c}\left(\frac{b\vec{B} + c\vec{C}}{b+c} - \vec{A}\right) + \vec{A} = \frac{a\vec{A} + b\vec{B} + c\vec{C}}{a+b+c}.$$

□

11.2 Area Formulas

Theorem 11.27 (Triangle area formulas)**.** Let $\triangle ABC$ have edges $a = BC, b = CA, c = AB,$ semiperimeter $s,$ inradius r and circumradius $R.$ Then the area of $\triangle ABC$ is

$$\begin{aligned}[ABC] &= rs \\ &= \sqrt{s(s-a)(s-b)(s-c)} \text{ (Heron's formula)} \\ &= \frac{1}{2}ab\sin C \\ &= \frac{abc}{4R}.\end{aligned}$$

Proof. We will prove the formulas in succession as each will help to prove the next one.

1. The first formula was already proven more generally for all tangential polygons, but we will review the proof here in the special case of a triangle. Letting the incenter be $I,$

$$[ABC] = [BIC] + [CIA] + [AIB] = \frac{ra}{2} + \frac{rb}{2} + \frac{rc}{2} = r\left(\frac{a+b+c}{2}\right) = rs.$$

2. This proof of Heron's formula is due to Miles Dillon Edwards [4], who published it relatively recently. Let the incircle touch AB, BC, CA at X, Y, Z respectively. By Ravi substitution, let

$$\begin{aligned}u &= AY = AZ = s - a, \\ v &= BZ = BX = s - b, \\ w &= CX = XY = s - c.\end{aligned}$$

Since the incenter is the point of concurrency of the angle bisectors, SAS congruence yields the congruence of the following pairs of right triangles

$$\begin{aligned}\triangle AIY &= \triangle AIZ, \\ \triangle BIZ &= \triangle BIX, \\ \triangle CIX &= \triangle CIY.\end{aligned}$$

So we can label the equal pairs of angles

$$\begin{aligned}\alpha &= \angle AIY = \angle AIZ, \\ \beta &= \angle BIZ = \angle BIX, \\ \gamma &= \angle CIX = \angle CIY.\end{aligned}$$

Then $2\alpha + 2\beta + 2\gamma = 2\pi$ or $\alpha + \beta + \gamma = \pi.$ By applying the definitions of cosine and sine to $\alpha = \angle AIY$ in $\triangle AIY,$ we can compute the two components of the complex number

$$r + iu = AI(\cos\alpha + i\sin\alpha) = AIe^{i\alpha}.$$

Similarly, we express two more complex numbers in rectangular and trigonometric form

$$r + iv = BI e^{i\beta},$$
$$r + iw = CI e^{i\gamma}.$$

Multiplying the three equations yields

$$(r+iu)(r+iv)(r+iw) = AI \cdot BI \cdot CI \cdot e^{i(\alpha+\beta+\gamma)} = -AI \cdot BI \cdot CI.$$

This complex number is real, so its imaginary part is

$$\begin{aligned} 0 &= \operatorname{Im}((r+iu)(r+iv)(r+iw)) \\ &= \operatorname{Im}(r^3 + ir^2(u+v+w) - r(uv+vw+wu) - iuvw) \\ &= r^2(u+v+w) - uvw. \end{aligned}$$

Therefore, the area of $\triangle ABC$ is

$$[ABC] = rs = s\sqrt{\frac{uvw}{u+v+w}} = \sqrt{s(s-a)(s-b)(s-c)}.$$

As a side note, expanding Heron's formula leads us to an algebraic identity

$$(a+b+c)(-a+b+c)(a-b+c)(a+b-c) = 2a^2b^2 + 2b^2c^2 + 2c^2a^2 - a^4 - b^4 - c^4,$$

which has remarkably few terms in the expanded side.

3. Since $\angle ACB \in (0,\pi)$, we know that $\sin C \in (0,1)$. So $\sin\angle C = \sqrt{1-\cos^2 C}$, where we are able to take the positive square root. By the cosine law and the difference of squares factorization,

$$\begin{aligned} \frac{1}{2}ab\sin C &= \frac{1}{2}ab\sqrt{1-\cos^2 C} \\ &= \frac{1}{2}ab\sqrt{1-\left(\frac{a^2+b^2-c^2}{2ab}\right)^2} \\ &= \frac{1}{4}\sqrt{(2ab)^2-(a^2+b^2-c^2)^2} \\ &= \frac{1}{4}\sqrt{(2ab-a^2-b^2+c^2)(2ab+a^2+b^2-c^2)} \\ &= \frac{1}{4}\sqrt{(c^2-(a-b)^2)((a+b)^2-c^2)} \\ &= \frac{1}{4}\sqrt{(a+b+c)(-a+b+c)(a-b+c)(a+b-c)} \\ &= \sqrt{s(s-a)(s-b)(s-c)}. \end{aligned}$$

This is Heron's formula, so we are done.

4. By the extended law of sines, $\dfrac{c}{\sin C} = 2R$, so

$$[ABC] = \frac{1}{2}ab\sin C = \frac{1}{2}ab \cdot \frac{c}{2R} = \frac{abc}{4R}.$$

□

Corollary 11.28. Let $\triangle ABC$ have edges $a = BC, b = CA, c = AB$, semiperimeter s, inradius r, circumradius R, height h_a emanating from A, and exradius r_a of the A-excircle. Then

$$\begin{aligned} r &= \frac{[ABC]}{s}, \\ R &= \frac{abc}{4[ABC]}, \\ h_a &= \frac{2[ABC]}{a}, \\ r_a &= \frac{[ABC]}{s-a}, \end{aligned}$$

where each expression can be written purely in terms of a, b, c using Heron's formula for $[ABC]$.

Problem 11.29. Given $\triangle ABC$, let the inradius be r and the exradii of the excircles tangent to BC, CA, AB be r_a, r_b, r_c respectively. Show that

$$[ABC] = \sqrt{r r_a r_b r_c}.$$

Theorem 11.30 (Bretschneider's formula)**.** Let $ABCD$ be a generalized quadrilateral. Let $AB = a, BC = b, CD = c, DA = d$, let the interior angles at A, B, C, D be $\alpha, \beta, \gamma, \delta$ respectively, and let the semiperimeter be $s = \frac{a+b+c+d}{2}$. Then

$$[ABCD] = \sqrt{(s-a)(s-b)(s-c)(s-d) - abcd\cos^2\left(\frac{\alpha+\gamma}{2}\right)}.$$

Proof. There are a few configurations to consider:

- $ABCD$ is a convex polygon.
- $ABCD$ is a concave polygon, which is equivalent to one of the four vertices lying in the interior of the triangle formed by the other three vertices.
- $ABCD$ is not a polygon, which is equivalent to some three (consecutive) vertices being collinear.

The reader should verify that, thanks to the fact that sine is positive on $(0, \pi)$, negative on $(\pi, 2\pi)$, and 0 at π, it is true in all configurations that

$$[ABCD] = \frac{1}{2}ad\sin\alpha + \frac{1}{2}bc\sin\gamma.$$

Since the target formula has an overarching square root, we square our equation to get

$$4[ABCD]^2 = (ad)^2\sin^2\alpha + (bc)^2\sin^2\gamma + 2abcd\sin\alpha\sin\gamma.$$

By the cosine law,

$$BD^2 = a^2 + d^2 - 2ad\cos\alpha = b^2 + c^2 - 2bc\cos\gamma .$$

We rearrange this equation and square it to get

$$\begin{aligned}\left(\frac{a^2 + d^2 - b^2 - c^2}{2}\right)^2 &= (ad\cos\alpha - bc\cos\gamma)^2\\ &= (ad)^2\cos^2\alpha + (bc)^2\cos^2\gamma - 2abcd\cos\alpha\cos\gamma ,\end{aligned}$$

which is a cosine counterpart of our first equation. Adding the two equations gives

$$\begin{aligned}&4[ABCD]^2 + \left(\frac{a^2 + d^2 - b^2 - c^2}{2}\right)^2\\ &= ((ad)^2 + (bc)^2)(\sin^2 + \cos^2) + 2abcd(\sin\alpha\sin\gamma - \cos\alpha\cos\gamma)\\ &= (ad)^2 + (bc)^2 - 2abcd\cos(\alpha + \gamma)\\ &= (ad + bc)^2 - 4abcd\left(\frac{\cos(\alpha + \gamma) + 1}{2}\right)\\ &= (ad + bc)^2 - 4abcd\cos^2\left(\frac{\alpha + \gamma}{2}\right),\end{aligned}$$

where we have used the Pythagorean identity, an angle sum identity and a half angle identity. By using the difference of squares factorization several times in the same manner used to derive Heron's formula, we find that

$$\begin{aligned}[ABCD]^2 &= \left(\frac{ad + bc}{2}\right)^2 - \left(\frac{a^2 + d^2 - b^2 - c^2}{4}\right)^2\\ &= (s - a)(s - b)(s - c)(s - d).\end{aligned}$$

Note that we could replace $\cos^2\left(\frac{\alpha+\gamma}{2}\right)$ with $\cos^2\left(\frac{\beta+\delta}{2}\right)$ because

$$\cos^2\left(\frac{\beta+\delta}{2}\right) = \cos^2\left(\pi - \frac{\alpha+\gamma}{2}\right) = \left[-\cos\left(\frac{\alpha+\gamma}{2}\right)\right]^2 = \cos^2\left(\frac{\alpha+\gamma}{2}\right).$$

□

Corollary 11.31 (Brahmagupta's formula). If a, b, c, d are the edges of a cyclic quadrilateral $ABCD$ and the semiperimeter is $s = \frac{a+b+c+d}{2}$, then

$$[ABCD] = \sqrt{(s - a)(s - b)(s - c)(s - d)}.$$

Proof. This follows immediately from Bretschneider's formula because opposite angles in a cyclic quadrilateral are supplementary, and $\cos^2\frac{\pi}{2} = 0$. So the extra term in Bretschneider's formula vanishes. □

Chapter 12

Conics

> "I was almost driven to madness in considering and calculating this matter. I could not find out why the planet would rather go on an elliptical orbit. Oh, ridiculous me! As the liberation in the diameter could not also be the way to the ellipse. So this notion brought me up short, that the ellipse exists because of the liberation. With reasoning derived from physical principles, agreeing with experience, there is no figure left for the orbit of the planet but a perfect ellipse."
>
> – *Johannes Kepler*

While conic sections have been studied since antiquity, we will study conics from the perspective of algebraic equations. Beyond the basic definitions, our goals are three-fold. First, we will show that every conic is congruent to a conic that arises from a convenient type of bivariate quadratic, which are called standard forms. Secondly, there is a pair of uniqueness results that will interest us. Specifically, we will determine when two bivariate quadratics can produce the same conic, and we will find all directrices, eccentricities, and foci for fixed conics. Lastly, we will look at geometric properties of the curves that are conics.

12.1 Bivariate Quadratics

Conic sections were originally defined as the curves on the boundaries of cross-sections of any "double" right cone that extends infinitely. Despite the fact that this is how conic sections came to have their name, we have preferred a definition from analytic geometry involving a focus, an eccentricity, and a directrix.

We have already seen an example of a conic: circles. The general equation of a circle is

$$(x - a)^2 + (y - b)^2 = r^2.$$

This can be expanded and written in the form

$$x^2 + y^2 + Ax + By + C = 0,$$

though not all equations of this form lead to a circle. While circles are technically a type of conic, the following definition that we will use for a conic will not cover it. As such, whenever we write about conics, it will be in reference to non-circular conics.

Definition 12.1. A **DEF-construction** is an ordered triple of a line ℓ called the **directrix**, a positive real number e called the **eccentricity**, and a point F outside ℓ called the **focus**. In a given DEF-construction, the distance from the focus to the directrix is called the **focal parameter**. A **constructible conic** is a set of points that, for some DEF-construction (ℓ, e, F), is equal to the locus of all points P such that the distance from P to F is equal to e times the distance from P to ℓ. In our case, all references to conics will be to those that are constructible.

Note that the same constructible conic might arise from more than one DEF-construction. In fact, the question of classifying all DEF-constructions of a conic that has at least one DEF-construction will later occupy us.

Definition 12.2. Let e be the eccentricity in a DEF-construction of a conic. Then the conic is called an

$$\begin{cases} \text{ellipse} & \text{if } 0 < e < 1 \\ \text{parabola} & \text{if } e = 1 \\ \text{hyperbola} & \text{if } e > 1 \end{cases}.$$

Some sources fit a circle within this classification scheme by referring to it as a limiting case with an eccentricity of $e = 0$, but we will not do so because a circle has no directrix in the plane.

As we will see, the distinction between ellipses, parabolas and hyperbolas is justified because the algebra and geometry play out somewhat differently in each case.

Theorem 12.3. In a DEF-construction, let the directrix be given by $ax + by + c = 0$, let (x_0, y_0) be the focus, and let $e > 0$ be the eccentricity. Then (x, y) is a point on the resulting conic if and only

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0,$$

where the coefficients are

$$\begin{aligned} A &= a^2(1 - e^2) + b^2, \\ B &= -2abe^2, \\ C &= a^2 + (1 - e^2)b^2, \\ D &= -2(ace^2 + (a^2 + b^2)x_0) \\ E &= -2(bce^2 + (a^2 + b^2)y_0) \\ F &= (a^2 + b^2)(x_0^2 + y_0^2) - e^2c^2. \end{aligned}$$

This is one tuple of coefficients that works, and we will later show that all other such tuples are its non-zero scalar multiples.

Proof. The point (x, y) is on the conic if and only if the distance from (x, y) to (x_0, y_0) is equal to e times the distance from (x, y) to the line given by $ax + by + c = 0$. By the point-point and point-line distance formulas, this relation is given by the equation

$$\sqrt{(x - x_0)^2 + (y - y_0)^2} = \frac{e|ax + by + c|}{\sqrt{a^2 + b^2}}.$$

Neither side can be negative, so squaring is a reversible step. After that, it is a matter of taking everything to one side, expanding, and collecting like terms:

$$\begin{aligned} 0 &= (a^2 + b^2)[(x - x_0)^2 + (y - y_0)^2] - e^2(ax + by + c)^2 \\ &= (a^2 + b^2)(x^2 + y^2 - 2x_0x - 2y_0y + x_0^2 + y_0^2) \\ &\quad - e^2(a^2x^2 + b^2y^2 + c^2 + 2abxy + 2acx + 2bcy) \\ &= (a^2 + b^2 - e^2a^2)x^2 - 2e^2abxy + (a^2 + b^2 - e^2b^2)y^2 \\ &\quad + [-2x_0(a^2 + b^2) - 2e^2ac]x + [-2y_0(a^2 + b^2) - 2e^2bc]y \\ &\quad + [(a^2 + b^2)(x_0^2 + y_0^2) - e^2c^2], \end{aligned}$$

which matches the desired result after some minor algebraic fiddling. □

Definition 12.4. A **bivariate quadratic** is a two-variable function of the form

$$Q(x, y) = Ax^2 + Bxy + Cy^2 + Dx + Ey + F,$$

where the coefficients are real.

Example. As we saw, every constructible conic is the zero set of some bivariate quadratic. However, it is not true that the zero set of every bivariate quadratic is a constructible conic. As an extreme example, there are no real points (x, y) such that

$$x^2 + y^2 + 1 = 0.$$

There are also bivariate quadratics which have non-empty zero sets that are not constructible conics. Those curves are called **degenerate conics**. Henceforth, by "**conic**," we shall only refer to a constructible one, as in one that arises from a DEF-construction.

The next result will show the existence of a conic as the graph of a convenient equation for each possible pair of an eccentricity and a focal parameter.

Theorem 12.5. Given $e > 0$ and $f > 0$, there exists a conic with eccentricity e and focal parameter f. Specifically, we can choose it to be the graph of one of the following equations, which we call the **standard form** of an ellipse, parabola or hyperbola.

1. If $0 < e < 1$ then we choose $\frac{x^2}{p^2} + \frac{y^2}{q^2} = 1$ where $p = \frac{fe}{1 - e^2} > 0$ and $q = p\sqrt{1 - e^2} > 0$.
2. If $e = 1$ then we choose $y^2 = 4px$ where $p = \frac{f}{2} > 0$.
3. If $e > 1$ then we choose $\frac{x^2}{p^2} - \frac{y^2}{q^2} = 1$ where $p = \frac{fe}{e^2 - 1} > 0$ and $q = p\sqrt{e^2 - 1} > 0$.

Along the way, we will show that, for a conic that is a zero set of a bivariate quadratic in standard form, there is at least one DEF-construction of a parabola, and at least two DEF-constructions of an ellipse or hyperbola.

Proof. Suppose we are given $e > 0$ and $f > 0$. We construct the conic in each case as follows.

1. Suppose $0 < e < 1$. We choose the focus to be $(pe, 0)$ and the corresponding directrix to be $x = \frac{p}{e}$ where $p = \frac{fe}{1-e^2}$. The focal parameter is

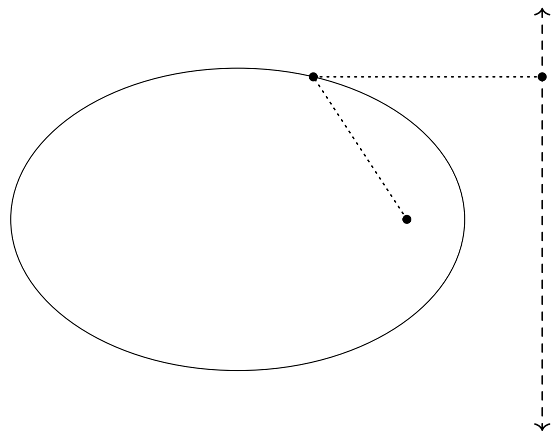
$$\frac{p}{e} - pe = p\left(\frac{1-e^2}{e}\right) = f$$

as desired. Then (x, y) lies on the ellipse with this focus-directrix pair and eccentricity e if and only if

$$\sqrt{(x-pe)^2 + y^2} = e\left|x - \frac{p}{e}\right|.$$

Neither side can be negative, so squaring both sides is reversible. We do this and rearrange the terms to get

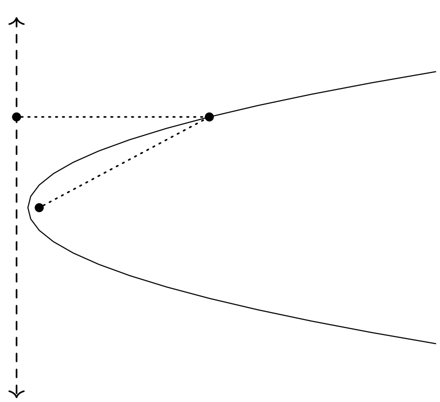
$$\frac{x^2}{p^2} + \frac{y^2}{p^2(1-e^2)} = 1.$$



2. Suppose $e = 1$. We choose the focus to be $(p, 0)$ and the corresponding directrix to be $x = -p$ where $p = \frac{f}{2}$. The focal parameter is f as desired. Then (x, y) lies on the parabola with this focus-directrix pair and eccentricity $e = 1$ if and only if

$$\sqrt{(x-p)^2 + y^2} = |x + p|.$$

Squaring and rearranging the terms yields $y^2 = 4px$.



3. Suppose $e > 1.$ As with an ellipse, we choose the focus to be $(pe, 0)$ and the corresponding directrix to be $x = \frac{p}{e}$ where $p = \frac{fe}{e^2 - 1}.$ The focal parameter is

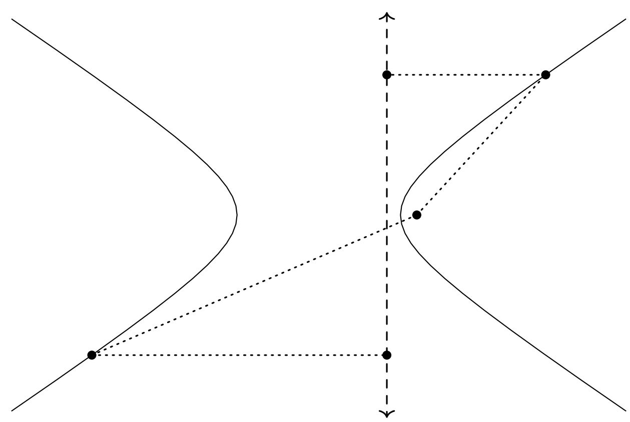
$$pe - \frac{p}{e} = p\left(\frac{e^2 - 1}{e}\right) = f$$

as desired. Then (x, y) lies on the hyperbola with this focus-directrix pair and eccentricity e if and only if

$$\sqrt{(x - pe)^2 + y^2} = e\left|x - \frac{p}{e}\right|.$$

Squaring both sides and rearranging the terms yields

$$\frac{x^2}{p^2} - \frac{y^2}{p^2(e^2 - 1)} = 1.$$



In fact, for ellipses or hyperbolas in standard form, the focus-directrix pair of $(-pe, 0)$ and $x = -\frac{p}{2}$ works with the same eccentricity and gives the same focal parameter. Admittedly, the choices of foci and directrices may seem to be lacking motivation. In Theorem 12.8, we will describe a naturally-arising Euclidean isometry that maps a conic that is the graph of a generic bivariate quadratic to a conic that is the graph of an equation in standard form. □

If a conic is the graph of an equation in standard form, we will sometimes abuse language to say that the conic itself is in standard form, though the meaning is clear.
Inspired by the fact that two polygons are congruent if and only if the image of one under a Euclidean isometry is the other, we make the following general definition.

Definition 12.6. Two subsets of $\mathbb{R}^2$ are said to be **congruent** if the image of one of the sets under some Euclidean isometry is equal to the other set.

Example. This definition applies to subsets of the plane that are conics, which gives us the notion of congruent conics. As we will see, congruence of conics is a useful concept because many of the parameters that we will define for conics are preserved under congruence, and conversely some combinations of matching parameters imply congruence.

Theorem 12.7. The image of a conic S under a Euclidean isometry k is a congruent set $k(S)$ that is a conic that has the same eccentricity and focal parameter as those of S. Moreover, each focus-directrix pair of S is mapped to a focus-directrix pair of $k(S)$. Conversely, two conics with the same eccentricity and the same focal parameter are congruent.

Proof. For the first direction, let the eccentricity and focal parameter of S be e and f respectively, and let F be a focus of S and ℓ be the corresponding directrix of $S.$ Then k maps F to a point $k(F)$ and ℓ to a line $k(\ell)$, where the latter is known from Theorem 3.13. That theorem also states that Euclidean isometries preserve the distance between a point and a line, so the distance between $k(F)$ and $k(\ell)$ is the same as the (non-zero) distance between F and $\ell.$ Let S' be the conic with eccentricity e that has $k(F)$ as a focus and $k(\ell)$ as the corresponding directrix. By the preceding statement, S' has focal parameter $f.$ Now it suffices to show that $k(S) = S'.$ This idea of constructing one object and showing that it turns out to be the same as another object is called using a "phantom" object, where the object is most commonly just a point, like in the proof of the converse of Ceva's theorem (Theorem 10.12). Suppose $P \in S'.$ This is true if and only if

$$d(P, k(F)) = e \cdot d(P, k(\ell)),$$

where d denotes the distance between the enclosed pair of elements; we will be using this notation later as well, so take heed. Since k is a Euclidean isometry, Theorem 3.2 says that it has an inverse k^{-1} that is also a Euclidean isometry. Euclidean isometries preserve the distance between a pair of points and the distance between a point and a line, so the last equation is true if and only if

$$d(k^{-1}(P), F) = e \cdot d(k^{-1}(P), \ell).$$

This is equivalent to saying that $k^{-1}(P)$ lies on $S,$ which is true if and only if $P \in k(S).$ So, $P \in S'$ if and only if $P \in k(S).$ Thus, $k(S)$ is not only congruent to S as set of points, but $k(S)$ is S', which is a conic that has the same eccentricity and focal parameter as $S.$ Moreover, $k(F)$ and $k(\ell)$ form a focus-directrix pair of $k(S).$

In the other direction, suppose S and S' are conics with the same eccentricity and focal parameter. Since they have the same eccentricity, it suffices to show that the image of a focus-directrix pair of S under some Euclidean isometry is a focus-directrix pair of $S'.$ Let F and ℓ be a focus-directrix pair of $S,$ and let F' and ℓ' be a focus-directrix pair of $S'.$ Let D be the foot of the perpendicular from F to $\ell,$ and let D' be the foot of the perpendicular from F' to $\ell'.$ Since S and S' have the same focal parameter, $FD = F'D'.$ Thus, a translation followed by a rotation will map F to F' and D to $D'.$ This causes ℓ to get mapped to ℓ' as well since there is exactly one line that goes through a given point and is perpendicular to a given line segment. □

The last result shows that conics arising from DEF-constructions that have the same eccentricity and focal parameter are unique up to congruence.

Theorem 12.8. Suppose S is a constructible conic. Let

$$Q(x, y) = Ax^2 + Bxy + Cy^2 + Dx + Ey + F$$

be a bivariate quadratic whose zero set is $S.$ If $B \neq 0,$ then rotating S counterclockwise by θ around the origin, where θ satisfies $\cot 2\theta = \frac{C-A}{B},$ yields a congruent conic that is the zero set of a bivariate quadratic, the coefficient of whose xy term is $0.$ Then an appropriate translation and/or reflection of this conic is a congruent conic that is the zero set of a bivariate quadratic in standard form.

Proof. The counterclockwise rotation of S around the origin by some θ is the graph of

$$Q(x\cos\theta + y\sin\theta, y\cos\theta - x\sin\theta).$$

We will not show the messy details, but the coefficient of the xy term after expanding and collecting like terms is

$$2(A-C)\cos\theta\sin\theta + B(\cos^2\theta - \sin^2\theta).$$

We want to choose θ so that this expression is $0.$ Using the double angle identities for sine and cosine, the xy term vanishes if and only if

$$(C-A)\sin(2\theta) = B\cos(2\theta).$$

If $A = C,$ then we can choose $\theta = \frac{\pi}{4}.$ Otherwise, we can apply the arctan function to $\tan(2\theta) = \frac{B}{C-A}$ and find a suitable angle θ in the range of $\arctan,$ which is $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right).$ If this angle is negative, we can add 2π to make it positive so that the rotation is still counterclockwise if desired.

After the xy term is eliminated, we can complete the square in whichever of the variables x and/or y still have a term of second degree left and then apply an appropriate translation (vertical and/or horizontal), possibly along with a reflection so that it is the graph of an equation in standard form. This reflection could be across a coordinate axis, or across the lines $x = y$ to swap the roles of the variables. □

We will now work through a number of preliminary results before we can classify the coefficient tuple of bivariate quadratics whose infinite zero set is a given constructible conic.

Lemma 12.9. Let ℓ be a line in the plane, parametrized by

$$(x, y) = (r, s) + t(v, w).$$

Let

$$Q(x, y) = Ax^2 + Bxy + Cy^2 + Dx + Ey + F$$

be a bivariate quadratic. Then either ℓ intersects the zero set $Q(x, y) = 0$ at $0, 1,$ or 2 distinct points, or every point on ℓ lies on $Q(x, y) = 0$.

Proof. By substituting

$$x = r + tv,$$
$$y = s + tw$$

into $Q(x; y)$, the intersections of the line with the conic occur at the values of t that satisfy

$$\begin{aligned} 0 &= Q(r + tv; s + tw) \\ &= A(r + tv)^2 + B(r + tv)(s + tw) + C(s + tw)^2 + D(r + tv) + E(s + tw) + F: \end{aligned}$$

Upon expanding this and collecting like terms in the parameter t, we get a polynomial equation

$$0 = \quad t^2 + \quad t +$$

of degree at most 2. If $\ ^2 + \ ^2 + \ ^2 = 0$, then every t gives a solution, making ` a subset of the curve cut out by $Q(x; y) = 0$. Otherwise,

$$t^2 + \quad t +$$

has at least one non-zero coefficient, leaving $0; 1;$ or 2 distinct solutions t, by the fundamental theorem of algebra from Volume 1. □

Problem 12.10 (Bivariate quadratic identity theorem). If there are two bivariate quadratics

$$\begin{aligned} Q_1(x; y) &= A_1x^2 + B_1xy + C_1y^2 + D_1x + E_1y + F_1; \\ Q_2(x; y) &= A_2x^2 + B_2xy + C_2y^2 + D_2x + E_2y + F_2 \end{aligned}$$

that give the same output for all real x and y, use the identity theorem for polynomials from Volume 1 to prove that corresponding coefficients match. That is, we have the equality

$$(A_1; B_1; C_1; D_1; E_1; F_1) = (A_2; B_2; C_2; D_2; E_2; F_2):$$

Lemma 12.11 (Bivariate quadratic decomposition). Let

$$Q(x; y) = Ax^2 + Bxy + Cy^2 + Dx + Ey + F$$

be a bivariate quadratic and let

$$L_1(x; y) = a_1x + b_1y + c_1;$$

be a line, where $a_1^2 + b_1^2 \neq 0$. This means $a_1 \neq 0$ or $b_1 \neq 0$, which splits the result into two (overlapping) cases:

- Case 1: If $a_1 \neq 0$, then there exists a linear function

  $$L_2(x; y) = a_2x + b_2y + c_2$$

  such that

  $$Q(x; y) = L_1(x; y)L_2(x; y) + M(y);$$

  where $M(y)$ is a unique quadratic whose only variable is y.

- Case 2: If $b_1 \neq 0$, then there exists a linear function

  $$L_3(x; y) = a_3x + b_3y + c_3$$

  such that

  $$Q(x; y) = L_1(x; y)L_3(x; y) + N(x);$$

  where $N(x)$ is a unique quadratic whose only variable is x.

Proof. First we handle Case 1. Suppose $a_1 \neq 0$. If the stated decomposition exists, then

$$\begin{aligned}
&Ax^2 + Bxy + Cy^2 + Dx + Ey + F - M(y) \\
&= Q(x,y) - M(y) \\
&= L_1(x,y)L_2(x,y) \\
&= (a_1x + b_1y + c_1)(a_2x + b_2y + c_2) \\
&= (a_1a_2)x^2 + (a_1b_2 + a_2b_1)xy + (b_1b_2)y^2 + (a_1c_2 + a_2c_1)x + (b_1c_2 + b_2c_1)y + c_1c_2.
\end{aligned}$$

By Problem 12.10, since $M(y)$ does not contain the variable x, the coefficients of x^2, xy, x can be matched in the two expressions to give

$$\begin{aligned}
A &= a_1a_2, \\
B &= a_1b_2 + a_2b_1, \\
D &= a_1c_2 + a_2c_1.
\end{aligned}$$

Since it was assumed that $a_1 \neq 0$, we get

$$\begin{aligned}
a_2 &= \frac{A}{a_1}, \\
B = a_1b_2 + \frac{Ab_1}{a_1} \quad &\Longrightarrow \quad b_2 = \frac{Ba_1 - Ab_1}{a_1^2}, \\
D = a_1c_2 + \frac{Ac_1}{a_1} \quad &\Longrightarrow \quad c_2 = \frac{Da_1 - Ac_1}{a_1^2}.
\end{aligned}$$

This determines

$$L_2 = a_2x + b_2y + c_2.$$

Finally, $M(y)$ is a well-defined and uniquely determined quadratic in the variable y because it is determined as $Q(x,y) - L_1(x,y)L_2(x,y)$, due to how a_2, b_2, c_2 were engineered:

$$M(y) = Q(x,y) - L_1(x,y)L_2(x,y) = Cy^2 + Ey + F.$$

Note that the proof relied on division by $a_1 \neq 0$. The proof of Case 2 is symmetrical, with reliance on division by b_1, which is acceptable due to the assumption that $b_1 \neq 0$ in Case 2. □

Lemma 12.12. Let

$$Q(x,y) = Ax^2 + Bxy + Cy^2 + Dx + Ey + F$$

be a bivariate quadratic, and let

$$L_1(x,y) = a_1x + b_1y + c_1$$

be a linear function with $a_1^2 + b_1^2 \neq 0$. If every point on $L_1(x,y) = 0$ lies on $Q(x,y) = 0$, then there exists a linear function

$$L_1(x,y) = a_2x + b_2y + c_2$$

such that

$$Q(x,y) = L_1(x,y)L_2(x,y).$$

In fact, this result strengthens itself by holding even if we know only that the zero sets $L_1(x,y) = 0$ and $Q(x,y) = 0$ intersect in at least 3 distinct points.

Proof. Since $a_1^2 + b_1^2 \neq 0$, we know that $a_1 \neq 0$ or $b_1 \neq 0$. If $a_1 \neq 0$, then Lemma 12.11 says that there exists a linear function

$$L_2(x; y) = a_2 x + b_2 y + c_2$$

and $M(y)$ such that

$$Q(x; y) = L_1(x; y) L_2(x; y) + M(y).$$

For each $y \in \mathbb{R}$, there exists exactly one $x \in \mathbb{R}$ (call it $\varphi(y)$) such that

$$L_1(x; y) = 0,$$

since we can divide by the non-zero coefficient a_1 of x to get

$$a_1 x + b_1 y + c = 0 \implies x = -\frac{b_1 y + c}{a_1} = \varphi(y).$$

Since it is assumed that every point on $L_1(x; y) = 0$ also lies on $Q(x; y) = 0$, it means Q vanishes wherever L_1 vanishes. As a result,

$$\begin{aligned} 0 &= Q(\varphi(y); y) \\ &= L(\varphi(y); y) L_2(\varphi(y); y) + M(y) \\ &= 0 \cdot L_2(\varphi(y); y) + M(y) \\ &= M(y). \end{aligned}$$

Since this says that $M(y)$ vanishes for all real y, the identity theorem for polynomials from Volume 1 says that the coefficients of M are all 0. This leaves us with

$$Q(x; y) = L_1(x; y) L_2(x; y),$$

as desired.

The second case, where $b_1 \neq 0$, follows from a symmetrical argument.

For the strengthening, suppose that, instead of knowing that $L_1(x; y) = 0$ is a subset of $Q(x; y) = 0$, we instead only knew that there are at least 3 points of intersection between $L_1(x; y) = 0$ and $Q(x; y) = 0$. Then Lemma 12.9 says that more than two points of intersection implies that $L_1(x; y) = 0$ is a subset of $Q(x; y) = 0$. The rest has already been argued. □

Lemma 12.13. If $Q(x; y) = 0$ is a constructible conic, meaning it arises from a DEF-construction, then the conic $Q(x; y) = 0$ cannot contain a whole line. By Lemma 12.12, this implies that there do not exist linear functions

$$\begin{aligned} L_1(x; y) &= a_1 x + b_1 y + c_1, \\ L_2(x; y) &= a_2 x + b_2 y + c_2 \end{aligned}$$

such that

$$Q(x; y) = L_1(x; y) L_2(x; y).$$

Proof. Since the conic is constructible, let a focus be $P = (x_0, y_0)$ and its corresponding directrix be

$$ax + by + c = 0.$$

For contradiction, suppose the conic $Q(x, y) = 0$ contains a line, within which there are points $V = (x_1, y_1)$ and $W = (x_2, y_2)$ and let U be the midpoint of V and W. Then $\triangle VPW$ has PU as a median. By Apollonius's theorem (Problem 10.6), which is simply the median case of Stewart's theorem (Theorem 10.5), along with the definition of a DEF-construction (Definition 12.1),

$$\begin{aligned}
&4PU^2 + VW^2 = 2PV^2 + 2PW^2 \\
&4 \cdot \frac{\left|a\left(\frac{x_1+x_2}{2}\right) + b\left(\frac{x_1+x_2}{2}\right) + c\right|^2}{a^2+b^2} + (x_1 - x_2)^2 + (y_1 - y_2)^2 \\
&\quad = 2 \cdot \frac{|ax_1 + by_1 + c|^2}{a^2+b^2} + 2 \cdot \frac{|ax_2 + by_2 + c|^2}{a^2+b^2}.
\end{aligned}$$

Clearing the denominator $a^2 + b^2$ makes this

$$\begin{aligned}
&[a(x_1 + x_2) + b(y_1 + y_2) + 2c]^2 + (a^2 + b^2)[(x_1 - x_2)^2 + (y_1 - y_2)^2] \\
&\quad = 2(ax_1 + by_1 + c)^2 + 2(ax_2 + by_2 + c)^2.
\end{aligned}$$

Rearranging the left side and expanding it gives

$$\begin{aligned}
&[(ax_1 + by_1 + c) + (ax_2 + by_2 + c)]^2 + (a^2 + b^2)[(x_1 - x_2)^2 + (y_1 - y_2)^2] \\
&\quad = 2(ax_1 + by_1 + c)^2 + 2(ax_2 + by_2 + c)^2 \\
&(ax_1 + by_1 + c)^2 + 2(ax_1 + by_1 + c)(ax_2 + by_2 + c) + (ax_2 + by_2 + c)^2 \\
&\quad + (a^2 + b^2)[(x_1 - x_2)^2 + (y_1 - y_2)^2] \\
&\quad = 2(ax_1 + by_1 + c)^2 + 2(ax_2 + by_2 + c)^2.
\end{aligned}$$

Cancelling common terms from both sides and completing a square, we get

$$\begin{aligned}
&(a^2 + b^2)[(x_1 - x_2)^2 + (y_1 - y_2)^2 = [(ax_1 + by_1 + c) - (ax_2 + by_2 + c)]^2 \\
&(a^2 + b^2)[(x_1 - x_2)^2 + (y_1 - y_2)^2] = [a(x_1 - x_2) + b(y_1 - y_2)]^2.
\end{aligned}$$

By the real number Cauchy-Schwarz inequality from Volume 1,

$$(a^2 + b^2)[(x_1 - x_2)^2 + (y_1 - y_2)^2] \geq [a(x_1 - x_2) + b(y_1 - y_2)]^2,$$

where equality holds if and only if $(a, b) = (0, 0)$ (which is impossible since $ax + by + c$ is a line), or (a, b) and $(x_1 - x_2, y_1 - y_2)$ are linearly dependent. Since (a, b) is a normal vector to the directrix $ax + by + c = 0$ and $(x_1 - x_2, y_1 - y_2)$ is a direction vector of the line though VW, the line through VW must be perpendicular to the stated directrix. Subsequently, the line through VW intersects the directrix, so let the point of intersection be R. But then the eccentricity times the shortest distance between R and the directrix must be 0, which is equal to the distance between R and the focus. So R is the focus, which makes the focus lie on the directrix. This is a contradiction to the definition of a DEF-construction.

Finally, if it is possible to cause a decomposition

$$Q(x; y) = L_1(x; y)L_2(x; y);$$

then $Q(x; y) = 0$ if and only if $L_1(x; y) = 0$ or $L_2(x; y) = 0$, which would cause the zero set of Q to contain both lines. As we saw above, the zero set of a constructible conic cannot contain a line, so Q cannot split multiplicatively. □

Theorem 12.14. Suppose S is a constructible conic (which therefore has infinitely many points, and so is non-degenerate). If Q_1 is a bivariate quadratic whose zero set is S, then the same is true for Q_1 for any real $\neq 0;$ which we call a scalar multiple of $Q_1:$ Conversely, if Q_1 and Q_2 are bivariate quadratics whose zero sets are both $S;$ then there exists a real $\neq 0$ such that $Q_2 = Q_1:$ More explicitly, this is saying that, given

$$Q_1(x; y) = A_1x^2 + B_1xy + C_1y^2 + D_1x + E_1y + F_1;$$
$$Q_2(x; y) = A_2x^2 + B_2xy + C_2y^2 + D_2x + E_2y + F_2;$$

these two bivariate quadratics have the same constructible conic as their zero set if and only if there exists a real $\neq 0$ such that

$$(A_1; B_1; C_1; D_1; E_1; F_1) = (A_2; B_2; C_2; D_2; E_2; F_2):$$

Proof. The first direction is easy because a point $(x; y)$ satisfies $Q_1(x; y) = 0$ if and only if it satisfies $Q_2(x; y) = Q_1(x; y) = 0$, given $\neq 0$. So the two zero sets are the same.

The second direction is more challenging, and it is the goal towards which we have been building with the preceding lemmas. Since $Q_1(x; y)$ arises from a constructible conic, Lemma 12.13 says that it cannot be decomposed into the product of two linear functions. By the contrapositive of Lemma 12.12, every line intersects $Q_1(x; y) = 0$ in at most 2 points.

Since $Q_1(x; y) = 0$ is a constructible conic, it has infinitely many points. From the curve, we pick 5 distinct points $P_i = (x_i; y_i)$ for $i = 1; 2; 3; 4; 5$. The zero sets of $Q_1(x; y) = 0$ and $Q_2(x; y) = 0$ are assumed to be the same, so the P_i also lie on the latter. Then we obtain the following system of equations, represented using matrix multiplication:

$$\begin{bmatrix} x_1^2 & x_1y_1 & y_1^2 & x_1 & y_1 & 1 \\ x_2^2 & x_2y_2 & y_2^2 & x_2 & y_2 & 1 \\ x_3^2 & x_3y_3 & y_3^2 & x_3 & y_3 & 1 \\ x_4^2 & x_4y_4 & y_4^2 & x_4 & y_4 & 1 \\ x_5^2 & x_5y_5 & y_5^2 & x_5 & y_5 & 1 \end{bmatrix} \begin{bmatrix} A_2 \\ B_2 \\ C_2 \\ D_2 \\ E_2 \\ F_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}:$$

Since the five $(x_i; y_i)$ coordinate pairs are known, each of the five rows represents a linear equation in 6 variables $(A_2; B_2; C_2; D_2; E_2; F_2)$. We wish to show that all such 6-tuples are determined as non-zero scalar multiples of $(A_1; B_1; C_1; D_1; E_1; F_1)$. With this goal in mind, if we can show that the rows are linearly independent, then the rows will span a 5-dimensional subspace of the 6-dimensional Euclidean vector space. The set of solutions for $(A_2; B_2; C_2; D_2; E_2; F_2)$ forms the "null space" of this system, meaning it is the set of 6-component real vectors that are individually sent to 0 by each row of the matrix. By the well-known rank-nullity theorem (which we have not covered) from linear algebra, this null space has dimension $6 - 5 = 1$. So, all $(A_2, B_2, C_2, D_2, E_2, F_2)$ are scalar multiples of each other, and, in particular, they are generated as the set of scalar multiples of the known tuple $(A_1, B_1, C_1, D_1, E_1, F_1)$.

As a result of the preceding argument, our penultimate step should be to prove that the rows of the matrix are linearly independent. For the sake of contradiction, suppose that, without loss of generality, the fifth row is a linear combination of the first four rows. Let the row vectors be r_1, r_2, r_3, r_4, r_5 and let a hypothetical solution to the system be

$$s = (A_2, B_2, C_2, D_2, E_2, F_2).$$

So, there exist real numbers $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ such that

$$r_5 = \lambda_1 r_1 + \lambda_2 r_2 + \lambda_3 r_3 + \lambda_4 r_4,$$

Taking the dot product of both sides by s, we get

$$\begin{aligned} r_5 \cdot s &= (\lambda_1 r_1 + \lambda_2 r_2 + \lambda_3 r_3 + \lambda_4 r_4) \cdot s \\ &= \lambda_1 (r_1 \cdot s) + \lambda_2 (r_2 \cdot s) + \lambda_3 (r_3 \cdot s) + \lambda_4 (r_4 \cdot s) \\ &= 0. \end{aligned}$$

This means that, for any $s = (A_2, B_2, C_2, D_2, E_2, F_2)$ such that

$$Q_2(x, y) = A_2 x^2 + B_2 xy + C_2 y^2 + D_2 x + E_2 y + F_2 = 0$$

for each $(x, y) = (x_i, y_i)$ $(i = 1, 2, 3, 4)$, it is also true that (x_5, y_5) lies on the conic $Q_2(x, y) = 0$ by satisfying the same equation.

Let $a_{12}x + b_{12}y + c_{12} = 0$ be the line that runs through (x_1, y_1) and (x_2, y_2), and let $a_{34}x + b_{34}y + c_{34} = 0$ be the line that runs through (x_3, y_3) and (x_4, y_4). Then expanding and collecting the like terms of

$$(a_{12}x + b_{12}y + c_{12})(a_{34}x + b_{34}y + c_{34}) = 0$$

yields a bivariate quadratic that that goes through (x_i, y_i) for $i = 1, 2, 3, 4$. By the above argument, (x_5, y_5) also lies on it, so at least one of the following equations is true:

$$\begin{aligned} a_{12}x_5 + b_{12}y_5 + c_{34} &= 0, \\ a_{34}x_5 + b_{34}y_5 + c_{34} &= 0. \end{aligned}$$

This would imply that three collinear points, that is (x_i, y_i) for $i = 1, 2, 5$ or $i = 3, 4, 5$ lie in the zero set of $Q_1(x, y) = 0$. Finally, this contradicts the fact that we derived at the beginning of the proof: every line intersects $Q_1(x, y) = 0$ in at most 2 points. □

Problem 12.15. In a bivariate quadratic that represents a constructible conic, prove that at least one of the coefficients of the quadratic terms x^2, xy, y^2 is non-zero.

So far, we know that every constructible conic can be transformed by a Euclidean isometry to be the zero set of a bivariate quadratic in standard form (Theorem 12.8). We also know that its DEF-constructions get mapped to those of the image conic and without changing the eccentricity or focal parameter (Theorem 12.7). We found that there is *at least* one DEF-construction for parabolas, and *at least* two DEF-constructions for ellipses and hyperbolas (Theorem 12.5). If we can deduce that there is also *at most* one DEF-construction for a parabola and *at most* two for ellipses and hyperbolas (it suffices to look at standard forms), then we will know that there are *exactly* as many DEF-constructions. When the eccentricity and focal parameter turn out to be the same in every DEF-construction for a constructible conic, we will finally be able to refer to “the” eccentricity and “the” focal parameter of a given conic.

Theorem 12.16. A parabola has a unique eccentricity, and exactly one focus-directrix pair. As a result, it has exactly one possible focal parameter.

Proof. Due to preservations under isometry, it suffices to look at the standard form

$$0x^2 + 0xy + 1y^2 + (-4p)x + 0y + 0 = 0$$

for $p > 0$. With the “at least” direction established in Theorem 12.5, we will show that at most one DEF-construction exists. Suppose there is a DEF-construction with the directrix $ax + by + c = 0$ where $a^2 + b^2 \neq 0$, eccentricity $e > 0$, and focus (x_0, y_0). By Theorem 12.14, we can scale these coefficients to be equal to those in Theorem 12.3:

$$\begin{bmatrix} 0 \\ 0 \\ 1 \\ -4p \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} A \\ B \\ C \\ D \\ E \\ F \end{bmatrix} = \begin{bmatrix} a^2(1-e^2) + b^2, \\ 2abe^2, \\ a^2 + (1-e^2)b^2, \\ 2(ace^2 + (a^2+b^2)x_0) \\ 2(bce^2 + (a^2+b^2)y_0) \\ (a^2+b^2)(x_0^2+y_0^2) - e^2c^2 \end{bmatrix}.$$

The second equation gives us $-2abe^2 = 0$, which is equivalent to $a = 0$ or $b = 0$ since we require $e > 0$. The first equation says

$$a^2(1-e^2) + b^2 = 0.$$

So, in the hypothetical case that $a = 0$, we would also get $b = 0$, contradicting $a^2+b^2 \neq 0$. We have deduced $b = 0$. Moreover, since (a, b, c) can be scaled without changing the directrix, we may choose $a = 1$ without loss of generality. Then the third equation becomes

$$= a^2 + (1-e^2)b^2 = 1.$$

The first equation yields

$$0 = a^2(1-e^2) + b^2 = 1 - e^2 \implies e = 1.$$

We update the fourth, fifth, and sixth equations to say

$$\begin{bmatrix} -4p \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} D \\ E \\ F \end{bmatrix} = \begin{bmatrix} 2(c + x_0) \\ 2y_0 \\ x_0^2 + y_0^2 - c^2 \end{bmatrix}.$$

Then $y_0 = 0$ and we can further simplify the system to

$$\begin{aligned} 2p &= c + x_0, \\ 0 &= x_0^2 - c^2 = (x_0 - c)(x_0 + c) = (x_0 - c)2p. \end{aligned}$$

The last equation yields $x_0 = c$, since $p > 0$. Combining this with $2p = c + x_0$ gives $c = x_0 = p$. Therefore, there is at most one DEF-construction, which is

$$((a : b : c), e, (x_0, y_0)) = ((1 : 0 : p), 1, (p, 0)),$$

where the colons indicate that multiplying all the components of $(a : b : c)$ by a non-zero scalar does not change the ratio. □

Theorem 12.17. An ellipse has a unique eccentricity, and exactly two focus-directrix pairs, each with the same focal parameter.

Proof. Due to preservations under isometry, it suffices to look at the standard form

$$\frac{1}{p^2}x^2 + 0xy + \frac{1}{q^2}y^2 + 0p + 0q + (-1) = 0$$

for $p > 0$ and $q > 0$. Since the "at least" direction has already been established in Theorem 12.5, we will show that at most two DEF-constructions exist, which will turn out to have equal eccentricities and equal focal parameters. Suppose there is a DEF-construction with the directrix $ax + by + c = 0$ where $a^2 + b^2 \neq 0$, eccentricity $e > 0$, and focus (x_0, y_0). By Theorem 12.14, we can scale these coefficients to be equal to those in Theorem 12.3:

$$\begin{bmatrix} 1/p^2 \\ 0 \\ 1/q^2 \\ 0 \\ 0 \\ -1 \end{bmatrix} = \begin{bmatrix} A \\ B \\ C \\ D \\ E \\ F \end{bmatrix} = \begin{bmatrix} a^2(1-e^2) + b^2, \\ -2abe^2, \\ a^2 + (1-e^2)b^2, \\ -2(ace^2 + (a^2+b^2)x_0) \\ -2(bce^2 + (a^2+b^2)y_0) \\ (a^2+b^2)(x_0^2+y_0^2) - e^2c^2 \end{bmatrix}.$$

Suppose, for contradiction that $p = q$. Then subtracting the third equation from the first yields

$$0 = \frac{1}{p^2} - \frac{1}{q^2} = -e^2(a^2+b^2) \implies e = 0,$$

since $a^2 + b^2 \neq 0$. This value of e is not allowed in a DEF-construction, so we know that $p \neq q$. Before we do casework on $q > p$ and $p > q$, note that the second equation $0 = -2abe^2$ leads to $a = 0$ or $b = 0$, which is a key deduction. In fact, since $(a : b : c)$ can be scaled without changing the directrix and $a^2 + b^2 \neq 0$, we can choose $a = 1$ if $b = 0$, or $b = 1$ if $a = 0$.

- Case 1: Suppose $q > p$. Suppose, for contradiction, that $b = 0$ and we take $a = 1$. Then dividing the first equation by the third yields

$$1 < \frac{q^2}{p^2} = -\frac{a^2(1-e^2)+b^2}{a^2+(1-e^2)b^2} = \frac{1-e^2}{1} \implies e^2 < 0,$$

which is a contradiction. So, $a = 0$ and we take $b = 1$. Then, the first equation becomes

$$\frac{\quad}{p^2} = a^2(1 - e^2) + b^2 = 1 \implies \quad = p^2.$$

The third equation yields

$$\frac{p^2}{q^2} = \frac{\quad}{q^2} = a^2 + (1 - e^2)b^2 = 1 - e^2 \implies e = \sqrt{1 - \frac{p^2}{q^2}},$$

which exists in the real numbers because $q > p$. We update the fourth, fifth, and sixth equations to say

$$\begin{bmatrix} 0 \\ 0 \\ -p^2 \end{bmatrix} = \begin{bmatrix} D \\ E \\ F \end{bmatrix} = \begin{bmatrix} -2x_0 \\ -2(ce^2 + y_0) \\ x_0^2 + y_0^2 - e^2c^2 \end{bmatrix}.$$

The first of these equations immediately gives $x_0 = 0$. The second equation gives

$$ce = -\frac{y_0}{e} \implies c^2e^2 = \frac{y_0^2}{e^2},$$

which can be substituted into the third equation to give

$$-p^2 = y_0^2 - e^2c^2 = y_0^2 - \frac{y_0^2}{e^2} = y_0^2\left(1 - \frac{1}{e^2}\right).$$

Substituting $e = \sqrt{1 - \frac{p^2}{q^2}}$ into this and algebraic manipulations yield the isolated solutions

$$y_0 = \pm\sqrt{q^2 - p^2}.$$

Finally, we can compute

$$c = -\frac{y_0}{e^2} = \mp\frac{q^2}{\sqrt{q^2 - p^2}}.$$

Therefore, there are at most two DEF-constructions, which are

$$((a : b : c); e; (x_0; y_0)) = \left(\left(0 : 1 : \mp\frac{q^2}{\sqrt{q^2 - p^2}}\right); \sqrt{1 - \frac{p^2}{q^2}}; \left(0; \pm\sqrt{q^2 - p^2}\right)\right).$$

The eccentricity is the same in both cases, and the focal parameter is equal to

$$\frac{q^2}{\sqrt{q^2 - p^2}} - \sqrt{q^2 - p^2} = \frac{p^2}{\sqrt{q^2 - p^2}}$$

in both cases, which were easy to compute because the directrices are horizontal.

- Case 2: Suppose $p > q$. In an argument that is symmetrical to that for Case 2, we find that

$$((a : b : c); e; (x_0; y_0)) = \left(\left(1 : 0 : \mp\frac{p^2}{\sqrt{p^2 - q^2}}\right); \sqrt{1 - \frac{q^2}{p^2}}; \left(0; \pm\sqrt{p^2 - q^2}\right)\right).$$

Again, the eccentricity is the same in both cases, and the focal parameter is equal to

$$\frac{p^2}{\sqrt{p^2 - q^2}} - \sqrt{p^2 - q^2} = \frac{q^2}{\sqrt{p^2 - q^2}}$$

in both cases, which were easy to compute because the directrices are vertical.

□

Theorem 12.18. A hyperbola has a unique eccentricity, and exactly two focus-directrix pairs, each with the same focal parameter.

Proof. Due to preservations under isometry, it suffices to look at the standard form

$$\frac{1}{p^2}x^2 + 0xy + \left(-\frac{1}{q^2}\right)y^2 + 0p + 0q + (-1) = 0$$

for $p > 0$ and $q > 0$. As we have already established the "at least" direction in Theorem 12.5, we will show that at most two DEF-constructions exist, which will turn out to have equal eccentricities and equal focal parameters. Suppose there is a DEF-construction with the directrix $ax + by + c = 0$ where $a^2 + b^2 \neq 0$, eccentricity $e > 0$, and focus (x_0, y_0). By Theorem 12.14, we can scale these coefficients to be equal to those in Theorem 12.3:

$$\begin{bmatrix} 1/p^2 \\ 0 \\ -1/q^2 \\ 0 \\ 0 \\ -1 \end{bmatrix} = \begin{bmatrix} A \\ B \\ C \\ D \\ E \\ F \end{bmatrix} = \begin{bmatrix} a^2(1-e^2) + b^2, \\ -2abe^2, \\ a^2 + (1-e^2)b^2, \\ -2(ace^2 + (a^2+b^2)x_0) \\ -2(bce^2 + (a^2+b^2)y_0) \\ (a^2+b^2)(x_0^2 + y_0^2) - e^2c^2 \end{bmatrix}.$$

The second equation $-2abe^2 = 0$ implies $a = 0$ or $b = 0$. We investigate these two cases separately:

- Case 1: Suppose $a = 0$. Since $a^2 + b^2 \neq 0$ and $(a : b : c)$ may be rescaled without changing the directrix, we may choose $b = 1$. Then the first equation gives

$$\frac{1}{p^2} = a^2(1-e^2) + b^2 = 1 \implies 1 = p^2.$$

  This leads to the third equation yielding

$$-\frac{p^2}{q^2} = a^2 + b^2(1-e^2) = 1 - e^2 \implies e = \sqrt{1 + \frac{p^2}{q^2}}.$$

  We update the fourth, fifth, and sixth equations to say

$$\begin{bmatrix} 0 \\ 0 \\ -p^2 \end{bmatrix} = \begin{bmatrix} D \\ E \\ F \end{bmatrix} = \begin{bmatrix} -2x_0 \\ -2(ce^2 + y_0) \\ x_0^2 + y_0^2 - e^2c^2 \end{bmatrix}.$$

As with the first case for ellipses, the first of these equations immediately gives $x_0 = 0$, and the second equation gives

$$ce = -\frac{y_0}{e} \implies c^2e^2 = \frac{y_0^2}{e^2}.$$

This can be substituted into the third equation to give

$$p^2 = y_0^2 - e^2c^2 = y_0^2 - \frac{y_0^2}{e^2} = y_0^2\left(1 - \frac{1}{e^2}\right).$$

Substituting $e = \sqrt{1 + \frac{p^2}{q^2}}$ into this and algebraically manipulating it yields

$$y_0^2 = -(q^2 + p^2) < 0.$$

This is a contradiction, meaning Case 1 (where the directrix is horizontal due to $a = 0$) is empty for the standard form.

- Case 2: Suppose $b = 0$. Since $a^2 + b^2 \neq 0$ and $(a : b : c)$ may be rescaled without changing the directrix, we may choose $a = 1$. Then the third equation gives

$$\frac{\lambda}{q^2} = a^2 + b^2(1 - e^2) = 1 \implies \lambda = q^2.$$

Then the first equation yields

$$\frac{q^2}{p^2} = a^2(1 - e^2) + b^2 = 1 - e^2 \implies e = \sqrt{1 + \frac{q^2}{p^2}}.$$

We update the fourth, fifth, and sixth equations to say

$$\begin{bmatrix} 0 \\ 0 \\ -q^2 \end{bmatrix} = \begin{bmatrix} D \\ E \\ F \end{bmatrix} = \begin{bmatrix} 2(ce^2 + x_0) \\ 2y_0 \\ x_0^2 + y_0^2 - e^2c^2 \end{bmatrix}.$$

The first of these equations gives $y_0 = 0$, and the second equation can be contorted into

$$ce = -\frac{x_0}{e} \implies c^2e^2 = \frac{x_0^2}{e^2}.$$

This can be substituted into the third equation to give

$$-q^2 = x_0^2 - e^2c^2 = x_0^2 - \frac{x_0^2}{e^2} = x_0^2\left(1 - \frac{1}{e^2}\right).$$

Substituting $e = \sqrt{1 + \frac{q^2}{p^2}}$ into this and algebraically manipulating it yields

$$x_0^2 = q^2 + p^2 \implies x_0 = \pm\sqrt{p^2 + q^2}.$$

Finally, we compute

$$c = \frac{x_0}{e^2} = \frac{p^2}{\sqrt{p^2+q^2}}.$$

Therefore, there are at most two DEF-constructions, which are

$$((a:b:c); e; (x_0, y_0)) = \left(\left[1 : 0 : \pm\frac{p^2}{\sqrt{p^2+q^2}}\right]; \sqrt{1+\frac{q^2}{p^2}}; \left(\pm\sqrt{p^2+q^2}, 0\right)\right).$$

The eccentricity is the same in both cases, and the focal parameter is equal to

$$\sqrt{p^2+q^2} - \frac{p^2}{\sqrt{p^2+q^2}} = \frac{q^2}{\sqrt{p^2+q^2}}$$

in both cases, which were easy to compute because the directrices are vertical.

□

Corollary 12.19. For a given constructible conic, the eccentricity and focal parameter are equal in every DEF-construction that leads to the conic. In other words, these parameters are inherent to the curve that is the conic and does not depend on how the curve is constructed. Thus, we can refer to “the” eccentricity and “the” focal parameter of a conic.

Problem 12.20. The discriminant of a bivariate quadratic

$$Q(x, y) = Ax^2 + Bxy + Cy^2 + Dx + Ey + F$$

is $\Delta(Q) = B^2 - 4AC.$ Suppose S is a conic arising from a specific DEF-construction with eccentricity e and that Q is a bivariate quadratic whose zero set is $S.$ Prove that

$$\operatorname{sgn}(\Delta(Q)) = \operatorname{sgn}(e - 1).$$

Thus, the sign of $\Delta(Q)$ determines the type of the conic.

12.2 Metric Properties

Problem 12.21. For an ellipse, parabola or hyperbola in standard form, determine, in terms of a and $b,$ the set of x values that have a corresponding y value such that (x, y) satisfies the equation; for each such x value, determine the number of y values that correspond to this x value. Repeat the exercise with the roles of x and y interchanged.

Definition 12.22. Equipped with a complete classification of the focus-directrix pairs of each type of conic, we can define the following geometric components of a conic:

1. The **principal axis** of a conic is the line through a focus that is perpendicular to the directrix corresponding to that focus. For an ellipse or hyperbola, the principal axis runs through both foci and is perpendicular to both directrices. For a parabola, it is called the **axis of symmetry**. The principal axis of a conic is a line of symmetry in the sense that reflecting the conic across the line yields the same set of points.

2. Ellipses and hyperbolas have a **secondary line of symmetry** that is perpendicular to the axis of symmetry and runs through the midpoint of the segment connecting the foci. This is a name that we have coined for this book, as the line is unnamed in the literature to the best of our knowledge. Reflecting an ellipse or hyperbola across its secondary line of symmetry yields the same set of points. Note that it is parallel to the directrices.

3. The two regions of a hyperbola are called its **branches**. For a hyperbola that is the graph of an equation in standard form, the two branches are the points $(x;y)$ on the hyperbola such that x p and the points $(x;y)$ on the hyperbola such that x $p:$ For any hyperbola, there is one branch and one focus-directrix pair on each half-plane defined by the secondary axis of symmetry. Points in each branch are strictly closer to the focus in its half-plane than the focus in the other half-plane.

4. The **center** of an ellipse or hyperbola is the point of intersection of its principal axis and its secondary line of symmetry. This happens to be the midpoint of the segment joining the two foci. The center is a point of symmetry in the sense that reflecting the conic across it yields the same set of points.

5. The **linear eccentricity** of an ellipse or hyperbola is the distance between the center and a focus. This distance is the same for either of the two choices of a focus in an ellipse or hyperbola.

6. A **chord** of a conic is a line segment whose endpoints are two distinct points on the conic. A **latus rectum** of a conic is, for a particular focus-directrix pair, the chord that contains the focus and is parallel to the directrix. The length of such a chord is the same regardless of the focus-directrix pair chosen. A **semi-latus rectum** is either of the two segments resulting from splitting a latus rectum at the focus that is its midpoint.

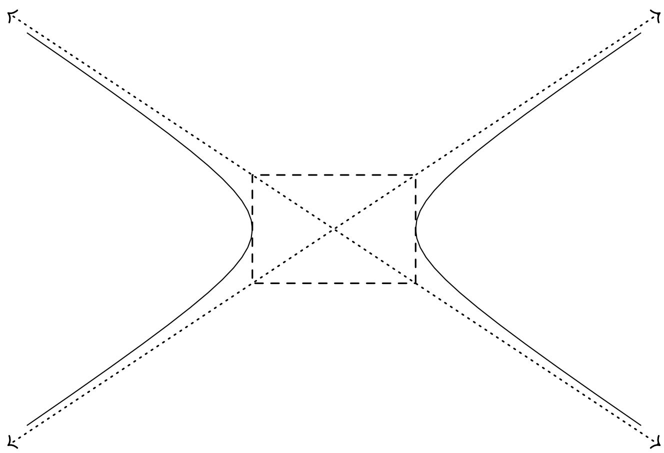
7. The points at which the principal axis of a conic intersects the conic are called the **vertices** of the conic. Parabolas have one vertex, whereas ellipses and hyperbolas have two vertices. Note that a parabola in standard form has its vertex as the origin, which is where its extreme x-coordinate occurs. The vertices of an ellipse in standard form occupy the two extreme x-coordinates if $p > q$ or the two extreme y-coordinates if $p < q$, and the vertices of a hyperbola in standard form occupy the extreme x-coordinates of each branch.

8. For an ellipse or hyperbola, the unique chord lying on the principal axis is called the **major axis** in an ellipse and the **transverse axis** in a hyperbola. The endpoints of this chord happen to be the two vertices. This is the longest chord of an ellipse and the shortest chord connecting the two branches of a hyperbola. A **semi-major axis** is either of the two segments resulting from splitting the major axis at its midpoint, which is the center of the conic.

9. The **minor axis** of an ellipse is the unique chord lying on the secondary line of symmetry. This is the shortest chord of the ellipse. A **semi-minor axis** is either of the two

segments resulting from splitting the minor axis at its midpoint, which is the center of the conic. The analogous idea for a hyperbola is the **conjugate axis**, which is the segment perpendicular to the principal axis whose midpoint is the center of the hyperbola, and whose length is $\frac{2fe}{\sqrt{e^2-1}}$ where f is the focal parameter and e is the eccentricity. Though it is not a chord of the hyperbola, it does have significance in the context of the asymptotes of a hyperbola, which is the next definition.

10. The **asymptotes** of a hyperbola are the two lines that run through the diagonals of the rectangle that is defined by having the transverse and conjugate axes as its medians. Both asymptotes come arbitrarily close to both branches of the hyperbola as $x \to \infty$ and $x \to -\infty$ but the asymptotes do not intersect the hyperbola. In standard form, the axes are $y = \pm\frac{q}{p}x$.



We will not prove that these notions are well-defined or that the various assertions made about them are true, but the reader is encouraged to pursue the proofs as they are not difficult for the most part. All of these objects and lengths have the expected preservation properties under Euclidean isometries. As such, it suffices to consider only the standard forms for the proofs and then apply a suitable arbitrary Euclidean isometry to them.

Problem 12.23. A **rectangular hyperbola** is a conic whose eccentricity is $\sqrt{2}$.

1. Show that a hyperbola is rectangular if and only if its asymptotes are perpendicular.

2. Show that every rectangular hyperbola in standard form has an equation of the form $x^2 - y^2 = p^2$.

3. Determine an equation of the rectangular hyperbola resulting from rotating a rectangular hyperbola in standard form counterclockwise by $\frac{\pi}{4}$ around the origin.

Problem 12.24. Determine an equation whose graph is the resulting of rotating the graph of a conic in standard form by $\frac{\pi}{2}$, π, or $\frac{3\pi}{2}$ counterclockwise around the origin. Work with each type of conic separately.

Theorem 12.25. Let S be a conic and let its eccentricity be e and focal parameter be f: Then the following relations hold.

1. Suppose S is the graph of an equation in standard form. If S is an ellipse then the length of the semi-major axis is p and the length of the semi-minor axis is q. If S is a hyperbola then the length of the semi-transverse axis is a and the length of the semi-conjugate axis is b:

2. If the length of the semi-latus rectum is l then $e = \frac{l}{f}$: This is true for all parabolas, ellipses and hyperbolas.

3. If S is an ellipse or hyperbola, let the linear eccentricity be r; the length of the semi-major axis or semi-transverse axis be p, and the length of the semi-minor axis or semi-conjugate axis be q: Then
$$e = \frac{r}{p} = \frac{p}{f + r} = \frac{f + \sqrt{f^2 + 4p^2}}{2p}:$$

4. In the above notation, $p^2 - q^2 = r^2$ if the conic is an ellipse, and $p^2 + q^2 = r^2$ if the conic is a hyperbola.

Proof. All of these objects and lengths are preserved under Euclidean isometries so it suﬃces to prove the relations for only conics that are the graphs of equations in standard form. This is helpful because all the lengths involved are then parallel or perpendicular to the coordinate axes, making their computations easier. The results all follow from expressing the quantities in terms of the coeﬃcients p and q in the standard form. We chose $p > q$ for ellipses.

	Ellipse	Parabola	Hyperbola
Eccentricity	$\sqrt{1 - \left(\frac{q}{p}\right)^2}$	1	$\sqrt{1 + \left(\frac{q}{p}\right)^2}$
Linear eccentricity	$\sqrt{p^2 - q^2}$	Undefined	$\sqrt{p^2 + q^2}$
Focal parameter	$\frac{q^2}{\sqrt{p^2 - q^2}}$	$2p$	$\frac{q^2}{\sqrt{p^2 + q^2}}$
Semi-major axis Semi-transverse axis	p	Undefined	p
Semi-minor axis Semi-conjugate axis	q	Undefined	q
Semi-latus rectum	$\frac{q^2}{p}$	$2p$	$\frac{q^2}{p}$

We leave it to the reader to compute the entries of this table and verify the stated relations, but we note that it will be helpful to remember the coordinates of the foci and the equations of the corresponding directrices for conics in standard forms. □

Unlike most elementary texts, we have not used the following characterizations of ellipses and hyperbolas as our primary definition because DEF-constructions provide a more unified framework. They are still interesting properties to work out.

Example 12.26. Given two distinct points F_1 and F_2 and a real number $d > F_1F_2$, let S be the set of points P such that $PF_1 + PF_2 = d$. Show that there exists an ellipse E such that $S \subseteq E$, and that F_1 and F_2 are the foci of E and that d is the length of the major axis of E. On the other hand, show that if E' is some ellipse with foci F_1' and F_2', then there exists a real number $d' > F_1'F_2'$ such that for all points P' on E', $P'F_1' + P'F_2' = d'$. As a result, $E \subseteq S$ as well, proving that $S = E$.

Solution. We apply a Euclidean isometry to the configuration so that F_1 and F_2 are opposite points on the x-axis. Let $F_1 = (r, 0)$ and $F_2 = (-r, 0)$ for some $r > 0$. Since we expect d to be the length of the major axis, we define p so that $d = 2p$. Then we are seeking all points $P = (x, y)$ such that

$$\sqrt{(x+r)^2 + y^2} + \sqrt{(x-r)^2 + y^2} = 2p.$$

Now, we perform a sequence of manipulations, the reversibility of which do not matter since we only want to establish a subset relation for the time being:

$$\begin{aligned} \sqrt{(x+r)^2 + y^2} &= 2p - \sqrt{(x-r)^2 + y^2} \\ (x+r)^2 + y^2 &= 4p^2 + (x-r)^2 + y^2 - 4a\sqrt{(x-r)^2 + y^2} \\ p^2 - rx &= p\sqrt{(x-r)^2 + y^2} \\ (p^2 - rx)^2 &= p^2((x-r)^2 + y^2) \\ p^2(p^2 - r^2) &= x^2(p^2 - r^2) + p^2y^2 \\ 1 &= \frac{x^2}{p^2} + \frac{y^2}{p^2 - r^2}. \end{aligned}$$

Since $p = \frac{d}{2} > \frac{F_1F_2}{2} = r$, we can define $q > 0$ to satisfy $q^2 = p^2 - r^2$, which means that S is a subset of an ellipse E in standard form. Due to the relation $r^2 = p^2 - q^2$ being satisfied, r is the linear eccentricity, and so F_1 and F_2 are the foci. Of course, $d = 2a$ is the length of the major axis.

For the second part, let the eccentricity of E' be e and let P' be a point on E'. Let the directrices of E' be d_1 and d_2. By using the focus-directrix definition of an ellipse and the fact that the two directrices are parallel, we get

$$P'F_1' + P'F_2' = e \cdot d(P', d_1) + e \cdot d(P', d_2) = e \cdot d(d_1, d_2).$$

Using the equations of the directrices in standard form, the distance between the directrices is

$$e \cdot d(d_1, d_2) = e\left(\frac{p}{e} - \left(-\frac{p}{e}\right)\right) = 2p,$$

which is the desired constant $d' > F_1'F_2'$, since the segment between the foci is strictly contained within the major axis in an ellipse. (Note that here p is the symbol used in the standard form of an ellipse.)
Finally, going back to the first part, S is easily shown to be non-empty, say by showing that either of the two vertices of E satisfies the condition required to lie in S. Thus, the second part implies that $E \subseteq S$, proving $S = E$. □

Problem 12.27. Given two distinct points F_1 and F_2 and a real number $F_1F_2 > d > 0$, let S be the set of points P such that $|PF_1 - PF_2| = d$. Show that there exists a hyperbola H such that $S \subseteq H$, and that F_1 and F_2 are the foci of H and that d is the length of the transverse axis of H. On the other hand, show that if H' is some hyperbola with foci F_1' and F_2', then there exists a positive real number $d' < F_1'F_2'$ such that for all points P' on H', $|P'F_1' - P'F_2'| = d$. More specifically, show that $P'F_2' - P'F_1' = d'$ if P' is on the branch closer to F_1' and $P'F_1' - P'F_2' = d'$ if P' is on the branch closer to F_2'. As a result, $H \subseteq S$ as well, proving that $S = H$. You may use the solution to Example 12.26 as a template.

There is a wonderful algebraic development of the theory of conics in [5]. The enterprising reader is encouraged to explore that book.

Chapter 13

Cross Product

"Plato said God geometrizes continually."
– *Plutarch, Convivialium disputationum*

The analogue in three-dimensional Cartesian space of a line in the two-dimensional Cartesian plane is a plane. The cross product of three-dimensional Euclidean vectors is a natural outgrowth of the study planes in three dimensions. We will study planes, their normal vectors, such as the cross product, and the algebraic and geometric properties of the cross product.

13.1 Formulas

Now we will define lines in higher dimensions by extending the definition of lines in $\mathbb{R}^2$ from Chapter 1 into a definition for $\mathbb{R}^n$.

Definition 13.1. Recalling Lemma 1.8, we define a **line** in $\mathbb{R}^n$ (as opposed to just $\mathbb{R}^2$) to be a set

$$\ell = \{p + tv : t \in \mathbb{R}\}$$

parametrized by the real variable $t,$ where $p \in \mathbb{R}^n$ is a point and $v \in \mathbb{R}^n$ is a non-zero point that controls the direction. The variable t is called a **parameter**. This form includes lines in every direction, including those that run through a coordinate axis.

Definition 13.2. If $(0,v)$ and $(0,w)$ are non-zero position vectors in $\mathbb{R}^n$ (here, 0 is the n-dimensional origin) such that $tv = w$ for some (necessarily non-zero) real $t,$ then we say that the vectors $[v]$ and $[w]$ are **parallel**. In this case, we also define that the displacement vectors in $[v]$ are **parallel** to the displacement vectors in $[w].$ By Theorem 1.36, lines in $\mathbb{R}^2$ that have position vectors v, w that are just scaled versions of each other are parallel or coincident. Subsequently, we define that two lines

$$\{p + tv : t \in \mathbb{R}\} \text{ and } \{q + tw : t \in \mathbb{R}\}$$

in $\mathbb{R}^n$ are **parallel** if v and w are parallel as vectors but the lines are not exactly the same set of points; if they are the same set of points, we use the word **coincident** instead of parallel.

We have been working on the 2-dimensional plane so far, which is like a flat piece of paper, with no height, that extends infinitely in all directions. In 3-dimensional space, there are infinitely many copies of the plane, as should be intuitively clear. This is clearly defined as follows.

Definition 13.3. Inspired by Problem 1.31, in Euclidean spaces of dimension $n \geq 3$, a plane is defined as a translation of the collection of all arrowheads of the set of position vectors

$$\{av + bw : a \in \mathbb{R}, b \in \mathbb{R}\},$$

where v and w are linearly independent position vectors. By a **translation**, we mean that we add the same point $z \in \mathbb{R}^3$ to all the arrowheads to get a subset of $\mathbb{R}^3$ that we then call a plane. By a **plane**, we might also refer to the collection of linear combinations of two linearly independent vectors, if we do not want to talk about a particular translation.

For the remainder of the chapter, we will work only on three dimensions. We begin our exploration of the cross product by asking a natural question: what can be said about a vector that is "perpendicular" to an entire plane in three dimensions?

Theorem 13.4. Suppose u is a vector that is orthogonal to two linearly independent vectors v and w. Then u is orthogonal to every vector in the plane spanned by v and w, and u is orthogonal to no other vectors.

Proof. If $\langle u, v\rangle = 0$ and $\langle u, w\rangle = 0$, then for any real a, b,

$$\langle u, av + bw\rangle = a\langle u, v\rangle + b\langle u, w\rangle = a \cdot 0 + b \cdot 0 = 0,$$

so u is orthogonal to every vector in the plane of v, w. Now suppose z is a vector that is orthogonal to u. We want to show that there exist real a, b such that $z = av + bw$. Working backwards, if such a, b exist, then

$$\langle z, v\rangle = \langle av + bw, v\rangle = a\langle v, v\rangle + b\langle w, v\rangle,$$
$$\langle z, w\rangle = \langle av + bw, w\rangle = a\langle v, w\rangle + b\langle w, w\rangle.$$

This system of equations can be written as the matrix equation

$$\begin{bmatrix} \langle v, v\rangle & \langle w, v\rangle \\ \langle v, w\rangle & \langle w, w\rangle \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \langle z, v\rangle \\ \langle z, w\rangle \end{bmatrix}.$$

The matrix on the far left has its determinant equal to

$$\|v\| \cdot \|w\| - \langle v, w\rangle^2,$$

which is strictly positive by the Cauchy-Schwarz inequality (Theorem 4.8) since v, w are linearly independent. Since the determinant is non-zero, we can multiply both sides by the inverse of the matrix and uniquely solve for a and b. We want to prove that $z = av + bw$. Unfortunately, the proof of which we are aware depends on a result from linear algebra, so we will provide only an outline of it: Let $x = z - (av + bw)$. Our aim is to prove that $x = 0$. By reversing our steps, we know that x is orthogonal to u, v, w, so x is orthogonal to three linearly independent vectors. All vectors in $\mathbb{R}^3$ can be generated by such three vectors, so x is orthogonal to all vectors in $\mathbb{R}^3$. By the weak cancellation rule (Theorem 4.5), $x = 0$ and so $z = av + bw$. The reader will be able to fill in the gaps in a course in linear algebra. □

Definition 13.5. Inspired by Theorem 13.4, we say that a vector v is **orthogonal** to a plane if v is orthogonal to two vectors that generate the plane with their linear combinations. A vector v that is orthogonal to a plane is called a **normalized** orthogonal vector if it is, well, normalized so that it has a magnitude of 1. The process of normalizing v is to divide v by $\|v\|$, since

$$\left\|\frac{v}{\|v\|}\right\| = \frac{\|v\|}{\|v\|} = 1.$$

Next comes the question of whether every plane has a normal vector. The answer is "yes" and such a vector is a key aspect of analyzing planes, especially their equations.

Theorem 13.6 (Orthogonal projection to a plane). Let v and w_1, w_2 be vectors in $\mathbb{R}^3$, where w_1, w_2 are linearly independent. Then there exist unique vectors u, z such that u is in the plane of w_1, w_2 and z is orthogonal to this plane and $v = u + z$. Letting the w_1, w_2 plane be W, we call u the **projection** of v *to* W and denote it by $\mathrm{proj}_W v$, and we call z the **rejection** of v *from* W and denote it by $\mathrm{oproj}_W v$.

Proof. Working backwards, suppose such u, z exist. In other words,

$$\langle z, w_1\rangle = \langle z, w_2\rangle = 0,$$

and there exist real a, b such that $u = aw_1 + bw_2$, and $v = u + z$. Then

$$\begin{aligned} 0 = \langle z, w_1\rangle = \langle v - u, w_1\rangle = \langle v - aw_1 - bw_2, w_1\rangle &= \langle v, w_1\rangle - a\langle w_1, w_1\rangle - b\langle w_2, w_1\rangle \\ \langle v, w_1\rangle &= a\langle w_1, w_1\rangle + b\langle w_2, w_1\rangle \end{aligned}$$

and similarly,

$$\begin{aligned} 0 = \langle z, w_2\rangle = \langle v - u, w_2\rangle = \langle v - aw_1 - bw_2, w_2\rangle &= \langle v, w_2\rangle - a\langle w_1, w_2\rangle - b\langle w_2, w_2\rangle \\ \langle v, w_2\rangle &= a\langle w_1, w_2\rangle + b\langle w_2, w_2\rangle. \end{aligned}$$

This system of two equations in two variables a, b may be written as the matrix equation

$$\begin{bmatrix} \langle w_1, w_1\rangle & \langle w_2, w_1\rangle \\ \langle w_1, w_2\rangle & \langle w_2, w_2\rangle \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \langle v, w_1\rangle \\ \langle v, w_2\rangle \end{bmatrix}.$$

Since w_1, w_2 are linearly independent, the Cauchy-Schwarz inequality (Theorem 4.8) implies that the determinant of the matrix on the far left is strictly positive, and so non-zero. So we can multiply both sides by the inverse of the matrix, which proves that, if a, b exist, then they are unique. That would make $u = aw_1 + bw_2$ unique and $z = v - u$ unique as well. We must show that these a, b actually work, meaning $v - aw_1 - bw_2$ is orthogonal to w_1 and w_2. Since our steps were reversible, we can go through them in the reverse order and find that

$$\langle v - aw_1 - bw_2, w_1\rangle = \langle v - aw_1 - bw_2, w_2\rangle = 0.$$

So we can set $u = aw_1 + bw_2$ and $z = v - u$, and these u, z will satisfy our conditions.

Here is a side note for those inclined to look further into linear algebra: the "basis" consisting of w_1 and w_2 for the plane that they generate would be called **orthonormal** if the norms of w_1 and w_2 are both 1 and the two vectors are orthogonal to each other. In that case, computing a, b is easy because

$$\begin{pmatrix} \langle w_1, w_1 \rangle & \langle w_2, w_1 \rangle \\ \langle w_1, w_2 \rangle & \langle w_2, w_2 \rangle \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \implies \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \langle v, w_1 \rangle \\ \langle v, w_2 \rangle \end{pmatrix}.$$

□

Corollary 13.7. The vector Pythagorean theorem (Theorem 4.14) for two dimensions extends to three dimensions by an identical proof. To be precise, if v is a vectors and W is a plane, let $u = \text{proj}_W v$ and let $z = \text{oproj}_W v$. Then

$$\|u\|^2 + \|z\|^2 = \|v\|^2.$$

Definition 13.8. In Theorem 13.6, if we can assume that v can be chosen "outside" the plane so that z is a non-zero vector (we have not yet proven that such a v exists), then every plane in $\mathbb{R}^3$ has a normal vector, meaning a non-zero vector that is orthogonal to every vector in the plane (and no other vectors by Theorem 13.4). Though we will not prove it, *one* normal vector can characterize a plane in $\mathbb{R}^3$ by itself and with a point that is on the plane, instead of us generating a plane using the linear combinations of *two* linearly independent vectors. So, while a plane in $\mathbb{R}^3$ was originally defined as a translation of the set of linear combinations of two linearly independent vectors with possibly a translation (Definition 13.3), it may equivalently be defined as the set that is orthogonal to some normal vector and runs through some given point.

Definition 13.9. Two planes in $\mathbb{R}^3$ are said to be:

1. Intersecting if there is at least once point that lies on both of them.
2. Coincident if they consist of the same set of points
3. Parallel if they do not intersect
4. Perpendicular if they have orthogonal normal vectors

There are statements that can be made about the various configurations of intersections of two or three lines and/or planes in space, but we will not look into them in detail here. For example, two planes intersect if and only if they have non-parallel normal vectors or the planes are coincident. The contrapositive of this biconditional statement is also useful: Two planes are parallel if and only if they have parallel normal vectors and the planes are not coincident.

Theorem 13.10 (Equation of a plane)**.** If $n = (a, b, c)$ is a non-zero vector and $p = (x_0, y_0, z_0)$ is a point in space, then the unique plane that contains p and that has n as a normal vector (we will assume that such a plane exists) is given by the set

$$\{(x, y, z) \in \mathbb{R}^3 : ax + by + cz + d = 0\},$$

where $d = -(ax_0 + by_0 + cz_0)$. In the other direction, if $(a, b, c) \neq 0$ and d is some real constant, then the set of points (x, y, z) that satisfy the equation $ax + by + cz + d = 0$ is a plane that has (a, b, c) as a normal vector. We call $ax + by + cz + d = 0$ a standard equation of the plane, much like the standard form of the equation of a line in the Cartesian plane.

Proof. The key idea is that a point (x, y, z) lies on the unique plane if and only if the vector

$$((x_0, y_0, z_0), (x, y, z)) \sim ((0, 0, 0), (x - x_0, y - y_0, z - z_0))$$

is orthogonal to the non-zero vector (a, b, c). Using the dot product, this is true if and only if

$$\begin{aligned} 0 &= (a, b, c) \cdot (x - x_0, y - y_0, z - z_0) \\ &= a(x - x_0) + b(y - y_0) + c(z - z_0) \\ &= ax + by + cz - (ax_0 + by_0 + cz_0) \\ &= ax + by + cz + d. \end{aligned}$$

For the second assertion, suppose $(a, b, c) \neq 0$. If we can find a point (x_0, y_0, z_0) that satisfies the equation $ax + by + cz + d = 0$, then we can go through the steps above in backwards fashion to rewrite the equation as

$$(a, b, c) \cdot (x - x_0, y - y_0, z - z_0) = 0.$$

Equivalently,

$$((x_0, y_0, z_0), (x, y, z)) \sim ((0, 0, 0), (x - x_0, y - y_0, z - z_0)) \perp ((0, 0, 0)(a, b, c)).$$

Thus, the existence of a point that satisfies the equation implies that the set of points that satisfy the equation is a plane with a normal vector (a, b, c). To find a point (x_0, y_0, z_0), we will use the fact that at least one of a, b, c is non-zero. If $a \neq 0$, then we set $y = z = 0$ and isolate $x = -\frac{d}{a}$. The $b \neq 0$ and $c \neq 0$ cases follow similarly. In any of the three cases, this process finds a point in the set. □

Corollary 13.11 (Point-plane perpendicular distance). If a plane is defined by the equation $Ax + By + Cz + D = 0$ and (x_0, y_0, z_0) is a point in space, then the perpendicular distance from the point to the plane is given by

$$\frac{|Ax_0 + By_0 + Cz_0 + D|}{\sqrt{A^2 + B^2 + C^2}}.$$

Proof. Let (x_1, y_1, z_1) be a point on the plane. We will proceed by the same clever method as in Theorem 4.17, by finding the length of the projection of the displacement vector

$$((x_1, y_1, z_1), (x_0, y_0, z_0))$$

to the plane's normal vector (A, B, C). By the formula in Theorem 4.13, our answer is

$$\begin{aligned} &\left\| \frac{\langle (x_0 - x_1, y_0 - y_1, z_0 - z_1), (A, B, C) \rangle}{\langle (A, B, C), (A, B, C) \rangle} (A, B, C) \right\| \\ &= \frac{|\langle (x_0 - x_1, y_0 - y_1, z_0 - z_1), (A, B, C) \rangle|}{\|(A, B, C)\|} \\ &= \frac{|Ax_0 - Ax_1 + By_0 - By_1 + Cz_0 - Cz_1|}{\sqrt{A^2 + B^2 + C^2}} \\ &= \frac{|Ax_0 + By_0 + Cz_0 + D|}{\sqrt{A^2 + B^2 + C^2}}, \end{aligned}$$

where we have used the fact that $Ax_1 + By_1 + Cz_1 + D = 0$ in the last step. As before, this length is the one we seek due to the triangle inequality and parallelogram law. □

Theorem 13.12. "The" shortest distance between two parallel planes in three dimensions is a well-defined concept because:

1. Suppose we are given a plane and a point in space that does not lie on the plane. Then there exists a unique line segment of shortest length with one endpoint on the point and the other endpoint on the plane. The segment is perpendicular to any direction vectors on the plane.

2. The distance from a point on one plane to the closest point on a parallel plane is a constant, regardless of the first point.

3. The shortest distance from any point on the first plane to the second plane is the same as the shortest distance from any point on the second plane to the first plane.

Proof. The proofs are very similar to the proofs of the corresponding parts of Lemma 7.17, showing that perpendicular distance once again converges with shortest distance. We encourage the reader to follow through with a write-up. It will be helpful to use the formula for the perpendicular distance between a point and a plane in space (Corollary 13.11). □

Problem 13.13. Prove that, if two points p, q are on a plane, then the line

$$\ell = \{p + t(q - p) : t \in \mathbb{R}\}.$$

through the two points lies entirely on the plane.

One may wonder about the extent to which the equation of a plane is unique. Certainly, we may multiply through an equation

$$ax + by + cz + d = 0$$

by a non-zero constant t to get

$$(ta)x + (tb)y + (tc)z + (td) = 0$$

without altering the set of points that satisfy the equation because the step is reversible. So any non-zero multiple of a normal vector to a plane is also a normal vector. But do any other equations in standard form yield the same plane? If we can show that all normal vectors to a plane are non-zero multiples of each other, then it will mean that all equations of the plane are non-zero multiples of each other. This is because if the equations

$$ax + by + cz + d_1 = 0,$$
$$(ta)x + (tb)y + (tc)z + d_2 = 0$$

are satisfied by the same point (x, y, z), then

$$td_1 - d_2 = t(ax + by + cz) - ((ta)x + (tb)y + (tc)z) = 0,$$

so $td_1 = d_2$. Therefore, let us show that all normal vectors to a plane are non-zero multiples of each other. This line of investigation naturally leads to the cross product as follows.

Theorem 13.14 (Derivation of the cross product). Let $a = (a_1, a_2, a_3)$ and $b = (b_1, b_2, b_3)$ be linearly independent (and, so, non-zero) position vectors. Then the set of normal vectors to the plane generated by a, b is precisely the set of non-zero multiples of the vector

$$n = (a_2b_3 - a_3b_2, a_3b_1 - a_1b_3, a_1b_2 - a_2b_1).$$

This vector is 0 if and only if $a = 0$ or $b = 0$ or a, b are non-zero linearly dependent vectors (so they point in the same or opposite directions). More briefly, n is non-zero if and only if a, b are linearly independent.

Proof. Following the definition of n in terms of $a, b,$ we will aim to show that $\pm\frac{n}{\|n\|}$ are the only unit normal vectors of the plane generated by $a, b.$ This will prove that all non-zero multiples of n are normal vectors, and that there are no further normal vectors beyond their multiples because normalizing any others to have magnitude 1 would produce these just two. First we will check that

$$n \cdot a = n \cdot b = 0.$$

Then we will naturally deduce $n,$ thereby proving that n is a normal vector that is unique up to multiplication by a non-zero real. Indeed,

$$\begin{aligned}\langle n, a\rangle &= \langle (a_2b_3 - a_3b_2, a_3b_1 - a_1b_3, a_1b_2 - a_2b_1), (a_1, a_2, a_3)\rangle \\ &= a_1(a_2b_3 - a_3b_2) + a_2(a_3b_1 - a_1b_3) + a_3(a_1b_2 - a_2b_1) \\ &= 0,\end{aligned}$$

and a similar computation shows that $\langle n, b\rangle = 0.$ Now we will compute the norm of n to check that it is non-zero and determine the conditions under which it is zero. The crucial idea is to use a sum-of-squares identity, called **Lagrange's identity**: for complex n-tuples

$$(a_1, a_2, \ldots, a_n) \text{ and } (b_1, b_2, \ldots, b_n),$$

it holds that

$$\left(\sum_{k=1}^{n} a_k^2\right)\left(\sum_{k=1}^{n} b_k^2\right) = \sum_{i=1}^{n-1}\sum_{j=1}^{n}(a_ib_j - a_jb_i)^2.$$

The real $n = 3$ case (which the reader should be able to check manually) then states that

$$\begin{aligned}\|n\|^2 &= (a_2b_3 - a_3b_2)^2 + (a_3b_1 - a_1b_3)^2 + (a_1b_2 - a_2b_1)^2 \\ &= (a_1^2 + a_2^2 + a_3^2)(b_1^2 + b_2^2 + b_3^2) - (a_1b_1 + a_2b_2 + a_3b_3)^2 \\ &= \|a\|^2\|b\|^2 - \langle a, b\rangle^2.\end{aligned}$$

The Cauchy-Schwarz inequality (Theorem 4.8) returns! By Cauchy-Schwarz, $\|n\| = 0$ if and only if a, b are linearly dependent, which is equivalent to our stated condition. Since our a, b are linearly independent, $\|n\|$ is strictly positive, so n is a non-zero vector and we can divide by its norm. This proves that every plane has a non-zero normal vector to it.

Next we will deduce all possibilities for $n.$ We know from a moment ago that every plane has a non-zero orthogonal vector, so suppose $m = (m_1, m_2, m_3)$ is one such normal vector.

Without loss of generality, we may assume that that m is a unit vector, since, as we said, m may be scaled to $\frac{m}{\|m\|}$. This yields the following relations:

$$
\begin{aligned}
0 &= \langle m, a \rangle = m_1 a_1 + m_2 a_2 + m_3 a_3, \\
0 &= \langle m, b \rangle = m_1 b_1 + m_2 b_2 + m_3 b_3, \\
1 &= \langle m, m \rangle = m_1^2 + m_2^2 + m_3^2.
\end{aligned}
$$

We wish to determine (m_1, m_2, m_3) from this system of equations. We do not know which variables might be zero, so we will have to avoid isolating variables by division. Instead, we will use elimination via multiplication. We will multiply the first equation by b_1, b_2, b_3 and the second equation by a_1, a_2, a_3, which produces the array of equations

$$
\begin{aligned}
0 &= m_1 a_1 b_1 + m_2 a_2 b_1 + m_3 a_3 b_1, & 0 &= m_1 b_1 a_1 + m_2 b_2 a_1 + m_3 b_3 a_1, \\
0 &= m_1 a_1 b_2 + m_2 a_2 b_2 + m_3 a_3 b_2, & 0 &= m_1 b_1 a_2 + m_2 b_2 a_2 + m_3 b_3 a_2, \\
0 &= m_1 a_1 b_3 + m_2 a_2 b_3 + m_3 a_3 b_3, & 0 &= m_1 b_1 a_3 + m_2 b_2 a_3 + m_3 b_3 a_3.
\end{aligned}
$$

Equating the equations that are beside each other yields the following three equations, after cancelling common terms and factoring a bit:

$$
\begin{aligned}
m_3(a_3 b_1 - a_1 b_3) &= m_2(a_1 b_2 - a_2 b_1), \\
m_1(a_1 b_2 - a_2 b_1) &= m_3(a_2 b_3 - a_3 b_2), \\
m_2(a_2 b_3 - a_3 b_2) &= m_1(a_3 b_1 - a_1 b_3).
\end{aligned}
$$

As we have previously verified that n is non-zero, we may define

$$
(\ , \ , \ ) = \frac{n}{\|n\|} = \frac{1}{\|n\|}(a_2 b_3 - a_3 b_2, a_3 b_1 - a_1 b_3, a_1 b_2 - a_2 b_1).
$$

This allows us to rewrite the equations as

$$
\begin{aligned}
m_3 \ &= m_2 \ , \\
m_1 \ &= m_3 \ , \\
m_2 \ &= m_1 \ .
\end{aligned}
$$

Moreover, $\ ^2 + \ ^2 + \ ^2 = 1$ holds. We will use these four equations to show that $m = \ (\ , \ , \ )$. Since $(\ , \ , \ ) \neq 0$ at least one of the three components has to be non-zero. Suppose it is $\ $. Then

$$
m_3 = m_1 \frac{\ }{\ } \text{ and } m_2 = m_1 \frac{\ }{\ }.
$$

By substitution, we get

$$
\begin{aligned}
1 &= m_1^2 + m_2^2 + m_3^2 \\
&= m_1^2 \left(1 + \frac{\ ^2}{\ ^2} + \frac{\ ^2}{\ ^2} \right) \\
&= m_1^2 \left(\frac{\ ^2 + \ ^2 + \ ^2}{\ ^2} \right) \\
&= \frac{m_1^2}{\ ^2}.
\end{aligned}
$$

Thus, $m_1 = \quad :$ Choosing $m_1 = \quad$ leads to $m_2 = \quad$ and $m_3 = \quad ;$ whereas choosing $m_1 = \quad$ leads to $m_2 = \quad$ and $m_3 = \quad :$ Now recall that we were only dealing with the case $\neq 0$; the cases where $\neq 0$ and $\neq 0$ follow similarly and lead to the same two options for $m:$ We encourage the reader to write out at least one of them. Therefore, the only two possible unit vectors that are normal to the plane generated by a and b are

$$(\; ; \; ; \;) = \quad \frac{n}{knk}:$$

This completes the proof, in accordance with our original strategy. □

Definition 13.15. The cross product of two position vectors $a = (a_1; a_2; a_3)$ and $b = (b_1; b_2; b_3)$ in R^3 is denoted by and defined as the vector

$$a \quad b = (a_2b_3 \quad a_3b_2; a_3b_1 \quad a_1b_3; a_1b_2 \quad a_2b_1):$$

This vector is the zero vector if and only if $a = 0$ or $b = 0$ or $a; b$ are non-zero linearly dependent vectors. As it is a complicated formula, the student may be interested in knowing the following mnemonic for remembering it in the form of a “formal” 3 3 determinant (recall determinants from Definition 1.27, where we mentioned the Rule of Sarrus memory trick)

$$\begin{aligned} a \quad b = \det \begin{bmatrix} i & j & k \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{bmatrix} &= ia_2b_3 + ja_3b_1 + ka_1b_2 \quad ka_2b_1 \quad ja_1b_3 \quad ia_3b_2 \\ &= (a_2b_3 \quad a_3b_2)i + (a_3b_1 \quad a_1b_3)j + (a_1b_2 \quad a_2b_1)k; \end{aligned}$$

where we define

$$\begin{aligned} i &= (1; 0; 0); \\ j &= (0; 1; 0); \\ k &= (0; 0; 1): \end{aligned}$$

Note that the definition for the cross product requires us to use position vectors, as opposed to arbitrary displacement vectors.

This formula for the cross product tell us an explicit way of finding a non-zero normal vector to any given plane. Every non-zero vector generates a plane to which it is normal (given a point on the plane), and every plane has a non-zero vector that is unique up to multiplication of the normal vector by a non-zero scalar.

13.2 Properties

Theorem 13.16. Given three distinct non-collinear points in space, there exists a plane that runs through them.

Proof. It suffices to find an equation of the form $ax + by + cz + d = 0$ that is satisfied by all three terms. We know that, if two linearly independent vectors lie on a plane, then their cross product is a normal to the plane. After that, it is a matter of knowing one point on the plane to determine the constant term of the equation. If the three points are $p, q, r,$ then the vectors (r, p) and (r, q) lie on the plane and are linearly independent by Lemma 9.26, since they are non-collinear. Then we can take the cross product

$$(r, p) \times (r, q) = (0, p - r) \times (0, q - r) = (a, b, c)$$

to get a normal vector and use the point $r = (r_1, r_2, r_3)$ to get the constant term $d = -(ar_1 + br_2 + cr_3).$ Thus, the equation $ax + by + cz + d = 0$ works. □

Theorem 13.17 (Algebraic properties of cross product)**.** Let a, b, c be vectors in $\mathbb{R}^3$ and r be a real number. Then the cross product satisfies the following properties:

1. Anticommutative: $a \times b = -(b \times a)$
2. Distributive over addition: $a \times (b + c) = a \times b + a \times c$
3. Compatible with scalar multiplication: $(ra) \times b = r(a \times b)$

It follows from the first and third properties that $a \times (rb) = r(a \times b)$ as well. We leave the algebraic verification of these properties to the reader as an independent exercise.

Problem 13.18. Let i, j, k be defined as in Definition 13.15. Compute each of

$$i \times j, \; j \times k, \; k \times i.$$

Lemma 13.19. Given two parallel lines in space, there exists a plane that goes through both lines. (Recall that parallel lines do not include the coincident configuration.)

Proof. Let the lines be $p_1 + sq$ and $p_2 + tq,$ where p_1, p_2 are distinct points, s, t are real parameters, and q is the common direction of the lines. We will show that the plane through $p_1, p_1 + q, p_2$ has the same equation as the plane through $p_2, p_2 + q, p_1.$ This will prove that there exists a plane that runs through all four points; by Problem 13.13, the line through $p_1, p_1 + q$ and the line through $p_2, p_2 + q$ will both lie on the plane. The first plane has a normal vector

$$(p_1 - p_2) \times (p_1 - (p_1 + q)) = (p_1 - p_2) \times (-q) = q \times (p_1 - p_2).$$

The second plane has a normal vector

$$(p_2 - (p_2 + q)) \times (p_2 - p_1) = (-q) \times (p_2 - p_1) = q \times (p_1 - p_2).$$

Since both planes have the same normal vector and both planes contain the point p_1 (and p_2), they satisfy the same standard equation, and so they must be the same plane. □

Theorem 13.20 (Trigonometric cross product)**.** If θ is the non-reflex angle between vectors a and $b,$ then

$$\|a \times b\| = \|a\| \, \|b\| \sin \theta.$$

Proof. According to the application of Lagrange's identity in the proof of Theorem 13.14 and the trigonometric dot product (Theorem 4.10),

$$\begin{aligned} \|a \times b\|^2 &= \|a\|^2 \|b\|^2 - \langle a, b\rangle^2 \\ &= \|a\|^2 \|b\|^2 - \|a\|^2 \|b\|^2 \cos^2\theta \\ &= \|a\|^2 \|b\|^2 (1 - \cos^2\theta) \\ &= \|a\|^2 \|b\|^2 \sin^2\theta . \end{aligned}$$

Since θ is non-reflex, $\sin\theta$ must lie in $[0, 1]$ making it non-negative. So we may take the square root of both sides to get

$$\|a \times b\| = \|a\| \|b\| |\sin\theta| = \|a\| \|b\| \sin\theta .$$

□

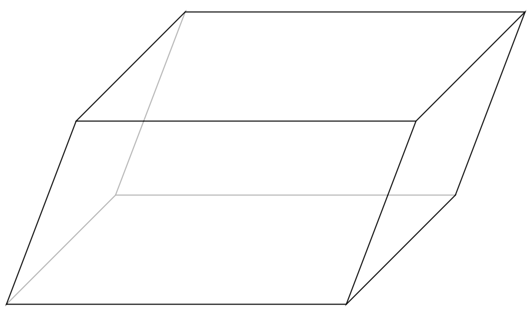
Corollary 13.21. An added benefit of the trigonometric cross product is that, by the sine area formula for a triangle,

$$\|a\| \|b\| \sin\theta$$

is the area of the parallelogram that has displacement vector representatives of a and b as adjacent sides with a common tail.

Definition 13.22. A **parallelepiped** is the three-dimensional analogue of a parallelogram. More precisely, there are at least three ways of characterizing it:

1. A 6-faced polyhedron such that each face is a parallelogram.
2. A 6-faced polyhedron such that there are three pairs of parallel faces.
3. A parallelogram-based prism (not necessarily right).



To construct a parallelepiped, it is enough to know three vectors that are edges that protrude from one vertex of a parallelepiped. The volume of a parallelepiped may be computed as the area of any base times the corresponding height, which is the distance between that base and its twin parallel base.

Theorem 13.23. Let P be the parallelepiped that has a vertex at the origin and the position vectors a, b, c as adjacent edges coming out of the origin. Then the volume of P is given by

$$V = |c \cdot (a \times b)|.$$

Subsequently, a, b, c are **coplanar**, meaning they inhabit the same plane, if and only if $V = 0$; this will be proven rigorously, but it should make geometric sense as well.

Proof. We recommend that the reader draws a diagram and adds to it while going through the proof. Let B be the area of the a, b base and h be the displacement normal vector from the a, b base to the arrowhead of c; this means that $\|h\|$ is the distance between these two parallel faces. By the trigonometric dot product, we know that

$$B = \|a\| \cdot \|b\| \cdot \sin\theta = \|a \times b\|,$$

where θ is the non-reflex angle between the vectors $a, b.$ Let α be the non-reflex angle between the normal vector $a \times b$ and $c.$ By Lemma 13.19, there exists a plane that runs through $a \times b$ and $h.$ By Problem 13.13, c also lies on this plane since it attaches the tail of $n = a \times b$ to the arrowhead of h. Let β be the non-reflex angle between h and $c.$ By the theorems on angles between parallel lines and a transversal, regardless of the direction of $a \times b$ in comparison to the direction of $h,$

$$|\cos\alpha| = |\cos\beta|$$

since $\cos x$ and $\cos(\pi - x)$ are negatives of each other. By the unit circle definition of cosine,

$$|\cos\beta| = \frac{\|h\|}{\|c\|} \implies \|h\| = \|c\| \cdot |\cos\beta|.$$

By the trigonometric dot product, the volume of the parallelepiped is

$$\begin{aligned} V = B \cdot \|h\| &= \|c\| \cdot \|a \times b\| \cdot |\cos\beta| \\ &= \|c\| \cdot \|a \times b\| \cdot |\cos\alpha| \\ &= |c \cdot (a \times b)|. \end{aligned}$$

For the corollary about a, b, c being coplanar, recall that $a \times b$ is orthogonal to the a, b plane, and a vector is orthogonal to $a \times b$ if and only if it lies on the a, b plane. Therefore, $|c \cdot (a \times b)| = 0$ if and only if c lies on the a, b plane, meaning a, b, c are coplanar. □

Definition 13.24. A computation of the form $a \cdot (b \times c),$ where there is the dot product of a vector with a cross product, is called a **scalar triple product**.

Theorem 13.25. If we have three position vectors

$$a = (a_1, a_2, a_3), b = (b_1, b_2, b_3), c = (c_1, c_2, c_3),$$

then

$$a \cdot (b \times c) = \det \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix}.$$

As a consequence,

$$a \cdot (b \times c) = b \cdot (c \times a) = c \cdot (a \times b).$$

Thanks to the anticommutativity of the cross product, the volume of parallelepiped can also be computed as

$$-a \cdot (c \times b) = -b \cdot (a \times c) = -c \cdot (b \times a).$$

Proof. The determinant identity may be proven by direct expansion of both sides, which we ask the reader to verify independently. The consequences follow because transposing any two rows of a matrix only changes change the sign of the determinant, and so applying two transpositions leaves the determinant intact (Theorem 1.28). □

Problem 13.26. For any vectors a, b in three dimensions, prove that

$$a \cdot (a \times b) = b \cdot (a \times a) = 0.$$

Problem 13.27 (Lagrange's formula). Let a, b, c be any three vectors in three dimensions. Since $b \times c$ is orthogonal to b and c, the vector triple product $a \times (b \times c)$ is orthogonal to $b \times c$. So $a \times (b \times c)$ must lie in the b, c plane. Prove that

$$a \times (b \times c) = (a \cdot c)b - (a \cdot b)c.$$

Problem 13.28 (Jacobi's identity). For any three vectors a, b, c in three dimensions, prove that

$$a \times (b \times c) + b \times (c \times a) + c \times (a \times b) = 0.$$

Chapter 14

Three Dimensions

> "I've always been passionate about geometry and the study of three-dimensional forms."
>
> – *Ernő Rubik*

The two basic measures of objects in three dimensions are volume and surface area. The objects that we will study are prisms and pyramids, followed by their curved counterparts, which are cylinders and cones, along with spheres. The material presented here about three-dimensional geometry will partly be about observing formulas whose rigorous proofs are beyond the presented level of exposition. However, one should be familiar with the definitions and formulas. In the cases where we derive the formulas with complete or partial proofs, it is important to understand the logic in addition to knowing the formulas.

14.1 Prisms and Pyramids

Definition 14.1. In the 2D plane, recall from Definition 2.16 that a half-plane is either part of the plane resulting from splitting the plane with a line. For our purposes, we include the splitting line as a part of each of the two resulting half-planes. The 3D analogue of a half-plane is a half-space, which is defined as either side of a plane in 3D space. Similar to how we include the line in a half-plane defined by the line, we include the plane in a half-space defined by the plane.

Similar to Definition 2.16, half-planes are a useful concept because it is possible to show that one formulation of the definition of a convex polygon is a bounded non-empty region in the 2D plane that is the intersection of finitely many half-planes, where we never include both half-planes resulting from any one line. This allows us to easily define the 3D analogue of convex polygons, as follows.

Definition 14.2. A convex polyhedron (the plural is polyhedra) is a bounded non-empty region in 3D space that is the intersection of finitely many half-spaces, where we never include both half-spaces resulting from any one plane. This region's boundary has three types of components:

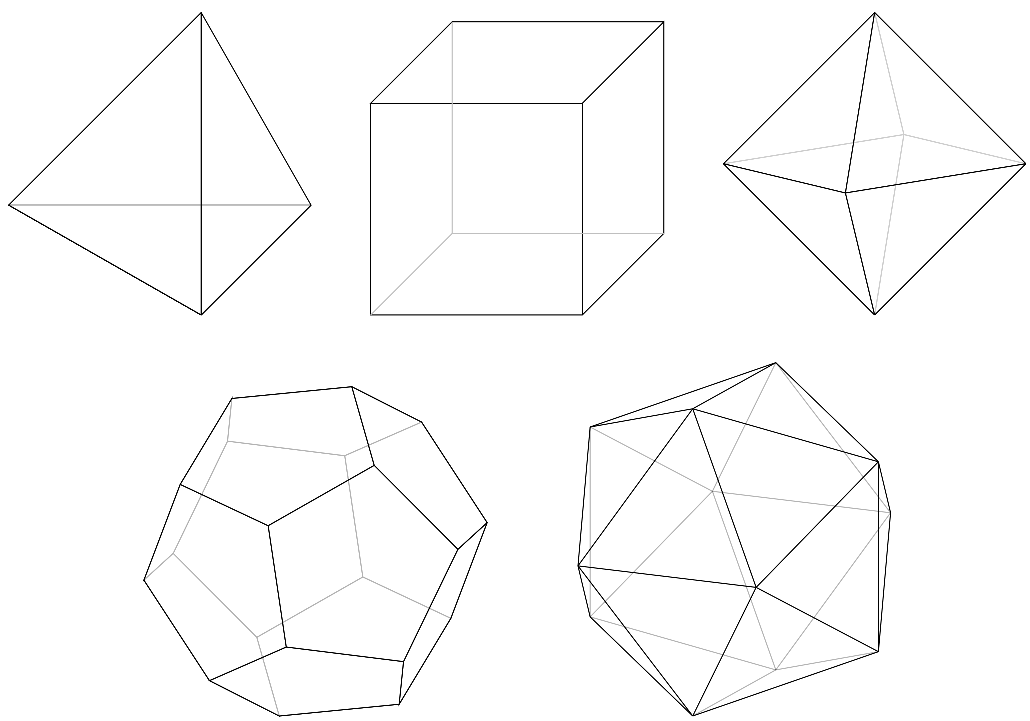
- Flat faces that are polygons resulting from each plane being cut by other planes
- Straight edges that are maximal line segments on the boundaries of each face
- Vertices that are corners at the endpoints of edges; these are also the vertices of faces

Unfortunately, it is much more difficult to define non-convex polyhedrons, so we will not attempt to characterize them. In fact, there is some controversy over several competing definitions that are not equivalent.

Definition 14.3. In a convex polyhedron, the **degree of a vertex** is the number of edges emanating from it. For convex polyhedra, it is clear that the degree of a vertex is also equal to the number of faces meeting at the vertex.

Definition 14.4. A **regular polyhedron**, otherwise known as a **Platonic solid**, is a convex polyhedron whose faces are congruent regular polygons and each vertex has the same degree.

Example. There are exactly five distinct Platonic solids: tetrahedron, hexahedron (cube), octahedron, dodecahedron, and icosahedron.



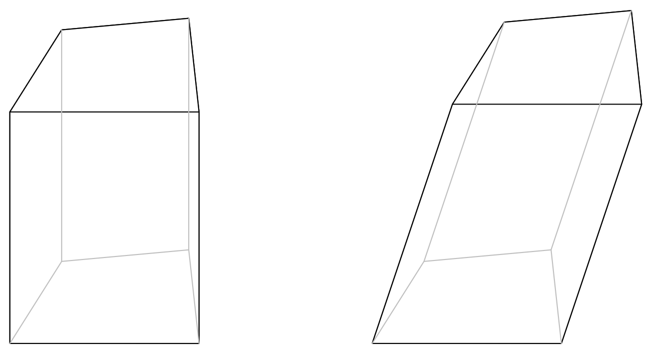
We classified them and studied the properties of their skeletons using the tools of graph theory in Volume 2.

Definition 14.5. There are two standard ways of measuring the size of a polyhedron, analogous to the perimeter and area of a polygon:

- The **surface area** of a polyhedron is the sum of the areas of its faces.
- Intuitively, the **volume** of a polyhedron is the amount of space it takes up in 3D space. This is the number of $1 \times 1 \times 1$ unit cubes that would fit inside it, including partial unit cubes.

Definition 14.6. A **prism** is a convex polyhedron whose construction begins with two congruent faces called bases that are translations of each other (i.e. no other transformations like rotation are involved) and that lie on parallel but different planes. The other faces, called **lateral faces**, are constructed by drawing line segments between corresponding vertices of the bases. If the bases are n-gons, we call the polyhedron an n-**gonal prism**. There are some related concepts:

- By Theorem 13.12, there is a constant perpendicular distance between the two bases because they lie on parallel planes. This is the prism's **height** or **altitude**.
- A **right prism** is a prism in which the lateral faces are perpendicular to the bases. Otherwise, it is called an **oblique prism**.



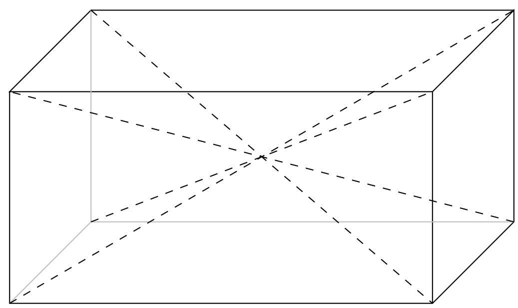
Example. The most common prism is a right **rectangular prism**, otherwise known as a **box**; the word "right" is usually dropped because it is presumed that a rectangular prism is right. This polyhedron has three disjoint sets of two parallel faces each, where any two faces from different sets lie on perpendicular planes. A box has three disjoint sets of parallel edges that each have four edges of equal length. These three lengths are called its **dimensions**, distinguished as the **length**, **width** and **height**, though there is no canonical way of making the distinction. A special case of a rectangular prism is a **cube**, whose three dimensions are all equal. If the dimensions of a box are $a, b, c,$ then its volume is defined as abc; this formula is consistent with the fact that if a, b, c are positive integers, then exactly abc unit cubes could be fitted together without overlaps or gaps to produce a box with dimensions $a \times b \times c.$

Example 14.7. Find a formula for the surface area of a rectangular prism with dimensions $a \times b \times c$ and use it to find the surface area of a cube of edge length $s.$

Solution. A rectangular prism has six faces, which come in three pairs of congruent faces. The sum of the areas of the six faces is $2(ab + bc + ca)$. Subsequently, since each face of a cube of edge length s is a square of side length s, its surface area is $6s^2$. □

Problem 14.8 (Square-cube law). Defining the similarity of polyhedrons in a sensible way, if the ratio of corresponding edges is k, then the ratio of surface areas is k^2 and the ratio of volumes is k^3. Prove this for boxes. This is the three-dimensional version of Corollary 9.24.

Problem 14.9. A space diagonal of a box is a line segment connecting one of the three pairs of opposite vertices. By "a pair of opposite vertices," we mean that there does not exist a face such that the two vertices both belong to that face. If a box has dimensions $a \times b \times c$, then determine the lengths of its space diagonals.



Theorem 14.10. We can compute the following:

1. The surface area of a right prism is $S = 2B + ph$, where B is the area of a base, p is the perimeter of a base, and h is the height of the prism.

2. The volume of a prism is $V = Bh$, where B is the area of a base and h is the height. A consequence of this formula is that the volume remains unchanged if the bases are translated to anywhere on the respective planes running through them, since the distance between two parallel planes is constant.

Proof. In a right prism, all the lateral faces are rectangles with height h. If the side lengths of the base are $s_1, s_2, \ldots, s_n$, then the sum of the areas of the lateral faces is

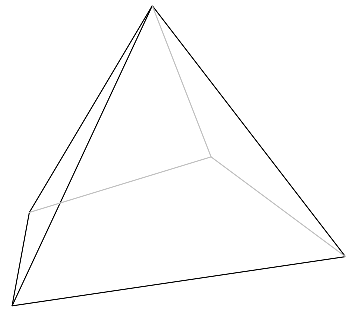
$$s_1h + s_2h + \cdots + s_nh = ph.$$

Then we simply add the areas of the bases, which is $2B$. We do not provide a proof of the volume formula. □

Definition 14.11. A pyramid is a convex polyhedron whose construction begins with a face called the base, and a vertex called the apex that lies outside the plane of the base. The other faces, called lateral faces, are constructed by drawing the line segments between the vertices of the base and the apex. If the base is an n-gon, then we call the polyhedron an n-gonal pyramid. There are some related concepts:

- The perpendicular distance between the apex and the plane running through the base is the height or altitude of the pyramid.

- For our purposes, we define a regular pyramid to be a pyramid such that its base is a regular polygon and the foot of its height is the center of the base. For our purposes, we will call non-regular pyramids oblique.



- The lateral faces of a regular pyramid are congruent triangles, and a lateral face height emanating from the apex is called a lateral height or slant height of the pyramid. The Pythagorean theorem tells us that the slant height is $\sqrt{r^2 + h^2}$ where r is the inradius of the base (this is the perpendicular distance from the center of the regular polygon that is the base to an edge of the base), and h is the height of the pyramid.

Example. A triangular pyramid, is a pyramid whose base is a triangle. Notice that any of the faces of a triangular pyramid can be taken as a base. A square-based pyramid is a pyramid with a square base.

Theorem 14.12. We can compute the following:

1. The surface area of a regular pyramid is $S = B + \frac{p\ell}{2}$, where B is the area of the base, p is the perimeter of the base, and ℓ is the slant height.

2. The volume of a pyramid is $V = \frac{Bh}{3}$, where B is the area of the base and h is the height. A consequence of this formula is that the volume remains unchanged if the apex is relocated to anywhere on the plane that runs through it and is parallel to the base, and if the base is translated to anywhere on the plane running through it. This is because the distance between two parallel planes is constant.

Proof. In a regular pyramid, let the slant height be l. If the side lengths of the base are $s_1, s_2, \ldots, s_n$, then the sum of the areas of the lateral faces is

$$\frac{s_1\ell}{2} + \frac{s_2\ell}{2} + \cdots + \frac{s_n\ell}{2} = \frac{p\ell}{2}.$$

Then we simply add the area of the base, which is B. We do not prove the volume formula, which can be proven by integration from calculus. □

Problem 14.13. Let V be a vertex if a cube of side length s and let A, B, C be the three vertices that are attached to V by an edge. The tetrahedron resulting from drawing $\triangle ABC$ is sliced off the cube. Determine the height of the tetrahedron corresponding to the base $\triangle ABC$.

Example 14.14. Find the surface area and volume of a tetrahedron in terms of its edge length s.

Solution. We find the surface area, followed by the volume:

1. A tetrahedron is a regular triangular pyramid, so it has four faces that are all equilateral triangles of side length s. The area of each equilateral triangle is $\frac{\sqrt{3}s^2}{4}$. Thus, the surface area of a tetrahedron with edge length s is
$$4\left(\frac{\sqrt{3}s^2}{4}\right) = \sqrt{3}s^2.$$

2. To find the volume, we only need to determine the height of the tetrahedron, as we already know the area of a base. The foot of the height is the center of the equilateral triangle that is the base because a tetrahedron is a regular pyramid. The center of an equilateral triangle cuts each height of the equilateral triangle in a 2 : 1 ratio and each height of the equilateral triangle has length $\frac{\sqrt{3}s}{2}$, so the distance from the foot of the height of the pyramid to a vertex of the base is
$$\frac{2}{3}\left(\frac{\sqrt{3}s}{2}\right) = \frac{\sqrt{3}s}{3}.$$
By the Pythagorean theorem, the height of the tetrahedron is
$$\sqrt{s^2 - \left(\frac{\sqrt{3}s}{3}\right)^2} = \frac{\sqrt{2}s}{\sqrt{3}}.$$
Therefore, the volume of a tetrahedron with side length s is
$$\frac{1}{3}\left(\frac{\sqrt{3}s^2}{4}\right)\left(\frac{\sqrt{2}s}{\sqrt{3}}\right) = \frac{\sqrt{2}s^3}{12}.$$

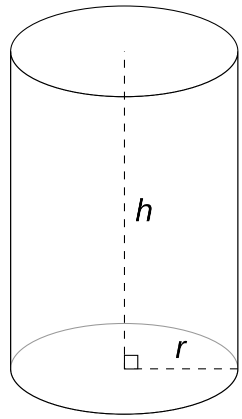
□

Problem 14.15. Find the surface area and volume of an octahedron in terms of its edge length s.

14.2 Curved Objects

Definition 14.16. A **cylinder** is like a prism, except its bases are closed disks. A **height** is defined analogously. There are some related concepts:

- The radius of the bases is called the **radius** of the cylinder.
- A **right cylinder** is a cylinder such that the line segment connecting the centers of the two bases is a height. Otherwise, it is called an **oblique cylinder**.



Theorem 14.17. We can compute the following:

1. The surface area of a right cylinder with radius r and height h is $2\ \ r(r+h)$.
2. The volume of a cylinder with radius r and height h is $\ \ r^2h$.

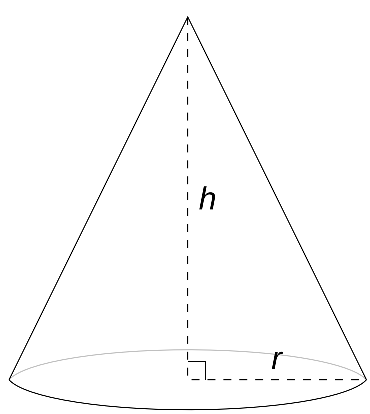
Proof. A right cylinder's bases can be popped out and the lateral surface can be unrolled along a height on the lateral surface to produce a rectangle with dimensions $2\ \ r$ and h, since one of the dimensions is the circumference of a base. This yields the surface area

$$2\ \ r^2 + 2\ \ rh = 2\ \ r(r+h).$$

If we assume that the volume formula for prisms holds with circular bases, then, since the area of a base is $\ \ r^2$, the volume is $\ \ r^2h$. □

Definition 14.18. A **cone** is like a pyramid, except its base is a closed disk. The **height** and the **apex** are defined analogously. The **lateral surface** of a cone is the part of the surface that excludes the base. There are some related concepts:

- The radius of the base is called the **radius** of the cone.
- A **right cone** is a cone in which the line segment connecting the apex to the center of the base is the height. In a right cone, the **lateral height** is the distance from the apex to a point on the circumference of the base (this distance is the same for all such points).



Theorem 14.19. We can compute the following:

1. The surface area of a right cone with radius r and height h is $\pi r(r + \sqrt{r^2 + h^2}).$
2. The volume of a cone with radius r and height h is $\frac{\pi r^2 h}{3}.$

Proof. The Pythagorean theorem tells us that a right cone has a constant slant height of

$$\ell = \sqrt{r^2 + h^2}.$$

So popping out the base and unrolling the lateral surface along a slant height produces a sector of a circle with radius $\sqrt{r^2 + h^2}.$ The sector has arc length equal to the circumference of the base, which is $2\pi r.$ In radians, the sector has measure

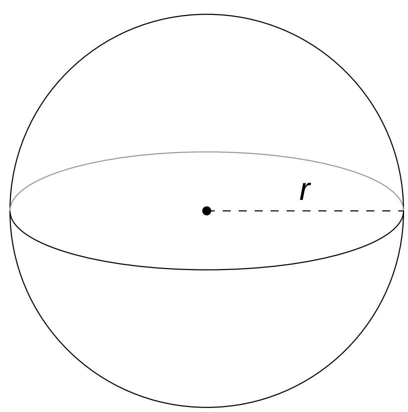
$$2\pi \cdot \frac{2\pi r}{2\pi \ell} = \frac{2\pi r}{\ell}.$$

Therefore the surface area is

$$\pi r^2 + \frac{1}{2} \cdot \frac{2\pi r}{\ell} \cdot \ell^2 = \pi r(r + \ell).$$

If we assume that the volume formula for pyramids holds with a circular base, then since the area of the base is πr^2, we get that the volume is $\frac{\pi r^2 h}{3}.$ □

Definition 14.20. A **sphere** is a 3D analogue of a circle. It is the collection of all points in 3D space that are at a fixed distance called the **radius**, from a certain point called the **center**.



Theorem 14.21. The following formulas hold for a sphere of radius r:

1. It has surface area $4\pi r^2$.
2. It has volume $\frac{4}{3}\pi r^3$.

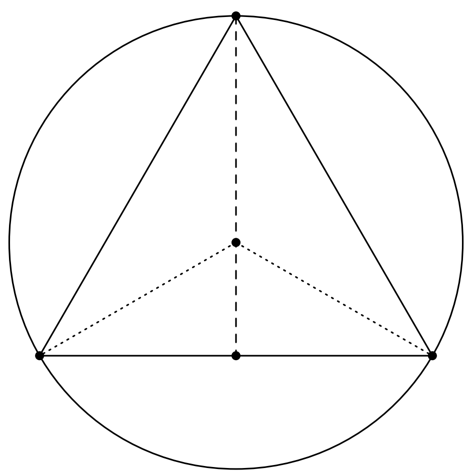
We do not have the means to prove either formula rigorously, as the standard proofs involve integrals from calculus.

Definition 14.22. A general technique that is useful in 3D problems is to convert the problem into a 2D problem. One instance of this technique is to cut through a 3D object with a plane to get a 2D shape that is the **cross section**.

Example. A cross section was taken to find the volume of an octahedron in the solution to Problem 14.15.
Every cross section of a sphere is a circle. A special case is that, if the cutting plane goes through the center of the sphere, then we call the resulting cross-section a **great circle**. As a side note, drawing the segments from the center of a sphere to the boundary of a non-great circle resulting from a cross-section produces a right cone. (Do you see why?)
We have already seen another famous example of cross sections in the conic sections, which are curves on the boundaries of cross-sections of cones. These are ellipses, parabolas, and hyperbolas. These were studied from a largely algebraic perspective in Chapter 12.

Example 14.23. A cone whose lateral height equals the diameter of its base is "inscribed" in a sphere, meaning each point of the circumference of the base of the cone touches the sphere and the apex of the cone also touches the sphere. Determine the ratio of the volume of the cone to the volume of the sphere.

Solution. Let the radius of the sphere be $R.$ We take a cross section of the configuration such that the cutting plane goes through the apex of the cone, the center of the sphere and a diameter of the base of the cone. Then the cross section consists of a great circle, which necessarily has radius $R,$ with an equilateral triangle whose vertices all lie on the great circle. We want to find the side length of this triangle because it equals the lateral height and the diameter of the base of the cone.



Let the side length of the triangle be $a.$ It was proven in Theorem 11.27 that the area of the triangle is $\frac{a^3}{4R}.$ Another way of finding the area of the triangle is $\frac{ah}{2},$ where h is the height. By equating these two formulas and using the Pythagorean theorem,

$$\frac{a^3}{4R} = \frac{ah}{2} \implies \frac{a^2}{2R} = h = \sqrt{a^2 - \left(\frac{a}{2}\right)^2} = \frac{\sqrt{3}a}{2} \implies a = \sqrt{3}R.$$

Thus the radius of the cone is

$$\frac{a}{2} = \frac{\sqrt{3}R}{2}$$

and the height of the cone is

$$h = \frac{\sqrt{3}a}{2} = \frac{\sqrt{3}\sqrt{3}R}{2} = \frac{3R}{2};$$

so the volume of the cone is

$$\frac{\pi}{3}\left(\frac{\sqrt{3}R}{2}\right)^2 \frac{3R}{2} = \frac{3}{8}\pi R^3.$$

Therefore, the ratio of the volume of the cone to the volume of the sphere is

$$\left(\frac{3}{8}\pi R^3\right) \Big/ \left(\frac{4}{3}\pi R^3\right) = \frac{9}{32}.$$

□

Appendices

Appendix A

Solutions

"Most of the time, I know what to do. I don't have to figure it out. I don't have to sit there, calculate for forty-five minutes, an hour, to know what is the right move. I just, usually I can just feel it immediately... I have to, you know, verify my opinion, see that I haven't missed anything. But a lot of the time, it's fairly useless because I know what I'm going to do, and then I sit there for a long time, and I do what I immediately wanted to do."

– Magnus Carlsen, 60 Minutes Overtime

Solution 1.18. The idea is that the two lines in the system turn out to the be same line, and so any point on the common line is a solution. If $p = (p_1, p_2)$ and $q = (q_1, q_2)$ are two solutions, we conjecture that all elements of the line $\{p + t(q - p) : t \in \mathbb{R}\}$ are solutions. A generic element of this set is

$$p + t(q - p) = (p_1, p_2) + t(q_1 - p_1, q_2 - p_2) = ((1 - t)p_1 + tq_1, (1 - t)p_2 + tq_2).$$

So we want it to be true that

$$\begin{aligned} a((1 - t)p_1 + tq_1) + b((1 - t)p_2 + tq_2) &= c, \\ \ ((1 - t)p_1 + tq_1) + \ ((1 - t)p_2 + tq_2) &= \ . \end{aligned}$$

We know that

$$\begin{aligned} ap_1 + bp_2 = c &\implies a(1 - t)p_1 + b(1 - t)p_2 = c(1 - t), \\ \ p_1 + \ p_2 = \ &\implies \ (1 - t)p_1 + \ (1 - t)p_2 = \ (1 - t), \end{aligned}$$

and

$$\begin{aligned} aq_1 + bq_2 = c &\implies atq_1 + btq_2 = ct, \\ \ q_1 + \ q_2 = \ &\implies \ tq_1 + \ tq_2 = \ t. \end{aligned}$$

Adding the first and third equations yields

$$a((1 - t)p_1 + tq_1) + b((1 - t)p_2 + tq_2) = c$$

and adding the second and fourth equations yields

$$\ ((1 - t)p_1 + tq_1) + \ ((1 - t)p_2 + tq_2) = \ .$$

This solution would have been cleaner and clearer using matrices, but we have not yet introduced the algebra of matrices.

Solution 1.31. Let $v = (v_1, v_2)$ and $w = (w_1, w_2)$ be points in $\mathbb{R}^2$ such that

$$((0,0),(v_1,v_2)) \text{ and } ((0,0),(w_1,w_2))$$

are linearly independent position vectors. We want to show that for every point $z = (z_1, z_2)$, there exist real numbers a, b such that $av + bw = z$. Expanding it out in terms of coordinates, we want

$$(av_1 + bw_1, av_2 + bw_2) = (z_1, z_2).$$

This can be written as the matrix system

$$\begin{pmatrix} v_1 & w_1 \\ v_2 & w_2 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}.$$

Since v and w are linearly independent, the determinant $v_1w_2 - v_2w_1$ is non-zero. So, we can multiply both sides by the inverse of the matrix $\begin{pmatrix} v_1 & w_1 \\ v_2 & w_2 \end{pmatrix}$ (whose determinant is $v_1w_2 - w_1v_2 \neq 0$) to get

$$\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} v_1 & w_1 \\ v_2 & w_2 \end{pmatrix}^{-1} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \frac{1}{v_1w_2 - w_1v_2} \begin{pmatrix} w_2 & -w_1 \\ -v_2 & v_1 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}.$$

This means the constants a, b exist, and we can solve for them in general in this way through matrix inversion and multiplication.

Solution 1.32. We know that $v \neq 0$. By Lemma 1.25, the linear dependence of u, v implies there exists $t \in \mathbb{R}$ such that $u = tv$. Similarly, the linear dependence of v, w implies there exists $s \in \mathbb{R}$ such that $w = sv$. Then $u = 0$ or $u \neq 0$. If $u \neq 0$, then $u = tv$ implies $t \neq 0$, in which case

$$w = sv = \frac{s}{t}u.$$

So $u = 0$ or, for $r = \frac{s}{t} \in \mathbb{R}$, $w = ru$. Therefore, u, w are linearly dependent.

Solution 2.27. By our complex criteria for lines being parallel or perpendicular (Theorem 2.24), we get the separate equations

$$\frac{z - f}{a - b} + \frac{\overline{z} - \overline{f}}{\overline{a} - \overline{b}} = 0,$$
$$\frac{f - a}{b - a} = \frac{\overline{f} - \overline{a}}{\overline{b} - \overline{a}}.$$

We want to find an expression for f, and as with many computations involving complex numbers, the challenge is to remove $\overline{f}$ from the equation. Isolating $\overline{f}$ in each equation yields the equations

$$-\left(\frac{z - f}{a - b}\right)(\overline{a} - \overline{b}) + \overline{z} = \overline{f} = \left(\frac{f - a}{b - a}\right)(\overline{b} - \overline{a}) + \overline{a}.$$

After equating the left and right sides, it is a matter of going through several lines of algebra to eliminate f; which we leave to the reader.

Solution 3.4. Since $f : X \to Y$ is bijective, we know that f^{-1} exists and maps Y to X bijectively. By the definition of inverses,

$$f(x_0) = y_0 \implies f^{-1}(y_0) = f^{-1}(f(x_0)) = x_0;$$
$$f^{-1}(y_0) = x_0 \implies f(x_0) = f(f^{-1}(y_0)) = y_0.$$

So $(x_0; y_0)$ is on the graph of f if and only if $(y_0; x_0)$ is on the graph of f^{-1}. Now it suffices to show that $(x_0; y_0)$ and $(y_0; x_0)$ are mirror images across the line $x = y$. By the reflection formula, the image of reflecting $z = x_0 + iy_0$ across $1 \cdot x + (-1) \cdot y + 0 = 0$ can be computed as $y_0 + ix_0$. This is what we wanted to see.

Solution 3.11. If $AB < AC$ and $BC < AC$ then

$$AB < AB + BC < AC + BC;$$
$$BC < BC + AB < AC + AB.$$

These are two triangle inequalities. If it turned out that $AC < AB + BC$ as well, Theorem 3.10 says that $A; B; C$ would form a non-degenerate triangle, contradicting the fact that they are collinear. So it must be be otherwise. That is, $AC = AB + BC$; which proves that B lies strictly between A and C, by our definition of "betweenness."

Solution 3.15. It suffices to verify that z_3 lies on the line though z_1 and z_2 and that z_3 is equidistant from z_1 and z_2; as this will force z_3 to lie on the line segment between z_1 and z_2. To do this, we compute that

$$\frac{z_2 - z_3}{z_1 - z_3} = \frac{z_2 - \frac{z_1+z_2}{2}}{z_1 - \frac{z_1+z_2}{2}} = -1 \implies |z_3 - z_1| = |z_3 - z_2|.$$

This means that the distance from z_3 to z_2 is equal to the distance from z_3 to z_1. Moreover, since $-1 = e^{i\pi}$, the first equation means that rotating z_1 counterclockwise around z_3 by π causes z_1 to coincide with z_2; so the three points are collinear; more formally, we can simply use Theorem 2.24.

Solution 3.16. Let w be the reflection of z across the line through a and b; and let f be the point at which the segment through z and w intersects the line through a and b. Since f is the midpoint of the segment between z and w, the complex midpoint formula (Problem 3.15) tells us that

$$f = \frac{z + w}{2} \implies w = 2f - z.$$

Since f is the foot of the perpendicular from z to the line through a and b, we can use the complex foot formula (Problem 2.27) to get

$$w = 2\left(\frac{z(\bar{a} - \bar{b}) + \bar{z}(a - b) + \bar{a}b - a\bar{b}}{2(\bar{a} - \bar{b})}\right) - z = \frac{\bar{z}(a - b) + \bar{a}b - a\bar{b}}{\bar{a} - \bar{b}}.$$

Solution 4.7. Let two adjacent sides be given by the position vectors v and w. By the parallelogram law (Theorem 1.39), the length of one diagonal is $\|v+w\|$ and the length of the other diagonal is $\|v-w\|$. By using the relation between the Euclidean norm and dot product, followed by expansions and cancellations, we get

$$\begin{aligned}
&\|x+y\|^2 + \|x-y\|^2 \\
&= \langle x+y, x+y\rangle + \langle x-y, x-y\rangle \\
&= \langle x, x\rangle + \langle x, y\rangle + \langle y, x\rangle + \langle y, y\rangle + \langle x, x\rangle + \langle x, -y\rangle + \langle -y, x\rangle + \langle -y, -y\rangle \\
&= \langle x, x\rangle + 2\langle x, y\rangle + \langle y, y\rangle + \langle x, x\rangle - 2\langle x, y\rangle + \langle y, y\rangle \\
&= 2\langle x, x\rangle + 2\langle y, y\rangle \\
&= 2\|x\|^2 + 2\|y\|^2.
\end{aligned}$$

This is the identity that we wanted to see.

Solution 4.12. Using the fact that the dot product distributes over addition, we compute that

$$\begin{aligned}
(v-w)\cdot(v+w) &= (v-w)\cdot v + (v-w)\cdot w \\
&= v\cdot v - w\cdot v + v\cdot w - w\cdot w \\
&= v\cdot v - w\cdot w \\
&= \|v\|^2 - \|w\|^2.
\end{aligned}$$

Solution 5.22. Since $\triangle ABC$ is isosceles with $CA = CB$, we also know that $\angle CAB = \angle CBA$. Since they cannot both be right or both be obtuse, they must both be acute. By Theorem 5.21, F lies in the interior of AB.

Solution 5.29. The sum of each exterior and interior angle is 180 degrees and there are n such pairs for a total of $180^\circ n$. Subtracting the sum of the interior angles from this yields

$$180^\circ n - 180^\circ(n-2) = 360^\circ.$$

This is a constant! It is in contrast to the sum of the interior angles formula that is monotonically increasing in n as we saw in Theorem 5.28.

Solution 5.31. The individual interior angles of equiangular n-gons measure

$$\frac{180^\circ(n-2)}{n} = 180^\circ\left(1 - \frac{2}{n}\right),$$

which approaches 180° as $n \to \infty$. The individual exterior angles of equiangular n-gons measure $\frac{360^\circ}{n}$, which approaches 0 as $n \to \infty$.

Solution 5.32. With the base case for triangles taken for granted, assume that there exists an integer $n \geq 3$ such that Pick's theorem holds for generalized lattice n-gons. Suppose P is a generalized lattice $(n+1)$-gon. After finding an ear V of P, let T be the induced triangle of V, which is easily seen to be a lattice triangle, and let Q be the generalized lattice n-gon resulting from clipping V. For any generalized polygon O, such as T, Q or P, let I_O be the number of lattice points in the interior of O, and let B_O be the number of lattice points on the boundary of O, including vertices. Let c be the number of lattice points on the edge connecting the neighbours U and W of V in P, including U and W. Then

$$\begin{aligned} I_P &= I_T + I_Q + (c - 2), \\ B_P &= B_T + B_Q - 2c + 2. \end{aligned}$$

By the base case and the induction hypothesis,

$$\begin{aligned} [T] &= I_T + \frac{B_T}{2} - 1, \\ [Q] &= I_Q + \frac{B_Q}{2} - 1. \end{aligned}$$

Summing these two equations and applying the equations that resulted from studying the absorption effect on the lattice points on UW, we get

$$\begin{aligned} [P] = [T] + [Q] &= I_T + I_Q + \frac{B_T + B_Q}{2} - 2 \\ &= I_P - (c - 2) + \frac{B_P + 2c - 2}{2} - 2 \\ &= I_P + \frac{B_P}{2} - 1. \end{aligned}$$

This completes the proof by induction.

It is reasonable to wonder how the result would be proven for triangles in order to establish the base case. Elementary proofs typically start off by proving the result for rectangles and then right triangles, neither of which are difficult. The trouble is the next step, when a minimal bounding rectangle (whose edges are parallel to the axes) is drawn around an arbitrary generalized lattice triangle, and right triangles are sliced off from the rectangle to produce the original triangle. The issues with such proofs is not only the fact it is necessary to verify Pick's theorem in the numerous cases of the last step, but that it is hard to be sure that all cases have been covered. However, if we move away from such an elementary proof in favour of rigour, a preliminary step that is usually taken is to prove that every primitive lattice triangle has area $\frac{1}{2}$, where a primitive generalized lattice polygon is defined as a generalized lattice polygon that has no lattice points in its interior and the only lattice points on its boundary are its vertices. There are other steps involved, but this is the most non-elementary as it involves using linear algebra to formalize the idea of tiling the plane with copies of this triangle.

Solution 6.3. Our proof of the second result will use the first result.

1. Let ABC be a triangle and let M be the midpoint of AB. For one direction, suppose $\angle ACB = 90^\circ$. We will use coordinates, specifically $C = (0, 0)$, $A = (0, a)$, and $B =$

$(b, 0)$. Then the $M = \left(\frac{a}{2}, \frac{b}{2}\right)$. This easily leads to

$$CM^2 = \left(\frac{a}{2} - 0\right)^2 + \left(\frac{b}{2} - 0\right)^2 = \frac{(a-0)^2 + (0-b)^2}{2^2} = \frac{AB^2}{2^2},$$

$$CM = \frac{AB}{2}.$$

Conversely, suppose $CM = \frac{AB}{2}$. As M is the midpoint of AB, we get $MA = MB = MC$. So the circle with center M and radius MA goes through A, B, C. Since AB is a diameter of this circle, Thales's theorem implies that $\angle ACB = 90^\circ$.

2. Let Γ be the circle with diameter UV and let W be a point on the plane that is distinct from U and V. Let M be the midpoint of UV. By the last part, $\angle UWV = 90^\circ$ if and only if $MW = \frac{UV}{2}$. Since

$$\frac{UV}{2} = MU = MV,$$

the condition $MW = \frac{UV}{2}$ is equivalent to $MW = MU = MV$. Finally, M is equidistant from U, V, W if and only if W lies on Γ, since Γ is the circle with diameter UV.

Solution 6.8. Let R be the circumradius of $\triangle ABC$. By the sum-to-product identities and the extended sine law,

$$\begin{aligned}\frac{\tan\left(\frac{A-B}{2}\right)}{\tan\left(\frac{A+B}{2}\right)} &= \frac{2 \cdot \sin\left(\frac{A-B}{2}\right) \cdot \cos\left(\frac{A+B}{2}\right)}{2 \cdot \sin\left(\frac{A+B}{2}\right) \cdot \cos\left(\frac{A-B}{2}\right)} \\ &= \frac{\sin A - \sin B}{\sin A + \sin B} \\ &= \frac{\frac{a}{2R} - \frac{b}{2R}}{\frac{a}{2R} + \frac{b}{2R}} \\ &= \frac{a-b}{a+b}.\end{aligned}$$

Solution 6.13. The second part will follow from the first.

1. The common tangent line is perpendicular to both radii that touch the point of tangency of the circles. Since the two radii are both perpendicular to the same line, the the lines through the two radii are parallel or the same line. Since the two radii share a common point (that is, the point of tangency of the two circles), the same line goes through them. This line contains both centers and the point of tangency of the circles.

2. By the previous part, the line through the centers contains the point of tangency of the circles. This line contains a radius from each circle. Since the tangent line through the point of tangency is perpendicular to each radius, it is perpendicular to the line through the centers.

Solution 6.17. If P lies on the major arc $\overset{\frown}{AB}$, then by the inscribed angle theorem, $\angle APB$ has the same measure as half the measure of the intercepted arc of the chord. According to the preceding lemma, this is exactly the acute angle between the chord and the tangent line. Similarly, if P lies on the minor arc $\overset{\frown}{AB}$, then by the inscribed angle theorem, $\angle APB$ has the same measure as half the measure of the arc opposite to the intercepted arc of the chord. According to the second part of the preceding lemma, this is exactly the obtuse angle between the chord and the tangent line.

Solution 6.19. We treat the three cases separately:

- Secant-secant: There are four intersection points. We connect opposing intersection points (meaning two intersection points that lie on opposite sides of the line through the other two) to create two inscribed angles that are equal alternate interior angles of a transversal. The result follows from the inscribed angled theorem.

- Secant-tangent: There are three intersection points. We connect the lone intersection point of the tangent with one of the secant's intersection points to produce two equal alternate interior angles of a transversal. The result follows from the inscribed angle theorem and the chord-tangent arc theorem.

- Tangent-tangent: There are two intersection points, one from each tangent. We know that, if a chord is perpendicular to a tangent and touches that tangent's point of tangency, then the chord is a diameter, by the chord-tangent arc theorem. The same must be true for the other chord corresponding to the other tangent. Since the tangents are parallel, the chords are either parallel or they coincide. Two diameters cannot be parallel as they always share the center, so they must coincide. So connecting the two intersection points produces a diameter, which cuts off two semicircles.

Solution 6.20. This is immediately true from applying the cosine law. One of the directions is a computation of length. The other direction uses the fact that cosine is bijective on the interval $(0^\circ, 180^\circ)$. Work out the details for yourself.

Solution 7.6. Let the centers of the circles be P and Q, let the common chord be AB, and let M be the midpoint of AB. By Theorem 7.5, we know that PM and QM are both perpendicular to AB; thus, $\angle PMQ$ is a straight angle, so the line through PQ is perpendicular to AB and runs through M.

Solution 7.7. Since cosine is strictly decreasing on the interval $(0, \pi)$ and the interior angles of a triangle lie in this interval, $\angle A > \angle B$ if and only if $\cos\angle B > \cos\angle A$. By the cosine law, we can take a sequence of reversible algebraic steps:

$$\frac{c^2+a^2-b^2}{2ac} > \frac{b^2+c^2-a^2}{2bc}$$

$$bc^2+a^2b-b^3 > ab^2+c^2a-a^3$$

$$(a-b)(a+b+c)(a+b-c) > 0.$$

Certainly $a + b + c > 0$ and the triangle inequality tells us that $a + b > c.$ So the inequality is true if and only if $a > b.$
This result tells us that $a > b$ and $a > c$ if and only if $\angle A > \angle B$ and $\angle A > \angle C$, so the longest side is opposite the largest angle. Similarly, $a < b$ and $a < c$ if and only if $\angle A < \angle B$ and $\angle A < \angle C$, so the shortest side is opposite the smallest angle.

Solution 7.14. If $\triangle ABC$ is equilateral, then all of the sides are equal and we have

$$\frac{AB}{BC} = \frac{BC}{CA} = \frac{CA}{AB} = 1,$$

so SSS similarity yields $\triangle ABC \sim \triangle BCA.$ In the other direction, suppose $\triangle ABC \sim \triangle BCA.$ Then there exists a positive constant k such that

$$\frac{AB}{BC} = \frac{BC}{CA} = \frac{CA}{AB} = k.$$

As a consequence,

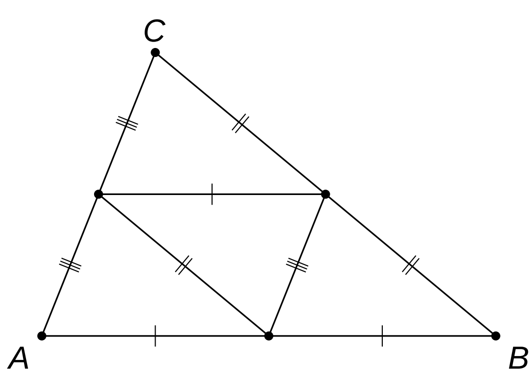
$$AB = k \cdot BC = k^2 \cdot CA = k^3 \cdot AB,$$

and so $k = 1.$ This proves that

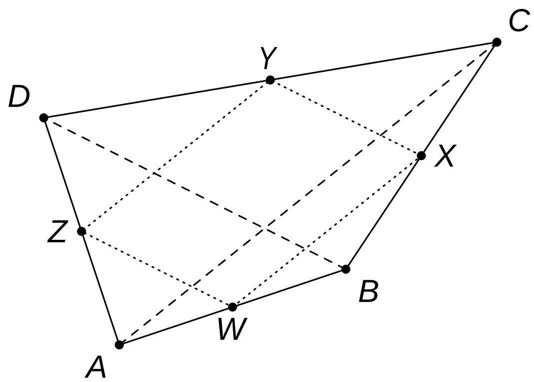
$$AB = BC = CA,$$

so $\triangle ABC$ is equilateral.

Solution 7.16. By SAS similarity, the nested triangles are similar with similarity ratio $\frac{1}{2}$ and have the same orientation. Thus, the sides opposite to the common angle are in a $1:2$ ratio and are parallel by Theorem 7.15. The four triangles that result from connecting the midpoints of all the edges are congruent by SSS congruence because, if the original triangle has side lengths $a, b, c,$ then the four triangles each have side lengths $\frac{a}{2}, \frac{b}{2}, \frac{c}{2}$, as shown in the diagram.



Solution 7.20. First we draw the two diagonals AC and BD of the convex quadrilateral $ABCD.$ Let the midpoints of AB, BC, CD, DA be W, X, Y, Z respectively. By Problem 7.16, WX and YZ are parallel to $AC,$ and XY and ZW are parallel to $BD.$ By Lemma 2.15, XY and ZW are parallel, and YZ and WX are parallel. Thus, $WXYZ$ is a parallelogram.



Solution 7.23. Suppose we have a rhombus. Since it is a parallelogram, its diagonals bisect each other due to Theorem 7.22. Drawing the diagonals and using the fact that they bisect each other and that all sides have equal length, we find using SSS congruence that the diagonals produce four equal right angles at their intersection.
Conversely, suppose the diagonals of a convex quadrilateral perpendicularly bisect each other. By SAS congruence, the four right triangles produced by the diagonals intersecting are congruent. So the sides of the quadrilateral (which are the hypotenuses) are all equal, making the convex quadrilateral a rhombus.

Solution 7.24. Suppose we have a rectangle $ABCD$. Every rectangle is a parallelogram, so its diagonals bisect each other due to Theorem 7.22. We know that the four interior angles are right angles, and that opposite sides of a parallelogram have equal length. By SAS congruence, $4ABC = 4DCB;$ so the diagonals AC and DB have equal length.
Conversely, suppose the diagonals of a quadrilateral are equal in length and bisect each other. Then we can draw a circle whose center is the intersection point of the diagonals and the endpoints of vertices of the quadrilateral are on the circle. This means the diagonals are diameters, so Thales's theorem (Theorem 6.2) says that the four interior angles of the quadrilateral are all 90 . Therefore, the quadrilateral is a rectangle.

Solution 8.6. Let $ABCD$ be a cyclic quadrilateral, and for the sake of brief notation, let $AB = a; BC = b; CD = c; DA = d:$ Let be the measure of the interior angle $\backslash ABC$ and let be the measure of the interior angle $\backslash ADC:$ By cyclicity, $+ = ;$ so

$$\frac{a^2 + b^2 \quad AC^2}{2ab} = \cos \quad = \quad \cos \quad = \frac{c^2 + d^2 \quad AC^2}{2cd}:$$

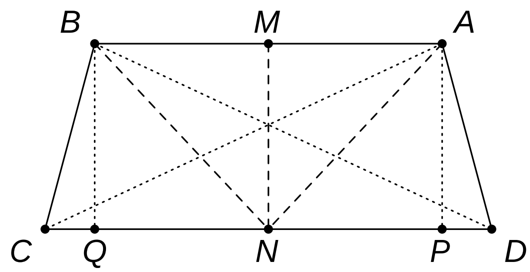
Clearing the denominators and isolating AC^2 yields

$$AC^2 = \frac{a^2cd + b^2cd + c^2ab + d^2ab}{ab + cd} = \frac{(ac + bd)(ad + bc)}{ab + cd}:$$

By a similar derivation,

$$BD^2 = \frac{(ac + bd)(ab + cd)}{ad + bc}:$$

Solution 8.8. Suppose $ABCD$ is a trapezoid. We will show that a pair of base angles are equal if and only if connecting the midpoints of the bases creates a segment perpendicular to both bases. Let $AB \parallel CD$ without loss of generality, and let M be the midpoint of AB and N be the midpoint of CD.



- Suppose MN is perpendicular to AB and CD. After drawing AN and BN, SAS congruence tells us that $\triangle ANM \cong \triangle BNM$. This means $AN = BN$ and $\angle ANM = \angle BNM$. Then

$$\angle AND = 90^\circ - \angle ANM = 90^\circ - \angle BNM = \angle BNC$$

  Combined with $DN = CN$ and $AN = BN$, this tells us that $\triangle AND \cong \triangle BNC$ by SAS congruence. Thus, the base angles $\angle ADN$ and $\angle BCN$ are equal.

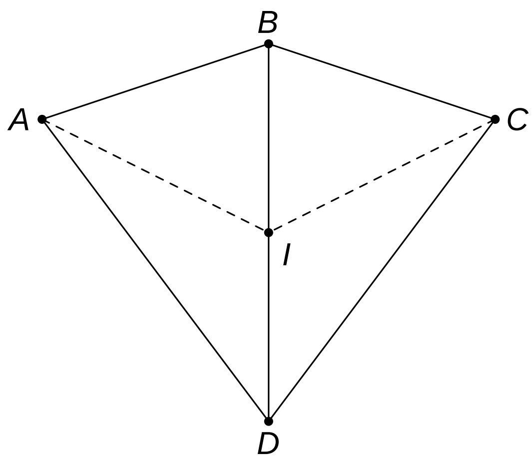
- Suppose a pair of base angles are equal; we know that this means the other pair of base angles are also equal. Let the feet of the perpendiculars from A and B to the line through CD be P and Q, respectively. Since $AB \parallel CD$ and base angles are equal, $\angle ADC = \angle BCD$ are acute and so are $\angle ACD$ and $\angle BDC$. So P and Q both lie on the segment CD. Then $AP = BQ$ because parallel lines have a constant perpendicular distance between them (Lemma 7.17), and $\angle DAP = \angle CBQ$ since $\angle ADP = \angle BCQ$. Then AAS congruence yields $\triangle DAP \cong \triangle CBQ$ and so $DA = CB$. By SAS congruence, $\triangle ADN \cong \triangle BCN$, so $AN = BN$. Since $AM = BM$, SSS congruence yields $\triangle AMN \cong \triangle BMN$. Thus, $\angle AMN = \angle BMN = 90^\circ$ and, since alternate interior angles of a transversal are equal, $\angle DNM = \angle CNM = 90^\circ$ as well.

Solution 8.22. Let I be the incenter of the tangential polygon. Let V_i be the vertex connecting edges s_{i-1} and s_i. Let the foot of the perpendicular segment from I to the line through each s_i be X_i; we know that X_i lies in the interior of s_i since we are handling a tangential polygon. Drawing IV_i produces right triangles $\triangle IV_iX_{i-1}$ and $\triangle IV_iX_i$, implies $\angle X_iIV_i = \angle X_{i-1}IV_i$. As the two perpendicular distances IX_{i-1} and IX_i are equal and the right triangles share the hypotenuse IV_i, HL congruence tells us that the triangles are congruent. As this can be done with every vertex, the lengths that we seek are $t_i = X_{i-1}V_i = V_iX_i$. More simply, we can just use the fact that tangent segments to a circle from a given external point are equal in length.

Solution 8.23. Let the sides of the tangential polygon be $s_0, s_1, \ldots, s_{n-1}$. Drawing the segment between the incenter and each vertex partitions the polygon into n triangles, each of which has a base s_i and corresponding height $r.$ Thus, the area of the polygon is

$$\sum_{i=0}^{n-1} \frac{rs_i}{2} = r\left(\frac{1}{2}\sum_{i=0}^{n-1} s_i\right) = rs.$$

Solution 8.25. First we label the vertices of the kite in clockwise order as $ABCD$ where $AB = BC$ and $CD = DA.$ We want to find a point that lies on all of the interior angle bisectors. By SSS congruence, $\triangle ABD \cong \triangle CBD,$ so every point on the segment BD lies on the bisectors of the interior angles $\angle ABC$ and $\angle ADB.$ Let I be the point at which the bisector of the interior angle $\angle BAD$ intersects $BD.$ What we need to do is show that CI is the bisector of the interior angle $\angle BCD.$ By SAS congruence, $\triangle ABI \cong \triangle CBI$ and $\triangle ADI \cong \triangle CDI.$



So,

$$\begin{aligned}\angle AIB &= \angle CIB,\\ \angle AID &= \angle CID.\end{aligned}$$

As a result,

$$\begin{aligned}\angle ICB &= 180^\circ - \angle IBC - \angle CIB\\ &= 180^\circ - \angle IBA - \angle AIB\\ &= \angle BAI\\ &= \angle DAI\\ &= 180^\circ - \angle IDA - \angle AID\\ &= 180^\circ - \angle IDC - \angle CID\\ &= \angle ICD.\end{aligned}$$

Thus all kites are tangential, and so are rhombuses since rhombuses are kites where all the sides are equal.

Solution 8.27. The sum of the interior angles in a generalized n-gon is $180^\circ(n-2)$; so each interior angle in a regular n-gon is $180^\circ \cdot \frac{n-2}{n} < 180^\circ$; which makes regular n-gons convex. Let the vertices of a regular n-gon be $V_0, V_1, \ldots, V_{n-1}$ in clockwise or counterclockwise order. Let the interior angle bisectors at V_i and V_{i+1} intersect at I_i for each index $0 \le i \le n-1$; where indices are reduced modulo n. Since the interior angles are all equal, this produces n isosceles triangles $\triangle V_i I_i V_{i+1}$. By ASA congruence, these n triangles are congruent for all i because the edges of the polygon are all equal in length and the base angles are all equal to $180^\circ \cdot \frac{n-2}{2n}$. Thus, all the $V_i I_i$ are equal and so all the I_i are the same point I. This point I lies on the angle bisectors of the all the interior angles, so it is the incenter. Moreover, since all the $V_i I$ are equal, I is also the circumcenter.

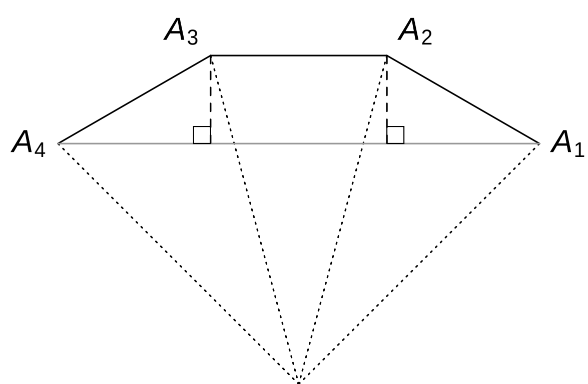
Solution 9.9. We know that the diagonals intersect perpendicularly, so let c be split into c_1 and c_2; and d be split into d_1 and d_2. Since there are four right triangles produced, the area of the orthodiagonal convex quadrilateral is the sum of the four areas, which is

$$\frac{c_1 d_1 + c_2 d_1 + c_1 d_2 + c_2 d_2}{2} = \frac{(c_1 + c_2)(d_1 + d_2)}{2} = \frac{cd}{2}.$$

Solution 9.17. First we calculate using the usual method that the interior angles of a regular dodecagon measure

$$\frac{12 \cdot 180^\circ - 360^\circ}{12} = 150^\circ.$$

Then we drop perpendiculars from A_2 and A_3 to A_1A_4. It is easy to verify that the two right triangles produced are 30°-60°-90° triangles with the angles at A_2 and A_3 measuring $150^\circ - 90^\circ = 60^\circ$.



Therefore,

$$A_1A_4 = s + 2 \cdot \frac{\sqrt{3}s}{2} = s \cdot (1 + \sqrt{3}).$$

Solution 10.2. Each of the two new right triangles have a right angle, as well as a shared angle with the original triangle. By AA similarity, the two new triangles are similar to the original larger triangle.

Solution 10.6. Let D be the foot of either cevian emanating from A, and let $BD = m, CD = n, AD = d$. In the case of a median, $m = n = \frac{a}{2}$. By Stewart's theorem,

$$d^2a + \frac{a^3}{4} = \frac{b^2a}{2} + \frac{c^2a}{2}.$$

By isolating d, we get

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

Solution 10.7. Suppose a^2, b^2, c^2 form an arithmetic sequence. Then the common difference is equal to both of

$$b^2 - a^2 = c^2 - b^2,$$

which is equivalent to

$$a^2 + c^2 = 2b^2.$$

We can use this with Apollonius's theorem to get

$$\begin{aligned}
4m_a^2 &= 2b^2 + 2c^2 - a^2 = (a^2 + c^2) + 2c^2 - a^2 = 3c^2, \\
4m_b^2 &= 2c^2 + 2a^2 - b^2 = 2 \cdot 2b^2 - b^2 = 3b^2, \\
4m_c^2 &= 2a^2 + 2b^2 - c^2 = 2a^2 + (a^2 + c^2) - c^2 = 3a^2.
\end{aligned}$$

Therefore, (a, b, c) is scaled by a factor of $\frac{\sqrt{3}}{2}$ to produce (m_c, m_b, m_a).
In the other direction, suppose there exists a scale factor k such that

$$(ka, kb, kc) = (m_c, m_b, m_a).$$

Then

$$\begin{aligned}
4k^2c^2 &= 4m_a^2 = 2b^2 + 2c^2 - a^2, \\
4k^2b^2 &= 4m_b^2 = 2c^2 + 2a^2 - b^2, \\
4k^2a^2 &= 4m_c^2 = 2a^2 + 2b^2 - c^2.
\end{aligned}$$

Adding them yields

$$\begin{aligned}
4k^2(a^2 + b^2 + c^2) &= 3(a^2 + b^2 + c^2) \\
4k^2 &= 3.
\end{aligned}$$

This turns the first equation above into

$$\begin{aligned}
3c^2 &= 2b^2 + 2c^2 - a^2 \\
a^2 + c^2 &= 2b^2 \\
b^2 - a^2 &= b^2 - c^2,
\end{aligned}$$

which proves that (a^2, b^2, c^2) is an arithmetic sequence, thereby establishing the second direction of the result.

Solution 10.8. Let D be the foot of either cevian emanating from A, and let $BD = m, CD = n, AD = d$. In the case of an angle bisector, the challenge is to express m and n in terms of a, b, c. The angle bisector theorem (Theorem 10.4) tells us that $bm = cn$. By some clever algebra,

$$\frac{a}{n} = \frac{m+n}{n} = \frac{m}{n} + 1 = \frac{c}{b} + 1 = \frac{c+b}{b} \implies n = \frac{ab}{b+c},$$
$$\frac{a}{m} = \frac{n+m}{m} = \frac{n}{m} + 1 = \frac{b}{c} + 1 = \frac{b+c}{c} \implies m = \frac{ac}{b+c}.$$

Now we substitute these expressions for m and n into Stewart's theorem to get

$$\begin{aligned} d^2 &= \frac{b^2m + c^2n}{a} - mn \\ &= bc - \frac{a^2bc}{(b+c)^2} \\ &= \frac{bc((b+c)^2 - a^2)}{(b+c)^2}. \end{aligned}$$

Taking the square root of both sides yields the final formula

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

If desired, the difference of squares factorization can be applied inside the square root.

Solution 10.14. Following the proofs of Ceva's theorem and van Aubel's theorem, we find that

$$\frac{PX}{AX} = \frac{[PBX]}{[ABX]} = \frac{[PCX]}{[ACX]} = \frac{[PBX] + [PCX]}{[ABX] + [ACX]} = \frac{[PBC]}{[ABC]}.$$

By similar derivations,

$$\frac{PY}{BY} = \frac{[PCA]}{[BCA]},$$
$$\frac{PZ}{CZ} = \frac{[PAB]}{[CAB]}.$$

Summing the three expressions yields Gergonne's theorem. The alternate form comes from

$$\begin{aligned} \frac{AP}{AX} + \frac{BP}{BY} + \frac{CP}{CZ} &= \frac{AX - PX}{AX} + \frac{BY - PY}{BY} + \frac{CZ - PZ}{CZ} \\ &= 3 - \left(\frac{PX}{AX} + \frac{PY}{BY} + \frac{PZ}{CZ} \right) \\ &= 3 - 1 = 2. \end{aligned}$$

Solution 10.15. Suppose AD, BE, CF concur. Using Ceva's theorem and its converse, along with the given fact $CD = DB$, we can get the equivalent condition

$$\begin{aligned}
\frac{AE}{EC} \cdot \frac{CD}{DB} \cdot \frac{BF}{FA} &= 1 \\
\frac{AE}{EC} \cdot \frac{BF}{FA} &= 1 \\
\frac{EC}{AE} &= \frac{FB}{AF} \\
\frac{AE + EC}{AE} &= \frac{AF + FB}{AF} \\
\frac{AC}{AE} &= \frac{AB}{AF}.
\end{aligned}$$

Using this ratio of lengths and the fact that $\angle FAE = \angle BAC$, SAS similarity tells us that $\triangle FAE \sim \triangle BAC$. This is equivalent to $\angle AFE = \angle ABC$ and $\angle AEF = \angle ACB$, which is equivalent to FE being parallel to BC, via the F-angle theorem. The steps are reversible, so both directions of the results have been established.

Solution 10.17. By the trigonometric Ceva's theorem (Theorem 10.16), we wish to obtain that

$$\frac{\sin BAX}{\sin XAC} \cdot \frac{\sin ACZ}{\sin ZCB} \cdot \frac{\sin CBY}{\sin YBA} = 1,$$

so we will work on each of the fractions. By a rearranged form of the ratio lemma (Theorem 10.9) that isolates the quotient of the sines,

$$\begin{aligned}
\frac{\sin BAX}{\sin XAC} &= \frac{FX}{XE} \cdot \frac{AE}{FA}, \\
\frac{\sin ACZ}{\sin ZCB} &= \frac{EZ}{ZD} \cdot \frac{CD}{EC}, \\
\frac{\sin CBY}{\sin YBA} &= \frac{DY}{YF} \cdot \frac{BF}{DB}.
\end{aligned}$$

Multiplying these equations together yields

$$\begin{aligned}
\frac{\sin BAX}{\sin XAC} \cdot \frac{\sin ACZ}{\sin ZCB} \cdot \frac{\sin CBY}{\sin YBA} &= \left(\frac{FX}{XE} \cdot \frac{AE}{FA}\right)\left(\frac{EZ}{ZD} \cdot \frac{CD}{EC}\right)\left(\frac{DY}{YF} \cdot \frac{BF}{DB}\right) \\
&= \left(\frac{FX}{XE} \cdot \frac{EZ}{ZD} \cdot \frac{DY}{YF}\right)\left(\frac{AE}{FA} \cdot \frac{CD}{EC} \cdot \frac{BF}{DB}\right).
\end{aligned}$$

The first product is equal to 1 by Ceva's theorem (Theorem 10.12), as is the second one because

$$\frac{AE}{FA} \cdot \frac{CD}{EC} \cdot \frac{BF}{DB} = \frac{AE}{EC} \cdot \frac{CD}{DB} \cdot \frac{BF}{FA} = 1.$$

Solution 11.3. Since the foot of each median is the midpoint of an edge, $AZ = ZB$ and $AY = YC$. By van Aubel's theorem (Theorem 10.13),

$$\frac{AP}{PX} = \frac{AY}{YC} + \frac{AZ}{ZB} = 1 + 1 = 2.$$

In the same way, $\frac{BP}{PY} = 2$ and $\frac{CP}{PZ} = 2.$

As for the second part of the problem, some of these triangles can be immediately seen to have equal areas because they have equal bases running through the same line with shared corresponding heights. So we let

$$= [BPX] = [CPX];$$
$$= [CPY] = [APY];$$
$$= [APZ] = [BPZ].$$

Now we need to show that $\quad = \quad = \quad .$ Using the previous part of the problem,

$$\frac{2}{\quad} = \frac{[BCP]}{[PCY]} = \frac{BP}{PY} = 2 \implies \quad = \quad ;$$
$$\frac{2}{\quad} = \frac{[ACP]}{[APZ]} = \frac{CP}{PZ} = 2 \implies \quad = \quad .$$

Solution 11.9. Let $BC = a; CA = b; AB = c.$ Using Ravi substitution,

$$s \quad a = AY = AZ;$$
$$s \quad b = BZ = BX;$$
$$s \quad c = CX = CY.$$

The result follows from the converse of Ceva's theorem because

$$\frac{BX}{XC} \frac{CY}{YA} \frac{AZ}{ZB} = \frac{s \quad b}{s \quad c} \frac{s \quad c}{s \quad a} \frac{s \quad a}{s \quad b} = 1.$$

Solution 11.13. In order to apply the converse of Ceva's theorem, we need to compute

$$\frac{BI_{AA}}{CI_{AA}} \frac{CI_{BB}}{AI_{BB}} \frac{AI_{CC}}{BI_{CC}}.$$

Let $b = CA$ and $c = AB.$ In the notation of Example 11.12,

$$BI_{AA} = BI_{AB} = AI_{AB} \quad AB = s \quad c;$$
$$CI_{AA} = CI_{AC} = AI_{AC} \quad AC = s \quad b.$$

Computing the other lengths in the same way, we get

$$\frac{BI_{AA}}{CI_{AA}} \frac{CI_{BB}}{AI_{BB}} \frac{AI_{CC}}{BI_{CC}} = \frac{s \quad c}{s \quad b} \frac{s \quad a}{s \quad c} \frac{s \quad b}{s \quad a} = 1.$$

Solution 11.18. Using the base-height formula for the area of a triangle,

$$\frac{1}{h_a} + \frac{1}{h_b} + \frac{1}{h_c} = \frac{a}{2[ABC]} + \frac{b}{2[ABC]} + \frac{c}{2[ABC]} = \frac{s}{[ABC]} = \frac{s}{sr} = \frac{1}{r};$$

where s is the semiperimeter and we have used the area formula rs for tangential polygons. By Example 11.12, the other computation is

$$\frac{1}{r_a}+\frac{1}{r_b}+\frac{1}{r_c}=\frac{s-a}{sr}+\frac{s-b}{sr}+\frac{s-c}{sr}=\frac{3s-2s}{sr}=\frac{1}{r}.$$

Solution 11.20. Let $\triangle ABC$ be equilateral. By Example 11.19, the altitude and interior angle bisector emanating from A are the same cevian, and the perpendicular bisector of BC contains both. Since the perpendicular bisector of BC goes through A, it must contain the median emanating from A, which then is the same cevian as the aforementioned altitude and angle bisector. Thus, the perpendicular bisector of BC contains a cevian that is simultaneously the median, interior angle bisector and altitude emanating from A. By a symmetric argument the analogous statements are true for the perpendicular bisectors of CA and AB. The perpendicular bisectors of the edges intersect at the circumcenter. Since the centroid, incenter and orthocenter are known to exist for any triangle, and distinct lines cannot have more than one intersection point, we find that all four centers are the same point.

Solution 11.21. By the definition of the orthocenter, H is the orthocenter if and only if

$$\overrightarrow{AH}\perp\overrightarrow{BC},\ \overrightarrow{BH}\perp\overrightarrow{CA},\ \overrightarrow{CH}\perp\overrightarrow{AB}$$

if and only if

$$\begin{aligned}0&=(\vec{H}-\vec{A})\cdot(\vec{C}-\vec{B})\\&=(\vec{H}-\vec{B})\cdot(\vec{A}-\vec{C})\\&=(\vec{H}-\vec{C})\cdot(\vec{B}-\vec{A})\end{aligned}$$

if and only if, letting the circumcenter O be the origin,

$$\begin{aligned}0&=(\overrightarrow{OH}-\overrightarrow{OA})\cdot(\overrightarrow{OC}-\overrightarrow{OB})\\&=(\overrightarrow{OH}-\overrightarrow{OB})\cdot(\overrightarrow{OA}-\overrightarrow{OC})\\&=(\overrightarrow{OH}-\overrightarrow{OC})\cdot(\overrightarrow{OB}-\overrightarrow{OA}).\end{aligned}$$

There can be only one orthocenter and one circumcenter, since concurrent lines cannot intersect at more than one place. So, there can be only one vector $\overrightarrow{OH}$. It suffices to substitute

$$\overrightarrow{OH}=\overrightarrow{OA}+\overrightarrow{OB}+\overrightarrow{OC}$$

into these three equations and check that they all hold. Indeed, by using

$$\|OA\|=\|OB\|=\|OC\|=R,$$

we find that

$$
\begin{aligned}
(\overrightarrow{OH} - \overrightarrow{OA}) \cdot (\overrightarrow{OC} - \overrightarrow{OB}) &= (\overrightarrow{OC} + \overrightarrow{OB}) \cdot (\overrightarrow{OC} - \overrightarrow{OB}) \\
&= \|\overrightarrow{OC}\|^2 - \|\overrightarrow{OB}\|^2 = R^2 - R^2 = 0; \\
(\overrightarrow{OH} - \overrightarrow{OB}) \cdot (\overrightarrow{OA} - \overrightarrow{OC}) &= (\overrightarrow{OA} + \overrightarrow{OC}) \cdot (\overrightarrow{OA} - \overrightarrow{OC}) \\
&= \|\overrightarrow{OC}\|^2 - \|\overrightarrow{OB}\|^2 = R^2 - R^2 = 0; \\
(\overrightarrow{OH} - \overrightarrow{OC}) \cdot (\overrightarrow{OB} - \overrightarrow{OA}) &= (\overrightarrow{OB} + \overrightarrow{OA}) \cdot (\overrightarrow{OB} - \overrightarrow{OA}) \\
&= \|\overrightarrow{OB}\|^2 - \|\overrightarrow{OA}\|^2 = R^2 - R^2 = 0;
\end{aligned}
$$

where we have used Problem 4.12. For the computation of OH^2, the relation between the Euclidean dot product and norm yields

$$
\begin{aligned}
OH^2 = \|\overrightarrow{OH}\|^2 &= \langle \overrightarrow{OH}, \overrightarrow{OH} \rangle \\
&= \langle \overrightarrow{OA} + \overrightarrow{OB} + \overrightarrow{OC}, \overrightarrow{OA} + \overrightarrow{OB} + \overrightarrow{OC} \rangle \\
&= \|\overrightarrow{OA}\|^2 + \|\overrightarrow{OB}\|^2 + \|\overrightarrow{OC}\|^2 + 2\langle \overrightarrow{OA}, \overrightarrow{OB} \rangle + 2\langle \overrightarrow{OB}, \overrightarrow{OC} \rangle + 2\langle \overrightarrow{OC}, \overrightarrow{OA} \rangle.
\end{aligned}
$$

So we need to compute these components. The first three are easy because, by the definition of the circumcenter,

$$\|\overrightarrow{OA}\|^2 = \|\overrightarrow{OB}\|^2 = \|\overrightarrow{OC}\|^2 = R^2.$$

In a more involved argument, we find using the trigonometric dot product, inscribed angle theorem, a trigonometric double angle identity, and the extended law of sines that

$$
\begin{aligned}
2\langle \overrightarrow{OA}, \overrightarrow{OB} \rangle &= 2 \cdot R \cdot R \cdot \cos 2C \\
&= 2R^2(1 - 2\sin^2 C) \\
&= 2R^2 \left(1 - 2\left(\frac{c}{2R}\right)^2\right) \\
&= 2R^2 - c^2, \\
2\langle \overrightarrow{OB}, \overrightarrow{OC} \rangle &= 2R^2 - a^2 \\
2\langle \overrightarrow{OC}, \overrightarrow{OA} \rangle &= 2R^2 - b^2.
\end{aligned}
$$

Adding these all up, we find that

$$OH^2 = 3R^2 + (2R^2 - a^2) + (2R^2 - b^2) + (2R^2 - c^2) = 9R^2 - a^2 - b^2 - c^2,$$

as expected.

Solution 11.23. Suppose $\triangle ABC$ is oriented counterclockwise, as shown. Then a counterclockwise rotation by $\frac{\pi}{3}$ of CB around the center C produces D, of AC around the center A produces E, and of BA around the center F produces F. Denoting each capital letter vertex's complex coordinates by the corresponding small letter and letting $x = e^{\frac{\pi}{2} i}$, this yields the equations

$$
\begin{aligned}
d - c = (b - c)x &\implies d = xb + (1 - x)c, \\
e - a = (c - a)x &\implies e = xc + (1 - x)a, \\
f - b = (a - b)x &\implies f = xa + (1 - x)b.
\end{aligned}
$$

By the complex centroid formula (Theorem 11.22),

$$3p = b + c + d;$$
$$3q = c + a + e;$$
$$3r = a + b + f.$$

We wish to prove that $\triangle PQR$ is equilateral, so it suffices to prove that a rotation of QP by $\frac{\pi}{3}$ radians around the center R produces exactly PR. As an equation, we wish to prove that

$$\frac{r-p}{q-p} = x.$$

To that end, we compute

$$\begin{aligned}
\frac{r-p}{q-p} &= \frac{3r-3p}{3q-3p}; \\
&= \frac{(a+b+f)-(b+c+d)}{(c+a+e)-(b+c+d)} \\
&= \frac{a+f-c-d}{a+e-b-d} \\
&= \frac{a+[xa+(1-x)b]-c-[xb+(1-x)c]}{a+[xc+(1-x)a]-b-[xb+(1-x)c]} \\
&= \frac{(1+x)a+(1-2x)b+(x-2)c}{(2-x)a+(-1-x)b+(2x-1)c}
\end{aligned}$$

Since three counterclockwise rotations by $\frac{\pi}{3}$ produce a reflection across the point of rotation, we find that $x^3 = -1$. So the above can be multiplied by $-x^3 = 1$ to get

$$\begin{aligned}
\frac{-x^3}{-1} \cdot \frac{(1+x)a+(1-2x)b+(x-2)c}{(2-x)a+(-1-x)b+(2x-1)c} &= x \cdot \frac{(x^2+x^3)a+(x^2-2x^3)b+(x^3-2x^2)c}{(x-2)a+(1+x)b+(1-2x)c} \\
&= x \cdot \frac{(x^2-1)a+(x^2+2)b+(-1-2x^2)c}{(x-2)a+(1+x)b+(1-2x)c}
\end{aligned}$$

Since

$$x = e^{\frac{\pi}{3}i} = \cos\frac{\pi}{3} + i \cdot \sin\frac{\pi}{3} = \frac{1+i\sqrt{3}}{2};$$

which is a root of $f(t) = t^2 - t + 1$, we can rearrange $x^2 - x + 1 = 0$ to get equality of the corresponding coefficients in the numerator and denominator

$$x^2 - 1 = x - 2;$$
$$x^2 + 2 = 1 + x;$$
$$-1 - 2x^2 = 1 - 2x.$$

Therefore, the above expression is $x - 1 = x$, as desired. This proves the fact about the outer Napoleon triangle being equilateral.

The inner Napoleon triangle is also equilateral because it replaces the rotation by $\frac{\pi}{3}$ counterclockwise with a rotation by $\frac{\pi}{3}$ clockwise, so we simply have to replace all $x = e^{\frac{\pi}{3}i}$ with $y = e^{-\frac{\pi}{3}i}$, which still works because this is the other root $\frac{1 - i\sqrt{3}}{2}$ of $f(t) = t^2 - t + 1$.

Solution 11.29. Let $a = BC; b = CA; c = AB$. By Heron's formula,

$$\begin{aligned} r r_a r_b r_c &= \frac{[ABC]}{s} \cdot \frac{[ABC]}{s-a} \cdot \frac{[ABC]}{s-b} \cdot \frac{[ABC]}{s-c} \\ &= \frac{[ABC]^4}{s(s-a)(s-b)(s-c)} \\ &= [ABC]^2. \end{aligned}$$

Solution 12.10. By subtracting Q_1 from Q_2, it suffices to prove that any bivariate quadratic

$$Q(x; y) = Ax^2 + Bxy + Cy^2 + Dx + Ey + F$$

that identically vanishes (meaning, the output is 0 for all real inputs $(x; y)$) has all zero coefficients

$$(A; B; C; D; E; F) = (0; 0; 0; 0; 0; 0).$$

First, substituting $y = 0$ gives

$$Q(x; 0) = Ax^2 + Dx + F.$$

This is a univariate quadratic that identically vanishes, so the identity theorem for polynomials from Volume 1 says that $A = D = F = 0$. Secondly, substituting $x = 0$ gives

$$Q(x; 0) = Cy^2 + Ey + F.$$

This is another univariate quadratic that identically vanishes, so $C = E = F = 0$. Only B is left, which we handle by substituting $(x; y) = (1; 1)$ into $Q(x; y)$, all of the other coefficients of which have disappeared by now. This gives

$$0 = Q(1; 1) = B \cdot 1 \cdot 1 = B.$$

Solution 12.15. By Theorem 12.3 and Theorem 12.14, the coefficients of $x^2; xy; y^2$, respectively, form a scalar multiple of

$$(A; B; C) = ((1 - e^2)a^2 + b^2; -2e^2ab; a^2(1 - e^2)b^2);$$

it suffices to prove that

$$A^2 + B^2 + C^2 \neq 0.$$

To this end, we compute

$$
\begin{aligned}
A^2 + B^2 + C^2 &= [(1 - e^2)a^2 + b^2]^2 + (-2abe^2)^2 + [a^2(1 - e^2)b^2]^2 \\
&= [(1 - e^2)^2 a^4 + 2(1 - e^2)a^2b^2 + b^4] + 4a^2b^2e^4 \\
&+ [a^4 + 2a^2b^2(1 - e^2) + (1 - e^2)^2 b^4] \\
&= (a^4 + b^4)(1 - e^2)^2 + (a^4 + b^4) + 4a^2b^2(1 - e^2) + 4a^2b^2e^4 \\
&= (a^4 + b^4)[(1 - e^2)^2 + 1] + 4a^2b^2[(1 - e^2)^2 + e^2],
\end{aligned}
$$

which consists of two non-negative terms added together. In fact, since a and b cannot simultaneously be zero, $a^4 + b^4 0 >$. We also know that $(1 - e^2)^2 + 1 \geq 1 > 0$ by the trivial inequality, so $A^2 + B^2 + C^2$ is strictly positive.

Solution 12.20. Let $ax + by + c = 0$ be a directrix of S and let (x_0, y_0) be the corresponding focus of S. Then we know that

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + F$$

is a bivariate quadratic whose zero set is S, where

$$
\begin{aligned}
A &= a^2(1 - e^2) + b^2, \\
B &= -2abe^2, \\
C &= a^2 + (1 - e^2)b^2.
\end{aligned}
$$

Since there exists a non-zero real λ such that

$$Q(x, y) = \lambda(Ax^2 + Bxy + Cy^2 + Dx + Ey + F),$$

we can compute the discriminant of Q to be

$$
\begin{aligned}
(\lambda B)^2 - 4(\lambda A)(\lambda C) &= \lambda^2(B^2 - 4AC) \\
&= \lambda^2((-2abe^2)^2 - 4(a^2(1 - e^2) + b^2)(a^2 + (1 - e^2)b^2)) \\
&= 4\lambda^2(a^2 + b^2)^2(e + 1)(e - 1).
\end{aligned}
$$

Since $4\lambda^2(a^2 + b^2)^2(e + 1)$ is always positive, the sign of the discriminant is the sign of $e - 1$. This sign determines the type of the conic because conics are classified into ellipses, parabolas and hyperbolas according to whether e lies in $(0, 1)$, is equal to 1, or is greater than 1, respectively.

It turns out that the discriminant of a bivariate quadratic is invariant under Euclidean isometries. Since a Euclidean isometry can be decomposed into translations, rotations around the origin and conjugations, the result can be proven by taking the discriminant of the three standard forms. This is still computationally intensive, so we will skip over it, but the reader may be interested in independently verifying it.

Solution 12.21. The answers can be produced by isolating y in terms of x or isolating x in terms of y in each of the three equations.

- For an ellipse in standard form, an x value has a corresponding y value if and only if $-p \le x \le p$; for each such x value, there exist two corresponding y values that happen to be negations of each other. Similarly, a y value has a corresponding x value if and only if $-q \le y \le q$; for each such y value, there exist two corresponding x values that happen to be negations of each other. This shows the minimum and maximum value of each coordinate.

- For a parabola in standard form, an x value has a corresponding y value if and only if $x \ge 0$; for each such x value, there exist two corresponding y values that are negatives of each other. All real y values have exactly one corresponding x value.

- For a hyperbola in standard form, an x value has a corresponding y value if and only if $x \le -p$ or $x \ge p$; for each such x value, there exist two corresponding y values that happen to be negations of each other. All real y values have exactly two corresponding x values, which happen to be negations of each other.

Solution 12.23. The solution to each part leads to the next one:

1. Every hyperbola is congruent to a hyperbola in standard form. Since Euclidean isometries preserve angles, the original hyperbola has perpendicular asymptotes if and only if the congruent hyperbola in standard form has perpendicular asymptotes. Thus, it suffices to prove the assertion for only hyperbolas in standard form. The asymptotes of a hyperbola in standard form are $y = \pm\frac{q}{p}x.$ They are perpendicular if and only if the slopes $\frac{q}{p}$ and $-\frac{q}{p}$ are opposite reciprocals, which is equivalent to $p = q.$ If the eccentricity is e then $q = p\sqrt{e^2 - 1}$ by the definition of $q,$ which means $p = q$ if and only if if $e = \sqrt{2}.$

2. As we saw in the first part, a hyperbola in standard form is rectangular if and only if $p = q.$ By clearing the equal denominators of the standard form

$$\frac{x^2}{p^2} - \frac{y^2}{p^2} = 1$$

   means an equation for such a conic is $x^2 - y^2 = p^2$.

3. Let $Q(x, y) = x^2 - y^2 - p^2.$ Rotating the graph of $Q(x, y) = 0$ by θ around the origin yields the graph of

$$Q(x\cos\theta + y\sin\theta, y\cos\theta - x\sin\theta) = 0.$$

   For $\theta = \frac{\pi}{4},$ this reduces to $2xy - p^2 = 0.$ Thus, the resulting conic is the graph of $xy = \frac{p^2}{2}.$ This means the graph of the usual reciprocal function $f(x) = \frac{1}{x}$ is a rectangular hyperbola.

Solution 12.24. Let Q be a bivariate quadratic whose zero set is a conic. By Lemma 3.5, rotating the graph of $Q(x; y) = 0$ by θ around the origin is the graph of

$$Q(x\cos\theta + y\sin\theta; y\cos\theta - x\sin\theta) = 0.$$

We compute that

$$(\cos\theta; \sin\theta) = \begin{cases} (0; 1) & \text{if } \theta = \frac{\pi}{2} \\ (-1; 0) & \text{if } \theta = \pi \\ (0; -1) & \text{if } \theta = \frac{3\pi}{2} \end{cases}$$

and so

$$Q(x\cos\theta + y\sin\theta; y\cos\theta - x\sin\theta) = \begin{cases} Q(y; -x) & \text{if } \theta = \frac{\pi}{2} \\ Q(-x; -y) & \text{if } \theta = \pi \\ Q(-y; x) & \text{if } \theta = \frac{3\pi}{2} \end{cases}.$$

This allows us to produce the following table of bivariate quadratics.

	Ellipse	Parabola	Hyperbola
$Q(x; y)$	$\frac{x^2}{p^2} + \frac{y^2}{q^2} - 1$	$y^2 - 4px$	$\frac{x^2}{p^2} - \frac{y^2}{q^2} - 1$
$Q(y; -x)$	$\frac{x^2}{q^2} + \frac{y^2}{p^2} - 1$	$x^2 - 4py$	$\frac{y^2}{p^2} - \frac{x^2}{q^2} - 1$
$Q(-x; -y)$	$\frac{x^2}{p^2} + \frac{y^2}{q^2} - 1$	$y^2 + 4px$	$\frac{x^2}{p^2} - \frac{y^2}{q^2} - 1$
$Q(-y; x)$	$\frac{x^2}{q^2} + \frac{y^2}{p^2} - 1$	$x^2 + 4py$	$\frac{y^2}{p^2} - \frac{x^2}{q^2} - 1$

Note that rotating an ellipse or hyperbola counterclockwise around its center by π yields the same conic.

In the standard form of an ellipse, $p > q$ and the major axis is horizontal. Upon a counterclockwise rotation by $\frac{\pi}{2}$ around the origin, the major axis becomes vertical and this table shows that p and q switch places. This tells us what the graph of $\frac{x^2}{p^2} + \frac{y^2}{q^2} = 1$ looks like for $p < q$.

In the standard form of a hyperbola, the branches open left and right. Upon a counterclockwise rotation by $\frac{\pi}{2}$ around the origin, the branches open down and up and this table shows that the signs of the two non-constant terms are switched. This tells us what the graph of $\frac{y^2}{q^2} - \frac{x^2}{p^2} = 1$ looks like for any $p; q$.

Solution 12.27. As with the analogous proof for ellipses, we apply a Euclidean isometry to the configuration so that $F_1 = (r, 0)$ and $F_2 = (-r, 0)$ for some $r > 0.$ With some foresight, we define p so that $d = 2p.$ Then we are seeking all points $P = (x, y)$ such that

$$\left|\sqrt{(x+r)^2 + y^2} - \sqrt{(x-r)^2 + y^2}\right| = 2p.$$

Once again we perform a sequence of manipulations without paying attention to their reversibility:

$$\begin{aligned}
\sqrt{(x+r)^2 + y^2} - \sqrt{(x-r)^2 + y^2} &= \pm 2p \\
\sqrt{(x+r)^2 + y^2} &= \sqrt{(x-r)^2 + y^2} \pm 2p \\
(x+r)^2 + y^2 &= (x-r)^2 + y^2 + 2p^2 \pm 4p\sqrt{(x-r)^2 + y^2} \\
rx - p^2 &= \pm p\sqrt{(x-r)^2 + y^2} \\
(rx - p^2)^2 &= p^2((x-r)^2 + y^2) \\
p^2(r^2 - p^2) &= x^2(r^2 - p^2) - p^2y^2 \\
1 &= \frac{x^2}{p^2} - \frac{y^2}{r^2 - p^2}.
\end{aligned}$$

Since $p = \frac{d}{2} < \frac{F_1F_2}{2} = r,$ we can define $q > 0$ to satisfy $q^2 = r^2 - p^2,$ which means that S is a subset of a hyperbola H in standard form. Due to the relation $r^2 = p^2 + q^2$ being satisfied, r is the linear eccentricity, and so F_1 and F_2 are the foci. And $d = 2p$ is the length of the transverse axis.

For the second part, let the eccentricity of H' be e and let P' be a point on H' that is on the branch closer to F_1' than to $F_2'.$ Let the directrices of H' be d_1 and $d_2.$ By using the focus-directrix definition of a hyperbola and the fact that the two directrices are parallel, we get

$$P'F_2' - P'F_1' = e \cdot d(P', d_2) - e \cdot d(P', d_1) = e \cdot d(d_2, d_1) = 2p,$$

since the standard form shows that the distance between the directrices is $\frac{2p}{e}$. (Note that here p is the symbol used in the standard form of a hyperbola.) Similarly, if P' were on the branch closer to F_2' than to F_1' then we can show that $P'F_1' - P'F_2' = 2p.$ In either case, $|P'F_1' - P'F_2'| = 2p.$ Thus, $2p$ is the desired positive real constant $d' < F_1'F_2',$ since the segment between the foci strictly contains the transverse axis in a hyperbola.

Going back to the first part, S can be shown to be non-empty by showing that either of the vertices of H satisfies the condition required to be in $S.$ Thus, the second part implies that $H \subseteq S,$ proving $S = H.$

Solution 13.13. Let an equation of the plane be $ax + by + cz + d = 0,$ and let the two points be

$$p = (p_1, p_2, p_3) \text{ and } q = (q_1, q_2, q_3).$$

Every point on the line through p and q is parametrized by

$$\begin{aligned}
p + t(q - p) &= (1 - t)p + tq \\
&= ((1 - t)p_1 + tq_1, (1 - t)p_2 + tq_2, (1 - t)p_3 + tq_3).
\end{aligned}$$

Plugging this generic point into $ax + by + cz + d$ yields

$$\begin{aligned} ax + by + cz + d &= a((1-t)p_1 + tq_1) + b((1-t)p_2 + tq_2) + c((1-t)p_3 + tq_3) + d \\ &= (1-t)(ap_1 + bp_2 + cp_3) + t(aq_1 + bq_2 + cq_3) + d \\ &= (1-t)(-d) + t(-d) + d \\ &= -d + td - td + d = 0. \end{aligned}$$

So the generic point on the line lies on the plane.

Solution 13.18. It follows directly from the formula for the cross product that

$$i \times j = k, \; j \times k = i, \; k \times i = j.$$

The details of plugging the coordinates into the formula are left to the reader.

Solution 13.26. These identities can be proven by plugging the coordinates of a and b into the expressions and using the dot product and cross product formulas. However, the determinant form of the scalar triple product provides an easier process. This is because the determinant remains unchanged when one row is subtracted from another row. Consequently,

$$a \cdot (a \times b) = \det \begin{bmatrix} a_1 & a_2 & a_3 \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{bmatrix} = \det \begin{bmatrix} 0 & 0 & 0 \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{bmatrix} = 0.$$

A similar proof can be used to show that $b \cdot (a \times a) = 0$, though we can alternatively use the fact that $a \times a = 0$, so $b \cdot (a \times a) = b \cdot 0 = 0$.

Solution 13.27. For each position vector v, we will denote the components of its arrowhead by (v_1, v_2, v_3). Our proof of Lagrange's formula will utilize brute computations as we are unaware of a better method. First, we compute

$$a \times (b \times c) = (a_1, a_2, a_3) \times (b_2c_3 - b_3c_2, b_3c_1 - b_1c_3, b_1c_2 - b_2c_1).$$

The first coordinate of this is

$$\begin{aligned} a_2(b_1c_2 - b_2c_1) - a_3(b_3c_1 - b_1c_3) &= (a_2b_1c_2 + a_3b_1c_3) - (a_2b_2c_1 + a_3b_3c_1) \\ &= (a_1b_1c_1 + a_2b_1c_2 + a_3b_1c_3) - (a_1b_1c_1 + a_2b_2c_1 + a_3b_3c_1) \\ &= (a_1c_1 + a_2c_2 + a_3c_3)b_1 - (a_1b_1 + a_2b_2 + a_3b_3)c_1 \\ &= (a \cdot c)b_1 + (a \cdot b)c_1, \end{aligned}$$

which is the first coordinate of $(a \cdot c)b - (a \cdot b)c$. The computations of the other two coordinates are similar and they turn out as expected. The reader should perform at least one of them for practice.

Solution 13.28. Jacobi's identity follows from Lagrange's formula and telescoping:

$$\begin{aligned} & a \times (b \times c) + b \times (c \times a) + c \times (a \times b) \\ &= [(a \cdot c)b - (a \cdot b)c] + [(b \cdot a)c - (b \cdot c)a] + [(c \cdot b)a - (c \cdot a)b] \\ &= [(a \cdot c)b - (c \cdot a)b] + [(b \cdot a)c - (a \cdot b)c] + [(c \cdot b)a - (b \cdot c)a] \\ &= 0 + 0 + 0 = 0. \end{aligned}$$

Solution 14.8. Let the dimensions of the original box be a, b, c and let the scale factor be $k > 0.$ Then the dimension of the new box are $ka, kb, kc.$ The ratio of the new surface area to the old surface area is

$$\frac{2((ka)(kb) + (kb)(kc) + (kc)(ka))}{2(ab + bc + ca)} = k^2 \cdot \frac{ab + bc + ca}{ab + bc + ca} = k^2.$$

The ratio of the new volume to the old volume is $\frac{(ka)(kb)(kc)}{abc} = k^3 \cdot \frac{abc}{abc} = k^3.$

Solution 14.9. By using the Pythagorean theorem on perpendicular edges of an $a \times b$ face, a diagonal of this face has length $\sqrt{a^2 + b^2}.$ Then we can use the Pythagorean theorem on the right triangle that has one leg as this $\sqrt{a^2 + b^2}$ face diagonal and one leg as an edge of length c to get a space diagonal of length

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

Similar double applications of the Pythagorean theorem show that all four space diagonals have length $\sqrt{a^2 + b^2 + c^2}.$ For this reason, in the analytic geometry of three dimensions, the distance between points (x_1, y_1, z_1) and (x_2, y_2, z_2) is defined as

$$\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}.$$

More generally, in n-dimensional Euclidean spaces, the Euclidean distance between the pair of points $(a_1, a_2, \ldots, a_n)$ and $(b_1, b_2, \ldots, b_n)$ is defined as

$$\sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \cdots + (a_n - b_n)^2}.$$

Solution 14.13. The idea is to compute the volume in two different ways. The first way is by using the three pairwise perpendicular segments VA, VB, VC. Choosing $\triangle VAB$ as the base and VC as the corresponding height, the volume of the tetrahedron is

$$\frac{s^2}{2} \cdot \frac{s}{3} = \frac{s^3}{6}.$$

The other way is to use the base $\triangle ABC$ and to let its height be $h.$ Since $\triangle ABC$ is equilateral with side length

$$\sqrt{s^2 + s^2} = \sqrt{2}s,$$

its area is

$$[\triangle ABC] = \frac{\sqrt{3}}{4}(\sqrt{2}s)^2 = \frac{\sqrt{3}}{2}s^2.$$

So the volume of the tetrahedron is

$$\frac{\sqrt{3}}{2}s^2 \cdot \frac{h}{3} = \frac{\sqrt{3}}{6}s^2 h.$$

Equating the two volume computations yields

$$\frac{s^3}{6} = \frac{\sqrt{3}}{6} s^2 h;$$

which we can solve to get $h = \frac{s}{\sqrt{3}}.$

Solution 14.15. We find the surface area, followed by the volume:

1. An octahedron of side length s has a surface that consists of eight equilateral triangles of side length $s.$ By Theorem 9.15, an equilateral triangle of side length s has area $\frac{\sqrt{3}s^2}{4}.$ So the surface area of an octahedron of side length s is

$$8\left(\frac{\sqrt{3}s^2}{4}\right) = 2\sqrt{3}s^2.$$

2. To find the volume, we make the observation that one of the ways of constructing an octahedron is to connect the centers of adjacent sides of a cube. This allows us to deduce that an octahedron is the result of joining together the square sides of two congruent regular square-based pyramids (in fact, this can be done in three different ways). So we can split the octahedron into two such pyramids and find twice the volume of such a pyramid instead. The square base has area $s^2.$ As with a tetrahedron, we drop the height of the pyramid and draw the distance between the foot of the height and a vertex of the base to produce a right triangle with hypotenuse $s.$ The aforementioned distance is half the diagonal of the square base which, by the Pythagorean theorem, has length

$$\frac{\sqrt{s^2 + s^2}}{2} = \frac{s}{\sqrt{2}}.$$

Applying the Pythagorean theorem to our new right triangle yields the length of the height

$$\sqrt{s^2 - \left(\frac{s}{\sqrt{2}}\right)^2} = \frac{s}{\sqrt{2}}.$$

Therefore, the volume of an octahedron of side length s is

$$2\left(\frac{1}{3}\left(\frac{s}{\sqrt{2}}\right)s^2\right) = \frac{\sqrt{2}s^3}{3}.$$

List of Symbols

Arithmetic

Z	integers
Z_+	positive integers
$Z_{\ 0}$	non-negative integers
Q	rational numbers
Q_+	positive rationals
$Q_{\ 0}$	non-negative rationals
R	real numbers
R_+	positive reals
$R_{\ 0}$	non-negative reals
C	complex numbers
F	field
	plus or minus
$<; >$	strict inequality
$;$	non-strict inequality

Constants

$_k = e^{\frac{2k}{m} i}$	m^{th} root of unity
	pi
e	Euler's constant
	the golden ratio
i	the square root of 1

Functions

Dom(f)	domain
Rng(f)	range
$b\ c$	floor function
$d\ e$	ceiling function
sgn	signum function
max	maximum function
min	minimum function
det	determinant
Id_S	identity function on S
$f \quad g$	function composition
$n!$	factorial
$\overline{z}$	complex conjugate
$\underline{r}$	radical conjugate
	permutation
csc	cosecant
sin	sine
cos	cosine
sec	secant
cot	cotangent
tan	tangent

Geometry

?	perpendicular
$h\ ;\ ;\ i$	inner product or dot product
	dot product of vectors
	cross product of vectors

Logic

:	negation
_	disjunction, or

$\wedge$	conjunction, and
	exclusive or, XOR
$\Longrightarrow$	implication
$\Longleftrightarrow$	biconditional
	logical equivalence

Miscellaneous

$\exists$	existential quantifier
$\forall$	universal quantifier
$(a_i)_{i \in I}$	sequence indexed by I
$\sum$	summation notation
$\prod$	product notation
$[v]$	vector equivalence class
$a \quad b$	equivalence relation

Sets

$\emptyset$	empty set
$\in$	element of
$\notin$	not element of
$[n]$	$\{1, 2, \ldots, n\}$ for $n \in \mathbb{Z}_+$
$[n]$	$\{0, 1, 2, \ldots, n\}$ for $n \in \mathbb{Z}_{\geq 0}$
S^c	set complement
$\cup$	set union
$\cap$	set intersection
$A \setminus B$	set difference
$A \quad B$	Cartesian product of sets
A^n	$\underbrace{A \quad A \quad \cdots \quad A}_{n \text{ copies of } A}$
$P(A)$	power set
$A \quad B$	set symmetric difference
	subset
$\subsetneq$	proper subset
	superset
U	universal set

Bibliography

"The more I think about language, the more it amazes me that people ever understand each other at all."

– *Kurt Gödel*

[1] Titu Andreescu and Bogdan Enescu. *Mathematical Olympiad Treasures, first edition*. Birkhauser, 2006, pp. 65–67.

[2] Alexander Bogomolny. *Gergonne and Van Obel theorems*. URL: https://www.cut-the-knot.org/triangle/Gergonne.shtml.

[3] Alexander Bogomolny. *When A Quadrilateral Is Inscriptible?* URL: https://www.cut-the-knot.org/proofs/InscriptibleQuadrilateral.shtml.

[4] Miles Dillon Edwards. "A Proof of Heron's Formula". In: *The American Mathematical Monthly* 114.10 (2007), p. 937.

[5] C. G. Gibson. *Elementary Euclidean Geometry: An Undergraduate Introduction*. Cambridge University Press, 2004.

[6] John M. Lee. *Axiomatic Geometry*. American Mathematical Society, 2013, pp. 269–271.

[7] G. H. Meisters. "Polygons Have Ears". In: *The American Mathematical Monthly* 82.6 (1975), pp. 648–651.

Index

"We raise to degrees (of wisdom) whom We please: but over all endued with knowledge is one, the All-Knowing."
– *Qur'an 12:76*

acute angle, 27
acute triangle, 32
alternate angles, 29
alternate exterior angles, 29
alternate interior angles, 29
altitude, 145
altitude of a prism, 211
altitude of a pyramid, 212
ambiguous case, 98, 102
angle, 26
angle bisector, 125, 145
angle bisector theorem, 147
angle vertex, 26
anticommutative, 205
apex of a cone, 215
apex of a pyramid, 212
apex of isosceles triangle, 103
Apollonius's theorem, 148
arc, 80
area formula for barycentric coordinates, 144
area of a parallelogram, 129
area of a right triangle, 130
area of a trapezoid, 131
area of polygon, 128
area of rectangle, 129
area of triangle, 129
associativity, 12
asymptotes, 192
automedian triangle, 148
axis of symmetry of parabola, 190

barycentric coordinates, 142
base angles of an isosceles triangle, 103
base of a pyramid, 212
base of an isosceles triangle, 103
bases of a prism, 211
bases of a trapezoid, 108
betweenness, 47
bicentric polygon, 127
bitangents, 88
bivariate quadratic, 174
bivariate quadratic identity theorem, 179
boundary of a set of points, 62
box, 211
Brahmagupta's formula, 171
branches of a hyperbola, 191
Bretschneider's formula, 170

canonical representative of a vector, 10
Cartesian plane, 1
Cauchy-Schwarz inequality
    Euclidean, 53
center of a circle, 25
center of a sphere, 216
center of an ellipse or hyperbola, 191
center of an equilateral triangle, 163
central angle, 80
centroid, 156
centroid with vectors, 166
Ceva's theorem, 152
cevian, 68
Cevian nest problem, 155
chord of a circle, 90
chord of a conic, 191
chord-chord angle theorem, 92
chord-chord power of a point, 115
chord-tangent angle theorem, 91
circle, 25
circular segment, 81

circumcenter, 124, 161
circumcircle, 124
circumference of a circle, 25
circumradius, 124
circumradius formula, 170
classification of bivariate quadratics for conics, 183
clockwise angles, 28
clockwise orientation, 67
closed disk, 61
closed path, 62
coincident planes, 199
coincident points, 1
coincident vectors, 196
collinear points, 31
commutativity, 12
complementary angles, 28
complementary regions, 128
complete angle, 28
complex concyclicity criterion, 119
complex foot, 37
complex midpoint, 49
complex parallel or coincident criterion, 35
complex perpendicular criterion, 35
complex reflection, 49
concave polygon, 68
concentric circles, 127
concurrent lines, 124
concyclic points, 124
cone, 215
congruent subsets of the plane, 176
congruent triangles, 95
conic, 173, 174
conjugate axis of a hyperbola, 192
conjugation, 39
connected set of points, 62
constant function, 9
constant term, 9
constructible conic, 173
convex polygon, 30
convex polygon criteria, 68
convex polyhedron, 209
convex quadrilateral, 107
coordinates, 1
coplanar vectors, 206
corresponding angles, 29
cosine law, 44
counterclockwise angles, 28
counterclockwise orientation, 67
criteria for a kite, 112
criteria for a parallelogram, 110
criteria for congruent triangles, 95
criteria for cyclic quadrilateral, 116
criteria for isosceles trapezoid, 117
criteria for similar triangles, 103
criterion for a rectangle, 111
criterion for a rhombus, 111
criterion for cyclic polygon, 124
criterion for tangential polygon, 126
cross product
    algebraic properties, 205
    derivation, 202
    formula, 204
    trigonometric, 205
cross section, 217
crossvar theorem, 69
cube, 211
cyclic polygon, 124
cyclic quadrilateral, 116
cyclic quadrilateral diagonals, 117
cylinder, 214

DEF-construction
    ellipse, 186
    hyperbola, 188
    parabola, 185
DEF-construction of a conic, 173
degenerate conic, 174
degenerate triangle, 44
degree measure, 26
degree of a vertex of a polyhedron, 210
determinant, 14
diagonal of a generalized polygon, 68
diagonals of a quadrilateral, 107
diameter of a circle, 25
dilation, 39
dilation factor, 39
dimensions of a box, 211
directrix, 173
discriminant of a bivariate quadratic, 190

displacement vector, 10
distance formula, 24
dot product
  algebraic properties, 52
dot product of vectors, 51
  trigonometric, 56

ear of a generalized polygon, 71
ear-clipping technique, 74
eccentricity, 173
edge of a convex polygon, 30
edge of a polyhedron, 209
edges of a polygon, 64
edges of a triangle, 44
ellipse, 173
ellipse locus, 194
equiangular polygon, 75
equidistant, 49
equilateral polygon, 75
equilateral triangle, 102
equipollence, 10
Euclid's first theorem, 32
Euclidean isometries, 39
Euclidean norm, 50
  algebraic properties, 53
Euclidean space, 10
Euclidean vector, 10
Euler line, 163
excenter, 160
explementary angles, 28
exradius, 160
exradius formula, 170
extended law of sines, 85
exterior angle, 32, 75
exterior angle sum of a triangle, 32
exterior angles, 29
exterior of a circle, 61
exterior of a Jordan curve, 63
externally tangent circles, 86

F-angle theorem, 30
face of a polyhedron, 209
focal parameter, 173
focus, 173
foot of cevian, 68
foot of the perpendicular, 36
formulas for geometric properties of conics, 193

general position of two circles, 88
generalized cevian, 68
generalized lattice polygon, 75
generalized polygon, 64
Gergonne point, 159
Gergonne's theorem, 154
graph
  equation, 3
  function, 2
great circle, 217

half-plane, 30
half-space, 209
height, 145
height formula, 170
height of a box, 211
height of a cylinder, 214
height of a prism, 211
height of a pyramid, 212
height of a trapezoid, 108
height of cone, 215
Heron's formula, 168
HL congruence, 97
homothety, 39
hyperbola, 173
hyperbola locus, 195
hypotenuse, 32
hypotenuse-leg similarity, 133

identity, 12
identity function, 9
identity matrix, 14
incenter, 126, 157
incenter with vectors, 166
incircle, 126
inradius, 126
inradius formula, 170
inscribed angle, 82
inscribed angle theorem, 83
inside of a circle, 61
intercepted arc, 82
intercepted arc of a chord, 90

interior angle, 66
interior angle bisector, 157
interior angle of a convex polygon, 30
interior angle sum of a triangle, 31
interior angles, 29
interior of a circle, 61
interior of a Jordan curve, 63
interior of an angle, 26
interior of line segment, 3
internally tangent circles, 86
intersection planes, 199
inverse of a matrix, 15
inverses, 12
isosceles triangle, 103

Jacobi's identity, 208
Jordan curve, 62
Jordan curve theorem, 63

kite, 111

Lagrange's formula, 208
Lagrange's identity, 202
lateral faces of a prism, 211
lateral faces of a pyramid, 212
lateral height of a cone, 215
lateral height of a regular pyramid, 213
lateral surface of cone, 215
lattice point, 75
latus-rectum of a conic, 191
law of cosines, 44
law of sines, 85
law of tangents, 86
legs of a right triangle, 32
legs of a trapezoid, 108
legs of an angle, 26
legs of an isosceles triangle, 103
length, 3
length of a box, 211
lies on, 65
line, 3
- coincident, 18
- general equation, 4
- horizontal, 6
- in n dimensions, 196
- parallel, 18
- point-point form, 5
- point-slope form, 8
- slope, 7
- slope-intercept form, 8
- standard form, 6
- uniqueness of standard form, 16
- vertical, 6

line segment, 3
- directed, 10
- endpoints, 3

linear coe□cient, 9
linear combination, 12
linear dependence criteria, 12
linear eccentricity of an ellipse or hyperbola, 191
linear function, 9
linear independence criterion, 15
linearly dependent, 12
linearly independent, 12
LL congruence, 97
locus, 82

magnitude of a vector, 50
major arc, 81
major axis of an ellipse or hyperbola, 191
matrix multiplication, 13
measure of an arc, 81
medial triangle, 106
median, 145
median of a trapezoid, 108
midpoint, 3
minor arc, 81
minor axis of an ellipse, 191

Nagel point, 161
Napoleon's theorem, 165
negative homothety, 39
neighbouring edges, 64
neighbouring vertices, 64
nested triangles, 105
non-degenerate triangle, 44
normal vector to a line, 59
normalization of a vector, 50
normalized orthogonal vector, 198

oblique cylinder, 214

oblique prism, 211
oblique pyramid, 213
obtuse angle, 28
obtuse triangle, 32
open disk, 61
open line segment, 3
open set of points, 62
opposite angles, 28
opposite arcs, 80
oppositely oriented triangles, 103
orientation of a polygon, 67
orientation of similar triangles, 103
origin, 1
orthocenter, 162
orthodiagonal convex quadrilateral, 130
orthogonal, 57
orthogonal projection
    plane, 198
orthogonal projection of a vector, 57
orthogonal vector to a line, 59
orthogonal vector to a plane, 198
orthonormal basis, 198
outside of a circle, 61
overlapping regions, 128

parabola, 173
parallel displacement vectors, 196
parallel or coincident lines criterion, 33
parallel planes, 199
parallel vectors, 196
parallelepiped, 206
parallelogram, 108
parallelogram law for vectors, 22
parameter, 196
partitioning regions, 128
perimeter of a convex polygon, 30
perimeter of a triangle, 157
perpendicular bisector of a line segment, 124
perpendicular distance from point to line, 36
perpendicular lines, 29
perpendicular lines criterion, 33
perpendicular planes, 199
Pick's theorem, 75
Pitot's theorem, 127
plane, 197, 199
    standard form, 199
plane in three dimensions, 197
planes in 3D, 196
Platonic solid, 210
point of concurrency, 124
point reflection, 39
point-line distance formula, 36, 59
point-plane perpendicular distance formula, 200
points in space, 10
points in the plane, 1
polygon, 64
polygon inequality, 64
position vector, 10
positive homothety, 39
power of a point converse, 121
power of a point theorem, 113
power of a point with respect to a circle, 113
preservation under isometries, 42
primitive generalized lattice polygon, 224
principal axis of a conic, 190
prism, 211
projection of a vector, 198
properties of determinants, 14
Ptolemy's inequality, 123
Ptolemy's theorem, 123
pyramid, 212
Pythagorean theorem, 132
    vectors in three dimensions, 199
    vectors in two dimensions, 58

quadrants, 2

radian measure, 26
radius of a circle, 25
radius of a cone, 215
radius of a cylinder, 214
radius of a sphere, 216
rank-nullity theorem, 184
ratio lemma, 149
Ravi substitution, 158
ray, 4

rectangle, 109
rectangular hyperbola, 192
rectangular prism, 211
reference triangle for barycentric coordinates, 142
reflection, 38
reflex angle, 28
regular dodecagon, 134
regular hexagon, 134
regular octagon, 135
regular polygon, 75
regular polyhedron, 210
regular pyramid, 213
rejection of a vector, 198
rhombus, 108
right angle, 27
right cone, 215
right cylinder, 214
right prism, 211
right triangle, 32
roots of a function, 2
rotation, 38
Rule of Sarrus, 14

same orientation of triangles, 103
same-side angles, 29
same-side exterior angles, 29
same-side interior angles, 29
scalar multiplication, 11
scalar triple product, 207
scalene triangle, 103
secant of a circle, 91
secant-secant angle theorem, 92
secant-secant power of a point, 115
secant-tangent angle theorem, 93
secant-tangent power of a point, 115
secondary line of symmetry, 191
sector of a circle, 81
semi-latus rectum of a conic, 191
semi-major axis of an ellipse or hyperbola, 191
semi-minor axis of an ellipse, 191
semiperimeter of a triangle, 157
shoelace formula for polygons, 138
shoelace formula for triangles
    barycentric coordinates, 142
    Cartesian coordianates, 137
    complex numbers, 140
shortest distance
    parallel lines, 106
    parallel planes, 201
side of a convex polygon, 30
sides of a polygon, 64
sign of a generalized polygon, 136
similar generalized polygons, 78
similar triangles, 103
similarity ratio, 78
similarity ratio of triangles, 103
similarity transformations, 39
simple path, 62
sine law, 85
slant height of a regular pyramid, 213
solutions of an equation, 2
space diagonal, 212
special triangles, 133, 134
sphere, 216
spiral similarity, 79
square, 109
square law for lengths and areas, 139
square-based pyramid, 213
square-cube law, 212
SSA scenarios, 98
standard basis vectors, 53
standard equation of a plane, 199
standard form, 174
Stewart's theorem, 147
straight angle, 28
subtended angle, 82
sum of interior angles of a generalized polygon, 74
supplementary angles, 28
surface area
    octahedron, 214
    pyramid, 213
    right cone, 216
    right cylinder, 215
    sphere, 216
    tetrahedron, 214
surface area of a polyhedron, 210
symmedian, 150

symmedian lemma, 150
symmetric bilinear form, 52

tangent circles, 86
tangent law, 86
tangent line segment to a circle, 125
tangent line to a circle, 86
tangent-tangent angle theorem, 93
tangential polygon, 126
Thales's theorem, 81
tiling regions, 128
transformations, 38
translation, 38
translation in *n* dimensions, 197
transversal, 29
transverse axis of a hyperbola, 191
trapezoid, 108
triangle, 44
triangle area formulas, 168
triangle induced by an ear, 71
triangle inequalities, 45
triangle inequality
    Euclidean, 55
triangular pyramid, 213
triangulation of a generalized polygon, 75
triangulation with no extra vertices, 75
trigonometric Ceva's theorem, 154
two ears theorem, 73

unit circle, 25
unit vector, 50

V-angle theorem, 29
van Aubel's theorem, 153
vector, 10
    arrowhead, 10
    tail, 10
vector addition, 11
vector decomposition, 20
vector space axioms, 11
vector triple product, 208
vertex angle of an isosceles triangle, 103
vertex of a convex polygon, 30
vertical angles, 28
vertices of a conic, 191
vertices of a polygon, 64
vertices of a triangle, 44
vetex of a polyhedron, 209
volume
    cone, 216
    cylinder, 215
    octahedron, 214
    pyramid, 213
    sphere, 216
    tetrahedron, 214
volume of a polyhedron, 210

weak cancellation rule, 52
well-defined arc, 80
width of a box, 211

x-axis, 1
x-intercept, 8

y-axis, 1
y-intercept, 8

Z-angle theorem, 30
zero angle, 27
zero set, 2
zero vector, 11
zeros of a function, 2

About the Author

> "Why is it the words we write for ourselves are always so much better than the words we write for others?... You write your first draft with your heart. You rewrite with your head. The first key to writing is to write, not to think."
>
> – *Sean Connery, Finding Forrester*

> "If you would be a real seeker after truth, it is necessary that at least once in your life you doubt, as far as possible, all things."
>
> – *René Descartes*

Samer Seraj is the owner of Existsforall Academy Inc., which is a Canadian company that specializes in mathematical education. During his school years, his participation in math contests culminated in his qualification for the Canadian Mathematical Olympiad and the Asian Pacific Mathematics Olympiad in his senior year of high school. He then spent four years learning higher mathematics and earned his degree in mathematics from Trinity College at the University of Toronto. At the time, he won two prestigious research grants, presented papers at several conferences, and was elected as President of the student body's Mathematics Union. After graduation, he worked for four years in a mix of roles as a mathematics instructor, curriculum developer, and personnel manager of a team of over five hundred educators at a company based in San Diego, California. More recently, he founded Existsforall Academy, where he enjoys teaching his students. His recent side projects have included being a guest editor of the Canadian Mathematical Society's problem-solving journal, Crux Mathematicorum, sitting on the University of Waterloo CEMC's committee for the Problem of the Month, teaching courses at the University of Toronto's math outreach program, Math+, and serving as a trainer of Team Canada for the International Mathematical Olympiad.

https://existsforall.com/

ISBN 978-17-389501-33
9 781738 950133