Deep Thoughts on Area
(Version 4.0)

Jim Farmer

21 August 2024

Contents

1 Introduction

This material was originally formatted as a series of web pages. Web support for mathematical typing and complex diagrams is poor and it was proving difficult to get a good quality appearance that was consistent across browsers and operating systems. I've converted it to a pdf document created using the LaTeX software.

This is a work in progress. It probably contains typographical errors.

1.01 Why I wrote this book

Syllabus contents can vary by country, and even by state, territory or province within a country. You might find the following comments don't apply in your region, but when it comes to geometry, many regions follow the following pattern.

Many useful theorems are provided to students before they learn how to prove them. For example, students may learn methods to calculate the areas of rectangles and triangles before they have the skills to prove those methods correct.

In a later school year, students learn how to construct geometric proofs. However, often the syllabus does NOT require students go back and prove the formulae for calculating areas that they were given in earlier years. Rather, it moves onto proving new properties that students have never seen before. Common exercises here include proving many esoteric properties of circles, such as the Alternate Angle Theorem, and the relationship between angles at the centre and circumference subtended by the same arc.

The result of this is that students learn how to prove several esoteric results that most will never see again after they leave school, let alone use in a practical situation. But they never see a formal proof of many basic area properties they they may use frequently after leaving school.

This book tries to fill this gap, by providing formal proofs of all the common area formulae.

This book does not include any results requiring trigonometry. By the time students encounter trigonometry they have learned about how to formally construct proofs, so when textbooks present the area formulae that employ trigonometry they do formally prove them. Rather, this textbook tries to collect together all the classical area proofs that the syllabus skips over in many regions.

1.02 Who is this book for?

This book is for students who want to know why things work. If you get irritated if your teacher hands you a formula and expects you to trust them when they tell you it works, then hopefully you'll enjoy this book.

The book is for adventurous students who don't mind having advanced concepts thrown at them. If you just want to do the minimum required to pass an exam, this book isn't for you.

This book is for students who think they might want to be a professional mathematician, or work in one of the many related fields that heavily rely on mathematics, such as statistics, engineering, physics, actuarial studies or epidemiology. In all these mathematical fields, it is important that you can think like a mathematician. Whenever the opportunity arises, this book tries to demonstrate useful techniques that mathematicians employ to prove interesting results.

1.03 How to use this book

Most geometry textbooks feature tedious repetitive exercises.

For example, if a textbook has a section that explains the formula for calculating the area of a rectangle, it's likely to follow this with a set of about ten exercises which give you the lengths of two adjacent sides of a rectangle and set you the task of calculating the area. It might follow this with a few interesting harder questions, but most of the problem set involves the mindless repetitive task of multiplying two numbers together and writing down the answer, with the appropriate units attached.

If you find this approach incredibly boring, I have good news. I'm not going to do that. This book has no repetitive exercises. It only contains deep challenging questions.

> I refer to them as challenges, and they appear in shaded boxes like this one.

All the challenges are followed by a page break. When you think you have solved the challenge, turn the page and check your answer.

These challenges are not like the repetitive questions mentioned above where your brain is figuring out the answers faster than your hand can write them down. Many of these questions require deep thought, so don't panic if you've been staring at a problem for five minutes without making any progress. School mathematics tests seem to place an unhealthy emphasis on being able to perform routine tasks quickly and accurately. By contrast, real life mathematical work involves tackling complex problems that take time to figure out.

When totally stuck on a challenge, some people find it useful to take a rest from mathematics and go and do something else for an hour or so. Do some work on some other subject. Get some exercise. When you return to the problem after a break, sometimes you immediately see a new way to approach it, almost as if your brain was still thinking about it on some subconscious level during the break and came up with some new ideas that it was ready to hand back to the conscious part of your brain whenever you eventually returned to the problem.

1.04 There is some assumed knowledge

A geometry textbook tries to present ideas and proofs in a logical sequence. Once a theorem has been proved true, it can be used as an input to other proofs appearing later in the book. To state this another way, in a geometry textbook, a proof can use results proved earlier in the book.

But this book is only about area. It (mostly) doesn't include proofs of general geometric properties that don't involve calculating areas. This creates a problem. It isn't clear where these area concepts would occur in a general geometry textbook, so it's not clear what other theorems and definitions we can assume as already known.

So when I challenge you to prove something, sometimes I'll need to indicate that that some particular geometric result can be assumed to be already proved.

2 Area and perimeter

2.01 What is area?

A two dimensional shape is one which can be drawn on a plane, but not on a line. Planes have two dimensions while lines only have one. Two dimensional shapes include triangles, squares, quadrilaterals, pentagons and circles.

Area is a quantity which answers the question: How big is the region of the plane that falls inside this two dimensional shape?

Area can be used to answer questions like:

- How many carpet tiles of a known fixed size do I need to cover a floor? The larger the floor area, the more tiles I need.
- How much paint do I need to paint a room? I could just buy a large tin, start painting, and see what happens. But I might find that I've got lots of paint left over when I finish, which means I wasted money buying too much paint. Alternatively, I might run out of paint before I finish, which means I have to make an extra trip to the store to buy more. Ideally I calculate how much paint I need in advance. I can do this by calculating the areas of the walls and ceiling that need painting. The larger that area is, the more paint I need. The information shown on a tin of paint will usually include a statement of how much area it will cover, so I can buy the right number of tins of the right size. The room is a three dimensional shape, but each individual wall is a two dimensional shape, as is the ceiling.
- How large does a computer monitor need to be to contain a certain number of pixels of a certain size? If the size of a pixel is fixed, a larger monitor can display more pixels.
- How much seed do I need to buy if I'm planting crops in a field. The larger the area of the field, the more seed I need. Agricultural scientists do experiments to determine the optimum amount of seed to use per unit area of field. That amount varies by crop (wheat, barley, etc) and can also vary by other properties, such as the expected amount of rainfall for the field or the volume of water available for irrigation, and the latitude of the farm, which affects how much solar energy reaches the field.

2.02 What is perimeter?

The perimeter of a two dimensional shape is the length of its outline.

The perimeter of a polygon is simply the sum of the lengths of the line segments that are its edges.

If you have a rectangle field, where its area can tell you have much seed to buy to sow crops in the field, its perimeter can tell you how much fencing wire you need to build a fence around the border of the field.

The term "perimeter" also applies to shapes involving curves, such as circles and ellipses. The perimeter of a circle is called its circumference.

To measure the perimeter of a polygon, we can use a ruler to measure each edge add up the results. Rulers aren't so useful when it comes to measuring distances along curves. We need a flexible ruler. Tape measures are rulers printed on cloth. If a tailor is making a pair of pants to fit a particular person they can use a tape measure to find that person's hip and waist measurements. To measure your waist, you can wrap a tape measure around your body at the narrowest point between your hips and chest.

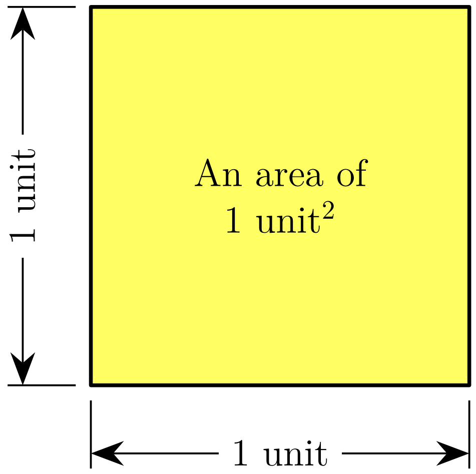
A circle is a two dimensional shape, so it's not possible to wrap a tape measure around it, so let's instead try a three dimensional shape that has a circular cross section. Go to your pantry or refrigerator and find a cylindrical can, perhaps a can of soup or tinned fruit. The can has a circular cross section. You can measure the circumference of the circle by wrapping the tape measure around the tin. If you don't have access to a tape measure, an alternative is wrap a piece of string around the tin, mark a point on the string to indicate the distance around the can, and then hold the string against a ruler to read off the distance.

If you try these experiments you will find that the perimeter of the circle that is the cross section of the tin is a little over three times its diameter.

Books dealing with area have to mention perimeter at some point because there are simple formulae that link the area of regular polygons and circles to their perimeter. There is also a formula for the area of a triangle that refers to its perimeter.

3 One Square Unit

3.01 Definitions



Definition: A unit square is square in which the length of every side is one unit.

Definition: The area of any unit square is defined to be one square unit, which is commonly written as 1 unit^2.

'Unit' is a placeholder term used to indicate that the definition works for any unit that you require. For example:

- A square with side length one metre (abbreviation: 1m) has an area of one square metre (1m^2).
- A square with side length one centimetre (1cm) has an area of one square centimetre, (1cm^2).
- A square with side length one kilometre (1km) has an area of one square kilometre, (1km^2).

This book uses the metric system. For those unfortunate enough to be using other systems of measurement, similar results apply. For example, a square with side length one inch has an area of one square inch.

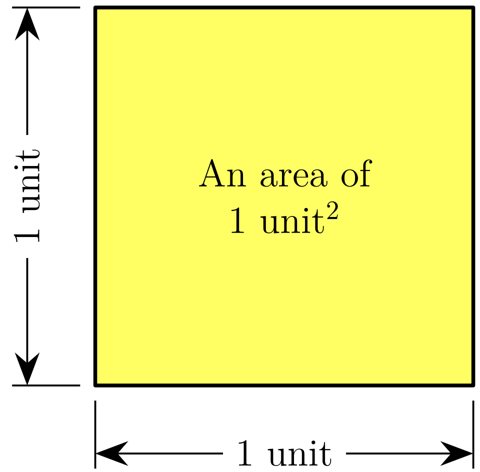
Using this definition of one square unit as our starting point, we will develop formulae for the areas of more complex shapes, starting with rectangles, and then moving on to triangles and various special quadrilaterals.

3.02 Does the area include the line segments?

In this book, I often repeat diagrams, so that you can see everything you need to understand the text on the page without having to flip back to an earlier page. But not always. Sometimes it's just not possible to squash all the relevant ideas into a single page and still have room for the diagrams, so later chapters will sometime require some page flipping to check earlier diagrams.

In this diagram I drew the unit square in black and coloured the region inside it yellow. Students sometimes ask, when we refer to "the area of the unit square," do we mean only the area of the yellow region *inside* the square, or do we also include the area of four black line segments that form the square?

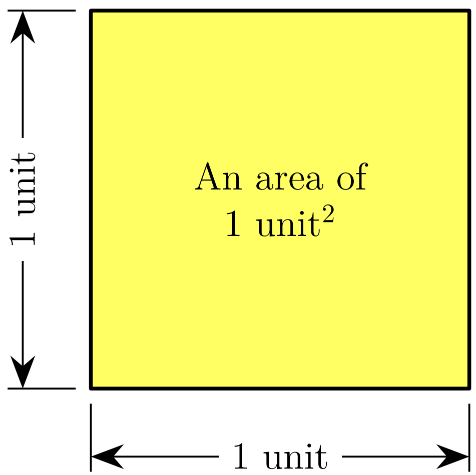
Perhaps another issue which prompts this question is that textbooks can sometimes be inconsistent in how they phrase references to area. You might find a book sometimes says "the area of a square" but at other times says "the area within a square," so it's natural to wonder whether these mean different things, with the first version including the four black line segments and the second version excluding them.



Does the area of the square include the area of the black line segments? This question can be answered using one of the properties of lines and line segments. Can you spot the relevant property and answer the question?

Lines and line segments have no thickness. We can't actually draw a line segment, because any writing implement we use produces marks that do have thickness. If the marks had no thickness we wouldn't be able to see them! We might draw a straight mark on a page and call it a line segment, but really we should be calling it a imperfect representation of a line segment.

When I created this image of the unit square, I made the edges 0.5mm wide, though they they may look different to you depending due on the magnification factor in your software and the pixel resolution of your screen. If I had given them zero width, they would be invisible. Again, they are not true line segments. They are a visible and hence imperfect representation of an invisible line segment.



Since lines and line segments have no thickness, they have no area, so we don't need to worry about whether the area of the square includes or excludes the area of the four line segments. Since line segments have no area, we would get the same result for the area either way.

Strictly speaking, "the square" means the four line segments, so we really should say "the area inside the square." Textbooks often abbreviate this to "the area of the square." This is lazy, but doesn't introduce any ambiguity, so it is a common practice and one which I will follow.

4 Properties of Area

If you are not yet familiar with the difference between a "postulate" and a "theorem", this is a good time to read Appendix B, which explains the meanings of these two words.

4.01 Congruent figures have equal area

In the previous chapter I defined one square unit to be the area of any unit square. This definition includes an implicit assumption, which is that all squares with a particular side length have equal area. If two squares both with a side length of one metre could have different areas, then this definition would make no sense!

Most traditional textbooks would have stated this assumption *before* giving the definition of one square unit. I didn't, because I wanted to be able to ask: When you read the definition, did you *notice* it made this implicit assumption?

Most readers don't notice the assumption, because they simply regard it as *obvious* that two squares with the same side length should have equal area. I hope you'll accept this property as self-evident, because I can't prove it. It is not a theorem. It is a postulate, or to be more precise, it is one particular case of a more general postulate.

I also hope you will regard it as self-evident that:

- Two equilateral triangles both with side length 3cm will have equal area.
- Two rectangles both with dimensions 5 metres by 2 metres will have equal area.

The more general postulate covers all these cases.

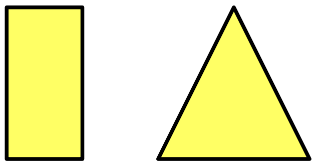
Postulate 4.1: Congruent figures have equal area.

I'm going to write a bit more about why this postulate is plausible, but this only constitutes supporting evidence, not proof. Theorems are proved; postulates are not.

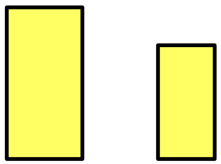
Definitions of congruence usually either refer to shape and scale, or refer to transformations.

Two geometric figures are congruent if they have the same shape and scale.

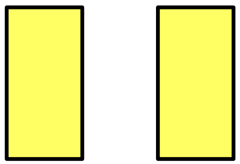
These two figures have different shapes.



These two figures have the same shape but different scale.



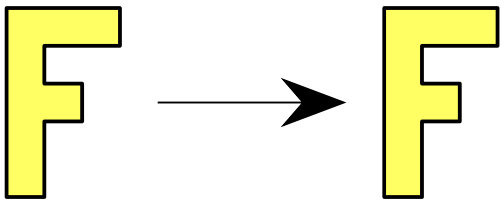
These two figures have the same shape and scale and hence are congruent.

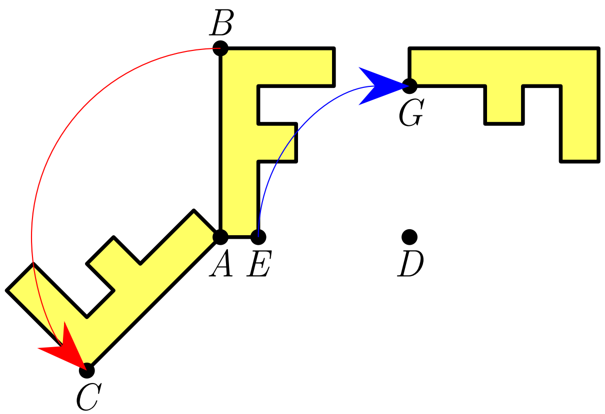


Alternatively, two geometric shapes are congruent if we can move one of them to make it exactly coincide with the other. If they can be made to exactly coincide, it sounds reasonable that all their properties would match exactly, including their area. "Moving" a figure means we can employ any or all of the three transformations that preserve shape and scale: translation, rotation and reflection.

For example, if we draw say a capital F, and then create a copy of it by moving the original some fixed distance to the right, the original and the copy are congruent.

We would expect these two figures to have equal area. If you had to paint the capital F, it wouldn't matter where you place it on the page. It would still require the same amount of paint.





Similarly, if we draw a capital F, and then create a copy of it by rotating the original around a point, the original and the copy are congruent. The diagram shows two different rotations of the original upright F.

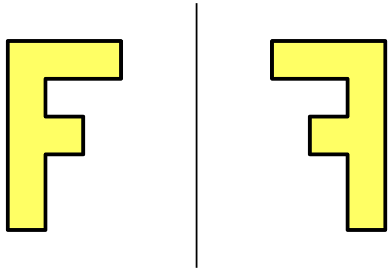
The F is rotated 135° anticlockwise around the point A, which is the bottom left corner of the F. Hence the original figure and this transformation share a common corner at the point of rotation A. The red arc shows the point B at the top left corner of the F, rotating around the point A to become the point C.

The F is also rotated 90° clockwise around the point D. The blue arc shows the movement of the point E around the point D to become the point G.

These three congruent capital Fs will have equal area. If you had to paint the capital F, it wouldn't matter how you orient it on the page. It would still require the same amount of paint.

In the previous diagram the two rotations shown were chosen to ensure the three figures don't overlap. This was done for convenience, to ensure the three figures can all be clearly seen. If I had put the point of rotation inside the original F, that F and its transformation would still be congruent and have equal area, but this would be harder to see this due to the figures overlapping.

A figure is also congruent with its mirage image. Here I have reflected the F in the vertical line. The original and the mirror image are congruent.

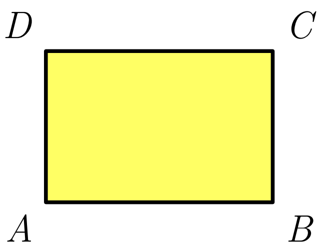


Imagine you copy the outline of the F figure onto a sheet of transparent plastic. If you turn the sheet over, you will see the back to front F which results from a reflection. If I now asked you to paint the F yellow, it should take the same amount of paint no matter which side of the plastic you chose to paint. That is, it would take the same amount of paint to fill in a normal F and a mirror image F, so the original F and its mirror image should have equal area.

Incidentally, I chose to use capital F as the figure in these examples because it is not symmetrical, so we can easily tell the original and its reflection apart.

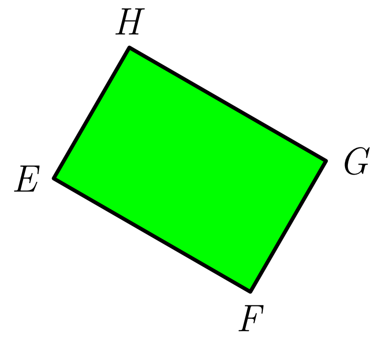
The postulate states that congruent figures have equal area. An equivalent statement is that area is invariant, meaning unchanged, under the transformations of translation, rotation and reflection. That is, subjecting a figure to any sequence of translations, rotations or reflections will not change the area of the figure.

This means that when we need to find the area of a particular figure we can move the figure to a more convenient position and orientation, because doing so does not change its area.



For example, if we are asked to derive a formula for the area of a rectangle with dimensions m units by n units, we can simply choose to draw it in whatever position and orientation we find most convenient. We could choose to draw it with its edges parallel to the edges of the page we are using, like the yellow rectangle $ABCD$ shown here.

Any formula we derive for the yellow rectangle will also work for a rectangle with the same dimensions drawn in any other orientation, because all rectangles with those particular dimensions are congruent with each other and thus have equal area. For example, it will also work for the green rectangle $EFGH$. In terms of transformations, if we rotate the green rectangle 30° anticlockwise, it will have the same orientation as the yellow rectangle, and then it will be possible to translate the green rectangle so it exactly coincides with the yellow rectangle.



Since I have mentioned three of the four simple transformations, I should also mention the fourth: scaling. Area is NOT invariant under scaling. Scaling transformations are split into expansions and contractions.

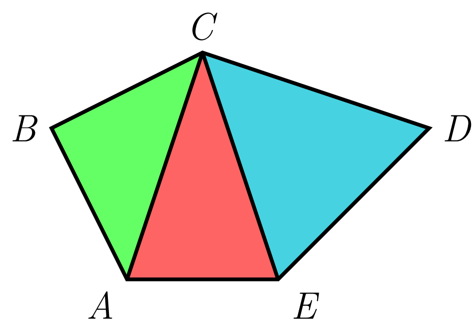
- Expansions will make a figure bigger, increasing its area.
- Contractions will make a figure smaller, reducing its area.

4.02 The Area Sum Postulate

If we divide a region into smaller pieces, it seems sensible to suggest that the area of the original region will be equal to the sum of the areas of the smaller pieces.

For example, here is a pentagon divided into three triangles. It feels reasonable to say that the area of the pentagon is equal to the sum of the areas of the three triangles.

When we talk about dividing a region, we mean the resulting smaller regions should not overlap, other than perhaps for common edges. For example, we could not divide the pentagon $ABCDE$ into the two quadrilaterals $ABCE$ and $ACDE$, since they overlap. If we summed the areas of these two quadrilaterals we would be counting the area of $\triangle ACE$ twice, so that would not be a valid way to find the area of the pentagon.



By contrast, overlap is not an a problem if we sum the areas of the three triangles. The only overlap between $\triangle ABC$ and $\triangle ACE$ is the line segment AC. We don't have to worry about whether summing the area of the two triangles will double count the area of line segment AC because line segments have no area. Similar comments apply to $\triangle ACE$ and $\triangle CDE$, which have overlap in the line segment CE.

Here is a formal statement of our postulate.

Postulate 4.2 — The Area Sum Postulate: If a region is divided into two or more smaller regions, the area of the original region is equal to the sum of the areas of the smaller regions.

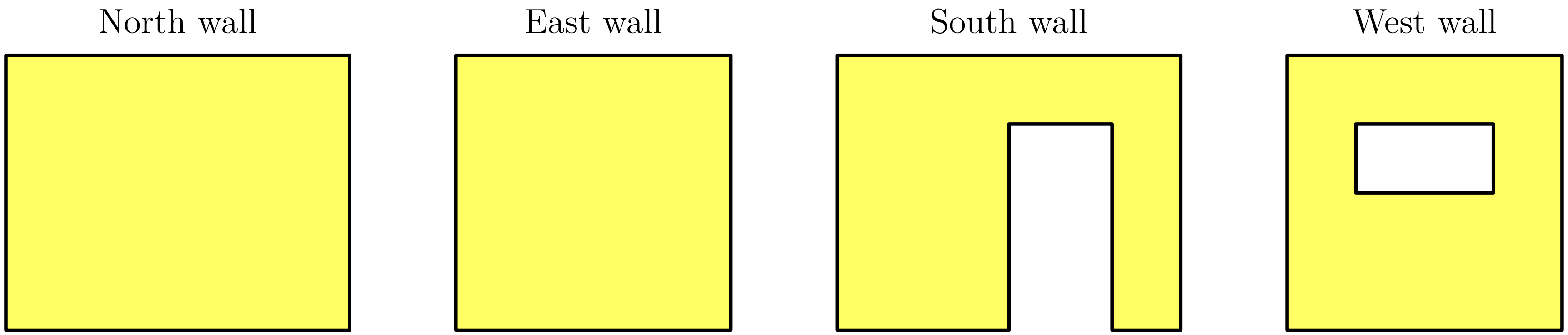
Most school geometry textbooks that take an axiomatic approach use a postulate like that shown above, and that is the approach I will be following in this book. However, a few adopt a different approach, using a weaker version of the above postulate that refers to subdividing a region into exactly two regions. They then use that weaker postulate to prove a corresponding theorem that covers subdivision into two or more regions.

If you are interested in this alternative method, you can find it in Appendix D. The proof uses the Principle of Mathematical Induction. If you are familiar with mathematical induction, you can go straight to Appendix D. If not, I suggest you ignore Appendix D for now. Chapter 4 includes an introduction to mathematical induction in a simpler scenario. Appendix D will make more sense after you master Chapter 4.

4.03 Applications

Let's look at some simple examples that use the Area Sum Postulate.

This diagram shows the four walls of a room, currently painted yellow. I want to paint them some less obnoxious colour. The South wall has a door, which does not need repainting, so I exclude the door when calculating how much paint we need to paint the walls. That is, the area of the South wall does not include the doorway. The West wall contains a window, which obviously I do not want to paint. That is, the area of the West wall does not include the window.



The 4 walls all have the same height. The floor and ceiling are rectangles, so the North and South walls have the same width as each other and the East and West walls have the same width as each other.

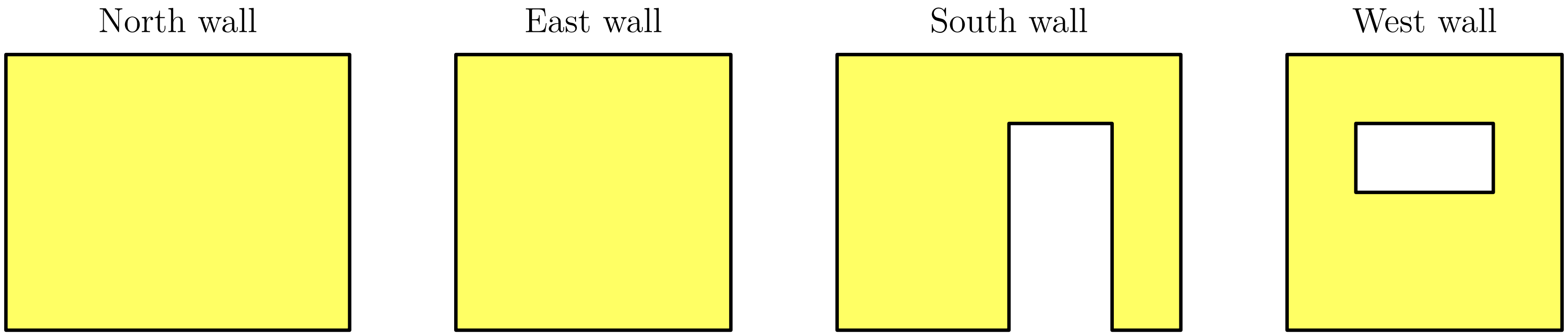
For each of the following pairs of walls, state which wall requires more paint. Explain your reasoning. Rather than just saying something like "This one *looks* bigger", try to explain it using the Area Sum Postulate.

(a) North wall or South Wall?

(b) East wall or West wall?

(c) North wall or East Wall?

Why is it difficult to compare the East and South walls?



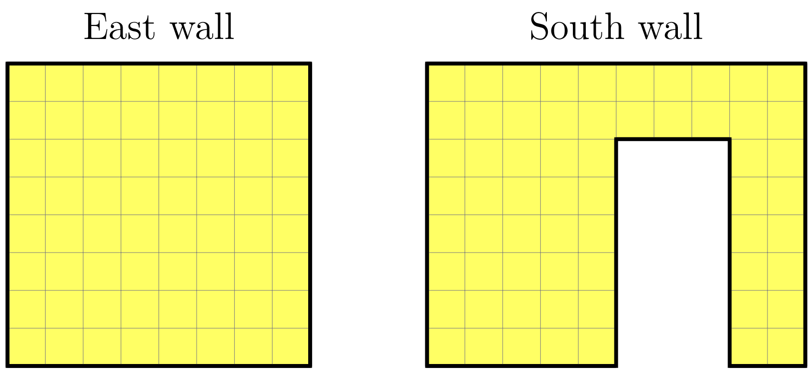
(a) The North and South walls have the same width and height. If the doorway didn't exist, the diagram would show them as two congruent rectangles, which would have equal area. But the doorway does exist. We can divide the North wall into two regions, one being congruent to the diagram of the South wall and the other congruent to the doorway. By the Area Sum Postulate, the area of the North wall is equal to the area of the South wall plus the area of the doorway. Hence the area of the North wall is larger than the area of the South wall, so the North wall requires more paint.

(b) The East wall and West wall have the same width and height, so a similar argument shows that the area of the East wall equals the area of the West wall plus the area of the window. Hence the East wall has greater area than the West wall and requires more paint than it.

(c) The North and East walls have the same height but the North wall has greater width. We could draw a vertical line down the North dividing it into two rectangles, one of which has the same width as the East wall. The area of the North wall will equal the sum of the areas of those two rectangles, but one of those rectangles has the same area as the East wall. Thus the North wall has greater area than the East wall and will require more paint.

The East and South walls have the same height. The South wall is wider, which increases the amount of paint required, but then it also has a doorway removed from its area, which reduces the amount of paint required. Without more information, it seems difficult to reliably guess which area is larger.

Here I've redrawn the East and South walls with a grid on them. The grid divides the walls into small squares of equal size.



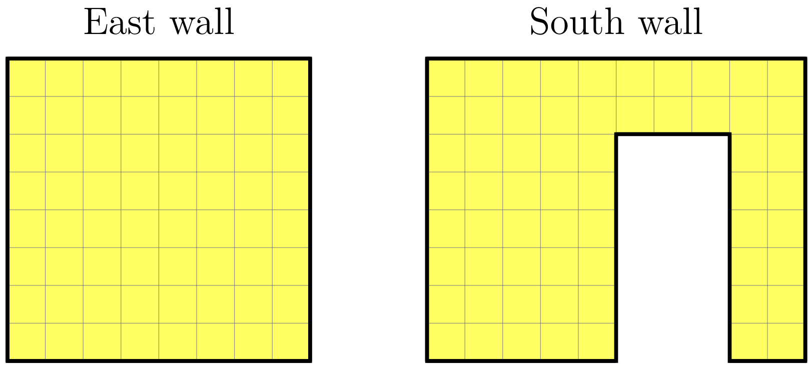
At this point you could rightfully interject with all sorts of difficult questions. For example, what do we do if the width of the wall was such that it could not be evenly divided into squares, but instead left a narrow column of rectangles narrower than the squares? What if the same thing happened for the height. What if the placement of the doorway in the South wall caused similar problems?

We will get to issues like these in later chapters. For the moment, I'm just showing how the Area Sum Postulate applies to areas, so I manufactured an artificially easy example where such difficulties do not arise. First I decided I wanted to use a grid where the lines are 5mm apart, and then I chose wall and doorway dimensions that guaranteed all the edges of the paintable area fell exactly on the grid. (It's 5mm in my software. Depending on screen resolutions and zoom settings, it may be different on your screen.)

You might also ask how we can be sure the grid divides these shapes into squares? They *look* like squares, but it's possible they might be slightly off. Is there a way to construct a grid which absolutely guarantees that the small shapes are squares? For the moment, let's assume the small shapes are all congruent squares and thus they all do have exactly the same area, and see where that leads us. We will get to formal derivations in the next few chapters.

Given the assumption that the small squares have equal area, which we will denote as a square units, can you now tell which of these two walls has the larger area?

The East wall contains 64 small squares. Perhaps you counted them individually, but hopefully you noticed a shortcut. There are 8 columns of squares with 8 squares in each column, giving $8 \times 8 = 64$ squares. The Area Sum Postulate tells us the area of the East wall is equal to the sum of the areas of the 64 small squares. Each square has area a square units, so the East wall has area $64a$ square units.

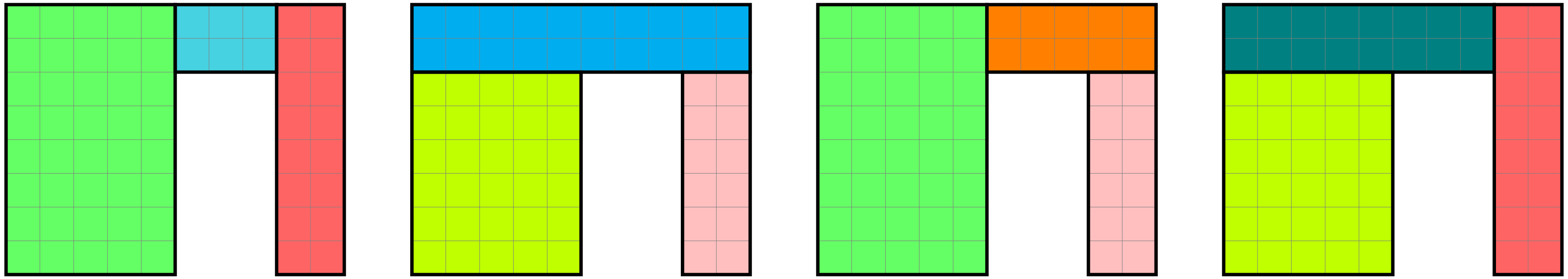


By contrast, the South wall contains 62 small squares, so it has area $62a$ square units. That is, the East wall has a larger area than the South wall, so the East wall requires more paint.

Perhaps you counted those 62 squares in the South wall individually. Perhaps you found a more efficient method to count the squares by dividing the figure into three rectangles.

> If you haven't already done so, show how to divide the South wall into three rectangles and thus determine the number of small squares it contains. If you'd like a harder challenge, find four different ways to divide the South wall into three rectangles.

Here are the 4 different subdivisions. For clarity, I used different colours for the rectangular subdivisions.



In the first solution, the green rectangle contains 5 columns of squares, each column containing 8 squares, so the green rectangle contains $5 \times 8 = 40$ squares. Similar arguments can be applied to the blue and red rectangles. Then the area of the figure will be the sum of the areas of the green, blue and red rectangles. The number of squares contained by the figure can be found as follows.

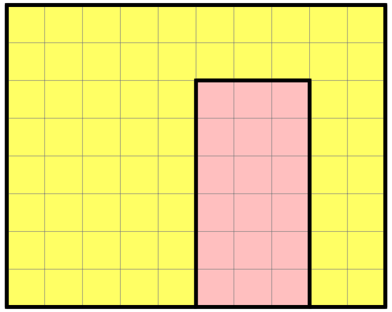
$$5 \times 8 + 3 \times 2 + 2 \times 8 = 40 + 6 + 16 = 62$$

You can also verify that the other 3 methods of subdividing the figure into rectangles give the same final total.

When I drew the grid on the South wall I deliberately placed it only on the figure we were going to paint, omitting the doorway. This lead us towards the solution of dividing the figure into 3 rectangles.

Here I've extended the grid over the doorway, which I've coloured pink. Use this diagram, to find another method for counting the number of yellow squares in the South wall.

South wall



It is now easy to count the number of squares taken up by the doorway. Earlier I noted that by the Area Sum Postulate,

$$\text{Area of North wall} = \text{Area of South wall} + \text{Area of doorway}$$

Rearranging this equation gives:

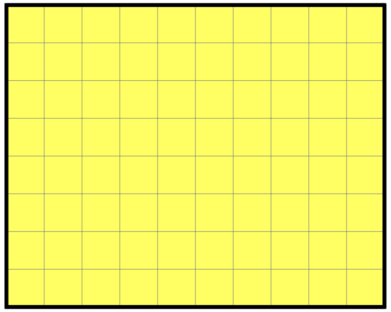
$$\text{Area of South wall} = \text{Area of North wall} - \text{Area of doorway}$$

Thus the number of squares in the South wall is

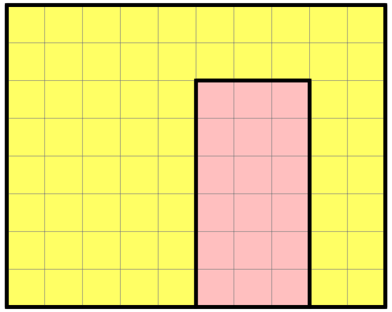
$$10 \times 8 - 3 \times 6 = 62$$

This agrees with our previous calculations that subdivided the South wall into three rectangles.

North wall



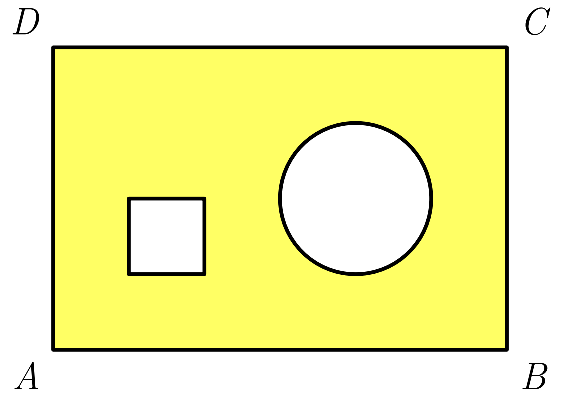
South wall



4.04 Regions with holes

In the previous section, the West wall contained a window that we did not want to paint. That is, for the purpose of calculating the quantity of paint required, the West wall contains a rectangular region that was not part of the wall.

This is an example of a more general idea. A region of the plane might contain within itself one or more regions that are not regarded as part of the original region. The excluded regions might be called “holes.” The diagram here shows a rectangular region containing two holes. One hole is a square and the other is a circle.



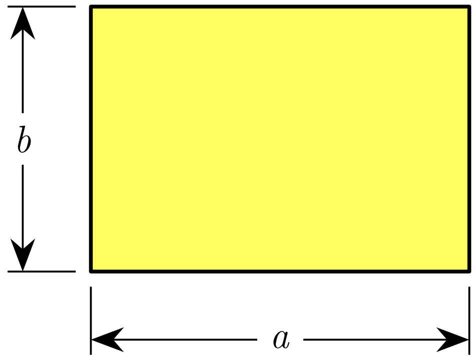
By the The Area Sum Postulate, the area of rectangle $ABCD$ will equal the sum of the areas of the yellow shaded region, the square and the circle. Thus if we need the area of the yellow shaded region, we can calculate it as

$$\text{Area of rectangle } ABCD - \text{Area of Square } - \text{Area of Circle}$$

This approach can be adapted to deal with any region containing holes. Thus, for the rest of this book I will essentially ignore regions with holes. I will derive formulae for the area of common shapes such as rectangles, square and circles which do not contain holes. When needed, these formulae can combined to find the areas of regions with any of those shapes containing holes with any of those shapes.

5 Joining rectangles with a common dimension

5.01 Some jargon from algebra



Here is a rectangle with dimensions a units by b units. In the next chapter, we will prove that we can calculate the rectangle's area by multiplying those two dimensions together, so its area is ab square units.

If you have prior experience with algebra, hopefully that last paragraph makes sense. If so, you can skip the rest of this section, which have been included to assist any readers who are new to algebra.

In the above, a and b are known as variables. When two variables are written consecutively, it means they are multiplied together, so ab denotes $a \times b$. They are called variables because their values can vary. Sometimes variables can only vary over some restricted range. Here we will required the variables to be greater than zero. Subject to that restriction, whatever values they take, the area of the resulting rectangle in square units can always be found as $a \times b$.

For example, if we have a rectangle with dimensions 5 units by 3 units, then we can say $a = 5$ and $b = 3$. Then if we want to find the area of the rectangle we evaluate the expression ab using those particular values of the variables a and b.

$$ab = a \times b = 5 \times 3 = 15$$

Hence the area of the rectangle is 15 square units.

Another common practice is to let A denote the area of the rectangle in square units. Then we can write the equation $A = ab$. In algebra, case is important. The lower case a and upper case A denote different things.

In mathematics, a "formula" is an expression or equation that tells us how to calculate some quantity in terms of one or more variables. Thus the expression ab is a formula for calculating the area of a rectangle, as is the equation $A = ab$. However, some authors disagree with this statement, claiming that a formula must be an equation, and thus they would not call the expression ab a formula.

My opinion is that the context matters. If I simply wrote the algebraic expression ab without any explanation of what the two symbols mean, then it is definitely not a formula. If I provide the diagram above or the explanation in words that a and b denote the dimensions of a rectangle, then I'm still uncomfortable describing ab as a formula, because there's no explanation of what it's supposed to represent. But if I go the extra step and write: "The rectangle's area is ab square units," then I'll happily describe ab as a formula, since the surrounding words make its purpose clear.

When a formula is written as a equation, the left hand side should only contain the thing being calculated. Thus $A = ab$ is a formula for the area of the rectangle shown above. A true equation remains true if the same amount is added to both sides, so while it is true that $A + 42 = ab + 42$, this new equation would not be described as a formula for the area.

A formula has to contain at least one variable, meaning it will have some sort of general applicability. The formula $A = ab$ tells us how to find the area given two variables a and b, so it can be applied to any rectangle with known dimensions. By contrast $4 = 2 + 2$ is a true equation, but it has no variables, so we don't call it a formula.

The word "formula" comes from Latin, so purists argue we should also use the corresponding Latin plural, which is "formulae," the third syllable being pronounced as rhyming with tea. That is, we can refer to one formula or to two or more formulae. US-English usually leads the world when it comes to ignoring the many difficult to remember Latin and Greek plural forms, so most US dictionaries only list the plural "formulas." Australian and UK dictionaries tend to list both "formulae" and "formulas" as options.

In Australia, the word "pronumeral" may be encountered when studying algebra, though its use is declining. A "pronumeral" is a letter used to denote a number. When doing mathematics in English, pronumerals are usually selected from the 26 letters of the English alphabet, with both upper and lower case being allowed, giving 52 possible pronumerals. The example above contained two variables which were denoted by the pronumerals a and b. Greek letters are also frequently used. For example, the ratio of the circumference of a circle to its diameter is traditionally denoted by the Greek letter π, which is spelt "pi" and pronounced the same as "pie". No matter what size circle we choose, the ratio of its circumference to its diameter always takes the same value, so π is a constant rather than a variable. I can't give an exact decimal expression for π because it is irrational, but an approximation is $\pi \approx 3.14159$.

Apparently the word "pronumeral" is an Australian invention and is not in common use elsewhere. Outside Australia, if an author has to distinguish between a variable or constant, and the pronumeral being used to denote that variable or constant, they usually refer to the latter as a symbol. But the word "symbol" also applies to symbols denoting operators, such as the + symbol for addition and the × symbol for multiplication. That is, the word pronumeral has a narrower meaning than symbol, which does make it useful.

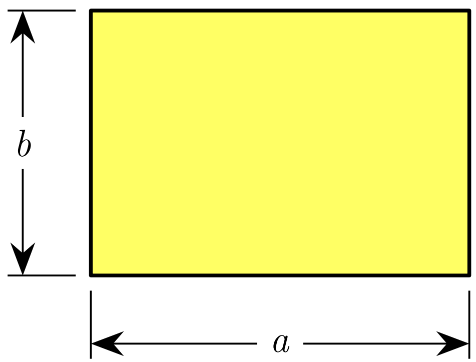
5.02 Rectangle jargon: width and height

Textbooks often state the formula for the area of a rectangle using words rather than pronumerals, but different textbooks can use different words. For example, you may find textbooks claiming:

Area = Width × Height   Area = Length × Width   Area = Length × Breadth

What's going on? Can these formula all be right?

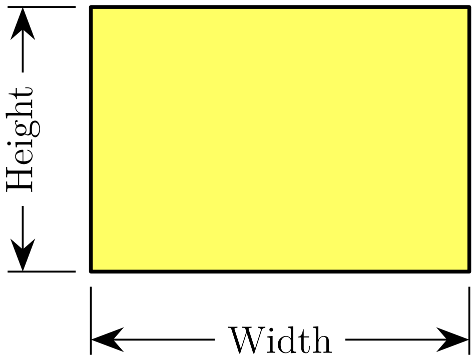
These formula are all intended to *mean* the same thing. If they are applied to a rectangle with dimensions a units by b units, the authors all mean that the area is ab square units. The problem here is that the authors are choosing different words to describe the two dimensions of the rectangle.



Inconsistent names are confusing. In this chapter and the next I'll often need to refer to the two dimensions of a rectangle by name, and I want to choose two names for those two dimensions that make it clear which dimension is which.

If I place a rectangle in a horizontal orientation, like the floor of a room, and refer to the rectangle's length, it can be unclear which of the two dimensions I mean. So let's not do that!

Instead, I'm going to place a rectangle in a vertical orientation with its base horizontal, like the wall of a room. In common English, the word "height" refers to a length measured in a vertical direction. So if I tell you that I'm going to name the two dimensions of this rectangle as "width" and "height", hopefully you will have no trouble remembering that I am using those terms as shown in this diagram.



At least, this should be easy to remember if you are reading this book on a computer screen which is oriented vertically, since that means the height dimension really is vertical.

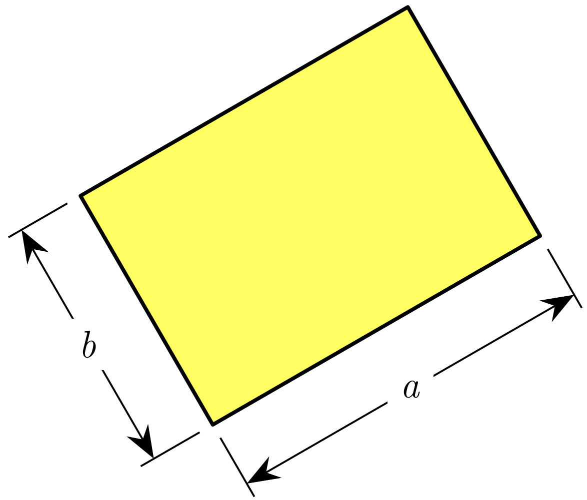
I'm guessing most readers will be viewing this book on a vertical computer screen. If you aren't, just take a moment to move your screen into a vertical orientation with its base horizontal, and verify that the dimension labelled "height" in the above rectangle is vertical. If you are reading this on a tablet computer that is lying flat on your desk, prop it up vertically. If you are using a notebook computer that unfolds to expose its keyboard and screen, make sure you have unfolded it by 90° so the screen is vertical.

I will consistently use the labels "width" and "height" in the orientation shown above in all the proofs in this chapter and the next, and by the end of next chapter we will prove that the area of the above rectangle can be found as:

$$\text{Area} = \text{Width} \times \text{Height}$$

The opposite sides of a rectangle have equal length, so this formula tells us that the area of a rectangle in this orientation is the product of the lengths of any two adjacent sides.

Once we have proved this result, proving that it also holds in other orientations is easy, because congruent shapes have equal area. If I use any combination of translations, rotations and/or reflections to move this rectangle to some other orientation, such as the sloping orientation show at right, this doesn't change its area or the lengths of its sides, so its area can still be found as the product of the lengths of any two adjacent sides.



In the new orientation shown here, none of the sides are aligned vertically, so rather than referring to the rectangle's dimensions as "height" and "width", I might instead choose to revert to labelling the side lengths as a and b, as I did earlier in this section, and then the area will be ab square units.

5.03 Alternative jargon: length, width, height, breadth and depth

This section contains further discussion about the conflicting jargon in use for the dimensions of rectangles. If extended discussions of the alternative meanings of words don't interest you, skip ahead to the next section.

Here are the five words I have seen used to describe a dimension of a rectangle: length, width, height, breadth and depth.

So what are the best two words to use for the two dimensions of a rectangle? Unfortunately, the answer is that it varies.

In common English the word "height" is usually reserved for things measured in a vertical direction. This interpretation will help us remember how I'm using the word in this chapter and the next. Unfortunately, when we reach the chapter on triangles, we're going to find that mathematicians sometimes use the word "height" in orientations that aren't vertical.

When architects design buildings, their drawings include:

- "floor plans" which show the positions of the floors and walls of the building as viewed from above, and
- "elevation plans" which show front and side views of the building.

Architects are careful to only to use the word “height” for the vertical measurements in the elevation plans. If we consider a rectangular wall of a room, it makes sense to call the vertical dimension height. But if we consider a rectangular floor of a room, the word height doesn’t fit the context because the rectangle is horizontal. The dimensions of the rectangular floor of a room would be described with other words, such as length and width.

While not universally true, most people regard “width” and “breadth” as synonyms. Thus it’s common for the dimensions of a horizontally oriented rectangle such as a floor or ceiling to be called length and width, and it’s also common for them to be called length and breadth, but they are seldom called width and breadth.

Say we’ve decided to call the dimensions of a horizontal rectangle length and width. Which is which? Unfortunately, there are still conflicting interpretations. Some say length should be the longer dimension and width the shorter. Others say that it depends on how the rectangle is drawn on the page: length runs from top to bottom on the page while width runs left to right, an explanation that doesn’t provide any help for sloping rectangles.

Some authors reserve the word depth for describing a vertical dimension that is measured downwards from ground level. So say a construction team is installing a flagpole. Initially, when the flagpole is lying on the ground they might say it has a length of 6 metres, because they are measuring that distance horizontally. They dig a 1 metre hole. They could describe the hole as having a depth of 1 metre, since that dimension is measured vertically, but it extends below ground level. They drop the flagpole into the hole and cement it in place. Now the flagpole would be described as having a height of 5 metres because it extends 5 metres vertically above ground level.

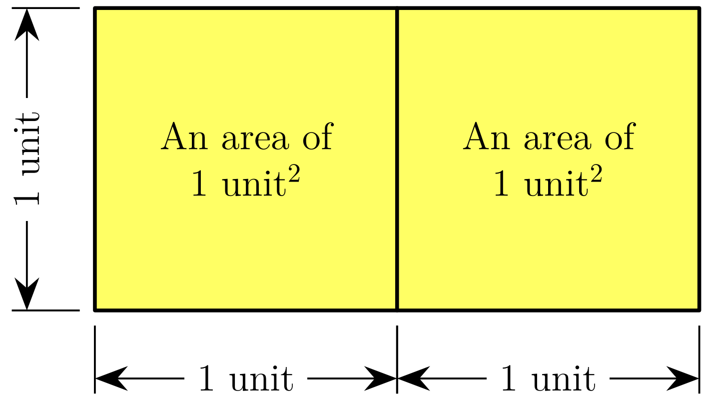
To confuse things further, liquids are often described as having depth rather than height, irrespective of where the liquid is relative to ground level. Thus if we construct a circular wall 1 metre high, line it with waterproof plastic and then fill it with water to 20cm from the top, then we have an above ground swimming pool which would probably be described as having a “height” of 1 metre while the water would be described as having a “depth” of 80cm. It’s as if we think of our natural position in a swimming pool as being with our head above water, so if someone ask us to measure the distance from between the pool floor and the top of the water, we naturally start at the top and measure down, and call the result a depth, rather than starting at the bottom and measuring up, which might make us more likely to call it a height.

Unfortunately, the word depth is also used in a different context with a conflicting meaning. The dimensions of rectangular household appliances such as TVs, refrigerators and microwave ovens are routinely labelled height, width and depth. These appliances all have a face which everyone would agree is its front. For the TV this is the face that displays the picture and for the other two examples it is the face containing the door(s). Height is measured vertically. Width is measured from left to right. Depth is measured from front to back. Thus the dimensions of the rectangle that is the top face of a microwave oven are being labelled width and depth. In this example, depth is measured horizontally, which is in conflict with the previous paragraph where depth was measured vertically downwards.

5.04 The purpose of this chapter

In chapter 2 we defined the area of a square with sides of length 1 unit to be 1 square unit. Using this as our starting point, we will derive formulae for the areas of more complex shapes, starting with rectangles.

For example, if we take two copies of a square unit and join them side by side, it looks like we obtain a rectangle 2 units wide and 1 unit high. Since each copy of the square unit has an area of 1 square unit, it seems the area of a rectangle 2 units wide and 1 unit high will be 2 square units.



But how can we be *sure* that the result of joining the two squares is a rectangle. Is it possible that the figure is really a hexagon that is only quite close to being rectangular? Can we *prove* that it is a rectangle?

Rather than proving that, the purpose of this chapter is to prove more general theorems that relate to joining rectangles that have a common dimension. Then in the following chapter, we will use those theorems to formally derive the formula for the area of a rectangle.

When a mathematician refers to a “more general theorem”, they mean a theorem that is more powerful because it applies to extra cases not covered by the original theorem. Consider the following two hypotheses. I’m calling them hypothesis rather than theorems, because they sound plausible, but we haven’t yet proved that they are true.

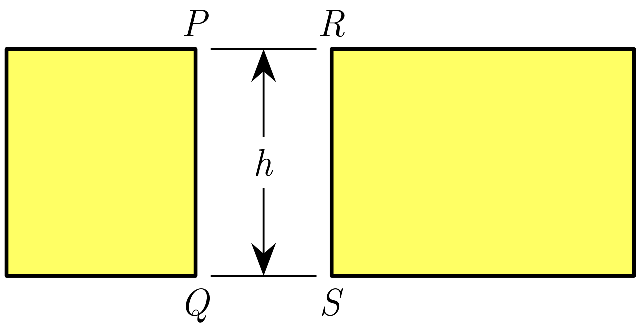
Hypothesis 1: Joining two squares with equal side length side by side produces a rectangle.
Hypothesis 2: Joining two rectangles of equal height side by side produces a rectangle.

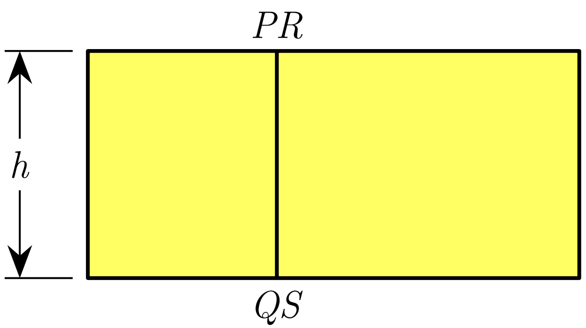
All squares are rectangles, so two squares of equal side length *are* two rectangles with equal height. This means that hypothesis 2 covers all the cases covered by hypothesis 1. But it also covers cases where the two rectangles of equal height are *not* squares, so it covers extra cases that are not covered by hypothesis 1. So we say that hypothesis 2 is more general than hypothesis 1. We can also say that hypothesis 1 is a special case of hypothesis 2, which means that hypothesis 1 is less powerful because it only covers some of the cases covered by hypothesis 2.

It is possible to prove the more general hypothesis 2 without first proving hypothesis 1, so we will do that because it saves time. In fact, we're going to prove a more informative form of hypothesis 2 that also tells us how the dimensions of the resulting rectangle compare to the dimensions of the two rectangles that were joined.

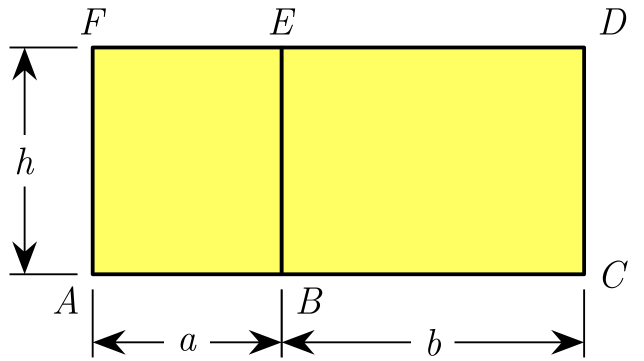
5.05 Joining two rectangles with equal height

Consider two rectangles both having height h. I've labelled two sides as PQ and RS. Since $|PQ| = |RS| = h$, we could slide the two rectangles together, so points P and R coincide and points Q and S coincide. Area is invariant (unchanged) under translation, so sliding the rectangles together doesn't change their area.





P and R are now a single point, as are Q and S. It's confusing having two different labels for a single point. We also need labels for the other vertices of the rectangles. Let's just attach a new label to every point.



Let the widths of the two rectangles be a units and b units.

At first glance, this new combined shape $ABCDEF$ looks like a rectangle. Let's try to prove it.

To prove that the new combined shape is a rectangle we need to:

1. Prove that ABC is one continuous line segment, rather than AB and BC being segments of two different lines with a bend at B. I can't see a bend in the diagram, but maybe there is a very small bend, too small to be easily noticed.

2. Similarly, prove that DEF is one continuous line segment, so there is no bend at E. If we succeed in these first two steps, then we will have proved that the new combined shape $ABCDEF$ is not a hexagon, but rather it is a quadrilateral that we more simply refer to as $ACDF$.

3. If we complete the previous steps, prove that the quadrilateral $ACDF$ is a rectangle.

If this book was a complete geometry textbook rather than just a book about areas, it would have defined a rectangle before deriving a formula for its area. We can't find the area of a rectangle if we don't know what a rectangle is! Hence, for the current proof you are allowed to use the definition of a rectangle. There is more than one way to define a rectangle. Here is probably the most popular definition in use.

Definition: A rectangle is a quadrilateral in which all four angles are right angles.

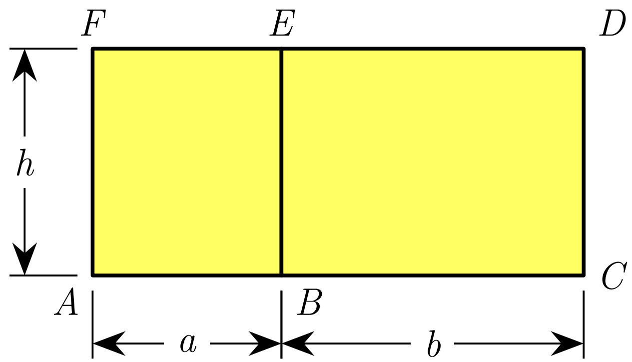
Use the definition to complete the three steps of the proof.

If you weren't able to complete the proof, the following extra hint may help.

We know that $ABEF$ and $BCDE$ are both rectangles.

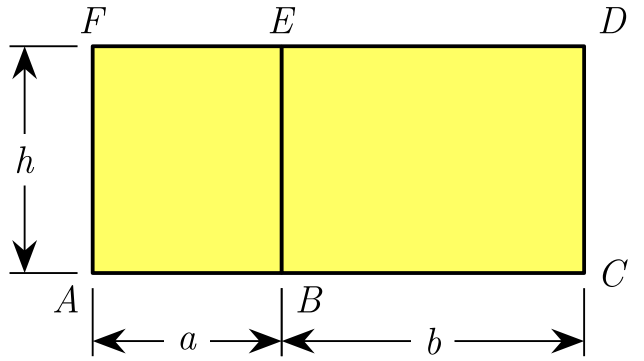
The definition of a rectangle gives the following two properties.

- If a quadrilateral is a rectangle, then all four of its angles are right angles.
- If all four angles of a quadrilateral are right angles, then it is a rectangle.



Use *all* of the above facts and properties to complete the three steps of the proof.

Since $ABEF$ is a rectangle, all four of its angles are right angles. The same is true for rectangle $BCDE$. In particular, $\angle ABE = \angle EBC = 90°$. Hence



$$\begin{aligned}\angle ABC &= \angle ABE + \angle EBC \\ &= 90° + 90° \\ &= 180°\end{aligned}$$

Hence ABC is a single line segment, which we can refer to simply as AC. That is, there is no "bend" at B.

Similarly $\angle DEF = \angle DEB + \angle BEF = 90° + 90° = 180°$, so DEF is also a single line segment DF. There is no "bend" at E.

Hence the new combined shape $ABCDEF$ is not a hexagon. It is correctly described as the quadrilateral $ACDF$.

Finally, all four angles of the quadrilateral $ACDF$ are angles of either rectangle $ABEF$ or rectangle $BCDE$, and so they are all right angles. Since all four angles of the quadrilateral $ACDF$ are right angles, it is a rectangle.

The new rectangle has height h, the common height from the two original rectangles, and width $a+b$, which is the sum of widths of the two rectangles that were joined.

Let's state what we have proved as a theorem.

Theorem 5.1a: Two rectangles with equal height can be joined side by side to form a new rectangle with that same height and with width equal to the sum of the widths of the two rectangles.

It's also possible to reference pronumerals in a theorem.

Theorem 5.1b: Two rectangles with widths a and b, and both with height h, can be joined side by side to form a rectangle with width $a + b$ and height h.

Both forms of the theorem are valid. They are effectively saying the same thing, which is why I've labelled them as 5.1a and 5.1b, two different variants of the first theorem of chapter 5. We can refer to these theorems collectively as theorem 5.1. In this book, if a challenge asks you to "Use theorem 5.1" to prove something, I mean it doesn't matter whether you use 5.1a or 5.1b, so you can pick whichever version you prefer.

So how do we choose which form to use when writing a theorem? Sometimes it just comes down to the author's personal preference, but the following factors are worth considering.

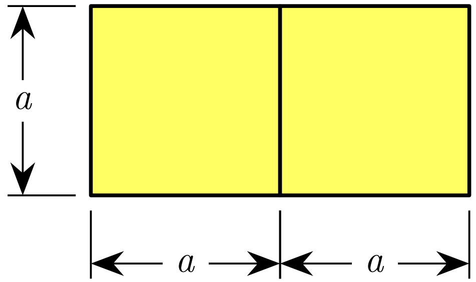
- If a theorem includes a complicated calculation, it's usually clearest to use pronumerals since this lets you describe the calculation with an algebraic formula which is much clearer than trying to explain the calculation in words.
- On the other hand, understanding the meaning of the pronumerals take time, so if there aren't any complicated calculations involved the theorem might be easier to understand without the pronumerals. For theorem 4.1, the expression $a + b$ is certainly shorter than saying "the sum of the widths of the two rectangles", but when the reader sees $a + b$ they need to refer back to the earlier part of the sentence to realise that it is indeed summing the widths of the original two rectangles, so here the presence of the pronumerals might mean the reader takes longer to understand what the theorem is saying.
- If your theorem has pronumerals in it, it can be confusing if you have to apply it to a problem where those same pronumerals have been used with different meanings. I'll give an example of this later in this chapter.

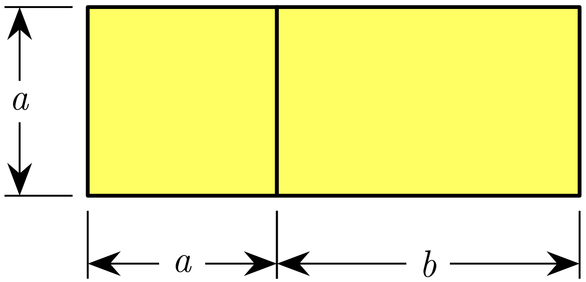
When I'm writing about mathematics my personal preference is to make the reader aware of when they have options, so I made a point of demonstrating both ways of writing the theorem. I know this irritates some readers, who would prefer to only ever see one way of doing something, so that they don't have to make a choice. But mathematics isn't a solitary occupation. If you want to use mathematics to make the world a better place, you need to be able to clearly communicate the results of your mathematics to others. When it comes to communication, having options to express an idea in different ways is always useful.

5.06 A special case: some rectangles are squares

Where the theorem refers to "two rectangles", it's possible that both of those rectangles are squares.

That is, if two squares both with width a are joined side by side, they will form a rectangle with width $2a$ and height a.





It's also possible that only one of the rectangles is a square.

If a square of width a and a rectangle of width b and height a are joined side by side, they will form a rectangle with width $a + b$ and height a.

There is nothing here that needs proving. These results following directly from theorem 5.1. All squares are rectangles, so any theorem about rectangles must also be true for squares. Students sometimes find this type of assertion surprising and they may raise one of the following objections.

1. But look at the pronumerals you used when proving that theorem. The rectangles have widths a and b, and they both have height h. None of those three dimensions can be equal because they are denoted by three different pronumerals, so neither rectangle can be a square.

2. But in the diagram used to prove this theorem it's clear that $a < h < b$, so neither rectangle can be a square.

What do you think of these two objections?

Neither objection is valid.

1. When you introduce a pronumeral, it can take any value at all unless you explicitly state otherwise. This includes the possibility that it can take the same value as any other pronumerals being used within the question. So it is possible that $a = h$, meaning that the first rectangle can be a square, or $b = h$, meaning the second rectangle is a square, or $a = b = h$, meaning both rectangles are squares. The only restriction on the values taken by these pronumerals is that they must all be greater than zero.

2. In geometry, if there is some important restrictions on the relative sizes of pronumerals, we should state these restrictions explicitly. Readers shouldn't have to guess what the restrictions are based on how we drew the diagram.

For example, I can't draw a rectangle without making some assumption about the relative sizes of the two dimensions. In my diagram above, I drew the first rectangle with $a < h$, but I haven't explicitly stated that as a restriction. Thus anyone reading my proof should assume the argument still works when $a = h$ or $a > h$.

If I believed my proof was only valid when $a < h$, I would have explicitly stated this when I introduced those symbols. To make the proof as instructive as possible, I probably would have stated which step(s) in the proof required that restriction, and explained why they failed if the restriction didn't hold. I would also have repeated the restriction in the statement of the theorem.

I did none of those things, so *I* believe my proof is valid for all three cases $a < h$, $a = h$ and $a > h$. Do you?

Reread the proof of theorem 5.1. Can you find any step in the proof that fails for any of those three cases? Similarly, try to find any step that fails for any of the three cases $b < h$, $b = h$ and $b > h$.

You should find the proof valid for all the cases listed above. The proof relies solely on properties about the angles of rectangles. If we change the values taken by a, b or h, the rectangles are still rectangles, the angles are all still right angles and so the proof remains valid. This is subject to the constraint that the pronumerals must remain positive. If we let any pronumeral reduce to zero, the rectangle collapses to a line segment.

Can you spot the problem with the following suggestion?

The proof only gave one diagram. Why don't we give three different diagrams for the three cases $a < h$, $a = h$ and $a > h$? Surely the proof would be clearer if we gave diagrams that cover all the possible cases.

The problem is that those three diagrams don't actually show *all* the possible cases. They only show the three possible cases for the relative sizes of a and h. We could split each of those three cases into three more cases, with $b < h$, $b = h$ and $b > h$, so that would give us nine different diagrams. But we're still not done, because some of those nine cases could be subdivided further. For example, the case where $a < h$ and $b < h$, could be further subdivided into $a < b < h$, $a = b < h$ and $b < a < h$. Then, in the next section, we want to extend the theorem to cover three rectangles, which would require even more diagrams.

Now if you are not yet convinced that the proof will work for all these different cases, you should definitely draw *some* of the possible different diagrams. But don't feel obliged to find *every* possible diagram. You can stop as soon as you convince yourself that the proof only relies on the sizes of the angles, and that the relative size of the side is irrelevant. Once you realise that, it is hopefully clear that if you are asked to provide a formal proof of theorem 5.1, it is fine to only provide one diagram.

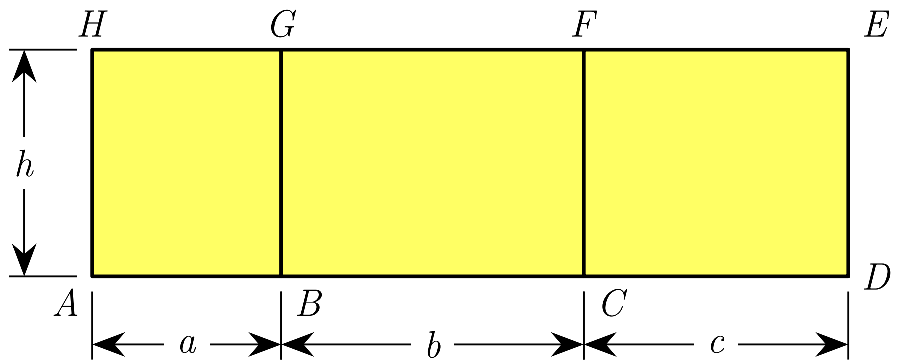
There are harder problems in geometry where it *is* necessary to provide two or more different diagrams to prove a theorem, because the argument required is different for those different cases. We will encounter an example of this later in this book.

5.07 Joining three rectangles of equal height

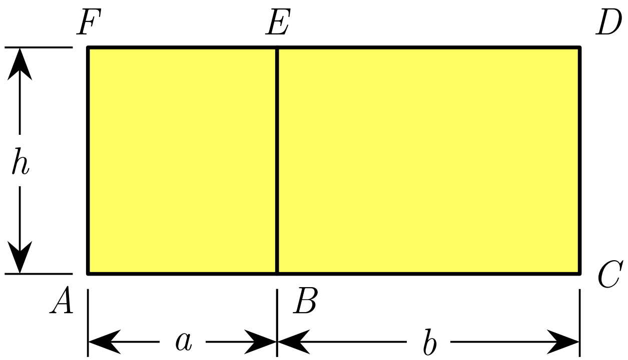
Eventually we want a theorem that talks about joining together two or more rectangles of common height, but we'll approach that problem gradually.

Theorem 5.1 covers joining two rectangles with common height. The next step is to consider joining three rectangles with common height h. Let their widths be a, b and c.

It looks like we can join these side by side to give a new rectangle of height h and width $a + b + c$. But how would we prove the new combined shape is a rectangle?

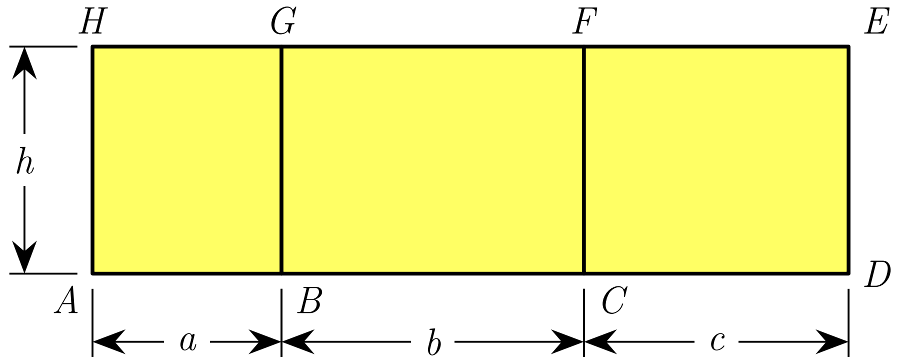


We could work through the same process we did in the previous proof, but it would be harder this time.



Recall that in the previous proof where we joined two rectangles, we had to prove the new figure was a rectangle rather than a hexagon. We proved that ABC was a single line segment with no bend at B, and that DEF was a single line segment with no bend at E. That is, we eliminated the possibility that there were bends at two particular points.

We could use that same process again for our new problem with three rectangles, but now we'd have to eliminate the possibility of bends at four points, the points labelled B, C, F and G in the diagram shown right. Admittedly the same argument is used over and over again, but it's tedious to work through, and becomes even more tedious if we want to extend it to four rectangles.

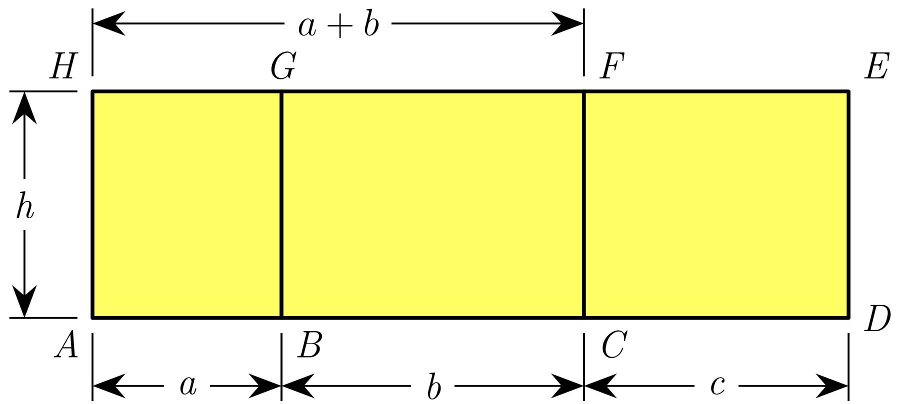


There's an easier method that avoids repeating all the detailed work of the previous proof. Instead, it only employs the resulting theorem. Can you see how to do this?

We use the theorem twice. Here is the theorem again, in the form without pronumerals.

Theorem 5.1a: Two rectangles with equal height can be joined side by side to form a new rectangle with that same height and with width equal to the sum of the widths of the two rectangles.

First we apply this theorem to rectangles $ABGH$ and $BCFG$. Since these are rectangles with a common height of h, the theorem tells us that when they are joined together the resulting figure $ABCFGH$ is a rectangle, so we can refer to it simply as $ACFH$. The theorem also tells us this new rectangle will have height h and width equal to the sum of the widths of the two rectangles $ABGH$ and $BCFG$. That is $ACFH$ is a rectangle with width $a + b$ and height h.



Since we have proved $ACFH$ is a rectangle with height h, and we know from the data that $CDEF$ is a rectangle of height h, we can now apply the theorem to those two rectangles. The theorem tells us that when these two rectangles are joined, the resulting figure $ACDEFH$ is a rectangle, so we can refer to it more simply as $ADEH$. The theorem also tells us that this new rectangle will have height h and width equal to the sum of the widths of the two rectangles $ACFH$ and $CDEF$, which is $a + b + c$.

Theorem 5.1a applied to two rectangles. We have now proved the analogous theorem for three rectangles.

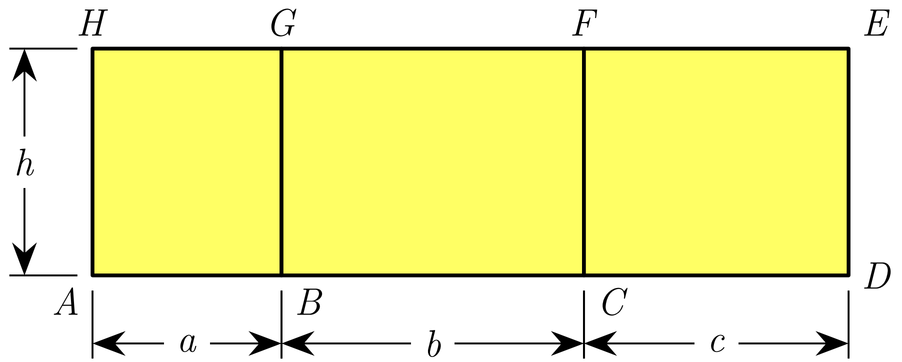
Theorem 5.2a: Three rectangles with equal height can be joined side by side to form a new rectangle with that same height and with width equal to the sum of the widths of the three rectangles.

5.08 Once more with pronumerals

In the previous section we used theorem 5.1a, the version that doesn't contain pronumerals. What changes if we instead use theorem 5.1b?

Theorem 5.1b: Two rectangles with widths a and b, and both with height h, can be joined side by side to form a rectangle with width $a+b$ and height h.

Let's walk through the proof again.



First we join rectangle $ABGH$ and rectangle $BCFG$. That step is easy because the pronumerals used for these two rectangles exactly match those in theorem 5.1b. The theorem immediately tells us the result of joining those two rectangles is a rectangle of width $a+b$ and height h, the rectangle $ACFH$.

But is theorem 5.1b still useful when we join rectangle $ACFH$ to rectangle $CDEF$? The pronumerals in the theorem don't match that scenario.

In the theorem, we can replace the variables a and b by any number we like, such as 42, $\frac{3}{4}$ or $\sqrt{2}$. Well, not quite *any* number. They represent lengths, so they do have to be greater than zero. They can't be zero, since that would cause a rectangle to collapse down to a line segment. It would no longer *be* a rectangle. So we can replace a and b by any number greater than zero.

We can also replace the the pronumerals a and b by other pronumerals, provided we similarly restrict those pronumerals so that the dimensions of the rectangles are always greater than zero. For example, we could replace every occurrence of the pronumeral a by x, or even an expression containing multiple pronumerals, such as $p+q$. That might seem an odd thing to do, since it seems to be making the theorem more complex than it needs to be, but the resulting theorem would be valid.

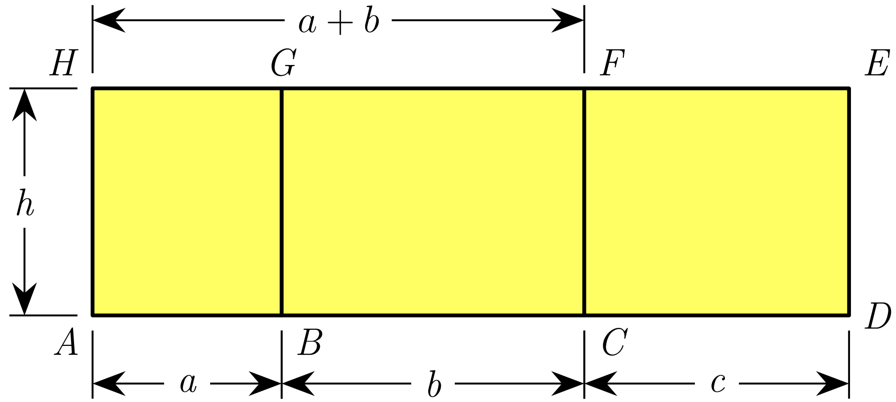
In theorem 5.1b, what should we replace a and b by to make the theorem applicable to joining rectangle $ACFH$ and rectangle $CDEF$?

Hint: When you have to replace two pronumerals, the order in which you replace them can be important. If you're stuck, try switching the order.

If you found this challenge difficult, you are not alone! Many people prefer to use theorem 5.1a — the version that does not refer to pronumerals — because it avoids the need to figure out what replacements are required here.

First replace every occurrence of b by c. Then replace every occurrence of a by $a+b$. Then the theorem will say:

Two rectangles with widths $a+b$ and c, and both with height h, can be joined side by side to form a rectangle with width $a+b+c$ and height h.



Hence the result of joining rectangle $ACFH$ to rectangle $CDEF$ is a rectangle with width $a+b+c$ and height h. We can state this result as a theorem which is equivalent to theorem 5.2a, but which refers to pronumerals.

Theorem 5.2b: Three rectangles with widths a, b and c, and all with height h, can be joined side by side to form a rectangle with width $a+b+c$ and height h.

But why stop there? Say someone now hands us a 4^{th} rectangle that had width d and height h. We've already proved that the joining the first three rectangles produced the rectangle $ADEH$ which has width $a+b+c$ and height h. So our theorem about joining two rectangles with common height says that joining rectangle $ADEH$ and the 4^{th} rectangle will produce a rectangle with height h and with width equal to the sum of the width of rectangle $ADEH$ and the 4^{th} rectangle. That is, it will have width $a+b+c+d$, which also happens to be the sum of the widths of the four rectangles it contains.

It feels like we can extend this argument to cover as many rectangles as we like. We can construct a formal proof of this by using a technique called mathematical induction. In Australia, mathematical induction is not usually encountered until the senior years of high school, which is odd given given that it is a relatively simple idea which allows us to prove several results encountered in earlier years of the syllabus.

5.09 The principle of mathematical induction

This principle is useful for proving theorems that involve a pronumeral that can take evenly spaced values, most commonly whole numbers. Here is how it works.

- Make a statement setting out the property to be proved. The property will contain a variable, which I will denote by the pronumeral n. To write the argument succinctly, it is useful to have a symbol that denotes this statement. I will use $S(n)$.
- Indicate the values of n for which you hope to prove statement $S(n)$ true. The values must be evenly spaced. Most commonly the values are $0, 1, 2, \ldots$ or $1, 2, 3, \ldots$, but for the first scenario where we will use mathematical induction, they will be the numbers $2, 3, 4, \ldots$. That is, we aim to prove the statements $S(2), S(3), S(4), \ldots$ are all true statements.
- The initial step: Find a way to prove $S(n)$ true for the smallest possible value of n. In our first scenario, we want to prove $S(2)$ is true. That is, prove our statement $S(n)$ is true when $n = 2$.
- The induction step: Find a way to prove that if $S(k)$ is true then $S(k+1)$ must be true. That is, prove that if the statement $S(n)$ is true when $n = k$, it must be true when $n = k + 1$.
- If you succeed in doing the above, the Principle of Mathematical Induction allows you to conclude that $S(n)$ is true for all the required values.

The principle works because the induction step lets us sequentially run through all the relevant values of n. The initial step proved $S(2)$ is true. Now apply the result of the induction step over and over again. Because $S(2)$ is true, $S(3)$ must be true. Because $S(3)$ is true, $S(4)$ must be true. Because $S(4)$ is true, $S(5)$ must be true. We can keep going as long as we like. Thus $S(n)$ is true for $n = 2, 3, 4, \ldots$

5.10 Generalising to two or more rectangles

This section employs pronumerals with subscripts. If you aren't familiar with these, this would be a good time to read Appendix C which provides an introduction to this notation.

Earlier in this chapter we derived a theorem about joining together two rectangles with equal height, and then we used that theorem twice to prove a corresponding theorem about joining three rectangles with equal height. It felt like the process could be extended to cover any number of rectangles. The principle of mathematical inductions lets us prove this.

For some readers, this will be the first example of mathematical induction they have ever read, so just a warning that this next proof is going to explain the process in excruciatingly fine detail.

First we write a statement $S(n)$ that refers to joining n rectangles.

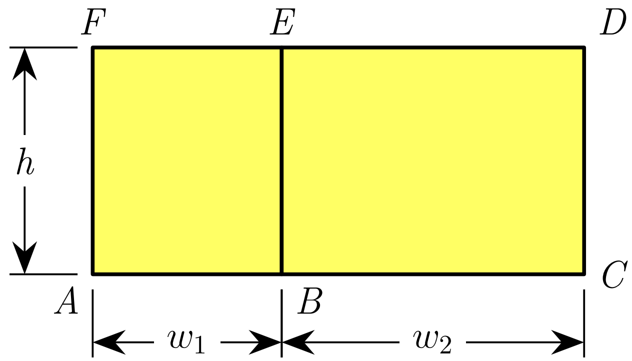
$S(n)$: If n rectangles with widths $w_1, w_2, w_3, \ldots, w_n$, all with height h, are joined side by side they form a rectangle with width $w_1 + w_2 + w_3 + \ldots + w_n$ and height h.

If you are reading that aloud, the colon immediately after $S(n)$ can be read as "states that" or "is the statement".

We want to prove the statement $S(n)$ is true for $n = 2, 3, 4, \ldots$

The initial step is to prove the statement $S(2)$ is true. That is, prove that $S(n)$ is true when $n = 2$. How do we do this?

Initial step: Here is $S(2)$, which is $S(n)$ with n replaced by 2, and some minor grammar adjustments to improve clarity.



$S(2)$: If two rectangles with widths w_1 and w_2, both with height h, are joined side by side they form a rectangle with width $w_1 + w_2$ and height h.

Earlier we proved:

Theorem 5.1a: Two rectangles with equal height can be joined side by side to form a new rectangle with that same height and with width equal to the sum of the widths of the two rectangles.

That theorem covers $S(2)$ perfectly, so we can simply write: By theorem 5.1a, $S(2)$ is true.

Here is the alternative form of the theorem that uses pronumerals.

Theorem 5.1b: Two rectangles with widths a and b, and both with height h, can be joined side by side to form a rectangle with width $a + b$ and height h.

The substitutions $a = w_1$ and $b = w_2$ demonstrate that this theorem proves $S(2)$.

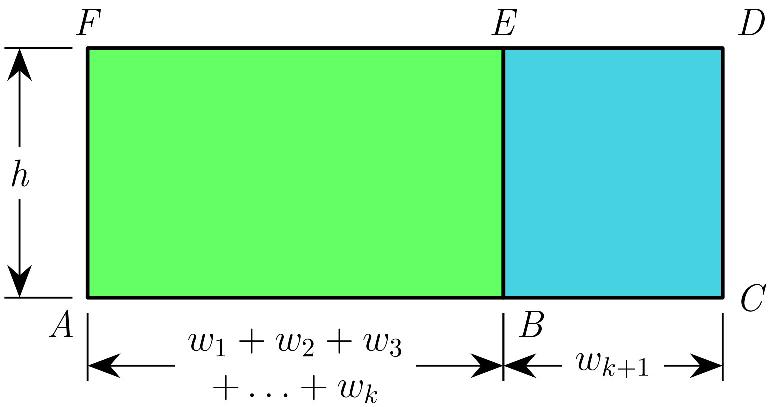
Now look at the induction step. For $k = 2, 3, 4, \ldots$, prove that if $S(k)$ is true, then $S(k+1)$ must be true.

Induction step: Here are the two relevant statements.

$S(k)$: If k rectangles with widths $w_1, w_2, w_3, \ldots, w_k$, all with height h, are joined side by side they form a rectangle with width $w_1 + w_2 + w_3 + \ldots + w_k$ and height h.

$S(k+1)$: If *k+1* rectangles with widths $w_1, w_2, w_3, \ldots, w_k, w_{k+1}$, all with height h, are joined side by side they form a rectangle with width $w_1 + w_2 + w_3 + \ldots + w_k + w_{k+1}$ and height h.

Style hint: When a terminating sequence of terms includes an ellipsis, (the three dots), we often list only the final term after the ellipsis, but we *can* list more if it improves clarity. Here, in the statement of $S(k+1)$, I chose to also list the second last term, w_k, since this more clearly shows the difference between $S(k)$ and $S(k+1)$.



If $S(k)$ is true then the first k of the $k+1$ rectangles mentioned in $S(k+1)$ can be joined side by side to form the green rectangle shown here. The $(k+1)^{th}$ rectangle is shown in blue. The green and blue rectangles both have height h, so by theorem 5.1a they can be joined side by side as shown to produce a rectangle of height h and with width equal to the sum of the widths of the green and blue rectangles, which is $w_1 + w_2 + w_3 + \ldots + w_k + w_{k+1}$. This proves that if $S(k)$ is true then $S(k+1)$ is true.

If you prefer to employ theorem 4.1b, the form that contains pronumerals, then the substitutions $a = w_1 + w_2 + w_3 + \ldots + w_k$ and $b = w_{k+1}$ will provide the required result.

We have proved $S(2)$ is true, and that if $S(k)$ is true then $S(k+1)$ is true. Hence by the principle of mathematical induction $S(n)$ is true for $n = 2, 3, 4, \ldots$.

Here is the resulting theorem, written both with and without pronumerals.

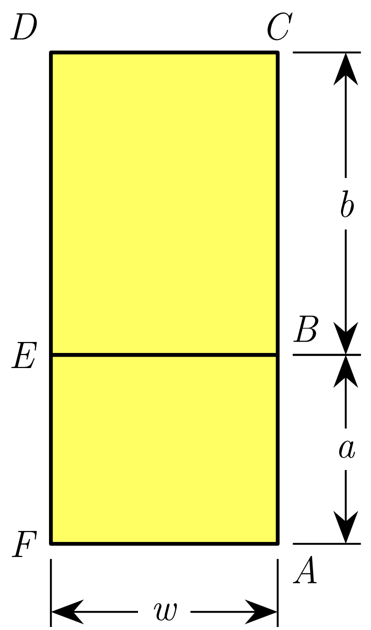
Theorem 5.3a: Two or more rectangles with equal height can be joined side by side to form a new rectangle with that same height and with width equal to the sum of the widths of the rectangles being joined.

Theorem 5.3b: For $n \in \{2, 3, 4, \ldots\}$, n rectangles with widths $w_1, w_2, w_3, \ldots, w_n$ and with common height h can be joined side by side to form a new rectangle with width $w_1 + w_2 + w_3 + \ldots + w_n$ and height h.

Since this theorem works for two or more rectangles, it supersedes the theorem 5.1 which only worked for two rectangles and theorem 5.2, which only worked for three rectangles. That is, we no longer *need* theorems 5.1 and 5.2, because we now have a more general theorem that covers both those cases and many more.

5.11 Rotating the result

So far in this chapter we've been taking rectangles with a common height and joining them side by side. It's natural to wonder whether similar results can be obtained if we take rectangles with a common width and join them top to bottom. It is indeed possible to do this. Before we do, let's revise a few facts about how shapes behave under rotation.



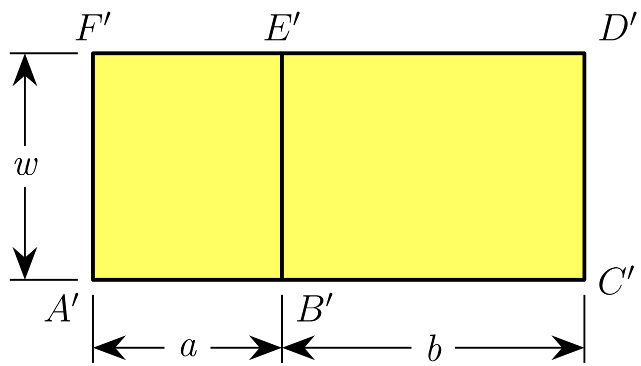
Consider two rectangles with common width w and with heights a and b. Since they have the same width, we can join them top to bottom as shown here.

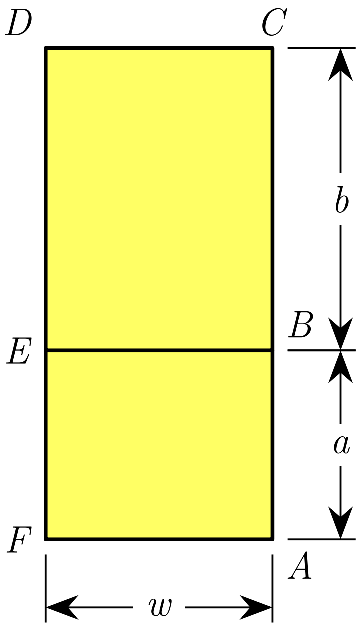
At this point you are probably guessing that the resulting shape $ABCDEF$ is a rectangle, which can be referred to more simply as $ACDF$. But how do we prove that?

We could prove this result by a similar process to how we proved theorem 5.1a, with everything adjusted to this new orientation. If you like, you can do that for practice. But there's an easier way.

If we rotate the above diagram 90° clockwise, it looks like this. I've added prime symbols to the labels for the vertices to distinguish them from the vertices in the original diagram.

Complete the proof. Hint: If we rotate any shape, the original shape and the shape resulting from the rotation are congruent.



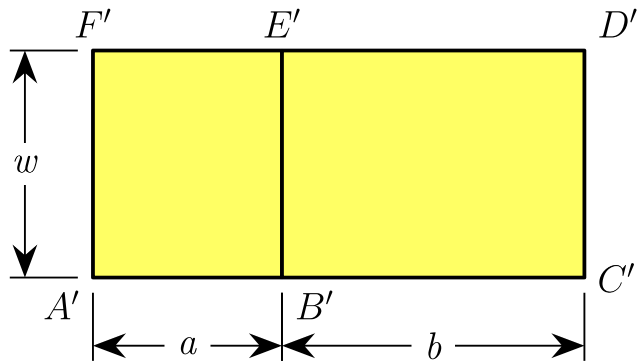


Because a figure is congruent with its image under rotation:

- $ABEF$ is congruent with $A'B'E'F'$. We know that $ABEF$ is a rectangle with width w and height a, so $A'B'E'F'$ is also a rectangle with dimensions w and a, though due to the rotation w is now the height and a the width.
- $BCDE$ is congruent with $B'C'D'E'$. We know that $BCDE$ is a rectangle with width w and height b, so $B'C'D'E'$ is also a rectangle with dimensions w and b, though due to the rotation w is now the height and b the width.
- The shape $ABCDEF$ is congruent to the shape $A'B'C'D'E'F'$.

The first two points above imply $A'B'E'F'$ and $B'C'D'E'$ are rectangles with a common height. Hence theorem 5.3 applies — or if you prefer you could use the superseded theorem 5.1 — and tells us the result of joining them, $A'B'C'D'E'F'$, is a rectangle of width $a + b$ and height w, which can be described more simply as $A'C'D'F'$.

Then the third point implies $ABCDEF$ must also be a rectangle with dimensions $a + b$ and w, which we can simply call it the rectangle $ACDF$. This proves that when two rectangles with equal width are joined top to bottom the resulting shape is a rectangle with that same width and height equal to the sum of the heights of the two rectangles.



While the above argument demonstrated the rotations with two rectangles, the process can be generalised to any number of rectangles we like. Hence theorem 5.3, which deals with joining rectangles with equal height side by side, can be converted to a corresponding theorem that deals with joining rectangles with equal width top to bottom. Here is the result, both with and without pronumerals.

Theorem 5.4a: Two or more rectangles with equal width can be joined top to bottom to form a new rectangle with that same width and with height equal to the sum of the heights of the rectangles being joined.

Theorem 5.4b: For $n \in \{2, 3, 4, \ldots\}$, n rectangles with common width w and with heights $h_1, h_2, h_3, \ldots, h_n$ can be joined top to bottom to form a new rectangle with width w and height $h_1 + h_2 + h_3 + \ldots + h_n$.

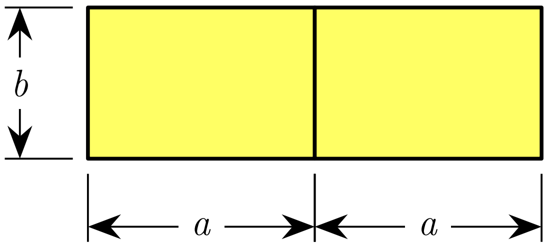
Incidentally, we now have theorems to cover two particular orientations: rectangles with common height joined horizontally and rectangles with common width joined vertically. These are the two particular orientations we will need in the next chapter.

We proved the second case by using a 90° rotation that brought it in line with first case. If we needed a similar theorem in some other orientation, we could prove it in a similar manner using some other rotation. For example, the theorem could also be adjusted to apply to rectangles that are rotated at say 45° to those shown above, or to rectangles sitting in a horizontal plane rather than a vertical plane.

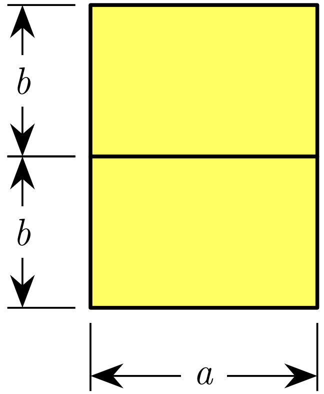
5.12 A special case: Identical rectangles

We know that two or more rectangles with the same height can be joined side by side to form a rectangle, and that two or more rectangles with the same width can be joined top to bottom to form a rectangle.

An interesting special case arises when the rectangles have both the same width and the same height, which means they are identical rectangles. When this happens, you can stack them in either direction.



For example, consider two identical rectangles with width a and height b. These could be joined side by side to form a rectangle with width $2a$ and height b. They can also be joined top to bottom to form a rectangle with width a and height $2b$.



5.13 Summary of theorems carried forward

We've derived several theorems in this chapter. Some theorems superseded earlier theorems. A general theorem that works for two or more rectangles supersedes a less general theorem that only holds for two rectangles. This is a good point to summarise the general theorems that we will use in the next chapter.

Theorem 5.3a: Two or more rectangles with equal height can be joined side by side to form a new rectangle with that same height and with width equal to the sum of the widths of the rectangles being joined.

Theorem 5.3b: For $n \in \{2, 3, 4, \ldots\}$, n rectangles with widths $w_1, w_2, w_3, \ldots, w_n$ and with common height h can be joined side by side to form a new rectangle with width $w_1 + w_2 + w_3 + \ldots + w_n$ and height h.

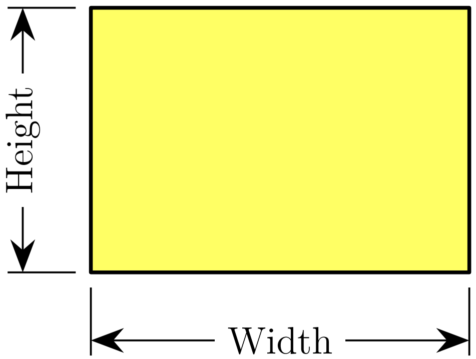
Theorem 5.4a: Two or more rectangles with equal width can be joined top to bottom to form a new rectangle with that same width and with height equal to the sum of the heights of the rectangles being joined.

Theorem 5.4b: For $n \in \{2, 3, 4, \ldots\}$, n rectangles with common width w and with heights $h_1, h_2, h_3, \ldots, h_n$ can be joined top to bottom to form a new rectangle with width w and height $h_1 + h_2 + h_3 + \ldots + h_n$.

While they have been stated as four separate theorems, remember that they come in pairs which really only state the result in a different way. For example, theorems 5.3a and 5.3b state the same result, but one uses pronumerals and the other doesn't. If I refer to theorem 5.3, I mean that you can use either form.

6 Rectangle

6.01 Orientation



As in Chapter 5, for most of this chapter I will use rectangles aligned vertically, such as the walls of a room, and use the words height and width to describe its two dimensions. After we have derived the formula for the area of rectangles in this orientation, we can consider how the formula translates to rectangles in other orientations.

6.02 A rectangle of height one

Let's bring together the relevant theorem and definition.

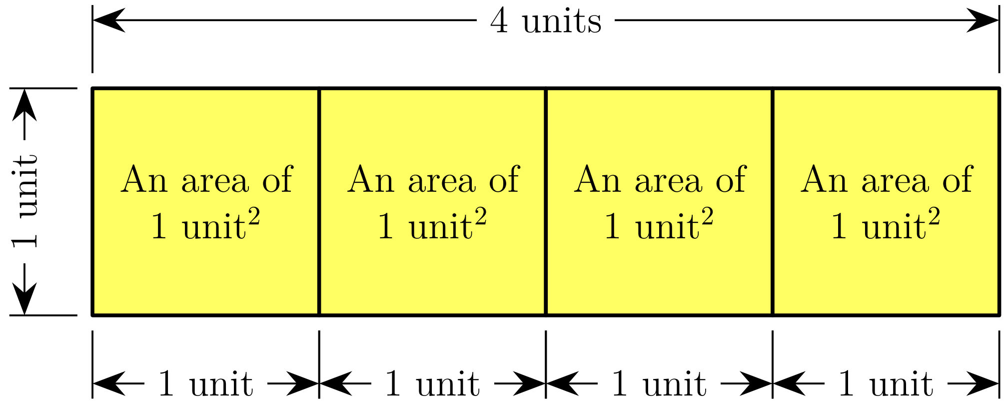
We have proved the following theorem, which I will show in both forms we developed.

Theorem 5.3a: Two or more rectangles with equal height can be joined side by side to form a new rectangle with that same height and with width equal to the sum of the widths of the rectangles being joined.

Theorem 5.3b: For $n \in \{2, 3, 4, \ldots\}$, n rectangles with widths $w_1, w_2, w_3, \ldots, w_n$ and with common height h can be joined side by side to form a new rectangle with width $w_1 + w_2 + w_3 + \ldots + w_n$ and height h.

In Chapter 3 we introduced the unit square, a square with a side length of one unit. We defined its area to be one square unit. We know that all squares are rectangles, so the unit square can be used in theorem 5.3.

For example, take four unit squares. Join them side by side. Theorem 5.3 implies the result is a rectangle with width four units and height one unit. The Area Sum Postulate implies the area of the rectangle is the sum of the areas of the four unit squares. Since every unit square has an area of one square unit, the area of the rectangle must be four square units. That is, the area of a rectangle of width four units and height one unit is four square units.



This argument is easily generalised.

Take w unit squares. Join them side by side. Theorem 5.3 implies the result is a rectangle with width w units and height 1 unit. The Area Sum Postulate implies the area of this resulting rectangle is the sum of the areas of the w units squares. Since every unit square has an area of one square unit, this gives w square units.

Flawed conclusion: The area of a rectangle w units wide and 1 unit high is w square units.

I've used the label "flawed conclusion" rather than theorem, because there is something wrong with it. Can you spot what is missing?

We failed to specify the values of w for which the conclusion holds. Someone reading that conclusion might think it holds for all possible values of w. The argument we used doesn't work for all values of w.

I am not saying that the statement isn't true for all possible values of w. I am saying that, so far, we have not *proved* that it is true for all possible values of w. So, there are some values of w for which we now know the statement to be true. There are other values of w for which we don't yet know whether the statement is true or false.

For which values of w can we claim the statement to be proved?

Theorem 5.3a referred to joining together 2 or more rectangles. Theorem 5.3b referred to joining n rectangles where $n \in \{2, 3, 4, \ldots\}$. These are equivalent, so it doesn't matter which theorem we use. Our argument is valid for $w \in \{2, 3, 4, \ldots\}$.

Also, if $w = 1$ then a rectangle of width w and height one is simply the unit square, which by definition has area one square unit. That is, the flawed conclusion stated above does also give the correct answer when $w = 1$.

Putting those two cases together, our conclusion is valid for any $w \in \{1, 2, 3, \ldots\}$, which is the set of positive integers. If you are not familiar with the word "integer", see Appendix A.

Here is the resulting theorem, properly qualified to clearly state the range of values for which we are sure it works.

Theorem 6.1: For any positive integer w, the area of a rectangle with width w units and height 1 unit is w square units.

6.03 General Integer Dimensions

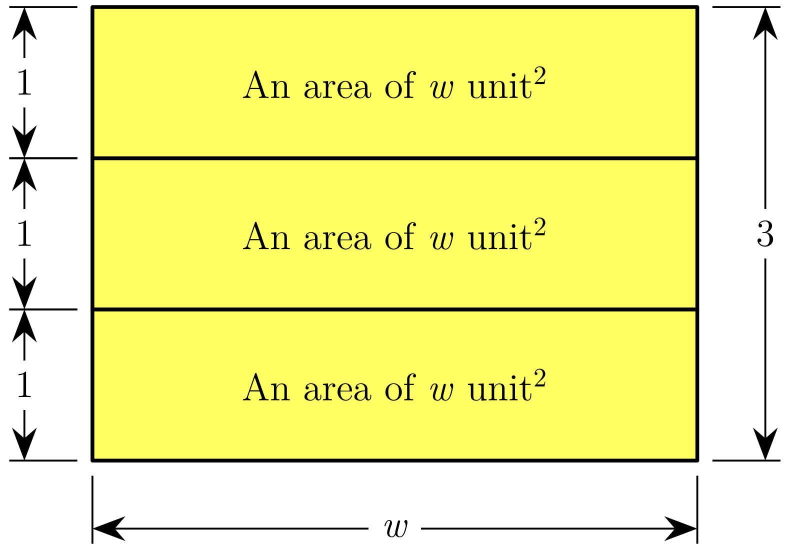
Theorem 6.1 is not very powerful, since it only works for rectangles with positive integer width and height one.

How can we use Theorem 6.1 to ascertain the area of a rectangle with height 3 units, but with the width still being any positive integer w?

After that, the next step is to extend the theorem to cover rectangles where the height and width are both positive integers, so if you can see how to jump directly to that more general case, feel free to do so.

Take 3 rectangles of positive integer width w units and height 1 unit. Join them top to bottom as shown in the diagram. Theorem 6.1 says that they will each have area w units^2. Theorem 5.4 says that when they are joined top to bottom as shown here they form a rectangle which will have width w units and height 3 units. The Area Sum Postulate says the area of this new rectangle will be the sum of the areas of 3 rectangles that were joined.

That is, a rectangle with positive integer width w units and height 3 units has an area of $3w$ units^2.



Generalise the above argument so that it joins h rectangles rather than 3 rectangles. State the result as a theorem. Consider carefully the values of h you include in your theorem. In particular, consider whether it should include the case $h = 1$.

Consider h rectangles each with positive integer width w units and height 1 unit, where $h \in \{2, 3, 4, \ldots\}$. Theorem 6.1 says that they will each have area w units^2. Theorem 5.4 says that when they are joined top to bottom they form a rectangle which will have width w units and height h units. The Area Sum Postulate says the area of this new rectangle will be the sum of the areas of the h rectangles that were joined, so the area is wh units^2.

That is, a rectangle with positive integer width w units and height h units, where $h \in \{2, 3, 4, \ldots\}$, has an area of wh units^2.

The argument given above relies on Theorem 5.4. Theorem 5.4a refers to joining two or more rectangles. Theorem 5.4b refers to joining n rectangles where $n \in \{2, 3, 4, \ldots\}$. These are equivalent. Theorem 5.4 doesn't allow $h = 1$, because trying to joining one rectangle makes no sense. Joining requires at least two rectangles. But if we substitute $h = 1$ into the previous paragraph, it matches Theorem 5.1, which we have already proved to be true. Thus we can also include the $h = 1$ case, so the statement is valid for any $h \in \{1, 2, 3, \ldots\}$, meaning any positive integer h. Let's state that result as a theorem.

Theorem 6.2: For any positive integers w and h, the area of a rectangle with width w units and height h units is wh square units.

When $h = 1$ theorem 6.2 reproduces theorem 6.1. That is, theorem 6.1 is a special case of theorem 6.2, or we could say that theorem 6.2 is more general than theorem 6.1. This means that theorem 6.2 supersedes 6.1. That is, now that we have theorem 6.2, we no longer *need* theorem 6.1, because theorem 6.2 will deal with every case that theorem 6.1 could handle.

6.04 One dimension at a time

One technique for attacking a hard problem is to try break it up into two or more simpler problems. That is what we have done here. We started with the definition of the unit square. We now have a formula for the area of a rectangle with positive integer dimensions w and h, but we got there by splitting the problem into two simpler steps. The two steps are actually the same process of stacking rectangles, just performed in different directions.

1. In Section 6.2 we stacked rectangles — which happened to be unit squares — horizontally to find the area of a rectangle that was w unit squares wide and one unit square high.

2. In section 6.3 we stacked copies of that rectangle vertically, to find the area of a rectangle that was exactly h of those rectangles high.

Incidentally, when deriving the formula for volumes of rectangular prisms, the time saved by this approach is even greater. Area is a concept that applies to two dimensional shapes, and above we developed an argument around stacking in one dimension, and then essentially reused the argument to deal with the second dimension. Volume is a concept that applies to three dimensional solids, and the early steps of developing formulae for volumes are analogous to what we have done so far with areas. We define a unit cube that has a volume of one cubic unit. We stack several cubes in a row to find the volume of a square prism of integer width w, and height and depth 1. Then we get to repeat that argument twice to cover the other two dimensions, producing a formula for the volume of a rectangular prism with integer dimensions w, h and d.

So far our theorem is only proved for positive integer height and width. Our next step will be to investigate whether we can generate a similar result when the two dimensions are positive rational numbers, and we'll deal with that one dimension at a time. So first we'll consider positive rational widths, while still requiring the height to be a positive integer, and we *will* go through that process in painstaking detail, because it contains a lot of new ideas. Essentially we'll be doing some clever stacking in the horizontal direction. But when that is complete, we can go through the step of dealing with positive rational heights far more quickly, because we just need to explain that we are repeating the clever stacking in the vertical direction.

6.05 Rational width examples

If you are not familiar with the term “rational number”, refer to Section A.03 in Appendix A.

Here again is what we have proved so far.

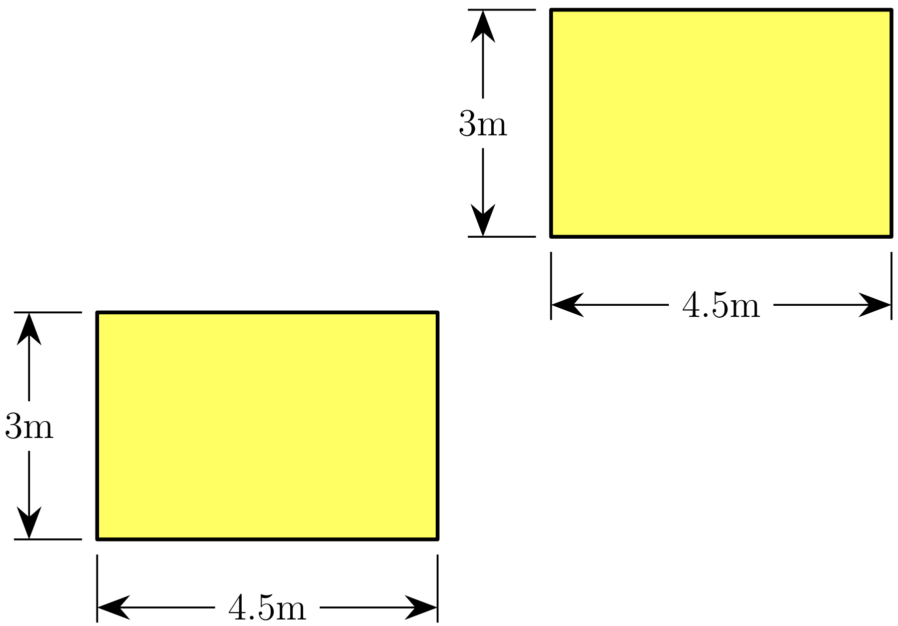
Theorem 6.2: For any positive integers w and h, the area of a rectangle with width w units and height h units is wh square units.

We will now investigate whether the theorem remains valid for any positive rational number w? Before trying to answer that, let’s look at some examples. In these examples, we want the height to still be a positive integer, but the width to be a positive rational number that is not an integer. For variety, the first example will specify particular units.

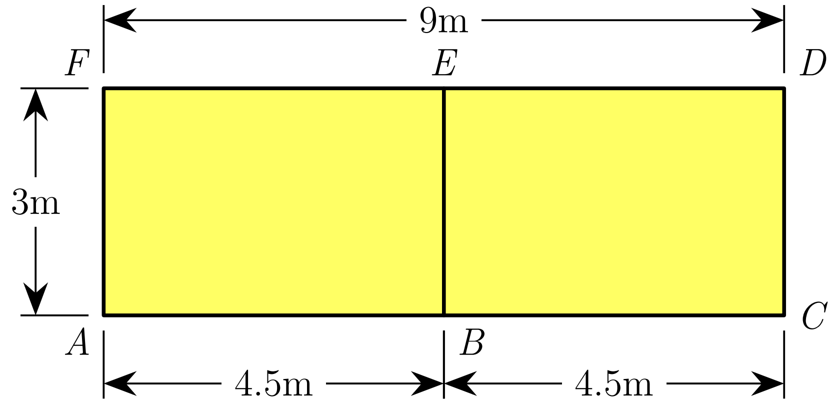
How can we use theorem 6.2, which only works for positive integer width, to calculate the area of rectangle 4.5 metres wide and 3 metres high?

For those who didn't manage to solve the previous challenge, here's an extra hint.

Could you do anything useful if I gave you two copies of the rectangle?



The two rectangles have the same height, so we can join them side by side. Label the vertices as shown.



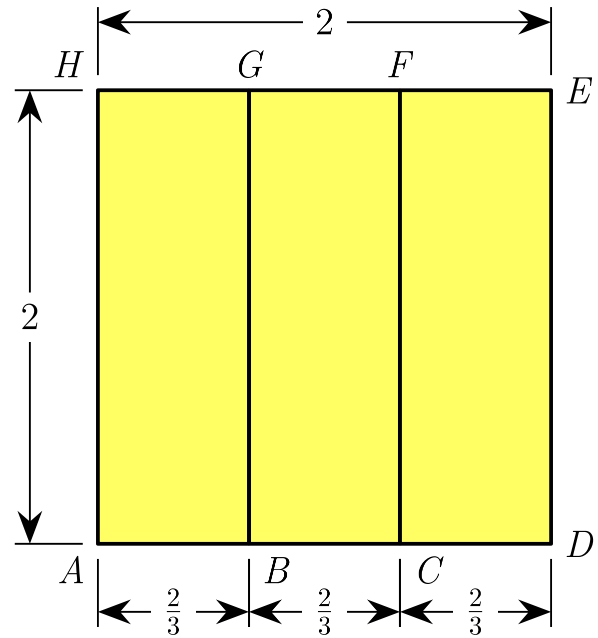
Since $ABEF$ and $BCDE$ are rectangles with the same height, theorem 5.3 applies and tells us that the result of joining them is a rectangle with width 9 metres and height 3 metres. This is the rectangle $ACDF$. Since its dimensions in metres are positive integers, theorem 6.2 applies and tells us its area can be found as the product of its width and height, so its area is 27m^2.

The Area Sum Postulate tells us that the area of rectangle $ACDF$, is also the sum of the areas of rectangles $ABEF$ and $BCDE$. But rectangles $ABEF$ and $BCDE$ are congruent, so they have equal area. Hence each one must have an area of half of 27 square units, which is 13.5 square metres. This happens to be the product of the width and height of each small rectangle.

Let's try another example and see if the pattern still holds.

Use theorem 6.2, which only works if the width and height are positive integers, to calculate the area of rectangle of width $\frac{2}{3}$ units and height 2 units.

This time we need to use three copies of the rectangle joined side by side, so that the width in metres of the new rectangle formed is a positive integer.



Theorem 5.3 still applies, and tells us that the result of joining the three narrow rectangles with constant height is a rectangle with that same height of 2 units and a width in units of $3 \times \frac{2}{3} = 2$. That is, rectangle $ADEH$ is also a square. But at the moment the important thing is that it is a rectangle with dimensions that are positive integers, so theorem 6.2 applies, telling us it has area equal to the product of its width and height, which is 4 square units.

The 3 narrow rectangles are congruent and so have equal area. The Area Sum Postulate says their areas must sum to 4 square units, so each narrow rectangle has area $\frac{4}{3}$ square units. Once again, this is the product of the width and height of each narrow rectangle.

It seems plausible to propose the hypothesis that theorem 6.2 doesn't just hold for positive integer widths, but also holds for positive rational widths. It's time to stop playing with particular examples and try to prove this hypothesis.

6.06 General rational width

Consider a rectangle with positive rational width w and positive integer height h units. Since w is a positive rational number we can express it as $w = \frac{p}{q}$, where p and q are positive integers.

In fact, we can be slightly more restrictive than that. Theorem 6.2 covers the cases where w is a positive integer, so for this next proof we only need to consider the cases where w is a positive rational number that is not also an integer.

> How would that extra restriction affect the values taken by p or q? Once we make that adjustment, how many copies of the rectangle should we join together horizontally to produce a rectangle which is certain to have positive integer dimensions?

The positive integers $\{1, 2, 3, \ldots\}$ can be written as the fractions $\{\frac{1}{1}, \frac{2}{1}, \frac{3}{1}, \ldots\}$. So if we now want to consider the positive rational numbers $w = \frac{p}{q}$ that are not integers, we can exclude the cases where $q = 1$.

That is, we need to allow p to be any positive integer, so $p \in \{1, 2, 3, \ldots\}$, but we only need to consider cases where $q \in \{2, 3, 4, \ldots\}$.

Recall that theorem 5.3 relates to joining together two or more rectangles. It doesn't make sense to talk about joining one rectangle together. We need at least two to do any joining. Luckily the previous paragraph says that q is at least two, so it is valid to join together q copies of the rectangle. Since the copies are all congruent, they have equal height, so theorem 5.3 applies. It tells us that the result of joining together these q rectangles is a new rectangle. The height of the new rectangle is the same as the rectangles being joined, which is h. The width of the new rectangle is the sum of the widths of the rectangles being joined. All q rectangles have width $w = \frac{p}{q}$, so the new rectangle has width $q \times \frac{p}{q} = p$.

Since the new rectangle has width p and height h, both of which are positive integers, theorem 6.2 applies, telling us that its area is ph.

The q rectangles we joined are congruent and so must have equal area, and their combined area is ph, so each of the congruent rectangles has area $\frac{ph}{q} = \frac{p}{q} \times h = wh$.

This proves a rectangle has area wh when the width w is a positive non-integer rational number and the height h is a positive integer. Theorem 6.2 says the area is wh when w and h are both positive integers. Combining those two cases allows us to state a theorem that covers all positive rational widths and positive integer heights.

Theorem 6.3: For any positive rational number w and positive integer h, the area of a rectangle with width w units and height h units is wh square units.

Theorem 6.3 supersedes theorem 6.2. The set of positive integers is a subset of the set of positive rational numbers. Theorem 6.3 allows the width to be any positive rational number, which includes all the cases where the width is a positive integer, so it also covers all those cases covered by theorem 6.2. That is, now that we have theorem 6.3, we no longer need theorem 6.2.

6.07 General Rational Dimensions

Theorem 6.3 allows the width to be a positive rational number, but still restricts the height to be a positive integer. The obvious next step is to relax that restriction, allowing the height to also be any positive rational number. This is the point where we simply rerun the argument of the previous section in the other dimension, so a shorter explanation will suffice.

Consider a rectangle with width w and height h, where w is any positive rational number. Theorem 6.3 already covers the cases where h is a positive integer. We now want to deal with the cases where h is a positive rational number but not an integer, so we let $h = \frac{p}{q}$, where $p \in \{1, 2, 3, \ldots\}$ and $q \in \{2, 3, 4, \ldots\}$.

Can you complete the proof from here?

Take q copies of the rectangle. Theorem 5.4 says that if we join them together vertically they form a rectangle with width w and height $qh = q \times \frac{p}{q} = p$. This new rectangle has positive rational width and integer height, so theorem 6.3 applies, telling us it has area wp. Thus each of the q identical rectangles has area $\frac{wp}{q} = w \times \frac{p}{q} = wh$, as required. Now combine theorem 6.3, which covers positive integer height, with this new argument, that covers heights which are positive rational numbers but not positive integers. The result is a theorem covering all positive rational heights.

Theorem 6.4: For any positive rational numbers w and h, the area of a rectangle with width w units and height h units is wh square units.

Theorem 6.4 supersedes theorem 6.3. That is, all the cases there were covered by theorem 6.3 are covered by theorem 6.4, so we no longer need theorem 6.3.

6.08 A particular irrational width

Theorem 6.4 covers rectangles with positive rational dimensions. We will now show that the theorem will also hold true for positive irrational dimensions, and so it holds for all positive real dimensions. In case you are not familiar with them, the terms terms "irrational number" and "real number" are explained in more detail in Appendix A.

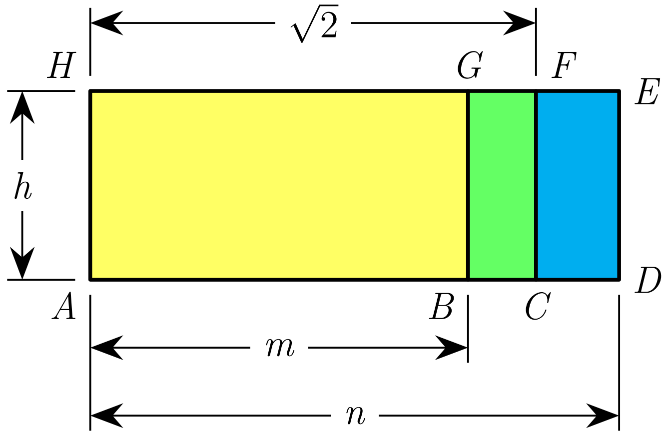
The arguments we will look at in this section are more difficult than those in the previous sections of this chapter. If you are having trouble with them, you can defer reading this section for a while, since they won't be needed for the next few chapters. We will need them when we find the area of a circle, (and I haven't written that chapter yet). If you do decide to skip these arguments for the moment, jump ahead to section 6.11.

Because the method for dealing with irrational numbers is quite challenging, let's start by looking at a particular irrational width, $\sqrt{2}$. Then later we will generalise the argument to use an arbitrary irrational number w.

Expressed as a decimal, $\sqrt{2} = 1.414213562373\ldots$. Because it is irrational, the decimal does not terminate and the digits never settle into a pattern that repeats forever.

The argument in this section hinges on the fact that we can always find rational numbers arbitrarily close to any irrational number. Appendix A includes an explanation of this property.

Consider three rectangles all with rational height h, but different widths. The three rectangles have been stacked together so that their left-hand sides overlap on the line AH. They have been stacked in increasing order of width, so the widest is at the back.



The narrowest (least wide) rectangle $ABGH$ has rational width m and has been coloured yellow. Since it is at the front of the stack it is fully visible. Since both its dimensions are rational, theorem 6.4 applies and so we know its area is mh square units.

Behind that is rectangle $ACFH$ which has width $\sqrt{2}$. It is has been coloured green, but most of it is hidden behind the yellow rectangle. We can only see the narrow green strip $BCFG$ that extends beyond the yellow rectangle. Since its width is irrational, theorem 6.4 does not apply, so we don't know what its area is. We might *suspect* that its area will be $\sqrt{2}h$, but so far we haven't *proved* that. I'll call this the target rectangle, since our task here is to derive a formula for its area. Let its area be x square units.

Behind that is the widest rectangle $ADEH$ which has rational width n. It has been coloured blue, but most of it is hidden the yellow and green rectangles. Since both its dimensions are rational, theorem 6.4 applies and so we know its area is nh square units.

By the Area Sum Postulate, the area of $ACFH$ equals the sum of the areas of $ABGH$ and $BCFG$. This means $ACFH$ has a larger area than $ABGH$. Similarly the area of $ADEH$ equals the sum of the areas of $ACFH$ and $CDEF$. This means $ADEH$ has a larger area than $ACFH$. That proves something which you might regard as obvious: if rectangles have the same height, those with a larger width have greater area.

Let's write some inequalities summarising these results. The widths tell us $m < \sqrt{2} < n$. The areas tell us $mh < x < nh$. I'll call these the width inequality and area inequality.

Note that $m < \sqrt{2} < n \implies mh < \sqrt{2}h < nh$. Thus x, the area of the target rectangle, is constrained to the range $mh < x < nh$, and that constraint allows the possibility that $x = \sqrt{2}h$.

While we don't know the value of x, we could estimate it using appropriate values of m and n that meet the restrictions $m < \sqrt{2} < n$, with m and n both rational.

Earlier I noted that $\sqrt{2} = 1.4142135623 73\ldots$, so $1.4142 < \sqrt{2} < 1.4143$. We need to ensure m and n are rational numbers, and terminating decimals are rational numbers, so set $m = 1.4142$ and $n = 1.4143$. That is, I've set m and n as close as I can to $\sqrt{2}$ while only using 4 decimal places. The area inequality tells us $1.4142h < x < 1.4143h$. That restricts x to a fairly narrow range, a range that happens to include the value $\sqrt{2}h$.

But if that range isn't accurate enough for you, I can make it more accurate by including another decimal place in the values of m and n. Let's now reset m and n using 5 decimal places, still keeping them as close to $\sqrt{2}$ as possible.

$1.41421 < \sqrt{2} < 1.41422$. Set $m = 1.41421$ and $n = 1.41422$. The area inequality now tells us $1.41421h < x < 1.41422h$, a range that still includes the value $\sqrt{2}h$. In terms of the diagram, I've made the yellow rectangle wider, though still narrower than the target rectangle, and I've made the blue rectangle narrower, though still wider than the target rectangle. As a result of this, I've reduced the range of possible values for x to $\frac{1}{10}$ of what is was two paragraphs ago.

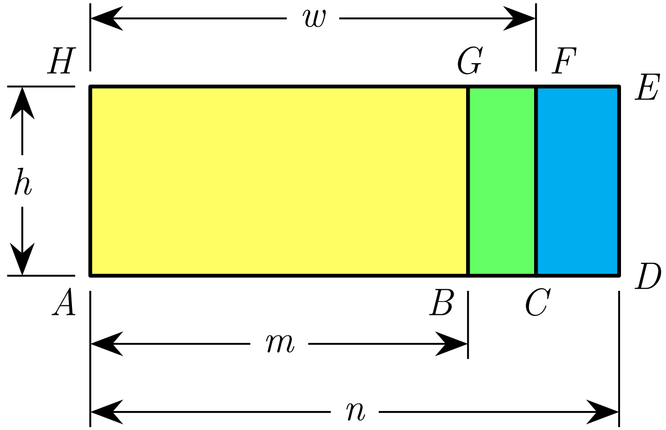
And I can keep going. Every time I include an extra decimal place in the values of m and n:

- The gap between m and n shrinks to $\frac{1}{10}$ its previous value, while keeping $m < \sqrt{2} < n$. That is, m and n are constrained closer and closer to $\sqrt{2}$.
- The range of possible values for x shrinks to $\frac{1}{10}$ its previous value, but the range always includes $\sqrt{2}h$. That is, the range of possible values for the area of the target rectangle is constrained closer and closer to $\sqrt{2}h$.

The mathematical jargon for this is: As m and n converge on $\sqrt{2}$, the range of possible values for x converges on $\sqrt{2}h$.

Because I can keep narrowing these ranges indefinitely I can show the area falls in a range arbitrarily close to $\sqrt{2}h$. This can only happen if the area of the target rectangle is $\sqrt{2}h$. That is, the area of a rectangle with irrational width $\sqrt{2}$ and rational height h is the product of its width and height.

6.09 A general irrational width



Now let's replace $\sqrt{2}$ by an arbitrary positive irrational number w. That is, consider a rectangle $ACFH$ with irrational width w and rational height h.

As before, let the area of the target rectangle $ACFH$ be x square units. Set m and n to rational numbers close to w with $m < w < n$. Theorem 6.4 still applies to the rectangles with widths m and n, giving their areas to be mh and nh, so $mh < x < nh$.

Note that $m < w < n \implies mh < wh < nh$. Hence x, the area of the target rectangle, is constrained to the range $mh < x < nh$, and that constraint definitely allows the possibility that $x = wh$.

Now we want to gradually reset m and n closer and closer to w, while still restricting them to rational numbers. One way to do this is to first set m and n as close as possible to w while using only 4 decimal places. Then reset them as close as possible while using 5 figures decimal places.

Keep going. Every time I include an extra decimal place in the values of m and n:

- The gap between m and n shrinks to $\frac{1}{10}$ its previous value, while keeping $m < w < n$. That is, m and n are constrained closer and closer to w.
- The range of possible values for x shrinks to $\frac{1}{10}$ its previous value, but the range always includes wh. That is, the range of possible values for the area of the target rectangle is constrained closer and closer to wh.

That is, as m and n converge on w, the range of possible values for x converges on wh.

Because we can keep narrowing these ranges indefinitely we can show the area falls in a range arbitrarily close to wh. This can only happen if the area of the target rectangle is wh. That is, the area of a rectangle with irrational width w and rational height h is wh.

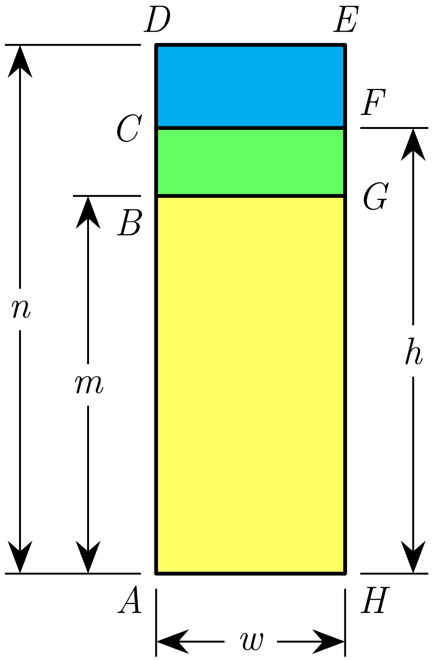
Theorem 6.4 states that for any positive rational numbers w and h, the area of a rectangle with width w units and height h units is wh square units. We've just proved the result also holds for positive irrational w. Putting those two cases together, the result holds for any positive real number w. Here is the resulting theorem, which supersedes theorem 6.4.

Theorem 6.5: For any positive real number w and positive rational number h, the area of a rectangle with width w units and height h units is wh square units.

6.10 Irrational Height

The final step is to allow the height to be irrational. Since this is the third time we are going to take an argument used in the horizontal direction and reuse it in the vertical direction, it would be tedious to do it all again in full detail. So here is just a very quick summary of the argument. If you are already happy that the argument will work in the vertical direction, you can skip the next three paragraphs entirely.

Consider three rectangles all with positive real width w. Let their heights be m, h and n, with $m < h < n$. Make m and n rational, so by theorem 6.5 the areas of the shortest and tallest rectangles are wm and wn. Let h be irrational. The rectangle with height h is the target rectangle. Let its area be y, so $wm < y < wn$. Note that $m < h < n \implies wm < wh < wn$. Hence y, the area of the target rectangle, is constrained to the range $wm < y < wn$, and that range definitely allows the possibility that $y = wh$.



Move m and n closer and closer towards h, but taking only rational values so their areas can still be found by theorem 6.5. As this happens, wm and wn move closer together, always maintaining the relationship $wm < y < wn$. That is, as m and n converge on w, the range of possible values for y converges on wh.

Because we can keep narrowing these ranges indefinitely we can show the area falls in a range arbitrarily close to wh. This can only happen if the area of the target rectangle is wh. That is, the area of a rectangle with positive real width w and irrational height h is wh.

Theorem 6.5 states that for any positive real number w and any positive rational number h, the area of a rectangle with width w units and height h units is wh square units. We've just proved the result also holds for positive irrational h. Putting those two cases together, the result holds for any positive real number h. We could state the resulting theorem like this.

For any positive real numbers w and h, the area of a rectangle with width w units and height h units is wh square units.

But instead, I'm going to write it as follows.

Theorem 6.6: The area of a rectangle with width w units and height h units is wh square units.

Why is it "safe" to omit the references to "positive real numbers"?

In a theorem, placing a restriction on a pronumeral is a way to warn the reader that the theorem is not guaranteed to be safe to use in all situations. Here is a theorem from earlier in the chapter.

Theorem 6.4: For any positive rational numbers w and h, the area of a rectangle with width w units and height h units is wh square units.

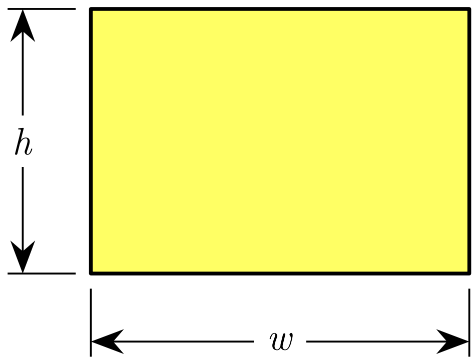
When we wrote that theorem we had only proved the wh formula valid for positive rational dimensions. But rectangles with positive irrational dimensions do exist. If we hadn't included the restriction that the formula should only be used for rational dimensions, there was a risk that someone might inadvertently apply it to a rectangle with irrational dimensions, which could have had adverse consequences if the formula didn't work for irrational values.

But now we have proved that the wh formula *does* work for all positive real dimensions. The dimensions of a rectangle are *always* positive real numbers, so theorem 6.6 works for every rectangle. There isn't any risk that someone might accidentally apply this theorem to a rectangle that has a zero dimension or a negative dimension, because such rectangles do not exist. Thus its safe for the theorem to omit the restriction that the pronumerals must be greater than zero.

Theorem 6.6 supersedes theorem 6.5. Theorem 6.6 works for every rectangle that can exist, so we don't need any of the earlier theorems in this chapter that only worked for some rectangles.

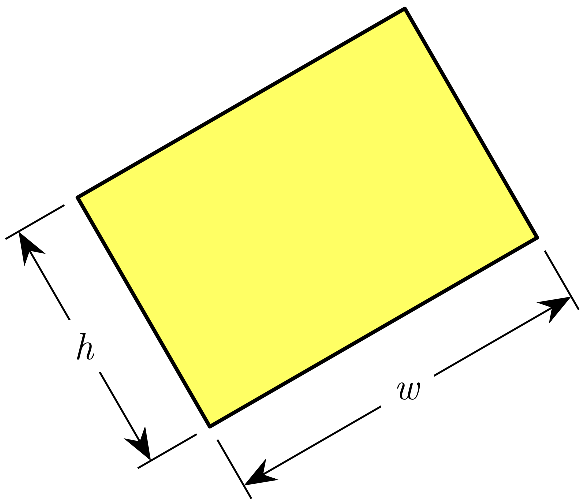
6.11 Changing the orientation

This chapter used a rectangle that was in a vertical plane with its base horizontal. Provided you are reading this book on a computer screen oriented vertically, the left and right sides of the rectangles do run vertically.



I chose this orientation because, if I label the two dimensions of the rectangle as “width” and “height,” there is no confusion about which is which. The height is the length of the sides running vertically.

We proved that, if the rectangle has width w units and height h units, its area is wh square units.



Any other rectangle with dimensions w by h will be congruent to this rectangle, and so will also have area wh square units. This is true no matter what orientation that other rectangle takes. It could still be in a vertical plane, but rotated so that none of its edges are horizontal, as shown here. It could be sitting in a horizontal plane or a sloping plane, rather than in a vertical plane. If that happens, we might choose to use different words to describe the lengths of its sides and denote those lengths by different pronumerals.

Whatever pronumerals we use, the area can be found as the product of those two pronumerals. This result applies in any orientation, so some people prefer to state the resulting theorem in a way that doesn't imply a particular orientation. This can be done as follows.

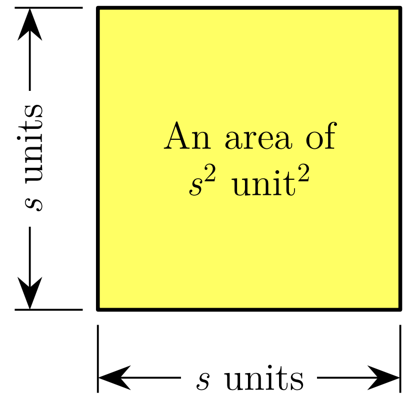
Theorem 6.7: The area of a rectangle is the product of the lengths of any two adjacent sides.

6.12 It gets easier

The formula for the area of a rectangle is simple, so students are often surprised by how much work is involved in proving the formula valid. We had to separately work through cases with integer dimensions, then rational dimensions, and then irrational dimensions. The good news is that for this book, that is the last time we have to separately derive formulae for those three cases.

The next step is to use the rectangle formula to develop formulae for the areas of other common shapes. Since the rectangle formula works for all positive real dimensions, the formulae we develop from it also automatically work for all positive real dimensions. For example, when we deal with triangles, we will *not* have to consider separate cases for triangles with integer, rational and irrational dimensions.

7 Square



All squares are rectangles, so the formula for the area of a rectangle will also be valid for the area of a square. That is, the area of a square will be the product of the length of any two adjacent sides. But for a square all four sides have the same length, leading to the following theorem.

Theorem 6.1: The area of a square of with side length s units is s^2 square units.

This theorem is so pervasive that it has even influenced the way we pronounce a particular exponent.

Recall that x^n is pronounced "x to the power n" or "x to the n^{th} power." For example, x^5 could be pronounced "x to the 5^{th} power." If we are in a hurry, we might even abbreviate that to "x to the 5^{th}."

To be consistent, x^2 would be pronounced "x to the 2^{nd} power," but instead we pronounce it "x squared," because it gives the area of a square of side length x.

Similarly x^3 is pronounced "x cubed," since it gives the volume of a cube of side length x.

8 Area unit conversions

8.01 Generic square units

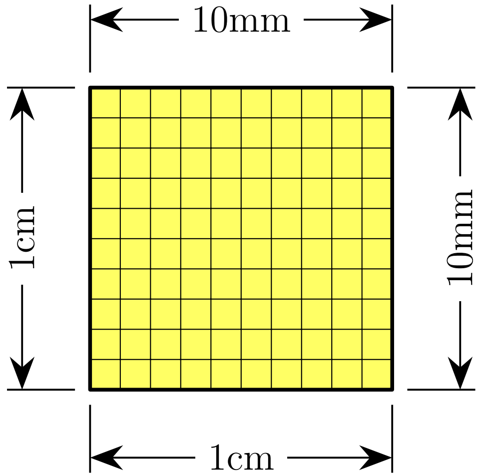
If we square both sides of the equation $a = b$, we obtain the result $a^2 = b^2$.

A similar process holds for equations that include units, though we must remember to also square the units. For example:

$$\begin{aligned} 10\text{mm} = 1\text{cm} &\implies 100\text{mm}^2 = 1\text{cm}^2 \\ 100\text{cm} = 1\text{m} &\implies 10,000\text{cm}^2 = 1\text{m}^2 \\ 1,000\text{m} = 1\text{km} &\implies 1,000,000\text{m}^2 = 1\text{km}^2 \end{aligned}$$

While the above argument may be mathematically valid, not everyone finds it convincing. Some people say that the process leaves them with a nagging suspicion that the author has done something dodgy in the argument, though they can't quite pinpoint exactly what the dodgy part is. These reactions are important. A proof shouldn't just be valid—it should be convincing. If someone finds a proof unconvincing, a good mathematician will try to find a more convincing explanation. Here, the formula for the area of a square allows us to do this.

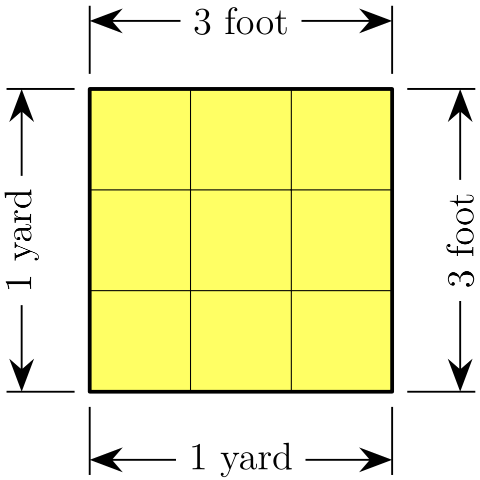
Consider a square with a side length of 1cm, which is also 10mm. The diagram shows it an enlarged scale for clarity. If you are concerned that this scaling might make the following proof invalid, you can make your own copy of this diagram at an accurate scale and use that instead.



We know the area of a square with side length s units is s^2 square units. This is true for any unit we choose to use.

If we use centimetres as the unit, the square has a side length of 1cm and hence has an area of 1 square centimetre. If we use millimetres as the unit, the square has a side length of 10mm and hence has an area of 10^2 square millimetres. Hence 1 square centimetre must equal 100 square millimetres, which is consistent with the conclusion given above

We can also draw a 1mm grid on the diagram, clearly splitting the square centimetre into 100 squares, each with area 1 square millimetre. Anyone who was not convinced by the previous paragraph can be invited to count the squares. Individually counting the 100 squares is tedious, but given that units in the metric system go in steps of 10, this is the simplest metric example we can give.



We can produce an easier example if we delve into pre-metric dimensions, where two common units of length were the foot, plural feet, and the yard. There are 3 feet in a yard. Hence one square yard contains 3^2 or 9 square feet.

Here is a similar diagram showing the subdivision of a square yard into 9 square feet, scaled down considerably to fit on the page.

The ruler most commonly used in school has the unusual length of 30cm. Why not 25cm or 50cm? This is a hangover from pre-metric days when the common measuring device was the "foot rule". One foot is about 30.5cm. While they are becoming rare, you might still be able to find a 30cm ruler that includes an alternate scale showing 1 foot subdivided into 12 inches.

8.02 Special area measures

Most area units are just squares of length units. For example, lengths can be measured in centimetres, metres or kilometres. Areas can be measured in square centimetres, square metres and square kilometres. But the metric system has one unit specifically defined for areas: the hectare. Note that it is just "the hectare", not "the square hectare".

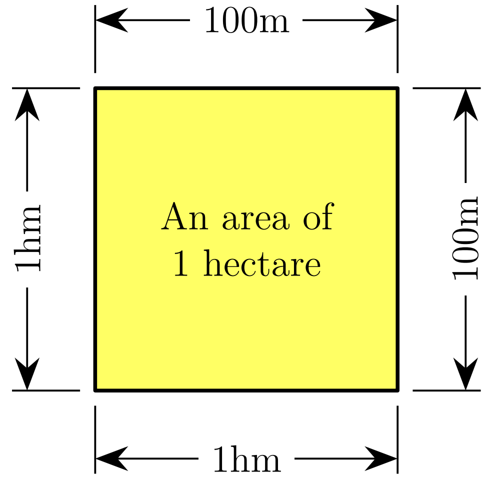
One hectare is defined to be 10,000 square metres.

In Australia, the areas of farms are usually measured in hectares. If you don't plan on owning a farm, you will probably never use hectares in real life. Residential housing blocks are measured in square metres, because the hectare is too large to be convenient.

The rarely used metrix prefix "hecto" denotes 100. For example, a hectogram is 100 grams, though you are unlikely to ever see that unit being used. In common practice, masses are measure in milligrams, grams or kilograms.

Given that the prefix hecto denotes a factor of 100 rather than 10,000, can you spot why it appears, slightly truncated, in the word hectare?

Note that $10,000 = 100^2$. A hectometre is 100 metres. A square with a side length of 100 metres has an area of 10,000 square metres. Thus the hectare is the area of a square with a side length of 1 hectometre. The truncation of the prefix from “hecto” to “hect” is simply because “hectoare” would be hard to pronounce.



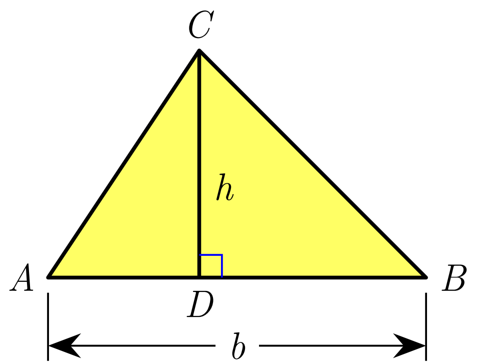
Prior to the introduction of the metric system in Australia, the available units for area included the rood, which was the area of a rectangle of 1 furlong by 1 rod, and the acre which was the area of a rectangle of 1 furlong by 1 chain. These units were chiefly used for measuring fields in agriculture.

9 Triangle

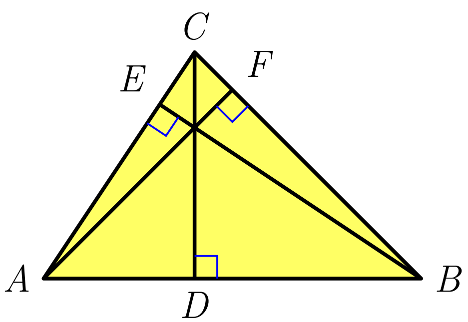
9.01 Triangle jargon

Reminders: A right angle is 90°. An acute angle is less than 90°. An obtuse angle is between 90° and 180°. The three angles of a triangle sum to 180°.

This diagram shows triangle ABC, an "acute triangle", meaning that all three angles are acute. Older texts call this an "acute angled triangle", but over time this was abbreviated.



CD is drawn perpendicular to side AB. Line segment CD is called an altitude of $\triangle ABC$. More specifically, it is the altitude on the base AB. Later we show there is an easy way to calculate the area of this triangle given the two measurements $|AB|$ and $|CD|$. When we do this, it is common to denote the length of the base AB by the pronumeral b and to call the length of the altitude BD the height of the triangle, denoted by the pronumeral h.



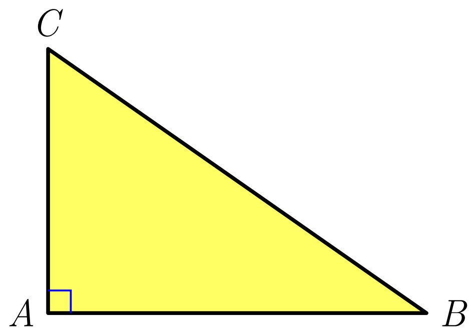
Triangles have three altitudes. This diagram shows the three altitudes of $\triangle ABC$. Line segment BE is the altitude on the base AC. Line segment AF is the altitude on the base BC.

Curiously, the 3 altitudes meet in a single point, called the orthocentre. While that is interesting, it isn't relevant in deriving a formula for the area, so I won't prove it here.

In earlier chapters I oriented rectangles in a vertical plane with their bases horizontal, so the word "height" referred to a distance measured vertically. By contrast, for triangles there is a tradition of using "height" to mean the length of any useful altitude, no matter what its orientation.

For example, if the distances $|AF|$ and $|BC|$ are both known, we can use them to calculate the area of $\triangle ABC$. When doing this calculation, the tradition is to call $|AF|$ the height of the triangle, even though line segment AF is not vertical.

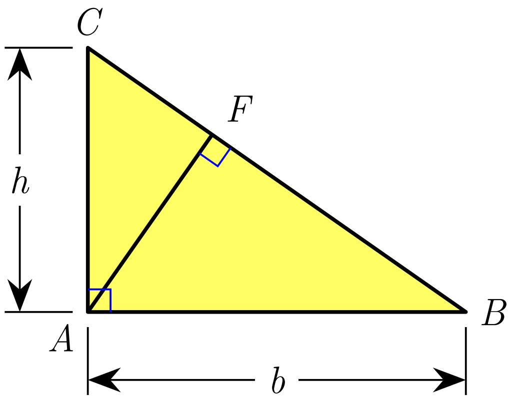
In the acute triangle shown above, drag the point C to the left until $\angle BAC$ becomes a right angle, producing the triangle shown here. A triangle in which one angle is a right angle is called a "right triangle", though older texts call it a "right-angled triangle".



The side opposite the right angle, in this case the line segment BC, is named the hypotenuse. The other two sides are most commonly called the legs of the right triangle, though some texts call them the arms.

The three angles of a triangle sum to 180°, so if one angle is 90° the other two must both be acute.

Draw a rough sketch showing this right triangle's three altitudes.



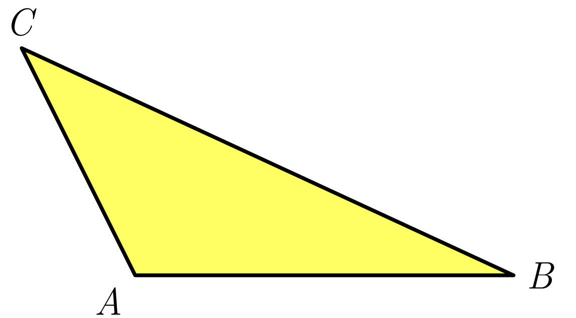
To find the altitude on the base AB we draw a line segment from C meeting the base AB at right angles, but this gives the line segment AC. That is AC is the altitude on the base AB. If we were using this base and altitude to find the area, we would assign the pronumerals b and h as shown in this diagram.

There's another way to think about this. Look back at the earlier diagram showing the three altitudes of the acute triangle. To make a right triangle I dragged the point C to the left until $\angle BAC$ became a right angle. As I drag B to the left, the point D must also move to the left, staying directly below C, to ensure that $\angle CDB$ remains a right angle. The acute triangle becomes a right triangle when C is directly above A, but since D is directly below C, this makes points A and D coincide. That is, the altitude CD becomes the edge CA.

Similarly, the altitude on the base AC is the side AB. AF is the altitude on the base BC. The three altitudes again meet at a single point, the vertex A.

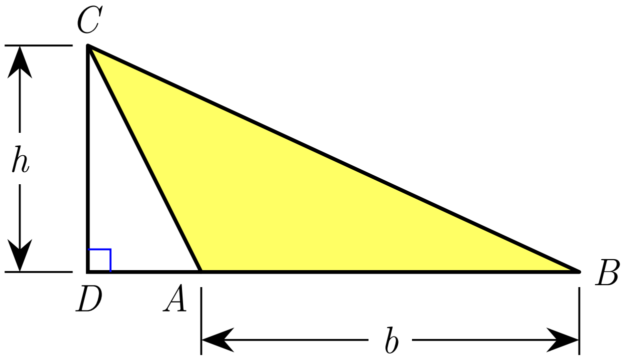
Now let's drag the point C even further to the left, so that $\angle BAC$ is greater than 90°, as shown here. A triangle containing an obtuse angle is called an "obtuse triangle", though older texts call it an "obtuse angled triangle".

The three angles of a triangle sum to 180°, so if one angle is obtuse, the other two must both be acute.



How do we find the altitude on the base AB? Earlier, in the acute triangle we drew a perpendicular from C to the opposite side AB, meeting it at D. Does that work here?

If we try that here, we can't find a suitable point D on side AB, meaning *line segment* AB, but we can find one on *line* AB. The altitude CD now falls outside the triangle.

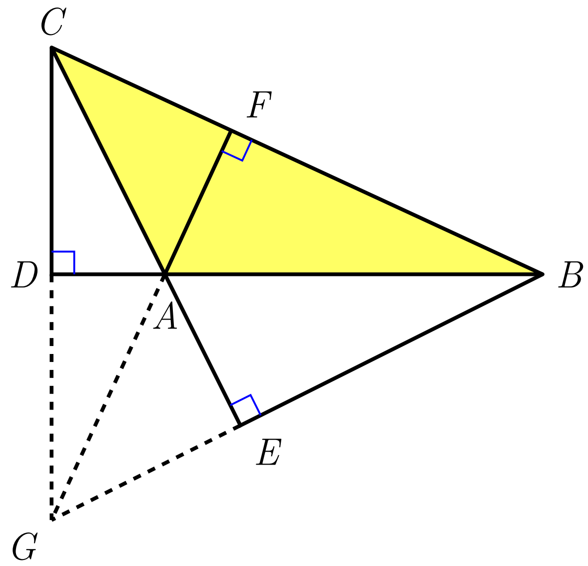


We produced the right triangle and obtuse triangle by starting with the acute triangle and dragging the point C to the left. As we did that, the point D always had to stay directly below C, to ensure CD stays perpendicular to the base AB. When the triangle was an acute triangle, this meant D would fall between A and B. When the triangle became a right triangle it meant D coincided with A. When the triangle became obtuse, it means D is left of A, meaning it falls outside the triangle. As point D moves, so does the altitude AD, being inside the acute triangle, on an edge of the right triangle and outside the obtuse triangle.

Later we will find that the lengths $|AB|$ and $|CD|$ are still useful for finding the area of obtuse triangle ABC, so we still call them the base and height and assign the pronumerals b and h as shown here.

A triangle has three altitudes. For this obtuse triangle, do the other two altitudes fall inside or outside the triangle? Do they intersect?

The altitude AF on the base BC falls within the triangle. The altitude BE on the base AC falls outside the triangle.



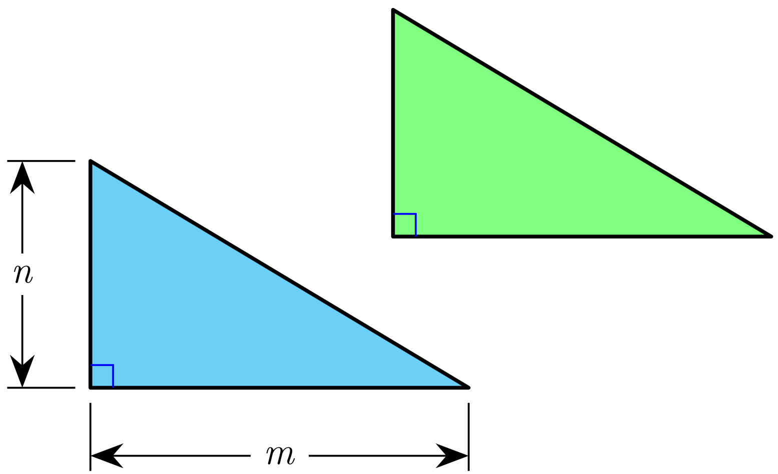
Altitudes are line segments, so the three altitudes of this obtuse triangle do not intersect, but the three lines containing those three altitudes still meet in a single point, labelled G in this diagram, and we still call this point the orthocentre. In general, the orthocentre falls inside an acute triangle, at the right angle in a right triangle and outside an obtuse triangle. While interesting, this result isn't relevant to calculating the area, so I won't be proving it.

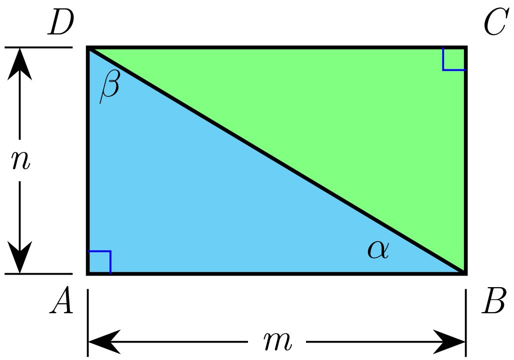
9.02 Area of right triangle given leg lengths

Consider the blue right triangle which has legs of length m units and n units. We seek a formula for the area of this triangle in terms of the pronumerals m and n.

Take a copy of the triangle and shade it green.

Arrange the blue and green right triangles to form a shape of known area.





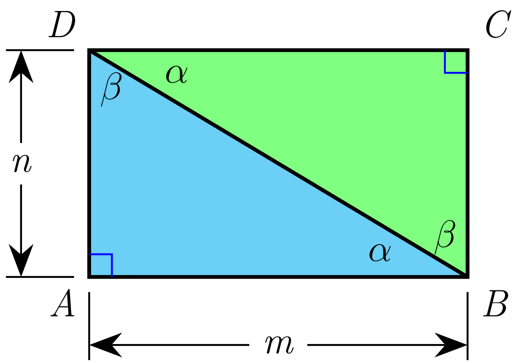
The green rectangle is a copy of the blue triangle, so the two triangles are congruent. This means the two triangles have hypotenuse of equal length. Hence if we rotate the green triangle 180° we can join the two triangles along their hypotenuse.

Quadrilateral $ABCD$ looks suspiciously like a rectangle. If we can prove it is, then we already have a formula to calculate its area.

For convenience, let $\angle ABD = \alpha$ and $\angle ADB = \beta$. Angles are often denoted by lower case Greek letters.

What does the angle sum of a triangle tell you about the relationship between α and β? Identify any other angles that will equal α or β. Hence prove that quadrilateral $ABCD$ is a rectangle.

The three angles of any triangle sum to 180°. In $\triangle ABD$, $\angle BAD = 90°$. Hence $\alpha + \beta = 90°$.



$\triangle BCD$ is a copy of $\triangle DAB$ so the corresponding angles must be equal. Hence

- $\angle BCD = \angle DAB = 90°$
- $\angle CDB = \angle ABD = \alpha$
- $\angle CBD = \angle ADB = \beta$

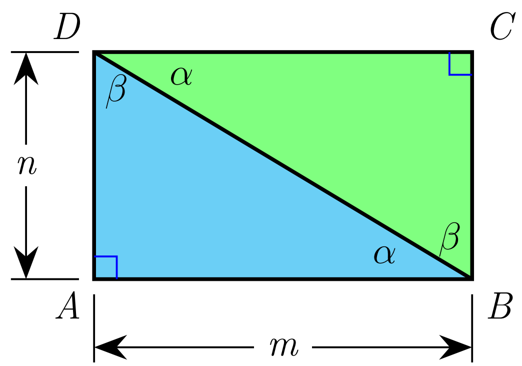
These angles are shown on the diagram. Finally, consider the four angles of quadrilateral $ABCD$.

- The angles at A and C have already been shown to be right angles
- $\angle ABC = \angle ABD + \angle DBC = \alpha + \beta = 90°$
- $\angle CDA = \angle CDB + \angle BDA = \alpha + \beta = 90°$

Since all four angles of quadrilateral $ABCD$ are right angles, it is a rectangle.

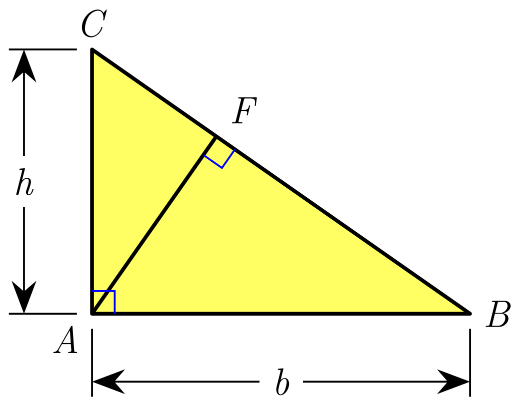
Use the congruency of the two triangles to derive a formula for the area of right triangle ABD.

Since $\triangle ABD$ and $\triangle CDB$ are congruent, they have equal area. By the Area Sum Postulate, the sum of their areas is the area of rectangle $ABCD$. By theorem 6.7, the area of rectangle $ABCD$ is mn square units, so each triangle has area $\frac{1}{2}mn$ square units. The lengths m and n are the leg lengths of the right triangle ABD. Here is the result stated as a theorem.



Theorem 9.1: The area of a right triangle is ½ the product of the leg lengths.

Recall that in a right triangle, as in any triangle, any of the three sides may be chosen as the base, and each base has a corresponding altitude.

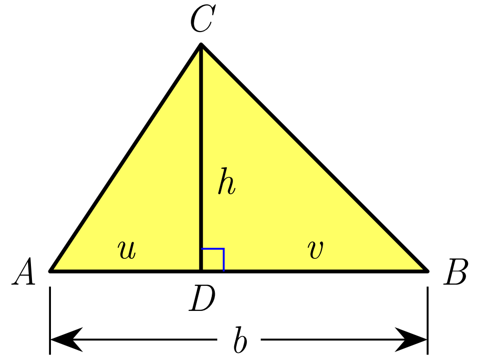


In two of these choices the altitude coincides with an edge of the triangle. AB is the altitude for the base AC and AC is the altitude for the base AB. That is, in these two cases, the base and altitude are the two legs of the right triangle. Theorem 9.1 tells us how to calculate the area of the triangle given those two leg lengths.

The third option is to use the hypotenuse BC as the base. Its altitude is AF, which falls inside the triangle rather than coinciding with a leg. We don't yet know how to calculate the area of the triangle given the lengths $|BC|$ and $|AF|$. It turns out the method to derive a formula for the area of a triangle using an internal altitude is the same for right triangles, acute triangles and obtuse triangles, so in the next section we will combine these three cases.

9.03 Area of a triangle using an internal altitude

Say we know the height of a triangle measured at an altitude falling within the triangle, and we also know the length of the corresponding base. Orient the triangle with the base horizontal at the bottom of the diagram and label it as shown here.

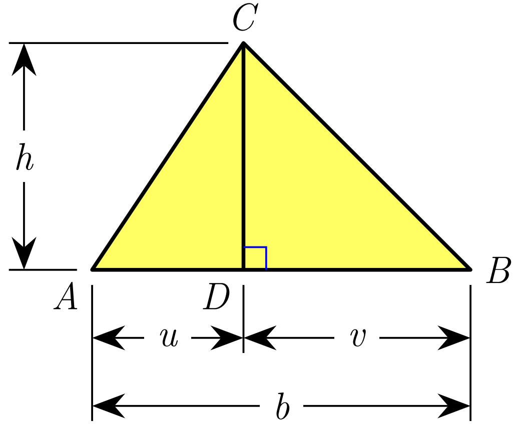


We seek a formula for the area of $\triangle ABC$ in terms of b, the length of base AB, and h, the height measured at the altitude CD.

Let $|AD| = u$ and $|DC| = v$.

These two diagrams are two different ways of displaying the same information.

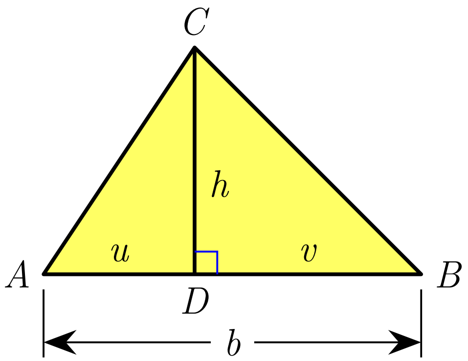
Most mathematicians favour the approach of the first diagram, writing the symbols h, u and v close to the line segments they relate to. The meaning is usually clear enough provided all the symbols refer to the shortest available line segment with labelled end-points. For example, people reading this diagram will almost certainly correctly interpret u as referring to $|AD|$ rather than $|AB|$. However, some people would just write the b pronumeral somewhere below the point D and hope the meaning was clear. I find that far too vague, so I added the extra guidelines and arrows to make it clear that $b = |AB|$.



Most engineers and architects favour the second approach, stacking two sets of dimensions below the diagram. This approach removes any risk of ambiguity. However, it does involve many extra guidelines and arrows, so people who are not used to this approach can find the diagram a little overwhelming at first glance.

You can try these variations in your own diagrams and see which form you prefer. In very complex diagrams you may find using a mix of the two approaches gives the best result.

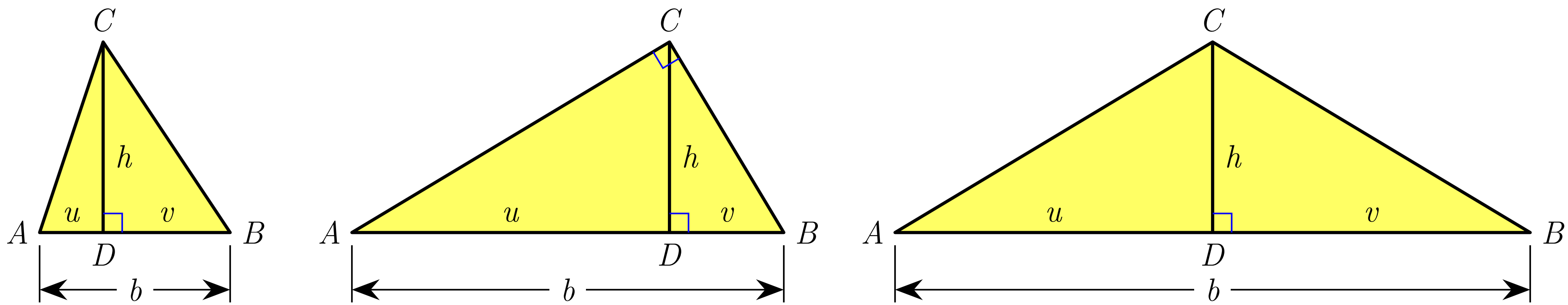
By considering the areas of $\triangle ACD$ and $\triangle BCD$, develop a formula for the area of $\triangle ABC$.



Since $\triangle ACD$ and $\triangle BCD$ are right triangles, theorem 9.1 applies to them, and tells us their areas are half the product of their leg lengths, giving $\frac{1}{2}uh$ and $\frac{1}{2}vh$ respectively. By the Area Sum Postulate:

$$\begin{aligned}\text{Area of } \triangle ABC &= \text{Area of } \triangle ACD + \text{Area of } \triangle BCD \\ &= \tfrac{1}{2}uh + \tfrac{1}{2}vh \\ &= \tfrac{1}{2}(u+v)h \\ &= \tfrac{1}{2}bh\end{aligned}$$

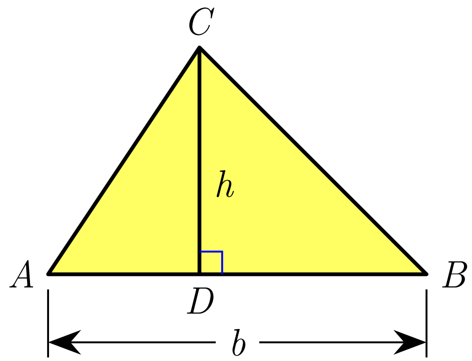
Note that this derivation makes no assumption about the size of angle $\angle ACB$. That is, the result will hold for acute triangles, right triangles and obtuse triangles as shown in the following three diagrams. In the first diagram, $\triangle ABC$ is an acute triangle, in the second it is a right triangle and in the third it is an obtuse triangle. You can reread the derivation given above and verify that it is valid for all three triangles. The three triangles have the same height, but the value of b varies between them, so they do have different areas. But in each case, the area of the triangle can be found as $\frac{1}{2}bh$, using that triangle's value of b.

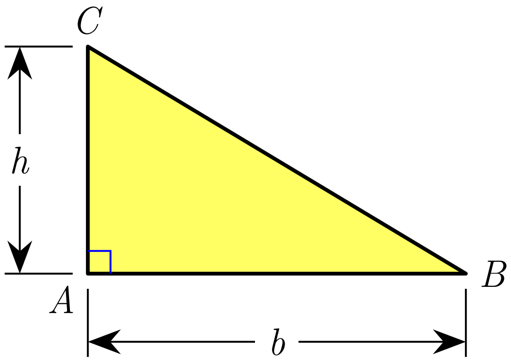


The three triangles also show a variety of relative sizes of u and v. In the first, $u < v$, so D is nearer A than B. The second diagram reverses this with $u > v$ and D nearer B. The third has $u = v$ so D is equidistant between A and B, which makes the triangle isosceles with $|AC| = |BC|$. These changes have no impact on the derivation. The derivation only requires that D is somewhere between A and B, so that the altitude CD does fall within the triangle.

Let's summarise what we now know about the area of triangles with a base length b and height h.

In this section we found that if the altitude used to measure h falls inside the triangle, such as the altitude BD shown to the right, the triangle has area $\frac{1}{2}bh$.





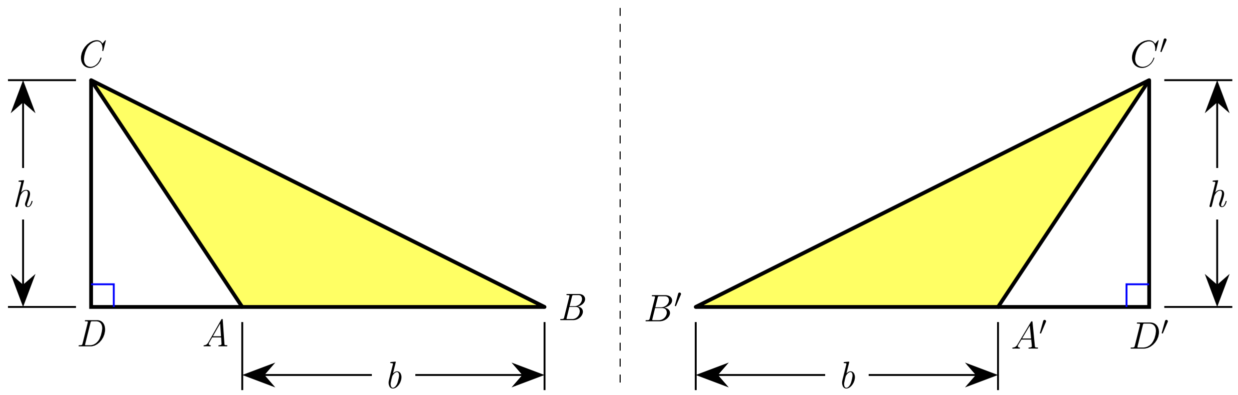
We have previously noted that in the right triangle shown at left, side AC is also the altitude on the base AB. Hence we can use $b = |AB|$ with $h = |AC|$. In the previous section we ascertained that area of a right triangle is ½ of the product of the leg lengths. This means the area is once again $\frac{1}{2}bh$. That's suspicious!

Let's investigate the final case, where the altitude falls outside the triangle, to see whether the $\frac{1}{2}bh$ result also applies there.

9.04 Area of a triangle using an external altitude

Say we know the values of h, the height of a triangle measured at an altitude falling outside the triangle, and b, the length of the corresponding base. Orient the triangle with the base horizontal at the bottom of the diagram. When we do this, the altitude could fall left or right of the triangle.

In this diagram, $\triangle ABC$ has been reflected in the vertical dashed line to produce $\triangle A'B'C'$. In $\triangle ABC$ the altitude CD is left of the triangle. In $\triangle A'B'C'$, the altitude at $C'D'$ is to the right.



A figure and its mirror image are congruent. Hence these two triangles have the same height as each other, the same base length as each other and the same area as each other. Thus a formula that gives the area of $\triangle ABC$ in terms of b and h must also be valid for $\triangle A'B'C'$.

This means that for this derivation, we do not need to separately consider two separate cases of "left altitude" and "right altitude". We can just assume the altitude falls on the left, provide a diagram of that case, derive the required formula, and it will automatically be valid for the mirror image case where the altitude falls on the right.

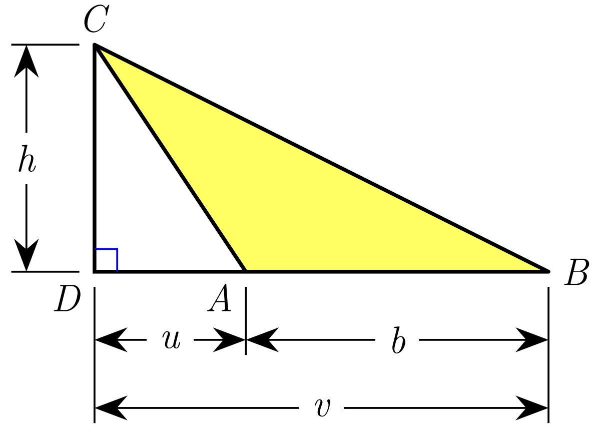
This idea doesn't just apply to areas of triangles. It is used throughout the study of Euclidean geometry. Whenever a valid Euclidean geometry proof uses a diagram, the proof continues to be valid if you replace the diagram by a congruent diagram.

- It will still be valid if you translate the diagram, drawing it in a different position of the page.
- It will still be valid if you rotate the diagram, drawing it in a different orientation on the page.
- It will still be valid if you replace the diagram by its mirror image, though ideally you don't replace any letters identifying vertices or lengths by their mirror images, since that would make it difficult to read.

Since this idea applies throughout Euclidean geometry, mathematicians often don't even mention it when they use it. To demonstrate how it could be explained relatively quickly, let's start the derivation from the beginning.

Say we know the values of h, the height of a triangle measured at an altitude falling outside the triangle, and b, the length of the corresponding base. Orient the triangle with the base horizontal at the bottom of the diagram.

When we do this, the altitude could fall left or right of the triangle. If it falls to the right, reflect the triangle in a vertical line, so that the altitude will fall to the left, as shown here.



Label the triangle's vertices ABC as shown here.

Extend the base BA to the point D so that $\angle CDA = 90°$. This means CD is the altitude on the base AB.

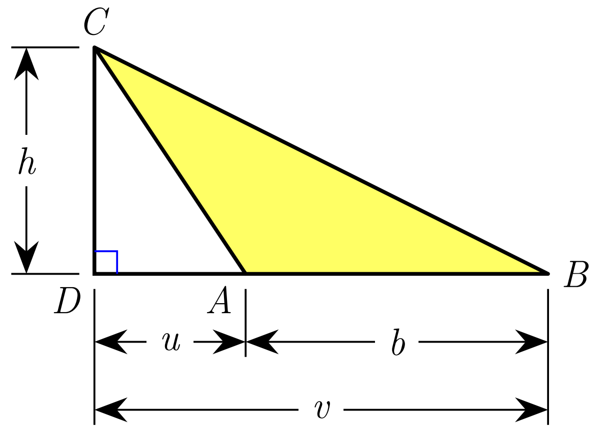
Reminder: In the terse language that mathematicians use to describe geometric constructions, the phrase "extend BA" implies a direction. It means to extend it beyond the 2nd of the two points listed, so we extend it beyond the point A, not B.

We seek a formula for the area of $\triangle ABC$ in terms of b, the length of base AB, and h, the height measured at the altitude CD.

Let $|AD| = u$ and $|BD| = v$.

Identify the two right triangles in the diagram and give formulae for their areas. Hence find the area of $\triangle ABC$.

$\angle BDC = 90°$. Hence $\triangle ACD$ and $\triangle BCD$ are right triangles. By theorem 9.1 their areas are half the product of their leg lengths, which gives $\frac{1}{2}uh$ and $\frac{1}{2}vh$ respectively.

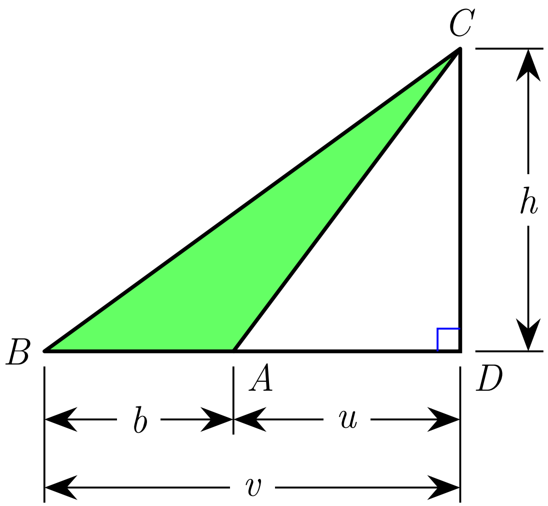


By the Area Sum Postulate, the area of $\triangle BCD$ equals the sum of the areas of $\triangle ACD$ and $\triangle ABC$. Rearranging gives:

$$\begin{aligned}
\text{Area of } \triangle ABC &= \text{Area of } \triangle BCD - \text{Area of } \triangle ACD \\
&= \tfrac{1}{2}vh - \tfrac{1}{2}uh \\
&= \tfrac{1}{2}(v-u)h \\
&= \tfrac{1}{2}bh
\end{aligned}$$

This is the same formula as found for the earlier triangles.

In the above derivation I explained why it was valid to only provide a diagram where the altitude falls left of the triangle. This isn't the only way to solve the problem.



If you prefer, you can include a second diagram where the altitude falls on the right, such as this green triangle. For variety, I've made the base length and height different to the yellow triangle, though I'm still denoting them by b and h

Label the triangles' vertices ABC as shown here. Note that in the yellow triangle above the vertices are labelled in anticlockwise order but in this green triangle they are labelled clockwise. Due to this trick, if you now reread the algebra given above for the yellow triangle, you can verify it still works for this green triangle.

Recall that "Extend the base BA" means extending it beyond the point A. In the yellow triangle that meant it was extended to the left. In the green triangle, because the points A and B are reversed, it means it is extended to the right.

9.05 Conclusion

We have had to adopt three different approaches depending on whether the height was measured at an altitude that was

- an edge of the triangle,
- inside the triangle, or
- outside the triangle.

The resulting formula was the same in all cases. The area is always half the product of the base length and height. Thus we can state a single theorem covering all triangles.

Theorem 9.2: The area of a triangle with base length b units and height h units is $\frac{1}{2}bh$ square units.

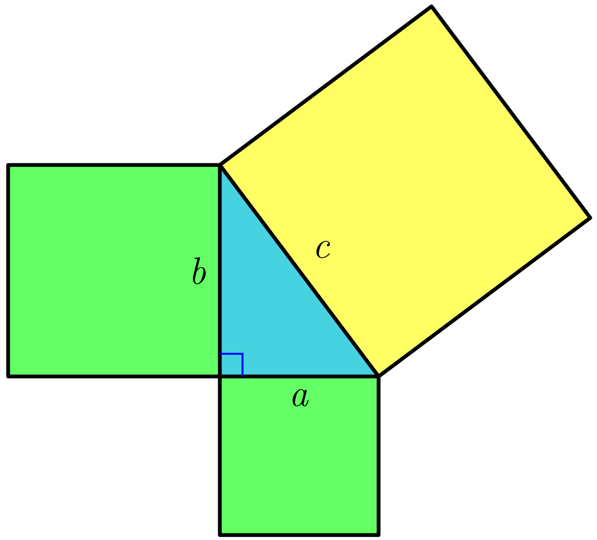
This theorem supersedes theorem 9.1, which only works for the first of the three cases, where the triangle is a right triangle and the two known lengths are the leg lengths.

9.06 Statement of Pythagoras' Theorem

Conflicting terminology alert: In North American this theorem is more commonly called the Pythagorean Theorem.

Theorem 9.3 — Pythagoras' Theorem: The area of the square on the hypotenuse of a right triangle is equal to the sum of the areas of two squares on the other two sides of the triangle.

To make sense of this theorem you need to understand that referring to the square on a side of a triangle means that we should construct a square which has side length equal to that side of the triangle. See the diagram, where squares have been constructed on the three sides of the blue right triangle.



Several popular shorter statements of the theorem exist, but they sacrifice some clarity for brevity. For example, it is common to omit the reference to area, giving the following statement.

The square on the hypotenuse of a right triangle is equal to the sum of the squares on the other two sides.

If we denote the length of the hypotenuse by c, and the lengths of the legs by a and b, then the theorem tells us $a^2 + b^2 = c^2$.

In the diagram shown right, the theorem states that the area of the yellow square is equal to the sum of the areas of the two green squares.

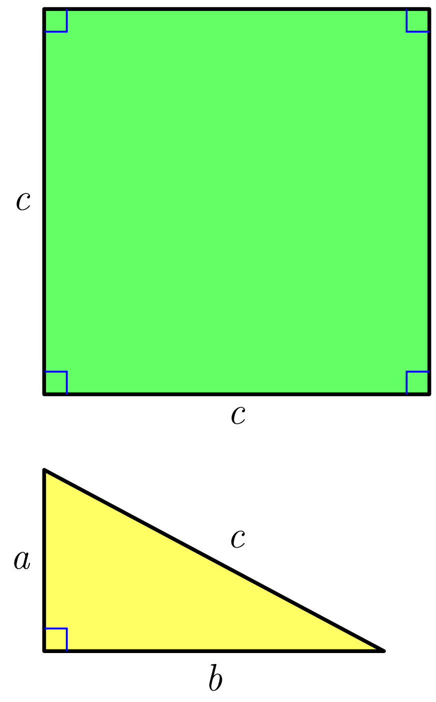
Pythagoras' theorem does not provide a practical method for calculating areas, but I included it in this book because we will need it to prove Heron's formula, which is a practical method for calculating the area of a triangle given the lengths of its three sides.

9.07 An area proof of Pythagoras' Theorem

There are several proofs of Pythagoras' theorem that employ area formulae. We will look at one such proof, since this does provide a practical demonstration of the usefulness of area formulae. There are also several ways to prove Pythagoras' theorem using similar triangles. We will look at one such proof in a later chapter on scaling and similarity.

The diagram shows a square of side length c and a right triangle with legs of length a and b and a hypotenuse of length c.

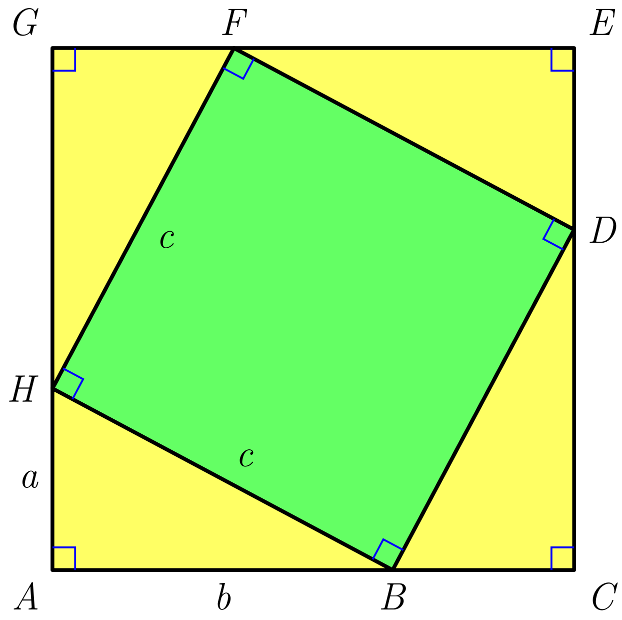
> Arrange one copy of the square and four copies of the right triangle to form a larger square.



The diagram here shows the green square and four copies of the yellow triangle forming alleged square $ACEG$. The next task is to prove this shape really is a square.

We know the hypotenuse of the right triangle is c, which matches the side length of the green square. Thus it is valid to join the hypotenuse of each right triangle to a side of the square, giving the shape $ABCDEFGH$.

Our task is to prove that this shape is a square, not an octagon. For example, we need to prove that ABC is a single line segment, rather than two separate line segments AB and BC with a bend occurring at B.



Prove that $\angle ABC = 180°$. You can use the fact that the three angles of a triangle sum to $180°$. Hence prove that $ABCDEFGH$ is a square that can more simply be labelled $ACEG$.

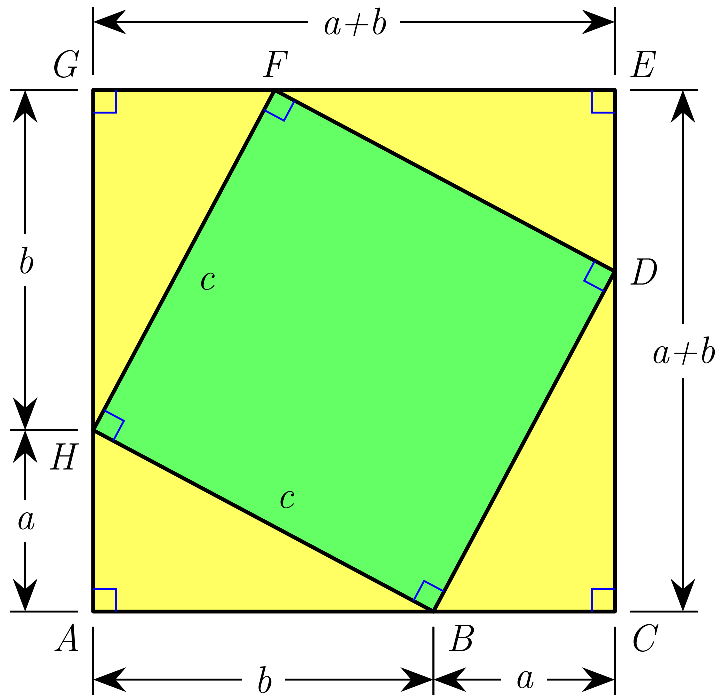
The four yellow right triangles are identical copies, and so are congruent.

$\angle DBC = \angle BHA$, because corresponding angles of congruent triangles are equal.

$\angle HBD = \angle HAB$, because both are right angles.

$$\begin{aligned}\angle ABC &= \angle ABH + \angle HBD + \angle DBC \\ &= \angle ABH + \angle HAB + \angle BHA \\ &= 180°, \text{ since the three angles of a triangle sum to } 180°\end{aligned}$$

That is, there is no bend at B, so ABC is a single line segment. The diagram has 90° rotational symmetry, so this argument can be repeated to show that CDE, EFG and GHA are also single line segments, so the resulting shape is a quadrilateral which we can simply refer to as $ACEG$.

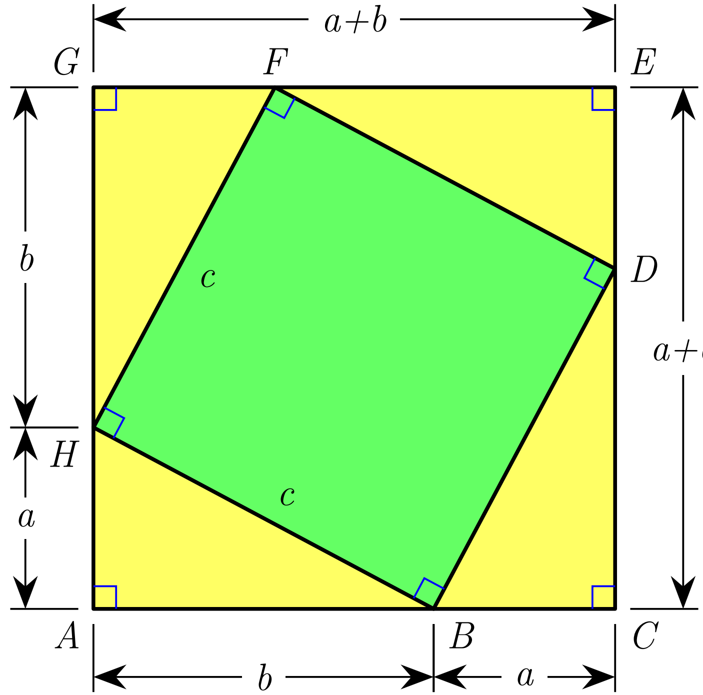


Since its four angles are the right angles from the four right triangles, $ACEG$ is a rectangle.

Since the four yellow triangles are congruent, their corresponding sides have equal length. Thus $|AB| = |CD| = |EF| = |GH| = a$ and $|BC| = |DE| = |FG| = |HA| = b$. Thus all sides of rectangle $ACEG$ have length $a + b$, so it is a square.

Use that side length to ascertain the area of square $ACEG$. What is the area of the green square $BDFH$? What is the area of each yellow triangle? Use the area sum postulate to write an equation that links these areas. Simplify your equation as far as possible. If you do this correctly, the result will prove Pythagoras' theorem.

The area of a square is simply the square of its side length. Hence square $ACEG$ has area $(a+b)^2$. Most readers will know the expansion of this expression from memory. For any readers still new to algebra, we can expand this square by using the distribute law of multiplication over addition.



$$\begin{aligned}(a+b)^2 &= (a+b)(a+b)\\ &= a(a+b)+b(a+b)\\ &= a^2+ab+ba+b^2\\ &= a^2+2ab+b^2\end{aligned}$$

The green square has area c^2. The area of a right triangle is half the product of the leg lengths. Hence each yellow rectangle has area $\frac{1}{2}ab$. The area sum postulate gives:

$$\begin{aligned}(a+b)^2 &= c^2+4\times\tfrac{1}{2}ab\\ a^2+2ab+b^2 &= c^2+2ab\\ a^2+b^2 &= c^2\end{aligned}$$

This proves Pythagoras' theorem.

9.08 Pythagoras of Samos

Pythagoras of Samos is thought to have lived from about 570BCE to about 495BCE. Most of the things commonly said about Pythagoras are wrong.

He is commonly said to have discovered the theorem that bears his name. He didn't. The theorem was already well-known to Mesopotanian mathematicians and particularly to surveyors over 1,000 years before Pythagoras was born. It is likely we will never know who first discovered it. Alas, much of the early history of mathematics is lost.

When Euclid wrote his famous book *Elements* around 300BCE, he identified his theorems by number, and called them "propostions". He labelled Pythagoras' theorem as "Proposition 47". The theorem is definitely important enough to deserve a more memorable name, but it remains a mystery who made the inappropriate decision to name it after Pythagoras. There is no evidence that Pythagoras ever claimed to have discovered the theorem or to have proved it true.

Euclid's great discovery, for which he definitely deserves fame, was the idea that all the important theorems of geometry known at the time could be proved by starting with a very small number of postulates. The purpose of *Elements* was to demonstrate a sequence of proofs that achieves this, so unfortunately it made no attempt to document who had originally discovered each proof. Historians believe Euclid was prolific when it came to discovering proofs of theorems and that he was the original discoverer of several of the proofs in *Elements*, but it also contains proofs that predate Euclid. *Elements* seems to be the oldest surviving document formally proving Proposition 47, but it is unclear if Euclid discovered his proof of this theorem 47 or copied it from an earlier source.

Pythagoras is also commonly said to be a great mathematician or a great geometer. This is also probably wrong.

Biographies of Pythagoras describe him as a philosopher, not a mathematician. Some historians believe he invented the term "philosopher". However the philosophy of the time was a strange amalgam of mystical religious beliefs which would have little if any overlap with modern philosophy. For example, Pythagoras is believed to be have been a practitioner of the philosophies of divination and prophecy, though as a vegetarian with concern for animal welfare, it seems he was not involved in the practice of prophesying the future by examining the entrails of sacrificed animals.

Many accounts describe Pythagoras as a charismatic speaker, who was politically influential. This likely explains why his documented history is frequently contradictory and sometimes clearly mythical. His supporters were prone to exaggerate his achievements, just as his detractors were prone to exaggerate his faults. The ridiculous claims made about him include that he was the son of the god Apollo, that he possessed a golden arrow that allowed him to fly, and that he could converse with non-human animals.

One of the most important contributions made be Pythagoras, which *should* be well-known, is not. He founded a school which freely admitted females at a time when most societies denied females access to education.

Pythagoras' students called themselves "Pythagoreans". While Pythagoras himself was not a great mathematician or geometer, it seems his school did some teaching in these areas. I mentioned earlier that in North America Pythagoras' theorem is more commonly known as the Pythagorean theorem. This was perhaps based on a belief that, while Pythagoras himself made no contribution to discovering or proving the theorem, perhaps some Pythagoreans did. While there is no strong evidence to support this claim, it cannot be absolutely refuted. While the theorem was proved before Pythagoras was born, new proofs were still being discovered as late as the 19th century.

Incidentally, one of the many conflicting calculation schemes within the occult pseudoscience of numerology is called the Pythagorean method. While Pythagoras' beliefs did incorporate many ideas that are now regarded as nonsense, there is no firm evidence that Pythagoras or any of the Pythagoreans had any involvement in the particular nonsense known as numerology. Sometimes those who promulgate pseudo-scientific beliefs name their theories after deceased famous people in an attempt to add unwarranted credibility to their incredible claims.

9.09 Heron's Theorem

Theorem 9.4 — Heron's Theorem: The area of a triangle with sides of length a, b and c, and with semiperimeter s, is $\sqrt{s(s-a)(s-b)s-c}$.

The semiperimeter is simply half the perimeter, so $s = \frac{1}{2}(a+b+c)$, or if you prefer, $2s = a+b+c$. Semiperimeter is a term that usually only arises for triangles. In this book, we are only ever going to see it in Heron's Theorem, but it does also arise in other more complex theorems about triangles.

There are some short straightforward proofs of Heron's Theorem that employ well-known trigonometry theorems. Unfortunately, this book is intended to make sense to students who have not yet encountered trigonometry, which means we are going to have to use a lengthy proof containing some difficult algebra. Let's be honest here. This proof is hard and ugly. Do *not* make any attempt to try to memorise it. Just follow along, verifying that each step in the algebra is correct.

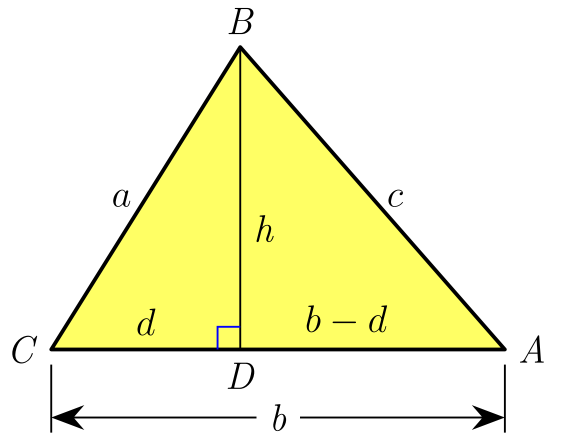
In the derivation of Pythagoras' Theorem, I used the distributive property of multiplication over addition to derive the well-known result $(a+b)^2 = a^2+2ab+b^2$. If you are not already familiar with the results, use a similar method to evaluate $(a-b)^2$ and $(a-b)(a+b)$.

$$\begin{aligned}(a-b)^2 &= (a-b)(a-b) \\ &= a(a-b) - b(a-b) \\ &= a^2 - ab - ba + b^2 \\ &= a^2 - 2ab + b^2\end{aligned}$$

$$\begin{aligned}&(a-b)(a+b) \\ &= a(a+b) - b(a+b) \\ &= a^2 + ab - ba - b^2 \\ &= a^2 - b^2\end{aligned}$$

The result $a^2 - b^2 = (a-b)(a+b)$ is commonly known as "the difference of two squares".

Consider $\triangle ABC$ with internal altitude BD. Every triangle has at least one altitude that falls inside the triangle, so it is always possible to label the vertices in a way that ensures BD is an internal altitude.



In my diagram, $\triangle ABC$ seems to be an acute triangle, but you can also produce diagrams where $\angle ABC$ is a right angle or an obtuse angle, and verify that the proof remains valid for those scenarios.

Let the sides of $\triangle ABC$ opposite vertices A, B and C have lengths a, b and c respectively. Let $|CD| = d$, so $|AD| = b - d$.

Note that this triangle has area $\frac{1}{2}bh$. Since the formula in Heron's Theorem does not contain h, it is useful to obtain an expression for h that depends only on a, b and c.

Applying Pythagoras' Theorem to $\triangle BCD$ gives $a^2 = d^2 + h^2$.

Applying Pythagoras' Theorem to $\triangle ABD$ gives $c^2 = h^2 + (b-d)^2 = h^2 + b^2 - 2bd + d^2$.

Taking the difference of the previous two results gives $c^2 - a^2 = (h^2 + b^2 - 2bd + d^2) - (d^2 + h^2) = b^2 - 2bd$

Rearranging gives: $2bd = a^2 + b^2 - c^2$, and so $d = \dfrac{a^2 + b^2 - c^2}{2b}$.

Rearranging $a^2 = d^2 + h^2$ gives $h^2 = a^2 - d^2 = (a-d)(a+d)$.

Substituting the previous expression for d into this result leads us into the following lengthy algebra.

$$
\begin{aligned}
h^2 &= \left(a - \frac{a^2+b^2-c^2}{2b}\right)\left(a + \frac{a^2+b^2-c^2}{2b}\right) \\
(2b)^2h^2 &= \{(2b)a - (a^2+b^2-c^2)\}\{(2b)a + (a^2+b^2-c^2)\} \\
4b^2h^2 &= \{(c^2 - (a^2 - 2ab + b^2)\}\{(a^2 + 2ab + b^2) - c^2\} \\
&= \{(c^2 - (a-b)^2\}\{(a+b)^2 - c^2\} \\
&= \{c - (a-b)\}\{c + (a-b)\}\{(a+b) - c\}\{(a+b) + c\} \text{ using the difference of two squares result, twice.} \\
&= (b+c-a)(a+c-b)(a+b-c)(a+b+c)\} \\
&= (2s-2a)(2s-2b)(2s-2c)2s \text{ since } 2s = a+b+c \\
&= 16s(s-a)(s-b)(s-c) \\
(\tfrac{1}{2}bh)^2 &= s(s-a)(s-b)(s-c) \\
\tfrac{1}{2}bh &= \sqrt{s(s-a)(s-b)(s-c)}
\end{aligned}
$$

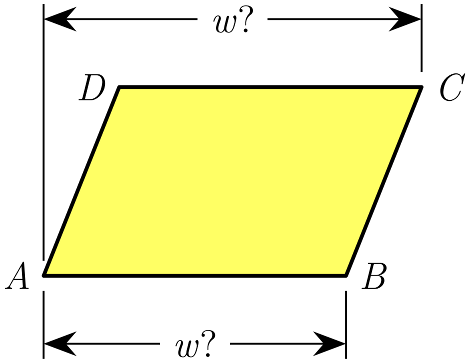
As stated earlier, the area of the triangle is $\frac{1}{2}bh$, so the conclusion above proves the theorem.

10 Parallelogram

We already have a formula for the area of a rectangle, so in this chapter we only need to find a formula for the area of a parallelogram that is not also a rectangle.

10.01 Jargon: base and height

Opposite sides of a parallelogram are parallel. For convenience, orient the parallelogram so that one pair of opposite sides run horizontally across the page.

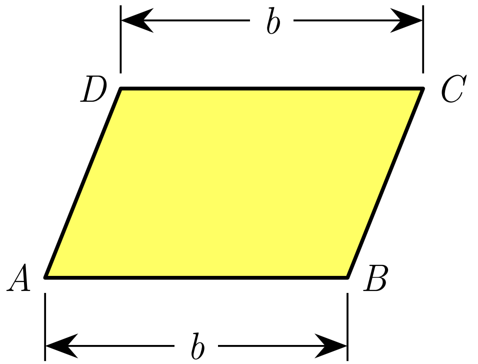


In Chapter 6, I used the word "width" for one of the rectangle's dimensions. For parallelograms, there's a problem with this word. If you draw a parallelogram and ask people to identify its width, people don't agree what it should be. Some choose the distance $|AB|$, as shown by the lower dimension here, but others disagree and say it should be the horizontal distance from the leftmost vertex A to the rightmost vertex C, as shown by the top dimension line.

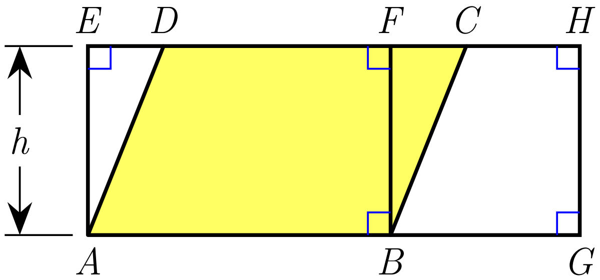
If we slide the line segment CD to the left until the four angles of the parallelogram become right angles the two interpretations become the same. That is, the word "width" isn't ambiguous for rectangles.

To avoid confusion, it's safer to avoid using the word "width" for the parallelogram, and also for the trapezium which we will see in the next chapter.

We will use the word "base" to mean the lower horizontal side of the parallelogram. In this diagram it is the line segment AB. Let the length of the base be b units. Opposite sides of a parallelogram have equal length, so $b = |AB| = |CD|$.



Define the parallelogram's height to be the distance between the parallel lines AB and DC. The distance between a pair of parallel lines is constant, so we can measure that distance wherever we like. We simply draw a line segment that joins line AB to line DC while being perpendicular to both. The height is the length of that line segment.



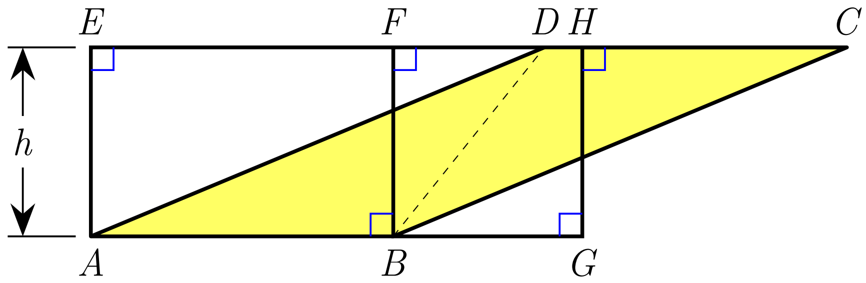
The line segment could start at a vertex of the parallelogram, such as the line segments AE and BF. But it could also be drawn at some completely arbitrary position. It could even fall entirely outside the parallelogram, such as the line segment GH.

We will denote this height by h, so in this diagram $h = |AE| = |BF| = |GH|$.

I defined the height of the parallelogram to be the distance between the parallel lines AB and DC. Why didn't I define it more simply as the distance between the sides AB and DC? Won't that give the same number?

The simpler wording gives the same number for the parallelogram shown above, but that isn't true for all parallelograms.

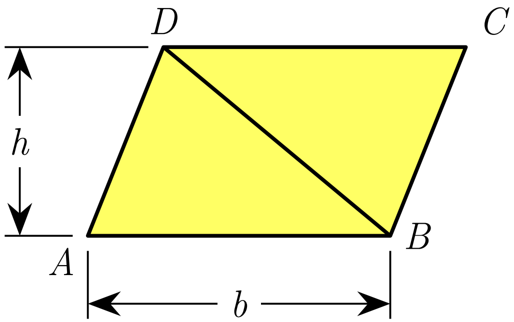
Move the side DC further to the right, until the point D is further right than the point F, as shown here. All points other than D and C were left undisturbed, so the distance between the parallel lines AB and DC has not changed. It is still $h = |AE| = |BF| = |GH|$.



The sides of a polygon are line segments, not lines. In this new parallelogram, the distance between the sides AB and DC, is the length of the shortest line segment joining line segments AB and DC. That is, it is the length of the dashed line segment BD. This exceeds h, so the distance between the sides AB and DC is not the same as the distance between the parallel lines AB and DC.

In the next section we will show that we can easily calculate the area of this parallelogram when given b and h, so it useful to define the height of this parallelogram to be the distance between the parallel lines AB and DC. The distance $|BD|$ is not useful for calculating the area.

10.02 Calculating the area

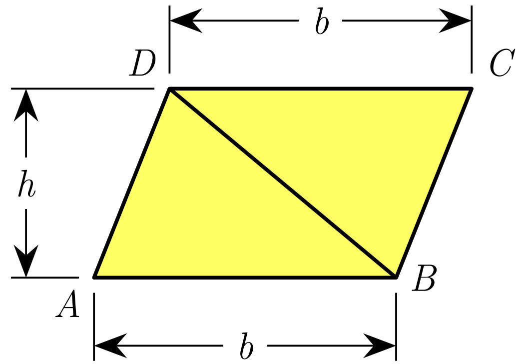


The formula for the area of a parallelogram is easily derived by drawing one diagonal, subdividing it into two triangles. Either diagonal can be used. For no particular reason, I have chosen to use BD.

> Use the formula for the area of a triangle to derive a formula for the area of the parallelogram.

Triangle ABD has base AB of length b. Its height is the length of the altitude at D drawn perpendicular to the base AB. This is also the distance between the parallel lines AB and DC, which is h, the height of parallelogram $ABCD$. By theorem 8.2 triangle ABD has area $\frac{1}{2}bh$.

Since the opposite sides of a parallelogram have equal length, side CD also had length b. So if we want the area of triangle BCD, it makes sense to regard BC as the base, since we know that side's length. We might regard triangle BCD as upside down, compared to the orientation in which we most often draw triangles. If we regard BC as its base, then the height will be the length of the altitude drawn at B, perpendicular to CD. This is again the distance between the parallel lines AB and DC, which is still h. Thus by theorem 8.2, triangle BCD also has area $\frac{1}{2}bh$.

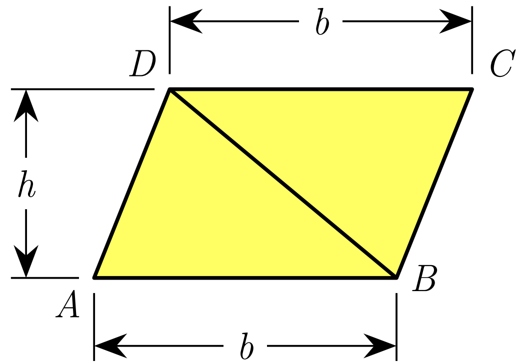


$$\begin{aligned}\text{Area of parallelogram } ABCD &= \text{Area of } \triangle ABD + \text{Area of } \triangle BCD \\ &= \tfrac{1}{2}bh + \tfrac{1}{2}bh \\ &= bh\end{aligned}$$

Are the two triangles congruent? If so, is that useful?

Line segment BD is common to both triangles. Opposite sides of a parallelogram have equal length, so $|AB| = |CD|$ and $|AD| = |BC|$. Hence by the SSS congruency test — three corresponding equal sides — $\triangle ABD$ and $\triangle CDB$ are congruent.

This provides an alternative method for deriving the area formula. After explaining why the area of $\triangle ABD$ is $\frac{1}{2}bh$, we can deduce $\triangle CDB$ has the same area since it is congruent.



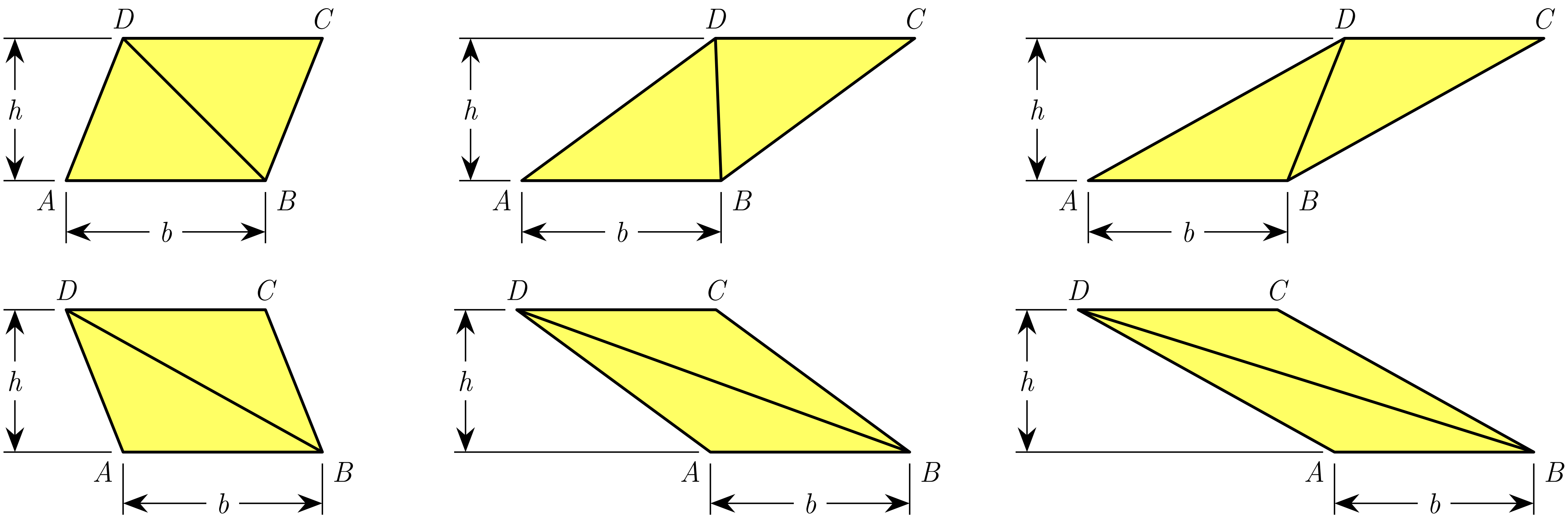
Before stating our result as a theorem, we should check if our diagram for this proof covers all possible scenarios. For example, does it work for the parallelogram shown earlier that has the top edge shifted further right relative to the bottom edge?

Also, this parallelogram leans to the right, meaning the upper edge CD is further right on the page than the lower edge AB. If a parallelogram leans to the left we could reflect it in a vertical line to produce a congruent parallelogram that leans to the right, with the same base length, height and area. This means we don't need to separately consider parallelograms that lean left. Any area formula that works for a right leaning parallelogram will also work for its left-learning reflection. Having said that, it might still be interesting to look at some left-leaning parallelograms just to see what happens.

Investigate other ways to draw a parallelogram with its base still horizontal. Can you find a diagram where the derivation given above will fail?

Here are six cases that might be worth considering. The first three diagrams have the top edge offset to the right. In the first diagram the horizontal position of D is still left of B while in the third it is right of B. It wasn't immediately clear to me whether something unusual might happen at the crossover between these two scenarios, so I also included the second diagram where D is directly above B.

The next three diagrams show the comparable three cases where the top edge is offset to the left instead of the right. Recall that we already derived a formula for the area of a rectangle in chapter 6, so at the start of this chapter I explicitly stated that this chapter would be looking at parallelograms that are not also rectangles. Hence I didn't include a 7th diagram of a parallologram that doesn't lean.



If you didn't find all six cases, check those that are new to you. Does the derivation given above work in all six cases?

The derivation works in all six cases. We can identify some features that change between these six diagrams, but none of them affect the derivation. For example

- In the first diagram the parallelogram is subdivided into two acute triangles, in the second diagram they are two right triangles and in the other four diagrams they are two obtuse triangles. However, in all six cases the two triangles both have base length b and height h, so they always have area $\frac{1}{2}bh$. If you preferred the alternative approach that identified congruent triangles, the SSS congruency test still works for the two triangles in each diagram.
- The parallelogram has two diagonals, AC and BD. I chose to use diagonal BD to split it into two triangles. BD is the shorter diagonal in the first three diagrams and the longer diagram in the last three diagrams. This has no effect on the derivation.

Since our derivation works in all cases, it's safe to state its result as a theorem.

Theorem 10.1: The area of a non-rectangular parallelogram with base length b units and height h units is bh square units.

That inclusion of "non-rectangular" is clumsy. We should look more closely at the rectangle special case. We found a formula for the area of a rectangle in Chapter 6. If the formula in Theorem 10.1 matches that earlier formula when the parallelogram happens to be a rectangle, then we don't have to exclude rectangles in Theorem 10.1.

10.03 Special case: rectangle

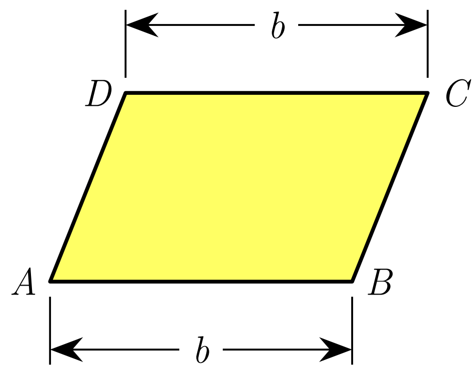
The rectangle is a special case of the parallelogram. If the angles of a parallelogram are all right angles, it is a rectangle. Here are the relevant theorems.

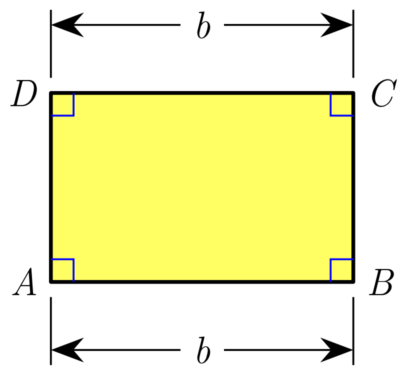
Theorem 6.6: The area of a rectangle with width w units and height h units is wh square units.

Theorem 10.1: The area of a non-rectangular parallelogram with base length b units and height h units is bh square units.

The formulae use different pronumerals, so it's tempting to say that can't be consistent, but we should look deeper. We should check whether the b in the parallelogram formula becomes equivalent to the w in the rectangle formula when the parallelogram is a rectangle. We should also check if the h pronumeral is used consistently. It *is* possible for a pronumeral to be given different meanings in different formulae.

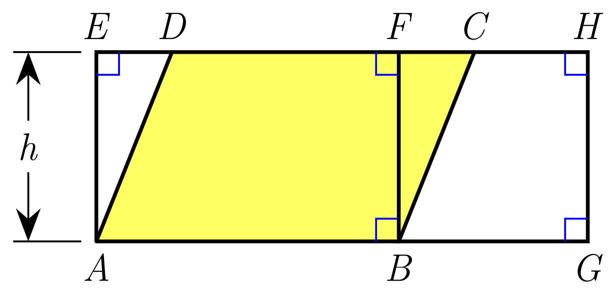
The pronumeral b denotes the length of the base of parallelogram $ABCD$. That is, it is defined as the length of line segment AB. The opposite sides of a parallelogram have equal length, so $b = |AB| = |CD|$

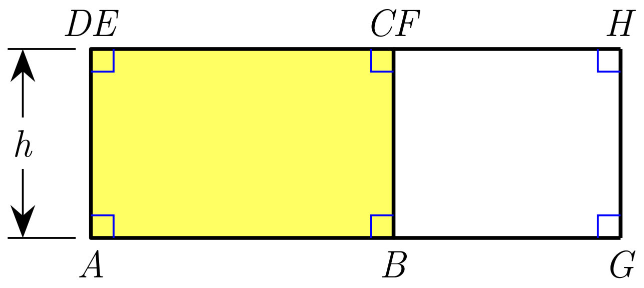




In the special case where the four angles of the parallelogram are right angles, making it a rectangle, this still gives $b = |AB| = |CD|$. This is consistent with the pronumeral w which we used in chapter 6 to denote the width of a rectangle, which could be measured as the length of its top or bottom side.

The pronumeral h denotes the height of the parallelogram $ABCD$. It is defined as the distance between lines (NOT sides or line segments) AB and CD. We can measure the distance at any convenient location. For example, $h = |AE| = |BF| = |GH|$. In general, this height is NOT equal to the length of sides AD and BC.





Here is the corresponding special case where the four angles of the parallelogram are right angles, making it a rectangle. As in the previous diagram $h = |AE| = |BF| = |GH|$. But the points D and E now coincide so $|AE| = |AD|$. Also points C and F coincide, so $|BF| = |BC|$. Thus in this special case $h = |AD| = |BC|$. That is, h denotes the length of either vertical side of the rectangle. This is consistent with how we defined the height of the rectangle in Chapter 6.

Thus the formula for the area of the parallelogram will give the correct result in the special case where the parallelogram is a rectangle. Hence we can simplify theorem 10.1 to include the rectangular special case. Here is again is theorem 10.1 and the resulting simpler replacement.

Theorem 10.1: The area of a non-rectangular parallelogram with base length b units and height h units is bh square units.

Theorem 10.2: The area of a parallelogram with base length b units and height h units is bh square units.

If someone asks you to calculate the area of a rectangle, I am not suggesting that you should explain that all rectangles are parallelograms and use Theorem 10.2. That would be needlessly complex. The formula for the area of a rectangle is very well-known, so you should use it when asked to find the area of a rectangle.

That is, this section is NOT included to give you a more complex way to calculate the areas of rectangles. Rather, its purpose is to justify simplifying the clumsy phrase “non-rectangular parallelogram” in Theorem 10.1, to the much shorter “parallelogram” that appears in Theorem 10.2.

Another situation where this special case could become important is if you are writing a computer program that includes a calculation of the area of a parallelogram given the base length and height. Perhaps the program has to perform this calculation for many different parallelograms, so your program contains a function whose sole purpose is to calculate the area given the two arguments base length and height. In most cases, the function simply multiplies together the two arguments it is given and return their product as the result.

But programmers also need to consider whether there are any special cases where the function might need to do something different. For example, the base length and height of a parallelogram have to be greater than zero, so you might decide that your function should test that both arguments meet that requirement. If either argument fails that test it is likely that there is an error somewhere in the code, so your function should probably issue a warning message to alert you that something has gone wrong, rather than just returning the product of the two arguments.

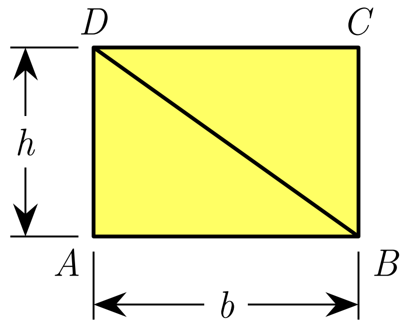
The programmer writing the function might also wonder whether they need to do something special if the parallelogram happens to be a rectangle. Theorem 10.2 says that they don't, because the parallelogram formula will give the correct answer in that special case.

10.04 An invalid approach

Section 10.02 derived a formula for the area of a non-rectangular parallelogram. Sometimes students wonder whether the method used in that section can be extended to cover rectangles. Recall that in chapter 6, to develop the formula for the area of a rectangle, we had to first consider integer dimensions, then rational dimensions, then irrational dimensions, and for each of those cases we had to separately consider both dimensions. That took a lot of work! The method used in section 10.02 is considerably shorter than the arguments presented in chapter 6, so it looks like that approach could get to the answer much faster.

Alas, this approach contains a logical error. I'll now present the approach in more detail. Try to spot where the error happens.

Section 10.02 gave six different diagrams showing plausible shapes for a parallelogram, three sloping left and three sloping right. Here again is the special case from section 10.03 where the parallelogram doesn't slope either way but instead stands upright, producing a rectangle. In this case, the distance between lines AB and CD, denoted by h, is also the height of the rectangle, which is also the length of either vertical edge. I'll continue to denote the base length by b, but remember that for this special case it is equal to the width w used in chapter 6.



$ABCD$ is a rectangle, so all its angles are right angles. This means the two triangles mentioned above are both right triangles, so by Theorem 9.1 their area is half the product of their leg lengths.

Now, simply repeat the relevant three lines of algebra from section 10.02, replacing the word "parallelogram" by "rectangle".

$$\begin{aligned}\text{Area of rectangle } ABCD &= \text{Area of } \triangle ABD + \text{Area of } \triangle BCD \\ &= \tfrac{1}{2}bh + \tfrac{1}{2}bh \\ &= bh\end{aligned}$$

Since $b = w$, this matches the chapter 6 result of wh. This proves the area of a rectangle is the product of the lengths of two adjacent sides.

Explain the logic error inherent in this alleged proof.

Here are the two relevant theorems.

Theorem 6.6: The area of a rectangle with width w units and height h units is wh square units.

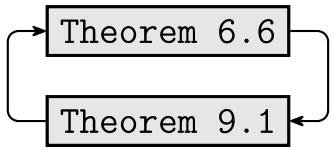
Theorem 9.1: The area of a right triangle is ½ the product of the leg lengths.

Over chapters 5 and 6 we derived increasingly more general theorems to eventually prove theorem 6.6. The invalid approach given above alleges this was unnecessary because theorem 6.6 is easily proved by calculating the area of a rectangle as the sum of the areas of two right triangles. Using theorem 9.1, both those triangles have area $\frac{1}{2}bh$. That is, this alleged proof of theorem 6.6 relies on theorem 9.1.

However, the derivation of theorem 9.1 in chapter 9 involved joining two right triangles to form a rectangle, with the area of that rectangle being found by theorem 6.6. That is, the proof of theorem 9.1 relies on theorem 6.6.

This is called circular reasoning or a circular argument. It is not a valid method of argument.

The diagram shown here illustrates the circular nature of the argument. From Chapter 9 we know that if theorem 6.6 is true, then theorem 9.1 is true. The invalid argument above shows that if theorem 9.1 is true then theorem 6.6 is true. This means that it's possible that both theorems are true, but it's also possible that both are false. Thus the invalid argument fails to prove either theorem true.

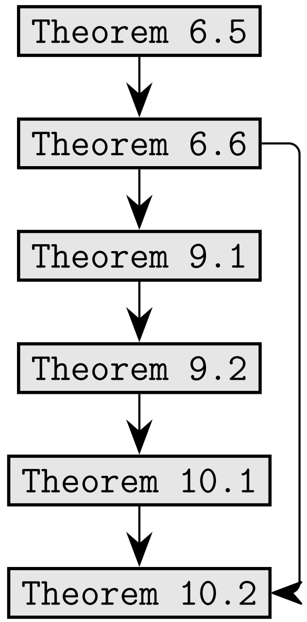


This particular invalid circular argument only cycles through two theorem. Longer loops are possible. If we use theorem 1 to prove theorem 2 true, then use theorem 2 to prove theorem 3 true, and finally use theorem 3 to prove theorem 1 true, then we have an invalid circular argument that cycles through 3 theorems.

It can be easy to fall into the trap of believing circular arguments because they do produce plausible results. For example, in the circular argument given above a rectangle of know area is split into two congruent right triangles to find the area of one such triangle, and then two copies of that triangle are joined together to recreate the original rectangle. If we started that process with a valid expression for the area of the rectangle, we will end up with the same valid expression. That doesn't make the circular argument valid logic. It just means we didn't make any other errors while working through the invalid circular argument!

The key to avoiding circular arguments is to keep track of the order in which theorems are derived. The proof of a theorem should only use theorems derived earlier in the sequence. Mathematicians like to number their theorems in the order derived, since this gives an easy check. In the invalid proof given above, if you notice that the alleged proof of theorem 6.6 is relying on a theorem 9.1 from 3 chapters *later*, this immediately flags that there is a risk of a circular argument being present.

By contrast, if we draw a "dependency chart" showing the order we used to derive the theorems being discussed, there are no circularities present.



Theorem 6.6 which gives the area of a rectangle was derived in chapter 6 from Theorem 6.5, not from a later theorem as happened in the circular argument.

Chapter 9 used Theorem 6.6 to derive Theorem 9.1 giving the area of right triangle given the leg lengths. Then Theorem 9.1 was used to derive the formula for the area of any triangle given a base length and the height of the corresponding altitude, resulting in Theorem 9.2.

In this chapter we used Theorem 9.2 to derive Theorem 10.1, which gives the area of all parallelograms that are NOT rectangles. That "NOT" is important, since it avoids introducing a circular argument.

We then showed that if we apply the formula given in Theorem 10.1 to a rectangle, it is consistent with the formula for rectangle that we derived in Theorem 6.6. So we combined Theorem 6.6, which deals with rectangles, with Theorem 10.1, which deals with non-rectangular parallelograms, to produce Theorem 10.2, which covers all parallelograms.

Incidentally, if you are building a dependency chart to demonstrate that your development of a set of theorems contains no circular arguments, there is no need to include postulates on the chart. We never attempt to prove postulates, so that cannot form part of a circular argument.

11 Trapezium

11.01 Inconsistent terminology alert

A trapezium (plural: trapezia) is a quadrilateral with a pair of opposite sites which are parallel. There is however disagreement between authors as to whether a trapezium is:

- a quadrilateral with *at least* one pair of parallel sides. This means that a parallelogram is a special case of the trapezium.
- a quadrilateral with *exactly* one pair of parallel sides. This means trapezia and parallelograms are mutually exclusive sets.

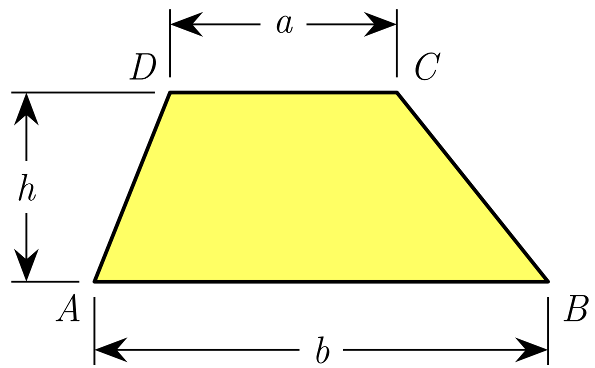
In U.S.A. the trapezium is known as a trapezoid. Curiously, this usage can be traced to an error in a single early mathematics dictionary published in U.S.A. that was subsequently copied into most mathematics textbooks published in U.S.A.

11.02 Area of trapezium

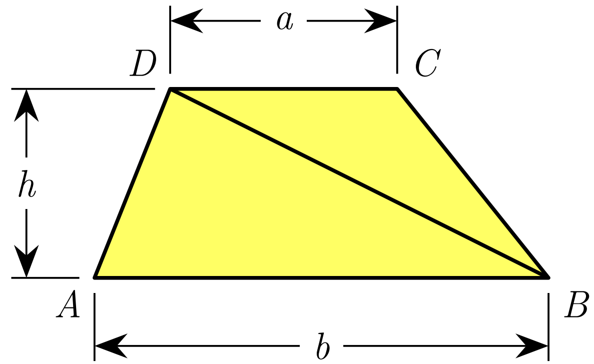
We already know how to calculate the area of a parallelogram, so consider a trapezium that is not a parallelogram. That is, it will have exactly one pair of parallel sides. Orient the trapezium so that its two parallel sides are horizontal with the shorter of these two sides at the top of the figure.

Label the vertices as shown. Let the lengths of the parallel sides be a and b, with $a < b$. As for the parallelogram, let the height of the trapezium be the distance between lines AB and DC, the lines containing the two parallel sides. Let h denote this height.

Using a similar method to that used to find the area of a non-rectangular parallelogram, derive a formula for the area of the trapezium.



Draw diagonal BD. I chose BD to be consistent with what I did with parallelograms in the previous chapter, but either diagonal will work.



Triangle ABD has base AB of length b and its height is the length of the altitude at D drawn perpendicular to the base AB. This happens to be the distance between the parallel lines AB and DC, which is h, the height of the trapezium. Thus by Theorem 9.2, $\triangle ABD$ has area $\frac{1}{2}bh$.

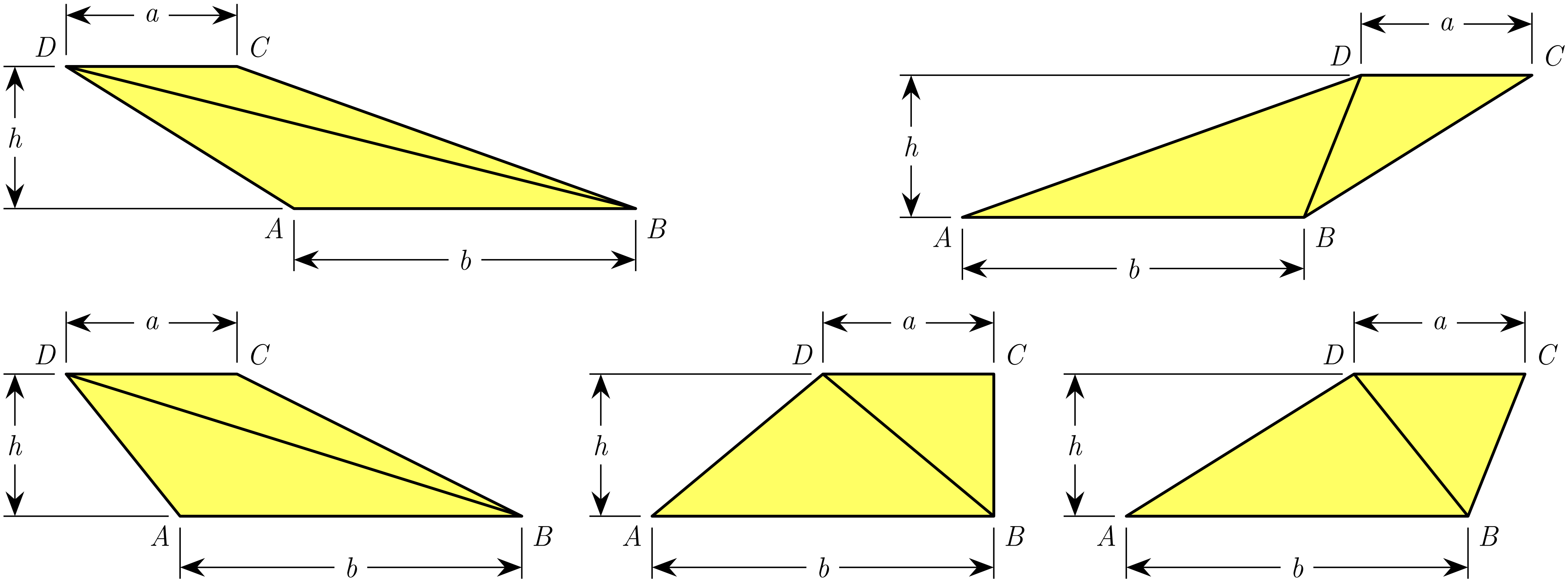
In $\triangle BCD$, we know the length of side CD, so regard that side as the base. The base length is a. Using BC as the base, the relevant height is the length of the altitude drawn at B, perpendicular to CD. This is again the distance between the parallel lines AB and DC, which is h. Thus by theorem 8.2, $\triangle BCD$ has area $\frac{1}{2}ah$.

$$\begin{aligned}\text{Area of trapezium } ABCD &= \text{Area of } \triangle ABD + \text{Area of } \triangle BCD \\ &= \tfrac{1}{2}bh + \tfrac{1}{2}ah \\ &= \tfrac{1}{2}(a+b)h\end{aligned}$$

We should consider whether there are any other shapes for the trapezium that might cause this derivation to fail. Remember that we oriented the trapezium with the parallel edges horizontal with the shorter edge at the top. We haven't yet thought about the possible positions of side DC relative to side AB.

Regard the values a, b and h as being fixed. Also regard the position of side AB as fixed. Think about sliding side DC to the left and right, always keeping it parallel to side AB. Can you find any trapezium for which the method given above fails? It may not be necessary to draw every possible shape. Draw enough variations so that you can get a feel as to whether any of the changes will affect the proof.

Here are some of the possible shapes for a trapezium. You can verify that in all these scenarios, the two triangles continue to have bases of length a and b and both have height h, so the proof is always valid.

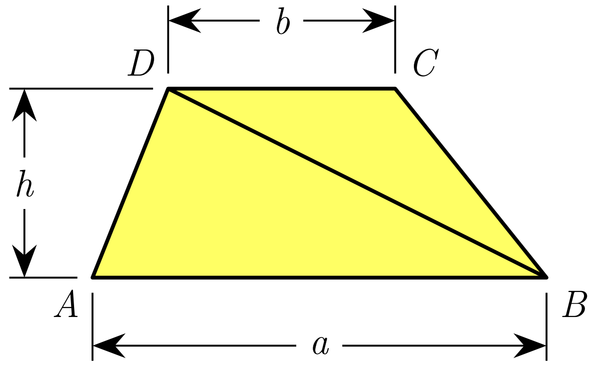


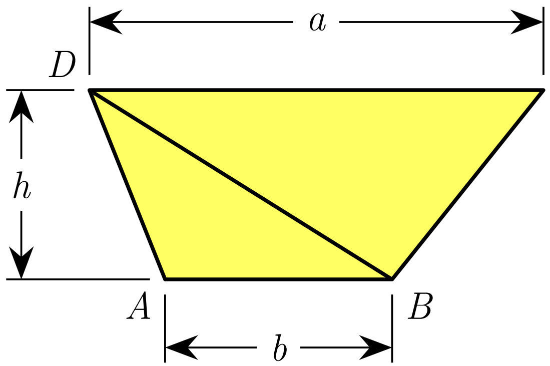
In our derivation we oriented the trapezium with the shorter of the two parallel sides at the top, denoting its length by a.

When we state the theorem for the area of a trapezium, should we include the constraint $a < b$? If someone labels the two parallel sides the other way around, with $b < a$, does the formula given above still work?

There are two ways to solve this challenge.

Say we continue to orient the trapezium with the shorter parallel edge at the top. Compared to the diagrams given above, the locations of the pronumerals a and b are reversed. If we run the through the derivation given above with this new diagram, it plays out the same way, but with every occurrence of a and b swapped. This means that rather than getting an answer of $\frac{1}{2}(a+b)h$ we would get $\frac{1}{2}(b+a)h$. But these two expressions are equal. This means the formula also provides the correct answer if $b < a$.





The other approach is to continue to orient the diagram with the edge of length a at the top. This means the top side is now longer than the short side. If you reread the derivation given earlier, you will find it continues to be valid for this new inverted diagram. You can also produce more diagrams similar to this new diagram but with the side DC moved further left or right, and the derivation continues to work.

So, our formula for the area of a trapezium works whether $a < b$ or $b < a$. Hence when we state our result as a theorem, we don't need to include a constraint that $a < b$.

I mentioned earlier that there are conflicting definitions of a trapezium. Some say that a trapezium is a quadrilateral with *exactly* one pair of parallel sides. At this point, these people would be happy to state the following theorem.

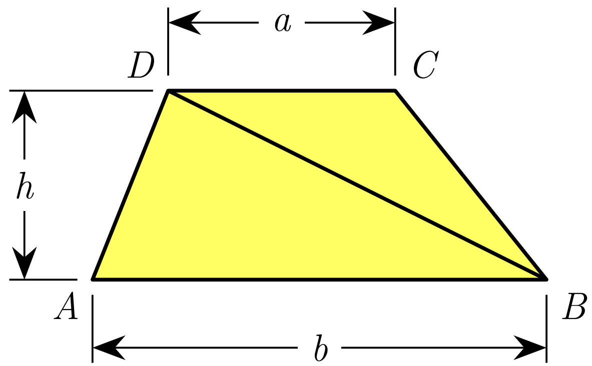
Theorem 11.1: The area of a trapezium with height h units and parallel sides of length a units and b units, where $a \neq b$, is $\frac{1}{2}(a+b)h$ square units.

But others say that a trapezium is a quadrilateral with *at least* one pair of parallel sides. These people may want to know whether the trapezium formula also holds for trapezia with two pairs of parallel sides, meaning they are parallelograms. We should check whether the formula derived above does work for that special case.

11.03 Special case: parallelogram

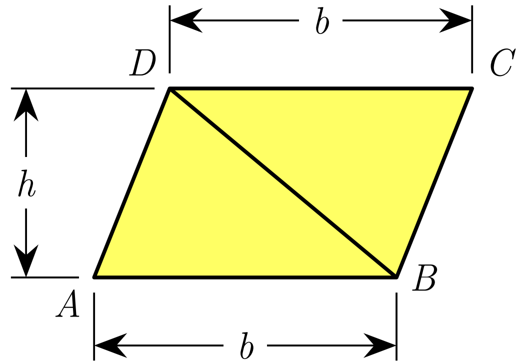
Investigate whether the trapezium formula is valid when applied to a parallelogram.

We derived the formula for the area of a trapezium by dividing it into two triangles and summing the areas of those triangles. This is the same method we used to find the area of the parallelogram, the only difference being that for the parallelogram the two triangles happened to have equal area. So it seems reasonable to expect the trapezium formula will still work for the parallelogram special case. But let's check the algebra and see what happens.



The trapezium split into two triangles with areas $\frac{1}{2}ah$ and $\frac{1}{2}bh$. In the special case where the trapezium becomes a parallelogram, the top and bottom sides have equal length, so $a = b$, so both triangles have area $\frac{1}{2}bh$ giving a total area of bh. Alternatively, when $a = b$ the trapezium area formula simplifies as follows.

$$\frac{1}{2}(a+b)h = \frac{1}{2}(b+b)h = bh$$



This matches the earlier result for the parallelogram, meaning the trapezium formula can correctly find the area of a parallelogram, so all is well. In fact, given that the parallelogram formula can correctly find the area of a rectangle, we can even say that the trapezium formula will correctly find the area of a rectangle!

Theorem 11.1 included the constraint $a \neq b$. We can omit that constraint, because the formula will still give the correct answer if applied to a parallelogram. This gives the following simpler version of the theorem.

Theorem 11.2: The area of a trapezium with height h units and parallel sides of length a units and b units, is $\frac{1}{2}(a+b)h$ square units.

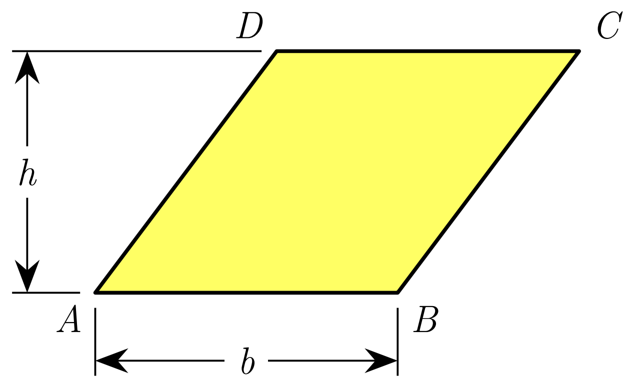
But I'm not suggesting here that you should habitually use the trapezium formula for those simpler special cases. If you want the area of a parallelogram, use the simpler parallelogram formula and if you want the area of a rectangle, use the simpler rectangle formula!

The comments in the previous chapter concerning computer programming are also relevant here. Say you are programming a function to calculate the area of a trapezium given the three values a, b and h. Your function does not need to test whether $a = b$ and do something different in that special case, because the formula $\frac{1}{2}(a + b)h$ does still give the correct area when $a = b$.

12 Rhombus

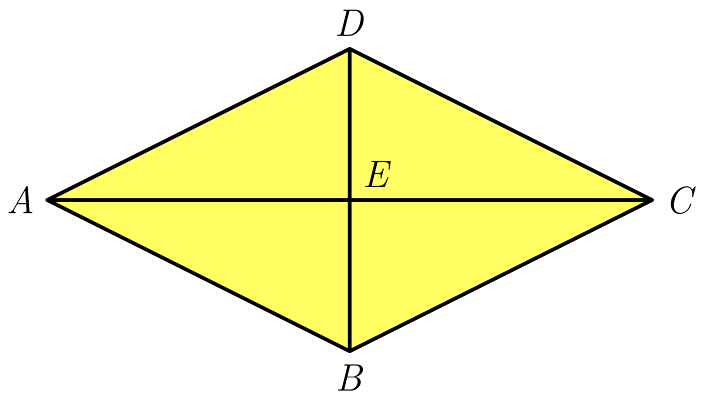
A rhombus is a parallelogram in which all four sides have equal length.

12.01 The parallelogram formula



The rhombus is a special case of a parallelogram. Hence the formula for the area of a parallogram will also work for a rhombus. That is, its area can be found as the product of the length of its base and its height.

12.02 The diagonals formula



We will start by proving that the diagonals of a rhombus bisect each other and intersect at right angles. That property will let us easily derive a formula for the area of a rhombus that uses the lengths of its two diagonals.

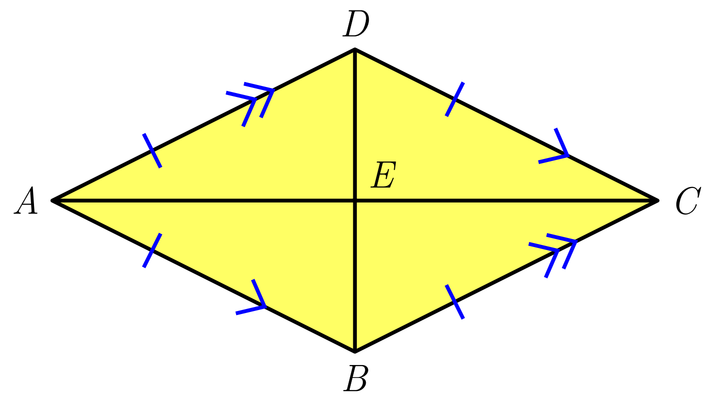
Draw the two diagonals AC and BD and let E be their point of intersection. Orient the rhombus standing on vertex B with AC horizontal.

Does the diagram look symmetrical? Do any triangles look like they might be congruent? I'm not asking for a formal proof. Just guess! A good strategy for geometric proofs is to draw an accurate diagram, use that to guess what is going on, and then try to prove your guess correct. For the moment, just do the initial guessing part.

I guess that if we reflect $\triangle ABD$ through the diagonal BD, it would overlap with $\triangle CBD$, making those two triangles congruent. Similarly, it looks like if we reflect $\triangle ABC$ in diagonal AC it overlaps with $\triangle ADC$, making those two triangles congruent. That is, it looks like the rhombus has mirror symmetry through both diagonals.

I'd also guess that the four smaller triangles that have E as a vertex are congruent. It looks like we could start with $\triangle ABE$, reflect it through BD to get $\triangle CBE$, then reflect that triangle through AC to get $\triangle CDE$, and reflect that through BD to get $\triangle ADE$

At this stage, these are just guesses; we haven't proved anything. Let's review what properties we have at our disposal to try to complete the proof. A rhombus is a parallelogram in which all four sides have equal length. A parallelogram is a quadrilateral in which opposite sides are parallel. Since this proof is heavily reliant on these properties, I'll indicate them on the diagram using the standard markers. I've made them blue for clarity. We know:



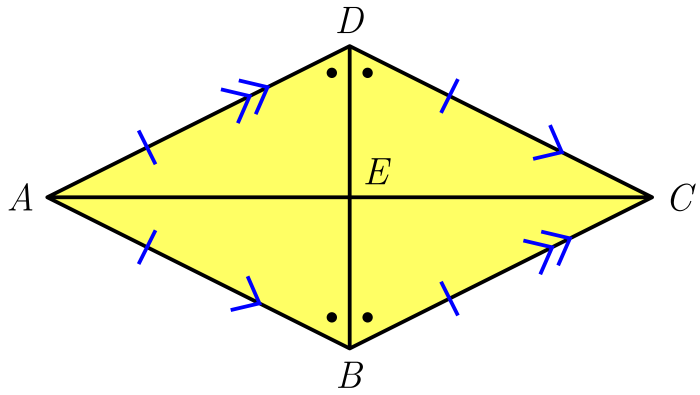
1. $|AB| = |BC| = |CD| = |DA|$, marked with a single blue stroke.
2. AB and DC are parallel, marked with a single blue angle bracket.
3. BC and AD are parallel, market with a double blue angle bracket.

If you're feeling adventurous, you can now use these properties to try to prove that diagonals AC and AD bisect each other at right angles. That is, AC and BD intersect at right angles, and the point of intersection E bisects both diagonals, so $|AE| = |EC|$ and $|BE| = |ED|$.

If you're not feeling quite that adventurous, the next few pages break the task into smaller steps with extra hints. There are many ways to complete the proof, so if you want to venture forth without looking at the following pages you might discover a different method to the one I'll be showing.

I'm going to prove the 4 smaller triangles are congruent. There are other methods which first prove that larger triangles such as $\triangle ABD$ and $\triangle CDB$ are congruent, and gradually work their way down to the smaller triangles, but the fastest approach is to go directly for the four small triangles.

Remember that when identifying congruent triangles, we should list the vertices in corresponding order. The four small triangles we suspect to be congruent can be listed as $\triangle ABE$, $\triangle CBE$, $\triangle CDE$ and $\triangle ADE$. So if we take the angle at the vertex listed first in the first triangle, it will correspond to the angle listed first in each of the other three triangles. The same applies to the angles at the vertices listed second in each triangle.

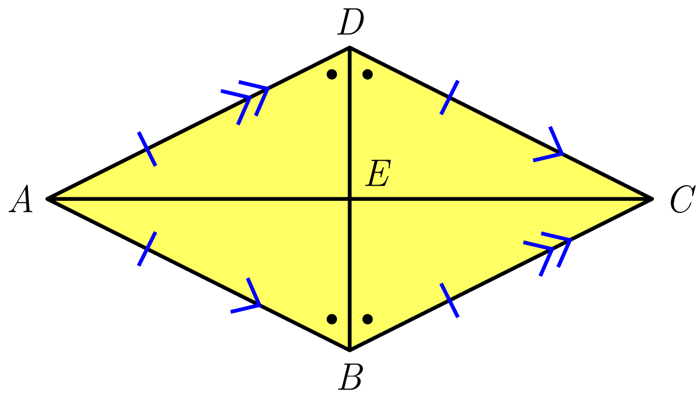


The angles at the vertices listed second have been marked by a dot in the diagram shown here. If the four triangles are congruent, then these four angles would be equal.

So let's investigate whether these four angles are equal. It we can prove them equal, it won't *prove* the four triangles are congruent, but it would be supporting evidence, and combining it with other arguments might give us a complete proof. Conversely, if our investigation proves those four angles are not necessarily equal, that would conclusively prove the four small triangles are not necessarily congruent, so our guess was wrong. Either outcome would be progress, so it *is* worth investigating whether these four angles are equal.

Try to prove the four angles marked by dots are equal. You'll need to employ two different types of arguments to cover all four angles. Here are two hints, one for each type of argument. What type of triangle is $\triangle ABD$? Parallel lines will be useful.

The four sides of a rhombus have equal length, so $|AB| = |AD|$. This means that $\triangle ABD$ is either:



- isosceles, which would make $\angle ABD = \angle ADB$, or
- equilateral, which makes all three angles equal.

In both cases, this proves the left two of the four dotted angles equal.

The same argument shows $|CB| = |CD|$, so $\triangle CBD$ is isosceles or equilateral, proving $\angle CBD = \angle CDB$. That is, the right two of the four dotted angles are equal to each other.

We need a different argument to prove that the left pair are equal to the right pair. The opposite sides of a rhombus are parallel. Line BD is a traversal on the parallel lines AB and DC. Hence $\angle ABD = \angle CDB$, because they are alternating angles on the parallel lines AB and DC.

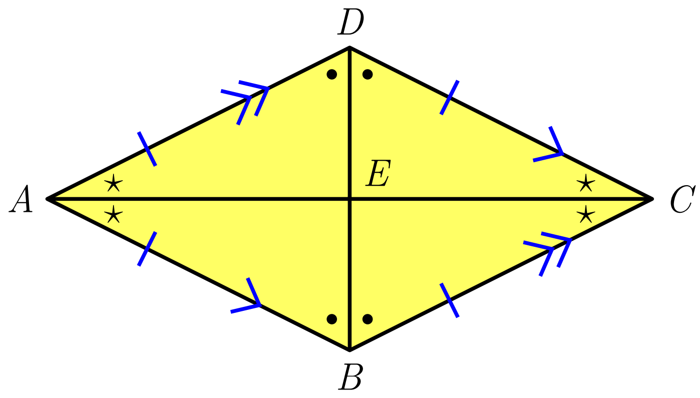
We have now proved $\angle ABD = \angle ADB$, $\angle ABD = \angle CDB$ and $\angle CBD = \angle CDB$. Combining these three equalities proves that all four dotted angles are equal.

Here are two other methods to prove the left two dotted angles equal the right two dotted angles.

- $\angle ADB = \angle CBD$, since they are alternating angles on the parallel lines AD and BC, with BD as the traversal.
- $\triangle ABD$ can be proved congruent to $\triangle CBD$ using the SSS congruency test.

Find a way to reuse this argument in a different orientation to identify some other equal angles. Hence prove the four small triangles are congruent.

We can reuse the previous arguments to show that the four angles marked with stars are equal.



Since the four sides of a rhombus have equal length, $|AB| = |CB|$, making $\triangle ABC$ isosceles or equilateral, which makes the lower two starred angles equal. Similarly $|AD| = |CD|$, making $\triangle ADC$ isosceles or equilateral, which makes the upper two starred angles equal. Finally, $\angle BAC = \angle ACD$, since they are alternating angles on the parallel lines AB and DC using AC as the traversal. This forces all four starred angles to be equal.

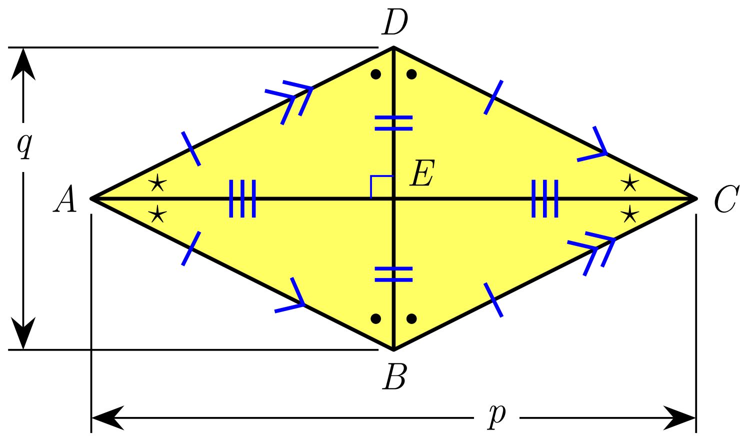
Each of the four small triangles has:

- one angle marked with a dot;
- one angle marked with a star;
- a side between those two angles which is a side of the rhombus, and the four sides of the rhombus have equal length.

Thus by the AAS congruency test — two angles and a corresponding side equal — the four small triangles are congruent.

Use the fact that the four small triangles are congruent to prove that the diagonals bisect each other at right angles.

Since the four small triangles are congruent, their third angles must also be equal. That is $\angle AEB = \angle CEB = \angle CED = \angle AED$. But these 4 angles sum to 360°, so each angle must be 90°. That is, the two diagonals intersect at right angles.



Also, since the four triangles are congruent, their other corresponding sides must be equal. Hence $|BE| = |ED|$, marked with blue double strokes on the diagram, and $|AE| = |EC|$, marked with blue triple strokes. That is, the point E bisects both diagonals.

Putting the conclusions of the last two paragraphs together gives the required theorem.

Theorem 12.1: The diagonals of a rhombus bisect each other at right angles.

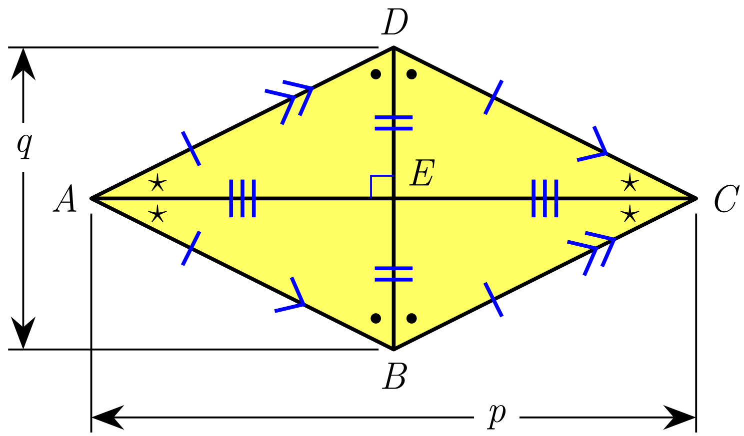
The rhombus is sometimes described as an "orthodiagonal" quadrilateral, though I don't recommend that since it is making the description more complex than it needs to be. "Orthodiagonal" means its diagonals are orthogonal. The word "orthogonal" is more general than perpendicular, having applications in abstract geometry in more than three dimensions. It also has meaning for curves, where "perpendicular" only has meaning for lines, line segments and rays. Two curves are orthogonal if they intersect at right angles, but they don't need to be perpendicular elsewhere. It is both simpler and more informative to say that the diagonals of a rhombus are perpendicular rather than orthogonal.

Let the lengths of the diagonals be $p = |AC|$ and $q = |BD|$, as shown in the diagram.

Derive a formula for the area of the rhombus in terms of the length of its diagonals, p and q.

By the Area Sum Postulate, the area of rhombus $ABCD$ is equal to the sum of the areas of the four small triangles. However we have shown that the four small triangles are congruent, and so they have equal area. Because the diagonals intersect at right angles, they are right triangles, and their area is easily found from the lengths of their legs. But the diagonals bisect each other, and so the leg lengths of the small right triangles are half the lengths of the diagonals.

$$\text{Area of rhombus } ABCD = 4 \times \text{Area of } \triangle AED = 4 \times \tfrac{1}{2} \times \tfrac{1}{2}p \times \tfrac{1}{2}q = \tfrac{1}{2}pq$$



Theorem 12.2: The area of a rhombus is the half the product of the lengths of its diagonals.

12.03 Special case: the square

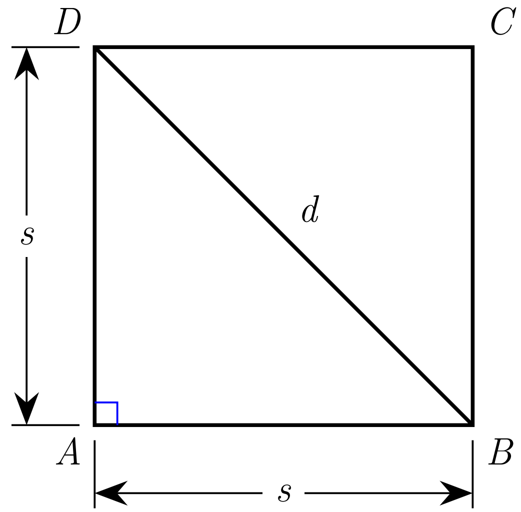
The area of a square is the square of its side length. The area of a rhombus is half the product of the lengths of its diagonals. The square is a special case of the rhombus, which means all squares are rhombi, so the rhombus area formula should work for squares.

Use Pythagoras' theorem to show that the rhombus area formula gives the correct answer when applied to squares.

Here are some extra hints for those who were unable to complete the previous challenge.

Consider a square with a side length of s units. According to the formula for the area of square, what is its area? According to the Pythagorean theorem, what are the lengths of its diagonals? According to the diagonals formula for the area of a rhombus, what is its area?

Here is a square with side length s units. The formula for the area of a square tells us this square has area s^2 square units.



Let the length of diagonal BD be d units. Adjacent sides of a square are perpendicular, so $\triangle ABD$ is a right triangle. Applying the Pythagorean theorem to this triangle gives $d^2 = s^2 + s^2$, so $d = s\sqrt{2}$. The sides of a square all have equal length, so applying this method to diagonal AC gives the same result. That is, the diagonals of a square have equal length.

Applying the diagonals formula for the area of a rhombus to the square gives its area in square units to be $\frac{1}{2} \times s\sqrt{2} \times s\sqrt{2} = s^2$.

That is, the diagonals formula is giving the same answer as the simpler formula for the area of a square, as it should.

13 Kite

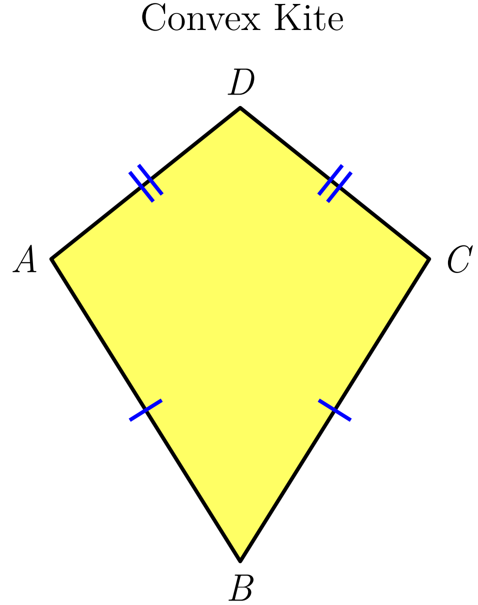
13.01 Definition

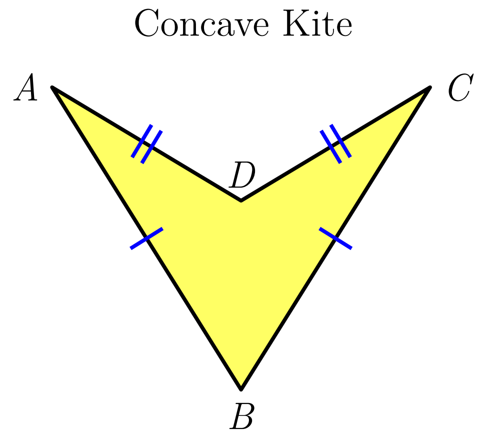
The kite has some interesting geometrical properties, but few real world practical applications. Some practical geometry books ignore it entirely, or mention it only briefly for completeness and don't delve into its properties.

A kite is a quadrilateral in which the four sides consist of two pairs of adjacent sides of equal length. In these diagrams, $|AB| = |BC|$ and $|AD| = |DC|$.

Some authors also require that the kite be convex. Others do not, and subdivide kites into convex and concave kites. I will follow the second approach.

A polygon is concave if one or more of its diagonals fall outside the polygon. For the concave kite shown here, the diagonal AC falls outside the kite. This kite may be described as being concave at D.

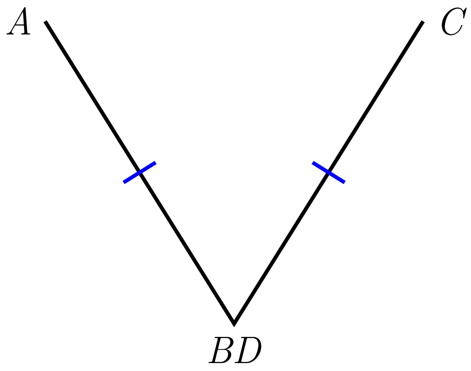




Some authors require that the two pairs of sides of equal length have different lengths to each other, so $|AB| \neq |AD|$. This means kites are distinct from rhombi. Others do not impose this constraint, and so regard the rhombus as a special case of the convex kite. I will follow the second approach.

The kite can only be concave at D if $|AD| = |DC| < |AB| = |CD|$. If we move the point D downward towards B we will maintain those constraints while increasing the length of sides AD and DC. What happens if we try to move D far enough to make the four sides equal in length? Can we change the concave kite into a concave rhombus?

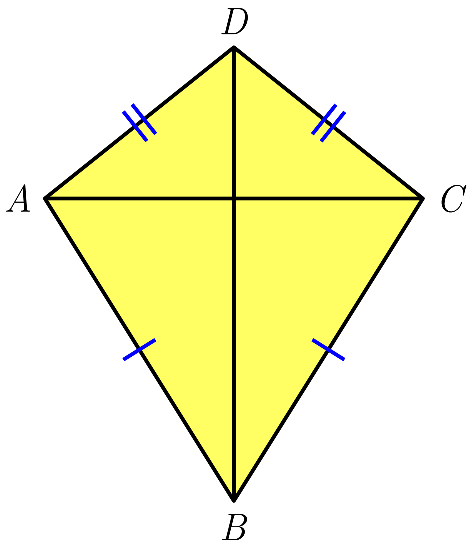
The four side would have equal length when the point D coincides with B. But then we would no longer have a quadrilateral. We would just have two joined line segments, each drawn twice. Hence there is no such thing as a concave rhombus.



So while the rhombus may be viewed as a special case of the convex kite, the concave kite has no comparable special case.

13.02 The diagonal property for convex kites

The area of a concave kite can be calculated given the lengths of its diagonals. The formula is easily derived once we have proved the following theorem.

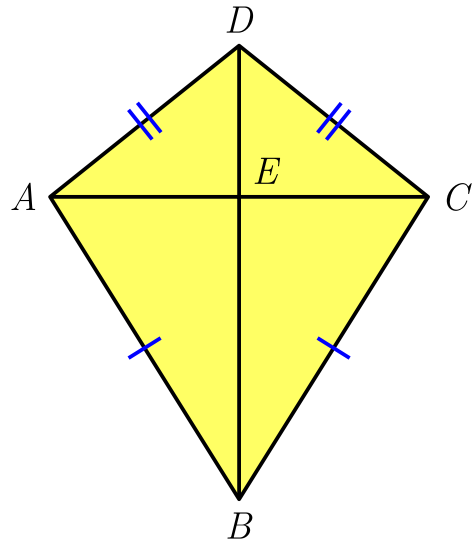


Theorem 13.1: If $ABCD$ is a convex kite with $|AB| = |BC|$ and $|AD| = |DC|$, then its diagonals intersect at right angles with diagonal BD bisecting diagonal AC.

If you are feeling adventurous, you can now try to prove that result. If you are less adventurous, work through the next few pages which break the task into smaller challenges.

There are multiple valid solutions, so if you found a proof without using the following hints, your solution may use a different method to that used below.

In the previous chapter, when we proved a similar theorem for the rhombus, we started by drawing an accurate diagram and made guesses about lines of symmetry and congruency of triangles. Then we tried to prove the guesses correct. We will try the same approach here.

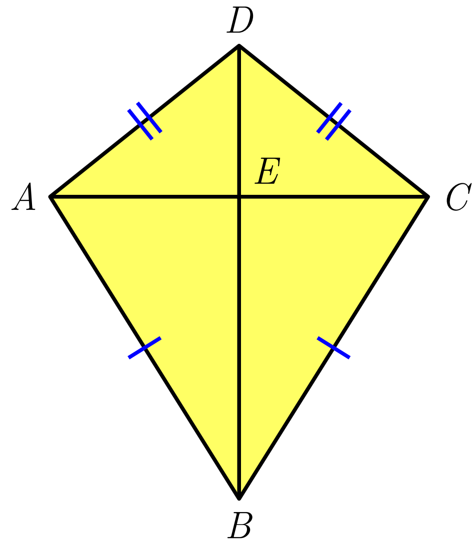


Quadrilateral $ABCD$ is a concave kite with $|AB| = |BC|$ and $|AD| = |DC|$. Both diagonals have been drawn and intersect at E.

> Can you spot any likely symmetries in the kite? Are there any triangles that look like they might be congruent?

It looks like the kite has line symmetry through the diagonal BD. If this guess is correct, then the diagram contains 3 pairs of congruent triangles.

- $\triangle ABD$ and $\triangle CBD$
- $\triangle AED$ and $\triangle CED$
- $\triangle AEB$ and $\triangle CEB$

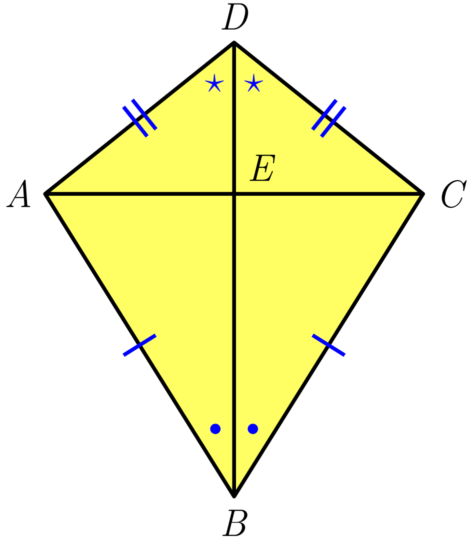


If we can prove the first pair of triangles in the above list are congruent, then reflecting $\triangle ABD$ through the diagonal BD gives $\triangle CBD$, which proves the kite is symmetrical through BD.

With the corresponding proof for the rhombus, we were able to directly prove the smaller triangles congruent, but that step relied on the opposite sides of a rhombus being parallel. Our kite has no parallel sides, so we are forced to adopt a longer approach. We will first need to prove the first pair of triangles listed above are congruent.

Prove $\triangle ABD$ and $\triangle CBD$ are congruent. Given that congruence, which angles can be shown to be equal?

Consider $\triangle ABD$ and $\triangle CBD$. They share the common edge BD. We know $|AB| = |BC|$ and $|AD| = |DC|$ because $ABCD$ is a kite.



Hence $\triangle ABD$ and $\triangle CBD$ are congruent by the SSS congruency test — three pairs of equal sides.

Corresponding angles of congruent triangles are equal. Hence:

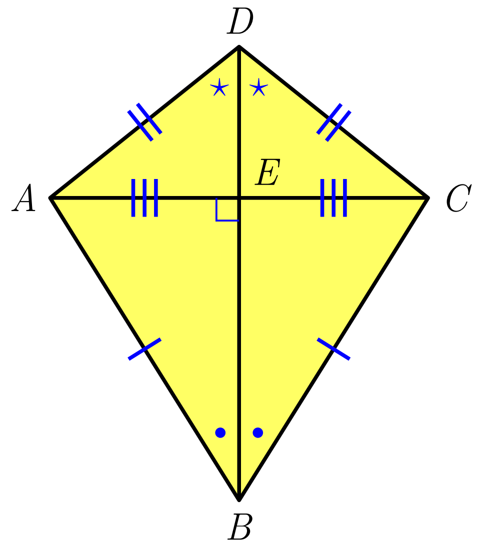
1. $\angle ABD = \angle CBD$ These angles marked with dots on the diagram.
2. $\angle ADB = \angle CDB$ These are marked with stars.
3. $\angle BAD = \angle BCD$

Our next step will be to prove the congruency of one of the pairs of smaller triangles. The third of the three angle equalities listed above refers to angles that don't belong to any of the smaller triangles, so it is unlikely to be useful. Hence I didn't even bother to mark those angles in the diagram.

For the next step, we could choose either the top or bottom pair of smaller triangles. I'll use the bottom pair. For this proof, it makes no difference which pair we choose, but later, we will work through the same process for a concave kite, and this choice will make that later proof slightly shorter.

Prove $\triangle AEB$ and $\triangle CEB$ are congruent. Use that congruence to prove that the kite's diagonals are perpendicular and that diagonal BD bisects diagonal AC.

$\triangle AEB$ and $\triangle CEB$ share the common edge BE. We know $|AB| = |CB|$. Also $\angle ABD = \angle CBD$, which in the context of these triangles we can write as $\angle ABE = \angle CBE$



Hence $\triangle AEB$ and $\triangle CEB$ are congruent by the SAS congruency test — two pairs of equal sides and equal included angles.

Corresponding sides of congruent triangles have equal length, so $|AE| = |CE|$, marked with blue triple strokes on the diagram. Hence diagonal BD bisects diagonal AC, which proves part of the theorem.

Corresponding angles of congruent triangles are equal. Hence $\angle AEB = \angle CEB$. Point E falls on diagonal AC, so $\angle AEC = 180°$. Hence

$$\begin{aligned} \angle AEB + \angle CEB &= \angle AEC \\ 2 \times \angle AEB &= 180° \\ \angle AEB = \angle CEB &= 90° \end{aligned}$$

Hence diagonals AC and BD are perpendicular, which proves the remaining part of the theorem and also proves $\angle AED = \angle CED = 90°$, which may be useful when it comes to calculating the kite's area. Here again is the full theorem we have proved.

Theorem 13.1: If $ABCD$ is a convex kite with $|AB| = |BC|$ and $|AD| = |DC|$, then its diagonals intersect at right angles with diagonal BD bisecting diagonal AC.

This theorem refers to labelled points. That lets us specify which diagonal is being bisected.

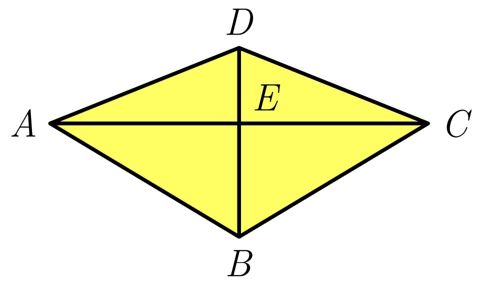
Here is a shorter version of the theorem that doesn't use labelled points, but it is faulty. Identify the fault.

Faulty Version of Theorem 13.1: The longer diagonal of a convex kite bisects the shorter diagonal at right angles.

For those who weren't able to complete the previous challenge, here's an extra hint. The faulty theorem works for the diagram given, but that diagram does not cover *all* the interesting cases.

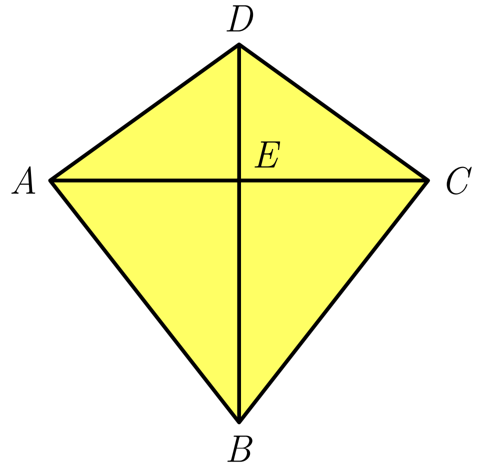
Find a diagram where the faulty theorem fails.

In the diagram we were using, diagonal AC was the shorter diagonal. Here are two other possible shapes for a convex kite. In each diagram, $|AB| = |BC|$ and $|AD| = |DC|$, which is consistent with the earlier diagram and the correct statement of theorem 13.1.



In the first diagram here, diagonal AC is longer than diagonal BD.

In the second diagram the two diagonals have equal length, so the phrase "shorter diagonal" has no meaning. You might need to use a ruler on the second diagram to verify this claim. Some people find that the sloping kite edges create an optical illusion that makes the vertical diagonal seem longer than the horizontal diagonal.



If you've been studying this book attentively, you already know what you need to do next. Make a copy of these two diagrams. Then go back and reread the entire proof of theorem 13.1, checking whether it remains valid for these alternative diagrams.

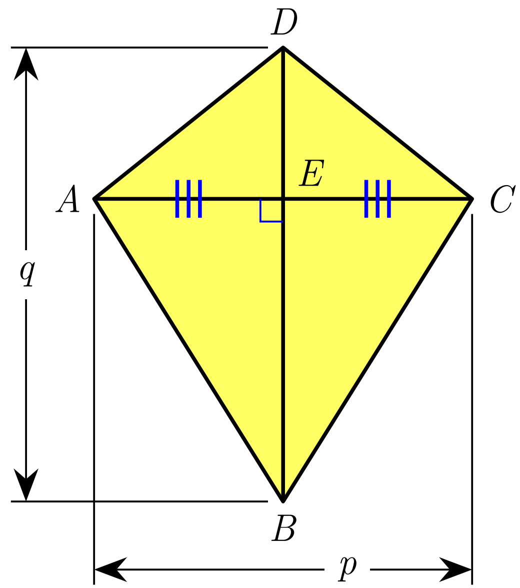
If you do this, you should find the proof is valid for both these alternative diagrams, so diagonal BD still bisects diagonal AC. That is, there are convex kites where the shorter diagonal bisects the longer diagonal, and there are kites where the diagonals have equal length, so referring to the shorter diagonal makes no sense. Either of these cases is sufficient to show that the alleged simpler form of the theorem is faulty.

13.03 Area of a Convex Kite

We will now derive a formula for the area of a convex kite in terms of the lengths of the two diagonals.

> Try it. There are more hints on the next page if you are stuck, but there several easy ways to derive the formula. If you've thoroughly understood the proof of theorem 13.1 there's a good chance you can find a valid method without looking at the hints.

I'll revert to the "original" concave kite diagram here, but you can check that the process we use also works for the alternative convex kites shown on the previous page. To make things a little easier, I've simplified the diagram, removing some symbols that indicated properties that we won't need in this final step.



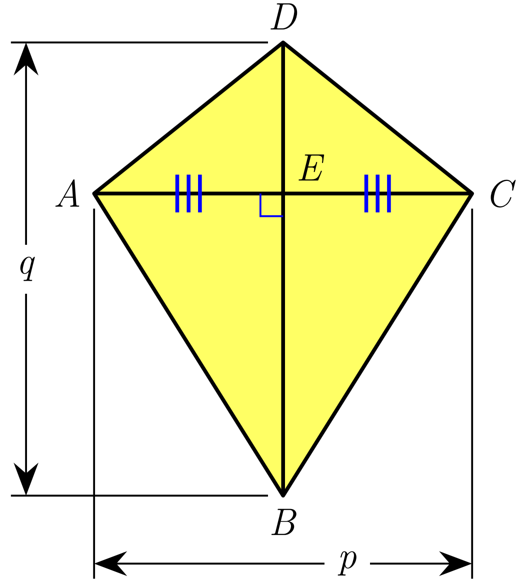
Let $p = |AC|$ and $q = |BD|$

The simplest approach to finding the area of the kite is use the areas of $\triangle ABD$ and $\triangle ABD$.

Our proof of theorem 13.1 showed that these triangles are congruent. How does that help? The theorem also proved the diagonals are perpendicular, with BD bisecting AC. How does that help? Find a formula for the area of the kite in terms of p and q.

Since $\triangle ABD$ and $\triangle CBD$ are congruent they have equal area.

To find the area of $\triangle ABD$, if we use line segment BD as the base, then the corresponding altitude is the perpendicular from point A to line BD. But from theorem 13.1 we know diagonals AC and BD are perpendicular, so the altitude must be line segment AE. The bisection property means $|AE| = \frac{1}{2}|AC|$.



$$
\begin{aligned}
&\text{Area of Kite } ABCD\\
&= \text{Area of } \triangle ABD + \text{Area of } \triangle CBD \text{ by the Area Sum Postulate}\\
&= 2 \times \text{Area of } \triangle ABD \text{ because the triangles are congruent}\\
&= 2 \times \tfrac{1}{2} \times |AE| \times |BD|\\
&= 2 \times \tfrac{1}{2} \times \tfrac{1}{2}p \times q \text{ because } BD \text{ bisects } AC\\
&= \tfrac{1}{2}pq
\end{aligned}
$$

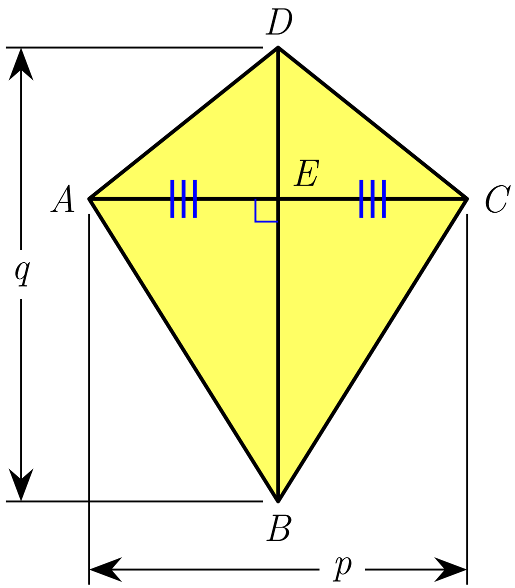
Here is the resulting theorem

Theorem 13.2: The area of a convex kite is the half the product of the lengths of its diagonals.

That is, the diagonal formula that we derived earlier for the rhombus applies more generally to all convex kites.

I stated that the method given above was the simplest way to find the area of the kite. If you are interested, verify that you get the same result if you use the areas of $\triangle ABC$ and $\triangle ADC$, and that you also get the same result if you use the four small right triangles.

As in the previous example, the fact that diagonals are perpendicular means that if we use line segment AC as the base of $\triangle ABC$ and $\triangle ADC$, the corresponding altitudes are BE and DE respectively.



Area of Kite $ABCD$

$= \text{Area of } \triangle ABC + \text{Area of } \triangle ADC$ by the Area Sum Postulate

$= \frac{1}{2} \times |AC| \times |BE| + \frac{1}{2} \times |AC| \times |DE|$

$= \frac{1}{2} \times |AC| \times (|BE| + |ED|)$

$= \frac{1}{2} \times |AC| \times |BD|$

$= \frac{1}{2}pq$

Using the four right triangles gives

Area of Kite $ABCD$

$= \text{Area of } \triangle ABE + \text{Area of } \triangle BCE + \text{Area of } \triangle CDE + \text{Area of } \triangle DAE$ by the Area Sum Postulate

$= \frac{1}{2} \times |AE| \times |BE| + \frac{1}{2} \times |EC| \times |BE| + \frac{1}{2} \times |EC| \times |ED| + \frac{1}{2} \times |AE| \times |ED|$

$= \frac{1}{2} \times (|AE| + |EC|) \times (|BE| + |ED|)$

$= \frac{1}{2} \times |AC| \times |BD|$

$= \frac{1}{2}pq$

13.04 Special case: Rhombus

As mentioned earlier, the rhombus is a special case of the convex kite. To be more precise, many authors hold this view, but some authors put an extra constraint in their definition of the kite to exclude the rhombus. I won't mention the latter group again in this section.

Because of this relationship, we'd expect that the formula for the area of a convex kite will also hold for a rhombus. This is consistent with what we found. Theorems 12.2 and 13.2 both refer to calculating the area as half of the product of the lengths of the diagonals.

But compare the following two theorems.

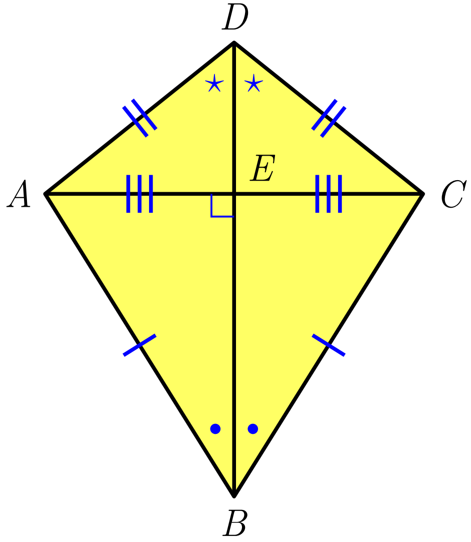
Theorem 12.1: The diagonals of a rhombus bisect each other at right angles.

Theorem 13.1: If $ABCD$ is a convex kite with $|AB| = |BC|$ and $|AD| = |DC|$, then its diagonals intersect at right angles with diagonal BD bisecting diagonal AC.

When shown these two theorems together, some people argue as follows. The kite theorem says only one diagonal is bisected. This is true for all convex kites, so it must be true for the rhombus, since the rhombus is a special case of the convex kite. But the rhombus theorem says the diagonals bisect each other. This is contradictory so one of these theorems is wrong.

Explain what these people have misunderstood.

Theorem 13.1 does not claim that only one diagonal is bisected. It labels the vertices, and then claims that one particular diagonal is bisected, the diagonal AC. It makes no claims about whether or not the other diagonal, diagonal BD, is also bisected.



I was very careful to write theorem 13.1 in a way that makes no claim about whether or not diagonal BD is bisected, because it could go either way. Most convex kites, like the one shown here, are not rhombi, and for those BD is not bisected. But some convex kites are rhombi, and for those diagonal BD is bisected.

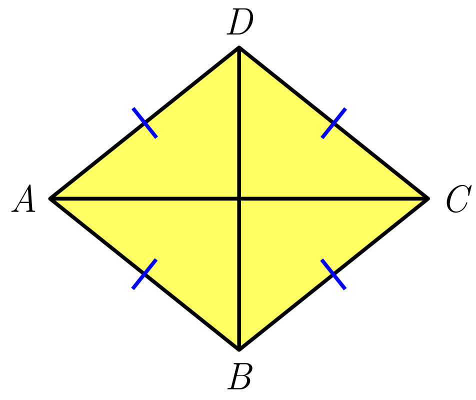
Thus the two theorems are not contradictory. In fact, it is even possible to use theorem 13.1 to prove theorem 12.1.

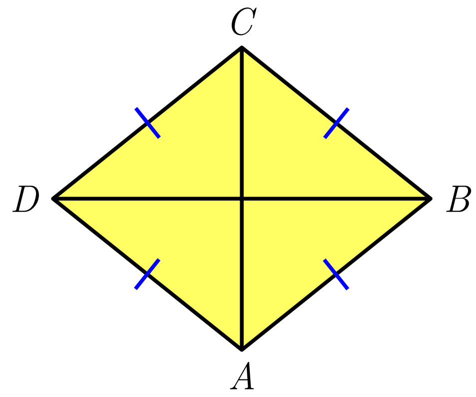
Use theorem 13.1 to prove theorem 12.1.

Hint: When you apply theorem 13.1 to a rhombus, you are *not* restricted to putting the labels in the same relative positions as shown in the kite above. For example, the lowest vertex does not have to be labelled B. The theorem refers to it as kite $ABCD$, so the labels do have to be assigned to the vertices in that order, going either clockwise or anticlockwise. Also, they have to be assigned in a way that ensures $|AB| = |BC|$ and $|AD| = |DC|$. Those are your only constraints.

Also, in an earlier topic we referred to the dangers of circular arguments. I said we could avoid the risk of introducing a circular argument into a proof by only using theorems proved earlier in the book. Using theorem 13.1 to prove theorem 12.1 breaks that ordering. This does not *automatically* create a circular argument, but it does mean we should be vigilant and carefully check for the presence of circular arguments.

The diagram shows two copies of a rhombus. The two rhombi are congruent, but they have been labelled differently. In both cases the labels progress in anticlockwise order, but they start at a different vertex. Since the shape is a rhombus, all four sides have equal length, so both label schemes comply with the constraints $|AB| = |BC|$ and $|AD| = |DC|$.





Since both label schemes comply with the constraints of theorem 13.1, we can apply that theorem to both diagrams, meaning in both diagrams diagonal BD bisects diagonal AC. In the first diagram this means the vertical diagonal bisects the horizontal diagonal, while in the second diagram this means the horizontal diagonal bisects the vertical diagonal. But while the labelling differs the two rhombi are congruent. They are two copies of the same rhombus, so both these properties have to apply to the rhombus at once, which means the two diagonals must bisect each other.

Whichever diagram we use, theorem 13.1 also tells us the diagonals meet at right angles. Hence we can conclude that in the special case where the convex kite happens to be a rhombus, the diagonals bisect each other at right angles, which proves theorem 12.1.

As flagged, we should now check for circularity. This is easily done. Just reread the proof of theorem 13.1 and verify that it does not use theorem 12.1. This demonstrates there are no circular arguments present.

I lied. While it is true that there are no circular arguments present, the process described in the previous paragraph is not sufficient to prove that. Can you spot the extra check we need to perform?

We used theorem 13.1 to prove theorem 12.1. Let's write that dependency as: 13.1 → 12.1.

We checked that the proof of theorem 13.1 does not use theorem 12.1. That is, we checked that we didn't accidentally use the circular argument: 13.1 → 12.1 → 13.1.

But a circular argument can involve more than two theorems, so we should also check we don't have the circular argument: 13.1 → 12.1 → 12.2 → 13.1

It's easy enough to check we don't have that circular argument. Theorem 13.1 does not rely on theorem 12.2.

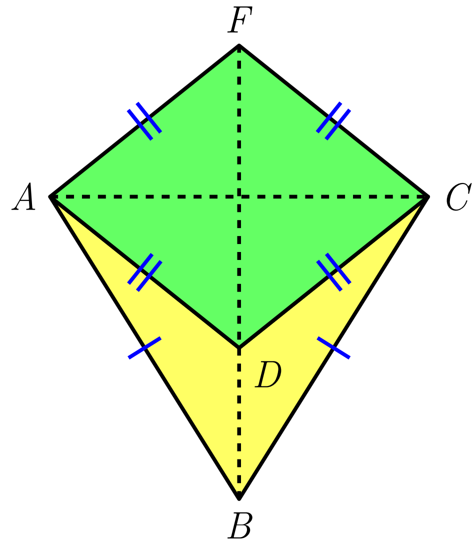
Another way to say this is that initially we only checked that theorem 13.1 does not *directly* use theorem 12.1. That is not sufficient. We also need to not check theorem 13.1 does not depend on theorem 12.1 *indirectly* via some other theorem that was developed between theorem 12.1 and theorem 13.1. We verified there was no such dependency, so all is well. Using this alternative approach of deriving 12.1 from theorem 13.1 does not create any circular arguments.

13.05 Area of concave kite

The simplest way to derive the formula for the area of the concave kite is to employ a convex kite and a rhombus. The following derivation uses this method but it contains an error. The error involves assuming a property to be true without proof. Try to spot where the error happens.

Let $ABCD$ be a concave kite, the concavity happening at vertex D, with $|AB| = |CB|$ and $|AD| = |CD|$. Its diagonals, AC and BD, have been shown as dashed lines.

Reflect point D in the diagonal AC, giving point F. Join F to A and to C.



Line segments AF and CF are reflections of sides AD and CD respectively, and we know those last two line segments have equal length. Since reflection preserves length, $|AD| = |CD| = |AF| = |CF|$, so quadrilateral $ADCF$ is a rhombus. Its diagonals, AC and DF are shown as dashed lines. By theorem 12.2 its area is $\frac{1}{2} \times |AC| \times |DF|$. Alternatively, since all rhombi are convex kites, you could use theorem 13.2 instead of theorem 12.2.

Since $|AB| = |CB|$ and $|AF| = |CF|$, quadrilateral $ABCF$ is a kite. Vertex D is the concave vertex of the concave kite $ABCD$, and F was found by reflecting D across the concave kite's external diagonal AC. This means F is definitely above diagonal AC, so $ABCF$ is definitely a convex kite. (Clearly in this diagram $ABCF$ is convex, but this argument shows it must be convex for any concave kite $ABCD$ that is concave at D, not just for the particular concave kite shown here.) This means theorem 13.2 applies to kite $ABCF$, giving its area as $\frac{1}{2} \times |AC| \times |BF|$.

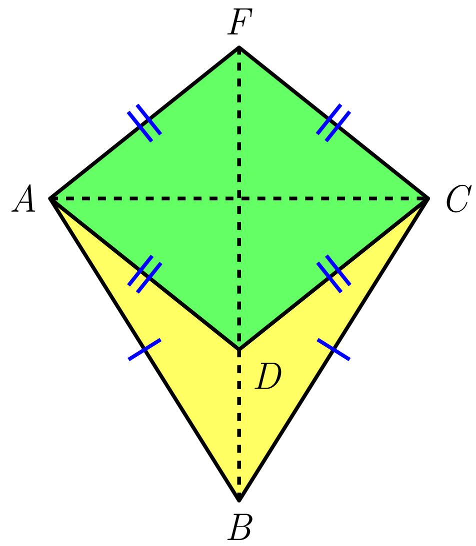
Area of convex kite $ABCF$ = Areas of rhombus $ADCF$ + Area of concave kite $ABCD$, by the area sum postulate. Hence:

$$\begin{aligned}
\text{Area of concave kite } ABCD &= \text{area of convex kite } ABCF - \text{area of rhombus } ADCF \\
&= \tfrac{1}{2} \times |AC| \times |BF| - \tfrac{1}{2} \times |AC| \times |DF| \\
&= \tfrac{1}{2} \times |AC| \times \{|BF| - |DF|\} \\
&= \tfrac{1}{2} \times |AC| \times |BD|
\end{aligned}$$

That is, the area of a concave kite is equal to half the product of the lengths of its diagonals.

Unfortunately, as flagged earlier, this derivation contains an unproven assumption. Identify it.

The error occurs in the very last step, which assumes $|BF| - |DF| = |BD|$. This is only true if the point D falls on line segment BF. We haven't proved that it does.



In the derivation I said the line segments BD and DF were drawn as dashed lines. I didn't draw line segment BF. The derivation assumed BF passes through D, so that we've effectively drawn it by drawing the line segments BD and DF.

Another way to say this is that the derivation assumed B, D and F are collinear. That is, it assumed BDF is a single line segment, meaning there is with no bend at D.

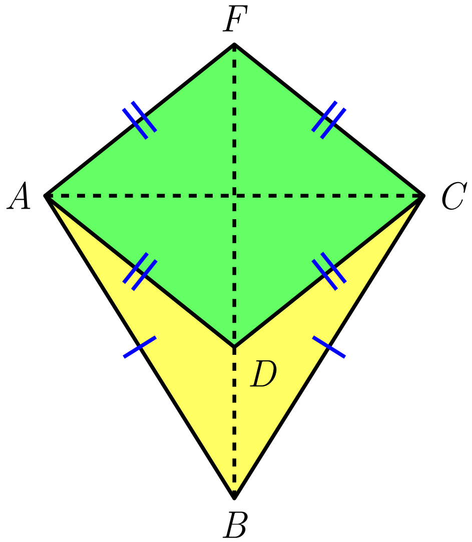
There are multiple ways to prove the assumption valid. Perhaps the simplest is to use the following theorem.

Given a line l and a point P not on that line, there is exactly one line through P perpendicular to l.

The unique line is described more briefly as "the perpendicular to l through P" or "the perpendicular from P to l."

Use this theorem to prove that B, D and F are collinear. You can also use theorems 12.1 and 13.1.

By the theorem, there is exactly one line through the point F perpendicular to line AC.



We proved $ADCF$ is a rhombus, so by theorem 12.1 its diagonals are perpendicular. Thus D must fall on the unique perpendicular to AC through F.

We proved $ABCF$ is a convex kite, so by theorem 13.1 its diagonals are perpendicular. Thus B must also fall on the unique perpendicular to AC through F.

Thus B, D and F are collinear.

This proves the assumption made in the earlier derivation, so we can now trust its conclusion: The area of a concave kite is half the product of the lengths of its diagonals. This is the same result as we obtained for convex kites. Thus rather than stating a new theorem that only works for concave kites, we can produce a single theorem that covers both convex and concave kites.

I mentioned earlier that some authors do not recognise the concave kite, defining the term “kite” so that it requires the shape be convex. To make it clear that I regard concave kites as valid and that I have verified that the formula applies to both types, I prefer to state that explicitly in the theorem. Here is the result.

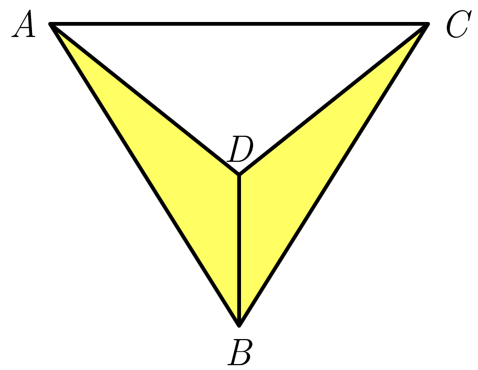
Theorem 13.3: The area of a kite is the half the product of the lengths of its diagonals. This works for both convex and concave kites.

13.06 The diagonal property for concave kites

Theorem 13.1 describes the relationship between the diagonals of convex kites. There is a similar theorem for the diagonals of concave kites, which will be given below and numbered 13.4.

We can use theorem 13.4 to derive the formula for the area of a concave kite. This is a longer process than that given in section 13.05, but some students prefer it because all the steps in the longer process are simpler and tend to be the obvious steps to take. By contrast, in the proof in section 13.05 the unstated assumption about the collinearity of points B, D and F was hard to spot, as was the simplest way to prove it valid.

The diagram shows concave kite $ABCD$, concave at D, with $|AB| = |BC|$ and $|AD| = |CD|$. Diagonal AC falls outside the kite, which means the diagonals do not intersect.



So far, when we have talked about line segments being perpendicular we described them as intersecting at right angles. Thus it's easy to fall into the trap of thinking the word "perpendicular" always requires intersection. For line segments, it doesn't, and diagonals are line segments, not lines. When dealing with line segments, the word "perpendicular" implies something special about the relative orientation of the line segments, but makes no claim about whether or not they intersect. Here, diagonals AC and BD are perpendicular but do not intersect.

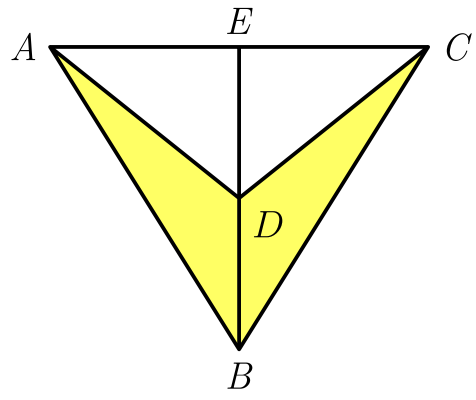
If geometry within the plane, if two lines are perpendicular, they will intersect. This means we can define the term "perpendicular" for line segments by referring to the lines that contain them.

If two lines intersect at right angles, they are perpendicular.

Two non-collinear line segments are perpendicular if the two lines that contain them are perpendicular.

A line is perpendicular to a non-collinear line segment if it is perpendicular to the line that contains that line segment.

These definitions suggest how we should proceed with our concave kite. We should extend diagonal BD until it meets diagonal AC. Label the point of intersection E. To prove that diagonal AC is perpendicular to diagonal BD it is sufficient to prove it perpendicular to line BDE. Two points are sufficient to determine a line, so line BDE can be called line BD or line BE if that is more convenient.



Based on the diagram, it seems plausible to guess point E bisects diagonal AC, which would be consistent with what happened for the convex kite.

Incidentally, when I introduced this kite I said it was concave at D. Thus it's valid to say that we find the point E by extending diagonal BD, which means we extend it beyond D. If the concavity had been at B rather than D, we would have said "extend diagonal DB", since we would need to extend it beyond B.

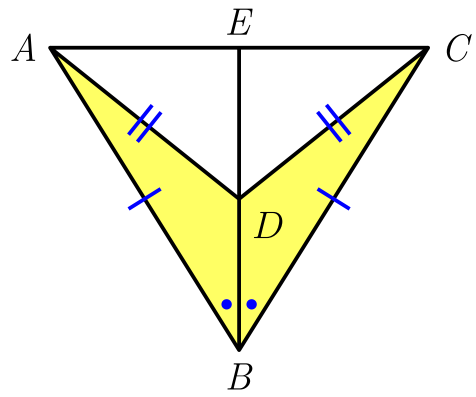
Theorem 13.1 applied to convex kites. Here is the corresponding theorem for concave kites.

Theorem 13.4: If $ABCD$ is a concave kite, concave at vertex D, with $|AB| = |BC|$ and $|AD| = |DC|$, then its diagonals are perpendicular. While the diagonals do not intersect, line BD bisects diagonal AC.

In the proof of theorem 13.1 we began by proving $\triangle ABD$ and $\triangle CBD$ congruent, which let us identify some equal angles. Try the same thing here. The next page will give hints on what to do after that, but if you feel confident, try to complete the proof of theorem 13.4 without reading those hints.

Consider $\triangle ABD$ and $\triangle CBD$. They share the common edge BD. We know the equal sides in kite $ABCD$ are $|AB| = |BC|$ and $|AD| = |DC|$.

Hence $\triangle ABD$ and $\triangle CBD$ are congruent by the SSS congruency test — three pairs of equal sides.



Corresponding angles of congruent triangles are equal. Hence:

1. $\angle ABD = \angle CBD$
2. $\angle ADB = \angle CDB$
3. $\angle BAD = \angle BCD$

Our aim is to prove the two angles at point E are right angles, so it would be useful if we can find a pair of congruent triangles that include angles at point E. The two pairs of triangles including E that could possibly be congruent are:

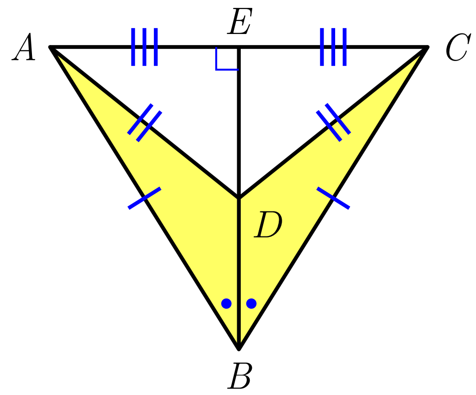
1. $\triangle AED$ and $\triangle CED$
2. $\triangle AEB$ and $\triangle CEB$

Compare the three known pairs of equal angles listed above with the two pairs of possibly congruent triangles. None of the angle pairs belong to the first pair of triangles, but the first angle pair belong to the second pair of triangles, so let's try to use that first angle pair to prove the second pair of triangles congruent. To help, I've marked that first angle pair with dots on the diagram.

Prove $\triangle AEB$ and $\triangle CEB$ are congruent. Use that congruence to prove that the kite's diagonals are perpendicular and that line BD bisects diagonal AC.

Consider $\triangle AEB$ and $\triangle CEB$. They share the common edge BE. As stated earlier, we know $|AB| = |CB|$ because $ABCD$ is a kite. We have proved $\angle ABD = \angle CBD$, which in the context of these triangles we can write as $\angle ABE = \angle CBE$

Hence $\triangle AEB$ and $\triangle CEB$ are congruent by the SAS congruency test — two pairs of equal sides and equal included angles.



Corresponding sides of congruent triangles have equal length, so $|AE| = |CE|$. That is, the point E bisects diagonal AC. But the point E is on line BD, so line BD bisects diagonal AC, which proves part of theorem 13.4.

Corresponding angles of congruent triangles are equal. Hence $\angle AEB = \angle CEB$. Point E falls on diagonal AC, so $\angle AEC = 180°$. Hence

$$\begin{aligned}\angle AEB + \angle CEB &= \angle AEC\\ 2 \times \angle AEB &= 180°\\ \angle AEB = \angle CEB &= 90°\end{aligned}$$

This means diagonals AC and BD are perpendicular, which proves the remaining part of the theorem. Here again is the resulting theorem.

Theorem 13.4: If $ABCD$ is a concave kite, concave at vertex D, with $|AB| = |BC|$ and $|AD| = |DC|$, then its diagonals are perpendicular. While the diagonals do not intersect, line BD bisects diagonal AC.

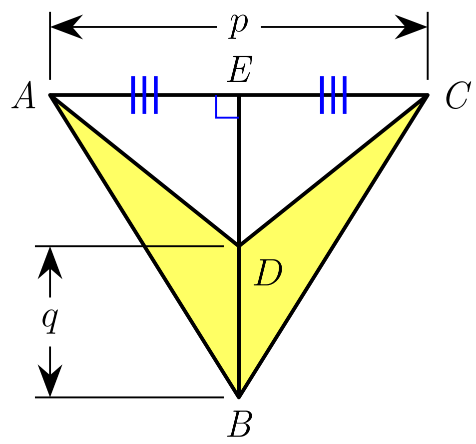
13.07 Area of concave kite

Let $p = |AC|$ and $q = |BD|$

Recall that $\triangle ABD$ and $\triangle CBD$ are congruent. Hence find a formula for the area of the concave kite in terms of p and q.

To make things a little easier, I've simplified the diagram, removing some symbols that indicated properties that we won't need in this final step.

Consider $\triangle ABD$. If we regard BD as its base, then the corresponding altitude is the line segment from A meeting line BD at right angles. But by theorem 13.4, AE is perpendicular to BD, so the required altitude is line segment AE. Since E bisects AC, $|AE| = \frac{1}{2}|AC| = \frac{1}{2}p$.



$$\begin{aligned} \text{Area of kite} &= \text{Area of } \triangle ABD + \text{Area of } \triangle CBD \\ &= 2 \times \text{Area of } \triangle ABD, \text{ because congruent triangles have equal area} \\ &= 2 \times \tfrac{1}{2} \times |BD| \times |AE| \\ &= 2 \times \tfrac{1}{2} \times q \times \tfrac{1}{2}p \\ &= \tfrac{1}{2}pq \end{aligned}$$

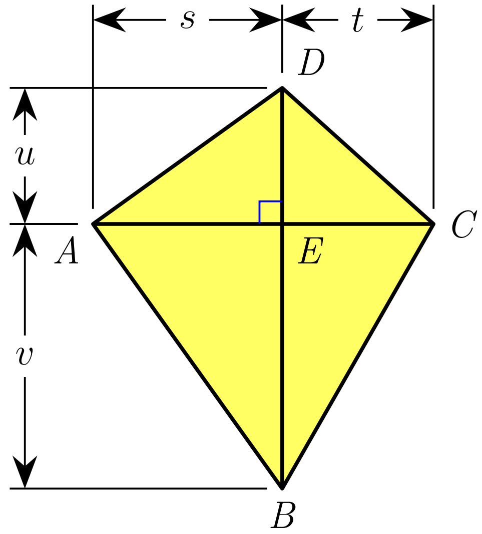
This is consistent with the result of the earlier derivation.

13.08 More general use of the diagonals formula

We have shown that the area of rhombi, and of both convex and concave kites can be found as half of the products of the lengths of the diagonals. In fact, this formula has even more general application. It works for any quadrilateral for which the diagonals are perpendicular.

As far as I am aware, this generalisation is not included in any Australian high school syllabus, but I have included it since is surprisingly easy to prove. Recall that in both this chapter and the earlier chapter on the rhombus, most of our effort was expended on proving these shapes had perpendicular diagonals. Once we know that the diagonals are perpendicular, deriving the formula for the area is easy. Thus if someone gives us a quadrilateral and they simply tell us that its diagonals are perpendicular and we are not required to prove that step, then it's possible that deriving a formula for its area might be very easy.

This diagram shows a convex quadrilateral with perpendicular diagonals. I've deliberately drawn it so that its four sides are of four different lengths and it has no symmetry. That is, it is neither a kite nor a rhombus, though the derivation we are going to use here will also be valid in those special cases.



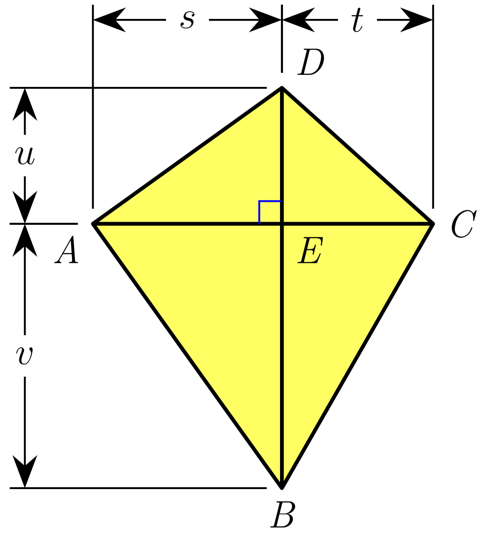
The quadrilateral has been aligned so that the diagonals are vertical and horizontal. The diagonals intersect at E. The variables s, t, u and v denote the distances between the vertices of the quadrilateral and E.

Show that the area of the quadrilateral is half the product of the lengths of its diagonals.

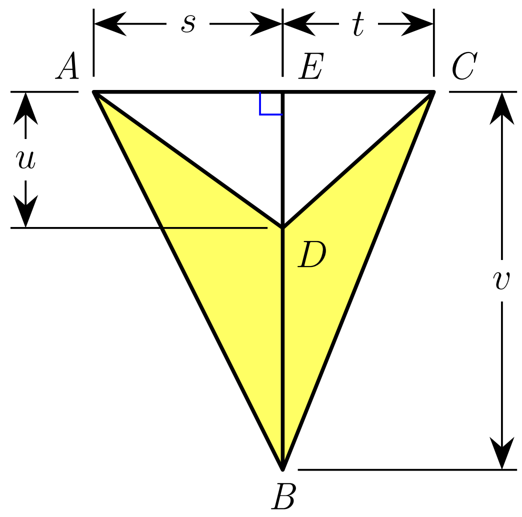
Using the Area Sum Postulate gives

$$\begin{aligned}
&\text{Area of Quadrilateral } ABCD \\
&= \text{Area of } \triangle ABE + \text{Area of } \triangle BCE + \text{Area of } \triangle CDE + \text{Area of } \triangle DAE \\
&= \tfrac{1}{2} \times |AE| \times |BE| + \tfrac{1}{2} \times |EC| \times |BE| + \tfrac{1}{2} \times |EC| \times |ED| + \tfrac{1}{2} \times |AE| \times |ED| \\
&= \tfrac{1}{2} \times \{sv + tv + tu + su\} \\
&= \tfrac{1}{2}(s+t)(u+v)
\end{aligned}$$

This is half the product of the lengths of the diagonals, as required.



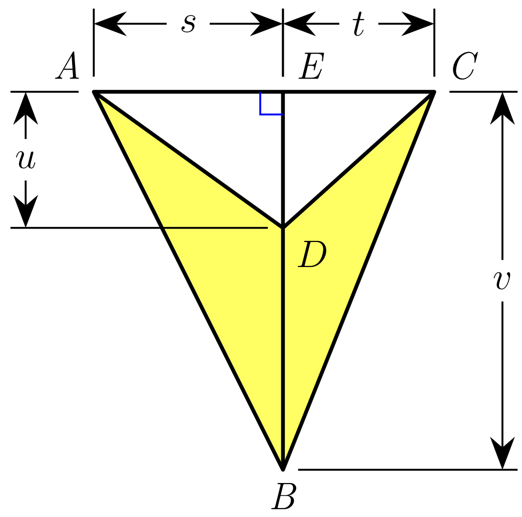
This diagram show a concave quadrilateral, concave at D, with diagonals that are perpendicular. They have been aligned so that the diagonals are vertical and horizontal. Diagonal BD is extended to meet diagonal AC at E. The variables s, t, u and v denote the distances between the vertices of the quadrilateral and E.



Adjust the derivation given above to show that the area of this quadrilateral is half the product of the lengths of its diagonals.

By considering two different ways of finding the area of $\triangle ABC$, we can write:

$$\text{Area of quadrilateral } ABCD + \text{Area of } \triangle ADE + \text{Area of } \triangle CDE$$
$$= \text{Area of } \triangle ABE + \text{Area of } \triangle CBE$$



Rearranging this gives

$$\begin{aligned}
&\text{Area of Quadrilateral } ABCD \\
&= \text{Area of } \triangle ABE + \text{Area of } \triangle BCE - \text{Area of } \triangle CDE - \text{Area of } \triangle DAE \\
&= \tfrac{1}{2} \times |AE| \times |BE| + \tfrac{1}{2} \times |EC| \times |BE| - \tfrac{1}{2} \times |EC| \times |ED| - \tfrac{1}{2} \times |AE| \times |ED| \\
&= \tfrac{1}{2} \times \{sv + tv - tu - su\} \\
&= \tfrac{1}{2}(s + t)(v - u)
\end{aligned}$$

Once again, this is half the product of the lengths of the diagonals, as required.

14 Scaling and Similarity

This chapter is very different to previous chapters in that most of this chapter is spent proving results that aren't about calculating areas. I've done this because most high school textbooks make no attempt to prove these results, so I can't just quote the result and suggest you look at your school textbooks if you want to see the proof. If you have already ventured beyond high school textbooks and have seen formal proofs about the scaling transformation and the theorems for testing for similarity, you may be able to skip over most sections in this chapter.

14.01 Use of translation and scaling in software

High school geometry courses usually examine four types of transformations: translation, rotation, reflection, and scaling. Scaling is considered the most complex and may only be introduced a year or two after the other three transformations. Pre-computers, this probably made sense, but if you have been using software from a young age, you may well be more experienced with translation and scaling than with rotation and reflection.

This book is distributed as a pdf file, so you are probably reading it on a computer. I read it using a free open-source application named xreader that can display pdf files and files in several other common formats.

No matter what software you are using, this document is far too large for your software to be able to display the entire document at once. You have to scroll through the document to read it. If you using a traditional computer, you probably scroll by turning a wheel on your mouse or other pointing device. If you are reading it on a tablet computer with a touch sensitive screen, you might scroll by moving a finger up or down on the screen.

Scrolling is a translation transformation. As we read this document, when we near the bottom of the display window, we translate the text upwards. Text near the bottom of the display window is translated further up the window, allowing space for later text that was not previously visible to appear at the bottom of the window. Text that was near the top of the window also moves upwards, and when this translation moves it above the top edge of the display window it disappears from our view.

Unfortunately, people disagree on how to describe the scrolling direction. From a mathematical perspective, the text is clearly translated upwards, so some people describe this is "scrolling up", but others describe it as "scrolling down", since they scrolling to see text that was further down in the document. So for clarity, let's avoid both of those labels and refer to translating the text upwards.

Similarly, if you want to see an earlier part of the document that is above the section being displayed, you can translate the text downwards, so text near the bottom of the display move down beyond the bottom of the display window, and is lost to view, while extra text from earlier in the document moves down into view.

Note that when you translate text up or down in the display window, the letters don't change shape or size, which is consistent with translation being a transformation that preserves shape and scale. Also, when you scroll up and down, the text does not rotate, which is consistent with translation and rotation being different transformations.

The other transformation you might use when viewing this document is scaling. Most applications for viewing pdf files display the scale of the current document somewhere in one of the status bars. If the scale is currently showing as 100%, it means the software believes that the document is being displayed at its real size.

This document contains many diagrams. When I built these diagrams, I had to specify how big things are. If I told the software to create a square of side length 4cm, and my pdf viewer application shows the scale to be 100%, then if I hold a ruler up to my screen I can verify the side length of the displayed square really is 4cm.

Also, if you print a page of this document to paper, and you tell your printer *not* to do any extra scaling in the print process, and then you hold the printed page next to your display screen, the text on the page *should* be the same size as the text on the screen. However, for this to work correctly, your computer must have accurate data about the size of your computer screen, and also have accurate data about your printer. Historically, this didn't always work perfectly, though nowadays computers almost always get the screen size correct, and errors in print size are becoming less common.

But sometimes, displaying this document at a scale of 100% isn't what you want.

- You might want to zoom-in for a closer look at one of the diagrams. If you do this you are performing an enlargement, a scaling transformation with a scaling factor greater than one.
- Alternatively, if you are viewing this book on a small screen, you might find that when you view it at 100% you can't see a whole page at a time, so you might zoom out until you can. This is performing a contraction, a scaling transformation with a scaling factor between zero and one.

14.02 Playing with OpenStreetMap

Maps never involve a scale of 100%. The whole purpose of a map is to let the user see a drawing of a large area mapped onto a much smaller area. Thus maps involve scales between zero and one, but much nearer zero. For example:

- Orienteering maps are most commonly drawn at a scale of 1:10,000 so a distance of 1cm on the map corresponds to a distance of 10,000cm or 100m in the real world.
- The distance across Australia from West to East is about 4,000km. A map with a scale of 1:5,000,000 would display Australia with a width of $\frac{4{,}000}{5{,}000{,}000}km$, which is 80cm. A map of Australia would also normally show a bit of the oceans beyond the Westmost and Eastmost points of land, giving a map width of about 1m, which is a good size for a classroom wall map of Australia.

Paper maps still have their uses, but computers revolutionised mapping, by allowing the user to adjust the scale whenever they like. If you haven't used online maps, then I suggest you do so. In fact, even if you have used one before, it would be useful to do so again just to remind yourself of their features.

Open your web browser. A web browser is an application for viewing web pages. Firefox and TOR Browser are two popular open source web browsers.

OpenStreetMap is a free online street map managed by an organisation that doesn't spy on you when you use their map. The OpenStreetMap slogan includes the phrase "built by people like you", so if you find an error in the map data, you can lodge a correction. Use you browser to view and play with the OpenStreetMap web site at www.openstreetmap.org. If your primary means of transport is bicycle, you might prefer to view their alternative map at www.opencyclemap.org which highlights off-road cycleways and bike repair shops, and flags roads that cyclists regard as hazardous. From here onwards I will only refer to OpenStreetMap, but everything will behave consistently if you are using OpenCycleMap instead.

You may find it easier to understand maps when they are displaying a region that you already know well, so use the search function in OpenStreetMap to find your suburb or your street address. We are going to perform some translation and scaling transformations on this map.

These experiment will work best if your computer has a mouse with a scroll wheel. If you are using a notebook computer with a track-pad or a tablet with a touch sensitive screen, try to connect a mouse to it. If you can't, ideally switch to a computer that does have a mouse.

The following explanation assumes you are using a normal mouse. If you are a left-handed mouser and have reversed your mouse assignments, you'll need to make the appropriate adjustment when the following explanation refers to the left mouse button.

To perform a translation, place the mouse cursor at a point on the map, hold down the left mouse button and drag the map to a new location. As you do this, some parts of the map will move beyond the window boundary and disappear from sight while new bits slide into view on the other side. Look carefully at the parts of the map that remain visible throughout the translation transformation. Verify that the shape and orientation of the streets shown does not change and the distance between two points shown on the map does not change. This confirms that translation preserves lengths, angles and shapes.

To perform a scaling transformation, place the mouse cursor at a point on the map and turn the mouse wheel. Moving the wheel forward performs an enlargement, also known as zooming in, and moving it backwards performs a contraction, also known as zooming out. The point of the map directly under your mouse cursor remains fixed. When you enlarge the map, zooming in, all the other points on the map move away from your mouse cursor, and some move beyond the window border, disappearing from sight. When you contract the map, zooming out, all the other points on the map move closer to the cursor, which allows previously invisible parts of the map to move into visible area. Mathematicians call the fixed point indicated by the mouse cursor the "scaling centre" or "centre of scaling".

When you zoom out on a map, it has to show more area so it does have to display less detail. Zoom in as far as the map allows. Pick a street that is not completely straight. The more irregular the shape, the better. Place your cursor on that street and zoom out one click at a time. After several clicks you may find the map starts discarding data and simplifying the shape of the street, because it can no longer clearly show all the streets curves at that scale. Keep scrolling out. Eventually the street will disappear entirely. As you zoom out, you'll find a scale where the map stops showing residential streets but still displays major highways, but if you keep zooming out even the highways eventually disappear.

In this experiment we are interested in how scaling affects shape, so if we zoom out so far that the map starts discarding data and simplifying shapes, that's going to break the experiment. So in this experiment, start with the map zoomed in as far as it will go, so you can clearly see the shape of an individual street, and then zoom out by only one click.

Find a straight street. Use the mouse wheel to scroll out by a single click and then zoom back in by a single click. Verify that the street stays straight and it maintains the same orientation on your screen.

Find a curved street. Zoom out one click then zoom back in. Verify the street retains the same curved shape.

Place the cursor at a point where two streets intersect. Zoom out and in. Verify that the angle between two intersecting streets doesn't seem to change.

These experiment demonstrate that scaling preserves angles and shapes. "Shape" includes both straight and curved streets.

Of course, as you zoom in and out, the distance between particular intersections on the map does change. That is, the scaling transformation does not preserve distances. This isn't surprising. The purpose of scaling is to make distances change. When we zoom in on a map we make the distances between locations bigger so that we can see more detail. When we zoom out, we make the distances between locations smaller so that we can see a larger region in a single view.

When you performed these experiments, you may have noticed that the OpenStreetMap does not scale *everything* it displays. It does not scale text. It tries to ensure that the text, which displays the names of things like streets and suburbs, always remains at a size which is easy to read. If it scaled the text then as we zoomed out the text would quickly become too small to read.

A less obvious cheat is that it does not accurately show the widths of streets. Provided you followed the advice to zoom in as far as possible and then zoom out only one click on the mouse wheel, you probably didn't notice this, because the street widths are reasonably accurate at those scales. However, if you then zoom out another two or three clicks, you will probably notice that the streets are not becoming appropriately narrower on the map. At these scales, OpenStreetMap cheats by imposing a minimum display width for streets that allows it to display the street name within the borders of the street. While I called this a cheat, I don't mean that as a criticism. The purpose of a map is to clearly provide information, and cartographers — people who make maps — are masters at using cheats to increase the readability of maps. However, I'm using these maps to introduce the mathematical concept of scaling, so I do need to mention that maps do not accurately scale everything.

For our final experiment, you will need to have the same map showing in two different tabs of your browser and you will need to have both maps showing at the same scale and showing exactly the same area. There is an easy way to do this. You may have noticed that as you dragged and zoomed your map, the URL being displayed in the address bar of your browser changed. The URL contains data about the current latitude, longitude and scale of the map. That means that you can bookmark your current URL, and when you use this bookmark you will return to exactly the same view of the map.

So, zoom in as far as you can on a map of your local streets, copy the URL and paste it into a new tab. Now swap between these two tabs and verify that you see the same image in both.

In one of these tabs, find a point location somewhere near the middle of the map that you can easily remember, such as the intersection of two streets. Place the cursor at that point and use the mouse wheel to zoom out one click. Switch to the other tab displaying the same map. Place the cursor about 1cm North of the point location you remembered from the first map and again use the mouse wheel to zoom out one click. We have now performed scaling transformations on both maps. The transformations have used the same scaling factor, since both times we rotated the mouse wheel one click, but they used different scaling centres, since we positioned the cursor differently when performing the zoom.

Switch back and forward between the two tabs. How do the images compare?

You should find that the two maps show *almost* the same image. They are clearly both using the same scale, which isn't surprising give that both were zoomed in by one mouse wheel click. But one of the maps seems to be displaced slightly North relative to the other. We could move the map in either tab slightly North or South to bring it back in line with the other tab.

Now reset both tabs so that they are both showing exactly the same image at maximum zoom. Provided you haven't copied anything else since the previous step, the URL you copied is still being stored in your clipboard, so you can just paste it back into the address bar on both tabs. In one tab, place the mouse cursor back on your recognisable point and zoom out one mouse wheel click. Go the to other tab. This time, place the mouse cursor about 1cm West (left) of the recognisable point and then zoom out one mouse wheel click.

Again, switch back and forward between the two tabs. How do the images compare?

You should find the maps in the two tabs are again at the same scale, but this time one map seems to be displaced West relative to the other. We could bring the maps back into line by performing a translation East or West translation as appropriate.

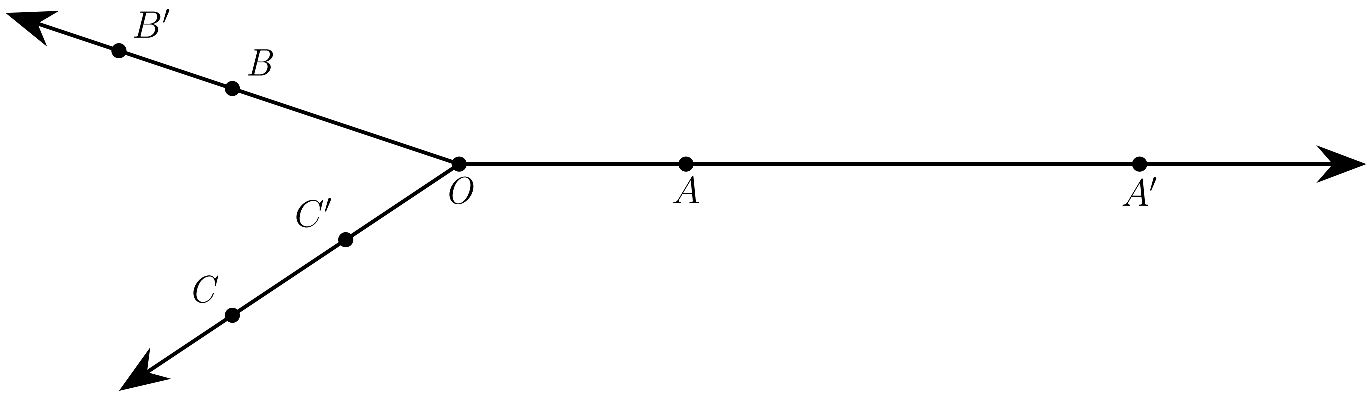
In general, it seems that if the same scaling factor is used, the difference between scaling with two different scaling centres is equivalent to a translation, and the direction of the translation seems to based on the direction of one scaling centre relative to the other. To be put it another way, a scaling transformation with a given scaling factor and scaling centre is equivalent to a transformation that involves a scaling by the same scaling factor but different scaling centre followed by an appropriate translation.

14.03 Defining the scaling transformation

A scaling transformation requires a point from which to apply the transformation, called the scaling centre, and a scaling factor, which we will require to be greater than zero. To apply a scaling transformation to a line or a shape, we can apply it to every point that makes up that line or shape, so to get started we need only define how to apply the scaling transformation to a point.

Let the scaling centre be point O and scaling factor be f, with $f > 0$. The result of applying this scaling transformation to point P, is the unique point P' on ray OP which satisfies the equation $|OP'| = f \times |OP|$.

This diagram shows three points being subject to scaling transformations. In all cases, the scaling centre is the point O, but three different scaling factors have been used. In practice, we seldom need to apply different scaling factors to points in a single diagram, but I have done so here to more clearly demonstrate the effect of changing the scaling factor.



Point A was transformed by applying a scaling factor of 3, giving point A'. To do this I drew the ray OA and placed A' at the unique point on this ray that satisfies the constraint $|OA'| = 3 \times |OA|$. It is important here to refer to ray OA rather than line OA. There is only one point on ray OA meeting the constraint $|OA'| = 3 \times |OA|$, but there are two points on line OA meeting that constraint, the other one being left of point O.

(To keep things as simple as possible, in this book, I will require the scaling factor to be positive. It is possible to give a sensible meaning to negative scaling factors, and that other point left of point O is the result of applying a scaling factor of -3. When the scaling factor is positive A and A' fall on the same side of O, so A' is on ray OA. When the scaling factor is negative, A and A' fall on the opposite sides of O.)

Point B was transformed by applying a scaling factor of 1.5, giving point B'. Thus B' falls on ray OB with $|OB'| = 1.5 \times |OB|$.

Point C was transformed by applying a scaling factor of 0.5, giving point C'. Thus C' falls on ray OC with $|OC'| = 0.5 \times |OC|$. When the scaling factor is between 0 and 1, the scaling transformation results in a point nearer the scaling centre. Thus C' falls between O and C. Thus I *could* have said that C' falls on line segment OC rather than ray OC. However, it's a good habit to always refer to a ray in this context, since that phrasing will work for all positive scaling factors.

When f denotes the scaling factor:

- If $0 < f < 1$, the scaling transformation is a contraction. A contraction moves a point nearer the scaling centre, and when applied to a shape, it makes the shape smaller.
- If $f = 1$, nothing changes. A transformation that doesn't change anything is called an "identity transformation". Other examples of identity transformations are a translation by a distance of zero and a rotation by an angle of zero.
- If $f > 1$, the scaling transformation is an enlargement. An enlargement moves a point further from the scaling centre, and when applied to a shape, it makes the shape bigger.

Less common synonyms for contraction are compression and reduction. Less common synonyms for enlargement are magnification, expansion and dilation, though some older textbooks use dilatation instead of dilation.

Conflicting jargon alert: There is a long history of mathematics textbooks from the United Kingdom referring to the scaling transformation as the "enlargement transformation", irrespective of whether the transformation is making things bigger or smaller, and some Australian textbooks followed this practice. This was always a poor choice of label since it conflicts with the common meaning of the word "enlargement", which is to make something bigger. Unfortunately the authors of the Australian Mathematics Curriculum have chosen to refer to this transformation as enlargement rather than scaling.

What effect does the scaling transformation have on the scaling centre? For example, what happens to point O if we apply a scaling transformation to it, using a scaling factor of 3, but using O as the scaling centre?

Applying a scaling factor of 3 to point A gives point A' where $|OA'| = 3 \times |OA|$. To be consistent, applying a scaling factor of 3 to O should give point O' where $|OO'| = 3 \times |OO|$. But $|OO|$ denotes the distance between point O and itself, which is zero. Hence $|OO'|$ must also be zero, meaning O' must coincide with O.

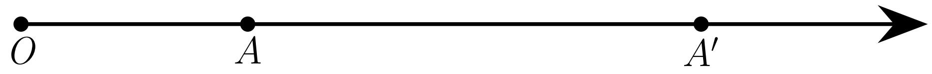
The same outcome results for any other positive scaling factor you choose. When applying a scaling transformation, the scaling centre remains fixed no matter what scaling factor is used. This is consistent with what we found when playing with OpenStreetMap. When zooming in or out by turning the scroll wheel, the point directly under the cursor did not move.

Other transformations also have fixed points that are unaltered by the transformation. When applying a reflection, every point on the reflection line is unchanged by the reflection and when applying a rotation, the rotation centre is unchanged by the rotation.

Most transformations have an inverse transformation that will cancel out the effect of the transformation, returning all points to their original position. The inverse of a translation is a translation by the same distance but in the opposite direction. The inverse of a clockwise rotation is an anticlockwise rotation by the same angle around the same rotation centre. In two dimensional geometry, the inverse of a reflection in a line is another reflection through the same line.

What is the inverse of a scaling transformation?

The inverse of a scaling transformation with scaling centre O and scaling factor f is a scaling transformation with the same scaling centre and scaling factor $\frac{1}{f}$.



To see this, note that the original transformation moves A to A', a point on ray OA with $|OA'| = f \times |OA|$. This means O, A and A' are collinear. When we apply the second transformation to A', we move it along ray OA', which is identical to ray OA. We move A' to the unique point A'' on this ray that satisfies $OA'' = \frac{1}{f} \times |OA'| = \frac{1}{f} \times f \times |OA| = |OA|$, which means A'' coincides with A. That is, we have reversed the original transformation.

The diagram above assumed $f > 1$. Naturally, you should also produce a diagram where $0 < f < 1$ and verify the argument is still valid for that diagram.

I mentioned early we will require the scaling factor to be greater than zero. What would happen to the points A, B and C in the earlier diagram if we applied a scaling factor of zero?

Applying a scaling factor of f to point A gives the point A' where $|OA'| = f \times |OA|$. Thus applying a scaling factor of zero gives $|OA'| = 0 \times |OA| = 0$. That is, A' must coincide with O. The same argument applies to any other point. That is, a scaling factor of zero causes all points to transform to the scaling centre. While this may be an interesting piece of trivia, a transformation that moves all points to a single location serves no practical purpose, so I won't consider it further.

14.04 Scaling in plane geometry

The real world has three dimensions and most real world objects are solids. By contrast, the geometry in this book is mostly two dimensional, meaning it lives within a plane, and we call it plane geometry.

Many geometrical ideas are introduced in the context of plane geometry, but when a student has mastered them in that context, they ideas can be expanded to more complicated scenarios.

For example, this book is about area, which is essentially a property of plane geometry. We have found how to calculate the areas of things like triangles and rectangles, which are plane figures.

But the concept of the area of a plane figure can be complicated into the concept of the surface area of a three dimensional solid. This could help answer questions like: How much paint would I need to paint the outside surface of this solid shape. For simple solids, such as cubes and rectangular prisms, this process quickly collapses back to a plane geometry problem, since we just add up the areas of the solid's faces, which are all squares or rectangles, plane figures. We encountered a problem like this earlier in this book, though we were thinking about painting some inside surfaces of a solid space rather than the outside. We considered painting the four walls of a room, and a room is a three dimensional shape, typically a rectangular prism. But then we calculated the area of each wall separately, and each wall was a plane figure. Some walls were rectangles, and others were rectangles from which other rectangles had been removed to make a door or a window. So while the room was three dimensional, the area calculations were problems in only two dimensions.

Similarly, some curved surfaces can be flattened to produce plane figures. Most tin cans are cylinders and their curved surface be flattened out to a rectangle. The two ends of the can are circles, which are also plane figures. We haven't yet studied how to find the area of a circle, but it is problem in plane geometry, not solid geometry.

When we come to the surface area of a sphere, this approach fails. We can't flatten out a sphere into a plane figure without stretching it, and stretching changes the area. The techniques needed to find a sphere's surface area have similarities to the methods needed to find its volume, so they are not encountered in books covering only plane geometry. We need a solid geometry textbook for that problem.

Transformations are also usually introduced in the context of plane geometry, but in later years of study they may be complicated in a way that allows can also have interpretations in solid geometry.

For example, in plane geometry we can rotate a shape such as a square around a point, called the rotation centre. When we do this, every point in the square stays within the plane. That is, this form of rotation "plays nicely" with plane geometry, in the sense that it can never move the square outside the plane.

But there is a more complex form of rotation that rotates around a line rather than a point, and this form does not "play nicely". Imagine drawing a square on the wall of a room. The wall is a rectangle, a plane figure. The base of this rectangle is the line where the line joins the floor. That line is part of the plane that the wall belongs to. But if we rotate that square on the wall by 90° around the line that is the base on the rectangular wall, the square can be transformed to a position on the floor of the room. That is, this form of rotation can break out of the plane that contained both the square and line around which we performed the rotation.

If we draw a square on a plane, and apply a scaling transformation using a scaling centre within the plane, the square stays within the plane when transformed. This is because the scaling transformation moves a point along a ray that starts at the scaling centre. When we do a scaling transformation on a square in plane geometry, the square and the scaling centre both live on the plane we are using. Every ray from the scaling centre to a point on the square also stays within that plane, so points on the square can never be transferred off the plane.

The scaling transformation does also exist in solid geometry. We can scale a cube to produce a bigger or smaller cube.

We can also scale a plane figure such as a square using a scaling centre that is not in that same plane. Take a piece of stiff wire and bend it into the shape of a square. Hold the wire square in a horizontal orientation while standing below a light bulb. The the light from the bulb projects a square shadow onto the floor. The light bulb is acting as the scaling centre, the wire frame is the original object and the shadow on the floor is the result of the scaling transformation. The shadow will be bigger than the wire square, so the scaling transformation was an enlargement. This is not a perfect analogy, since the wire has thickness, so our wire square is not a true two dimensional object, but it hopefully demonstrates that a scaling transformation can move a square to a different plane. But to make this happen we had to use a scaling centre in a different plane to the wire square. The light bulb was above the plane containing the horizontal wire square.

For the purpose of this book, the important point is that the scaling transformation "plays nicely" within the confines of plane geometry. If we draw a shape such as a square on a plane, and use a scaling centre that is a point on that same plane, a scaling transformation will never move the square off the plane into three dimensional space. Our only practical constraint is that while planes extend indefinitely in their two dimensions, we are usually doing our drawings on a piece of paper, which represents just a small section of a plane, so when doing enlargement transformations it is possible to transform our square beyond the edge of the page. The transformed square is still within the same plane. It's just not on the piece of paper that we can safely draw on. (You have been warned. If you draw the transformed square on your tabletop because it transformed beyond the edge of your page, don't blame me!)

14.05 Scaling consistently scales all distances

It is possible to prove all the results in this chapter from the basic postulates of Euclidean geometry, but that route is long and torturous, and so it is not usually attempted in Australian school geometry.

Instead, at school level, it is standard practice to pick one or more of the more believable theorems and simply present them as postulates, which means we will not prove them. For this book, the theorem that I'm going to convert into a postulate is the following, which essentially says that a scaling transformation effects all distances in a consistent manner.

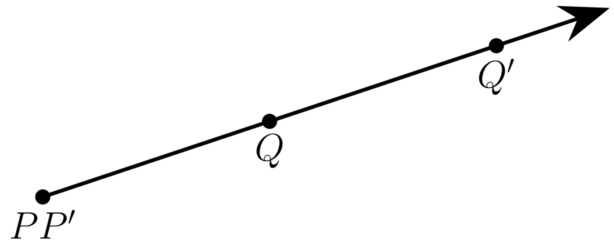
> **The Consistent Scaling Postulate**: No matter what scaling centre is used, a scaling transformation will multiply the distance between any two points by the scaling factor.

Well, that was so succinct that it can be a little difficult to understand what it's saying. Say we apply a scaling transformation using a scaling factor of f. Let P and Q be any two points subject to this transformation, with P' and Q' being the resulting points after applying the transformation. The postulate claims that, no matter what scaling centre is chosen, $|P'Q'| = f \times |PQ|$.

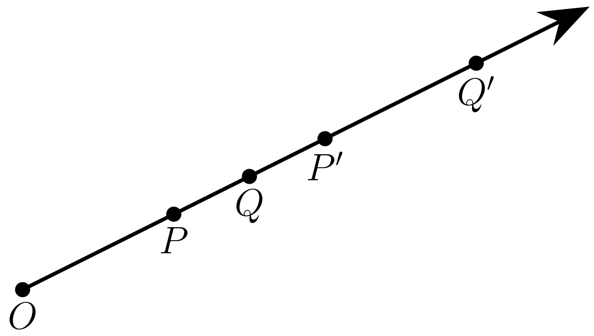
Since I am treating this as a postulate, we won't prove it, but I should still make some attempt to convince you that it is plausible to accept this claim as true. If you *are* already convinced that this postulate is true, then you can jump ahead to the next section. The remaining sections of this chapter use this postulate to derive a series of increasingly powerful results.

While it is difficult to prove this result *in general*, there a few *special cases* where it is simple to prove the postulate true, so we will start with those special cases. The diagrams for these next few proofs are going to assume the scaling factor $f > 1$, and you should also produce corresponding diagrams where $0 < f < 1$ and verify that the proofs are still valid for those cases.

The simplest special case is where one of the two points P and Q happens to be the scaling centre. Say P is the scaling centre. It will be unmoved by the transformation, so P and P' coincide. Using the earlier definition of how scaling works for points, we place Q' at the unique point on ray PQ for which $|P'Q'| = f \times |PQ|$, which is exactly what the postulate claims. Hence the postulate is true in this special case.



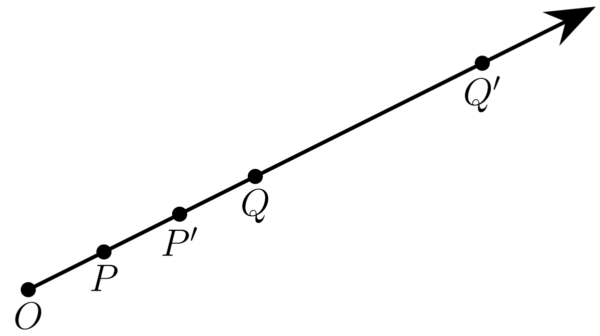
Another special case is where the scaling centre O does not coincide with P or Q, but O, P and Q happen to be collinear. First, let's look at the case where P falls between O and Q. Here, P' and Q' are placed on the rays OP and OQ respectively, but these are in fact the same ray since O, P and Q are collinear. They are placed so that $|OP'| = f \times |OP|$ and $|OQ'| = f \times |OQ|$. This gives two different possible orderings of the points, so I have shown a diagram for each. You can verify that in both cases:



$$|OQ| = |OP| + |PQ|$$
$$|OQ'| = |OP'| + |P'Q'|$$

Hence:

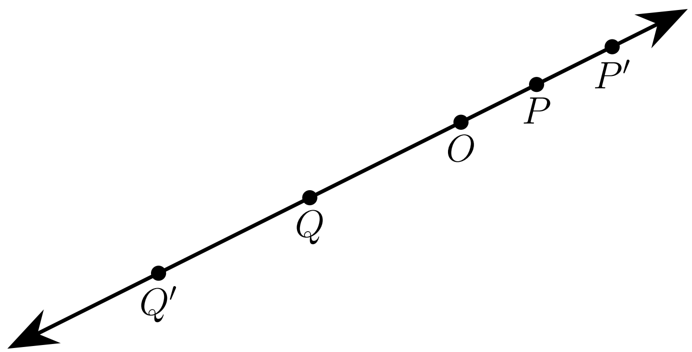
$$\begin{aligned}|P'Q'| &= |OQ'| - |OP'| \\ &= f \times |OQ| - f \times |OP| \\ &= f \times (|OQ| - |OP|) \\ &= f \times |PQ|\end{aligned}$$



For the special case where O, P and Q are collinear, this proves the postulate true when P falls between O and Q. If Q falls between O and P, the proof will be as above, but with all occurrences of P and Q swapped, which also means swapping P' and Q'.

Investigate what happens if O falls between P and Q.

While P' is placed on ray OP and Q' is placed on ray OQ, we know O, P and Q are collinear, so all five points are collinear. As in the previous case, $|OP'| = f \times |OP|$ and $|OQ'| = f \times |OQ|$. But the changed ordering of the points makes the final steps of this proof simpler than the previous case.



$$\begin{aligned} |P'Q'| &= |P'O| + |OQ'| \\ &= f \times |PO| + f \times |OQ| \\ &= f \times (|PO| + |OQ|) \\ &= f \times |PQ| \end{aligned}$$

Thus we have proved the postulate true for all scenarios where P and Q are collinear with the scaling centre. We now have to think about the far more common case where they are not.

Another useful thing you can do to convince yourself that the postulate seems reasonable is to draw some examples, measure the distances and see whether they comply with the postulate. So let's try that.

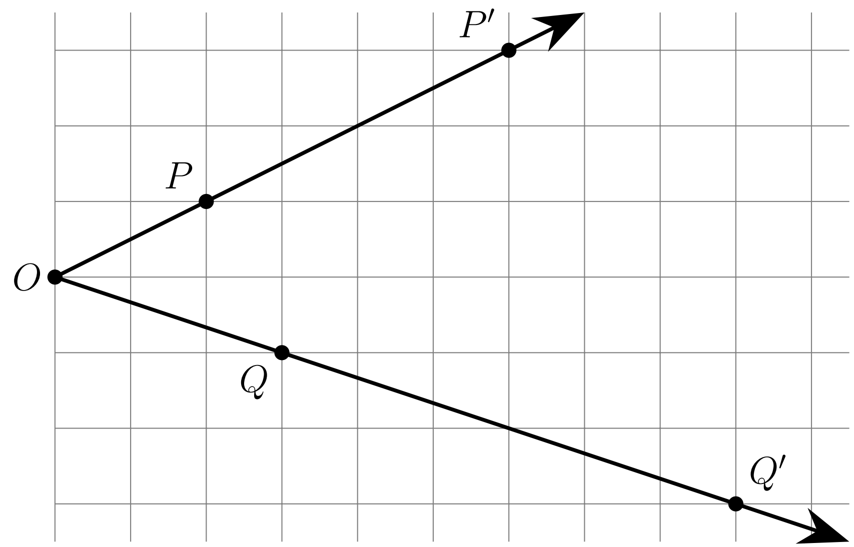
Ideally, do this experiment on graph paper that provides you with a square grid. If you place the scaling centre and the points P and Q at the intersection of grid lines, and then use a scaling factor which is a positive integer, you will find that when you locate the transformed points P' and Q', they also happen to land at the intersection of grid lines. This will help you keep your drawing accurate. Of course, if you make the scaling factor too big you might transform the points beyond the edge of your page, so perhaps try scaling factors of 2, 3 and 4 on various different points.

If we were constructing a formal proof, then only considering positive integer scaling factors would be a problem, since the proof might only be true for positive integer scaling factors and the property we are trying to prove might fail elsewhere. But this isn't a formal proof. The property we're testing here is a postulate, not a theorem, so we're just trying to convince ourselves that it seems plausible rather than trying to prove it. Thus testing it for particular positive integer scaling factors is a perfectly reasonable thing to try.

Here is my attempt at this experiment. My diagram here uses a scaling factor of 3. My grid lines are 1cm apart, though they may appear differently to you if you are not viewing this file at 100% scale.

My ruler tells me that $|PQ|$ is between 22mm and 23mm while $|P'Q'|$ is just a little over 67mm. These measurements are consistent with the postulate, which claims $|P'Q'| = 3 \times |PQ|$.

My diagram shows only one pair of points and a single scaling factor. Try using different pairs of points and different scaling factors. If you construct your diagram correctly, you should find your results are always consistent with the postulate, which hopefully convinces you this postulate is plausible.



Here is one final argument in favour of the postulate. Recall the last experiment we performed with OpenStreetMap. We compared the results of applying the same scaling factor, arising from a single click rotation of the mouse wheel, but using two different scaling factors. It seemed that the difference between the two results was a translation, with the translation being in the direction from one of the scaling centres to the other. We didn't formally prove that, but it certainly felt like a plausible assumption. Let's apply that assumption to this diagram.

The diagram above shows P and Q being transformed by a scaling factor of 3 using O as the scaling centre, giving the points P' and Q'. If the assumption is true, I should be able to get the same result by performing two successive transformations.

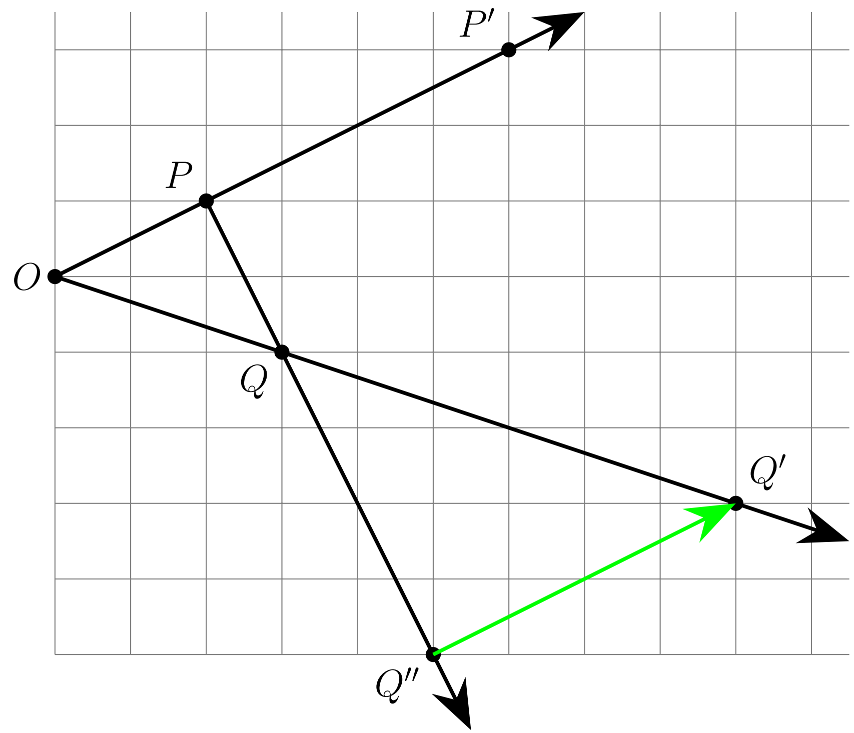
1. A scaling transformation using a scaling factor of 3 but using P as the scaling centre, then
2. A translation transformation in the direction from O to P.

The next diagram shows the result.

When we use P as the scaling centre, the scaling leaves P unmoved and Q is transformed along ray PQ to Q''. This scenario corresponds to the first special case given above, where one of the points being transformed *is* the scaling centre, and we did formally prove the postulate works in that special case. That is, given my diagram uses a scaling factor of 3, we know $|PQ''| = 3 \times |PQ|$.

Then we perform a translation in the direction of OP, choosing the translation distance that brings P into line with P'. When we perform that same translation on Q'' we find it moves to Q'. That is, the green arrow showing Q'' transformed to Q' is parallel to and the same length as line segment PP'.

Translations preserve distance so $|P'Q'| = |PQ''|$. Combining this with $|PQ''| = 3 \times |PQ|$ gives $|P'Q'| = 3 \times |PQ|$, just as the postulate predicted.



We will now use this postulate to prove various properties that are preserved by scaling transformations.

14.06 Scaling preserves linearity

Scaling a point is not very interesting. More commonly we want to scale line segments or shapes. In this section we will look at scaling a line.

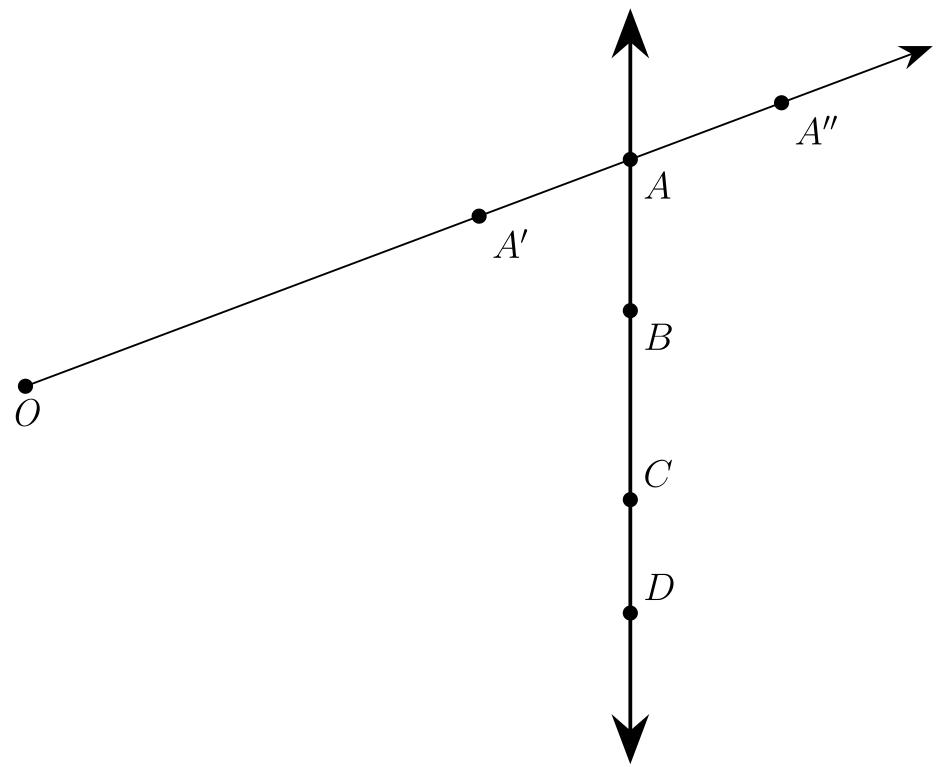
From the previous section, we know how to apply a scaling transformation to a point. A line is made up of points, so conceptually, to apply a scaling transformation to a line, we can just apply the scaling transformation to every point on the line. I said "conceptually" because there is a practical difficulty. A line contains an infinite number of points, so applying the scaling transformation to every point on the line takes an infinite amount of time. We need to find a smarter method.

When playing with OpenStreetMap we observed that when we zoomed out on a straight street, it seemed to stay straight. Let's do a paper and pencil investigation of that idea. We will apply a scaling transformation to four points on a straight line, to test whether the resulting four points also seem to be collinear. Actually, we'll do that twice with two different scaling factors, one enlargement and one contraction.

Pretend your teacher gives you a piece of paper showing a point to use as a scaling centre and a sloping line, and they asked you to perform a scaling transformation to that sloping line. You can rotate the page to make the sloping line run vertically and tape the page to a wall or whiteboard in that orientation. You can then carry out the required scaling transformation on the now vertical line. When finished you can take the page off the wall and hand it back to your teacher. If they choose to align the page with its edges vertical and horizontal, so the line is again sloping, that isn't a problem. Your construction remains valid in any orientation.

This is a general property of geometrical constructions and proofs. We can always choose to make our drawing in any orientation we like, so we can choose the orientation that makes things easiest for us. We'll use that approach here. Rather than dealing with a sloping line, we will choose to orient this task so that the line we use runs vertically down the page. Of course, if you don't find this argument convincing you can still choose to use a sloping line in your diagram and verify that you get a consistent result. I'm going to use a vertical line in my diagram, because the code needed to produce these diagrams is complicated, and making the line vertical reduces the complication.

Take a piece of paper. Draw a vertical line across it from top to bottom. Since it's a line rather than a line segment, I've placed arrows on both ends of my line to indicate that it extends indefinitely in both directions. Choose a point O left of the line to be the scaling centre. Choose 4 points on the line and label them A to D. Don't put any of them close to the edge of the page, because one of our scaling transformations will be an enlargement, and we don't want that transformation to move any of the points beyond the edge of the page.



Choose two different scaling factors, one between zero and one, and the other greater than one. I'm going to use 0.75 and 1.25. You can choose other numbers, but don't make the larger number so large that your transformation moves points beyond the page. Label the four points resulting from the smaller scaling factor as A' to D' and those resulting from the larger scaling factor as A'' to D''.

So far, my diagram here has only applied the scaling transformation to point A. That is, I have drawn ray OA and I placed points A' and A'' on that ray so that $|OA'| = 0.75 \times |OA|$ and $|OA''| = 1.25 \times |OA|$.

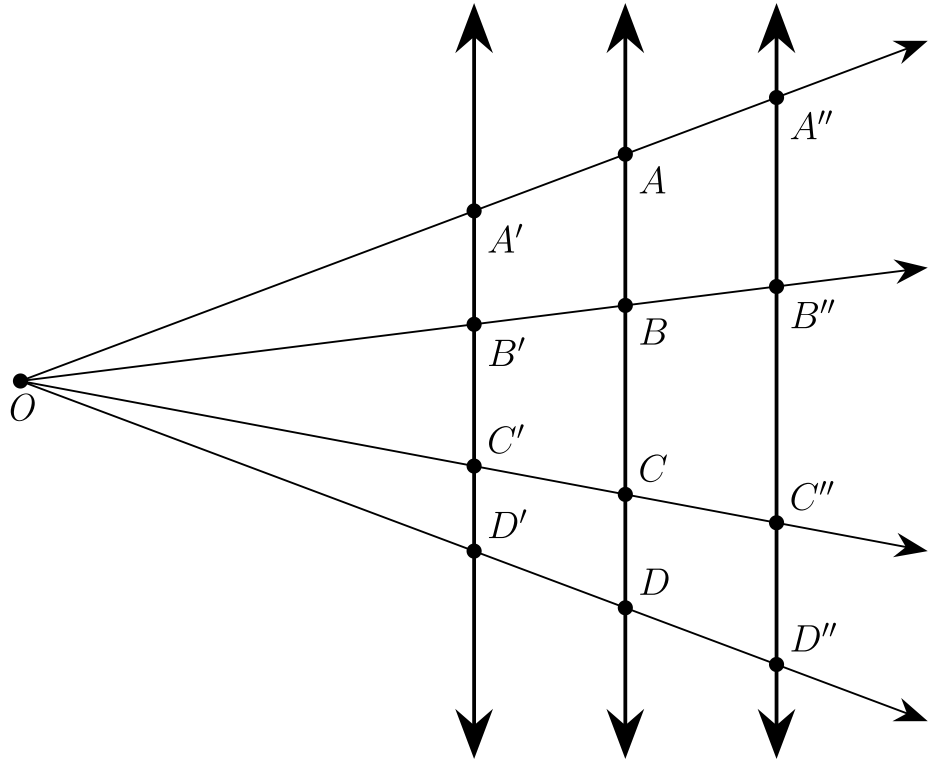
Complete both scaling transformations for all four points. What do you notice about the resulting points?

Here is the completed diagram. There are two important things to notice about the result.

The first is that for both transformations, the resulting four points appear to be collinear. That is, it seems that A', B', C' and D' all fall on a single line, and A'', B'', C'' and D'' fall on another line.

The second is that both transformed lines seem vertical, just like the original line. Of course, if you imagine rotating your screen so that the original line is slightly off vertical, the two transformed lines rotate by the same amount, so actually the important conclusion here is that the two transformed lines seem *parallel* to the original line, which means the scaling transformation preserves the *orientation* of lines.

We'll prove this second property in the next section. In this section we'll just deal with the first property, which is that the points resulting from each scaling transformation seem to be collinear.



In your diagram, the points might not appear *exactly* collinear. Don't panic! Your diagram is subject to the limited accuracy of your ruler. For example, when you measured the distance $|OA|$, you could probably only measure it to the nearest millimetre, so your calculations of the positions of A' and A'' will be slightly inaccurate. Realistically, the best that we can hope for in a diagram constructed with ruler and pencil is that the points A', B', C' and D' seem very close to being on a single line, so close that we suspect that a more accurate diagram *would* show them to be collinear. In my diagram shown here the positions were calculated by my computer, which can do much better than millimetre accuracy, and these four points do seem to be collinear, even if you zoom in on them.

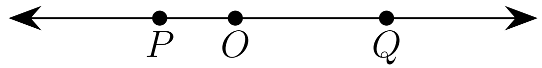
So, we suspect that applying a scaling transformation to a line produces another line. The brief way to say this is: scaling preserves linearity. A more casual way to express this is that scaling transformations preserve straightness.

Earlier I said that we know how to apply a scaling transformation to a point, so to apply it to a line we could theoretically apply the transformation to every point on the line, the problem being that we would have to transform an infinite number of points. If we can prove scaling preserves linearity then we have a much easier method to transform the line. We simply apply the scaling transformation to any two points P and Q on the line, giving the points P' and Q'. The result of applying the scaling transformation to line PQ is the line $P'Q'$, so just drawn the unique line that goes through P' and Q' and we're done!

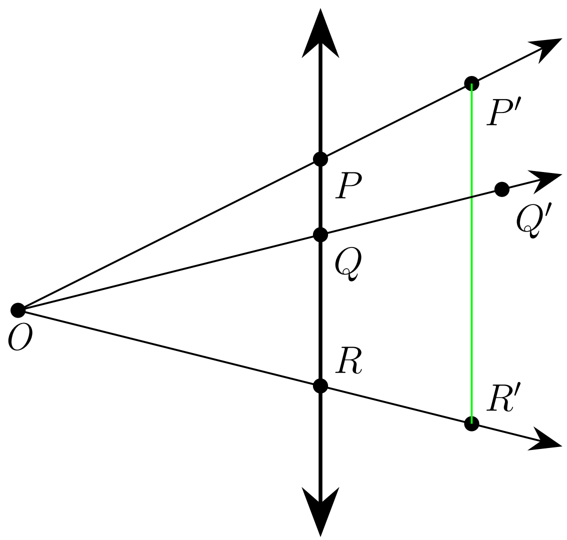
It's time to stop experimenting and formally prove that scaling preserves linearity. Before considering with the common scenario, we will deal with two easy special cases.

> What happens if we apply a scaling transformation to a line that passes through the scaling centre?

Consider a line passing through a scaling centre O. I've chosen to orient the diagram so that the line is horizontal. Let P be any point left of O. Any scaling transformation applied to P will transform it to a point on ray OP, which means it stays on the horizontal line. Similarly, let Q be any point on the line right of O. Any scaling transformation applied to Q will transform it to a point on ray OQ, which means it too stays on the horizontal line. Also, the scaling centre in unmoved by any scaling transformation. Thus any scaling transformation applied to a line that passes through the scaling centre merely reproduces the same line.



I mentioned earlier that if the scaling factor is one, then the scaling transformation doesn't change anything. This is the other easy special case. Applying a scaling factor of one to a line produces the same line. This is true whether the scaling centre falls on the line or not.



With the special cases out of the way, now consider applying a scaling transformation to a line that does not contain the scaling centre, and uses a positive scaling factor other than the trivial value of one. My diagram will assume the scaling factor $f > 1$, and you should also try making a comparable diagram where $0 < f < 1$ to verify the proof remains valid for those cases.

I'll choose to orient my diagram so that the line is vertical with the scaling centre to the left, as shown. We will apply a scaling transformation to the line, using the scaling factor f. Let P, Q and R be any three points on the line, labelled in that order. That is, Q falls between P and R. Let the results of transforming P, Q and R be P', Q' and R' respectively.

We hope to prove that P', Q' and R' must be collinear, but we haven't proved that yet, so I have deliberately drawn them as not being collinear. That is, Q' does not lie on the green line seqment $P'R'$.

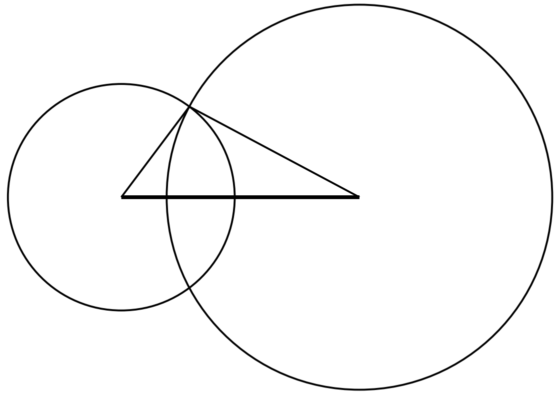
The consistent scaling postulate tells us that $|P'Q'| = f \times |PQ|$, $|Q'R'| = f \times |QR|$ and $|P'R'| = f \times |PR|$.

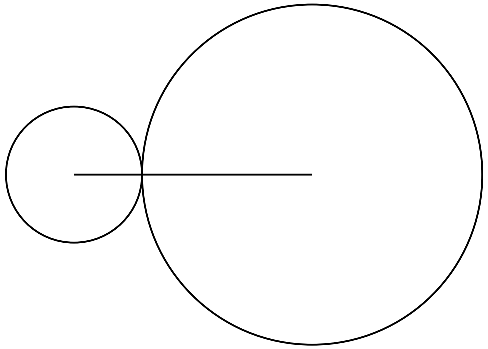
Since P, Q and R are collinear, $|PR| = |PQ| + |QR|$.

Combinging these results gives:

$$|P'R'| = f \times |PR| = f \times \{|PQ| + |QR|\} = f \times |PQ| + f \times |QR| = |P'Q'| + |Q'R|$$

If you can see how to complete the proof from here, do so. If not, think about the process we use to construct a triangle with side lengths of a, b and c, where c is the longest side. We draw a horizontal line segment of length c to be the base of the triangle, the thick line shown at right. Then we draw a circle of radius a centred on one end of the base and a circle of radius b centred on the other end. Provided $a + b > c$ these circles intersect at two points. We can choose the higher of these two points of intersection to be the apex of the triangle. Connect the apex to both ends of the base and we have a triangle with the required side lengths.

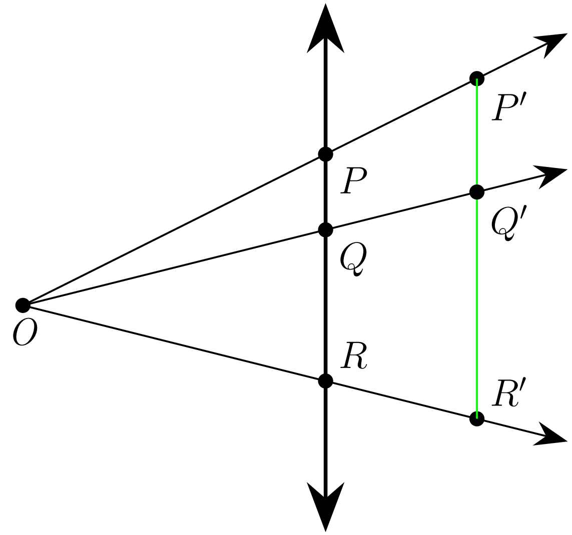




But what happens when $a + b = c$? That is, what happens if we try construct a triangle with side lengths a, b and $a + b$. We draw a horizontal base of length $a + b$. We construct the two circles on the end points, one with radius a, the other with radius b. We find the circles intersect in just one point that happens to fall on the base. That is, we fail to construct a triangle because the point that we hoped to use as the apex falls on the base, meaning our triangle has collapsed into being just a line segment.

With that insight, can you now complete the proof?

If $|P'R'| = |P'Q'| + |Q'R|$ then P', Q' and R' cannot be the vertices of a triangle. The only way that equation can be true is if those three points are collinear, with Q' falling between the other two points. That is, the earlier diagram which showed Q' falling to the right of the green line is wrong. Q' should actually be placed at the intersection of ray OQ and the green line segment $P'R'$, as shown in the revised diagram here.



Recall that P, Q and R represented *any* three points on the original line, the only constraint being that they occur in that order. This means that *all* points on the original line PR must transform to points on the new line $P'R'$.

From the above, we can conclude: The result of applying a scaling transformation to a line is a line.

Note that the phrasing I used covers both the earlier special cases, where transforming the line produces the same line, as well as the more common scenario where transforming the line gives a different line.

Can you see how to draw a comparable conclusion for line segments and rays?

The insight required to complete this is to note that points on a line must maintain their same order when subject to a scaling transformation.

P' falls on ray OP, Q' falls on ray OQ and R' falls on ray OR. These three rays intersect at O, and rays are straight, so that cannot ever intersect anywhere else. Therefore, since Q falls between P and R, Q' must fall between P' and R'. But Q can represent *any* point between P and R, so *every* points between P and R transforms to a point between P' and R', meaning line segment PR transforms to line segment $P'R'$.

That is, applying a scaling transformation to a line segment produces a line segment.

Similarly, line segment PQ transforms to line segment $P'Q'$.

Now think of P and Q as fixed, with R representing any point on ray PQ beyond point Q. Since the points maintain their order when transformed, R must transform to a point on ray $P'Q'$ beyond point Q'. Combining this with the result of the previous paragraph, ray PQ must transform to ray $P'Q'$.

That is, applying a scaling transformation to a ray produces a ray.

Since they are closely related, I'll state the three results of this section as a single theorem.

Theorem 14.1: When applying a scaling transformation, a line transforms to a line, a line segment transforms to a line segment and a ray transforms to a ray.

A immediate consequence of this theorem is the following.

Theorem 14.2: Under scaling, a polygon transforms to a polygon with the same number of sides.

This follows because the sides of a polygon are line segments, so under a scaling transformation they remain line segments. Thus to apply a scaling transformation to a polygon, we simply apply the transformation to its vertices, then use line segments to join each pair of transformed vertices that correspond to vertices which were joined in the original polygon.

14.07 Scaling preserves orientation

In an experiment in the previous section, we noted that transforming a line seemed to produce a line parallel to the original line. That is, scaling seems to preserve the orientation of a line. We will now prove that result.

The trivial special cases mentioned in the previous section, where the scaling centre falls on the line or the scaling factor is one, are not of interest, since the result of the transformation is the original line. Thus in this section our proof need only consider the common scenario where the line and its transformation are two *different* lines.

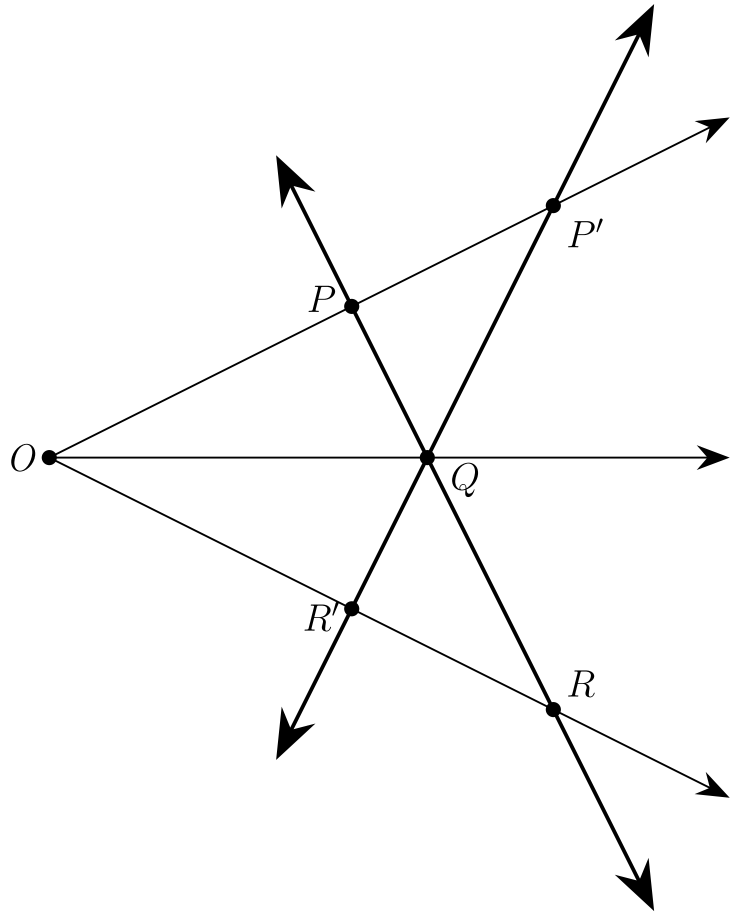
That is, we will deal with a line and a scaling centre which is not on the line. A line and a point not on that line uniquely determine a plane. In this proof we are doing geometry within that plane. As mentioned earlier the line resulting from the scaling transformation also falls entirely within that same plane.

Two different lines within a plane are either parallel or intersect at a single point. Can the original line and the line resulting from applying a scaling transformation intersect at a single point Q?

Hint: It's useful to try to draw a diagram that meets the constraints and then think about what various points on the original line imply about the size of the scaling factor. But it's also possible to build a valid proof just by thinking about the possible locations for point Q', the result of applying the scaling transformation to Q.

If you try to draw a scenario where the original line and its transformation intersect at a single point, you will find your diagram contains:

- Points like P and its transformation P' where $|OP'| > |OP|$, implying the scaling factor $f > 1$.
- A point like Q which is its own transformation, implying $f = 1$.
- Points like R and its transformation R' where $|OR'| < |OR|$, implying $0 < f < 1$.



A scaling transformation uses a single scaling factor, so it can't take different values like this. That is, we cannot produce a consistent diagram that has the line and its transformation intersecting at a single point.

For an alternative approach, let's follow the final hint in the question. In this alternative proof we only care about the alleged single point of intersection Q and its transformation Q'. We won't need to consider points like P, R and their transforms, so this approach doesn't require the diagram shown here.

- Q' must fall somewhere on ray OQ, because that's how scaling transformations work.
- It's location meets the constraint $|OQ'| = f \times |OQ|$.
- It can't coincide with O since the scaling factor must be greater than zero.
- It can't coincide with Q, since that would mean $f = 1$, which gives the trivial special case where the transformed line coincides with the original line rather than having just one point of intersection.

- So Q' must be a point on ray OQ that doesn't coincide with either O or Q. But Q was defined as the point of intersection of the line and its transformation, so now we know that Q and Q' are two distinct points on ray OQQ' that fall on the transformed line, meaning the transformed line must be line OQQ'. But the scaling centre O is unmoved by the transformation, so if it on the transformed line it must also be on the original line. We've already shown that in the special case where the scaling centre falls on a line, that line and its transformation are the same line rather than having just a single point of intersection.

Thus the line and its transformation cannot intersect at a single point, because if they intersect anywhere they are forced to coincide rather than just having a single point of intersection.

As mentioned earlier, if two distinct lines are contained on the same plane, they either intersect at a single point or they are parallel. We've shown that a line and its transformation under scaling cannot intersect at just one single point, so they must be parallel.

Another way to say this is that scaling a line preserves its orientation, a statement that also covers the trivial special cases mentioned earlier. Here is the resulting theorem.

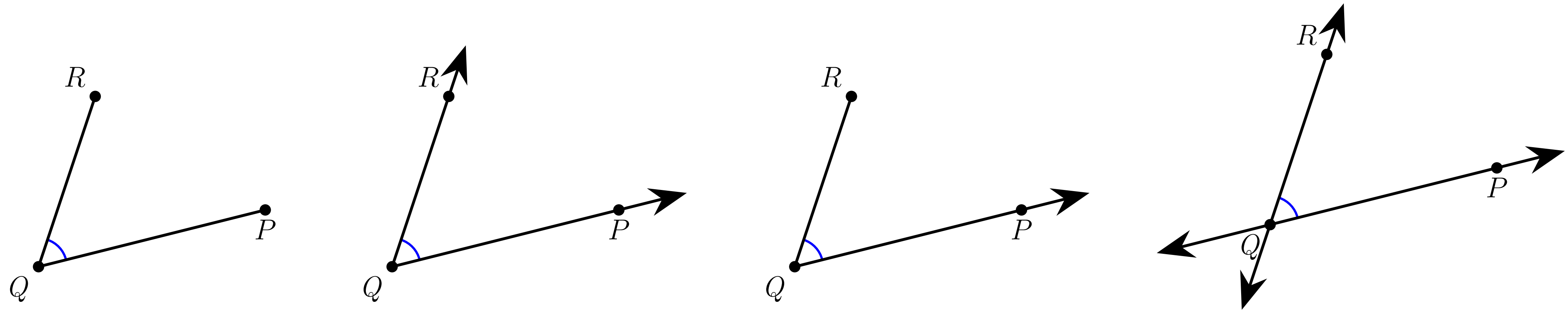
Theorem 14.3: Scaling a line preserves its orientation. That is, applying a scaling transformation to a line produces either the same line, or a new line parallel to the original line. The first outcome only occurs in the trivial special case where the scaling factor is one or in the special case where the line contains the scaling centre.

14.08 Scaling preserves angles

Whenever an angle is created by the intersection of lines, rays or line segments, applying a scaling transformation preserves the size of the angle. This is conceptually easy to prove using the property that scaling preserves orientation, but the process is made somewhat tedious but the need to consider several different possible diagrams and a few special cases.

We'll start by proving the most trivial special case and work our way up to more common cases.

Consider $\angle PQR$. The size of the angle does not depend on the lengths $|QP|$ and $|QR|$. It doesn't matter whether QP and QR are line segments or rays, as in the first two diagrams below, or even a mix of both, as in the third diagram. They could even be lines, as in the fourth diagram, which introduces more angles but doesn't change the size of $\angle PQR$. Thus we can freely switch between lines, rays and line segments during the proof if that is convenient.



We will apply a scaling transformation to $\angle PQR$ shown in the above four diagrams. To do this, we can apply the transformation, to P, Q and R, giving the points P', Q' and R' respectively. $Q'P'$ and $Q'R'$ can then be drawn, and it makes no difference whether we show them as line segments, rays or lines. Whatever choice we make, our aim is to prove $\angle PQR = \angle P'Q'R'$.

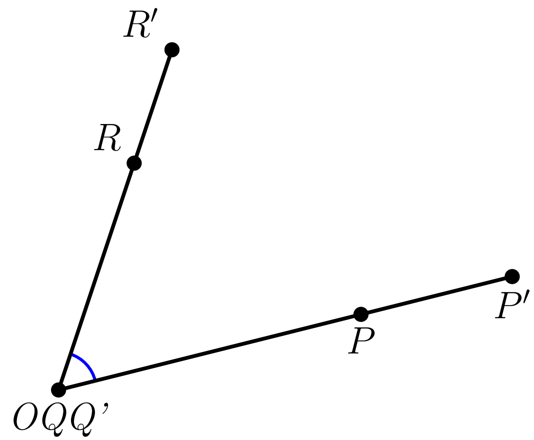
One easy special case is the trivial identity transformation where the scaling factor $f = 1$. This leaves the entire diagram unchanged, so clearly the size of $\angle PQR$ unchanged. That's all we need to say about that special case, so for the rest of this section, we will assume $f > 0$, but $f \neq 1$. My diagrams will show a case where $f > 1$, but the arguments will still work when $0 < f < 1$.

Another easy special case is where Q coincides with the scaling centre O. Draw a diagram of this case, and explain why $\angle PQR = \angle P'Q'R'$

The scaling centre is unmoved by a scaling transformation, so when Q coincides with O, they also both coincide with Q'.

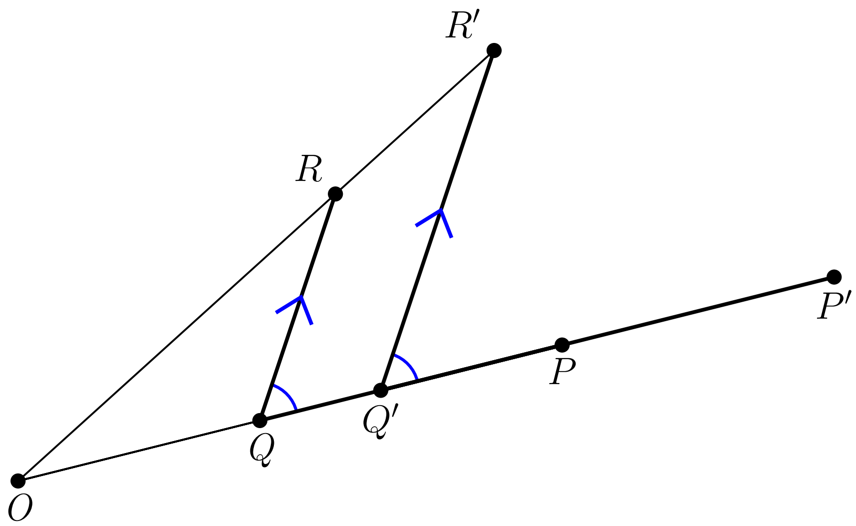
P' and R' fall on rays OP and OR respectively. This diagram shows an enlargement, where the transformation moves points away from O. A contraction moves them closer.

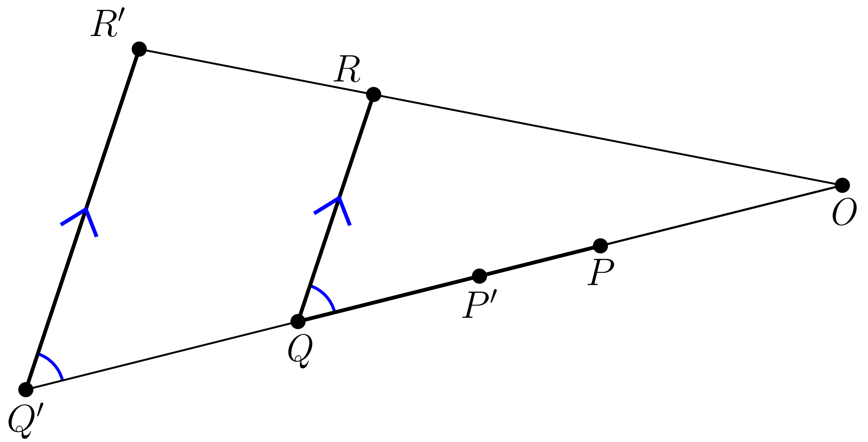
Either way $\angle PQR = \angle P'Q'R'$ because they are the same angle.



The next slightly less special case to consider is where the scaling centre does not coincide with Q, but is collinear with one arm of the angle. I'll consider the case where it falls on line PQ. A similar process can cope with the cases where it instead falls on line QR.

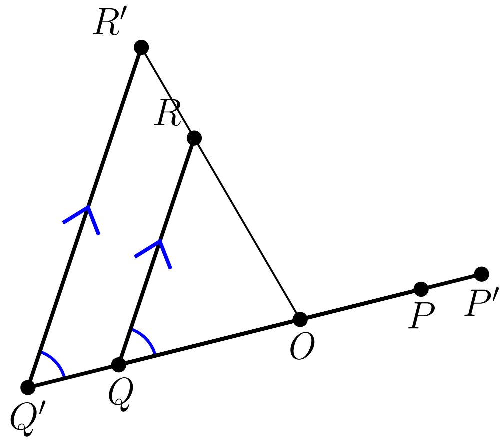
If O falls on line PQ, it can coincide with P or Q, fall between them, fall beyond P or fall beyond Q. We've already considered the case where it coincides with Q, so that leaves four cases to consider, shown here. While the diagrams differ, exactly the same argument applies to all four diagrams.

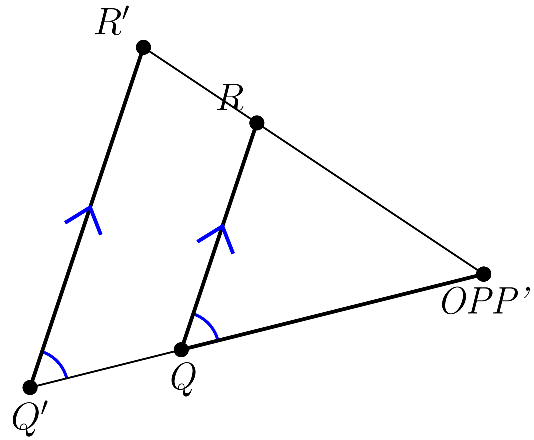




We use the property proved in the previous section: scaling preserves orientation. This means QR and $Q'R'$ are parallel.

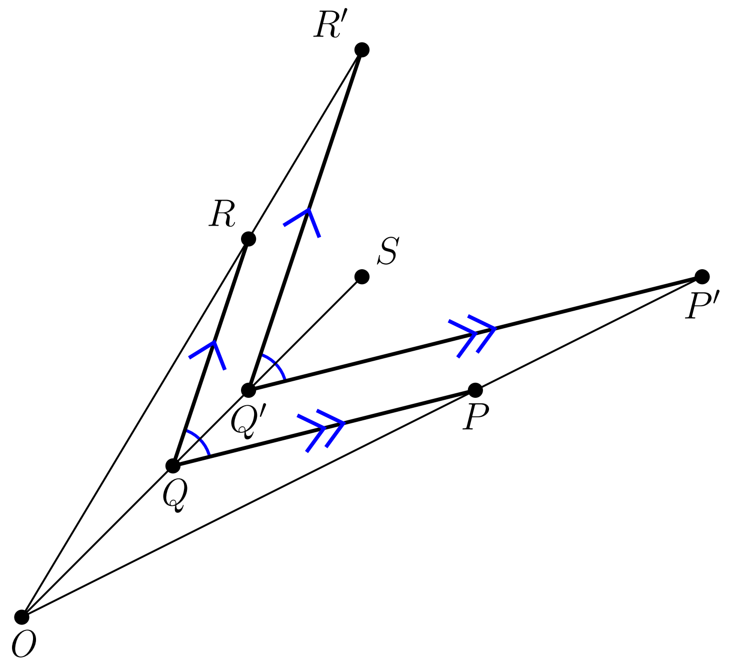
$\angle PQR = \angle P'Q'R'$ because they are corresponding angles on the parallel lines QR and $Q'R'$ using line PQ as the traversal.





Now we consider the most common scenario, where the scaling centre O does not fall on lines PQ or QR. These two intersecting lines split the plane into four regions. The scaling centre could be in any of these four regions, and unfortunately the proof plays out slightly different in each region.

Here is the first of the four cases. In this case, I have had to introduce an extra point S. This point is only needed for the purpose of identifying angles, so it's precise position does not matter. It just needs to be on ray OQ, somewhere beyond point Q'

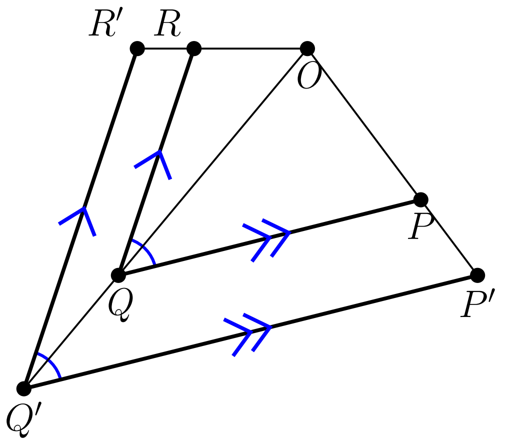


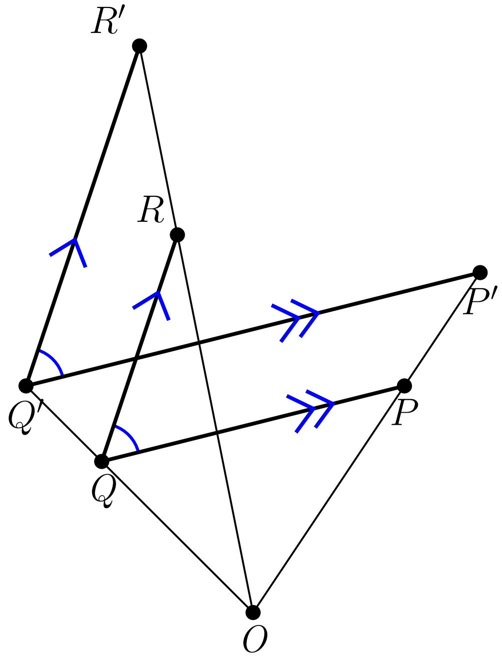
Since scaling preserves orientation, PQ is parallel to $P'Q'$ and QR is parallel to $Q'R'$. Using OS as the traversal, each pair of parallel lines gives a pair of corresponding angles which must be equal.

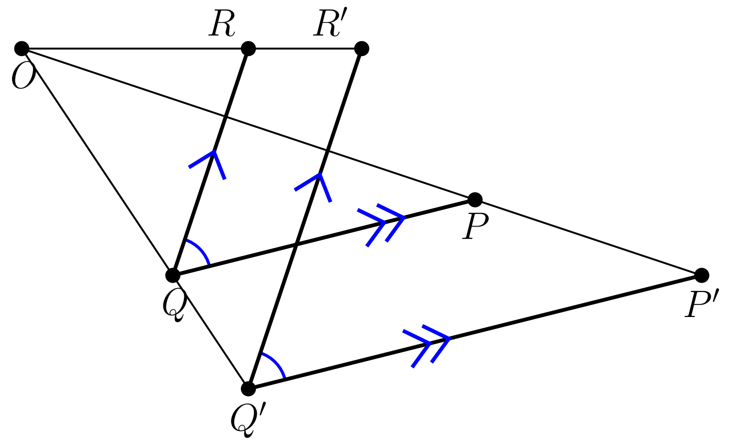
$PQ \parallel P'Q' \implies \angle PQS = \angle P'Q'S$

$QR \parallel Q'R' \implies \angle SQR = \angle SQ'R'$

Hence $\angle PQR = \angle PQS + \angle SQR = \angle P'Q'S + \angle SQ'R' = \angle P'Q'R'$ as required.



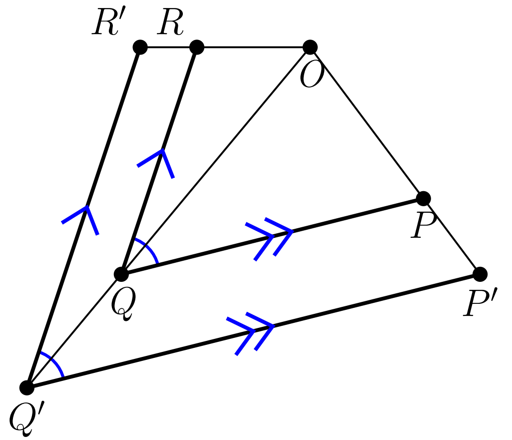


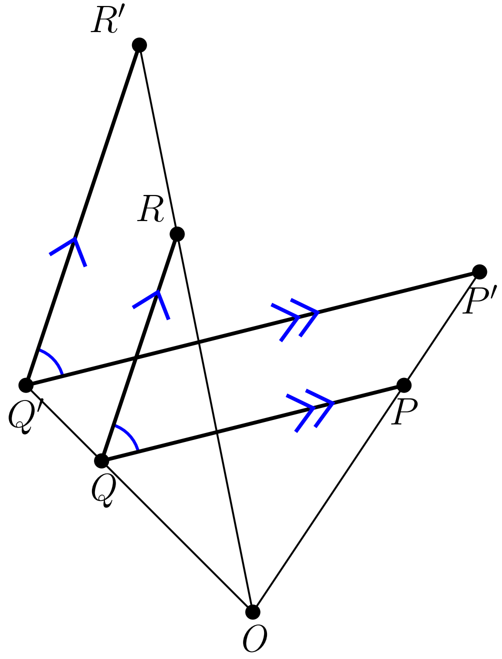


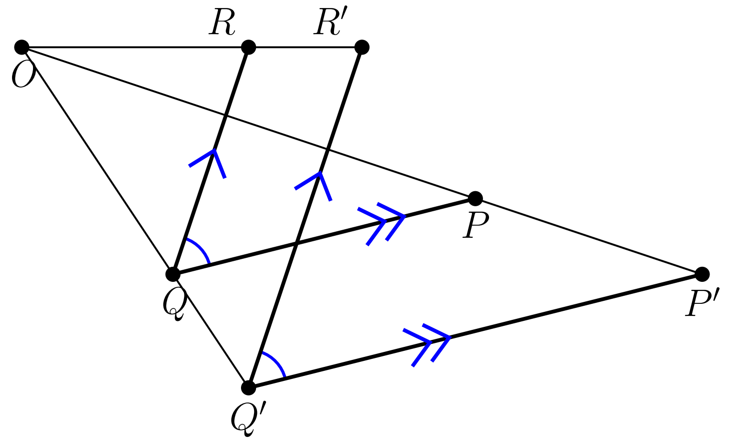
Shown above are diagrams for the remaining three cases. In each case, prove $\angle PQR = \angle P'Q'R'$.

Hints: In each case, since scaling preserves orientation, we know $PQ \parallel P'Q'$ and $QR \parallel Q'R'$. Again, each pair of parallel lines allows us to identify two pairs of corresponding angles which must be equal, but the angles differ by diagram, and the method for combining those angles may change. Curiously, we don't need the extra point S in any of these three cases.

Since the proofs are similar to the proof on the previous page, I haven't provided a full explanation of each case. Instead, I show the two pairs of angles which can be identified as being equal, due to being corresponding angles on the two pairs of parallel lines with OQ' as the traversal. Then I provide the equations that complete the proof.







$$\angle PQO = \angle P'Q'O$$
$$\angle OQR = \angle OQ'R'$$

$$\angle OQP = \angle OQ'P'$$
$$\angle OQR = \angle OQ'R'$$

$$\angle PQO = \angle P'Q'O$$
$$\angle OQR = \angle OQ'R'$$

$$\begin{aligned}\angle PQR &= \angle PQO + \angle OQR \\ &= \angle P'Q'O + \angle OQ'R' \\ &= \angle P'Q'R'\end{aligned}$$

$$\begin{aligned}\angle PQR &= \angle OQR - \angle OQP \\ &= \angle OQ'R' - \angle OQ'P' \\ &= \angle P'Q'R'\end{aligned}$$

$$\begin{aligned}\angle PQR &= \angle PQO - \angle OQR \\ &= \angle P'Q'O - \angle OQ'R' \\ &= \angle P'Q'R'\end{aligned}$$

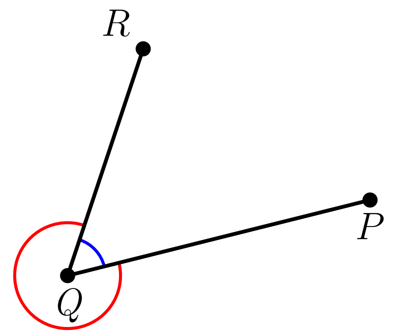
For each diagram, we have shown that $\angle PQR = \angle P'Q'R'$, which means that scaling preserves angles. This concludes the discussion of the different possible diagrams. Except ... there are two other issues that you might still be wondering about.

Firstly, all my diagrams involve enlargements, meaning the scaling factor $f > 1$. Earlier I did make the claim that the arguments given above will still work when dealing with a contraction, meaning $0 < f < 1$. If you don't believe this, I could make the suggestion that you create a new set of diagrams covering the contraction. But that would be a cruel suggestion, because there's a much easier way to show the proofs still work for contractions.

In an earlier section I mentioned that the scaling transformation with scaling factor f has an inverse transformation: a scaling transformation with scaling factor $\frac{1}{f}$ that uses the same scaling centre. In all the diagrams given in this section we used an enlargement transformation on points P, Q and R to produce points P', Q' and R'. So now we can just change our perspective and pretend these diagrams are showing the effect of a contraction transformation using the scaling factor $\frac{1}{f}$ on points P', Q' and R' to produce points P, Q and R. A contraction is still scaling transformation, so it must still preserve orientation. That means all the lines identified as parallel in the above diagrams must still be parallel, and all the pairs of corresponding angles identified above as being equal will still be equal. So without drawing *any* extra diagrams we can convince ourselves that all the proofs given above work for contractions as well as enlargements.

Secondly, the diagrams given all show the *same* angle. I did that deliberately, because I think that choice makes the proofs easier to follow. But once you've understood the proofs you might start wondering whether there's any particular special angles that might cause them to break.

The first of the three cases on the previous page uses the relationship $\angle PQR = \angle PQO + \angle OQR$. That relationship fails if the angle we are interested in is the *reflex* angle PQR. In case you haven't heard that last term before, here is a diagram. $\angle PQR$ is the angle marked in blue. The corresponding reflex angle is the angle marked in red that takes the long around. Reflex angles fall between 180° and 360°.



There also seem to be issues when $\angle PQR = 180°$, since it becomes unclear which side of the angle we are referring to. Also, in some of the earlier special cases, if $\angle PQR = 180°$, some previously parallel lines become collinear lines, which seems to break some steps in the proof.

It seems we should only claim the proof given above to be valid when $0° < \angle PQR < 180°$. I am not saying that scaling fails to preserve angles outside this range. It would be surprising if that happened. I am only saying that the arguments provided above fail to *prove* that scaling preserves angles outside this range, so we should try to find some other way to prove that angles outside this range are also preserved.

Luckily, there are easy ways to prove scaling preserves angles of $180°$, and also preserves reflex angles. Can you see how?

If $\angle PQR = 180°$ then PQR is a line. In the second section of this chapter we proved that scaling transformations preserve linearity, so if PQR is a line, so is $P'Q'R'$. Thus angles of 180° are preserved under scaling transformations.

An angle and its corresponding reflex angle combine to give a full revolution, so those two angles must sum to 360°. If $\angle PQR = \angle P'Q'R'$, then $360° - \angle PQR = 360° - \angle P'Q'R'$. We *have* proved that angles between 0° and 180° are preserved, so the corresponding reflex angles are also preserved.

Now we have covered all the possible scenarios, so it's safe to state our conclusion as a theorem.

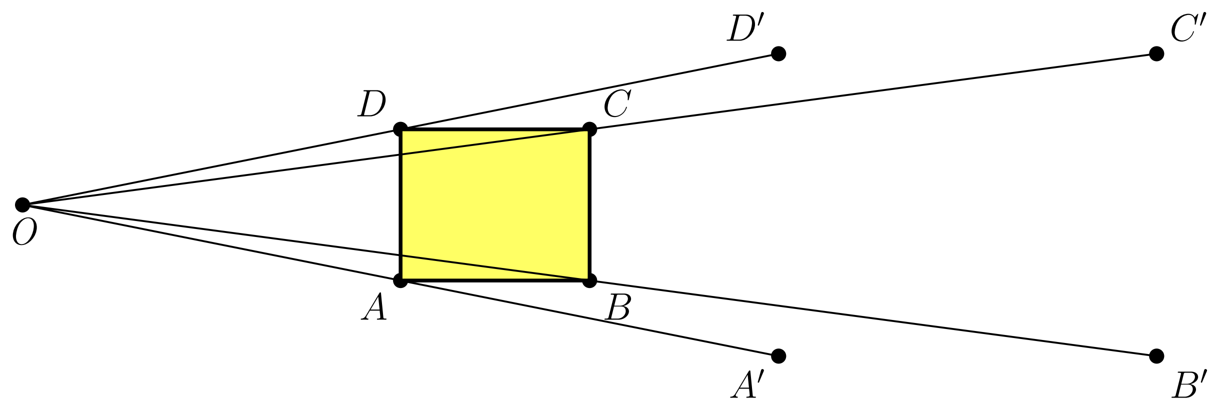
Theorem 14.4: Scaling transformations preserve the size of angles.

14.09 Effect of scaling on area

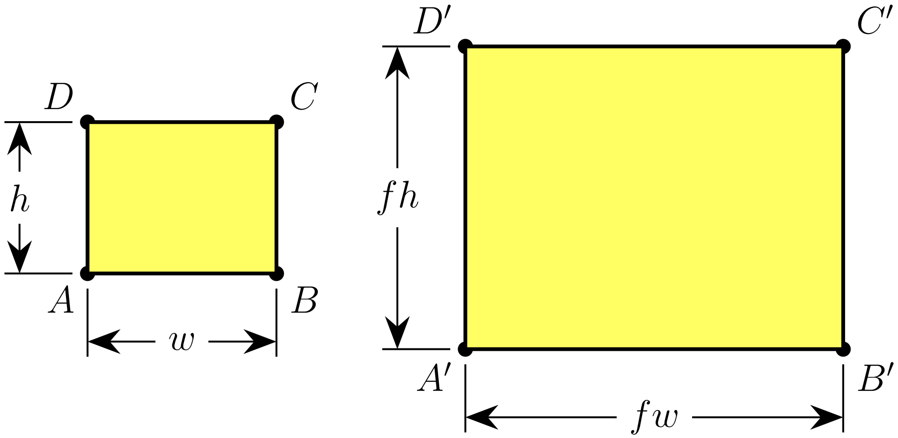
Consider rectangle $ABCD$ with width w units and height h units. We know its area is wh square units.

What happens to the rectangle and its area if we apply a scaling transformation with a scaling factor f?

Let the result of transforming points A, B, C and D be A', B', C' and D' respectively.



In this diagram, while AB and DC are parallel, at first glance they don't seem to be. This is an optical illusion created by the sloping lines. To avoid this mistaken perception, and to leave more room to clearly show the dimensions, I will stop showing the scaling centre and the sloping guidelines that start at that point.



From theorem 14.2, the result of scaling a quadrilateral is another quadrilateral.

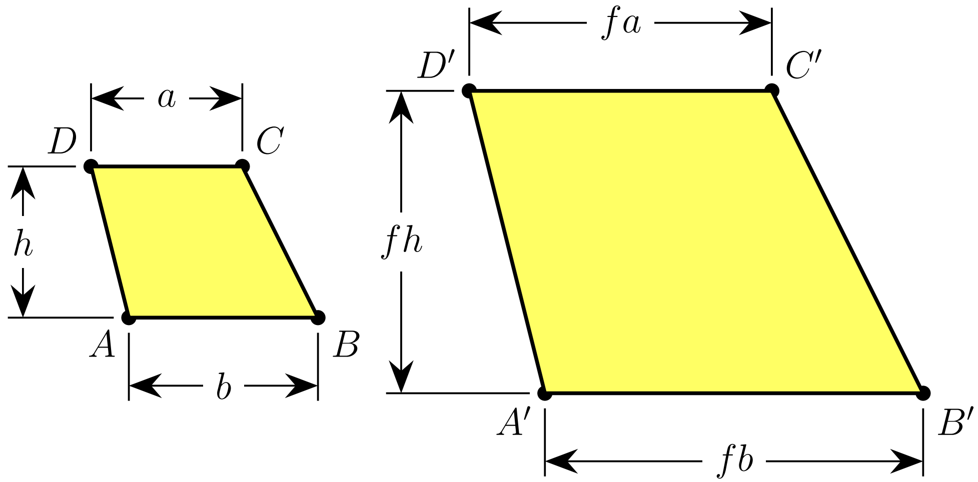
$ABCD$ is a rectangle so its adjacent sides are perpendicular. Scaling preserves angles, so adjacent sides of quadrilateral $A'B'C'D'$ are also perpendicular, meaning it is also a rectangle. This gives the unsurprising result that scaling a rectangle produces a rectangle.

Scaling adjust all distances consistently, so rectangle $A'B'C'D'$ has width fw units and height fh units.

Its area is the product of its width and height, which is whf^2 square units.

This type of argument can be performed for any of the polygons we considered in earlier chapters. Try it for the trapezium.

$ABCD$ is a trapezium with $AB \parallel DC$. The parallel sides have lengths a units and b units. Its perpendicular height is h units. We know its area is $\frac{1}{2}(a+b)h$ square units. Applying a scaling transformation to points A, B, C and D gives points A', B', C' and D' respectively.



Theorem 14.2 tells us that scaling trapezium $ABCD$ will give another quadrilateral, $A'B'C'D'$.

Theorem 14.3 tells us that scaling preserves orientation, so $AB \parallel A'B'$ and $DC \parallel D'C'$. Because $ABCD$ is a trapezium, $AB \parallel DC$. Combining these three properties proves $A'B' \parallel D'C'$, so quadrilateral $A'B'C'D'$ is a trapezium. That is, scaling a trapezium gives a trapezium.

A less efficient way to prove that last result is to consider co-interior angles. Since $AB \parallel DC$, the co-interior angles $\angle DAB$ and $\angle ADC$ on traversal AD must sum to 180°. Scaling preserves angles, so $\angle DAB = \angle D'A'B'$ and $\angle ADC = \angle A'D'C'$. Hence $\angle D'A'B'$ and $\angle A'D'C'$ also sum to 180°. Since they are cointerior angles on lines $A'B'$ and $D'C'$ with $A'D'$ as the traversal, the fact that they sum to 180° means $A'B' \parallel D'C'$, which matches the previous result.

Scaling transformations scale all lengths by the scaling factor f. Hence trapezium $A'B'C'D'$ has parallel sides of lengths fa and fb and has perpendicular height fh.

Its area in square units is $\frac{1}{2}(fa+fb)fh = \frac{1}{2}(a+b)hf^2$.

Clearly some of the points we've been making for scaling rectangles and trapezia apply more generally to all scaling transformations on polygons.

- By theorem 14.2, triangles transform to triangles, quadrilaterals transform to quadrilaterals, pentagons transform to pentagons, and so on.
- Scaling preserves orientation, so sides that are parallel have corresponding parallel sides in the transformed figure. Thus parallelograms will transform to parallelograms.

- Scaling preserves angles, so sides that were perpendicular have corresponding perpendicular sides in the transformed figure. Thus right triangles will transform to right triangles and rectangles transform to rectangles.
- The Consistent Scaling Postulate states that all distances are scaled consistently, so sides of equal length in the original polygon have corresponding sides of equal length in the transformed polygon. Thus isosceles triangles scale to isosceles triangles, equilateral triangles scale to equilateral triangles and kites scale to kites.
- Applying multiple properties from the above points shows rhombi transform to rhombi and squares transform to squares. We haven't yet defined regular polygons, but if you know what they are, you can probably see why a regular polygon with n sides transforms to a regular polygon with n sides.

In the above, note that I did not say that a triangle would transform to a *different* triangle, because under the trivial identity scaling transformation that uses a scaling factor of one, a triangle "transforms" to the same triangle. That is, it is unchanged.

It's suspicious that in both cases we've considered, applying a scaling factor of f resulted in the area being scaled by f^2. Let's check some other area formulae.

A triangle with base b and height h has area $\frac{1}{2}bh$. The scaled triangle will have base *fb* and height *fh*, giving area $\frac{1}{2}(fb)(fh) = \frac{1}{2}bhf^2$.

A kite with diagonals of length p and q has area $\frac{1}{2}pq$. It will scale to a kite with diagonals of length fp and fq, which will have area $\frac{1}{2}(fp)(fq) = \frac{1}{2}pqf^2$. Scaling transformations scale *all* distances consistently, diagonals as well as sides.

This strengthens our suspicion that a scaling factor of f always scales area by a factor f^2. But how do we prove that?

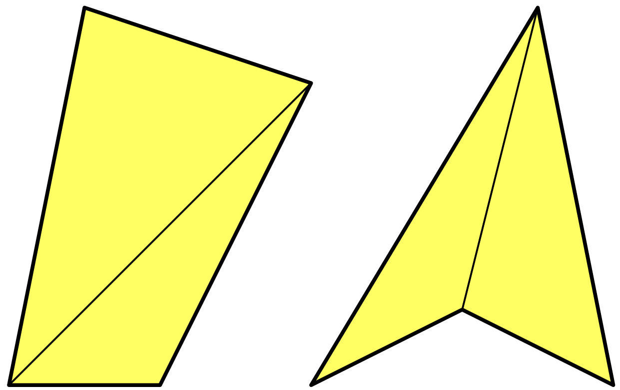
The good news is that we already have the triangle case proved, because the $\frac{1}{2}bh$ formula covers *all* triangles.

But we only have area formulae for several special quadrilaterals. If someone gives us an irregular quadrilateral that has no equal sides and no parallel sides, we don't yet have a formula for its area. We also don't yet have any area formulae for polygons with more than four sides.

There is a method to prove the result for all polygons, but while it is simple, it is not obvious. Can you see it? Hint: It uses a postulate that we haven't seen for several chapters.

Since we have the property proved for triangles, we dissect the polygon into triangles and use the Area Sum Postulate.

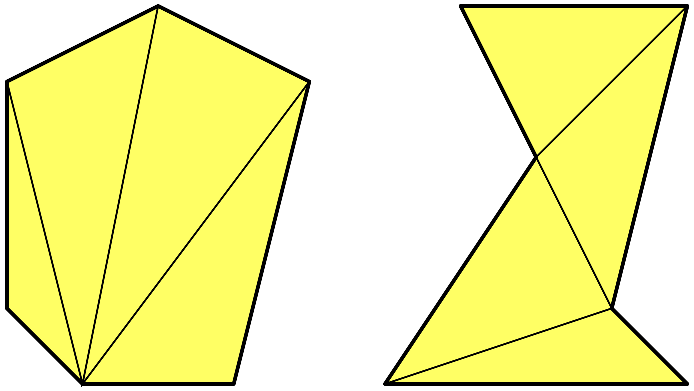
Quadrilaterals can always be dissected into two triangles by drawing one diagonal. For convex quadrilaterals either diagonal works. For concave quadrilaterals, choose the internal diagonal.



Let the area of the quadrilateral be Q and the areas of the two triangles be T_1 and T_2. Let the areas of the corresponding transformed figures be the same symbols with a dash attached. Our argument is

$$Q' = T_1' + T_2' = f^2 T_1 + f^2 T_2 = f^2(T_1 + T_2) = f^2 Q$$

The first and last steps use the Area Sum Postulate. The second step uses the result that scaling triangles by the scaling factor f does scale their area by the factor f^2.



For polygons with five or more sides, we just have to sum more than two triangles. It is always possible to dissect a polygon with n sides into $n - 2$ triangles. For convex polygons we simply choose one vertex and join it to every other vertex. For concave polygons there isn't a method that is so easily explained, but it is always possible to find a solution employing only diagonals, meaning line segments that join two vertices of the polygon. The diagram shows hexagons dissected into four triangles.

Theorem 14.5: When a polygon is scaled by the factor f, its area is scaled by the factor f^2.

In fact, this property also holds for curved figures such as circles and ellipses, but so far we have only proved it true for polygons.

14.10 Transitive Relationships

You are probably already aware that for three numbers a, b and c:

- If $a = b$ and $b = c$ then $a = c$.
- If $a < b$ and $b < c$ then $a < c$.

Mathematician have a word for relationships that behave in this way: transitive. That is, equality is a transitive relationship, as is the “less than” relationship.

Of the following four relationships that compare the size of numbers, which ones are transitive? $>, \leq, \geq, \neq$.

The $>$, $\leq$ and $\geq$ relationships are transitive. The $\neq$ relationship is not transitive, as can easily be demonstrated by providing a counterexample. If $a = 2$, $b = 4$ and $c = 2$, then $a \neq b$ and $b \neq c$, but we can *not* conclude from this that $a \neq c$, because they clearly *are* equal.

14.11 Review of congruence

This section provides a quick review of congruence concepts that I'm assuming you already know. I'm doing this because our next step will be to introduce similarity, which will involves several concepts analogous to the following congruence concepts.

The translation, reflection and rotation transformations preserve linearity, lengths and angles. Since they preserve linearity, applying them to a line, line segment or ray will produce another line, line segment or ray respectively. Polygons are composed of line segments, so applying one of these transformations to a polygon produces another polygon with the same number of sides.

Since any individual translation, any individual reflection and any individual rotation will preserve linearity, lengths and angles, any sequence of transformations selected from these three types will also preserve these three properties. For example, if we were to apply two translations, three reflections through different lines and five rotations around different centres to a pentagon, then no matter what order we used for ordering those ten transformations, the resulting figure would still be a pentagon with the same side lengths and same angles as the original pentagon.

Figures that share so many properties are very closely related to each other, so it's useful to have a word that describes this relationship. The word we use for the relationship is congruence. Two figures linked by this relationship are said to be congruent.

We can define congruence in terms of these three types of transformation.

Definition: Two figures are **congruent** if one can be made to exactly coincide with the other using only translation, rotation and reflection transformations.

There is no limit to the number of transformations used. If you can find a sequence of ten transformations that will make one pentagon exactly coincide with another, then those two pentagons are congruent. The only restriction is that each of the ten transformations must be a translation, a reflection or a rotation.

In the definition given above, “figures” includes polygons such as triangles, quadrilaterals and pentagons, but also shapes involving curves, like circles and ellipses.

When the shapes are polygons, the fact that translations, rotations and reflections preserve lengths and angles leads immediately to the following theorem.

Theorem 14.6: If two polygons are congruent then each angle of one is equal to the corresponding angle of the other and the length of each side of one is equal to the length of the corresponding side of the other.

The converse of this theorem is also true.

Theorem 14.7: If the corresponding angles of two polygons are equal and the length of each side of one is equal to the length of the corresponding side of the other, then the two polygons are congruent.

If you prefer, the previous two theorems could be combined into one theorem by using the “if and only if” wording.

Congruence is a transitive relationship. In the following explanation, an “allowed transformation” is any translation, rotation or reflection.

Consider three pentagons labelled A, B and C. Say we are told that pentagons A and B are congruent and pentagons B and C are congruent. The first congruence means that there exists a sequence of allowed transformations that can be applied to pentagon A to make it coincide with pentagon B. The second means there is sequence of allowed transformations that can be applied to pentagon B to make it coincide with pentagon C.

Make a new sequence of transformations that simply takes the first sequence mentioned in the previous paragraph followed by the second sequence. Clearly this new combined set of transformations will make pentagon A coincide with pentagon C. Since such a sequence exists, and the sequence does only include allowed transformations, pentagons A and C are congruent.

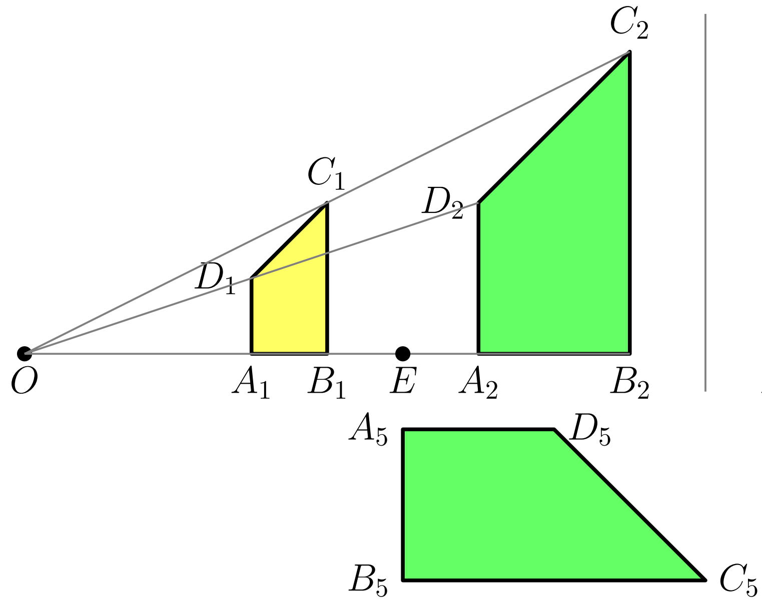
While I referred to pentagons to make the argument more concrete, the argument applies to any figure. Thus we can conclude that congruence is transitive.

Theorem 14.8: Congruence is a transitive relationship. That is, if figure A is congruent to figure B, which is congruent to figure C, then figure A is congruent to figure C.

14.12 Similarity

Scaling preservers linearity and angles, but does not preserve lengths, except in the trivial case where the scaling factor is one. Instead, it scales all lengths by the scaling factor.

Figures that are linked by scaling relationships are not as closely related as figures linked by translations, reflections and rotations, but the relationship between them is still important enough to deserve a special word. The word for the relationship is similarity. Two figures linked by this relationship are said to be similar.



The diagram shows five trapezia. In this discussion, I'm going to refer to each trapezium by the number used in the subscripts in the names of its vertices. Thus trapezium 3 means the trapezium with vertices $A_3B_3C_3D_3$.

Trapezium 2 is the result of applying a scaling transformation to trapezium 1, using point O as the scaling centre and a scaling factor of 2. Thus trapezium 1 and trapezium 2 are similar.

Trapezium 2 was reflected in the vertical gray line to produce trapezium 3. Trapezium 2 was also translated 8cm to the right to produce trapezium 4, and rotated 90° clockwise around point E to produce trapezium 5. That is, the green trapezia are congruent. Hence their corresponding sides have equal length and their corresponding angles are equal.

Because trapezia 1 and 2 are similar, their corresponding angles are equal. But the green trapezia are all congruent, so the corresponding angles of the those trapezia are also equal. Combining those two results, the corresponding angles of *all* the trapezia are equal.

Also, trapezia 1 and 2 are linked by a scaling transformation, so their correponding sides are in constant proportion. All sides of trapezium 2 are twice the length of the corresponding side of trapezium 1. But the green trapezia are all congruent, so their corresponding sides have equal length. This means the corresponding sides of all the trapezia are in constant proportion. When comparing any pair of green trapezia, the constant proportion is one. When comparing any green trapezium to the yellow trapezium, the constant proportion is two.

The important features of similarity are the preservation of angles and the scaling of sides. We don't want to define similarity in a way that only recognises trapezia 1 and 2 as similar, due to them being directly linked by a single transformation. Instead we want a definition that recognises that trapezium 1 is similar to *all* the green trapezia. We can do this by defining similarity by the same method used for defining congruence, but allowing one extra transformation: scaling. Here again is the definition of congruence and corresponding definition of similarity.

Definition: Two figures are **congruent** if one can be made to exactly coincide with the other using only translation, rotation and reflection transformations.

Definition: Two figures are **similar** if one can be made to exactly coincide with the other using only translation, rotation, reflection and scaling transformations.

The definition does *not* specify that the transformations used *must* include a scaling transformation. This means that if two figures are congruent, then they are also similar.

14.13 Similarity is a transitive relationship

The proof of this result is almost identical to the proof that congruence is transitive. We simply replace every occurrence of the word “congruent” by “similar” and add scaling to the list of allowed transformations which may be used. However, rather than starting with the more concrete example that referred to pentagons, this time I’ll go the more general argument that refers to any figure.

Let an “allowed transformation” be any translation, rotation, reflection or scaling transformation.

Consider three figures labelled A, B and C. Say we are told that figures A and B are similar and figures B and C are similar. The first relationship means that there exists a sequence of allowed transformations that can be applied to figure A to make it coincide with figure B. The second means there is sequence of allowed transformations that can be applied to figure B to make it coincide with figure C.

Make a new sequence of transformations that simply takes the first sequence mentioned in the previous paragraph followed by the second sequence. Clearly this new combined set of transformations will make figure A coincide with figure C. Since such a sequence exists, and the sequence does only include allowed transformations, figures A and C are congruent.

Theorem 14.9: Similarity is a transitive relationship. That is, if figure A is similar to figure B, which is similar to figure C, then figure A is similar to figure C.

It is also useful to think about how this type of argument applies when the congruence and similarity relationships are mixed.

If figure A is *congruent* to figure B, but figure B is *similar* to figure C, how are figures A and C related?

Since figures A and B are congruent, we can find a sequence of transformations that will make A coincide with B, with that sequence only containing translations, rotations and reflection. No scaling transformation is required. Figures B and C are similar. That means that there is a sequence of transformations that will make B coincide with C, but that list can contain translation, rotation, reflection and scaling transformations. When we join those two sequences, we get a set of transformations that will transform A to coincide with C. That list might contain scaling transformations, so we can only claim that A and C are similar, not congruent.

However, we can't claim that figures A and C are definitely *not* congruent. Congruence is a special case of similarity, so it is still possible that B and C are congruent as well as similar, which would then make A and C congruent.

Theorem 14.10: If figure A is congruent to figure B, which is similar to figure C, then figure A is similar to figure C.

Does your answer change if figure A is congruent to figure B, but figure B is similar to but *not* congruent to figure C?

The extra information tells us that B and C definitely do not have the same scale. A and B being congruent, do have the same scale, so A and C definitely have different scales. Thus we can conclude that A and C are still similar but are definitely not congruent.

This conclusion is consistent with the earlier comments about the five trapezia. Trapezium 2 was created from trapezium 1 using a scaling transformation with scaling factor two, so these two trapezia are similar but not congruent. Trapezium 2 is congruent with trapezia 3, 4 and 5. Hence trapezium 1 is similar to but not congruent to trapezia 3, 4 and 5.

In general, if you *know* that two figures are congruent, then it's not a great idea to tell someone that they are similar, and hide the fact that they are also congruent. We should strive to give the most informative description as possible, so if they are congruent, state that.

However, there are several important proofs that rely on identifying similar shapes, most frequently similar triangles. Because we have defined similarity in a manner that includes congruence as a special case, those proofs continue to function correctly in any special cases where the similar triangles are also congruent. By contrast, if we had defined similarity in a way that did not include congruence as a special case, then in all those proofs we would have to add extra wording to cover all the situations where congruence might arise. That is, we define similarity to include congruence since that makes our life easier when writing proofs.

14.14 Properties of similar polygons

By the definition of similarity, two figures are similar if one can be made to exactly coincide with the other using only translation, rotation, reflection and scaling transformations. All four of these transformation types conserve angles. The first three of these transformations conserve lengths, while the fourth scales all lengths by the scaling factor. Combining these properties gives:

Theorem 14.11: If two polygons are similar then each angle of one is equal to the corresponding angle of the other and the lengths of the sides of one have a constant ratio to the lengths of the corresponding sides of the other.

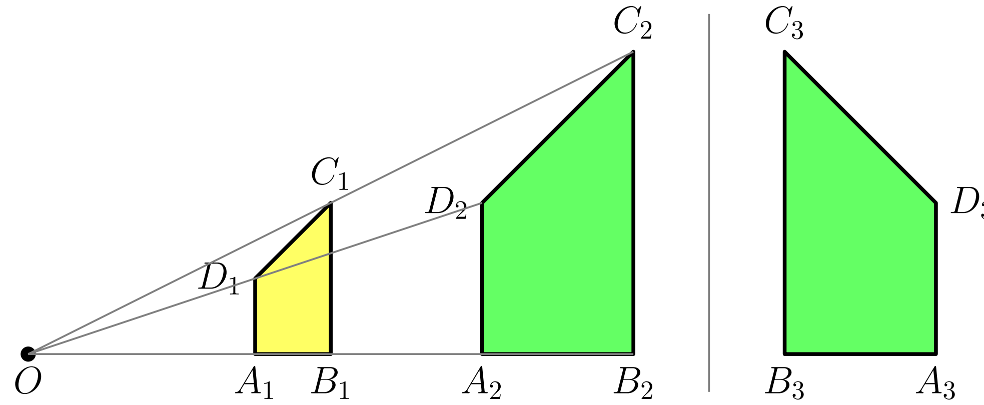
As for the analogous theorem relating to congruence, the converse of this theorem is also true. Our next step is to prove the converse theorem, which is:

Theorem 14.12: If the corresponding angles of two polygons are equal and the lengths of the sides of one have a constant ratio to the lengths of the corresponding sides of the other, then the two polygons are similar.

Here again are the first three trapezia from the previous diagram.

Recall that trapezium 2 was created by applying a scaling transformation to trapezium 1, and then trapezium 3 was created by applying a reflection transformation to trapezium 2. Thus the three trapezia are similar.

Imagine someone gave us a page containing a diagram that shows only trapezia 1 and 3. They tell us that the corresponding angles of the two trapezia are equal, and that the length of each side of trapezia 3 is twice the length of the corresponding side of trapezia 1. They ask us to prove that these two trapezia are similar.



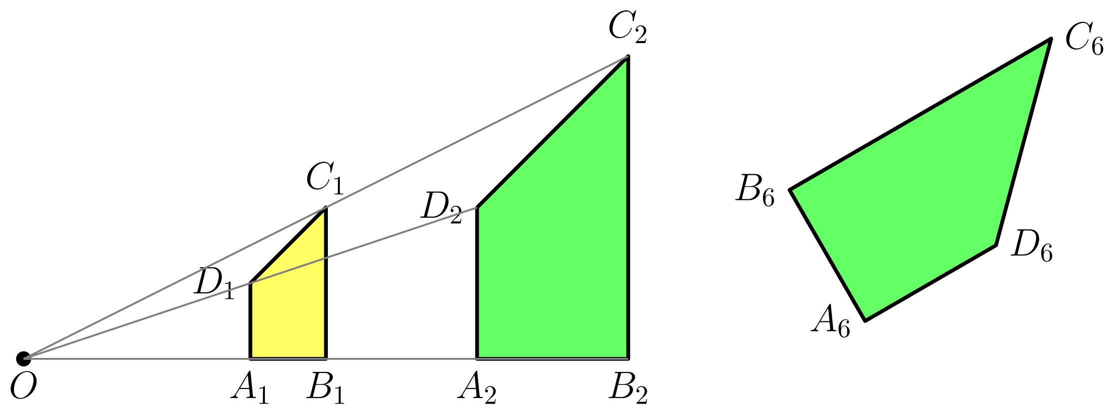
Now we have seen the version of this diagram that shows all three trapezia, so we know trapezium 2 is the "missing link" between the trapezia 1 and 3. So, on the page they gave us, we can construct trapezium 2 from trapezium 1 by applying a scaling transformation with scaling factor 2 and scaling centre O. Then we demonstrate that reflecting trapezium 2 in a vertical line produces trapezium 3. To do this, recall that if point A_2 reflects to point A_3, then the reflection line is the perpendicular bisector of line segment A_2A_3. So we could construct the perpendicular bisectors of each of the line segments A_2A_3, B_2B_3, C_2C_3 and D_2D_3, and we would find that they are the *same* vertical line, so reflecting trapezium 2 in that line will produce trapezium 3. We've now demonstrate that we can make trapezium 1 coincide with trapezium 3 by using a scaling transformation that gives trapezium 2, followed by a reflection transformation. Thus trapezium 1 and 3 are similar.

I've presented the above argument because for many people, it's the first approach they think of, so I want to discuss why it isn't the best approach. It has two problems.

Firstly, it relies on the construction of four perpendicular bisectors producing a single line. Any construction that uses a pair of compasses and a straight edge to draw arcs and lines with a pencil has limited accuracy. At best, we can claim that, as far as we can see, it's plausible that the four perpendicular bisectors are a single straight line, so this is not an irrefutable proof.

Secondly, it relied on us being able to figure out the sequence of transformations that would make trapezia 1 and 3 coincide. That was easy for us in the above example because, while we were given a page that only showed trapezia 1 and 3, I said that these were the trapezia from our diagram that also showed exactly where we should place trapezia 2 and gave the locations of a suitable scaling centre and reflection line. If someone gave us a page with a new trapezium, let's call it trapezium 6, it could be far from obvious how to generate a sequence of transformations to make trapezium 1 coincide with trapezium 6.

Let's start again and solve this problem a better way. Say someone gives a page showing trapezia 1 and 6 shown in the diagram at right. That is, when they give us the page, it does not yet contain trapezium 2. They tell us that each angle in trapezium 6 is equal to the corresponding angle in trapezium 1 and that each side of trapezium 6 has a length that is twice the length of the corresponding side in trapezium 1. This time we have no clues about how to use transformations to make trapezium 1 coincide with trapezium 6. How do we prove these two trapezia are similar?

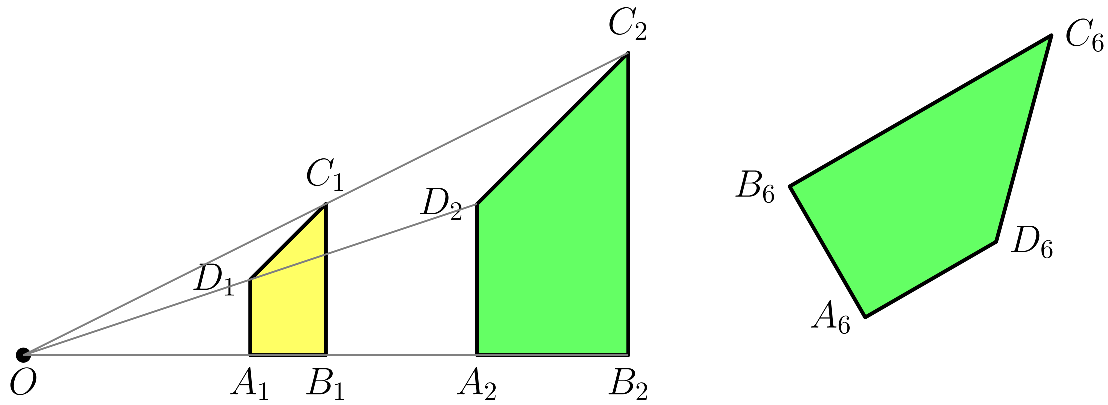


The first step is to scale trapezium 1 using a scaling factor of 2. We can use any scaling centre we like, but to keep things simple I've used the same scaling centre as in the previous diagram, giving trapezium 2.

Why is trapezium 2 similar to trapezium 1? Why is trapezium 2 congruent to trapezium 6? How does this prove the result?

Trapezium 2 was produced by applying a scaling transformation to trapezium 1, so these two trapezia satisfy the definition of "similar".

Scaling preserves angles, so the angles of trapezium 2 are equal to the corresponding angles of trapezium 1. But we were told that the angles of trapezium 6 are equal to the corresponding angles of trapezium 1. Hence the angles of trapezium 6 are also equal to the corresponding angles of trapezium 2.



Scaling transformations scale all distances by the scaling factor, so the lengths of all sides of trapezium 2 are twice the length of the corresponding side of trapezium 1. But we were told that the lengths of all sides of trapezium 6 are twice the length of the corresponding side of trapezium 1. Hence the lengths of all sides of trapezium 6 are equal to the length of the corresponding side of trapezium 2.

Combining the results of the previous two paragraphs, trapezium 6 has angles and side lengths equal to the corresponding angles and side lengths of trapezium 2, so by theorem 14.7 they are congruent.

We have proved that trapezium 6 is congruent to trapezium 2 which is similar to trapezium 1. Hence by theorem 14.10, trapezia 1 and 6 are similar. This proves the theorem.

14.15 Triangle similarity tests

Thereom 14.7 tells us that, in general, for two polygons to be congruent, each angle in one must equal the corresponding angle in the other and the length of each side in one must equal the length of the corresponding side in the other.

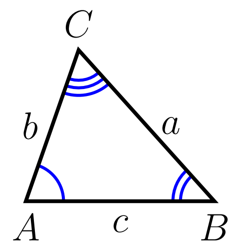
But triangles are simpler. For example, if the lengths of the sides of one triangle equal the lengths of the corresponding sides of a second triangle, this *forces* the corresponding angles of the two triangles to also be equal, so we know the triangles are congruent without having to test the angles. This is know as the SSS test for congruency of triangles. The other common tests for triangle congruency are labelled SAS, AAS and RHS. At school level, these four tests are usually treated as postulates. Some more advanced textbooks may present the first three as postulates and prove the RHS test as a theorem.

A similar thing happens for similarity. Testing triangles for similarity is much simpler than testing polygons with more than three sides.

We will examine three simple tests for similarity of triangles, known as the SSS, SAS and AAA similarity tests. As for the congruency tests, an A denotes equal corresponding angles. However, where an S in a congruency test denotes equal side lengths, in a similarity test it denotes sides being in a constant proportion.

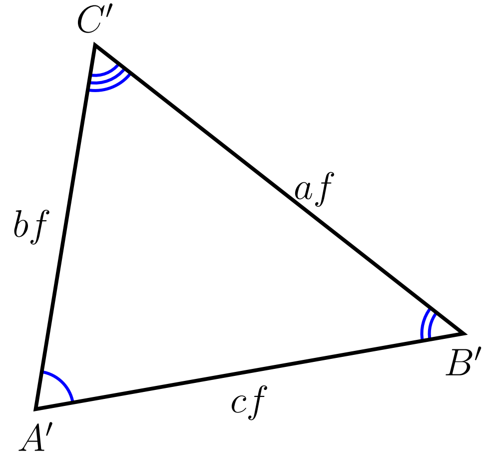
We actually won't need all three tests. But once you understand the method for proving any one of them, it's easy to do the others, so we might as well do all three of them. We start by essentially rerunning the argument from the previous section, but with triangles rather than trapezia.

Consider two triangles for which each angle in one equals the corresponding angle in the other and the length of each side of one is a constant proportion of the corresponding side in the other. Our aim is to prove that the triangles must be similar.



If the constant proportion mentioned happens to be one, then the corresponding side lengths and angles are equal, so the triangles are congruent by theorem 14.7. If they are congruent, they are also similar, so the result is proved. The rest of this proof considers the case where the constant proportion is not one, so the two triangles are different sizes.

Choose to label the triangles so that $\triangle ABC$ is the smaller and $\triangle A'B'C'$ is the larger. Label the side lengths of the smaller triangle as a, b and c as shown, with the lengths of the corresponding sides of the larger triangle being af, bf and cf, where $f > 1$. In my diagram $f = 2.3$.

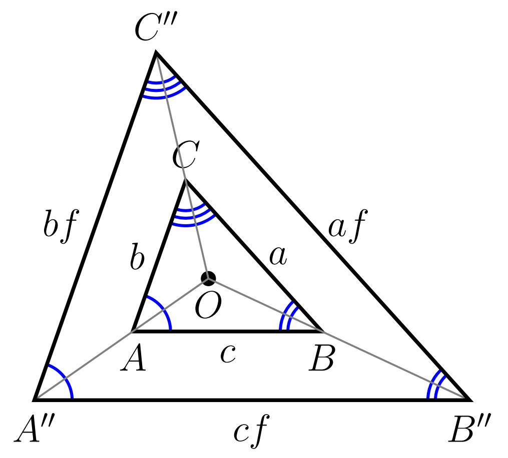


In my diagram, the larger triangle is rotated 10° relative to the smaller diagram. I hope that this amount of rotation is large enough to be clearly visible, so that it acts as a reminder that this proof does *not* require that the two triangles have the same orientation, but it also small enough that it is still easy to quickly identify corresponding angles and sides.

The proof *is* still valid for larger rotations, and is also valid if a reflection has made the second triangle back-to-front relative to the first one, but if the diagram showed those cases it would take a little longer to locate corresponding angles and sides. If you like, you can check this by making your own version of the second triangle in some other orientation.

In fact, the property we are proving does not even require that two triangles be in the same plane, so you could even read this proof using the smaller triangle shown at right which is probably on your vertical computer screen, but using your own drawing of the larger triangle that you have placed horizontally on your table.

Apply a scaling transformation to $\triangle ABC$ using a scaling factor of f, giving $\triangle A''B''C''$. To save space, I've put the scaling centre O inside triangle ABC, which makes the new triangle surround it, but any scaling centre works. If you prefer, you can make your own version of the diagram with the scaling centre outside the triangle. The proof still works.

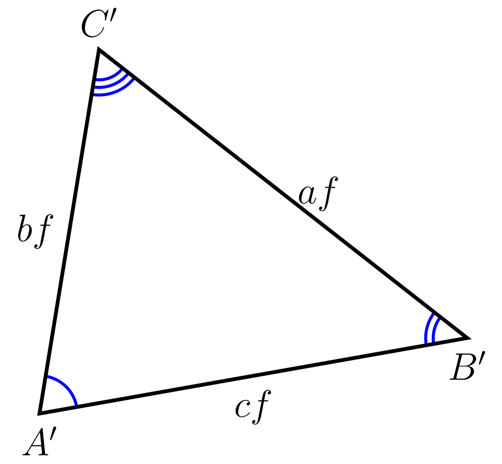


$\triangle A''B''C''$ was made by applying a scaling transformation to $\triangle ABC$ so these two triangles do meet the definition of "similar".

Scaling transformations preserve angles, so corresponding angles of $\triangle ABC$ and $\triangle A''B''C''$ are equal, as shown by the angle markings in the diagram.

Scaling transformations scale all distances by the scaling factor f, so the lengths of the sides of $\triangle A''B''C''$ are af, bf and cf, as shown in the diagram.

This means the corresponding angles and side lengths of $\triangle A'B'C'$ and $\triangle A''B''C''$ are equal, so by theorem 14.7 those two triangles congruent.



$\triangle A''B''C''$ is congruent to $\triangle A'B'C'$ which is similar to $\triangle ABC$, so by theorem 14.10 $\triangle A''B''C''$ is similar to $\triangle ABC$, proving the result.

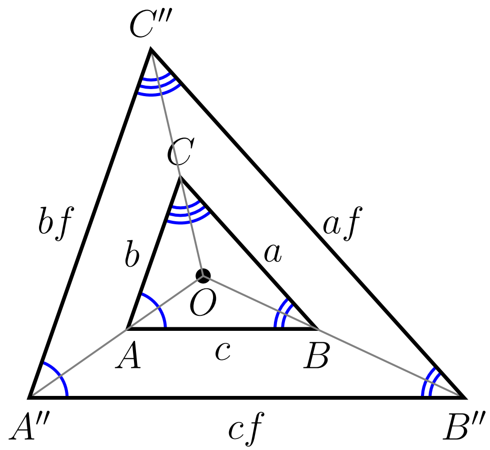
Two paragraphs ago I used theorem 14.7 to prove $\triangle A'B'C'$ and $\triangle A''B''C''$ congruent. To do that I needed six equalities, three pairs of equal angles and three pairs of equal side lengths. I could have reached the same conclusion by using any of three triangle congruency tests, the SSS test, the SAS test or the AAS test. Each of these tests requires only three equalities, not the six required by theorem 14.7. So these three triangle congruency tests can generate three triangle similarity tests that each require less data than theorem 14.12.

The next three sections develop these three triangle similarity tests. I won't rehash the entire proof each time. I still need to consider a trivial special case. After that, for the general case, I will just demonstrate that the reduced amount of data linking $\triangle A'B'C'$ and $\triangle A''B''C''$ is sufficient to prove those triangles are congruent. The diagrams will be almost identical to those shown here, the only difference being that the angle and side length data shown on $\triangle A'B'C'$ will be reduced to the smallest amount of data required to invoke the relevant triangle congruency test.

14.16 SSS similarity test

Theorem 14.13 - The SSS Similarity Test: If the lengths of the three sides of a triangle are a constant proportion of the three side lengths of another triangle, then the two triangles are similar.

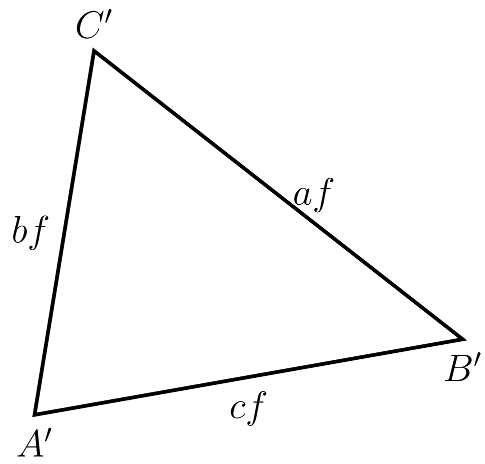
Trivial special case: If the constant proportion mentioned in the theorem is one, then the triangles are congruent by the SSS congruency test, and hence also similar. Now consider the more common case where we are dealing with two triangles of different size.



Then proceed as in the previous section, numbered 14.15.

This time we know that the side lengths of $\triangle A'B'C'$ are f times the lengths of the corresponding sides of $\triangle ABC$, but we have no information about how the angles of the two triangles compare. Hence in the diagram of $\triangle A'B'C'$ at right, the side lengths are shown as af, bf and cf, but there are no angle markings claiming equality with angles of the other two triangles.

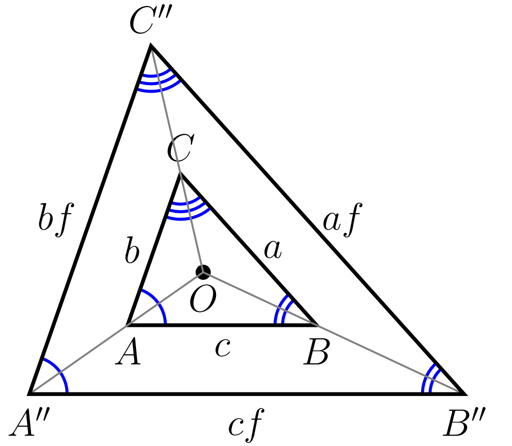
We have enough data to use the SSS congruency test to prove $\triangle A'B'C'$ and $\triangle A''B''C''$ congruent, so the method of section 14.15 still works, proving that $\triangle ABC$ and $\triangle A'B'C'$ are similar. This proves the theorem.



Incidentally, in the previous paragraph we proved that $\triangle A'B'C'$ and $\triangle A''B''C''$ are congruent. Because they are congruent, their corresponding angles must be equal, so we *could* add the appropriate blue angle arcs to $\triangle A'B'C'$ at right to indicate the equal angles, but I am not going to do that. I think this proof is easier to understand when the diagram of $\triangle A'B'C'$ shows the minimum information required to complete the proof. The key point of this theorem is that if we *only* know that the sides of $\triangle A'B'C'$ are all f times the corresponding sides of $\triangle ABC$, that *is* sufficient data for us to conclude that these two triangles are similar. The fact that the corresponding sides of the two triangles are in constant proportion does actually force the corresponding angles to be equal, but we don't need to know anything about the angles in order to invoke the theorem.

14.17 SAS similarity test

Theorem 14.14 - The SAS Similarity Test: If the lengths of two sides of a triangle are a constant proportion of two side lengths of another triangle, and the included angles are equal, then the two triangles are similar.

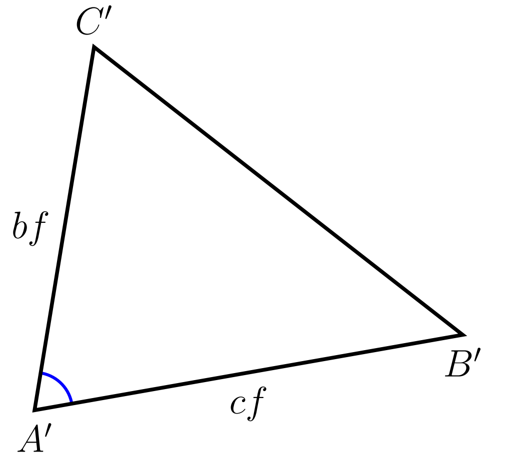


Trivial special case: If the constant proportion mentioned in the theorem is one, then the triangles are congruent by the SAS congruency test, and hence also similar. Now consider the more common case where the constant proportion is not one.

Label the vertices so that included angles mentioned in the theorem are at vertices A and A'. That is, the given data are $\angle BAC = \angle B'A'C'$ and $\frac{|A'B'|}{|AB|} = \frac{|A'C'|}{|AC|} = f > 1$.

Then proceed as in section 14.15.

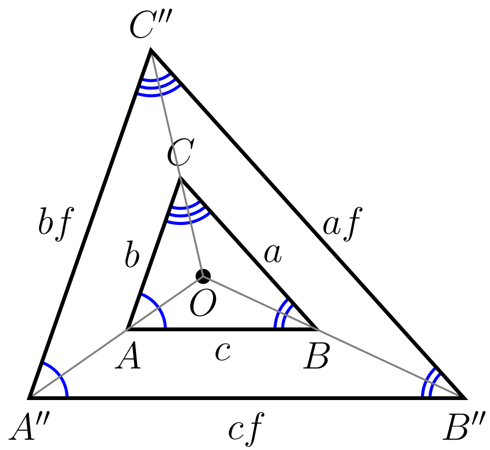
When we reach the point where we need to prove that $\triangle A'B'C'$ and $\triangle A''B''C''$ are congruent, we invoke the SAS congruency test. With that exception, everything plays out as in section 14.15, allowing us to conclude $\triangle ABC$ and $\triangle A'B'C'$ are similar. This proves the theorem.



As in the previous proof, the diagram of $\triangle A'B'C'$ at right shows only the minimum amount of data required to complete the proof, the lengths of sides $A'B'$ and $A'C'$, and the blue arc indicating $\angle B'A'C' = \angle B''A''C'' = \angle BAC$. Once we prove that $\triangle A'B'C'$ and $\triangle A''B''C''$ are congruent, we can conclude that $|B'C'| = cf$ and that the other two angles of $\triangle A'B'C'$ also match the corresponding angles in the other two triangles, but those facts are not shown on the diagram since they weren't required to complete the proof.

14.18 AAA similarity test

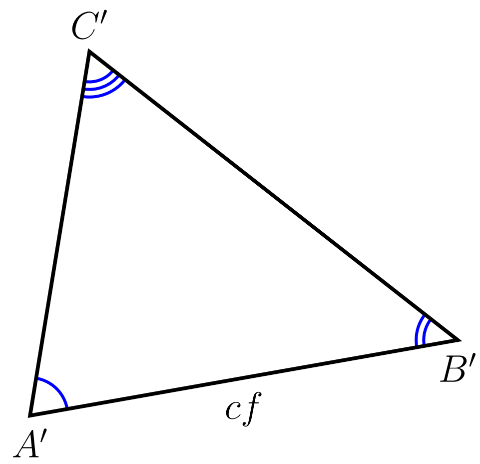
Theorem 14.15 - The AAA Similarity Test: If the three angles of a triangle are equal to the three angles of another triangle, then the two triangles are similar.



Trivial special case: If any side of one triangle happens to have the same length as the corresponding side of the other, then the triangles are congruent by the AAS congruency test, and hence also similar. Now consider the more common case where the triangles are not congruent.

The method presented in section 14.15 requires that we use a scaling factor f to produce $\triangle A''B''C''$, so we need to know how we obtain the value of f. In the proofs of the SSS and SAS Similarity tests f was based on a constant proportion of side lengths mentioned in the Test, but the AAA Similarity Test does not mention any side lengths. For the AAA test it turns out that it doesn't matter which side we choose. For no particular reason, I'll use side AB, which has length c. That is, let f take the value that makes $|A'C'| = cf$. Then, when use the scaling factor f to create $\triangle A''B''C''$, we know that $|A''B''| = |A'B'|$.

Then proceed as in section 14.15.



When we reach the point where we need to prove that $\triangle A'B'C'$ and $\triangle A''B''C''$ are congruent, we invoke the AAS congruency test, using $|A''B''| = |A'B'|$ and any two of the three pairs of corresponding equal angles. With that exception, everything plays out as in section 14.15, allowing us to conclude $\triangle ABC$ and $\triangle A'B'C'$ are similar. This proves the theorem.

The proofs of the previous two sections contain a nice pattern. We used the SSS Congruency Test in the proof of the SSS Similarity Test and we used the SAS Congruency Test in the proof of the SAS Similarity Test. Unfortunately this pattern breaks in this section, where we used the AAS Congruency Test in the proof the AAA Similarity Test. The letters in the test names don't match.

While it makes sense to have congruency test named “AAS”, it does not make sense to have a similarity test with that name. Can you see why?

In the names of the triangle congruency tests, an S denotes a pair of sides of equal length. Having a test name containing a single S makes sense. In the AAS congruency test, the S means one side of a triangle has the same length as the corresponding side of the other triangle, which is useful data to have.

But in the triangle similarity tests, an S denotes the ratio of the lengths of corresponding sides being equal to some other ratio, so there needs to be at least two S's in the name of the test to be meaningful. In the SAS similarity test, the two S's mean the ratio of one pair of corresponding sides is equal to the ratio of a second pair of corresponding sides, which is useful data. By contrast the S in an AAS similarity would be saying that the ratio of one pair of corresponding sides is equal to . . . Well, actually it isn't saying what that ratio is equal to, so it is not conveying any useful information. Hence there is no AAS similarity test.

Unfortunately, there is one extra complication I need to mention for the AAA Similarity Test. Mathematician don't all agree on its name! Some authors call this the AA similarity test rather than the AAA similarity test. I have called it the AAA Similarity test because that label seems more common, though not by a large margin.

Say two triangles both have an angle of size α and an angle of size β. The three angles of a triangle always sum to 180°, so the third angles of both triangles must be $180° - \alpha - \beta$. That is, if two angles of one triangle are equal to two angles of the another triangles, then the third angles are also equal.

So some authors call it the AAA Similarity Test, because it's impossible for two triangles to have *exactly* two pairs of pairs of equal corresponding angles. If you verify that two triangles do have two pairs of equal corresponding angles, they *must* have three pairs of equal corresponding angles.

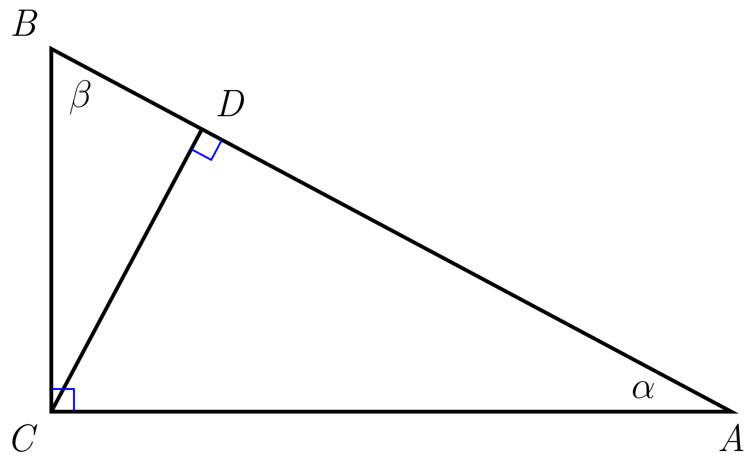
But other authors call this test the AA similarity test, because the smallest set of data that we need to invoke the test is two pairs of equal corresponding angles. That's sufficient data to be sure that they actually have three pairs of equal corresponding angles. Review the part of the proof where we use the AAS Congruency Test to prove that $\triangle A'B'C'$ and $\triangle A''B''C''$ are congruent. I stated that for the the AAS Congruency Test we could use any two of the three pairs of corresponding angles known to be equal. So if we had only been told that the two triangles shared two pairs of corresponding equal angles then we could use those two pairs in the AAS Congruency Test, and the proof remains valid.

Another relevant argument involves consistency between the test names. Consider the AAS Congruency Test. Since that involves two pairs of corresponding equal angles, the third pair must also be equal, but we still call it the AAS congruency test, not the

AAAS Congruency Test. To be consistent with that, we *should* use the name AA Similarity Test rather than AAA Similarity Test.

While that is an entirely valid argument, I think the AAA label is more common for a less logical reason. If we call it the AAA similarity test, then all the congruency and similarity tests are labelled by three letters, which is easier to remember, and looks neater if you happen to list all test labels in a column of a table. Even in mathematics, sometimes cosmetic features win out over pure logic!

14.19 Pythagoras' theorem: alternative proof



We already proved Pythagoras' Theorem in an earlier chapter. Here is an alternative proof using similar triangles.

Let $\triangle ABC$ be a right triangle, the right angle being at C.

Let D be the foot of the perpendicular from C to AB.

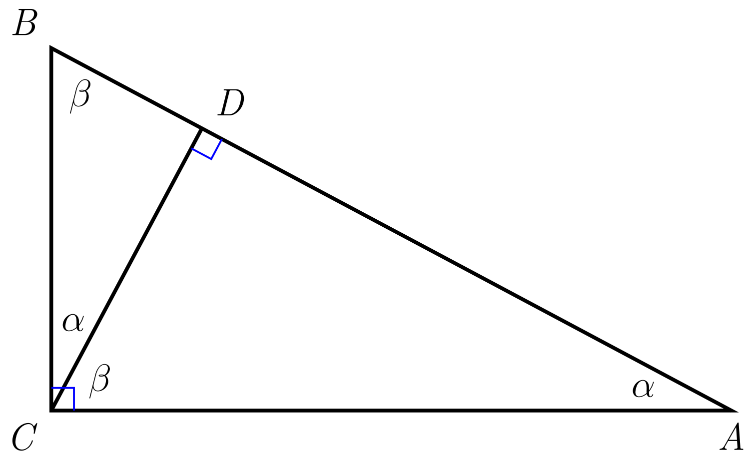
Hence $\angle ACB = \angle ADC = \angle BDC = 90°$.

Let $\angle CAB = \alpha$ and $\angle CBA = \beta$.

Use the angle sum of a triangle to ascertain the size of $\angle BCD$ and $\angle DCA$. Hence identify two triangles similar to $\triangle ABC$. Remember that when labelling similar triangles, the vertices should be listed in corresponding order.

The angle sum of a triangle is 180°. Thus from $\triangle ABC$ we know

$$\alpha + \beta + 90° = 180°.$$



To maintain this equality in $\triangle ACD$ we require $\angle ACD = \beta$, and to maintain it in $\triangle BCD$ we require $\angle BCD = \alpha$.

The AAA test for similarity states that if two triangles have three pairs of equal angles, they are similar triangles. Thus this diagram contains three similar triangles. We were asked to identify triangles similar to $\triangle ABC$, which lists the vertices in the order corresponding to the angles, α, β, right angle, so we should use that same angle order when identifying the other two triangles.

Thus $\triangle ABC$ is similar to $\triangle ACD$ and $\triangle CBD$.

Remember that we are trying to prove Pythagoras' theorem, which refers to the squares of the side lengths of a right triangle. It might be useful if we can use the similar triangles to develop expressions for the squares of $|AC|$ and $|BC|$.

Line segment AC is a side of the similar triangles $\triangle ABC$ and $\triangle ACD$. Corresponding sides of similar triangles have lengths in a constant ratio, so

$$\frac{|AC|}{|AD|} = \frac{|AB|}{|AC|}$$

Rearranging that result gives $|AC|^2 = |AB||AD|$.

That looks useful. Use a similar method to find an expression for $|BC|^2$. Use your result and the result stated above to give an expression for $|AC|^2 + |BC|^2$. If you do this correctly, simplifying your expression will prove the Pythagoras' theorem.

Line segment BC is a side of the similar triangles $\triangle ABC$ and $\triangle CBD$. Corresponding sides of similar triangles have lengths in a constant ratio, so

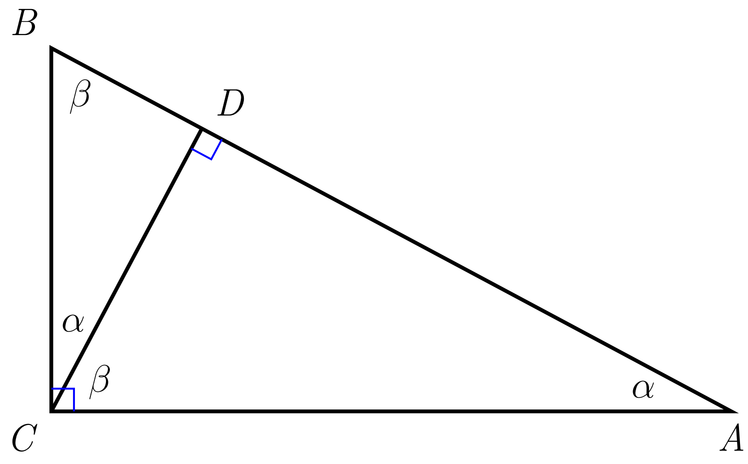
$$\frac{|BC|}{|BD|} = \frac{|BA|}{|BC|}$$

Rearranging that result gives $|BC|^2 = |AB||BD|$.

Combining this with the result from the previous page gives

$$\begin{aligned} |AC|^2 + |BC|^2 &= |AB||AD| + |AB||BD| \\ &= |AB| \times (|AD| + |BD|) \\ &= |AB|^2 \end{aligned}$$

This proves Pythagoras' theorem.



15 Where to next?

This book is incomplete. If I can find time to continue, the next two chapters will be:

- Regular polygon. There is a simple formula relating the area of a regular polygon to its perimeter. This is a useful stepping stone to deriving the formula for the area of a circle.
- Circle

A Types of numbers

A.01 Positive and negative

Positive numbers are those greater than zero. There are also numbers less than zero, called negative numbers. Zero is neither positive nor negative.

When measuring lengths or calculating areas, we only get positive numbers, so there is nothing in this book that requires you to understand how negative numbers work, but terms like "positive integer" will make a bit more sense if you know that negative numbers exist.

For every positive number, there is a corresponding negative number. For example, 42 is a positive number that is 42 greater than zero, so on a number line it would be shown as being 42 units to the right of zero. There is a corresponding negative number that is 42 less than zero. It is written as -42 and is pronounced "minus 42". On a number line it would be 42 units to the left of zero.

In the Celsius temperature scale used in most of the world, zero degrees is usually described as the temperature at which water freezes into ice. (If you study more science, you'll find it's a little more complex than that, but that description works fine for everyday use.) If it snows in your country, you may have experienced temperatures like "minus 5 degrees Celsius", written -5°C, meaning 5 degrees colder than zero degrees. It's also described as "5 degrees *below* zero". Old-fashioned thermometers were hung vertically because they relied on gravity to hold down the temperature-sensitive liquid inside them. Thus the negative temperatures were literally below the positive temperatures on the thermometer's scale.

Early school mathematics courses usually use a number line oriented horizontally, with numbers increasing from left to right. The thermometer scale is effectively a number line, but it is oriented vertically, with numbers increasing as you read up the scale.

(If you are only familiar with digital thermometers you might want to look at the Thermometer Wikipedia article at https://en.wikipedia.org/wiki/Thermometer which includes pictures of older style thermometers.)

A.02 Integers

The set of integers is $\{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$. The integers can be subdivided into:

- The negative integers. We can list them in decreasing order as $\{-1, -2, -3, \ldots\}$, or in increasing order as $\{\ldots, -3, -2, -1\}$. The order of elements in a set is irrelevant, so we can list them in whatever order we think is easiest to read.
- Zero, which is neither negative nor positive.
- The positive integers, $\{1, 2, 3, \ldots\}$.

Two other sets which may be mentioned in Australian primary schools are:

- The cardinal numbers, $\{0, 1, 2, 3, \ldots\}$.
- The natural numbers, $\{1, 2, 3, \ldots\}$, which is identical to the positive integers.

The difficulty with these last two labels is that they are not used consistently internationally. In some regions the set of cardinal numbers is said to exclude zero, so their cardinal numbers are our natural numbers, and they don't use the latter term. Some countries use the term "counting numbers", but that term is also used inconsistently, sometimes including zero and sometimes not. By contrast, the term "positive integers" is used consistently globally.

I want this book to be clear to students worldwide, so I'm going to refer to positive integers rather than natural numbers. That choice made it necessary to include this explanation of what an integer is, but that's the price we pay for international clarity.

A.03 Rational Numbers

The general definition of a rational number is a number that can be expressed as a fraction $\frac{p}{q}$ where p and q are both integers, with $q \neq 0$. That is, a **ratio**nal number can be expressed as the **ratio** of two integers. We need to exclude the case where $q = 0$ because division by zero does not make sense.

In English the word "rational" can also means sensible or reasonable, but that meaning arose from a different source. While a rational person is sensible, a rational number can be expressed as a ratio of integers and is no more or less sensible than the irrational numbers considered later.

Here is another way to explain this distinction. Say someone asks you to translate the word "rational" to some other language. For most languages you would need to ask that person if the context was "rational person" or "rational number" because most languages have two different words for those two different meanings. Only a few languages closely related to English use a single word to covers both contexts.

The labels "positive" and "negative" also apply to rational numbers. Positive rational numbers are greater than zero and negative rational numbers are less than zero.

It is also useful to have a definition of "positive rational number" in the same form as the definition of "rational number" used in the first sentence of this section. Here it is.

A positive rational number is a number which can be expressed as a fraction $\frac{p}{q}$ where p and q are both positive integers.

We don't need to include the restriction $q \neq 0$, since q is already restricted to being a positive integer and zero is not positive.

The definition refers to how a number *can* be expressed, rather than how it *is* expressed. Consider the decimal fraction 2.5. It is not expressed in the $\frac{p}{q}$ form, but it is still a positive rational number because it *can* be expressed as $\frac{5}{2}$, with 5 and 2 both being positive integers. It would still be a positive rational if we wrote it as the mixed fraction $2\frac{1}{2}$. Choosing to express it in a different form does not change the fact that it *can* be expressed $\frac{5}{2}$.

The positive integers $\{1, 2, 3, \ldots\}$ are all positive rational numbers because they can be expressed as the fractions $\{\frac{1}{1}, \frac{2}{1}, \frac{3}{1}, \ldots\}$. Thus the set of positive integers is a subset of the set of positive rational numbers. Similarly, the set of integers is a subset of the set of rational numbers.

If a rational number is expressed as decimal number, it will either:

- terminate at some point, such as the number 42.15, or
- be a recurring decimal, which means that at some point it starts repeating the same digits over and over again forever.

Recurring decimals may be indicated by placing a dot above the first and last digits of the repeating pattern. Another notation is to put a bar above the repeating block, but this notation is rare in English-speaking countries.

The repeating pattern might be a single digit. For example, if you want to express the fraction $\frac{1}{3}$ as a decimal, you can do so by dividing 1 by 3, a process that gives 0.333333. . . , which we write as $0.\dot{3}$ or $0.\overline{3}$.

The repeating pattern may involve multiple digits. For example, if you want to express the fraction $\frac{1}{7}$ as a decimal, you can do so by dividing 1 by 7, a process that gives 0.142857 142857 . . . , which we write as $0.\dot{1}4285\dot{7}$ or $0.\overline{142857}$.

Recurring decimals can be converted back to fractions by multiplying by an appropriate power of 10 and then subtracting the original fraction from that result. Rather than trying to explain this process precisely in words, it's easier to demonstrate the process by examples. I'll use the two examples given above.

$$\begin{aligned} \text{Let } x &= 0.3333\ldots \\ 10x &= 3.3333\ldots \\ 9x &= 3.3333\ldots - 0.3333\ldots = 3 \\ x &= \tfrac{3}{9} = \tfrac{1}{3} \end{aligned}$$

$$\begin{aligned} \text{Let } x &= 0.\dot{1}4285\dot{7} \\ 1,000,000x &= 142,857.\dot{1}4285\dot{7} \\ 999,999x &= 142,857.\dot{1}4285\dot{7} - 0.\dot{1}4285\dot{7} = 142,857 \\ x &= \tfrac{142,857}{999,999} = \tfrac{1}{7} \end{aligned}$$

Sometimes the recurring pattern won't commence immediately after the decimal point. For these, we can deal with the recurring and non-recurring parts separately. Here is an example that uses the fact that we now know that $0.\dot{3} = \frac{1}{3}$.

$$0.40\dot{3} = 0.4 + 0.0\dot{3} = 0.4 + 0.01 \times 0.\dot{3} = \frac{4}{10} + \frac{1}{100} \times \frac{1}{3} = \frac{121}{300}$$

When students first encounter recurring decimals, they sometimes wonder if there is a number $0.\dot{9}$ which is very close to but still less than one. Let's investigate.

$$\begin{aligned} \text{Let } x &= 0.9999\ldots \\ 10x &= 9.9999\ldots \\ 9x &= 9.9999\ldots - 0.9999\ldots = 9 \\ x &= 1 \end{aligned}$$

That is, it makes no sense to refer to $0.\dot{9}$ since it is simpler to write 1. Similarly $0.00\dot{9}$ should just be written as 0.01 and $42.4242\dot{9}$ should just be written as 42.4243. It nevers makes sense to have a recurring decimal where the recurring digits are all 9, though it is possible to have a set of recurring digits that include 9 along with other digits such as $0.\dot{9}2\dot{5}$.

A.04 Irrational Numbers

Irrational numbers are those which are not rational. Some examples: $\sqrt{2}$, $\sqrt{3}$ and $\sqrt{5}$. Unless a positive integer is a perfect square, such as 1, 4 and 9, its square root is irrational. Similarly, unless an integer is a perfect cube such as 1, 8 or 27, its cube root will be irrational.

An irrational number cannot be expressed as a ratio of two integers. When expressed as a decimal, it does not terminate, but it never settles into a pattern that repeats indefinitely.

A.05 Real Numbers

The union of the set of rational numbers and the set of irrational numbers is the set of real numbers.

Return Links: You may have come to this appendix from the section of Chapter 6 that first mentions

- integers
- rational numbers

- irrational numbers and real numbers

B Hypotheses, Postulates and Theorems

B.01 Hypothesis vs theorems

A hypothesis, also known as a conjecture, is a statement that is consistent with existing known data, but which has not yet been proved true or false.

If a hypothesis is proved true, it ceases to be a hypothesis and becomes a theorem. If a hypothesis is proved false, it ceases to be a hypothesis.

One way to disprove a hypothesis is to find a "counterexample", a case where the hypothesis fails. Occasionally it might be possible to salvage something from a failed hypothesis. Perhaps it only works in some smaller but still useful subset of cases. For example, perhaps a hypothesis made some claim about real numbers, but someone discovered a counterexample that involved irrational numbers. If the only counterexamples found relate to irrational numbers, then perhaps we should investigate a new version of the hypothesis that only claims the property to be true for rational numbers. But more often, if a counterexample proves a hypothesis faulty, there is nothing worth salvaging and the hypothesis is discarded.

Some hypotheses are short-lived, being proved or disproved quite quickly. But others survive for centuries. As an extreme example, Goldbach's conjecture was first stated by Christian Goldbach in 1742. It states that every even integer greater than two can be expressed as the sum of two prime numbers. Some examples: $4 = 2 + 2$, $6 = 3 + 3$, $8 = 3 + 5$ and $10 = 5 + 5 = 3 + 7$. The conjecture merely implies there is at least one way to express these numbers as the sum of two primes and sometimes, as with 10, there is more than one solution.

Mathematicians have expended considerable effort trying to prove Goldbach's conjecture true, without success.

The ability to quickly perform mathematical calculations with very large numbers is often regarded as a good test of the computational power of a computer, so computer scientists, often in collaboration with computer manufacturers, have spent considerable effort trying to find a counterexample to Goldbach's conjecture, also without success. As of 2024, they have tested even integers greater than two up to 4×10^{18} and have successfully expressed every one of those numbers as the sum of two primes.

Given the number of cases successfully tested, it would be truly surprising if further calculations do find a counterexample, so mathematicians regard Goldbach's conjecture as almost certainly true, but since it is still unproven, it remains a hypothesis/conjecture rather than a theorem.

The distinction between hypothesis and theorem is based on what the whole mathematical community has proven, not on what you and I have proven. In this book I usually clearly state a theorem before proving it. When I do this, I do *not* label it as a hypothesis, then show the proof, and then change its label from hypothesis to theorem after completing the proof. This book contains no new discoveries. All the theorems in this book were proved many centuries ago. Thus they should be identified as theorems, even at a point in the book when I haven't yet shown how to prove them.

B.02 Circular arguments

Most proofs work by taking some existing known theorems and combining them in clever ways to a produce a new theorem.

In doing this you need to be careful to not introduce any "loops" in your sequence of proofs. For example, say your proof of Theorem A employed some other theorems including Theorem B, but your proof of Theorem B employed some other theorems including Theorem A. Then you haven't actually proved anything. Your first proof is saying "If I can prove Theorem B is true then Theorem A is true" while your second proof says "If I can prove Theorem A is true then Theorem B is true," so you have not proved either. If you combine those two quoted sections, you are essentially saying "If I can prove Theorem A is true then theorem A is true," which is clearly of no value at all.

In that example the loop involved two theorems, but longer loops are also possible and just as faulty. If the proof of Theorem A relies on Theorem B, the proof of Theorem B relies on Theorem C and the proof of theorem C relies on Theorem A, then we haven't proved anything.

The traditional way of avoiding such "circular arguments" is to present your proofs in a linear order, and enforce a requirement that proofs may only employ theorems proved earlier in the sequence.

B.03 Theorems vs postulates

So can *every* theorem be proved using earlier theorems?

Well, no. We need to start somewhere. There has to be a first theorem, and it has no earlier theorems to use in its proof.

A proof cannot magically produce something out of nothing. A proof takes some inputs and combines them in a logical manner to prove a new theorem. For *most* theorems those inputs are other simpler theorems, but since that won't work for the first theorem there must also be other things that can be used as inputs. Those other things are called postulates.

A postulate is a statement which we take to be true, but which we will make no attempt to prove. Theorems are proved; postulates are not. Our first theorem has no other theorems available to use, so the proof of the first theorem will combine two or more postulates in a clever way to prove the first theorem. For every theorem after the first, the proof of the theorem can employ both postulates and theorems that have already been proved. For example, if we number our theorems as we prove them, theorem 20 can take as its inputs any postulates and any of the first 19 theorems.

I mentioned earlier that *most* theorems only combine other theorems. A fairly common pattern is that early theorems are very likely to directly use postulates, while later theorems are more likely to only directly refer to earlier theorems. However, *all* theorems do rely on the validity of the postulates, perhaps indirectly.

For example, it's possible that theorem 20 didn't *directly* use any postulates. It's possible its inputs were only theorem 4 and theorem 19. But those two theorem also rely on earlier theorems and/or postulates. If the inputs to theorem 19 were theorem 6 and theorem 18, then theorem 20 also relies indirectly on those two theorems. If we track the chain of dependencies all the way back through the earlier theorems, we ultimately find that *all* theorems rely on postulates. This is really just another way of saying something I already said earlier. We can't prove everything. Somewhere in our system of proofs and theorems there is a theorem 1 which has no earlier theorem to rely on, so it must rely on things that are not proved, the things we call postulates.

Postulates are the foundation on which we build all our theorems, so we want our postulates to be believable. We've got a problem if someone can point to one of our postulates and say: "That doesn't look sensible to me so I don't believe it." If there is a single postulate they don't believe then they also won't believe our set of theorems that relied on that postulate. In an ideal world, all our postulates would be self-evident statements.

Axiom is a synonym for postulate. Many now regard axiom as an archaic term and prefer the label postulate. This is curious, since the approach of setting out a set of unproven axioms/postulates and then deriving theorems from them is still called the axiomatic approach.

B.04 Postulates sets are not unique

If you read geometry textbooks by different authors you may find they use different postulates.

Sometimes when this happens, if you compare the postulates more carefully you might find the difference is only that the different authors have used slightly different words to express the same idea. The authors are all trying to express the same idea as clearly as they can, but authors often disagree as to what constitutes the clearest expression.

But sometimes it isn't just the same idea in different words. Sometimes the different postulates are different ideas. The issue here is that under the axiomatic approach to geometry, the set of postulates is not unique. That is, there are many different ways to select a set of postulates which will allow us to derive all the useful theorems of school geometry.

So how do authors choose which set of postulates to use?

Some mathematicians have tried to determine the simplest set of postulates required, which doesn't give a unique answer because they disagree as to what "simplest" means.

Others have tried to find the smallest number of postulates required. Unfortunately, this approach tends to result in postulates which are rather abstract and hard to understand, which clashes with the idea that postulates should be self-evident. Another failing of this approach is that the postulates are not very powerful, which means it can require a lot of proofs to get from these postulates to any theorems that have practical applications. Thus this approach tends to be only used in advanced university courses.

For school geometry courses, and even for early university courses, it is more useful to adopt a larger set of postulates, all of which are clearly reasonable. This approach allows us to start deriving practical theorems far more quickly.

Sets of postulates like these might be described as "not minimal" or the postulates in the set might be described as "not independent". This means that it is possible to use some of the postulates in the set to prove one of the other postulates, so that other postulate should really be called a theorem. However, such a proof is probably so long and difficult that we don't think it's worth attempting at this level, particularly so when the resulting theorem is a statement that appears so sensible that most of us will happily accept it as a postulate.

Like pretty much every geometry book ever written for school students, the postulates used in this book are not minimal.

Return Link: You might have come to this appendix from a link at the start of chapter 4

C Pronumerals with subscripts

This appendix has been written for any readers who are not familiar with using pronumerals with subscripts.

Attaching subscripts to pronumerals is useful when you have a large number of pronumerals.

Say you are given a problem about the heights of three adjacent buildings. You are given several facts about the heights, such as the sum of the three heights, and the difference between the heights of the middle building and each of the other two buildings. Your task is to use this data to find the height of each building. The first thing to do is to define pronumerals denoting the heights of the three buildings.

Let the heights of the three building from left to right be a metres, b metres and c metres.

Now lets change the problem so that instead of three building there are 20. We could use the pronumerals a to t for the 20 heights. If the problem gives us some information about the 11^{th} building, we have to count our way through the alphabet to figure out that the relevant pronumeral is k. That seems horribly inefficient.

What happens if I extended the problem to 40 buildings? If we need 40 pronumerals we run out of lower case letters, and have to start using upper case letters or perhaps Greek letters.

When we have to deal with a large number of pronumerals, we need a more efficient notation. The solution is to use pronumerals with subscripts.

The example above referred to building heights, so h would be a suitable base letter, so we'll add subscripts to that letter to produce 40 pronumerals. We can define all 40 pronumerals in a single sentence.

Let the height of the i^{th} building be h_i metres, $i = 1, 2, 3, \ldots, 40$.

Here are some examples of the notation in use.

- The height of the first building is h_1 metres.
- The height of the 7^{th} building is h_7 metres.

- The 2^{nd} last building is the 39^{th} so it has height h_{39} metres.
- The sum of the heights of the 40 buildings in metres is $h_1 + h_2 + h_3 + \ldots + h_{40}$.

If you have experience with computer programming, this approach may feel similar to using arrays. If we had to write a program to perform calculations with 40 building heights, we wouldn't create 40 different variables, each with a different name. We could create an array called perhaps building_height which contains 40 elements.

The advantages of subscripted pronumerals include the following.

- You never have to worry about running out of letters if you need more than 26 pronumerals.
- If we are given some data about the 11^{th} building, we don't have to count our way through the alphabet to ascertain the relevant pronumeral. We immediately know it is h_{11}.
- They can even cope with the more complex scenario where the number of different pronumerals we require is not a fixed number but is rather determined by some other pronumeral.

I'll illustrate that final point with a scenario from Chapter 5. In that chapter we derive some theorems about joining rectangles that have equal height, h.

Initially we consider joining just two rectangles and we let their widths be a and b. Then we extended the theorem to cover three rectangles, with widths a, b and c. But now we come to the tricky part. The next step is to generalise the theorem so that it works for two or more rectangles. That is, the theorem will refer to n rectangles, where n can be any number selected from $2, 3, 4, \ldots$

What pronumerals can we use given that we will need n different symbols for the widths of the n rectangles?

> The first time students see this problem, some suggest using $a, b, c, \ldots, n$. Does that work?

No. If you count through the alphabet, you'll find this gives 14 different symbols, not n.

There's also a second more subtle problem with this approach. In the sequence of pronumerals $a, b, c, \ldots, n$, the 8^{th} pronumeral is h, and I mentioned earlier that we're using h to denote the common height of the rectangles. That is, whenever $n \geq 8$, we'd be requiring that the 8^{th} rectangle be a square! Now all squares are rectangles, so if we have a general theorem about n rectangles, the theorem will remain valid if the 8^{th} rectangle happens to be a square, but we don't want to state the theorem is a way that *requires* that the 8^{th} rectangle always be a square.

If you need n different pronumerals, the correct approach is to use pronumerals with subscripts.

Here the n pronumerals denote widths, so w is a good choice for the letter. Then our n pronumerals are $w_1, w_2, w_3, \ldots, w_n$. This can also be written: w_i where $i \in \{1, 2, 3, \ldots, n\}$. Here are some examples of the notation in use.

- The first rectangle has width w_1.
- When $n \geq 7$, meaning there are at least seven rectangles, the 7^{th} rectangle has width w_7.
- The 2^{nd} last of the n rectangles will be the $(n-1)^{th}$ rectangle and it has width w_{n-1}. That's a general symbol that works for any value of n. If we are considering a particular value of n we can be more specific. For example, if $n = 10$, meaning we have 10 rectangles, then w_{n-1} simplifies to w_9. That is, the 2nd last of 10 rectangles is the 9^{th} rectangle.
- The sum of the widths of the n rectangles is $w_1 + w_2 + w_3 + \ldots + w_n$.

When listing a sequence of numbers that includes an ellipsis, that is the $\ldots$ dots, we list enough numbers to make the pattern clear, and tradition says we use at least 3 numbers before the ellipsis. For example, if we write $1, 2, 3, \ldots, 10$ we are referring to the first 10 positive integers, and $2, 4, 6, \ldots, 20$ refers to the first 10 positive even numbers. Sometimes three numbers isn't enough to clarify the pattern, particularly when there is no final number to give an extra hint. Writing $0, 1, 4, \ldots$ is a little confusing. It would be clearer to write $0, 1, 4, 9, 16, \ldots$ Also, if nothing in the preceding text gave the reader any hints that square numbers were going to arise, then the first time you use that notation it would be better to say something like "the square numbers $0, 1, 4, 9, 16, \ldots$," to make the meaning clear.

The tradition is to list at least 3 terms before the ellipsis, so I listed the n pronumerals as $w_1, w_2, w_3, \ldots, w_n$. But there's a surprise here. The pronumeral w_3 might not exist! I mentioned above that $n \in \{2, 3, 4, \ldots\}$. When $n = 2$ the list $w_1, w_2, w_3, \ldots, w_n$ has to be read as meaning just w_1, w_2. There is no w_3. Also, when $n = 3$ the third term w_3 is also the last term w_n, so $w_1, w_2, w_3, \ldots, w_n$ means w_1, w_2, w_3.

You might feel the above issues mean that the notation $w_1, w_2, w_3, \ldots, w_n$ is inherently flawed. I would agree with you. Unfortunately this notation has been in use for centuries so we're stuck with it.

The expression for the sum of the widths shown above has the same flaw. When $n = 2$, w_3 does not exist and $w_1 + w_2 + w_3 + \ldots + w_n$ means $w_1 + w_2$.

An aside: If you don't know how to use summation notation, just ignore this paragraph. It takes a lot of practice to become comfortable with summation notation, and this book has so few places where it is useful that I'm not going to explain how it works. If you are familiar with it, you can write the sum of the widths as $\sum_{i=1}^{n} w_i$, a notation that avoids the flaw. It would be wonderful if mathematicians recognised something like summation notation for listing rather than summing pronumerals!

Return Link: You might have come to this appendix from a link in the section that uses mathematical induction to prove a result about joining two or more rectangles.

D A variation of the Area Sum Postulate

Section 4.02 introduced the Area Sum Postulate.

Postulate 4.2 — The Area Sum Postulate: If a region is divided into two or more smaller regions, the area of the original region is equal to the sum of the areas of the smaller regions.

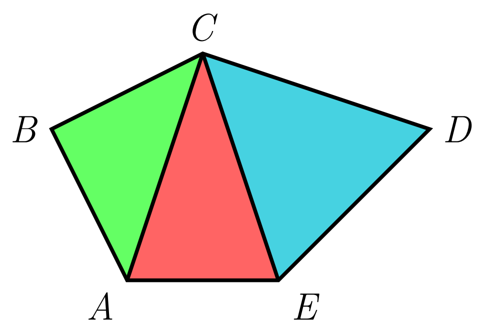
Some mathematicians like to keep their postulates as simple as possible. They might replace the above postulate by the simpler Postulate D.1 shown below, that subdivides a region into two regions, and then use that postulate to prove theorem D.1 refers to subdividing a region into two or more regions.

Postulate D.1 — The Area Sum Postulate: If a region is divided into two smaller regions, the area of the original region will be equal to the sum of the areas of the two smaller pieces.

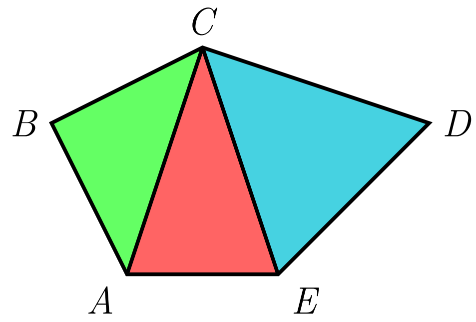
Theorem D.1 — The Area Sum Theorem: If a region is divided into two or more smaller regions, the area of the initial region is equal to the sum of the areas of those smaller regions.

Most authors of geometry textbooks are happy to use the stronger Postulate 4.2, because their readers have no trouble believing it, and because doing so bypasses the difficult proof of Theorem D.1. That's a sensible approach, so I adopted it in the main text. However, the alternative approach does contain an interesting use of mathematical induction, so I included it in this appendix for any interested readers.

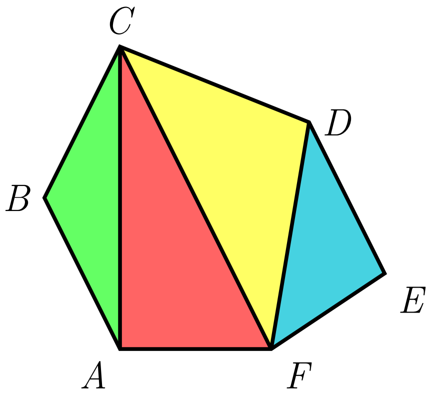
In chapter 4 we used Postulate 4.2 to justify the claim that the area of this pentagon can be found as the sum of the areas of the three triangles. How would you justify this conclusion if we can only use the weaker Postulate D.1?



We simply use the postulate twice. First, subdivide the pentagon into a triangle and a quadrilateral. Then, subdivide the quadrilateral into two triangles. Applying the postulate at each of those two steps gives:



$$\begin{aligned}&\text{Area of Pentagon } ABCDE\\&= \text{Area of } \triangle ABC + \text{Area of Quadrilateral } ACDE\\&= \text{Area of } \triangle ABC + \text{Area of } \triangle ACE + \text{Area of } \triangle CDE\end{aligned}$$



Similarly, if we have a hexagon subdivided into 4 triangles as shown here, we can use the postulate three times, giving:

$$\begin{aligned}&\text{Area of Hexagon } ABCDEF\\&= \text{Area of } \triangle ABC + \text{Area of Pentagon } ACDEF\\&= \text{Area of } \triangle ABC + \text{Area of } \triangle ACF + \text{Area of Quadrilateral } CDEF\\&= \text{Area of } \triangle ABC + \text{Area of } \triangle ACF + \text{Area of } \triangle CDF + \text{Area of } \triangle DEF\end{aligned}$$

It feels like this argument can be extended to cover as many regions as required. We can prove this formally using the principle of mathematical induction.

Let $S(n)$ be the statement: If a region is divided into n smaller regions, the area of the original region will be equal to the sum of the areas of the n smaller regions.

To make it easier to refer to the various regions, let's label the initial region R and number the smaller regions as region 1 through to region n.

We are about to encounter some pronumerals with subscripts. If you have not encountered this notation before, refer to Appendix C for an introduction to the concept.

Some people prefer to write the statement they are trying to prove more succinctly in an equation. To do that, we can define the symbol:

$$A_i = \text{ Area of region } i$$

Then we can write our statement as

$$S(n) : A_R = A_1 + A_2 + A_3 + \ldots + A_n$$

If you are reading that aloud, the colon can be read “states that” or “is the statement”.

Since we are dividing region R into smaller pieces, the smallest value of n that makes sense is $n = 2$. Hence we hope to prove $S(n)$ true for $n = 2, 3, 4, \ldots$

Try to complete the initial step, proving that the statement $S(2)$ is true. Remember that you can use Postulate D.1, but not Postulate 4.2

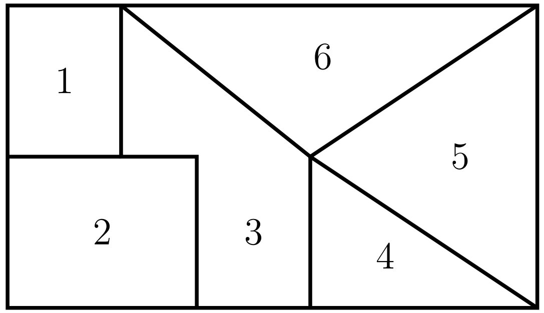
The initial step: If your reread statement $S(n)$ with every occurrence of n replaced by 2, then it refers to dividing region R into two smaller regions, region 1 and region 2, and it claims $A_R = A_1 + A_2$. It is saying the same thing as Postulate D.1. That is, we can simply write: By postulate D.1, $S(2)$ is true.

That was easy! When using mathematical induction, the initial step usually is easy. It's the induction step that can be challenging.

For the induction step, we are going to consider a region R divided into $k+1$ smaller regions. How do we draw a diagram of this? It's easy to show region R divided into say 6 regions numbered 1 to 6, but how do we show it divided into $k+1$ regions? Any diagram that shows and numbers *all* the resulting smaller regions is going to be assuming a particular value of k, where we want k to be a general pronumeral. If the diagram is to work for the general case, it's going to have to make some claim about how many smaller regions there are in some larger region but not actually show the individual smaller regions!

This idea can be confusing the first time you see it, so before looking at the general induction step, let's look at a particular case. Say we want to prove that if $S(5)$ is true, then $S(6)$ must be true. To do this I'll use two diagrams.

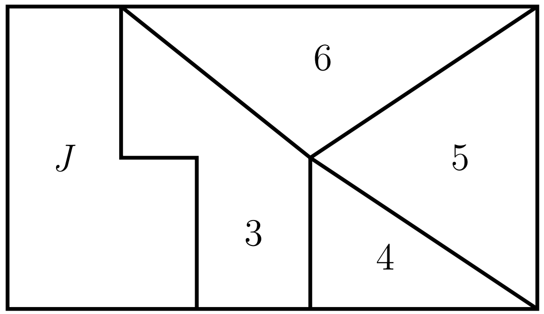
Region R



The first diagram shows a region R, divided into 6 smaller regions. I've chosen to make Region R a rectangle, but you can choose any shape you like. From the 6 smaller regions, choose any two regions that share a common border, and number them region 1 and region 2. Number the remaining regions 3 to 6 in any order you like.

To make the second diagram, take a copy of the first diagram, but erase the common border between regions 1 and 2, so they become a single new region which I will name region J.

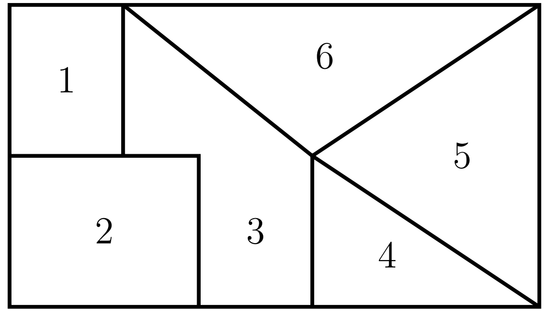
Region R



State with reasons how the area of region J compares to the areas of regions 1 and 2. If $S(5)$ is true, what can you say about the area of region R in this second diagram? Prove that if $S(5)$ is true, then $S(6)$ must be true.

Comparing the two diagrams, region J combines regions 1 and 2, so by postulate 3.2, the area of region J is the sum of the areas of regions 1 and 2.

Region R

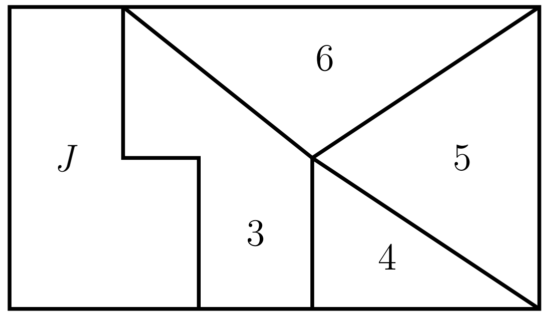


$$A_J = A_1 + A_2$$

In the second diagram, region R is divided into 5 smaller regions: region J and regions 3 to 6. $S(5)$ states that if a region is divided into 5 smaller regions, its area is the sum of the areas of those 5 regions. That is, if $S(5)$ is true, then

$$A_R = A_J + A_3 + A_4 + A_5 + A_6$$

Region R



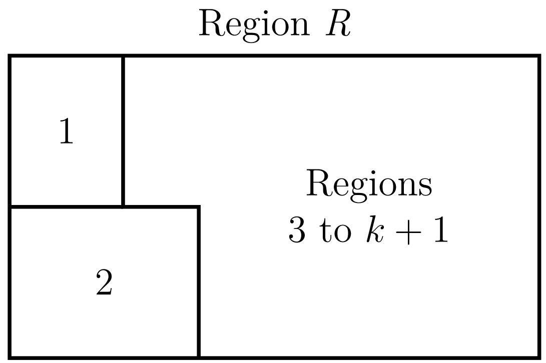
Substituting the earlier expression for A_J into this equation gives

$$A_R = A_1 + A_2 + A_3 + A_4 + A_5 + A_6$$

That is, the area of region R is equal to the sum of the areas of the 6 regions in the first diagram. This shows that if $S(5)$ is true, then $S(6)$ is true.

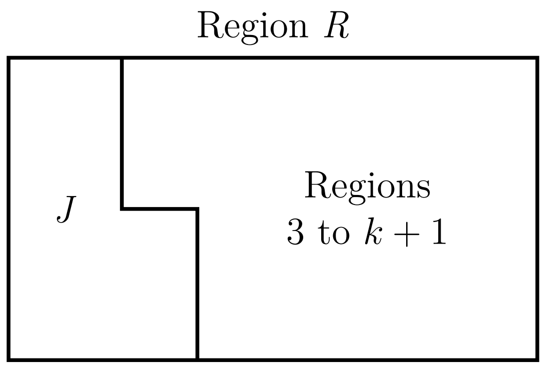
Now we try to generalise the above argument to prove the induction step. That is we now want to prove that if $S(k)$ is true, then $S(k+1)$ is true. While the previous argument did this for the case where $k = 5$, we now want to write a general argument that will work for any $k \in \{2, 3, 4, \ldots\}$. As before we need two diagrams.

Region R is now split into $k+1$ smaller regions. From these $k+1$ regions, choose any two regions that share a common border, and number them region 1 and region 2. The other regions can be numbered 3 to $k+1$ in any order.



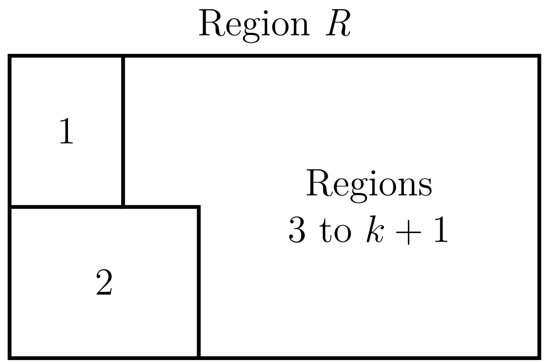
In the first diagram, draw region R and regions 1 and 2. We want this diagram to be valid for any $k \in \{2, 3, 4, \ldots\}$, but there's no way to draw all the other smaller regions individually without assuming k to be some particular number. So we don't drawn them individually. Instead we just label the remaining part of region R as containing "Regions 3 to $k+1$" with no subdivision shown.

As before, to make the second diagram, take a copy of the first diagram, but erase the common border between regions 1 and 2, so they become a single new region labelled region J.



State with reasons how the area of region J compares to the areas of regions 1 and 2. If $S(k)$ is true, what can you say about the area of region R in this second diagram? Complete the induction step. That is, prove that if $S(k)$ is true, then $S(k+1)$ must be true.

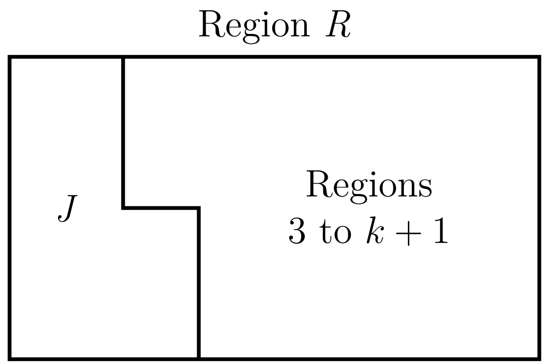
Comparing the two diagrams, region J combines regions 1 and 2, so by postulate 3.2, the area of region J is the sum of the areas of regions 1 and 2.



$$A_J = A_1 + A_2$$

In the second diagram, region R is divided into k pieces: region J and regions 3 to $k+1$. $S(k)$ tells us that when a region is divided into k pieces, its area is the sum of the areas of those k pieces. That is, if $S(k)$ is true, then

$$A_R = A_J + A_3 + A_4 + A_5 + \ldots + A_{k+1}$$



Substituting the earlier expression for A_J into this equation gives

$$A_R = A_1 + A_2 + A_3 + A_4 + A_5 + \ldots + A_{k+1}$$

That is, the area of region R is equal to the sum of the areas of the $k+1$ pieces in the first diagram. This shows that if $S(k)$ is true, then $S(k+1)$ is true, which completes the induction step. Hence by the Principle of Mathematical Induction, $S(n)$ is true for $n = 2, 3, 4, \ldots$

This proves Theorem D.1

The diagrams I used in the previous proof show one shape for region R and one possible position and shape for regions 1 and 2. You should verify that the method remains valid for other ways the diagram can be drawn. Some examples follow.

- My region R was a rectangle. Try other shapes.
- My diagram has placed regions 1 and 2 so that they share borders with region R. Try a scenario where regions 1 and 2 do not share any borders with region R. Regions 1 and 2 do still need to share a border with each other, since the proof relies on being able to erase their common border so they become a single region.

- My diagram only includes line segments, so all the regions are polygons. Try including some curves in your diagram. You will find the proof remains valid when some regions have curved borders, such as circles and ellipses, and also for shapes with a mixture of curved and straight borders, such as sectors or segments of a circle. For example, if a circle is subdivided into two or more sectors, the proof given above is still valid, so the area of the circle will equal the sum of the areas of the sectors.

Return Link: You might have come to this appendix from a link in the section on the Area Sum Postulate

E Converse

Let P and Q be two statements which can be true or false.

The statement $P \Rightarrow Q$, pronounced "P implies Q", means that if P is true then Q is true. We call the statement $P \Rightarrow Q$ a logical implication. Here is an example:

P: Quadrilateral $ABCD$ is a square. This is pronounced "P is the statement: Quadrilateral $ABCD$ is a square."

Q: Quadrilateral $ABCD$ is a rectangle.

At the moment, we no nothing about quadrilateral $ABCD$. We don't know whether it is a square or a rectangle or neither, so we don't know whether statements P and Q are true or false. But we do know that all squares are rectangles, so if quadrilateral $ABCD$ is a square it is also a rectangle. That is, if statement P is true, then statement Q must also be true. Hence it is valid to write $P \Rightarrow Q$.

The converse of the logical implication $P \Rightarrow Q$ is the logical implication $Q \Rightarrow P$. In general, if the logical implication $P \Rightarrow Q$ is true, then we cannot conclude $Q \Rightarrow P$ is true. That is $P \Rightarrow Q$ does not mean that $Q \Rightarrow P$.

In our example, all squares are rectangles, so $P \Rightarrow Q$. But not all rectangles are squares, so the statement $Q \Rightarrow P$ is invalid. If someone tells us Q is true then we know quadrilateral $ABCD$ is a rectangle. It's *possible* that it is also a square but it's also possible that it is not a square, so we don't know whether P is true or false and hence we can't claim that $Q \Rightarrow P$.

There are some pairs of statements P and Q for which the logical implication $P \Rightarrow Q$ and the converse $Q \Rightarrow P$ are both true. Then we can write $P \Leftrightarrow Q$. This means that P is true if and only if Q is true.

As an example, we'll keep P the same as the previous example, but alter Q.

P: Quadrilateral $ABCD$ is a square.

Q: Quadrilateral $ABCD$ is a rectangle in which all sides have equal length.

A rectangle with all sides of equal length *is* a square. Thus P is true if and only if Q is true. P and Q are effectively equivalent statements, so $P \Leftrightarrow Q$ may be pronounced as "P is equivalent to Q".

We still don't have any information about the shape of quadrilateral $ABCD$, so we still don't know whether P and Q are true or false, but we do know that they are either both true or both false.