(ii) Conclude that, if the angles in $\triangle ABC$ add to a straight angle, then $\angle ACB$ is a right angle. $\triangle$

Once we introduce the parallel criterion, and hence can prove that the three angles in any triangle add to a straight angle, Problem **141** will guarantee that "the angle subtended on the circumference by a diameter is always a right angle".

Problem 142 Show how to implement the basic ruler and compasses constructions:

(i) to construct the midpoint M of a given line segment $\underline{AB}$;

(ii) to bisect a given angle $\angle ABC$;

(iii) to drop a perpendicular from P to a line AB (that is, to locate X on the line AB, so that the two angles that PX makes with the line AB on either side of PX are equal).

Prove that your constructions do what you claim. $\triangle$

Problem 143 Given two points A and B.

(a) Prove that each point X on the perpendicular bisector of $\underline{AB}$ is equidistant from A and from B (that is, that $\underline{XA} = \underline{XB}$).

(b) Prove that, if X is equidistant from A and from B, then X lies on the perpendicular bisector of $\underline{AB}$. $\triangle$

Problem **143** shows that, given a line segment $\underline{AB}$, the perpendicular bisector of $\underline{AB}$ is the locus of all points X which are equidistant from A and from B. This observation is what lay behind the construction of the circumcentre of a triangle (back in Chapter 1, Problem **32**(a)):

> Given any $\triangle ABC$.
> Let O be the point where the perpendicular bisectors of $\underline{AB}$ and $\underline{BC}$ meet.
> Then $\underline{OA} = \underline{OB}$
> and $\underline{OB} = \underline{OC}$.
> $\therefore \underline{OA} = \underline{OB} = \underline{OC}$.
> Hence O is the centre of a circle passing through all three vertices A, B, C.
> Moreover O also lies on the perpendicular bisector of $\underline{CA}$.

This circle is called the *circumcircle* of $\triangle ABC$, and O is called the *circumcentre* of $\triangle ABC$. As indicated back in Problem **32**, the radius of the circumcircle of $\triangle ABC$ (called the *circumradius* of the triangle) is generally denoted by R. Later we will meet other circles and "centres" associated with a given triangle $\triangle ABC$.

Before moving on it is worth extending Problem **143** to three dimensions.

Problem 144 Given any two points N, S in 3D space, prove that the locus of all points X which are equidistant from N and from S form the plane perpendicular to the line NS and passing through the midpoint M of $\underline{NS}$. $\triangle$

The next two fundamental results are often neglected.

Problem 145 Given any $\triangle ABC$, if we extend the side $\underline{BC}$ beyond C to a point X, then the "exterior angle" $\angle ACX$ at C is greater than each of the "two interior opposite angles" $\angle A$ and $\angle B$. $\triangle$

Problem 146

(a) If in $\triangle ABC$ we have $\underline{AB} > \underline{AC}$, then $\angle ACB > \angle ABC$. ("In any triangle, the larger angle lies opposite the longer side.")

(b) If in $\triangle ABC$ we have $\angle ACB > \angle ABC$, then $\underline{AB} > \underline{AC}$. ("In any triangle, the longer side lies opposite the larger angle.")

(c) **(The triangle inequality)** Prove that in any triangle $\triangle ABC$,

$$\underline{AB} + \underline{BC} > \underline{AC}.$$

$\triangle$

The results in Problems **145** and **146** have surprisingly many consequences. For example, they allow one to prove the converse of the result in Problem **141**.

Problem 147 Suppose that in $\triangle ABC$, $\angle C = \angle A + \angle B$. Prove that C lies on the circle with diameter $\underline{AB}$.

(In particular, if the angles of $\triangle ABC$ add to a straight angle, and $\angle ACB$ is a right angle, then C lies on the circle with diameter $\underline{AB}$.) $\triangle$

We come next to a result whose justification is often fudged. At first sight it is unclear how to begin: there seems to be so little information to work with – just two points and a line through one of the points.

Problem 148 A circle with centre O passes through the point P. Prove that the tangent to the circle at P is perpendicular to the radius $\underline{OP}$. △

Problem **148** is an example of a result which implies its own converse – though in a backhanded way. Suppose a circle with centre O passes through the point P. If OP is perpendicular to a line m passing through P, then m must be tangent to the circle (because we know that the tangent at P is perpendicular to OP, so the angle between m and the tangent is "zero", which forces m to be equal to the tangent). This converse will be needed later, when we meet the *incircle*.

Problem 149 Let P be a point and m a line not passing through P. Prove that, among all possible line segments $\underline{PX}$ with X on the line m, a perpendicular from P to the line m is the shortest. △

The result in Problem **149** allows us to define the "distance" from P to the line m to be the length of any perpendicular from P to m. (As far as we know at this stage of the development, there could be more than one perpendicular from P to m.)

Note that all the results mentioned so far have avoided using the Euclidean "parallel criterion" (or – equivalently – the fact that the three angles in any triangle add to a straight angle). So results proved up to this point should still be "true" in any geometry where we have points, lines, triangles, and circles satisfying the congruence criteria – whether or not the geometry satisfies the Euclidean "parallel criterion".

The idea that there is only one "shortest" distance from a point to a line may seem "obvious"; but it is patently *false* on the sphere, where *every* line (i.e. 'great circle') from the North pole P to the equator is perpendicular to the equator (and all these lines have the same "length"). The proof that there is just one such perpendicular from P to m depends on the *parallel criterion* (see below) – a criterion which fails to hold for geometry on the sphere.

Euclid's *Elements* started with a few basic axioms that formalised the idea of ruler and compasses constructions. He then added a simple axiom that allowed one to compare angles in different locations. He made the forgivable mistake of omitting an axiom for congruence of triangles – imagining that it can be *proved*. (It can't.) However he then stated, and carefully developed the consequences of, a much more subtle axiom about parallel lines (two lines m, n in the plane are said to be *parallel* if they never meet, no matter how far they are extended). For reasons that remain unclear, instead of appreciating that Euclid's "parallel postulate" constituted a profound insight into the foundations of geometry, mathematicians in later ages saw the complexity of Euclid's postulate as some kind of flaw, and so tried to show that it could be derived from the other, simpler postulates. The attempt to "correct" this perceived flaw became a kind of Holy Grail.

The story is instructive, but too complicated to summarise accurately here. The situation was eventually clarified by two nineteenth century mathematicians (more-or-less at the same time, but working independently). In the revolutionary, romantic spirit of the nineteenth century, János Bolyai (Hungarian: 1802–1860) and Nikolai Lobachevski (Russian: 1792–1856) each allowed himself to consider what would happen if one adopted a *different* assumption about how "parallel lines" behave. Both discovered that one can then derive an apparently coherent theory of a completely novel kind, with its own beautiful results: that is, a geometry which seemed to be internally "consistent" – but different from Euclidean geometry. Lobachevski published brief notes of his work in 1829–30 (in Kazan); Bolyai knew nothing of this and published incomplete notes of his researches in 1832. Lobachevski published a more detailed booklet in 1840.

Neither mathematician got the recognition he might have anticipated, and it was only much later (largely after their deaths) that others realised how to show that the fantasy world they had each dreamt up was just as "internally consistent" as traditional Euclidean geometry. The story is further complicated by the fact that the dominant mathematician of the time – namely Gauss (1777–1855) – claimed to have proved something similar (and he may well have done so, but exactly what he knew has to be inferred from cryptic remarks in occasional letters, since he published nothing on the subject). If there is a moral to the story, it could be that success in mathematics may not be recognised, or may only be recognised after one's death: so those who spend their lives exploring the mathematical universe had better appreciate the delights of the mathematical journey, rather than being primarily motivated by a desire for immediate recognition and acclaim!

Two lines m, n in the plane are said to be *parallel* if they never meet – no matter how far they are extended. We sometimes write this as "$m \parallel n$".

Given two lines m, n in the plane, a third line p which crosses both m and n is called a *transversal* of m and n.

> **Parallel criterion**: Given two lines m and n, if some transversal p is such that the two "internal" angles on one side of the line p (that is the two angles that p makes with m and with n, and which lie between the two lines m and n) add to *less than* a straight angle, then the lines m and n must meet on that side of the line p.

If the internal angles on one side of p add to *more than* a straight angle, then internal angles on the *other side* of p add to less than a straight angle, so the lines m and n must meet on the other side of p. It follows

- that two lines m and n are parallel precisely when the two internal angles on one side of a transversal add to **exactly** a straight angle.

Parallel lines can be thought of as "all having the same direction"; so it is convenient to insist that "every line is parallel to itself" (even though it has lots of points in common with itself). It then follows

- that, given three lines k, m, n, if k is parallel to m and m is parallel to n, then k is parallel to n; and
- that given a line m and a point P, there is a unique line n through P which is parallel to m.

All this then allows one

- to conclude that, if m and n are any two lines, and p is a transversal, then m and n are parallel if and only if *alternate angles* are equal (or equivalently, if and only if *corresponding angles* are equal); and
- to extend the basic ruler and compasses constructions to include the construction:

  > "given a line AB and a point P,
  > construct the line through P which is parallel to AB"

  (namely, by first constructing the line PX through P, perpendicular to AB, and then the line through P, perpendicular to PX).

One can then prove the standard result about the angles in any triangle.

> **Claim** The angles in any triangle $\triangle ABC$ add to a straight angle.
>
> **Proof** Construct the line m through A that is parallel to BC. Then AB and AC are transversals, which cross both the line m and the line BC, and which make three angles at the point A on m:
>
> - one being just the angle $\angle A$ in the triangle $\triangle ABC$,
> - one being equal to $\angle B$ (alternate angles relative to the transversal AB) and
> - one being equal to $\angle C$ (alternate angles relative to the transversal AC).
>
> The three angles at A clearly add to a straight angle, so the three angles $\angle A$, $\angle B$, $\angle C$ also add to a straight angle. QED

Once we know that the angles in any triangle add to a straight angle, we can prove all sorts of other useful facts. One is a simple reformulation of the above **Claim**.

Problem 150 Given any triangle $\triangle ABC$, extend $\underline{BC}$ beyond C to a point X. Then the exterior angle

$$\angle XCA = \angle A + \angle B.$$

("In any triangle, each exterior angle is equal to the sum of the two interior opposite angles.") $\triangle$

Another important consequence is the result which underpins the sequence of "circle theorems".

Problem 151 Let O be the circumcentre of $\triangle ABC$. Prove that

$$\angle AOB = 2 \cdot \angle ACB. \qquad \triangle$$

Problem **151** implies that

> "the angles subtended by any chord $\underline{AB}$ on a given arc of the circle are all equal",

and are equal to exactly one half of the angle subtended by $\underline{AB}$ at the centre O of the circle. This leads naturally to the familiar property of *cyclic quadrilaterals*.

Problem 152 Let $ABCD$ be a quadrilateral inscribed in a circle (such a quadrilateral is said to be *cyclic*, and the four vertices are said to be *concyclic* – that is, they lie together on the same circle). Prove that opposite angles (e.g. $\angle B$ and $\angle D$) must add to a straight angle. (Two angles which add to a straight angle are said to be *supplementary*.) $\triangle$

These results have lots of lovely consequences: we shall see one especially striking example in Problem **164**. Meantime we round up our summary of the "circle theorems".

Problem 153 Suppose that the line XAY is tangent to the circumcircle of $\triangle ABC$ at the point A, and that X and C lie on opposite sides of the line AB. Prove that $\angle XAB = \angle ACB$. $\triangle$

Problem 154

(a) Suppose C, D lie on the same side of the line AB.

(i) If D lies inside the circumcircle of $\triangle ABC$, then $\angle ADB > \angle ACB$.

(ii) If D lies outside the circumcircle of $\triangle ABC$, then $\angle ADB < \angle ACB$.

(b) Suppose C, D lie on the same side of the line AB, and that $\angle ACB = \angle ADB$. Then D lies on the circumcircle of $\triangle ABC$.

(c) Suppose that $ABCD$ is a quadrilateral, in which angles $\angle B$ and $\angle D$ are supplementary. Then $ABCD$ is a cyclic quadrilateral. $\triangle$

Another result which follows now that we know that the angles of a triangle add to a straight angle is a useful additional congruence criterion – namely the **RHS-congruence criterion**. This is a 'limiting case' of the failed ASS-congruence criterion (see the example in Problem **140**). In the failed ASS criterion the given data correspond to two *different* triangles – one in which the angle opposite the first specified side (the first "S" in "ASS") is acute, and one in which the angle opposite the first specified side is obtuse. In the RHS-congruence criterion, the angle opposite the first specified side is a *right angle*, and the two possible triangles are in fact congruent.

RHS-congruence criterion: If $\angle ABC$ and $\angle A'B'C'$ are both right angles, and $\underline{BC} = \underline{B'C'}$, $\underline{CA} = \underline{C'A'}$, then

$$\triangle ABC \equiv \triangle A'B'C'.$$

Proof Suppose that $\underline{AB} = \underline{A'B'}$. Then

$\underline{AB} = \underline{A'B'}$,
$\angle ABC = \angle A'B'C'$,
$\underline{BC} = \underline{B'C'}$.

Hence we may apply the SAS-congruence criterion to conclude that $\triangle ABC \equiv \triangle A'B'C'$.

If on the other hand $\underline{AB} \neq \underline{A'B'}$, we may suppose that $\underline{BA} > \underline{B'A'}$. Now construct A'' on BA such that $\underline{BA''} = \underline{B'A'}$. Then

$\underline{A''B} = \underline{A'B'}$,
$\angle A''BC = \angle A'B'C'$,
$\underline{BC} = \underline{B'C'}$,
$\therefore \triangle A''BC \equiv \triangle A'B'C'$ (by SAS-congruence).
Hence $\underline{A''C} = \underline{A'C'} = \underline{AC}$, so $\triangle CAA''$ is isosceles.
$\therefore \angle CA''A = \angle CAA''$.

However, $\angle CA''A > \angle CBA$ (since the exterior angle $\angle CA''A$ in $\triangle CBA''$ must be greater than the interior opposite angle $\angle CBA$, by Problem **145**).

But then the two base angles in the isosceles triangle $\triangle CAA''$ are each greater than a right angle – so the angle sum of $\triangle CAA''$ is greater than a straight angle, which is impossible. Hence this case cannot occur. QED

RHS-congruence seems to be needed to prove the basic result (Problem **161** below) about the area of parallelograms, and this is then needed in the proof of Pythagoras' Theorem (Problem **18**). In one sense RHS-congruence looks like a special case of SSS-congruence (as soon as two pairs of sides in two right angled triangles are equal, Pythagoras' Theorem guarantees that the third pair of sides are also equal). However this observation cannot be used to justify RHS-congruence if RHS-congruence is needed to justify Pythagoras' Theorem.

Problem 155 Given a circle with centre O, let Q be a point outside the circle, and let QP, QP' be the two tangents from Q, touching the circle at P and at P'. Prove that $\underline{QP} = \underline{QP'}$, and that the line OQ bisects the angle $\angle PQP'$. △

Problem 156 You are given two lines m and n crossing at the point B.

(a) If A lies on m and C lies on n, prove that each point X on the bisector of angle $\angle ABC$ is equidistant from m and from n.

(b) If X is equidistant from m and from n, prove that X must lie on one of the bisectors of the two angles at B. △

Problem **156** shows that, given two lines m and n that cross at B, the bisectors of the two pairs of vertically opposite angles formed at B form the *locus* of **all** points X which are equidistant from the two lines m and n. This allows us to mimic the comments following Problem **143** and so to construct the *incentre* of a triangle.

> Given any $\triangle ABC$, let I be the point where the angle bisectors of $\angle ABC$ and $\angle BCA$ meet.
>
> Let the perpendiculars from I to the three sides AB, BC, CA meet the sides at P, Q, R respectively. Then
>
> $\underline{IP} = \underline{IQ}$ (since I lies on the bisector of $\angle ABC$) and
> $\underline{IQ} = \underline{IR}$ (since I lies on the bisector or $\angle BCA$).
>
> Hence the circle which has centre I and which passes through P also passes through Q and R.
> Moreover, I also lies on the bisector of $\angle CAB$; and since the radii $\underline{IP}, \underline{IQ}, \underline{IR}$ are perpendicular to the sides of the triangle, the circle is tangent to the three sides of the triangle (by the comments following Problem **148**).

This circle is called the *incircle* of $\triangle ABC$, and I is called the *incentre* of $\triangle ABC$. The radius of the incircle of $\triangle ABC$ is called the *inradius*, and is generally denoted by r.

A quadrilateral $ABCD$ in which $AB \parallel DC$ and $BC \parallel AD$ is called a *parallelogram*. A parallelogram $ABCD$ with a right angle is a *rectangle*. A parallelogram $ABCD$ with $\underline{AB} = \underline{AD}$ is called a *rhombus*. A rectangle which is also a rhombus is called a *square*.

Problem 157 Let $ABCD$ be a parallelogram.

(i) Prove that $\triangle ABC \equiv \triangle CDA$, so that each triangle has area exactly half of area$(ABCD)$.

(ii) Conclude that opposite sides of $ABCD$ are equal in pairs and that opposite angles are equal in pairs.

(iii) Let AC and BD meet at X. Prove that X is the midpoint of both $\underline{AC}$ and $\underline{BD}$. $\triangle$

Problem 158 Let $ABCD$ be a parallelogram with centre X (where the two main diagonals $\underline{AC}$ and $\underline{BD}$ meet), and let m be any straight line passing through the centre. Prove that m divides the parallelogram into two parts of equal area. $\triangle$

We defined a parallelogram to be "a quadrilateral $ABCD$ in which $AB \parallel DC$ and $BC \parallel AD$"; however, in practice, we need to be able to recognise a parallelogram even if it is not presented in this form. The next result hints at the variety of other conditions which allow us to recognise a given quadrilateral as being a parallelogram "in mild disguise".

Problem 159

(a) Let $ABCD$ be a quadrilateral in which $AB \parallel DC$, and $\underline{AB} = \underline{DC}$. Prove that $BC \parallel AD$, and hence that $ABCD$ is a parallelogram.

(b) Let $ABCD$ be a quadrilateral in which $\underline{AB} = \underline{DC}$ and $\underline{BC} = \underline{AD}$. Prove that $AB \parallel DC$, and hence that $ABCD$ is a parallelogram.

(c) Let $ABCD$ be a quadrilateral in which $\angle A = \angle C$ and $\angle B = \angle D$. Prove that $AB \parallel DC$ and that $BC \parallel AD$, and hence that $ABCD$ is a parallelogram. $\triangle$

The next problem presents a single illustrative example of the kinds of things which we know in our bones must be true, but where the reason, or proof, may need a little thought.

Problem 160 Let $ABCD$ be a parallelogram. Let M be the midpoint of $\underline{AD}$ and N be the midpoint of $\underline{BC}$. Prove that $MN \parallel AB$, and that MN passes through the centre of the parallelogram (where the two diagonals meet). $\triangle$

Problem 161 Prove that any parallelogram $ABCD$ has the same area as the rectangle on the same base $\underline{DC}$ and "with the same height" (i.e. lying between the same two parallel lines AB and DC). $\triangle$

The ideas and results we have summarised up to this point provide exactly what is needed in the proof of Pythagoras' Theorem outlined back in Chapter 1, Problem **18**. They also allow us to identify two more "centres" of a given triangle $\triangle ABC$.

Problem 162 Given any triangle $\triangle ABC$, draw the line through A which is parallel to BC, the line through B which is parallel to AC, and the line through C which is parallel to AB. Let the first two constructed lines meet at C', the second and third lines meet at A', and the first and third lines meet at B'.

(a) Prove that A is the midpoint of $\underline{B'C'}$, that B is the midpoint of $\underline{C'A'}$, and that C is the midpoint of $\underline{A'B'}$.

(b) Conclude that the perpendicular from A to BC, the perpendicular from B to CA, and the perpendicular from C to AB all meet in a single point H. (H is called the *orthocentre* of $\triangle ABC$.) $\triangle$

Let the foot of the perpendicular from A to BC be P, the foot of the perpendicular from B to CA be Q, and the foot of the perpendicular from C to AB be R. Then $\triangle PQR$ is called the *orthic triangle* of $\triangle ABC$. The "circle theorems" (especially Problems **151** and **154**(c)) lead us to discover that this triangle has two quite unexpected properties. As a partial preparation for one of the properties we digress slightly to introduce a classic problem.

Problem 163 My horse is tethered at H some distance away from my village V. Both H and V are on the same side of a straight river. How should I choose the shortest route to lead the horse from H to V, if I want to water the horse at the river en route? $\triangle$

Problem 164 Let $\triangle ABC$ be an acute angled triangle.

(a) Prove that, among all possible triangles $\triangle PQR$ inscribed in $\triangle ABC$, with P on BC, Q on CA, R on AB, the orthic triangle is the one with the shortest perimeter.

(b) Suppose that the sides of $\triangle ABC$ act like mirrors. A ray of light is shone along one side of the orthic triangle PQ, reflects off CA, and the reflected beam then reflects in turn off AB. Where does the ray of light next hit the side BC? (Alternatively, imagine the sides of the triangle as billiard table cushions, and explain the path followed by a ball which is projected, without spin, along PQ.) $\triangle$

We come next to the fourth among the standard "centres of a triangle".

Problem 165 Given $\triangle ABC$, let L be the midpoint of the side $\underline{BC}$. The line AL is called a *median* of $\triangle ABC$. (It is not at all obvious, but if we imagine the triangle as a lamina, having a uniform thickness, then $\triangle ABC$ would exactly balance if placed on a knife-edge running along the line AL.) Let M be the midpoint of the side $\underline{CA}$, so that BM is another median of $\triangle ABC$. Let G be the point where AL and BM meet.

(a)(i) Prove that $\triangle ABL$ and $\triangle ACL$ have equal area. Conclude that $\triangle ABG$ and $\triangle ACG$ have equal area.

(ii) Prove that $\triangle BCM$ and $\triangle BAM$ have equal area. Conclude that $\triangle BCG$ and $\triangle BAG$ have equal area.

(b) Let N be the midpoint of $\underline{AB}$. Prove that CG and GN are the same straight line (i.e. that $\angle CGN$ is a straight angle). Hence conclude that the three medians of any triangle always meet in a point G. $\triangle$

The point where all three medians meet is called the *centroid* of the triangle. For the geometry of the triangle, this is all you need to know. However, it is worth noting that the centroid is the point that would be the 'centre of gravity' of the triangle if the triangle is thought of as a thin lamina with a uniform distribution of mass.

Next we revisit, and reprove in the Euclidean spirit, a result that you proved in Problem **95** using coordinates – namely the *Midpoint Theorem*.

Problem 166 (**The Midpoint Theorem**) Given any triangle $\triangle ABC$, let M be the midpoint of the side $\underline{AC}$, and let N be the midpoint of the side $\underline{AB}$. Draw in $\underline{MN}$ and extend it beyond N to a point M' such that $\underline{MN} = \underline{NM'}$.

(a) Prove that $\triangle ANM \equiv \triangle BNM'$.

(b) Conclude that $\underline{BM'} = \underline{CM}$ and that $BM' \parallel CM$.

(c) Conclude that $MM'BC$ is a parallelogram, so that $\underline{CB} = \underline{MM'}$. Hence $\underline{MN}$ is parallel to $\underline{CB}$ and half its length. △

The *Midpoint Theorem* can be reworded as follows:

> Given $\triangle AMN$.
> Extend $\underline{AM}$ to C such that $\underline{AM} = \underline{MC}$ and extend $\underline{AN}$ to B such that $\underline{AN} = \underline{NB}$.
> Then $CB \parallel MN$ and $\underline{CB} = 2 \cdot \underline{MN}$.

This rewording generalizes SAS-congruence in a highly suggestive way, and points us in the direction of "SAS-similarity".

> **SAS-similarity** ($\times 2$): if $\underline{A'B'} = 2 \cdot \underline{AB}$, $\angle BAC = \angle B'A'C'$, and $\underline{A'C'} = 2 \cdot \underline{AC}$, then
> $\underline{B'C'} = 2 \cdot \underline{BC}$, $\angle ABC = \angle A'B'C'$, and $\angle BCA = \angle B'C'A'$.
>
> **Proof** Extend AB to the point B'' such that $\underline{AB''} = \underline{A'B'}$, and extend AC to the point C'' such that $\underline{AC''} = \underline{A'C'}$.
> Then $\triangle B''AC'' \equiv \triangle B'A'C'$ (by SAS-congruence), so
> $\underline{B''C''} = \underline{B'C'}$, $\angle B''C''A = \angle B'C'A'$, $\angle C''B''A = \angle C'B'A'$.
> By construction we have $\underline{AB''} = 2 \cdot \underline{AB}$, and $\underline{AC''} = 2 \cdot \underline{AC}$.
> Hence (by the Midpoint Theorem): $\underline{B''C''} = 2 \cdot \underline{BC}$ (so $\underline{B'C'} = 2 \cdot \underline{BC}$), and $BC \parallel B''C''$ (so $\angle BCA = \angle B''C''A$ and $\angle CBA = \angle C''B''A$).
> $\therefore \angle B''C''A = \angle B'C'A' = \angle BCA$,
> and $\angle C''B''A = \angle C'B'A' = \angle CBA$. QED

The SAS-similarity ($\times 2$) interpretation of the Midpoint Theorem is like the SAS-congruence criterion in that one pair of corresponding angles in $\triangle BAC$ and $\triangle B'A'C'$ are equal, while the sides on either side of this angle in the two triangles are related; but instead of the two pairs of corresponding sides being equal, the sides of $\triangle B'A'C'$ are double the corresponding sides of $\triangle BAC$.

In general we say that

> $\triangle ABC$ is *similar* to $\triangle A'B'C'$ (written as $\triangle ABC \sim \triangle A'B'C'$) with scale-factor m if each angle of $\triangle ABC$ is equal to the corresponding angle of $\triangle A'B'C'$, and if corresponding sides are all in the same ratio:
>
> $$\underline{A'B'} : \underline{AB} = \underline{B'C'} : \underline{BC} = \underline{C'A'} : \underline{CA} = m : 1.$$

If two triangles $\triangle A'B'C'$ and $\triangle ABC$ are similar, with (linear) scale factor $\underline{A'B'} : \underline{AB} = m : 1$, then the ratio between their areas is

$$\text{area}(\triangle A'B'C') : \text{area}(\triangle ABC) = m^2 : 1.$$

Two similar triangles $\triangle ABC$ and $\triangle A'B'C'$ give rise to six matching pairs:

- the three pairs of corresponding angles (which are equal in pairs), and
- the three pairs of corresponding sides (which are in the same ratio).

In the case of congruence, the **congruence criteria** tell us that we do not need to check all six pairs to guarantee that two triangles are congruent: these criteria guarantee that certain *triples* suffice. The **similarity criteria** guarantee much the same for similarity.

Suppose we are given triangles $\triangle ABC$, $\triangle A'B'C'$.

AAA-similarity: If

$$\angle ABC = \angle A'B'C',\ \angle BCA = \angle B'C'A',\ \angle CAB = \angle C'A'B',$$

then

$$\underline{A'B'} : \underline{AB} = \underline{B'C'} : \underline{BC} = \underline{C'A'} : \underline{CA},$$

so the two triangles are similar.

SSS-similarity: If

$$\underline{A'B'} : \underline{AB} = \underline{B'C'} : \underline{BC} = \underline{C'A'} : \underline{CA},$$

then

$$\angle ABC = \angle A'B'C',\ \angle BCA = \angle B'C'A',\ \angle CAB = \angle C'A'B',$$

so the two triangles are similar.

SAS-similarity: If

$$\underline{A'B'} : \underline{AB} = \underline{A'C'} : \underline{AC} = m : 1$$

and

$$\angle B'A'C' = \angle BAC,$$

then

$$\underline{B'C'} : \underline{BC} = \underline{A'B'} : \underline{AB} = \underline{A'C'} : \underline{AC}$$

and

$$\angle A'B'C' = \angle ABC, \ \ \angle B'C'A' = \angle BCA,$$

so the two triangles are similar.

Our rewording of the Midpoint Theorem gave rise to a version of the third of these criteria, with $m = 2$.

AAA-similarity in right angled triangles is what makes trigonometry possible. Suppose that two triangles $\triangle ABC$, $\triangle A'B'C'$ have right angles at A and at A'. If $\angle ABC = \angle A'B'C'$, then (since the angles in each triangle add to two right angles) we also have $\angle BCA = \angle B'C'A'$. It then follows (from AAA-similarity) that

$$\underline{A'B'} : \underline{AB} = \underline{B'C'} : \underline{BC} = \underline{C'A'} : \underline{CA},$$

so the trig ratio in $\triangle ABC$

$$\sin B = \frac{\underline{AC}}{\underline{BC}}$$

has the same value as the corresponding ratio in $\triangle A'B'C'$

$$\sin B' = \frac{\underline{A'C'}}{\underline{B'C'}}.$$

Hence this ratio *depends only on the angle* B, and not on the triangle in which it occurs. The same holds for $\cos\angle B$ and for $\tan\angle B$.

The art of solving geometry problems often depends on looking for, and identifying, similar triangles hidden in a complicated configuration. As an introduction to this, we focus on three classic properties involving circles, where the figures are sufficiently simple that similar triangles should be fairly easy to find.

Problem 167 The point P lies outside a circle. The tangent from P touches the circle at T, and a secant from P cuts the circle at A and at B. Prove that $\underline{PA} \times \underline{PB} = \underline{PT}^2$. △

Problem 168 The point P lies outside a circle. Two secants from P meet the circle at A, B and at C, D respectively. Prove in two different ways that

$$\underline{PA} \times \underline{PB} = \underline{PC} \times \underline{PD}.$$

△

Problem 169 The point P lies inside a circle. Two secants from P meet the circle at A, B and at C, D respectively. Prove that

$$\underline{PA} \times \underline{PB} = \underline{PC} \times \underline{PD}. \qquad \triangle$$

We end our summary of the foundations of Euclidean geometry by deriving the familiar formula for the area of a trapezium and its 3-dimensional analogue, and a formulation of the similarity criteria which is often attributed to Thales (Greek 6^{th} century BC).

Problem 170 Let $ABCD$ be a trapezium with $AB \parallel DC$, in which AB has length a and DC has length b.

(a) Let M be the midpoint of $\underline{AD}$ and let N be the midpoint of $\underline{BC}$. Prove that $MN \parallel AB$ and find the length of $\underline{MN}$.

(b) If the perpendicular distance between AB and DC is d, find the area of the trapezium $ABCD$. $\triangle$

Problem 171 A pyramid $ABCDE$, with apex A and square base $BCDE$ of side length b, is cut parallel to the base at height d above the base, leaving a frustum of a pyramid, with square upper face of side length a. Find a formula for the volume of the resulting solid (in terms of a, b, and d). $\triangle$

The following general result allows us to use "equality of ratios of line segments" whenever we have three parallel lines (without first having to conjure up similar triangles).

Problem 172 (**Thales' Theorem**) The lines AA' and BB' are parallel. The point C lies on the line AB, and C' lies on the line $A'B'$ such that $CC' \parallel BB'$. Prove that $\underline{AB} : \underline{BC} = \underline{A'B'} : \underline{B'C'}$. $\triangle$

Under certain conditions, the similarity criteria guarantee the equality of ratios of sides of two triangles. Thales' Theorem extends this "equality of ratios" to line segments which arise whenever two lines cross three parallel lines. One of the simplest, but most far-reaching, applications of this result is the tie-up between geometry and algebra which lies behind ruler and compasses constructions, and which underpins Descartes' (1596–1650) re-formulation of geometry in terms of coordinates (see Problem **173**).

Thales (c. 620–c. 546 BC) was part of the flowering of Greek thought having its roots in Milesia (in the south west of Asia Minor, or modern Turkey). Thales seems to have been interested in almost everything – philosophy, astronomy, politics, and also geometry. In Britain his name is usually attached to the fact that the angle subtended by a diameter is a right angle. On the continent, his name is more strongly attached to the result in Problem **172**. His precise contribution to geometry is unclear – but he seems to have played a significant role in kick-starting what became (300 years later) the polished version of Greek mathematics that we know today.

Thales' contributions in other spheres were perhaps even more significant than in geometry. He seems to have been among the first to try to "explain" phenomena in reductionist terms – identifying "water" as the single "element", or first principle, from which all substances are derived. Anaximenes (c. 586–c. 526 BC) later argued in favour of "air" as the first principle. These two elements, together with "fire" and "earth", were generally accepted as the four Greek "elements" – each of which was supposed to contribute to the construction of observed matter and change in different ways.

Problem 173 To define "length", we must first decide which line segment is deemed to have unit length. So suppose we are given line segments $\underline{XY}$ of length 1, $\underline{AB}$ of length a, (i.e. $\underline{AB} : \underline{XY} = a : 1$), and $\underline{CD}$ of length b.

(a) Use Problem **137** to construct a segment of length $a + b$, and if $a \geqslant b$, a segment of length $a - b$.

(b) Show how to construct a line segment of length ab and a segment of length $\frac{a}{b}$.

(c) Show how to construct a line segment of length $\sqrt{a}$. △

5.3. Areas, lengths and angles

Problem 174 A rectangular piece of fruitcake has a layer of icing on top and down one side to form a larger rectangular slab of cake (as shown in Figure 3).

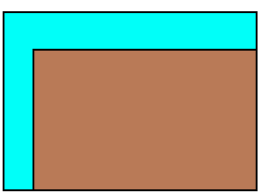


Figure 3: Icing on the cake

Describe how to make a single straight cut so as to divide both the fruitcake and the icing exactly in half. (The thickness of the icing on top is not necessarily the same as the thickness down the side.) △

Problem 175

(a) What is the angle between the two hands of a clock at 1:35? Can you find another time when the angle between the two hands is the same as this?

(b) How many times each day do the two hands of a clock 'coincide'? And at what times do they coincide?

(c) If we add a second hand, how many times each day do the three hands coincide? △

Problem 176 The twelve hour marks for a clock are marked on the circumference of a unit circle to form the vertices of a regular dodecagon $ABCDEFGHIJKL$. Calculate exactly (i.e. using Pythagoras' Theorem rather than trigonometry) the lengths of all the possible line segments joining two vertices of the dodecagon. △

Problem 177 Consider the lattice of all points (k, m, n) in 3-dimensions with integer coordinates k, m, n. Which of the following distances can be realised between lattice points?

$$\sqrt{1}, \sqrt{2}, \sqrt{3}, \sqrt{4}, \sqrt{5}, \sqrt{6}, \sqrt{7}, \sqrt{8}, \sqrt{9}, \sqrt{10}, \sqrt{11}, \sqrt{12}, \sqrt{13}, \sqrt{14}, \sqrt{15}, \sqrt{16}, \sqrt{17}.$$

△

Problem 178

(a) Five vertices A, B, C, D, E are arranged in cyclic order. However instead of joining each vertex to its two immediate neighbours to form a convex pentagon, we join each vertex to the *next but one* vertex to form a pentagonal star, or pentagram $ACEBD$. Calculate the sum of the five "angles" in any such pentagonal star.

(b) There are two types of 7-gonal stars. Calculate the sum of the angles at the seven vertices for each type.

(c) Try to extend the previous two results (and the proofs) to arbitrary n-gonal stars. △

Problem 179

(a) A regular pentagon $ABCDE$ with edges of length 1 is surrounded in the plane by five new regular pentagons – $ABLMN$ joined to $\underline{AB}$, $BCOPQ$ joined to $\underline{BC}$, and so on.

(i) Prove that M, N, X, Y lie on a line.

(ii) Prove that $MPSVY$ is a regular pentagon.

(iii) Find the edge length of this larger surrounding regular pentagon.

(b) Given a regular pentagon $MPSVY$, with edge length 1, draw the five diagonals to form the pentagram $MSYPV$. Let PY meet MV at A, and MS at B; let PV meet MS at C and SY at D; and let SY meet VM at E.

(i) Prove that $ABCDE$ is a regular pentagon.

(ii) Prove that A, B, and M are three vertices of a regular pentagon $ABLMN$, where L lies on $\underline{MP}$ and N lies on $\underline{MY}$.

(iii) Find the edge length of the regular pentagon $ABCDE$. △

5.4. Regular and semi-regular tilings in the plane

In Problem **36** we saw that a regular n-gon has a *circumcentre* O. If we join each vertex to the point O, we get n triangles, each with angle sum π. Hence the total angle sum in all n triangles is πn. Since the n angles around the point O add to 2π, the angles of the regular n-gon itself have sum $(n-2)\pi$. Hence each angle of the regular n-gon has size $\left(1-\frac{2}{n}\right)\pi$. (In the next chapter you will prove the general result that the sum of the angles in **any** n-gon is equal to $(n-2)\pi$ radians.)

Problem 180 A *regular tiling* of the plane is an arrangement of identical regular polygons, which fit together edge-to-edge so as to cover the plane with no overlaps.

(a) Prove that if a regular tiling of the plane with p-gons is possible, then $p = 3, 4$, or 6.

(b) Prove that a regular tiling of the plane exists for each of the values in (a). △

We refer to the arrangement of tiles around a vertex as the *vertex figure*. In a regular tiling all vertex figures are automatically identical, so it is natural to refer to the tiling in terms of its vertex figure. When $p = 3$, exactly $q = 6$ tiles fit together at each vertex, and we abbreviate "six equilateral triangles" as 3^6. In the same way we denote the tiling whose vertex figure consists of "four squares" as 4^4, and the tiling whose vertex figure consists of "three regular hexagons" as 6^3.

The natural approach in part (a) of Problem **180** is first to identify which *vertex figures* have no gaps or overlaps – giving a *necessary* condition for a regular tiling to exist. It is tempting to stop there, and to **assume** that this obvious *necessary* condition is also *sufficient*. The temptation arises in part because 2-dimensional regular tilings are so familiar. But it is important to recognize the distinction between a necessary and a sufficient condition; so the temptation should be resisted, and a construction given.

The procedure hidden in the solution to Problem **180** illustrates a key strategy, which dates back to the ancient Greeks, and which is called the *method of analysis*.

- First, we *imagine* that we have a typical solution to the problem – perhaps by giving it a name (even though we do not yet know anything about such a solution).
- We then use the given conditions to deduce features which any such solution must *necessarily* have.
- And we continue deriving more and more necessary conditions until we believe our list of derived conditions may also be *sufficient*.
- Finally we show that any configuration which satisfies our final derived list of necessary conditions is in fact a solution to the original problem, so that the list of necessary conditions is in fact *sufficient*, and we have effectively pinned down all possible solutions.

This is what we did in a very simple way in the solution to Problem **180**: the condition on vertex figures gave an evident necessary condition, which turned out to be sufficient to guarantee that such a tiling exists. The same general strategy guided our classification of *primitive Pythagorean triples* back in Problem **23**.

In the seventeenth century, this ancient Greek strategy was further developed by Fermat (1601–1665), and by Descartes (1596–1650). For example, Fermat left very few proofs; but his proof that the equation

$$x^4 + y^4 = z^4$$

has no solutions in positive integers x, y, z illustrated the method:

- Fermat started by supposing that a solution exists, and concluded that (x^2, y^2, z^2) would then be a Pythagorean triple.
- The known formula for such Pythagorean triples then allowed him to derive even stronger necessary conditions on x, y, z.
- These conditions were so strong they could never be satisfied!

Descartes developed a "method", whereby hard geometry problems could be solved by translating them into algebra – essentially using the *method of analysis*.

- Faced with a hard problem, Descartes first imagined that he had a point, or a locus, or a curve of the kind required for a solution.
- Then he introduced coordinates "x" and "y" to denote unknowns that were linked in the problem to be solved, and interpreted the given conditions as *equations* which the unknowns x and y would have to satisfy (i.e. as *necessary* constraints).
- The solutions to these equations then corresponded to possible solutions of the original problem.
- Sometimes the algebra did not quite generate a *sufficient* condition, giving rise to "pseudo-solutions" (values of x that satisfy the necessary conditions, but which did not correspond to actual solutions). So it was important to check each apparent solution – exactly as we did in Problem **180**(b), where we checked that we can construct tilings for each of the vertex figures that arise in part (a).

The importance of the final step in this process (checking that the list of necessary constraints is also sufficient) is underlined in the next problem where we try to classify certain "almost regular" tilings.

Problem 181 A *semi-regular tiling* of the plane is an arrangement of regular polygons (not necessarily all identical), which fit together edge-to-edge so as to cover the plane without overlaps, and such that the arrangements of tiles around any two vertices are congruent.

(a)(i) Refine your argument in Problem **180**(a) to list all possible vertex figures in a semi-regular tiling.

(ii) Try to find additional necessary conditions to eliminate vertex figures which cannot be realized, until your list of necessary conditions seems likely to be sufficient.

(b) The necessary conditions in part (a) give rise to a finite list of possible vertex figures. Construct all possible tilings corresponding to this list of possible vertex figures. △

Semi-regular tilings are often called *Archimedean* tilings. The reason for this name remains unclear. Pappus (c. 290–c. 350 AD), writing more than 500 years after the death of Archimedes (d. 212 BC), stated that Archimedes classified the semi-regular *polyhedra*. Now the classification of semi-regular polyhedra (Problem **190**) uses a similar approach to the classification of planar tilings, except that the sum of the angles at each vertex has sum *less than* (rather than exactly equal to) 360°. So it may be that the semi-regular tilings are named after Archimedes simply because he did something similar for polyhedra; or it may be that, since inequalities are harder to control than equalities, someone inferred (perhaps dodgily) that Archimedes must have known about semi-regular tilings as well as about semi-regular polyhedra.

Whatever the reason, tilings and polyhedra have fascinated mathematicians, artists and craftsmen for all sorts of unexpected reasons – as indicated by:

- the fact that the classification and construction of the five regular polyhedra appear as the culmination of the thirteen books of *Elements* by Euclid (flourished c. 300 BC);
- the ancient Greek attempt to link the five regular polyhedra with the four elements (earth, air, fire, and water) and the cosmos;
- the ceramic tilings to be found in Islamic art - for example, on the walls of the *Alhambra* in Grenada;
- the book *De Divina Proportione* by Luca Pacioli (c. 1445–1509), and the continuing fascination with the Golden Ratio;
- the geometric sketches of Leonardo da Vinci (1452–1519);
- the work of Kepler (1571–1630) who used the regular polyhedra to explain his bold theoretical cosmology in the *Astronomia Nova* (1609).

5.5. Ruler and compasses constructions for regular polygons

Euclid's *Elements* include methods for constructing the regular polygons that are required for the construction of the regular polyhedra (see Section 5.6). In one sense, Euclid is thoroughly modern: he is reluctant to work with entities that cannot be *constructed*. And for him, geometrical construction means construction "using ruler and compasses" only.

For each regular polygon, there are two related (and sometimes very different) construction problems:

- given two points A and B, construct the regular n-gon with $\underline{AB}$ as an *edge* of the regular polygon;
- given two points O and A, construct the regular n-gon $ABCD\cdots$ inscribed in the circle with centre O and passing through A, that is with *circumradius* $\underline{OA}$.

Before Problem **137** we saw how to construct an equilateral triangle ABC given the points A, B. And in Problem **36** we saw that every regular polygon has a circumcentre O.

Problem 182 Given points O, A, show how to construct the regular 3-gon ABC with circumcentre O. △

Problem 183

(a) Given two points O, A, show how to construct a regular 4-gon $ABCD$ with circumcentre O.

(b) Given points A, B, show how to construct a regular 4-gon $ABCD$. △

Problem 184

(a)(i) Given two points O, A, show how to construct a regular 6-gon $ABCDEF$ with circumcentre O.

(ii) Given two points O, A, show how to construct a regular 8-gon $ABCDEFGH$ with circumcentre O.

(b)(i) Given points A, B, show how to construct a regular 6-gon $ABCDEF$.

(ii) Given points A, B, show how to construct a regular 8-gon $ABCDEFGH$. △

Problem 185

(a)(i) Given two points O, A, show how to construct a regular 5-gon $ABCDE$ with circumcentre O.

(ii) Given points O, A, show how to construct a regular 10-gon $ABCDEFGHIJ$ with circumcentre O.

(b)(i) Given points A, B, show how to construct a regular 5-gon $ABCDE$.

(ii) Given points A, B, show how to construct a regular 10-gon $ABCDEFGHIJ$. △

We shall not prove it here, but it is impossible to construct a regular 7-gon, or a regular 9-gon, or a regular 11-gon using ruler and compasses. All constructions with ruler and compasses come down to two moves:

- if a is a known length, then $\sqrt{a}$ can be constructed (see Problem **173**(c));
- if an n-gon can be constructed, then the sides can be bisected to produce a $2n$-gon.

Put slightly differently, all ruler and compasses constructions involve solving linear or quadratic equations, so the only new points, or lengths we can construct are those which involve iterated square roots of expressions or lengths which were previously known.

This iterated extraction of square roots is linked to a fact first proved by Gauss (1777–1855), namely that the only regular p-gons (where p is a prime) that can be constructed are those where p is a Fermat prime – that is, a prime of the form $p = 2^k + 1$ (in which case k has to be a power of 2: see Problem **118**). Gauss proved (as a teenager, though it was first published in his book *Disquisitiones arithmeticae* in 1801):

> a regular n-gon can be constructed with ruler and compasses if and only if n has the form
>
> $$2^m \cdot p_1 \cdot p_2 \cdot p_3 \cdots p_k,$$
>
> where $p_1, p_2, p_3, \ldots, p_k$ are distinct Fermat primes.

As we noted in Chapter 2, the only known Fermat primes are the five discovered by Fermat himself, namely 3, 5, 17, 257, and 65 537.

5.6. Regular and semi-regular polyhedra

We have seen how regular polygons sometimes fit together edge-to-edge in the plane to create tilings of the whole plane. When tiling the plane, the angles of polygons meeting edge-to-edge around each vertex must add to 360°, or two straight angles. If the angles at a vertex add to *less than* 360°, then we are left with an empty gap and two free edges; and when these two free edges are joined, or glued together, the vertex figure rises out of the plane and becomes a 3-dimensional corner, or *solid angle.*

To form such a corner we need at least three polygons, or faces – and hence at least three edges and three faces meet around each vertex. For example, three squares fit nicely together in the plane, but leave a 90° gap. When the two spare edges are glued together, the result is to form a corner of a cube, where we have a vertex figure consisting of three regular 4-gons: so we refer to this vertex figure as 4^3.

Given a 3-dimensional corner, it may be possible to extend the construction, repeating the same vertex figure at every vertex. The resulting shape may then 'close up' to form a *convex* polyhedron. The assumption that in each vertex figure, the angles meeting at that vertex add to less than 360°, means that all the corners then project outwards – which is roughly what we mean when we say that the polyhedron is "convex".

A *regular polygon* is an arrangement of finitely many congruent line segments, with two line segments meeting at each vertex (and never crossing, or meeting internally), and with all vertices alike; a regular polygon can be inscribed in a circle (Problem **36**), and so encloses a convex subset of the plane. In the same spirit, a *regular polyhedron* is an arrangement of finitely many congruent regular polygons, with two polygons meeting at each edge, and with the same number of polygons in a single cycle around every vertex, enclosing a convex subset of 3-dimensional space (i.e. the polyhedron separates the remaining points of 3D into those that lie 'inside' and those that lie 'outside', and the line segment joining any two points of the polyhedral surface contains no points lying outside the polyhedron).

The important constraints here are the assumptions: that the polygons meet edge-to-edge with exactly two polygons meeting at each edge; that the same number of polygons meet around every vertex; and that the overall number of polygons, or faces, is *finite*. The assumption that the figure is convex should be seen as a temporary additional constraint, which means that the angles in polygons meeting at each vertex have sum less than 360°.

Problem 186 A vertex figure is to be formed by fitting regular p-gons together, edge-to-edge, for a fixed p. If there are q of these p-gons at a vertex, we denote the vertex figure by p^q. If the angles at each vertex add to less than 360°, prove that the only possible vertex figures are 3^3, 3^4, 3^5, 4^3, 5^3. △

The vertex figure 4^3 is realized by the way the positive axes meet at the vertex $(0, 0, 0)$, where

- the unit square $(0, 0, 0)$, $(1, 0, 0)$, $(1, 1, 0)$, $(0, 1, 0)$ in the xy-plane (with equation $z = 0$) meets
- the unit square $(0, 0, 0)$, $(1, 0, 0)$, $(1, 0, 1)$, $(0, 0, 1)$ in the xz-plane (with equation $y = 0$), and
- the unit square $(0, 0, 0)$, $(0, 1, 0)$, $(0, 1, 1)$, $(0, 0, 1)$ in the yz-plane (with equation $x = 0$).

If we include an eighth vertex $(1, 1, 1)$, and

- the unit square $(0, 0, 1)$, $(1, 0, 1)$, $(1, 1, 1)$, $(0, 1, 1)$ in the plane with equation $z = 1$,
- the unit square $(0, 1, 0)$, $(1, 1, 0)$, $(1, 1, 1)$, $(0, 1, 1)$ in the plane with equation $y = 1$,
- the unit square $(1, 0, 0)$, $(1, 1, 0)$, $(1, 1, 1)$, $(1, 0, 1)$ in the plane with equation $x = 1$,

we see that all eight vertices have the same vertex figure 4^3. Hence the possible vertex figure 4^3 in Problem **186** arises as the vertex figure of a regular polyhedron – namely the *cube.*

If we select the four vertices whose coordinates have odd sum $A = (1, 0, 0)$, $B = (0, 1, 0)$, $C = (0, 0, 1)$, $D = (1, 1, 1)$, then the distance between any two of these vertices is equal to $\sqrt{2}$, so each triple of vertices (such as $(1, 0, 0)$, $(0, 1, 0)$, $(0, 0, 1)$) defines a regular 3-gon ABC, with three such 3-gons meeting at each vertex of $ABCD$. Hence the possible vertex figure 3^3 in Problem **186** arises as the vertex figure of a regular polyhedron – namely the regular *tetrahedron* (*tetra* = four; *hedra* = faces).

Problem 187 With $A = (1, 0, 0)$ etc. as above, write down the coordinates of the six midpoints of the edges of the regular tetrahedron $ABCD$ (or equivalently, the six centres of the faces of the original cube). Each edge of the regular tetrahedron meets four other edges of the regular tetrahedron (e.g. $\underline{AB}$ meets $\underline{AC}$ and $\underline{AD}$ at one end, and $\underline{BC}$ and $\underline{BD}$ at the other end). Choose an edge $\underline{AB}$ and its midpoint P. Calculate the distance from P to the midpoints Q, R, S, T of the four edges which $\underline{AB}$ meets (namely the midpoints of $\underline{AC}$, $\underline{AD}$, $\underline{BD}$, $\underline{BC}$ respectively). Confirm that the triangles $\triangle PQR$, $\triangle PRS$, $\triangle PST$, $\triangle PTQ$ are all regular 3-gons, and that the vertex figure at P is of type 3^4 . Conclude that the possible vertex figure 3^4 in Problem **186** arises as the vertex figure of a regular polyhedron $PQRSTU$ – namely the *regular octahedron* (*octa* = eight; *hedra* = faces). △

Problem 188

(a) A *regular tetrahedron* $ABCD$ has edges of length 2, and sits with its base BCD on the table. Find the height of A above the base.

(b) A *regular octahedron* $ABCDEF$ has four triangles meeting at each vertex.

  (i) Let the four triangles which meet at A be ABC, ACD, ADE, AEB. Prove that $BCDE$ must be a square.

  (ii) Suppose that all the triangles have edges of length 2, and that the octahedron sits with one face BCF on the table – next to the regular tetrahedron from part (a). Which of these two solids is the taller? △

Problem 189 Let $O = (0, 0, 0)$, $A = (1, 0, 0)$, $B = (0, 1, 0)$, $C = (0, 0, 1)$ be four vertices of the cube as described after Problem **186** above. Draw equal and parallel line segments (initially of unknown length $1-2a$) through the centres of each pair of opposite faces – running in the three directions parallel to OA, or to OB, or to OC,

- from $N = (a, \frac{1}{2}, 0)$ to $P = (1-a, \frac{1}{2}, 0)$ and from $Q = (a, \frac{1}{2}, 1)$ to $R = (1-a, \frac{1}{2}, 1)$
- from $S = (\frac{1}{2}, 0, a)$ to $T = (\frac{1}{2}, 0, 1-a)$ and from $U = (\frac{1}{2}, 1, a)$ to $V = (\frac{1}{2}, 1, 1-a)$
- from $W = (0, a, \frac{1}{2})$ to $X = (0, 1-a, \frac{1}{2})$ and from $Y = (1, a, \frac{1}{2})$ to $Z = (1, 1-a, \frac{1}{2})$.

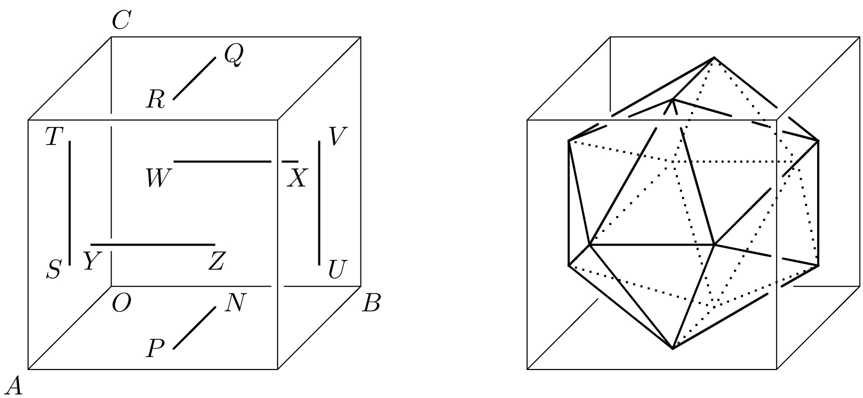


Figure 4: Construction of the regular icosahedron.

These are to form all 12 vertices and six of the 30 edges (of length $1-2a$) of a polyhedron, see Figure 4. The other 24 edges join each of these 12 vertices to its four natural neighbours on adjacent faces of the cube – to form the 20 triangular faces of the polyhedron: for example,

N joins: to S; to W; to X; and to U.

(i) Prove that $\underline{NS} = \underline{NW} = \underline{NX} = \underline{NU}$ and calculate the length of $\underline{NS}$.

(ii) Choose the value of the parameter a to guarantee that $\underline{NP} = \underline{NS}$, so that the five triangular faces meeting at the vertex N are all equilateral triangles, and each vertex figure of the resulting polyhedron then has vertex figure 3^5. △

The polyhedron is called the *regular icosahedron* (*icosa* = twenty, *hedra* = faces).

In the paragraph before Problem **187** we constructed the *dual* of the cube by marking the circumcentre of each of the six square faces of the cube, and then joining each circumcentre to its four natural neighbours. We now construct the *dual* of the regular icosahedron in exactly the same way. Each of the 20 circumcentres of the 20 triangular faces of a regular icosahedron has three natural neighbours (namely the circumcentres of the three neighbouring triangular faces). If we construct the 30 edges joining these 20 circumcentres, the five circumcentres of the five triangles in each vertex figure of the regular icosahedron form a regular pentagon, which becomes a face of the dual polyhedron – so we get 12 regular pentagons (one for each vertex of the regular icosahedron), with three pentagons meeting at each vertex of the dual polyhedron to give a vertex figure 5^3 at each of the 20 vertices, which form a *regular dodecahedron*.

Hence each of the five possible vertex figures in Problem **186** can be realised by a regular polyhedron. These are sometimes called the *Platonic solids* because Plato (c. 428–347 BC) often used them as illustrative examples in his writings on philosophy.

Constructing the five regular polyhedra is part of the essence of mathematics for everyone. In contrast, what comes next (in Problem **190**) may be viewed as "optional" at this stage. The ideas are worth noting, but the details may be best postponed for a rainy day.

Just as you classified semi-regular *tilings* in Section 5.4, so one can look for semi-regular *polyhedra*. A polyhedron is semi-regular if all of its faces are regular polygons (possibly with differing numbers of edges), fitting together edge-to-edge, with exactly the same ring of polygons around each vertex – the *vertex figure* of the polyhedron. Problem **190** uses "the method of analysis" - combining simple arithmetic, inequalities, and a little geometric insight – to achieve a remarkable complete classification of semi-regular polyhedra. There are usually said to be thirteen individual semi-regular polyhedra (excluding the five regular polyhedra); but one of these has a vertex figure that extends to a polyhedron in two different ways – each being the reflection of the other. There are in addition two infinite families – namely

- the n-gonal *prisms*, which consist of two parallel regular n-gons, with the top one positioned exactly above the bottom one, the two being joined by a belt of n squares (so with vertex figure $n \cdot 4^2$); and
- the n-gonal *antiprisms*, which consist of two parallel regular n-gons, but with the top n-gon turned through an angle of $\frac{\pi}{n}$ radians relative to the bottom one, the two being joined by a belt of $2n$ equilateral triangles (so with a vertex figure $n \cdot 3^3$).

Notice that the cube can also be interpreted as being a "4-gonal prism", and the regular octahedron can be interpreted as being a "3-gonal antiprism". Those interested in regular and semi-regular polyhedra are referred to the classic book *Mathematical models* by H.M. Cundy and A.P. Rollett.

Problem 190 Find possible combinations of three or more regular polygons whose angles add to less than 360°, and hence derive a complete list of possible vertex figures for a (convex) semi-regular polyhedron. Try to eliminate those putative vertex figures that cannot be extended to a semi-regular polyhedron. △

5.7. The Sine Rule and the Cosine Rule

Where given information, or a specified geometrical construction, determines an angle or length uniquely, it is sometimes – but not always – possible to find this angle or length using simple-minded angle-chasing and congruence.

Problem 191

(a) In the quadrilateral $ABCD$ the two diagonals $\underline{AC}$ and $\underline{BD}$ cross at X. Suppose $\underline{AB} = \underline{BC}$, $\angle BAC = 60^\circ$, $\angle DAC = 40^\circ$, $\angle BXC = 100^\circ$.

  (i) Calculate (exactly) $\angle ADB$ and $\angle CBD$.

  (ii) Calculate $\angle BDC$ and $\angle ACD$.

(b) In the quadrilateral $ABCD$ the two diagonals $\underline{AC}$ and $\underline{BD}$ cross at X. Suppose $\underline{AB} = \underline{BC}$, $\angle BAC = 70^\circ$, $\angle DAC = 40^\circ$, $\angle BXC = 100^\circ$.

  (i) Calculate (exactly) the size of $\angle BDC + \angle ACD$.

  (ii) Explain how we can be sure that $\angle BDC$ and $\angle ACD$ are uniquely determined, even though we cannot calculate them immediately. △

If it turns out that the simplest tools do not allow us to determine angles and lengths, this is usually because we are only using the most basic properties: the congruence criteria, and the parallel criterion. The general art of 'solving triangles' depends on the similarity criterion (usually via trigonometry). And the two standard techniques for 'solving triangles' that go beyond "angle-chasing" and congruence are the *Sine Rule*, which was established back in Problem **32** (and its consequences, such as the area formula $\frac{1}{2}ab\sin C$ – see Problem **33**), and the *Cosine Rule*.

The next problem invites you to use Pythagoras' Theorem to prove the Cosine Rule – an extension of Pythagoras' Theorem which applies to *all* triangles ABC (including those where the angle at C may not be a right angle).

Problem 192 (**The Cosine Rule**) Given $\triangle ABC$, let the perpendicular from A to BC meet BC at P. If $P = C$, then we know (by Pythagoras' Theorem) that $c^2 = a^2 + b^2$. Suppose $P \neq C$.

(i) Suppose first that P lies on the line segment $\underline{CB}$, or on $\underline{CB}$ extended beyond B. Express the lengths of $\underline{PC}$ and $\underline{AP}$ in terms of b and $\angle C$. Then apply Pythagoras' Theorem to $\triangle APB$ to conclude that

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

(ii) Suppose next that P lies on the line segment $\underline{BC}$ extended beyond C. Prove once again that

$$c^2 = a^2 + b^2 - 2ab\cos C. \qquad \triangle$$

Problem 193 Go back to the configuration in Problem **191**(b). The required angles are unaffected by scaling, so we may choose $\underline{AB} = \underline{BC} = 1$. Devise a strategy using the Sine Rule and the Cosine Rule to calculate $\angle BDC$ and $\angle ACD$ exactly. $\triangle$

It is worth reflecting on what the Cosine Rule really tells us:

(i) if in a triangle, we know any two sides (a and b) and the included angle (C), then we can calculate the third side (c); and

(ii) if we know all three sides (a, b, c), then we can calculate any angle (say C).

Hence if we know three sides, or two sides and the angle between them, we can work out all of the angles. The Sine Rule then complements this by ensuring that:

(iii) if we know any side and two angles (in which case we also know the third angle), then we can calculate the other two sides; and

(iv) if we know any angle A, and two sides – one of which is the side a opposite A, then we can calculate (one and hence) both the other angles (and hence the third side).

The upshot is that once a triangle is uniquely determined by the given data, we can "solve" to find all three sides and all three angles.

Trigonometry has a long and very interesting history (which is not at all easy to unravel). Euclid (flourished c. 300 BC) understood that corresponding sides in similar figures were "proportional". And he stated and proved the generalization of Pythagoras' Theorem, which we now call the Cosine Rule; but he did this in a theoretical form, without introducing cosines. Euclid's versions for acute-angled and obtuse-angled triangles involved correction terms with opposite signs, so he proved them separately (*Elements*, Book II, Propositions 12 and 13).

However, the development of trigonometry as an effective theoretical and practical tool seems to have been due to Hipparchus (died c. 125 BC), to Menelaus (c. 70–130 AD), and to Ptolemy (died 168 AD). Once trigonometry moved beyond the purely theoretical, the combination of

- the (exact) language of trigonometry, together with the Sine Rule and the Cosine Rule, and
- (approximate) "tables of trigonometric ratios" (nowadays replaced by calculators)

liberated astronomers, and later engineers, to calculate lengths and angles efficiently, and as accurately as they required.

In mathematics we either work with *exact* values, or we have to control errors precisely. But trigonometry can still be a valuable *exact* tool, provided we remember the lessons of working with fractions such as $\frac{2}{3}$, or with surds such as $\sqrt{2}$, or with constants such as π, and resist the temptation to replace them by some unenlightening approximate decimal. We can replace $\cos^{-1}\left(\frac{1}{2}\right)$ $\left(= \frac{\pi}{3}\right)$ and $\cos^{-1}\left(-\frac{1}{2}\right)$ $\left(= \frac{2\pi}{3}\right)$ by their exact values; but in general we need to be willing to work with, and to think about, *exact forms* such as "$\cos^{-1}\left(\frac{1}{3}\right)$" and "$\cos^{-1}\left(-\frac{1}{3}\right)$", without switching to some approximate evaluation.

Problem 194

(a) Let $ABCD$ be a regular tetrahedron with edges of length 2. Calculate the (exact) angle between the two faces ABC and DBC.

(b) We know that in 2D five equilateral triangles fit together *at a point* leaving just enough of an angle to allow a sixth triangle to fit. How many identical regular tetrahedra can one fit together, without overlaps *around an edge*, so that they all share the edge $\underline{BC}$ (say)? △

Problem 195

(a) Let $ABCDEF$ be a regular octahedron with vertices B, C, D, E adjacent to A forming a square $BCDE$, and with edges of length 2. Calculate the (exact) angle between the two faces ABC and FBC.

(b) How many identical regular octahedra can one fit together *around an edge*, without overlaps, so that they all share the edge $\underline{BC}$ (say)? △

Problem 196 Go back to the scenario of Problem **188**, with a regular tetrahedron and a regular octahedron both having edges of length 2, and both having one face flat on the table. Suppose we slide the tetrahedron across the table towards the octahedron. What unexpected phenomenon is guaranteed by Problems **194**(a) and **195**(a)? △

Problem 197 Consider the cube with edges of length 2 running parallel to the coordinate axes, with its centre at the origin $(0, 0, 0)$, and with opposite corners at $(1, 1, 1)$ and $(-1, -1, -1)$. The x-, y-, and z-axes, and the xy-, yz-, and zx-planes cut this cube into eight unit cubes – one sitting in each octant.

(i) Let $A = (0, 0, 1)$, $B = (1, 0, 0)$, $C = (0, 1, 0)$, $W = (1, 1, 1)$. Describe the solid $ABCW$.

(ii) Let $D = (-1, 0, 0)$, $X = (-1, 1, 1)$. Describe the solid $ACDX$.

(iii) Let $E = (0, -1, 0)$, $Y = (-1, -1, 1)$. Describe the solid $ADEY$.

(iv) Let $Z = (1, -1, 1)$. Describe the solid $AEBZ$.

(v) Let $F = (0, 0, -1)$ and repeat steps (i)–(iv) to obtain the four mirror image solids which lie beneath the xy-plane.

(vi) Describe the solid $ABCDEF$ which is surrounded by the eight identical solids in (i)–(v). △

Problem 198 Consider a single face $ABCDE$ of the regular dodecahedron, with edges of length 1, together with the five pentagons adjacent to it – so that each of the vertices A, B, C, D, E has vertex figure 5^3. Each vertex figure is rigid, so the whole arrangement of six regular pentagons is also rigid. Let V, W, X, Y, Z be the five vertices adjacent to A, B, C, D, E respectively. Calculate the dihedral angle between the two pentagonal faces that meet at the edge $\underline{AB}$. △

Problem 199 Suppose a regular icosahedron (Problem **189**) has edges of length 2. Position vertex A at the 'North pole', and let $BCDEF$ be the regular pentagon formed by its five neighbours.

(a)(i) Calculate the exact angle between the two faces ABC and ACD.

(ii) How many identical regular icosahedra can one fit together, without overlaps, around a single edge?

(b) Let $\mathcal{C}$ be the circumcircle of $BCDEF$, and let O be the circumcentre of this regular pentagon.

  (i) Prove that the three edge lengths of the right-angled triangle $\triangle BOA$ are the edge lengths of the regular hexagon inscribed in the circle $\mathcal{C}$, of the regular 10-gon inscribed in the circle $\mathcal{C}$, and of the regular 5-gon inscribed in the circle $\mathcal{C}$.

  (ii) Calculate the distance separating the plane of the regular pentagon $BCDEF$, and the plane of the corresponding regular pentagon joined to the 'South pole'. △

Notice that Problem **199**(b) shows that the regular icosahedron can be 'constructed' in the Euclidean spirit: part (b)(i) is essentially Proposition 10 of Book XIII of Euclid's *Elements*, and part (b)(ii) is implicit in Proposition 16 of the Book XIII. Once we are given the radius $\underline{OB}$, we can:

- construct the regular pentagon $BCDEF$ in the circle $\mathcal{C}$;
- bisect the sides of the regular pentagon and hence construct the regular 10-gon $BVC\cdots$ in the same circle;
- construct the vertical perpendicular at O, and transfer the length $\underline{BV}$ to the point O to determine the vertex A directly above O;
- transfer the radius $\underline{OB}$ to the vertical perpendicular at O to determine the plane directly below O, and hence construct the lower regular pentagon; etc..

It may be worth commenting on a common confusion concerning the regular icosahedron. Each regular polyhedron has a circumcentre, with all vertices lying on a corresponding sphere. If we join any triangular face of the regular icosahedron to the circumcentre O, we get a tetrahedron. These 20 tetrahedra are all congruent and fit together exactly at the point O "without gaps or overlaps". But they are **not** *regular* tetrahedra: the circumradius is *less than* the edge length of the regular icosahedron.

Problem 200 Prove that the only regular polyhedron that tiles 3D (without gaps or overlaps) is the cube. △

In one sense the result in Problem **200** is disappointing. However, since we know that there are all sorts of interesting 3-dimensional arrangements related to crystals and the way atoms fit together, the message is really that we need to look *beyond* regular tilings. For example, the construction in Problem **197** shows how the familiar *regular* tiling of space with cubes incorporates a *semi-regular* tiling of space with eight regular tetrahedra and two regular octahedra at each vertex.

5.8. Circular arcs and circular sectors

Length is defined for *straight* line segments, and area is defined in terms of *rectangles*; neither measure is defined for shapes with curved boundaries – unless, that is, they can be cunningly dissected and the pieces rearranged to make a straight line, or a rectangle.

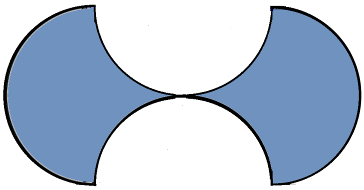


Figure 5: Dumbbell.

Problem 201 Four identical semicircles of radius 1 fit together to make the dumbbell shape shown in Figure 5. Find the exact area enclosed without using the formula for the area of a circle. △

In general, making sense of length and area for shapes with curved boundaries requires us to combine a little imagination with what we know about straight line segments and polygons. Our goal here is to lead up to the familiar results for the length of circular arcs and the area of circular sectors. But first we need to explore the perimeter and area of regular polygons, and the surface area of prisms and pyramids.

As so often in mathematics, to make sense of the perimeter and area of regular polygons we need to look beyond their actual values (which will vary according to the size of the polygon), and instead interpret these values as a function of some normalizing parameter – such as the *radius*. The calculations will be simpler if you first prove a general result.

Problem 202

(a) A regular n-gon and a regular $2n$-gon are inscribed in a circle of radius 1. The regular n-gon has edges of length $s_n = s$, while the regular $2n$-gon has edges of length $s_{2n} = t$. Prove that

$$t^2 = 2 - \sqrt{4 - s^2}.$$

(b) A regular "2-gon" inscribed in the unit circle is just a diameter (repeated twice), so has two identical edges of length $s_2 = 2$. Use the result in part (a) to calculate the edge length s_4 of a regular 4-gon, and the edge length s_8 of a regular 8-gon inscribed in the same circle.

(c) A regular 6-gon inscribed in the unit circle has edge length s_6 equal to the radius 1. Use the result in part (a) to calculate the edge length s_3 of a regular 3-gon inscribed in the unit circle, and the edge length s_{12} of a regular 12-gon inscribed in the unit circle.

(d) In Problem **185** we saw that a regular 5-gon inscribed in the unit circle has edge length

$$s_5 = \frac{\sqrt{10 - 2\sqrt{5}}}{2}.$$

Use the result in part (a) to calculate the edge length s_{10} of a regular 10-gon inscribed in the same circle. △

Problem 203

(a) A regular n-gon is inscribed in a circle of radius r.

(i) Find the exact perimeter p_n (in surd form): when $n = 3$; when $n = 4$; when $n = 5$; when $n = 6$; when $n = 8$; when $n = 10$; when $n = 12$.

(ii) Check that, for each n:

$$p_n = c_n \times r$$

for some constant c_n, where

$$c_3 < c_4 < c_5 < c_6 < c_8 < c_{10} < c_{12} \cdots$$

(b) A regular n-gon is circumscribed about a circle of radius r.

(i) Find the exact perimeter P_n (in surd form): when $n = 3$; when $n = 4$; when $n = 5$; when $n = 6$; when $n = 8$; when $n = 10$; when $n = 12$.

(ii) Check that, for each n:

$$P_n = C_n \times r$$

for some constant C_n, where

$$C_3 > C_4 > C_5 > C_6 > C_8 > C_{10} > C_{12} \cdots$$

(c) Explain why $c_{12} < C_{12}$. △

It follows from Problem **203** that

- the perimeters p_n and P_n of regular n-gons inscribed in, or circumscribed about, a circle of radius r all have the same form:

  $$\text{(inscribed) } p_n = c_n \times r; \text{ (circumscribed) } P_n = C_n \times r.$$

- The perimeters of inscribed regular n-gons all increase with n, but remain less than the perimeter of the circle, while
- the perimeters of the circumscribed regular n-gons all decrease with n, but remain greater than the perimeter of the circle.

Hence

- the perimeter P of the circle appears to have the form $P = K \times r$, where the ratio
  $$K = \frac{\text{perimeter}}{\text{radius}}$$
  satisfies
  $$c_3 < c_4 < c_5 < c_6 < c_8 < c_{10} < \cdots < K < \cdots < C_{10} < C_8 < C_6 < C_5 < C_4 < C_3.$$
  In particular, the value of the constant K lies somewhere between $c_{12} = 6.21\cdots$ and $C_{12} = 6.43\cdots$. If we now define the quotient K to be equal to "2π", we see that
  $$\text{(perimeter of circle of radius } r) = 2\pi r,$$
  where π denotes some constant lying between 3.1 and 3.22

  In this spirit one might reinterpret the first two bullet points as defining two sequences of constants "π_n" and "Π_n" for $n \geqslant 3$, such that

- (perimeter of a regular n-gon with circumradius r) $= 2\pi_n r$, where
  $$\pi_3 = \frac{3\sqrt{3}}{2} = 2.59\cdots, \pi_4 = 2\sqrt{2} = 2.82\cdots, \pi_5 = \frac{5\sqrt{10-2\sqrt{5}}}{4} = 2.93\cdots, \pi_6 = 3,$$
  etc.,

and

- (perimeter of a regular n-gon with inradius r)$= 2\Pi_n r$, where
  $$\Pi_3 = 3\sqrt{3} = 5.19\cdots, \Pi_4 = 4, \Pi_5 = 5\sqrt{5-2\sqrt{5}} = 3.63\cdots, \Pi_6 = 2\sqrt{3} = 3.46\cdots,$$
  etc..

Moreover

- $\pi_3 < \pi_4 < \pi_5 < \pi_6 < \pi_8 < \cdots < \pi < \cdots < \Pi_8 < \Pi_6 < \Pi_5 < \Pi_4 < \Pi_3.$

Problem 204 Find the exact length (in terms of π)

(i) of a semicircle of radius r;

(ii) of a quarter circle of radius r;

(iii) of the length of an arc of a circle of radius r that subtends an angle θ radians at the centre. △

In the next problem we follow a similar sequence of steps to conclude that the quotient

$$L = \frac{\text{area of circle of radius } r}{r^2}$$

is also constant. The surprise lies in the fact that this different constant is so closely related to the previous constant K.

Problem 205

(a) A regular n-gon is inscribed in a circle of radius r.

  (i) Find the exact area a_n (in surd form): when $n = 3$; when $n = 4$; when $n = 5$; when $n = 6$; when $n = 8$; when $n = 10$; when $n = 12$.

  (ii) Check that, for each n:

$$a_n = d_n \times r^2$$

  for some constant d_n, where

$$d_3 < d_4 < d_5 < d_6 < d_8 < d_{10} < d_{12} \cdots$$

(b) A regular n-gon is circumscribed about a circle of radius r.

  (i) Find the exact area A_n (in surd form): when $n = 3$; when $n = 4$; when $n = 5$; when $n = 6$; when $n = 8$; when $n = 10$; when $n = 12$.

  (ii) Check that, for each n:

$$A_n = D_n \times r^2$$

  for some constant D_n, where

$$D_3 > D_4 > D_5 > D_6 > D_8 > D_{10} > D_{12} \cdots$$

(c) Explain why $d_{12} < D_{12}$. △

It follows from Problem **205** that

- the areas a_n and A_n of regular n-gons inscribed in, or circumscribed about, a circle of radius r all have the same form:

$$\text{(inscribed) } a_n = d_n \times r^2; \text{ (circumscribed) } A_n = D_n \times r^2.$$

- The areas of inscribed regular n-gons all increase with n, but remain less than the area of the circle, while
- the areas of the circumscribed regular n-gons all decrease with n, but remain greater than the area of the circle, whence
- the area A of the circle appears to have the form $A = L \times r^2$, where the ratio
$$L = \frac{\text{area of circle of radius } r}{\text{radius squared}}$$
satisfies
$$d_3 < d_4 < d_5 < d_6 < d_8 < d_{10} \cdots < L < \cdots < D_{10} < D_8 < D_6 < D_5 < D_4 < D_3.$$
In particular, the value of L lies somewhere between $d_{12} = 3$ and $D_{12} = 12(2 - \sqrt{3}) = 3.21 \cdots$. The surprise lies in the fact that the constant L is exactly half of the constant K – that is, $L = \pi$, so
$$\text{(area of circle of radius } r) = \pi r^2.$$

The next problem offers a heuristic explanation for this surprise.

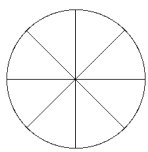


Figure 6: Circle cut into 8 slices.

Problem 206 A regular $2n$-gon $ABCDE \cdots$ is inscribed in a circle of radius r. The $2n$ radii $OA, OB, \ldots$ joining the centre O to the $2n$ vertices cut the circle into $2n$ sectors, each with angle $\frac{\pi}{n}$ (Figure 6).

These $2n$ sectors can be re-arranged to form an "almost rectangle", by orienting them alternately to point "up" and "down". In what sense does this "almost rectangle" have "height $= r$" and "width $= \pi r$"? △

Problem 207

(a) Find a formula for the surface area of a right cylinder with height h and with circular base of radius r.

(b) Find a similar formula for the surface area of a right prism with height h, whose base is a regular n-gon with inradius r. △

Problem 208

(a) Find the exact area (in terms of π)

(i) of a semicircle of radius r;

(ii) of a quarter circle of radius r;

(iii) of a sector of a circle of radius r that subtends an angle θ radians at the centre.

(b) Find the area of a sector of a circle of radius 1, whose total perimeter (including the two radii) is exactly half that of the circle itself. △

Problem 209

(a) Find a formula for the surface area of a right circular cone with base of radius r and slant height l.

(b) Find a similar formula for the surface area of a right pyramid with apex A whose base $BCDE \cdots$ is a regular n-gon with inradius r. △

Problem 210

(a) Find an expression involving "$\sin \frac{\pi}{n}$" for the ratio

$$\frac{\text{perimeter of inscribed regular } n\text{-gon}}{\text{perimeter of circumscribed circle}}.$$

(b) Find an expression involving "$\tan \frac{\pi}{n}$" for the ratio

$$\frac{\text{perimeter of circumscribed regular } n\text{-gon}}{\text{perimeter of inscribed circle}}.$$

△

Problem 211

(a) Find an expression involving "$\sin \frac{2\pi}{n}$" for the ratio

$$\frac{\text{area of inscribed regular } n\text{-gon}}{\text{area of circumscribed circle}}.$$

(b) Find an expression involving "$\tan \frac{\pi}{n}$" for the ratio

$$\frac{\text{area of circumscribed regular } n\text{-gon}}{\text{area of inscribed circle}}. \qquad \triangle$$

5.9. Convexity

This short section presents a simple result which to some extent justifies the assumptions made in the previous section – namely that the perimeter (or area) of a regular n-gon inscribed in a circle is less than the perimeter (or area) of the circle, and of the circumscribed regular n-gon.

Problem 212 A convex polygon P_1 is drawn in the interior of another convex polygon P_2.

(a) Explain why the area of P_1 must be less than the area of P_2.

(b) Prove that the perimeter of P_1 must be less than the perimeter of P_2. △

5.10. Pythagoras' Theorem in three dimensions

Pythagoras' Theorem belongs in 2-dimensions. But does it generalise to 3-dimensions? The usual answer is to interpret the result in terms of coordinates.

Problem 213

(a) Construct a right angled triangle that explains the standard formula for the distance from $P = (a, b)$ to $Q = (d, e)$.

(b) Use part (a) to derive the standard formula for the distance from $P = (a, b, c)$ to $Q = (d, e, f)$. △

This extension of Pythagoras' Theorem to 3-dimensions is extremely useful, but not very profound. In contrast, the next result is more intriguing, but seems to be a complete fluke of limited relevance. In 2D, a right angled triangle is obtained by

- taking one corner A of a rectangle $ABCD$, together with its two neighbours B and D;
- then "cutting off the corner" to get the triangle ABD.

This suggests that a corresponding figure in 3D might be obtained by

- taking one corner A of a cuboid, together with its three neighbours B, C, D;
- then cutting off the corner to get a pyramid $ABCD$, with the right angled triangle $\triangle ABC$ as base, and with apex D.

The obvious candidate for the "3D-hypotenuse" is then the sloping face BCD, and the three right angled triangles $\triangle ABC$, $\triangle ACD$, $\triangle ADB$ presumably correspond to the 'legs' (the shorter sides) of the right triangle in 2D.

Problem 214 You are given a pyramid $ABCD$ with all three faces meeting at A being right angled triangles with right angles at A. Suppose $\underline{AB} = b$, $\underline{AC} = c$, $\underline{AD} = d$.

(a) Calculate the areas of $\triangle ABC$, $\triangle ACD$, $\triangle ADB$ in terms of b, c, d.

(b) Calculate the area of $\triangle BCD$ in terms of b, c, d.

(c) Compare your answer in part (b) with the sum of the squares of the three areas you found in part (a). $\triangle$

More significant (e.g. for navigation on the surface of the Earth) and more interesting than Problem **214** is to ask what form Pythagoras' Theorem takes for "lines on a sphere".

For simplicity we work on a unit sphere. We discovered in the run-up to Problem **34** that lines, or shortest paths, on a sphere are arcs of great circles. So, if the triangle $\triangle ABC$ on the unit sphere is right angled at A, we may rotate the sphere so that the arc $\underline{AB}$ lies along the equator and the arc $\underline{AC}$ runs up a circle of longitude. It is then clear that, once the lengths c, b of $\underline{AB}$ and $\underline{AC}$ are known, the locations of B and C are essentially determined, and hence the length of the arc $\underline{BC}$ on the sphere is determined. So we would like to have a simple formula that would allow us to calculate the length of the arc $\underline{BC}$ directly in terms of c and b.

Problem 215 Given a spherical triangle $\triangle ABC$ on the unit sphere with centre O, such that $\angle BAC$ is a right angle, and such that $\underline{AB}$ has length c, and $\underline{AC}$ has length b.

(a) We have (rightly) referred to b and c as 'lengths'. But what are they really?

(b) We want to know how the inputs b and c determine the value of the length a of the arc $\underline{BC}$; that is, we are looking for a function with inputs b and c, which will allow us to determine the value of the "output" a. Think about the answer to part (a). What kind of standard functions do we already know that could have inputs b and c?

(c) Suppose $c = 0 \neq b$. What should the output a be equal to? (Similarly if $b = 0 \neq c$.) Which standard function of b and of c does this suggest is involved?

(d)(i) Suppose $\angle B = \angle C = \frac{\pi}{2}$, what should the output a be equal to?

(ii) Suppose $\angle B = \frac{\pi}{2}$, but $\angle C$ (and hence c) is unconstrained. The output a is then determined – but the formula must give this fixed output for different values of c. What does this suggest as the "simplest possible" formula for a? △

The answers to Problem **215** give a pretty good idea what form Pythagoras' Theorem must take on the unit sphere. The next problem proves this result as a simple application of the familiar 2D Cosine Rule.

Problem 216 Given any triangle $\triangle ABC$ on the unit sphere with a right angle at the point A, we may position the sphere so that A lies on the equator, with $\underline{AB}$ along the equator and $\underline{AC}$ up a circle of longitude. Let O be the centre of the sphere and let $\mathbf{T}$ be the tangent plane to the sphere at the point A. Extend the radii $\underline{OB}$ and $\underline{OC}$ to meet the plane $\mathbf{T}$ at B' and C' respectively.

(a) Calculate the lengths of the line segments $\underline{AB'}$ and $\underline{AC'}$, and hence of $\underline{B'C'}$.

(b) Calculate the lengths of $\underline{OB'}$ and $\underline{OC'}$, and then apply the Cosine Rule to $\triangle B'OC'$ to find an equation linking b and c with $\angle B'OC'$ $(= a)$. △

When "solving triangles" on the sphere the same principles apply as in the plane: right angled triangles hold the key – but Pythagoras' Theorem and trig in right angled triangles must be extended to obtain variations of the Sine Rule and the Cosine Rule for *spherical* triangles. The corresponding results on the sphere are both similar to, and intriguingly different from, those we are used to in the plane. For example, there are two forms of the Cosine Rule extending the result in Problem **216**.

Problem 217 Given a (not necessarily right angled) triangle $\triangle ABC$ on the unit sphere, apply the same proof as in Problem **216** to show (with the usual labelling) that:

$$\cos a = \cos b \cdot \cos c + \sin b \cdot \sin c \cdot \cos A$$ △

The other form of the Cosine Rule is "dual" to that in Problem **217** (with arcs and angles interchanged, and with an unexpected change of sign) – namely:

$$\cos A = -\cos B \cdot \cos C + \sin B \cdot \sin C \cdot \cos a.$$

The next two problems derive a version of the Sine Rule for spherical triangles.

Problem 218 Let $\triangle ABC$ be a triangle on the unit sphere with a right angle at A. Let A' lie on the arc BA produced, and C' lie on the arc BC produced so that $\triangle A'BC'$ is right angled at A'. With the usual labelling (so that x denotes the length of the side of a triangle opposite vertex X, with arc $AC = b$, arc $BC = a$, arc $BC' = a'$, and arc $A'C' = b'$, prove that:

$$\frac{\sin b}{\sin a} = \frac{\sin b'}{\sin a'}. \qquad \triangle$$

Problem 219 Let $\triangle ABC$ be a general triangle on the unit sphere with the usual labelling (so that x denotes the length of the side of a triangle opposite vertex X, and X is used both to label the vertex and to denote the size of the angle at X). Prove that:

$$\frac{\sin a}{\sin A} = \frac{\sin b}{\sin B} = \frac{\sin c}{\sin C}. \qquad \triangle$$

It is natural to ask (cf Problem **32**):

> "If the three ratios in Problem **219** are all equal, what is it that they are all equal to?"

The answer may not at first seem quite as nice as in the Euclidean 2-dimensional case: one answer is that they are all equal to

$$\frac{\sin a \cdot \sin b \cdot \sin c}{\text{volume of the tetrahedron } OABC}.$$

Notice that this echoes the result in the Euclidean plane, where the three ratios in the Sine Rule are all equal to $2R$, and

$$2R = \frac{abc}{2(\text{area of } \triangle ABC)}.$$

5.11. Loci and conic sections

This section offers a brief introduction to certain classically important *loci* in the plane. The word *locus* here refers to the set of all points satisfying some simple geometrical condition; and all the examples in this section are based on the notion of distance from a point and from a line.

Given a point O and a positive real r, the locus of points at distance r from O is precisely the circle of radius r with centre O. If $r < 0$, then the locus is empty; while if $r = 0$, the locus consists of the point O alone.

Given a line m and a positive real r, the locus of all points at distance r from the line m consists of a pair of parallel lines – one either side of the line m. Given a circle of radius r, and a positive real number $d < r$; the locus of points at distance d from the circle consists of two circles, each concentric with the given circle (one inside the given circle and one outside). If $d > r$, the locus consists of a single circle outside the given circle.

Given two points A and B, the locus of points which are equidistant from A and from B is precisely the perpendicular bisector of the line segment $\underline{AB}$. And given two lines m, n the locus of points which are equidistant from m and from n takes different forms according as m and n are, or are not, parallel.

- If m and n are parallel, then the locus consists of a single line parallel to m and n and half way between them.
- If m and n meet at X (say), then the locus consists of the pair of perpendicular lines through X, that bisect the four angles at X.

Problem 220 Given a point F and a line m, choose m as the x-axis and the line through F perpendicular to m as the y-axis. Let F have coordinates $(0, 2a)$.

(i) Find the equation that defines the locus of points which are equidistant from F and from m.

(ii) Does the equation suggest a more natural choice of axes – and hence a simpler equation for the locus? △

The locus, or curve, in Problem **220** is called a *parabola*; the point F is called the *focus* of the parabola, and the line m is called the *directrix*. In general, the ratio

"the distance from X to F" : "the distance from X to m"

is called the *eccentricity* of the curve. Hence the parabola has eccentricity $e = 1$.

The parabola has many wonderful properties: for example, it is the path followed by a projectile under the force of gravity; if viewed as the surface of a mirror, a parabola reflects the sun's rays (or any parallel beam) to a single point – the focus F. Since the only variable in the construction of the parabola is the distance "$2a$" between the focus and the directrix, we can scale distances to see that any two different-looking parabolas must in fact be *similar to one another* – just as with any two circles. (It is hard not to infer from the graphs that $y = 10x^2$ is a "thin" parabola, and that

$y = \left(\frac{1}{10}\right) x^2$ is a “fat” parabola. But the first can be rewritten in the form $10y = (10x)^2$, and the second can be rewritten in the form $\left(\frac{y}{10}\right) = \left(\frac{x}{10}\right)^2$, so each is a re-scaled version of $Y = X^2$.)

So far we have considered loci defined by some pair of distances being equal, or in the ratio $1 : 1$. More interesting things begin to happen when we consider conditions in which two distances are in a fixed ratio other than $1 : 1$.

Problem 221

(a) Given two points A, B, with $\underline{AB} = 6$. Find the locus of all points X such that $\underline{AX} : \underline{BX} = 2 : 1$.

(b) Given points A, B, with $\underline{AB} = 2b$ and a positive real number f. Find the locus of all points X such that $\underline{AX} : \underline{BX} = f : 1$. △

Problem 222

(a) Given points A, B, with $\underline{AB} = 2c$ and a real number $a > c$. Find the locus of all points X such that $\underline{AX} + \underline{BX} = 2a$.

(b) Given a point F and a line m, find the locus of all points X such that the ratio

distance from X to the point F : distance from X to the line m

is a positive constant $e < 1$.

(c) Prove that parts (a) and (b) give different ways of specifying the same curve, or locus. △

Problem 223

(a) Given points A, B, with $\underline{AB} = 2c$, and a positive real number a. Find the locus of all points X such that $|\underline{AX} - \underline{BX}| = 2a$.

(b) Given a point F and a line m, find the locus of all points X such that the ratio

distance from X to the point F : distance from X to the line m

is a constant $e > 1$.

(c) Prove that parts (a) and (b) give different ways of specifying the same curve, or locus. △

Problem **221** is sometimes presented in the form of a mild joke.

> Two dragons are sleeping, one at A and one at B. Dragon A can run twice as fast as dragon B. A specimen of *homo sapiens* is positioned on the line segment $\underline{AB}$, twice as far from A as from B, and cunningly decides to crawl quietly away, while maintaining the ratio of his distances from A and from B (so as to make it equally difficult for either dragon to catch him should they wake).

The locus that emerges generally comes as a surprise: if the man sticks to his imposed restriction, by moving so that his position X satisfies $\underline{XA} = 2\cdot\underline{XB}$, then he follows a circle and lands back where he started! The circle is called the *circle of Apollonius*, and the points A and B are sometimes referred to as its foci.

The locus in Problem **222** is an *ellipse* – with *foci* A (or $F = (-ae, 0)$) and B ($= (ae, 0)$), and with directrix m (the line $y = -\frac{a}{e}$; the line $y = \frac{a}{e}$ is the second directrix of the ellipse). The "focus-focus" description in part (a) is symmetrical under reflection in both the line AB and the perpendicular bisector of $\underline{AB}$. The "focus-directrix" description in (b) is clearly symmetrical in the line through F perpendicular to m; but it is a surprise to find that the equation

$$\frac{x^2}{a^2} + \frac{y^2}{a^2(1-e^2)} = 1$$

is also symmetrical under reflection in the y-axis. If we set $b^2 = a^2(1-e^2)$, the equation takes the form

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1,$$

which crosses the x-axis when $x = \pm a$, and crosses the y-axis when $y = \pm b$. In its standard form, we usually choose coordinates so that $b < a$: the line segment from $(-a, 0)$ to $(a, 0)$ is then called the *major axis*, and half of it (say from $(0, 0)$ to $(a, 0)$) – of length a – is called the *semi-major axis*; the line segment from $(0, -b)$ to $(0, b)$ is called the *minor axis*, and half of it (say from $(0, 0)$ to $(0, b)$) – of length b – is called the *semi-minor axis*.

The form of the equation shows that an ellipse is obtained from a unit circle by stretching by a factor "a" in the x-direction, and by a factor "b" in the y-direction. This implies that the area of an ellipse is equal to πab (since each small s by s square that arises in the definition of the "area" of the unit circle gets stretched into an "as by bs rectangle"). However, the equation tells us nothing about the *perimeter* of an ellipse. Attempts to pin down the perimeter of an ellipse gave rise in the 18$^{\text{th}}$ century to the subject of "elliptic integrals".

Like parabolas, ellipses arise naturally in many important settings. For example, Kepler (1571–1630) discovered that the planetary orbits are not circular (as had previously been believed), but are ellipses – with the Sun at one focus (a conjecture which was later explained by Isaac Newton (1642–1727)). Moreover, the tangent to an ellipse at any point X is equally inclined to the two lines XA and XB, so that a beam emerging from one focus is reflected at every point of the ellipse so that all the reflected rays pass through the other focus.

The curve in Problem **223** is a *hyperbola* – with *foci* A (or $F = (-ae, 0)$) and B $(= (ae, 0))$, and with *directrix* m (the line $y = -\frac{a}{e}$; the line $y = \frac{a}{e}$ is the second directrix of the hyperbola). The "focus-focus" description in part (a) is symmetrical under reflection in both the line AB and the perpendicular bisector of $\underline{AB}$. The "focus-directrix" description in (b) is clearly symmetrical in the line through F perpendicular to m; but it is a surprise to find that the equation

$$\frac{x^2}{a^2} - \frac{y^2}{a^2(e^2-1)} = 1$$

is also symmetrical under reflection in the y-axis. Like parabolas and ellipses, hyperbolas arise naturally in many important settings – in mathematics and in the natural sciences.

All these loci were introduced and studied by the ancient Greeks without the benefit of coordinate geometry and equations. They were introduced as planar *cross-sections of a cone* – that is, as natural extensions of straight lines and circles (since the doubly infinite cone is the surface traced out when one rotates a line about an axis through a point on that line). The equivalence of the focus-directrix definition in Problems **220**, **222**, and **223** and cross sections of a cone follows from the next problem. All five constructions in Problem **224** work with the doubly-infinite cone, which we may represent as $x^2+y^2 = (rz)^2$ – although this representation is not strictly needed for the derivations. The surface of the double cone extends to infinity in both directions, and is obtained by taking the line $y = rz$ in the yz-plane (where $r > 0$ is constant), and rotating it about the z-axis. Images of this rotated line are called *generators* of the cone; and the point they all pass through (i.e. the origin) is called the *apex* of the cone.

Problem 224 (**Dandelin's spheres**: Dandelin (1794–1847))

(a) Describe the cross-sections obtained by cutting such a double cone by a horizontal plane (i.e. a plane perpendicular to the z-axis). What if the cutting plane is the xy-plane?

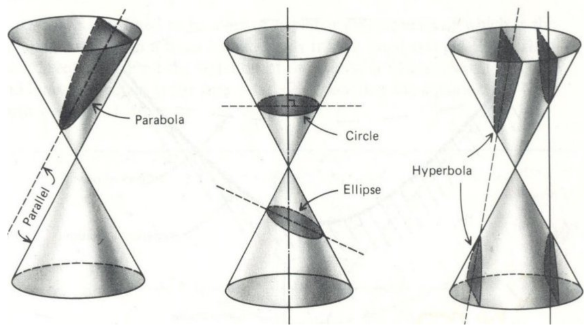


Figure 7: Conic sections.

(b)(i) Describe the cross-section obtained by cutting such a cone by a vertical plane through the origin, or apex.

(ii) What cross-section is obtained if the cutting plane passes through the apex, but is not vertical?

(c) Give a *qualitative* description of the curve obtained as a cross-section of the cone if we cut the cone by a plane which is parallel to a generator: e.g. the plane $y - rz = c$.

(i) What happens if $c = 0$?

(ii) Now assume the cutting plane is parallel to a generator, but does not pass through the apex of the cone – so we may assume that the plane cuts only the bottom half of the cone. Insert a small sphere inside the bottom half of the cone and above the cutting plane, and inflate the sphere as much as possible – until it touches the cone around a horizontal circle (the "contact circle with the cone"), and touches the plane at a single point F. Let the horizontal plane of the "contact circle with the cone" meet the cutting plane in the line m. Prove that each point of the cross-sectional curve is equidistant from the point F and from the line m – and so is a parabola.

(d)(i) Give a *qualitative* description of the curve obtained as a cross-section of the cone if we cut the cone by a plane which is *less steep* than a generator, but does not pass through the apex – and so cuts right across the cone.

(ii) We may assume that the plane cuts only the bottom half of the cone. Insert a small sphere inside the bottom half of the cone and above the cutting plane (i.e. on the same side of the cutting plane as the apex of the cone), and inflate the sphere as much as possible – until it touches the cone around a horizontal circle, and touches the plane at a single point F. Let the horizontal plane of the contact circle meet the cutting plane in the line m. Prove that, for each point X on the cross-sectional curve, the ratio

$$\text{"distance from } X \text{ to } F\text{"} : \text{"distance from } X \text{ to } m\text{"} = e : 1$$

is constant, with $e < 1$, and so is an ellipse.

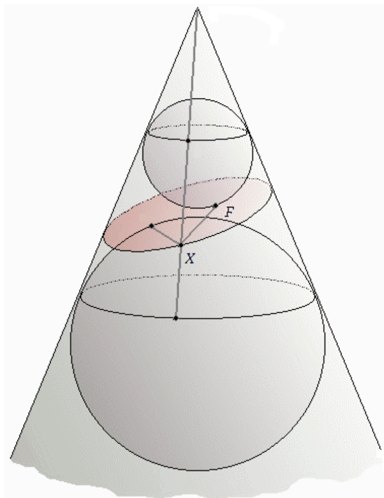


Figure 8: The conic section arising in Problem **224**(d).

(e)(i) Give a *qualitative* description of the curve obtained as a cross-section if we cut the cone by a plane which is *steeper* than a generator, but does not pass through the apex (and hence cuts both halves of the cone)?

(ii) We can be sure that the plane cuts the bottom half of the cone (as well as the top half). Insert a small sphere inside the bottom half of the cone and on the same side of the cutting plane as the apex, and inflate the sphere as much as possible – until it touches the cone around a horizontal circle, and touches the plane at a single point F. Let the horizontal plane of the contact circle meet the cutting plane in the line m. Prove that, for each point X on the cross-sectional curve, the ratio

$$\text{"distance from } X \text{ to } F\text{"} : \text{"distance from } X \text{ to } m\text{"} = e : 1$$

is constant, with $e > 1$, and so is a hyperbola. △

Problem **224** reveals a remarkable correspondence. It is not hard to show algebraically that any quadratic equation in two variables x, y represents either a point, or a pair of crossing (possibly identical) straight lines, or a parabola, or an ellipse, or a hyperbola: that is, by changing coordinates, the quadratic equation can be transformed to one of the standard forms obtained in this section. Hence, *the possible quadratic curves* are precisely the same as the possible cross-sections of a cone. This remarkable equivalence is further reinforced by the many natural contexts in which these conic sections arise.

5.12. Cubes in higher dimensions

This final section on elementary geometry seeks to explore fresh territory by going beyond three dimensions. Whenever we try to jump up to a new level, it can help to first take a step *back* and 'take a longer run up'. So please be patient if we initially take a step or two backwards.

We all know what a unit 3D-cube is. And – going backwards – it is not hard to guess what is meant by a unit "2D-cube": a unit 2D-cube is just another name for a unit *square*. It is then not hard to notice that a unit 3D-cube can be constructed from two unit 2D-cubes as follows:

- first position two unit 2D-cubes 1 unit apart in 3D space, with one directly above the other;
- make sure that each vertex of the lower 2D-cube is directly beneath a vertex of the upper 2D-cube;
- then join each vertex of the upper 2D-cube to the corresponding vertex below it.

Perhaps a unit 2D-cube can be constructed in a similar way from "unit 1D-cubes"! This idea suggests that a unit "1D-cube" is just another name for a unit *line segment*.

Take the unit 1D-cube to be the line segment from 0 to 1:

- position two such 1D-cubes in 2D (e.g. one joining $(0, 0)$ to $(1, 0)$, and the other joining $(0, 1)$ to $(1, 1)$);
- check that each vertex of the lower 1D-cube is directly beneath a vertex of the upper 1D-cube;
- then join corresponding pairs of vertices – one from the upper 1D-cube and one from the lower 1D-cube ($(0, 0)$ to $(0, 1)$, and $(1, 0)$ to $(1, 1)$) – to obtain a unit 2D-cube.

Having taken a step back, we repeat (and reformulate) the previous construction of a 3D-cube:

- position two such unit 2D-cubes in 3D: with one 2D-cube joining $(0, 0, 0)$ to $(1, 0, 0)$, then to $(1, 1, 0)$, then to $(0, 1, 0)$ and back to $(0, 0, 0)$, and the other 2D-cube joining $(0, 0, 1)$ to $(1, 0, 1)$, then to $(1, 1, 1)$, then to $(0, 1, 1)$, and back to $(1, 0, 0)$;

- with one 2D-cube directly above the other,
- then join corresponding pairs of vertices – one from the upper 2D-cube and one from the lower 2D-cube ($(0, 0, 0)$ to $(0, 0, 1)$, and $(1, 0, 0)$ to $(1, 0, 1)$, and $(1, 1, 0)$ to $(1, 1, 1)$, and $(0, 1, 0)$ to $(0, 1, 1)$) – to obtain a unit 3D-cube.

To sum up: a unit cube in 1D, or in 2D, or in 3D:

- has as "vertices" all points whose coordinates are all "0s or 1s" (in 1D, or 2D, or 3D)
- has as "edges" all the unit segments (or unit 1D-cubes) joining vertices whose coordinates differ in exactly one place
- and a unit 3D-cube has as "faces" all the unit 2D-cubes spanned by vertices with a constant value (0 or 1) in one of the three coordinate places (that is, for the unit 3D-cube: the four vertices with $x = 0$, or the four vertices with $x = 1$; or the four vertices with $y = 0$, or the four vertices with $y = 1$; or the four vertices with $z = 0$, or the four vertices with $z = 1$).

A 3D-cube is surrounded by six 2D-cubes (or faces), and a 2D-cube is surrounded by four 1D-cubes (or faces). So it is natural to interpret the two end vertices of a 1D-cube as being '0D-cubes'. We can then see that a cube in any dimension is made up from cubes of smaller dimensions. We can also begin to make a reasonable guess as to *what we might expect to find in a '4D-cube'*.

Problem 225

(a)(i) How many vertices (i.e. 0D-cubes) are there in a 1D-cube?

(ii) How many edges (i.e. 1D-cubes) are there in a 1D-cube?

(b)(i) How many vertices (or 0D-cubes) are there in a 2D-cube?

(ii) How many "faces" (i.e. 2D-cubes) are there in a 2D-cube?

(iii) How many edges (i.e. 1D-cubes) are there in a 2D-cube?

(c)(i) How many vertices (or 0D-cubes) are there in a 3D-cube?

(ii) How many 3D-cubes are there in a 3D-cube?

(iii) How many edges (i.e. 1D-cubes) are there in a 3D-cube?

(iv) How many "faces" (i.e. 2D-cubes) are there in a 3D-cube?

(d)(i) How many vertices (or 0D-cubes) do you expect to find in a 4D-cube?

(ii) How many 4D-cubes do you expect to find in a 4D-cube?

(iii) How many edges (i.e. 1D-cubes) do you expect to find in a 4D-cube?

(iv) How many "faces" (i.e. 2D-cubes) do you expect to find in a 4D-cube?

(v) How many 3D-cubes do you expect to find in a 4D-cube? △

Problem 226

(a)(i) Sketch a unit 2D-cube as follows. Starting with two unit 1D-cubes – one directly above the other. Then join up each vertex in the upper 1D-cube to the vertex it corresponds to in the lower 1D-cube (directly beneath it).

(ii) Label each vertex of your sketch with coordinates (x, y) $(x, y = 0$ or $1)$ so that the lower 2D-cube has the equation "$y = 0$" and the upper 2D-cube has the equation "$y = 1$".

(b)(i) Sketch a unit 3D-cube, starting with two unit 2D-cubes – one directly above the other. Then join up each vertex in the upper 2D-cube to the vertex it corresponds to in the lower 2D-cube (directly beneath it).

(ii) Label each vertex of your sketch with coordinates (x, y, z) (where each $x, y, z = 0$ or 1) so that the lower 2D-cube has the equation "$z = 0$" and the upper 2D-cube has the equation "$z = 1$".

(c)(i) Now sketch a unit 4D-cube in the same way – starting with two unit 3D-cubes, one "directly above" the other.

[**Hint:** In part (b) your sketch was a projection of a 3D-cube onto 2D paper, and this forced you to represent the lower and upper 2D-cubes as rhombuses rather than genuine 2D-cubes (unit squares). In part (c) you face the even more difficult task of representing a 4D-cube on 2D paper; so you must be prepared for other "distortions". In particular, it is almost impossible to see what is going on if you try to physically position one 3D-cube "directly above" the other on 2D paper. So start with the "upper" unit 3D-cube towards the top right of your paper, and then position the "lower" unit 3D-cube not directly below it on the paper, but below and slightly to the left, before pairing off and joining up each vertex of the upper 3D-cube with the corresponding vertex in the lower 3D-cube.]

(ii) Label each vertex of your sketch with coordinates (w, x, y, z) (where each $w, x, y, z = 0$ or 1) so that the lower 3D-cube has the equation "$z = 0$" and the upper 3D-cube has the equation "$z = 1$". △

Problem 227 The only possible path along the edges of a 2D-cube uses each vertex once and returns to the start after visiting all four vertices.

(a)(i) Draw a path along the edges of a 3D-cube that visits each vertex exactly once and returns to the start.

(ii) Look at the sequence of coordinate triples as you follow your path. What do you notice?

(b)(i) Draw a path along the edges of a 4D-cube that visits each vertex exactly once and returns to the start.

(ii) Look at the sequence of coordinate 4-tuples as you follow your path. What do you notice? △

5.13. Chapter 5: Comments and solutions

137. Note: The spirit of constructions restricts us to:

- drawing the line joining two known points
- drawing the circle with centre a known point and passing through a known point.

The whole thrust of this first problem is to find some way to "jump" from A (or B) to C. So the problem leaves us with very little choice; $\underline{AB}$ is given, and A and B are more-or-less indistinguishable, so there are only two possible 'first moves' – both of which work with the line segment $\underline{AC}$ (or $\underline{BC}$).

Join AC.

Then construct the point X such that $\triangle ACX$ is equilateral (i.e. use Euclid's *Elements*, Book I, Proposition 1). Construct the circle with centre A which passes through B; let this circle meet the line AX at the point Y, where either

(i) Y lies on the segment $\underline{AX}$ (if $\underline{AB} \leqslant \underline{AC}$), or

(ii) Y lies on $\underline{AX}$ produced (i.e. beyond X, if $\underline{AB} > \underline{AC}$).

In each case, $\underline{AY} = \underline{AB}$. Finally construct the circle with centre X which passes through Y. In case (i), let the circle meet the line segment $\underline{XC}$ at D; in case (ii), let the circle meet $\underline{CX}$ produced (beyond X) at D.

In case (i), $\underline{CX} = \underline{CD} + \underline{DX}$; therefore

$$\begin{aligned}\underline{CD} &= \underline{CX} - \underline{DX}\\ &= \underline{AX} - \underline{YX}\\ &= \underline{AY} = \underline{AB}.\end{aligned}$$

In case (ii),

$$\begin{aligned}\underline{CD} &= \underline{CX} + \underline{XD}\\ &= \underline{AX} + \underline{XY}\\ &= \underline{AY} = \underline{AB}.\end{aligned}$$

QED

138.

$$\begin{aligned}\angle AXC &= \angle AXB - \angle CXB\\ &= \angle CXD - \angle CXB \quad \text{(since the two straight angles}\\ &\qquad\qquad\qquad\qquad\quad \angle AXB \text{ and } \angle CXD \text{ are equal)}\\ &= \angle BXD.\end{aligned}$$

QED

139.

$\underline{AM} = \underline{AM}$
$\underline{MB} = \underline{MC}$ (by construction of the midpoint M)
$\underline{BA} = \underline{CA}$ (given).
$\therefore \triangle AMB \equiv \triangle AMC$ (by SSS-congruence)
$\therefore \angle AMB = \angle AMC$, so each angle is exactly half the straight angle $\angle BMC$.
Hence AM is perpendicular to BC. QED

140. Let $ABCDEF$ be a regular hexagon with sides of length 1. Then $\triangle ABC$ (formed by the first three vertices) satisfies the given constraints, with $\angle ABC = 120°$.

Let $\triangle B'C'D$ be an equilateral triangle with sides of length 2, and with A' the midpoint of $B'D$. Then $\triangle A'B'C'$ satisfies the given constraints with angle $\angle A'B'C' = 60°$.

141.

(i) Join $\underline{CO}$. Then $\triangle ACO$ is isosceles (since $\underline{OA} = \underline{OC}$) and $\triangle BCO$ is isosceles (since $\underline{OB} = \underline{OC}$).
Hence $\angle OAC = \angle OCA = x$ (say), and $\angle OBC = \angle OCB = y$ (say).
So $\angle C = x + y$ (and $\angle A + \angle B + \angle C = x + (x + y) + y = 2(x + y)$). QED

(ii) If $\angle A + \angle B + \angle C = 2(x + y)$ is equal to a straight angle, then $\angle C = x + y$ is half a straight angle, and hence a right angle. QED

142.

(i) Draw the circle with centre A and passing through B, and the circle with centre B passing through A. Let these two circles meet at C and D.

Wherever the midpoint M of $\underline{AB}$ may be, we know from Euclid Book I, Proposition 1 and Problem **139**:

> that $\triangle ABC$ is equilateral, and that CM is perpendicular to AB, and
> that $\triangle ABD$ is equilateral, and that DM is perpendicular to AB.

Hence CMD is a straight line.

So if we join CD, then this line cuts $\underline{AB}$ at its midpoint M. QED

(ii) We may suppose that $\underline{BA} \leqslant \underline{BC}$.

Then the circle with centre B, passing through A meets $\underline{BC}$ internally at A' (say).

Let the circle with centre A and passing through B meet the circle with centre A' and passing through B at the point D.

Claim BD bisects $\angle ABC$.
Proof

$$\begin{aligned} \underline{BA} &= \underline{BA'} \text{ (radii of the same circle with centre } B) \\ \underline{AD} &= \underline{AB} \text{ (radii of the same circle with centre } A) \\ &= \underline{A'B} = \underline{A'D} \text{ (radii of same circle with centre } A') \\ \underline{BD} &= \underline{BD} \end{aligned}$$

Hence $\triangle BAD \equiv \triangle BA'D$ (by SSS-congruence).
$\therefore \angle ABD = \angle A'BD$. QED

(iii) Suppose first that $\underline{PA} = \underline{PB}$. Then the circle with centre P and passing through A meets the line AB again at B. If we construct the midpoint M of $\underline{AB}$ as in part (i), then PM will be perpendicular to AB.

Now suppose that one of $\underline{PA}$ and $\underline{PB}$ is longer than the other. We may suppose that $\underline{PA} > \underline{PB}$, so B lies inside the circle with centre P and passing through A. Hence this circle meets the line AB again at A' where B lies between A and A'.

If we now construct the midpoint M of $\underline{AA'}$ as in part (i), then PM will be perpendicular to AA', and hence to AB. QED

143. Let M be the midpoint of $\underline{AB}$.

(a) Let X lie on the perpendicular bisector of $\underline{AB}$.

$\therefore \triangle XMA \equiv \triangle XMB$ (by SAS congruence, since $\underline{XM} = \underline{XM}$, $\angle XMA = \angle XMB$, $\underline{MA} = \underline{MB}$)

$\therefore \underline{XA} = \underline{XB}$.

(b) If X is equidistant from A and from B, then $\triangle XMA \equiv \triangle XMB$ (by SSS-congruence, since $\underline{XM} = \underline{XM}$, $\underline{MA} = \underline{MB}$, $\underline{AX} = \underline{BX}$).

$\therefore \angle XMA = \angle XMB$, so each must be exactly half a straight angle.

$\therefore X$ lies on the perpendicular bisector of $\underline{AB}$. QED

144. Let X lie on the plane perpendicular to $\underline{NS}$, through the midpoint M.
$\therefore \triangle XMN \equiv \triangle XMS$ (by SAS congruence, since $\underline{XM} = \underline{XM}$, $\angle XMN = \angle XMS$, $\underline{MN} = \underline{MS}$)
$\therefore \underline{XN} = \underline{XS}$.

Let X be equidistant from N and from S, then $\triangle XMN \equiv \triangle XMS$ (by SSS-congruence, since $\underline{XM} = \underline{XM}$, $\underline{MN} = \underline{MS}$, $\underline{NX} = \underline{SX}$).

$\therefore \angle XMN = \angle XMS$, so each must be exactly half a straight angle.

$\therefore X$ lies on the perpendicular bisector of $\underline{NS}$. QED

145. Let M be the midpoint of $\underline{AC}$. Join BM and extend the line beyond M to the point D such that $\underline{MB} = \underline{MD}$. Join CD. Then $\triangle AMB \equiv \triangle CMD$ (by SAS-congruence, since

$\underline{AM} = \underline{CM}$ (by construction of the midpoint M)
$\angle AMB = \angle CMD$ (vertically opposite angles)
$\underline{MB} = \underline{MD}$ (by construction)).

$\therefore \angle DCM = \angle BAM$.
Now

$$\begin{aligned} \angle ACX &= \angle DCM + \angle DCX \\ &> \angle DCM \\ &= \angle BAM \\ &= \angle A. \end{aligned}$$

Hence $\angle ACX > \angle A$.

Similarly, we can extend $\underline{AC}$ beyond C to a point Y. Let N be the midpoint of $\underline{BC}$. Join AN and extend the line beyond N to the point E such that $\underline{NA} = \underline{NE}$. Join CE.
Then $\triangle BNA \equiv \triangle CNE$ (again by SAS-congruence).
$\therefore \angle BCY > \angle BCE = \angle CBA = \angle B$. QED

146.

(a) Suppose that $\underline{AB} > \underline{AC}$.
Let the circle with centre A, passing through C, meet $\underline{AB}$ (internally) at X.
Then $\triangle ACX$ is isosceles, so $\angle ACX = \angle AXC$.
By Problem **145**, $\angle AXC > \angle ABC$.
$\therefore \angle ACB\ (> \angle ACX = \angle AXC) > \angle ABC$. QED

(b) Suppose the conclusion does not hold. Then either

(i) $\underline{AB} = \underline{AC}$, or
(ii) $\underline{AC} > \underline{AB}$.

(i) If $\underline{AB} = \underline{AC}$, then $\triangle ABC$ is isosceles, so $\angle ACB = \angle ABC$ – contrary to assumption.

(ii) If $\underline{AC} > \underline{AB}$, then $\angle ABC > \angle ACB$ (by part (a)) – again contrary to assumption.

Hence, if $\angle ACB > \angle ABC$, it follows that $\underline{AB} > \underline{AC}$. QED

(c) Extend $\underline{AB}$ beyond B to the point D, such that $\underline{BD} = \underline{BC}$.
Then $\triangle BDC$ is isosceles with apex B, so $\angle BDC = \angle BCD$.
Now

$$\angle ACD = \angle ACB + \angle BCD > \angle BCD = \angle BDC.$$

Hence, by part (b), $\underline{AD} > \underline{AC}$.
By construction,

$$\underline{AD} = \underline{AB} + \underline{BD} = \underline{AB} + \underline{BC},$$

so $\underline{AB} + \underline{BC} > \underline{AC}$. QED

147. Suppose $\angle C = \angle A + \angle B$, but that C does not lie on the circle with diameter $\underline{AB}$. Then C lies either inside, or outside the circle. Let O be the midpoint of $\underline{AB}$.

(i) If C lies outside the circle, then $\underline{OC} > \underline{OA} = \underline{OB}$.
$\therefore \angle OAC > \angle OCA$ and $\angle OBC > \angle OCB$ (by Problem **146**(a)).
$\therefore \angle C = \angle OCA + \angle OCB < \angle A + \angle B$ – contrary to assumption.
Hence C does not lie outside the circle.

(ii) If C lies inside the circle, then $\underline{OC} < \underline{OA} = \underline{OB}$.
$\therefore \angle OAC < \angle OCA$ and $\angle OBC < \angle OCB$ (by Problem **146**(a)).
$\therefore \angle C = \angle OCA + \angle OCB > \angle A + \angle B$ – contrary to assumption.
Hence C does not lie inside the circle.
Hence C lies on the circle with diameter $\underline{AB}$. QED

148. Suppose, to the contrary, that OP is **not** perpendicular to the tangent at P.

Drop a perpendicular from O to the tangent at P to meet the tangent at Q. Extend $\underline{PQ}$ beyond Q to some point X. Then $\angle OQP$ and $\angle OQX$ are both right angles.

Since Q ($\neq P$) lies on the tangent, Q lies outside the circle, so $\underline{OQ} > \underline{OP}$. Hence (by Problem **146**(a)) $\angle OPQ > \angle OQP = \angle OQX$ – contrary to the fact that $\angle OQX > \angle OPQ$ (by Problem **145**). QED

149. Let Q lie on the line m such that $\underline{PQ}$ is perpendicular to m.
Let X be any other point on the line m, and let Y be a point on m such that Q lies between X and Y.
Then $\angle PQX$ and $\angle PQY$ are both right angles.
Suppose that $\underline{PX} < \underline{PQ}$.
Then $\angle PQY = \angle PQX < \angle PXQ$ (by Problem **146**(a)), which contradicts Problem **145** (since $\angle PQY$ is an exterior angle of $\triangle PQX$).
Hence $\underline{PX} \geqslant \underline{PQ}$ as required. QED

150. $\angle A + \angle B + \angle C$, and $\angle XCA + \angle C$ are both equal to a straight angle. So $\angle A + \angle B = \angle XCA$.

151. Join $\underline{OA}$, $\underline{OB}$, $\underline{OC}$. Since these radii are all equal, this produces three isosceles triangles. There are **five** cases to consider.

(i) Suppose first that O lies on AB. Then AB is a diameter, so $\angle ACB$ is a right angle (by Problem **141**), $\angle AOB$ is a straight angle, and the result holds.

(ii) Suppose O lies on AC, or on BC. These are similar, so we may assume that O lies on AC.
Then $\triangle OBC$ is isosceles, so $\angle OBC = \angle OCB$.
$\angle AOB$ is the exterior angle of $\triangle OBC$, so

$$\angle AOB = \angle OBC + \angle OCB = 2 \cdot \angle ACB$$

(by Problem **150**).

(iii) Suppose O lies inside $\triangle ABC$.
$\triangle OAB$, $\triangle OBC$, $\triangle OCA$ are isosceles, so let $\angle OAB = \angle OBA = x$, $\angle OBC = \angle OCB = y$, $\angle OCA = \angle OAC = z$.
Then $\angle ACB = y + z$, $\angle ABC = x + y$, $\angle BAC = x + z$.
The three angles of $\triangle ABC$ add to a straight angle, so $2(x + y + z)$ equals a straight angle.
Hence, in $\triangle OBA$,

$$\angle AOB = 2(x + y + z) - (\angle OAB + \angle OBA) = 2(y + z) = 2 \cdot \angle ACB.$$

(iv) Suppose O lies outside $\triangle ABC$ with O and B on opposite sides of AC.
$\triangle OAB$, $\triangle OBC$, $\triangle OCA$ are isosceles, so let $\angle OAB = \angle OBA = x$, $\angle OBC = \angle OCB = y$, $\angle OCA = \angle OAC = z$.
Then $\angle ACB = y - z$, $\angle ABC = x + y$, $\angle BAC = x - z$.
The three angles of $\triangle ABC$ add to a straight angle, so $2x + 2y - 2z$ equals a straight angle.
Hence

$$2x + 2y - 2z = \angle AOB + \angle OAB + \angle OBA = \angle AOB + 2x,$$

so $\angle AOB = 2y - 2z = 2 \cdot \angle ACB$.

(v) Suppose O lies outside $\triangle ABC$ with O and B on the same side of AC.
$\triangle OAB$, $\triangle OBC$, $\triangle OCA$ are isosceles, so let $\angle OAB = \angle OBA = x$, $\angle OBC = \angle OCB = y$, $\angle OCA = \angle OAC = z$.
Then $\angle ACB = y + z$, $\angle ABC = y - x$, $\angle BAC = z - x$.
The three angles of ABC add to a straight angle, so $2y+2z-2x$ equals a straight angle.
Since C lies on the minor arc relative to the chord $\underline{AB}$, we need to interpret "the angle subtended by the chord $\underline{AB}$ at the centre O" as the *reflex* angle **outside** the triangle $\triangle AOB$ – which is equal to "$2x$ more than a straight angle", so $\angle AOB = 2y + 2z = 2 \cdot \angle ACB$. QED

152. The chord $\underline{AB}$ subtends angles at C and at D on the same arc. Similarly $\underline{BC}$ subtends angles at A and at D on the same arc. Hence (by Problem **151**) $\angle ACB = \angle ADB$, and $\angle BAC = \angle BDC$.
Hence

$$\angle ADC = \angle ADB + \angle BDC = \angle ACB + \angle BAC,$$

so $\angle ADC + \angle ABC$ equals the sum of the three angles in $\triangle ABC$. QED

153. Let O be the circumcentre of $\triangle ABC$, and let $\angle XAB = x$.
Then $\angle XAO$ is a right angle, and $\triangle OAB$ is isosceles. There are two cases.

(i) If X and O lie on opposite sides of AB, then $\angle OBA = \angle OAB = 90^\circ - x$.
$\therefore \angle AOB = 2x$.
$\therefore \angle ACB = x = \angle XAB$.

(ii) If X and O lie on the same side of AB, then Y and O lie on opposite sides of AB and of AC. Hence we can apply case (i) (with Y in place of X, AC in place of AB, $\angle YAC$ in place of $\angle XAB$, and $\angle ABC$ in place of $\angle ACB$) to conclude that $\angle YAC = \angle ABC$. Hence

$$\angle XAB + \angle BAC + \angle YAC = \angle XAB + \angle BAC + \angle ABC$$

are both straight angles, so $\angle XAB = \angle ACB$ (since the three angles of $\triangle ABC$ also add to a straight angle). QED

154.

(a)(i) Extend AD to meet the circumcircle at X. Then (applying Problem **145** to $\triangle DXB$), the exterior angle $\angle ADB > \angle DXB = \angle AXB = \angle ACB$.

(ii) We are told that the points C, Dlie "on the same side of the line AB". This"side" of the line AB (or "half-plane") is split into three parts by the half-lines"$\underline{AC}$produced beyond C" and "$\underline{BC}$produced beyond C".There are two very different possibilities.
Suppose first that Dlies in one of the two overlapping wedge-shaped regions"between $\underline{AB}$produced and$\underline{AC}$produced" or "between $\underline{BA}$produced and$\underline{BC}$produced". Then either DAorDBcuts the arc ABat a point X(say), and$\angle ACB = \angle AXB > \angle ADB$ (by Problem **145**applied to $\triangle AXD$or to $\triangle BXD$).
The only alternative is that Dlies in the wedge shaped region outside the circleat the point C,lying "between$\underline{AC}$produced and $\underline{BC}$produced". Then C liesinside $\triangle ADB$, so C lies inside the circumcircle of $\triangle ADB$. Hencepart (i) implies that $\angle ACB < \angle ADB$as required. QED

(b) If D does not lie on the circumcircle of $\triangle ABC$, then it lies either inside, or outside the circle – in which case $\angle ADB \neq \angle ACB$ (by part (a)).

(c) $\angle D$ is less than a straight angle, so D must lie outside $\triangle ABC$. Moreover, B, D lie on opposite sides of the line AC (since the edges $\underline{BC}$, $\underline{DA}$ do not cross internally, and neither do the edges $\underline{CD}, \underline{AB}$). Consider the circumcircle of $\triangle ABC$, and let X be any point on the circle such that X and D lie on the same side of the line AC. Then $ABCX$ is a cyclic quadrilateral, so $\angle ABC$ and $\angle CXA$ are supplementary. Hence $\angle CXA = \angle CDA = \angle D$. But then D lies on the circle (by part (b)).

155. $\triangle OPQ \equiv \triangle OP'Q$ (by RHS-congruence: $\angle OPQ = \angle OP'Q$ are both right angles, $\underline{OP} = \underline{OP'}$, $\underline{OQ} = \underline{OQ}$)
$\therefore \underline{QP} = \underline{QP'}$, and $\angle QOP = \angle QOP'$. QED

156.

(a) Let X be any point on the bisector of $\angle ABC$. Drop the perpendiculars XY from X to AB and XZ from X to CB. Then: $\angle XYB = \angle XZB$ are both right angles by construction; and $\angle XBY = \angle XBZ$ since X lies on the bisector of $\angle YBZ$; hence $\angle BXY = \angle BXZ$ (since the three angles in each triangle have the same sum; so as soon as two of the angles are equal in pairs, the third pair must also be equal). Hence $\triangle BXY \equiv \triangle BXZ$ (by ASA-congruence: $\angle YBX = \angle ZBX$, $\underline{BX} = \underline{BX}$, $\angle BXY = \angle BXZ$.)
$\therefore \underline{XY} = \underline{XZ}$. QED

(b) Suppose X is equidistant from m and from n. Drop the perpendiculars XY from X to m, and XZ from X to n.
Then $\triangle BXY \equiv \triangle BXZ$ (by RHS-congruence:

$\angle BYX = \angle BZX$ are both right angles
$\underline{XY} = \underline{XZ}$, since we are assuming X is equidistant from m and from n
$\underline{BX} = \underline{BX}$).

Hence $\angle XBY = \angle XBZ$, so X lies on the bisector of $\angle YBZ$. QED

157.

(i) Join $\underline{AC}$. Then $\triangle ABC \equiv \triangle CDA$ by ASA-congruence:

$\angle BAC = \angle DCA$ (alternate angles, since $AB \parallel DC$)
$\underline{AC} = \underline{CA}$
$\angle ACB = \angle CAD$ (alternate angles, since $CB \parallel DA$).

In particular, $\triangle ABC$ and $\triangle CDA$ must have equal area, and so each is exactly half of $ABCD$. QED

Note: Once we prove (Problem **161** below) that a parallelogram has the same area as the rectangle on the same base and lying between the same pair of parallels (whose area is equal to "base × height"), the result in part (i) will immediately translate into the familiar formula for the area of the triangle

$$\frac{1}{2}(\text{base} \times \text{height}).$$

(ii) $\triangle ABC \equiv \triangle CDA$ by part (i). Hence $\underline{AB} = \underline{CD}$, $\underline{BC} = \underline{DA}$; and $\angle B = \angle ABC = \angle CDA = \angle D$.
To show that $\angle A = \angle C$, we can either copy part (i) after joining BD to prove that $\triangle BCD \equiv \triangle DAB$, or we can use part (i) directly to note that $\angle A = \angle BAC + \angle DAC = \angle DCA + \angle BCA = \angle C$. QED

(iii) $\triangle AXB \equiv \triangle CXD$ by ASA-congruence:

$\angle XAB = \angle XCD$ (alternate angles, since $AB \parallel DC$)
$\underline{AB} = \underline{CD}$ (by part (ii))
$\angle XBA = \angle XDC$ (alternate angles, since $AB \parallel DC$).

Hence $\underline{XA}$ (in $\triangle AXB$) $=$ $\underline{XC}$ (in $\triangle CXD$), and $\underline{XB}$ (in $\triangle AXB$) $=$ $\underline{XD}$ (in $\triangle CXD$). QED

158. We may assume that m cuts the opposite sides AB at Y and DC at Z. $\triangle XYB \equiv \triangle XZD$ by ASA-congruence:

$\angle YXB = \angle ZXD$ (vertically opposite angles)
$\underline{XB} = \underline{XD}$ (by Problem **157**(iii))
$\angle XBY = \angle XDZ$ (alternate angles).

Therefore

$$\begin{aligned} \text{area}(YZCB) &= \text{area}(\triangle BCD) - \text{area}(\triangle XZD) + \text{area}(\triangle XYB) \\ &= \text{area}(\triangle BCD) \\ &= \frac{1}{2}\text{area}(ABCD). \end{aligned}$$

QED

159.

(a) Join AC. Then $\triangle ABC \equiv \triangle CDA$ by SAS-congruence:

$\underline{BA} = \underline{DC}$ (given)
$\angle BAC = \angle DCA$ (alternate angles, since $AB \parallel DC$)
$\underline{AC} = \underline{CA}$.

Hence $\angle BCA = \angle DAC$, so $AD \parallel BC$ as required. QED

(b) Join AC. Then $\triangle ABC \equiv \triangle CDA$ by SSS-congruence:

$\underline{AB} = \underline{CD}$ (given)
$\underline{BC} = \underline{DA}$ (given)
$\underline{CA} = \underline{AC}$.

Hence $\angle BAC = \angle DCA$, so $AB \parallel DC$; and $\angle BCA = \angle DAC$, so $BC \parallel AD$. QED

(c) $\angle A + \angle B + \angle C + \angle D = 2\angle A + 2\angle B = 2\angle A + 2\angle D$ are each equal to two straight angles.
$\therefore$ $\angle A + \angle B$ is equal to a straight angle, so $AD \parallel BC$; and $\angle A + \angle D$ is equal to a straight angle, so $AB \parallel DC$. QED

Note: The fact that the angles in a quadrilateral add to two straight angles is proved in the next chapter. However, if preferred, it can be proved here directly. If we imagine pins located at A, B, C, D, then a string tied around the four points defines their "convex hull" – which is either a 4-gon (if the string touches all four pins), or a 3-gon (if one vertex is inside the triangle formed by the other three). In the first case, either diagonal (AC or BD) will split the quadrilateral internally into two triangles; in the second case, one of the three 'edges' joining vertices of the convex hull to the internal vertex cannot be an edge of the quadrilateral, and so must be a diagonal, which splits the quadrilateral internally into two triangles.

160. $\underline{AM} = \underline{MD}$ (by construction of M as the midpoint), and $\underline{BN} = \underline{NC}$.
$\therefore$ $\underline{AM} = \underline{BN}$ (since $\underline{AD} = \underline{BC}$ by Problem **157**(ii)).
$\therefore$ $ABNM$ is a parallelogram (by Problem **159**(a)), so $MN \parallel AB$.
Let AC cross MN at Y.
Then $\triangle AYM \equiv \triangle CYN$ (by ASA-congruence, since

$\angle YAM = \angle YCN$ (alternate angles, since $AD \parallel BC$)
$\underline{AM} = \underline{CN}$
$\angle AMY = \angle CNY$ (alternate angles, since $AD \parallel BC$).

Hence $\underline{AY} = \underline{CY}$, so Y is the midpoint of $\underline{AC}$ – the centre of the parallelogram (where the two diagonals meet (by Problem **157**(iii))). QED

161.

Note: In the easy case, where the perpendicular from A to the line DC meets the side $\underline{DC}$ internally at X, it is natural to see the parallelogram $ABCD$ as the "sum" of a trapezium $ABCX$ and a right angled triangle $\triangle AXD$. If the perpendicular from B to DC meets DC at Y, then $\triangle AXD \equiv \triangle BYC$. Hence we can rearrange the two parts of the parallelogram $DCBA$ to form a rectangle $XYBA$.

However, a general proof cannot assume that the perpendicular from A (or from B) to DC meets $\underline{DC}$ internally. Hence we are obliged to think of the parallelogram in terms of **differences**. This is a strategy that is often useful, but which can be surprisingly elusive.

Draw the perpendiculars from A and B to the line CD, and from C and D to the line AB. Choose the two perpendiculars which, together with the lines AB and CD define a rectangle that completely contains the parallelogram $ABCD$ (that is, if AB runs from left to right, take the left-most, and the right-most perpendiculars). These will be either the perpendiculars from B and from D, or the perpendiculars from A and from C (depending on which way the sides AC and BD slope).

Suppose the chosen perpendiculars are the one from B – meeting the line DC at P, and the one from D – meeting the line AB at Q.

Then $BP \parallel QD$ (by Problem **159**(c)), so $BQDP$ is a parallelogram with a right angle – and hence a rectangle. Hence $\underline{BQ} = \underline{PD}$, and $\underline{BP} = \underline{QD}$ (by Problem **157**(ii)).

$\triangle QAD \equiv \triangle PCB$ (by RHS-congruence), so each is equal to half the rectangle on base $\underline{PC}$ and height $\underline{PB}$. Hence

$$\begin{aligned} \text{area}(ABCD) &= \text{area(rectangle } BQDP) \\ &\qquad - \text{area(rectangle on base } \underline{PC} \text{ with height } \underline{PB}) \\ &= \text{area(rectangle on base } \underline{CD} \text{ with height } \underline{DQ}). \end{aligned}$$

QED

162.

(a) $ABCB'$ is a parallelogram, so $AB' \parallel CB$ and $\underline{AB'} = \underline{BC}$.
Similarly, $BCAC'$ is a parallelogram, so $AC' \parallel CB$ and $\underline{AC'} = \underline{CB}$.
Hence $\underline{B'A} = \underline{AC'}$, so A is the midpoint of $\underline{B'C'}$.

Similarly B is the midpoint of $\underline{C'A'}$, and C is the midpoint of $\underline{A'B'}$.

(b) Let H be the circumcentre of $\triangle A'B'C'$ – that is, the common point of the perpendicular bisectors of $\underline{A'B'}$, $\underline{B'C'}$, and $\underline{C'A'}$. Then H is a common point of the three perpendiculars from A to BC, from B to CA, and from C to AB.

163. Consider any path from H to V. Suppose this reaches the river at X. The shortest route from H to X is a straight line segment; and the shortest route from X to V is a straight line segment.

If we reflect the point H in the line of the river, we get a point H', where HH' is perpendicular to the river and meets the river at Y (say).

Then $\triangle HXY \equiv \triangle H'XY$ (by SAS-congruence, since $\underline{HY} = \underline{H'Y}$, $\angle HYX = \angle H'YX$, and $\underline{YX} = \underline{YX}$). Hence $\underline{HX} = \underline{H'X}$, so the distance from H to V via X is equal to $\underline{HX} + \underline{XV} = \underline{H'X} + \underline{XV}$, and this is shortest when H', X, and V are collinear. (So to find the shortest route, reflect H in the line of the river to H', then draw $H'V$ to cross the line of the river at X, and travel from H to V via X.)

164.

(a) Let $\triangle PQR$ be any triangle inscribed in $\triangle ABC$, with P on BC, Q on CA, R on AB (not necessarily the orthic triangle). Let P' be the reflection of P in the side AC, and let P'' be the reflection of P in the side AB. Then $\underline{PQ} = \underline{P'Q}$, and $\underline{PR} = \underline{P''R}$ (as in Problem **163**).
$\therefore \underline{PQ} + \underline{QR} + \underline{RP} = \underline{P'Q} + \underline{QR} + \underline{RP''}$.

Each choice of the point P on $\underline{AB}$ determines the positions of P' and P''. Hence the shortest possible perimeter of $\triangle PQR$ arises when $P'QRP''$ is a straight line. That is, given a choice of the point P, choose Q and R by:

- constructing the reflections P' of P in AC, and P'' of P in AB;
- join $\underline{P'P''}$ and let Q, R be the points where this line segment crosses AC, AB respectively.

It remains to decide how to choose P on $\underline{BC}$ so that $\underline{P'P''}$ is as short as possible. The key here is to notice that A lies on both AC and on AB.
$\therefore \underline{AP} = \underline{AP'}$, and $\underline{AP} = \underline{AP''}$, so $\triangle AP'P''$ is isosceles.
Also $\angle PAC = \angle P'AC$, and $\angle PAB = \angle P''AB$.
$\therefore \angle P'AP'' = 2 \cdot \angle A$.

Hence, for each position of the point P, $\triangle AP'P''$ is isosceles with apex angle equal to $2 \cdot \angle A$. Any two such triangles are similar (by SAS-similarity).

Hence the triangle $\triangle AP'P''$ with the shortest "base" $\underline{P'P''}$ occurs when the legs $\underline{AP'}$ and $\underline{AP''}$ are as short as possible. But $\underline{AP} = \underline{AP'} = \underline{AP''}$, so this occurs when $\underline{AP}$ is as short as possible – namely when $\underline{AP}$ is perpendicular to BC.

Since the same reasoning applies to Q and to R, it follows that the required triangle $\triangle PQR$ must be the orthic triangle of $\triangle ABC$. QED

(b) Let $\triangle PQR$ be the orthic triangle of $\triangle ABC$, with P on BC, Q on CA, R on AB. Let H be the orthocenter of $\triangle ABC$.

$\angle BPH$ and $\angle BRH$ are both right angles, so add to a straight angle. Hence (by Problem **154**(c)), $BPHR$ is a cyclic quadrilateral. Similarly $CPHQ$ and $AQHR$ are cyclic quadrilaterals.
In the circumcircle of $CPHQ$, we see that the initial "angle of incidence" $\angle CQP = \angle CHP$. Also $\angle CHP = \angle AHR$ (vertically opposite angles); and in the circumcircle of $AQHR$, $\angle AHR = \angle AQR$.
Hence $\angle CQP = \angle AQR$, so a ray of light which traverses $\underline{PQ}$ will reflect at Q along the line $\underline{QR}$. Similarly one can show that $\angle ARQ = \angle BRP$, so that the ray will then reflect at R along $\underline{RP}$; and $\angle BPR = \angle CPQ$, so the ray will then reflect at P along $\underline{PQ}$. QED

165.

(a)(i) Triangles $\triangle ABL$ and $\triangle ACL$ have equal bases $\underline{BL} = \underline{CL}$, and the same apex A – so lie between the same parallels. Hence they have equal area (by Problems **157** and **161**).

Similarly, $\triangle GBL$ and $\triangle GCL$ have equal bases $\underline{BL} = \underline{CL}$, and the same apex G – so have equal area.

Hence the differences $\triangle ABG = \triangle ABL - \triangle GBL$ and $\triangle ACG = \triangle ACL - \triangle GCL$ have equal area.

(ii) [Repeat the solution for part (i) replacing A, B, C, L, G by B, C, A, M, G.]

(b) $\triangle ABL$ and $\triangle ACL$ have equal area (as in (a)(i)). Similarly $\triangle GBL$ and $\triangle GCL$ have the same area – say x (since $\underline{BL} = \underline{CL}$). Hence $\triangle ABG$ and $\triangle ACG$ have equal area.

In the same way $\triangle BCM$ and $\triangle BAM$ have equal area; and $\triangle GCM$ and $\triangle GAM$ have the same area – say y (since $CM = AM$). Hence $\triangle BCG$ and $\triangle BAG$ have equal area.

But then $\triangle ABG = \triangle ACG = \triangle BCG$ and $\triangle ACG = \triangle AMG + \triangle CMG = 2y$, $\triangle BCG = \triangle BLG + \triangle CLG = 2x$. Hence $x = y$, so $\triangle AMG$, $\triangle CMG$, $\triangle CLG$, $\triangle BLG$ all have the same area x, and $\triangle ABG$ has area $2x$.

The segment $\underline{GN}$ divides $\triangle ABG$ into two equal parts ($\triangle ANG$ and $\triangle BNG$), so each part has area x.
Hence $\triangle CAG + \triangle ANG$ has the same area ($3x$) as $\triangle CAN$. Hence $\angle CGN$ is a straight angle, and the three medians AL, BM, CN all pass through the point G.

Note: At first sight, the 'proof' of the result in (b) using vectors seems considerably easier. (If A, B, C have position vectors $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ respectively, then L has the position vector $\frac{1}{2}(\mathbf{b}+\mathbf{c})$, and M has position vector $\frac{1}{2}(\mathbf{c}+\mathbf{a})$, and it is easy to see that AL and BM meet at G with position vector $\frac{1}{3}(\mathbf{a}+\mathbf{b}+\mathbf{c})$. One can then check directly that G lies on CN, or notice that the symmetry of the expression $\frac{1}{3}(\mathbf{a}+\mathbf{b}+\mathbf{c})$ guarantees that G is also the point where BM and CN meet.

The inscrutable aspect of this 'proof' lies in the fact that all the geometry has been silently hidden in the algebraic assumptions which underpin the unstated axioms of the 2-dimensional vector space, and the underlying field of real numbers. Hence, although the vector 'proof' may seem simpler, the two different approaches cannot really be compared.

166.

(a) $\underline{AN} = \underline{BN}$ (by construction of N as the midpoint of $\underline{AB}$)
$\angle ANM = \angle BNM'$ (vertically opposite angles)
$\underline{NM} = \underline{NM'}$ (by construction).
$\therefore \triangle ANM \equiv \triangle BNM'$ (by SAS-congruence). QED

(b) $\underline{BM'} = \underline{AM} = \underline{CM}$.
$\angle NAM = \angle NBM'$ (since $\triangle ANM \equiv \triangle BNM'$)
$\therefore BM' \parallel MA$ (i.e. $BM' \parallel CM$). QED

(c) $BM'MC$ is a quadrilateral with opposite sides $\underline{CM}$, $\underline{BM'}$ equal and parallel.
$\therefore BM'MC$ is a (by Problem **158**(a)). QED

167. Since A and B are interchangeable in the result to be proved, we may assume that A is the point on the secant that lies between P and B.

In order to make deductions, we have to create triangles – so join AT and BT. This creates two triangles: $\triangle PAT$ and $\triangle PTB$, in which we see that:
$\angle TPA = \angle BPT$,
$\angle PTA = \angle PBT$ (by Problem **153**),
$\therefore \angle PAT = \angle PTB$ (since the three angles in each triangle add to a straight angle).
Hence $\triangle PAT \sim \triangle PTB$ (by AAA-similarity).
$\therefore \underline{PT} : \underline{PB} = \underline{PA} : \underline{PT}$, or $\underline{PA} \times \underline{PB} = \underline{PT}^2$. QED

168. Since A, B are interchangeable in the result to be proved, and C, D are interchangeable, we may assume that A lies between P and B, and that C lies between P and D.

(i) Let the tangent from P to the circle touch the circle at T.
Then Problem **167** guarantees that $\underline{PA} \times \underline{PB} = \underline{PT}^2$.
Replacing the secant PAB by PCD shows similarly that $\underline{PC} \times \underline{PD} = \underline{PT}^2$.
Hence $\underline{PA} \times \underline{PB} = \underline{PC} \times \underline{PD}$. QED

(ii) **Note:** The first proof is so easy, one may wonder why anyone would ask for a second proof. The reason lies in Problem **169** – which looks very much like Problem **168**, but with P *inside* the circle. Hence, when we come to the next problem, the easy approach in (i) will not be available, so it is worth looking for another proof of **168** which has a chance of generalizing.

Notice that $\underline{AC}$ is a chord which links the two secants PAB and PCD.
So join $\underline{AD}$ and $\underline{CB}$.

Then $\triangle PAD \sim \triangle PCB$ (by AAA-similarity: since $\angle APD = \angle CPB$, and $\angle PDA = \angle PBC$).
$\therefore \underline{PA} : \underline{PC} = \underline{PD} : \underline{PB}$, or $\underline{PA} \times \underline{PB} = \underline{PC} \times \underline{PD}$. QED

169. Join $\underline{AD}$ and $\underline{CB}$.
Then $\triangle PAD \sim \triangle PCB$ (by AAA-similarity: since
$\angle APD = \angle CPB$ (vertically opposite angles)
$\angle PDA = \angle PBC$ (angles subtended by chord $\underline{AC}$ on the same arc)
$\angle PAD = \angle PCB$ (angles subtended by a chord $\underline{BD}$ on the same arc)).
$\therefore \underline{PA} : \underline{PC} = \underline{PD} : \underline{PB}$, or $\underline{PA} \times \underline{PB} = \underline{PC} \times \underline{PD}$. QED

170.

(a) If $a = b$, then $ABCD$ is a parallelogram (by Problem **159**(a)).
$\therefore \underline{AD} = \underline{BC}$ (by Problem **157**(ii)).
$\therefore \underline{AM} = \underline{BN}$, so $ABNM$ is a parallelogram (by Problem **159**(a)).
$\therefore \underline{MN} = \underline{AB}$ has length a, and $MN \parallel AB$ (by Problem **160**).

If $a \neq b$, then $a < b$, or $b < a$. We may assume that $a < b$.
Construct the line through B parallel to AD, and let this line meet $\underline{DC}$ at Q.
Then $ABQD$ is a parallelogram, so $\underline{DQ} = \underline{AB}$, and $\underline{AD} = \underline{BQ}$. Hence $\underline{QC}$ has length $b - a$.
Construct the line through M parallel to QC (and hence parallel to BA), and let this meet $\underline{BQ}$ at P, and $\underline{BC}$ at N'.
Then $ABPM$ and $MPQD$ are both parallelograms.
$\therefore \underline{MP} = \underline{AB}$ has length a, and

$$\underline{BP} = \underline{AM} = \underline{MD} = \underline{PQ}.$$

Now $\triangle BPN' \sim \triangle BQC$ (by AAA-similarity, since $PN' \parallel QC$); so

$$\underline{BP} : \underline{BQ} = \underline{BN'} : \underline{BC} = 1 : 2.$$

Hence $N' = N$ is the midpoint of $\underline{BC}$, $MN \parallel BC$, and $\underline{MN}$ has length

$$a + \frac{b-a}{2} = \frac{a+b}{2}.$$

(b) Suppose first that $a = b$. Then $ABCD$ is a parallelogram (by Problem **159**(a)), so the area of $ABCD$ is given by $a \times d$ ("(length of base) $\times$ height"). (by Problem **161**). Hence we may suppose that $a < b$.

Solution 1: Extend $\underline{AB}$ beyond B to a point X such that $\underline{BX} = \underline{DC}$ (so $\underline{AX}$ has length $a + b$).
Extend $\underline{DC}$ beyond C to a point Y such that $\underline{CY} = \underline{AB}$ (so $\underline{DY}$ has length $a + b$).
Clearly $ABCD$ and $YCBX$ are congruent, so each has area one half of area($AXYD$).

Now $AX \parallel DY$, and $\underline{AX} = \underline{DY}$, so $AXYD$ is a parallelogram (by Problem **159**(a)).
Hence $AXYD$ has area "(length of base) $\times$ height" (by Problem **161**), so $ABCD$ has area $\frac{a+b}{2} \times d$.

Solution 2: [We give a second solution as preparation for Problem **171**.]
Now $a < b$ implies that $\angle BAD + \angle ABC$ is greater than a straight angle.

[**Proof.** The line through B parallel to AD meets DC at Q, and $ABQD$ is a parallelogram.
Hence $\underline{DQ} = \underline{AB}$, so Q lies between D and C, and

$$\angle ABC = \angle ABQ + \angle QBC > \angle ABQ.$$

$\therefore$ $\angle BAD + \angle ABC > \angle BAD + \angle ABQ$, which is equal to a straight angle.]

So if we extend $\underline{DA}$ beyond A, and $\underline{CB}$ beyond B, the lines meet at X, where X, D are on opposite sides of AB. Then $\triangle XAB \sim \triangle XDC$ (by AAA-similarity), whence corresponding lengths in the two triangles are in the ratio $\underline{AB} : \underline{DC} = a : b$. In particular, if the perpendicular from X to AB has length h, then $\frac{h}{h+d} = \frac{a}{b}$, so $h(b-a) = ad$.
Now

$$\begin{aligned}
\text{area}(ABCD) &= \text{area}(\triangle XDC) - \text{area}(\triangle XAB)\\
&= \frac{1}{2}b(h+d) - \frac{1}{2}ah\\
&= \frac{1}{2}bd + \frac{1}{2}h(b-a)\\
&= \frac{1}{2}bd + \frac{1}{2}ad\\
&= \left(\frac{a+b}{2}\right)d.
\end{aligned}$$

171. Note: The volume of a pyramid or cone is equal to

$$\frac{1}{3} \times \text{ (area of base) } \times \text{ height.}$$

There is no elementary general proof of this fact. The initial coefficient of $\frac{1}{3}$ arises because we are "adding up", or integrating, cross-sections parallel to the base, whose area involves a square x^2, where x is the distance from the apex (just as the coefficient $\frac{1}{2}$ in the formula for the area of a triangle arises because we are integrating linear cross-sections whose size is a multiple of x – the distance from the apex). Special cases of this formula can be checked – for example, by noticing that a cube $ABCDEFGH$ of side s, with base $ABCD$, and upper surface $EFGH$, with E above D, F above A, and so on, can be dissected into three identical pyramids

– all with apex E: one with base $ABCD$, one with base $BCHG$, and one with base $ABGF$. Hence each pyramid has volume $\frac{1}{3}s^3$, which may be interpreted as

$$\frac{1}{3} \times (\text{area of base} = s^2) \times (\text{height} = s).$$

To obtain the frustum of height d, a pyramid with height h (say) is cut off a pyramid with height $h + d$.

$$\begin{aligned}\therefore \text{volume(frustum)} &= \left[\frac{1}{3} \times b^2 \times (h+d)\right] - \left[\frac{1}{3} \times a^2 \times h\right] \\ &= \left[\frac{1}{3} \times b^2 \times d\right] + \left[\frac{1}{3} \times (b^2 - a^2) \times h\right].\end{aligned}$$

Let N be the midpoint of $\underline{BC}$, and let the line AN meet the upper square face of the frustum at M.
Let the perpendicular from the apex A to the base $BCDE$ meet the upper face of the frustum at Y and the base $BCDE$ at Z.
Then $\triangle AYM \sim \triangle AZN$ (by AAA-similarity), so $\underline{AY} : \underline{AZ} = \underline{YM} : \underline{ZN}$.
$\therefore \frac{h}{h+d} = \frac{a}{b}$, so $h(b-a) = ad$.

$$\begin{aligned}\therefore \text{volume(frustum)} &= \frac{1}{3}b^2d + \frac{1}{3}\left(b^2 - a^2\right)h \\ &= \frac{1}{3}b^2d + \frac{1}{3}(b+a)ad \\ &= \frac{1}{3}\left(b^2 + ab + a^2\right)d.\end{aligned}$$

172. Construct the line through A which is parallel to $A'B'$, and let it meet the line BB' at P.
Similarly, construct the line through B which is parallel to $B'C'$ and let it meet the line CC' at Q.
Then $\triangle ABP \sim \triangle BCQ$ (by AAA-similarity), so $\underline{AB} : \underline{BC} = \underline{AP} : \underline{BQ}$.
Now $AA'B'P$ is a parallelogram, so $\underline{AP} = \underline{A'B'}$, and $BB'C'Q$ is a parallelogram, so $\underline{BQ} = \underline{B'C'}$.
$\therefore \underline{AB} : \underline{BC} = \underline{A'B'} : \underline{B'C'}$. QED

173.

(a) Suppose $\underline{AB}$ has length a and $\underline{CD}$ has length b. Problem **137** allows us to construct a point X such that $\underline{AX} = \underline{CD}$, so $\underline{AX}$ has length b. Now construct the point Y where the circle with centre A and radius $\underline{AX}$ meets $\underline{BA}$ produced (beyond A). Then $\underline{YB}$ has length $a + b$.

If $a > b$, let Z be the point where the circle with centre A and passing through X meets the segment $\underline{AB}$ internally Then $\underline{ZB}$ has length $a - b$.

(b) Use Problem **137** to construct a point U (not on the line CD) such that $\underline{DU} = \underline{XY}$, and a point V on DU (possibly extended beyond U) such that $\underline{DV} = \underline{AB}$. Hence $\underline{DU}$ has length 1 and $\underline{DV}$ has length a.

Construct the line through V parallel to UC and let it meet the line DC at W. Then $\triangle DVW \sim \triangle DUC$, with scale factor $\underline{DW} : \underline{DC} = \underline{DV} : \underline{DU} = a : 1$. Hence $\underline{DW}$ has length ab.

To construct a line segment of length $\frac{a}{b}$ construct U, V as above. Then let the circle with centre D and radius $\underline{DU}$ meet CD at U'. Let the line through U' parallel to CV meet DV at X. The $\triangle DU'X \sim \triangle DCV$, with scale factor $\underline{DX} : \underline{DV} = \underline{DU'} : \underline{DC} = 1 : b$, so $\underline{DX}$ has length $\frac{a}{b}$.

(c) Use Problem **137** to construct a point G so that $\underline{AG} = \underline{XY} = 1$. Then draw the circle with centre A passing through G to meet BA extended at H. Hence $\underline{HA} = 1$, $\underline{AB} = a$.

Construct the midpoint M of $\underline{HB}$; draw the circle with centre M and passing through H and B.

Construct the perpendicular to $\underline{HB}$ at the point A to meet the circle at K.

Then $\triangle HAK \sim \triangle KAB$, so $\underline{AB} : \underline{AK} = \underline{AK} : \underline{AH}$. Hence $\underline{AK} = \sqrt{a}$.

174. The key to this problem is to use Problem **158**: a parallelogram (and hence any rectangle) is divided into two congruent pieces by any straight cut through the centre. If A is the centre of the rectangular piece of fruitcake, and B is the centre of the combined rectangle consisting of "fruitcake plus icing", then the line AB gives a straight cut that divides both the fruitcake and the icing exactly in two.

175.

(a) The minute hand is pointing exactly at "7", but the hour hand has moved $\frac{7}{12}$ of the way from "1" to "2": that is, $\frac{7}{12}$ of 30°, or $17\frac{1}{2}^\circ$. Hence the angle between the hands is $162\frac{1}{2}^\circ$.

The same angle arises whenever the hands are trying to point in opposite directions, but are off-set by $\frac{7}{12}$ of 30°, or $17\frac{1}{2}^\circ$. This suggests that instead of "35 minutes after 1" we should consider "35 minutes before 11", or 10:25.

(b) The two hands coincide at midnight. The minute hand then races ahead, and the hands do not coincide again until shortly after 1:00 – and indeed after 1:05.

More precisely, in 60 minutes, the minute hand turns through 360°, so in x minutes, it turns through $6x^\circ$. In 60 minutes the hour hand turns through 30°, so in x minutes, the hour hand turns through $\frac{1}{2}x^\circ$.

The hands overlap when

$$\frac{1}{2}x \equiv 6x \pmod{360},$$

or when $\frac{11}{2}x$ is a multiple of 360.

This occurs when $x = 0$ (i.e. at midnight); then not until

$$x = \frac{720}{11} = 65\frac{5}{11},$$

and the time is $5\frac{5}{11}$ minutes past 1: that is, after $1\frac{1}{11}$ hours.

It occurs again $1\frac{1}{11}$ hours, or $65\frac{5}{11}$ minutes, later – namely at $10\frac{10}{11}$ minutes past 2, and so on.

Hence the phenomenon occurs at midnight, and then 11 more times until noon (with noon as the 12th time; and then 11 more times until midnight – and hence 23 times in all (including both midnight occurrences).

(c) If we add a third hand (the 'second hand'), all three hands coincide at midnight. In x minutes, the second hand turns through $360x^\circ$.

We now know exactly when the hour hand and minute hand coincide, so we can check where the second hand is at these times. For example, at $5\frac{5}{11}$ minutes past 1, the second hand has turned through $(360 \times 65\frac{5}{11})^\circ$, and

$$360 \times 65\frac{5}{11} \equiv 360 \times \frac{5}{11} \pmod{360},$$

so the second hand is nowhere near the other two hands.

The k^{th} occasion when the hour hand and minute hands coincide occurs at $k \times 1\frac{1}{11}$ hours after midnight, when the two hands point in a direction $\left(\frac{360k}{11}\right)^\circ$ clockwise beyond "12" . At the same time, the second hand has turned through $(360 \times 65\frac{5}{11} \times k)^\circ$, and

$$360 \times 65\frac{5}{11} \times k \equiv 360 \times \frac{5}{11} \times k \pmod{360},$$

or five times as far round, and these two are never equal $\pmod{360}$.

176.

Note: One of the things that makes it possible to calculate distances exactly here is that the angles are all known exactly, and give rise to lots of right angled triangles.

The rotational symmetry of the clockface means we only have to consider segments with one endpoint at 12 o'clock (say A). The reflectional symmetry in the line AG means that we only have to find $\underline{AB}$, $\underline{AC}$, $\underline{AD}$, $\underline{AE}$, $\underline{AF}$, and $\underline{AG}$.

Clearly $\underline{AG} = 2$. If O denotes the centre of the clockface, then $\underline{AD}$ is the hypotenuse of an isosceles triangle $\triangle OAD$ with legs of length 1, so $\underline{AD} = \sqrt{2}$.

$\triangle ACO$ is isosceles ($\underline{OA} = \underline{OC} = 1$) with apex angle $\angle AOC = 60^\circ$, so the triangle is equilateral. Hence $\underline{AC} = \underline{OA} = 1$.

It follows that $ACEGIK$ is a regular hexagon, so $\underline{AE} = \sqrt{3}$ (if OC meets AE at X, then $\triangle ACX$ is a 30-60-90 triangle, and so is half of an equilateral triangle, whence $\underline{AX} = \frac{\sqrt{3}}{2}$).

It remains to find $\underline{AB}$ and $\underline{AF}$.

Let OB meet AC at Y. Then $\underline{OY} = \frac{\sqrt{3}}{2}$ (since $\triangle OYA$ is a 30-60-90 triangle).

$\therefore\ \underline{BY} = 1 - \frac{\sqrt{3}}{2}$, $\underline{AY} = \frac{1}{2}$, so $\underline{AB} = \sqrt{2 - \sqrt{3}}$.

Finally, $\underline{AB}$ subtends $\angle AOB = 30^\circ$ at the centre, whence $\angle OAB = \angle OBA = 75^\circ$ and $\angle AGB = 15^\circ$.

$\therefore\ \angle ABG = 90^\circ$ so we may apply Pythagoras' Theorem to $\triangle ABG$ to find $\underline{BG} = \underline{AF} = \sqrt{2 + \sqrt{3}}$.

177.

$\sqrt{1}$ is the distance from $(0,0,0)$ to $(1,0,0)$.
$\sqrt{2}$ is the distance from $(0,0,0)$ to $(1,1,0)$.
$\sqrt{3}$ is the distance from $(0,0,0)$ to $(1,1,1)$.
$\sqrt{4}$ is the distance from $(0,0,0)$ to $(2,0,0)$.
$\sqrt{5}$ is the distance from $(0,0,0)$ to $(2,1,0)$.
$\sqrt{6}$ is the distance from $(0,0,0)$ to $(2,1,1)$.
$\sqrt{7}$ cannot be realized as a distance between integer lattice points in 3D.
$\sqrt{8}$ is the distance from $(0,0,0)$ to $(2,2,0)$.
$\sqrt{9}$ is the distance from $(0,0,0)$ to $(3,0,0)$.
$\sqrt{10}$ is the distance from $(0,0,0)$ to $(3,1,0)$.
$\sqrt{11}$ is the distance from $(0,0,0)$ to $(3,1,1)$.
$\sqrt{12}$ is the distance from $(0,0,0)$ to $(2,2,2)$.
$\sqrt{13}$ is the distance from $(0,0,0)$ to $(3,2,0)$.
$\sqrt{14}$ is the distance from $(0,0,0)$ to $(3,2,1)$.
$\sqrt{15}$ cannot be realized as a distance between integer lattice points in 3D.
$\sqrt{16}$ is the distance from $(0,0,0)$ to $(4,0,0)$.
$\sqrt{17}$ is the distance from $(0,0,0)$ to $(4,1,0)$.

Note: The underlying question extends Problem **32**:

Which integers can be represented as a sum of three squares?

This question was answered by Legendre (1752–1833):

> **Theorem**. A positive integer can be represented as a sum of three squares if and only if it is not of the form $4^a(8b+7)$.

178.

(a) Let AD and CE cross at X. Join $\underline{DE}$. Then $\angle AXC = \angle EXD$ (vertically opposite angles). Hence $\angle CAD + \angle ACE = \angle ADE + \angle CED$, so the five angles of the pentagonal star have the same sum as the angles of $\triangle BED$. Hence the five angles have sum π radians.

(b) Start with seven vertices A, B, C, D, E, F, G arranged in cyclic order.

(i) Consider first the 7-gonal star $ADGCFBE$. Let GD and BE cross at X and let GC and BF cross at Y. Join $\underline{DE}$, $\underline{BG}$. As in part (a),

$$\angle BGC + \angle GBF = \angle BFC + \angle GCF,$$

and

$$\angle BGD + \angle GBE = \angle BED + \angle GDE,$$

so the angles in the 7-gonal star have the same sum as the angles in $\triangle ADE$. Hence the seven angles have sum π radians.

(ii) Similar considerations with the 7-gonal star $ACEGBDF$ show that its seven angles have sum 3π radians.

Note: Notice that the three possible "stars" (including the polygon $ABCDEFG$) have angle sums π radians, 3π radians, and 5π radians.

(c) if $n = 2k + 1$ is prime, we may join A to its immediate neighbour B (1-step), or to its second neighbor C (2-step), ..., or to its k^{th} neighbour (k-step), so there are k different stars, with angle sums

$$(n-2)\pi, (n-4)\pi, (n-6)\pi, \ldots, 3\pi, \pi$$

respectively.

If n is not prime, the situation is slightly more complicated, since, for each divisor m of n, the km-step stars break up into separate components.

179.

(a) (i) Let the other three pentagons be $CDRST$, $DEUVW$, $EAXYZ$.
At A we have three angles of $108°$, so $\angle NAX = 36°$.
$\triangle ANX$ is isosceles ($\underline{AN} = \underline{AB} = \underline{AE} = \underline{AX}$), so $\angle AXN = 72°$.
Hence $\angle AXY + \angle AXN = 180°$, so Y, X, N lie in a straight line.
Similarly M, N, X lie in a straight line.
Hence M, N, X, Y lie on a straight line segment $\underline{MY}$ of length $1 + \underline{YN} = 2 + \underline{NX}$.

(ii) In the same way it follows that MP passes through L and Q, that PS passes through O and T, etc. so that the figure fits snugly inside the pentagon $MPSVY$, whose angles are all equal to $108°$. Moreover, $\triangle ANX \equiv \triangle BQL$ (by SAS-congruence), so $\underline{XN} = \underline{LQ}$, whence $\underline{MP} = \underline{YM}$. Similarly $\underline{MP} = \underline{PS} = \underline{SV} = \underline{VY}$, so $MPSVY$ is a regular pentagon.

(iii) In the regular pentagon $EAXYZ$ the diagonal $EY \parallel AX$.
Moreover XAC is a straight line, and $\angle ACB + \angle NBC = 180°$, so $AC \parallel NB$.
Hence $YE \parallel NB$, and $\underline{YE} = \underline{NB} = \tau$ (the Golden Ratio $\frac{1+\sqrt{5}}{2}$), so $YEBN$ is a parallelogram.
Hence $\underline{YN} = \underline{EB} = \tau$, so $\underline{YM} = 1 + \tau = \tau^2$.

(b) (i)

$$\triangle MPY \equiv \triangle PSM \equiv \triangle SVP \equiv \triangle VYS \equiv \triangle YMV$$

(by SAS-congruence), so

$$\underline{YP} = \underline{MS} = \underline{PV} = \underline{SY} = \underline{VM}$$

Also $\angle PMS = \angle MPY = 36^\circ$;
$\therefore \triangle BMP$ has equal base angles, and so is isosceles.
Hence $\underline{BM} = \underline{BP}$, and $\angle MBP = 108^\circ = \angle ABC$.
Similarly $\angle AMY = \angle AYM = 36^\circ$.
$\therefore \triangle AMY \equiv \triangle BPM$ (by ASA-congruence), so

$$\underline{AY} = \underline{AM} = \underline{BM} = \underline{BP}.$$

$\triangle MAB \equiv \triangle PBC$ – by ASA-congruence:

$$\begin{aligned} \angle BPC &= 108^\circ - \angle BPM - \angle CPS = 36^\circ = \angle AMB, \\ \underline{MA} &= \underline{PB}, \text{ and} \\ \angle PBC &= \angle MBA = \angle MAB. \end{aligned}$$

Hence $\underline{AB} = \underline{BC}$.

Continuing round the figure we see that

$$\underline{AB} = \underline{BC} = \underline{CD} = \underline{DE} = \underline{EA}$$

and that

$$\angle A = \angle B = \angle C = \angle D = \angle E.$$

(ii) Extend DB to meet MP at L, and extend DA to meet MY at N.
Then $36^\circ = \angle DBC = \angle LBM$ (vertically opposite angles). Hence $\triangle LBM$ has equal base angles and so is isosceles: $\underline{LM} = \underline{LB}$. Similarly $\underline{NM} = \underline{NA}$.
Now $\triangle LBM \equiv \triangle NAM$ (by ASA-congruence, since $\underline{MA} = \underline{MB}$), so $\underline{LM} = \underline{LB} = \underline{NA} = \underline{NM}$.
In the regular pentagon $ABCDE$ we know that $\angle DBC = 36^\circ$; and in $\triangle LBM$, $\angle LBM = \angle DBC$ (vertically opposite angles), so $\angle MLB = 108^\circ$. Hence $\angle BLP = 72^\circ = \angle LBP$, so $\triangle PLB$ is isosceles: $\underline{PL} = \underline{PB}$.
In the regular pentagon $MPSVY$, $\triangle PMA$ is isosceles, so $\underline{PM} = \underline{PA}$.
$\therefore \underline{LM} = \underline{PM} - \underline{PL} = \underline{PA} - \underline{PB} = \underline{BA}$.
Hence $ABLMN$ is a pentagon with five equal sides. It is easy to check that the five angles are all equal.

(iii) We saw in (i) that the five diagonals of $MPSVY$ are equal. We showed in Problem **3** that each has length τ, and that $MPDY$ is a rhombus, so $\underline{DY} = \underline{PM} = 1$. Similarly $\underline{SE} = 1$.
Hence $\underline{SD} = \tau - 1$, and $\underline{DE} = \underline{SE} - \underline{SD} = 2 - \tau = (\tau^2)^{-1}$.

180.

(a) Since the tiles fit together "edge-to-edge", all tiles have the same edge length. The number k of tiles meeting at each vertex must be at least 3 (since the angle at each vertex of a regular n-gon $= \left(1 - \frac{2}{n}\right)\pi < \pi$), and can be at most 6 (since the smallest possible angle in a regular n-gon occurs when $n = 3$, and is then $\frac{\pi}{3}$).
We consider each possible value of k in turn.

- If $k = 6$, then $n = 3$ and we have six equilateral triangles at each vertex.
- If $k = 5$, then we would have $\left(1 - \frac{2}{n}\right)\pi = \frac{2\pi}{5}$, so $n = \frac{10}{3}$ is not an integer.
- If $k = 4$, then $n = 4$ and we have four squares at each vertex.
- If $k = 3$, then $n = 6$ and we have three regular hexagons at each vertex.

Hence $n = 3$, or 4, or 6.

(b) (i) If $k = n = 4$, it is easy to form the vertex figure. It may seem 'obvious' that it continues "to infinity"; but if we think it is obvious, then we should explain why: (choose the scale so that the common edge length is "1"; then let the integer lattice points be the vertices of the tiling, with the tiles as translations of the unit square formed by $(0, 0)$, $(1, 0)$, $(1, 1)$, $(0, 1)$).

(ii) If $k = 6$, $n = 3$, let the vertices correspond to the complex numbers $p + q\omega$, where p, q are integers, and where ω is a complex cube root of 1 (that is, a solution of the equation $\omega^3 = 1 \neq \omega$, or $\omega^2 + \omega + 1 = 0$), with the edges being the line segments of length 1 joining nearest neighbours (at distance 1).

(iii) If $k = 3$, $n = 6$, take the same vertices as in (ii), but eliminate all those for which $p + q \equiv 0 \pmod 3$, then let the edges be the line segments of length 1 joining nearest neighbours.

181.

(a) (i) As before, the number k of tiles at each vertex lies between 3 and 6. However, this time k does not determine the shape of the tiles. Hence we introduce a new parameter: the number of t of triangles at each vertex, which can range from 0 up to 6. The derivations are based on elementary arithmetic, for which it is easier to work with angles in degrees.

* If $t = 6$, then the vertex figure must be $\mathbf{3^6}$.
* If $t = 5$, the remaining gap of 60° could only take a sixth triangle, so this case cannot occur.
* If $t = 4$, we are left with angle of 120°, so the only possible vertex figure is $\mathbf{3^4.6}$.
* If $t = 3$, then we are left with an angle of 180°, so the only possible configurations are $\mathbf{3^3.4^2}$ (with the two squares together), or $\mathbf{3^2.4.3.4}$ (with the two squares separated by a triangle).
* If $t = 2$, we are left with an angle of 240°, which cannot be filled with 3 or more tiles (since the average angle size would then be at most 80°, and no more triangles are allowed), so the only possible vertex figures are $\mathbf{3^2.4.12}$, $\mathbf{3.4.3.12}$, $\mathbf{3^2.6^2}$, $\mathbf{3.6.3.6}$ (since $\mathbf{3^2.5.}\boldsymbol{n}$ or $\mathbf{3.5.3.}\boldsymbol{n}$ would require a regular n-gon with an angle of 132°, which is impossible).

Note: The compactness of the argument based on the parameter t is about to end. We continue the same approach, with the focus shifting from the parameter t to a new parameter s – namely the number of squares in each vertex figure.

* Suppose $t = 1$. We are left with an angle of 300°.
  If $s = 0$, the 300° cannot be filled with 3 or more tiles (since then the average angle size would be at most 100°, and no squares can be used), so there are exactly two additional tiles. Since each tile has angle $\leqslant 180^\circ$, we cannot use a hexagon, so the smallest tile has at least 7 sides; and since the average of the two remaining angle sizes is 150°, the smallest tile has at most 12 sides. It is now easy to check that the only possible vertex figures are **3.7.42**, **3.8.24**, **3.9.20**, **3.10.15**, $\mathbf{3.12^2}$.
  If $s = 1$, we would be left with an angle of 210°, which would require two larger tiles with average angle size 105°, which is impossible.
  If $s = 2$, the only possible vertex figures are $\mathbf{3.4^2.6}$, or **3.4.6.4**.
  Clearly we cannot have $s = 3$ (or we would be left with a gap of 30°); and $t = 1$, $s > 3$ is also impossible.
* Hence we may assume that $t = 0$, in which case s is at most 4.
  If $s = 4$, then the vertex figure is $\mathbf{4^4}$.
  If $s = 3$, then the remaining gap could only take a fourth square, so this case does not occur.
  If $s = 2$, we are left with an angle of 180°, which cannot be filled.
  If $s = 1$, we are left with an angle of 270°, so there must be exactly two additional tiles and the only possible vertex figures are **4.5.20**, **4.6.12**, or $\mathbf{4.8^2}$ (since a regular 7-gon would leave an angle of $141\frac{3}{7}^\circ$).

* Hence we may assume that $t = s = 0$, and proceed using the parameter f – namely the number of regular pentagons. Clearly f is at most 3, and cannot equal 3 (or we would leave an angle of 36°).
  If $f = 2$, we are left with an angle of 144°, so the only vertex figure is $\mathbf{5^2.10}$.
  If $f = 1$, we are left with an angle of 252°, which requires exactly two further tiles, whose average angle is 126°; but this forces us to use a hexagon – leaving an angle of 132°, which is impossible.

* Hence we may assume that $t = s = f = 0$. So the smallest possible tile is a hexagon, and since we need at least 3 tiles at each vertex, the only possible vertex figure is $\mathbf{6^3}$.
  Hence, the simple minded necessary condition (namely that the vertex figure should have no gaps) gives rise to a list of twenty-one possible vertex figures:

  $\mathbf{3^6}$, $\mathbf{3^4.6}$, $\mathbf{3^3.4^2}$, $\mathbf{3^2.4.3.4}$, $\mathbf{3^2.4.12}$, **3.4.3.12**, $\mathbf{3^2.6^2}$, **3.6.3.6**, **3.7.42**, **3.8.24**, **3.9.20**, **3.10.15**, $\mathbf{3.12^2}$, $\mathbf{3.4^2.6}$, **3.4.6.4**, $\mathbf{4^4}$, **4.5.20**, **4.6.12**, $\mathbf{4.8^2}$, $\mathbf{5^2.10}$, $\mathbf{6^3}$.

(ii) **Lemma**. The vertex figures $\mathbf{3^2.4.12}$, **3.4.3.12**, $\mathbf{3^2.6^2}$, **3.7.42**, **3.8.24**, **3.9.20**, **3.10.15**, $\mathbf{3.4^2.6}$, **4.5.20**, $\mathbf{5^2.10}$ do not extend to semi-regular tilings of the plane.

Proof. Suppose to the contrary that any of these vertex figures could be realized by a semi-regular tiling of the plane. Choose a vertex B and consider the tiles around vertex B.

In the first eight of the listed vertex figures we may choose a triangle $\boldsymbol{T} = ABC$, which is adjacent to polygons of different sizes on the edges $\underline{BA}$ (say) and $\underline{BC}$.

In the two remaining vertex figures, there is a face $\boldsymbol{T} = ABC\cdots$ with an **odd** number of edges, which has the same property – namely that of being adjacent to an a-gon on the edge $\underline{BA}$ (say) and a b-gon on the edge $\underline{BC}$ with $a \neq b$. (For example, in the vertex figure $\mathbf{3^2.4.12}$, $\boldsymbol{T} = ABC$ is a triangle, and the faces on $\underline{BA}$ and on $\underline{BC}$ are – in some order – a 3-gon and a 4-gon, or a 3-gon and a 12-gon.)

In each case let the face $\boldsymbol{T}$ be a p-gon.

If the face adjacent to $\boldsymbol{T}$ on the edge $\underline{BA}$ is an a-gon, and that on edge $\underline{BC}$ is a b-gon, then the vertex figure symbol must include the sequence "$\boldsymbol{\ldots a.p.b \ldots}$".

If we now switch attention from vertex B to the vertex A, then we know that A has the same vertex figure, so must include the sequence "$\boldsymbol{\ldots a.p.b \ldots}$", so the face adjacent to the other edge of $\boldsymbol{T}$ at A must be a b-gon. As one traces round the edges of the face $\boldsymbol{T}$, the faces adjacent to T are alternately a-gons and b-gons – contradicting the fact that $\boldsymbol{T}$ has an **odd** number of edges.

Hence none of the listed vertex figures extends to a semi-regular tiling of the plane. QED

(b) It transpires that the remaining eleven vertex figures

$$\mathbf{3^6},\ \mathbf{3^4.6},\ \mathbf{3^3.4^2},\ \mathbf{3^2.4.3.4},\ \mathbf{3.6.3.6},\ \mathbf{3.12^2},\ \mathbf{3.4.6.4},\ \mathbf{4^4},\ \mathbf{4.6.12},\ \mathbf{4.8^2},\ \mathbf{6^3}$$

can all be realized as semi-regular tilings (and one – namely $\mathbf{3^4.6}$ – can be realized in two different ways, one being a reflection of the other).

In the spirit of Problem **180**(b), one should want to do better than to produce plausible pictures of such tilings, by specifying each one in some canonical way. We leave this challenge to the reader.

182. [We construct a regular hexagon, and take alternate vertices.]

Draw the circle with centre O passing through A. The circle with centre A passing through O meets the circle again at X and Y.

The circle with centre X and passing through A and O meets the circle again at B; and the circle with centre Y and passing through A and O, meets the circle again at C.

Then $\triangle AOX$, $\triangle AOY$, $\triangle XOB$, $\triangle YOC$ are equilateral, so $\angle AXB = 120^\circ = \angle AYC$, and $\angle XAB = 30^\circ = \angle YAC$.

Hence $\triangle AXB \equiv \triangle AYC$, so $\underline{AB} = \underline{AC}$. $\triangle ABC$ is isosceles so $\angle B = \angle C$, with apex angle $\angle BAC = \angle XAY - \angle XAB - \angle YAC = 60^\circ$; hence $\triangle ABC$ is equiangular and so equilateral.

183.

(a) Draw the circle with centre O and passing through A.

Extend AO beyond O to meet the circle again at C.

Construct the perpendicular bisector of $\underline{AC}$, and let this meet the circle at B and at D.

Then $\underline{BA} = \underline{BC}$ (since the perpendicular bisector of $\underline{AC}$ is the locus of points equidistant from A and from C); similarly $\underline{DA} = \underline{DC}$.

$\triangle OAB$ and $\triangle OCB$ are both isosceles right angled triangles, so $\angle ABC$ is a right angle (or appeal to "the angle subtended on the circle by the diameter $\underline{AC}$"). Similarly $\angle A$, $\angle C$, $\angle D$ are right angles, so $ABCD$ is a rectangle with $\underline{BA} = \underline{BC}$, and hence a square.

Note: This construction starts with the regular 2-gon $\underline{AC}$ inscribed in its circumcircle, and doubles it to get a regular 4-gon, by constructing the perpendicular bisectors of the "two sides" to meet the circumcircle at B and at D.

(b) Erect the perpendiculars to AB at A and at B.

Then draw the circles with centre A and passing through B, and with centre B and passing through A.

Let these circles meet the perpendiculars to AB (on the same side of AB) at D and at C.

Then $\underline{AD} = \underline{BC}$, and $AD \parallel BC$, so $ABCD$ is a parallelogram (by Problem **159**(a)), and hence a (being a parallelogram with a right angle), and so a square (since $\underline{AB} = \underline{AD}$).

184.

(a) (i) First construct the regular 3-gon ACE with circumcentre O. Then construct the perpendicular bisectors of the three sides $\underline{AC}$, $\underline{CE}$, $\underline{EA}$, and let these meet the circumcircle at B, D, F.

Note: Here we emphasise the general step from inscribed regular n-gon to inscribed regular $2n$-gon – even though this may seem perverse in the case of a regular 3-gon (since we constructed the inscribed regular 3-gon in Problem **182** by first constructing the regular 6-gon and then taking alternate vertices).

(ii) First construct the regular 4-gon $ACEG$ with circumcentre O. Then construct the perpendicular bisectors of the four sides $\underline{AC}$, $\underline{CE}$, $\underline{EG}$, $\underline{GA}$, and let these meet the circumcircle at B, D, F, H.

(b) (i) Construct an equilateral triangle ABO.

Then draw the circle with centre O and passing through A and B. Then proceed as in **182**.

(ii) Extend AB beyond B, and let this line meet the circle with centre B and passing through A at X.

Now construct a square $BXYZ$ on the side $\underline{BX}$ as in **183**(a), and let the diagonal BY meet the circle at C.

Construct the circumcentre O of $\triangle ABC$ (the point where the perpendicular bisectors of $\underline{AB}$, $\underline{BC}$ meet).

Construct the next vertex D as the point where the circle with centre O and passing through A meets the circle with centre C and passing through B. The remaining points E, F, G, H can be found in a similar way.

185.

(a)(i) There are various ways of doing this – none of a kind that most of us might stumble upon. Draw the circumcircle with centre O and passing through A. Extend the line AO beyond O to meet the circle again at X.

Construct the perpendicular bisector of $\underline{AX}$, and let this meet the circle at Y and at Z. Construct the midpoint M of $\underline{OZ}$, and join $\underline{MA}$.

Let the circle with centre M and passing through A meet the line segment $\underline{OY}$ at the point F.

Finally let the circle with centre A and passing through F meet the circumcircle at B. Then $\underline{AB}$ is a side of the required regular 5-gon. (The vertex C on the circumcircle is then obtained as the second meeting point of the circumcircle with the circle having centre B and passing through A. The points D, E can be found in a similar way.)

The proof that this construction does what is claimed is most easily accomplished by calculating lengths.

Let $\underline{OA} = 2$. Then $\underline{OF} = \sqrt{5} - 1$, so $\underline{AF} = \sqrt{10 - 2\sqrt{5}} = \underline{AB}$. It remains to prove that this is the correct length for the side of a regular pentagon inscribed in a circle of radius 2. Fortunately the work has already been done, since $\triangle OAB$ is isosceles with apex angle equal to 72°. If we drop a perpendicular from O to AB, then we need to check whether it is true that

$$\sin 36^\circ = \frac{\sqrt{10 - 2\sqrt{5}}}{4}.$$

But this was already shown in Problem **3**(c).

(ii) To construct a regular 10-gon $ABCDEFGHIJ$, first construct a regular 5-gon $ACEGI$ with circumcentre O; then construct the perpendicular bisectors of the five sides, and so find B, D, F, H, J as the points where these bisectors meet the circumcircle.

(b)(i) The first move is to construct a line BX through B such that $\angle ABX = 108^\circ$. Fortunately this can be done using part (a), by temporarily treating B as the circumcentre, drawing the circle with centre B and passing through A, and beginning the construction of a regular 5-gon $AP\cdots$ inscribed in this circle.

Then $\angle ABP = 72°$; so if we extend the line PB beyond B to X, then $\angle ABX = 108°$. Let BX meet the circle with centre B through A at the point C. Then $\underline{BA} = \underline{BC}$ and $\angle ABC = 108°$, so we are up and running.

If we let the perpendicular bisectors of $\underline{AB}$ and $\underline{BC}$ meet at O, then the circle with centre O and passing through A also passes through B and C (and the yet to be located points D and E). The circle with centre C and passing through B meets this circle again at D; and the circle with centre A and passing through B meets the circle again at E.

(ii) To construct the regular 10-gon $ABCDEFGHIJ$, treat B as the point O in (a)(i) and construct a regular 5-gon $AXCYZ$ inscribed in the circle with centre B and passing through A.

Then $\angle ABC = 144°$, and $\underline{BA} = \underline{BC}$, so C is the next vertex of the required regular 10-gon. We may now proceed as in (a)(ii) to first construct the circumcentre O of the required regular 10-gon as the point where the perpendicular bisectors of $\underline{AB}$ and $\underline{BC}$ meet, then draw the circumcircle, and finally step off successive vertices D, E, ... of the 10-gon around the circumcircle.

186. The number k of faces meeting at each vertex can be at most five (since more would produce an angle sum that is too large). And $k \geqslant 3$ (in order to create a genuine corner.

- If $k = 5$, then the vertex figure must be $\mathbf{3^5}$ (or the angle sum would be $> 360°$).
- If $k = 4$, then the vertex figure must be $\mathbf{3^4}$ (or the angle sum would be too large).
- If $k = 3$, the angle in each of the regular polygons must be $< 120°$, so the only possible vertex figures are $\mathbf{5^3}$, $\mathbf{4^3}$, and $\mathbf{3^3}$.

187. The respective midpoints have coordinates:

of $\underline{AB}$: $\left(\frac{1}{2}, \frac{1}{2}, 0\right)$; of $\underline{AC}$: $\left(\frac{1}{2}, 0, \frac{1}{2}\right)$; of $\underline{AD}$: $\left(1, \frac{1}{2}, \frac{1}{2}\right)$; of $\underline{BC}$: $\left(0, \frac{1}{2}, \frac{1}{2}\right)$;

of $\underline{BD}$: $\left(\frac{1}{2}, 1, \frac{1}{2}\right)$; of $\underline{CD}$: $\left(\frac{1}{2}, \frac{1}{2}, 1\right)$.

$\therefore \underline{PQ} = \frac{1}{\sqrt{2}} = \underline{PR} = \underline{PS} = \underline{PT} = \underline{QR} = \underline{RS} = \underline{ST} = \underline{TQ}$.

188.

(a) There are infinitely many planes through the apex A and the base vertex B. Among these planes, the one perpendicular to the base BCD is the one that passes through the midpoint M of $\underline{CD}$. Let the perpendicular from A to the base, meet the base BCD at the point X, which must lie on $\underline{BM}$. Let $\underline{AX}$ have length h. To find h we calculate the area of $\triangle ABM$ in two ways.

First, $\underline{BM}$ has length $\sqrt{3}$, so area$(\triangle ABM) = \frac{1}{2}\left(\sqrt{3} \times h\right)$

Second, $\triangle ABM$ is isosceles with base $\underline{AB}$ and apex M, so has height $\sqrt{2}$.

$\therefore$ area$(\triangle ABM) = \frac{1}{2}(2 \times \sqrt{2}) = \sqrt{2}$.

If we now equate the two expressions for area$(\triangle ABM)$, we see that $h = \sqrt{\frac{8}{3}}$.

(b)(i) When constructing a regular octahedron (whether using card, or tiles such as Polydron®) one begins by arranging four equilateral triangles around a vertex such as A. This 'vertex figure' is not rigid: though we know that $\underline{BC} = \underline{CD} = \underline{DE} = \underline{EB}$, there is no *a priori* reason why the four neighbours B, C, D, E of A should form a square, or a rhombus, or a planar quadrilateral. We show that these four neighbours lie in a single plane: B, C, D, E are all distance 2 from A and distance 2 from F, so (by Problem **144**) they must all lie in the plane perpendicular to the line AF and passing through the midpoint X of $\underline{AF}$. Moreover, $\triangle ABX \equiv \triangle ACX$ (by RHS-congruence, since $\underline{AB} = \underline{AC} = 2$, $\underline{AX} = \underline{AX}$, and $\angle AXB = \angle AXC$ are both right angles). Hence $\underline{XB} = \underline{XC}$, so B, C lie on a circle in this plane with centre X. Similarly $\triangle ABX \equiv \triangle ADX \equiv \triangle AEX$, so $BCDE$ is a cyclic quadrilateral (and a rhombus), and hence a square – with X as the midpoint of both $\underline{BD}$ and $\underline{CE}$.

(ii) Let M be the midpoint of $\underline{BC}$ and N the midpoint of $\underline{DE}$.
Then NM and AF cross at X and so define a single plane. In this plane, $\triangle ANM \equiv \triangle FMN$ (by SSS-congruence, since $\underline{AN} = \underline{FM} = \sqrt{3}$, $\underline{NM} = \underline{MN}$, $\underline{MA} = \underline{NF} = \sqrt{3}$); hence $\angle ANM = \angle FMN$, so $AN \parallel MF$.
Similarly, if P is the midpoint of $\underline{AE}$ and Q is the midpoint of $\underline{FC}$, then $\triangle DPQ \equiv \triangle BQP$, so $DP \parallel QB$. Hence the top face DEA is parallel to the bottom face BCF, so the height of the octahedron sitting on the table is equal to the height of $\triangle FMN$. But this triangle has sides of lengths 2, $\sqrt{3}$, $\sqrt{3}$, so this height is exactly the same as the height h in part (a).

189.

(i) Let $L = \left(\frac{1}{2}, 0, 0\right)$ and $M = \left(\frac{1}{2}, \frac{1}{2}, 0\right)$. Then L lies on the line ST and M is the midpoint of $\underline{NP}$.
$\triangle LSM$ is a right angled triangle with legs of length $\underline{LS} = a$, $\underline{LM} = \frac{1}{2}$, so $\underline{MS} = \sqrt{a^2 + \frac{1}{4}}$.
$\triangle MNS$ is a right angled triangle with legs of length $\underline{MN} = \frac{1}{2} - a$, $\underline{MS} = \sqrt{a^2 + \frac{1}{4}}$. Hence $\underline{NS} = \sqrt{2a^2 - a + \frac{1}{2}}$.
Similarly $\underline{NU} = \sqrt{2a^2 - a + \frac{1}{2}}$.
Let $L' = \left(0, \frac{1}{2}, 0\right)$ and $M' = \left(0, \frac{1}{2}, \frac{1}{2}\right)$. Then M' is the midpoint of $\underline{WX}$ and L' lies immediately below M' on the line joining $(0,0,0)$ to $(0,1,0)$. In the right angled triangle $\triangle L'NM'$ we find $\underline{NM'} = \sqrt{a^2 + \frac{1}{4}}$, and in the right angled triangle $\triangle M'WN$ we then find $\underline{NW} = \sqrt{2a^2 - a + \frac{1}{2}}$. Similarly $\underline{NX} = \sqrt{2a^2 - a + \frac{1}{2}}$.

(ii) $\underline{NP} = \underline{NS}$ precisely when $1 - 2a = \sqrt{2a^2 - a + \frac{1}{2}} > 0$; that is, when $a = \frac{3-\sqrt{5}}{4}$, so all edges of the polyhedron have length $\frac{\sqrt{5}-1}{2} = \tau - 1$.

Note: The rectangle $NPRQ$ is a "1 by $\tau - 1$" rectangle, and $1 : \tau - 1 = \tau : 1$. Hence the regular icosahedron can be constructed from three congruent, and pairwise perpendicular, copies of a "Golden rectangle".

190. We mimic the classification of possible vertex figures for semi-regular tilings.

We are assuming that the angles meeting at each vertex add to $< 360°$, so the number k of faces at each vertex lies between 3 and 5. Because faces are regular, but not necessarily congruent, k does not determine the shape of the faces. Hence we let t denote the number of triangles at each vertex, which can range from 0 up to 5.

- If $t = 5$, then the vertex figure must be $\mathbf{3^5}$.
- If $t = 4$, the remaining polygons have angle sum $< 120°$, so the possible vertex figures are $\mathbf{3^4}$, $\mathbf{3^4.4}$, and $\mathbf{3^4.5}$.
- If $t = 3$, then the remaining polygons have angle sum $< 180°$, so the only possible vertex figures are $\mathbf{3^3}$, and $\mathbf{3^3}.\boldsymbol{n}$ (for any $n > 3$).
- If $t = 2$, then the remaining polygons have angle sum $< 240°$, so there are at most 2 additional faces in the vertex figure (since if there were 3 or more extra faces, the average angle size would then be at most $80°$, with no more triangles allowed). If there is just 1 additional face, we get the vertex figure $\mathbf{3^2}.\boldsymbol{n}$ for any $n > 3$. So we may assume that there are 2 additional faces – the smallest of which must then be a 4-gon or a 5-gon.

  If the next smallest face is a 4-gon, then we get the possible vertex figures

  $\mathbf{3^2.4^2}$ and **3.4.3.4**, $\mathbf{3^2.4.5}$ and **3.4.3.5**, $\mathbf{3^2.4.6}$ and **3.4.3.6**, $\mathbf{3^2.4.7}$ and **3.4.3.7**, $\mathbf{3^2.4.8}$ and **3.4.3.8**, $\mathbf{3^2.4.9}$ and **3.4.3.9**, $\mathbf{3^2.4.10}$ and **3.4.3.10**, $\mathbf{3^2.4.11}$ and **3.4.3.11**.

  If the next smallest face is a 5-gon, then we get the possible vertex figures

  $\mathbf{3^2.5^2}$ and **3.5.3.5**, $\mathbf{3^2.5.6}$ and **3.5.3.6**, $\mathbf{3^2.5.7}$ and **3.5.3.7**.

Note: Before proceeding further it is worth deciding which among the putative vertex figures identified so far seem to correspond to semi-regular polyhedra – and then to prove that these observations are correct.

- The vertex figure $\mathbf{3^5}$ corresponds to the regular icosahedron.
- The vertex figure $\mathbf{3^4}$ corresponds to the regular octahedron; $\mathbf{3^4.4}$ corresponds to the *snubcube*; $\mathbf{3^4.5}$ corresponds to the *snub dodecahedron* – which comes in left-handed and right-handed forms.
- The vertex figure $\mathbf{3^3}$ corresponds to the regular tetrahedron, and $\mathbf{3^3}.\boldsymbol{n}$ (for any $n > 3$) corresponds to the n-gonal *antiprism.*
- The vertex figure $\mathbf{3^2}.\boldsymbol{n}$ for any $n > 3$ does not seem to arise.
- The vertex figure $\mathbf{3^2.4^2}$ does not seem to arise; **3.4.3.4** corresponds to the *cuboctahedron*; $\mathbf{3^2.4.5}$ and **3.4.3.5**, $\mathbf{3^2.4.6}$ and **3.4.3.6**, $\mathbf{3^2.4.7}$ and **3.4.3.7**, $\mathbf{3^2.4.8}$ and **3.4.3.8**, $\mathbf{3^2.4.9}$ and **3.4.3.9**, $\mathbf{3^2.4.10}$ and **3.4.3.10**, $\mathbf{3^2.4.11}$ and **3.4.3.11** do not seem to arise.
- The vertex figure $\mathbf{3^2.5^2}$ does not seem to arise, whereas **3.5.3.5** corresponds to the *icosidodecahedron*; $\mathbf{3^2.5.6}$ and **3.5.3.6**, $\mathbf{3^2.5.7}$ and **3.5.3.7** do not seem to arise.

To avoid further proliferation of spurious 'putative vertex figures' we inject a version of the **Lemma** used for tilings somewhat earlier than we did for tilings, and then apply the underlying idea to eliminate other spurious possibilities as they arise.

Lemma. The vertex figures

> $\mathbf{3^2.4^2}$, $\mathbf{3^2.}\boldsymbol{n}$ $(n > 3)$, $\mathbf{3^2.4.5}$, **3.4.3.5**, $\mathbf{3^2.4.6}$, **3.4.3.6**, $\mathbf{3^2.4.7}$, **3.4.3.7**, $\mathbf{3^2.4.8}$, **3.4.3.8**, $\mathbf{3^2.4.9}$, **3.4.3.9**, $\mathbf{3^2.4.10}$, **3.4.3.10**, $\mathbf{3^2.4.11}$, **3.4.3.11**, $\mathbf{3^2.5^2}$, $\mathbf{3^2.5.6}$, **3.5.3.6**, $\mathbf{3^2.5.7}$, **3.5.3.7**

do not arise as vertex figures of any semi-regular polyhedron.

Proof outline. Each of these requires that the vertex figure of any vertex B includes a tile $\boldsymbol{T} = ABC\cdots$ with an **odd** number of edges, for which the edge $\underline{BA}$ is adjacent to an a-gon, the edge $\underline{BC}$ is adjacent to a b-gon (where $a \neq b$), and where the subsequent faces adjacent to $\boldsymbol{T}$ are forced to alternate – a-gon, b-gon, a-gon, ...– which is impossible. QED

For the rest we introduce the additional parameter s to denote the number of 4-gons in the vertex figure.

Suppose $t = 1$. Then the remaining polygons have angle sum $< 300^\circ$, so there are at most 3 additional faces in the vertex figure (since if there were 4 or more extra faces, the average angle size would be $< 75^\circ$, with no more triangles allowed).

If there are 3 additional faces, the average angle size is $< 100^\circ$, so $s > 0$.

- If $s > 1$, then the possible vertex figures are $\mathbf{3.4^3}$ (which corresponds to the *rhombicuboctahedron*), $\mathbf{3.4^2}$ (which corresponds to the 3-gonal prism), $\mathbf{3.4^2.5}$ (which is impossible as in the Lemma) and **3.4.5.4** (which corresponds to the *small rhombicosidodecahedron*).
- If $s = 1$, then the remaining faces have angle sum $< 210^\circ$, so there can only be one additional face, and every $\mathbf{3.4.}\boldsymbol{n}$ $(n > 4)$ is impossible as in the Lemma.
- If $s = 0$, then the remaining faces have angle sum $< 300^\circ$, so there are exactly two other faces and the smallest face has < 12 edges, so the only possible vertex figures are
  - $\mathbf{3.5.}\boldsymbol{n}$ $(4 < n)$, which is impossible as in the Lemma;
  - $\mathbf{3.6^2}$, which corresponds to the *truncated tetrahedron*;
  - $\mathbf{3.6.}\boldsymbol{n}$ $(6 < n)$, which is impossible as in the Lemma;
  - $\mathbf{3.7.}\boldsymbol{n}$ $(6 < n)$, which is impossible as in the Lemma;
  - $\mathbf{3.8^2}$, which corresponds to the *truncated cube*;
  - $\mathbf{3.8.}\boldsymbol{n}$ $(8 < n)$, which is impossible as in the Lemma;
  - $\mathbf{3.9.}\boldsymbol{n}$ $(8 < n)$, which is impossible as in the Lemma;
  - $\mathbf{3.10^2}$, which corresponds to the *truncated dodecahedron*;
  - $\mathbf{3.10.}\boldsymbol{n}$ $(10 < n)$, which is impossible as in the Lemma;
  - $\mathbf{3.11.}\boldsymbol{n}$ $(10 < n)$, which is impossible as in the Lemma.

Thus we may assume that $t = 0$ – in which case, $s < 4$.

- If $s = 3$, then the only possible vertex figure is $\mathbf{4^3}$, which corresponds to the *cube.*
- If $s = 2$, then the remaining faces have angle sum $< 180°$, so there is exactly one additional face, and every $\mathbf{4^2.n}$ ($n > 4$) corresponds to the n-gonal *prism.*
- If $s = 1$, then the remaining faces have angle sum $< 270°$, so there are at most 2 other faces with the smallest face having < 8 edges, so the only possible vertex figures are $\mathbf{4.5.n}$ (for $4 < n < 20$), $\mathbf{4.6^2}$, $\mathbf{4.6.n}$ (for $6 < n < 12$), $\mathbf{4.7^2}$, $\mathbf{4.7.n}$ (for $7 < n < 10$). Among these $\mathbf{4.5.n}$ is impossible as in the Lemma; $\mathbf{4.6^2}$ corresponds to the *truncated octahedron*; $\mathbf{4.6.n}$ with n odd is impossible as in the Lemma; **4.6.8** corresponds to the *great rhombicuboctahedron*; **4.6.10** corresponds to the *great rhombicosidodecahedron*; $\mathbf{4.7.n}$ is impossible as in the Lemma.
- If $s = 0$, there must be exactly three faces at each vertex, and the smallest must be a 5-gon, so the only possible vertex figures are $\mathbf{5^3}$, which corresponds to the *regular dodecahedron*; $\mathbf{5^2.n}$ (for $n > 5$), which is impossible as in the Lemma; $\mathbf{5.6^2}$, which corresponds to the *truncated icosahedron*; or **5.6.7**, which is impossible as in the Lemma.

191.

(a) (i) $\angle AXD = 100°$, so $\angle ADB = 40°$. In $\triangle ABD$ we then see that $\angle ABD = 40°$, so $\triangle ABD$ is isosceles with $\underline{AB} = \underline{AD}$. $\triangle ABC$ is isosceles with a base angle $\angle BAC = 60°$, so $\triangle ABC$ is equilateral. Hence $\angle CBD = 20°$.

(ii) $\triangle ABC$ is equilateral, so $\underline{AC} = \underline{AB} = \underline{AD}$.
Hence $\triangle ADC$ is isosceles, so $\angle ACD = 70° = \angle ADC$, whence $\angle BDC = 70° - \angle ADB = 30°$.

(b) (i) As before $\angle AXD = 100°$, so $\angle ADB = 40°$.
In $\triangle ABD$ we then see that $\angle ABD = 30°$, so $\triangle ABD$ is not isosceles (as it was in (a)).
$\triangle ABC$ is isosceles, so $\angle BCA = 70°$, whence $\angle CBD = 10°$.
In $\triangle XCD$, $\angle CXD = 80°$, so $\angle BDC + \angle ACD = 100°$, but there is no obvious way of determining the individual summands: $\angle BDC$ and $\angle ACD$.

(ii) No lengths are specified, so we may choose the length of $\underline{AC}$. The point B then lies on the perpendicular bisector of $\underline{AC}$, and $\angle CAB = 70°$ determines the location of B exactly. The line AD makes an angle of $40°$ with AC , and BD makes an angle of $80°$ with AC, so the location of D is determined. Hence, despite our failure in part (i), the angles are determined.

192.

(i) If P lies on $\underline{CB}$, then $\underline{PC} = b\cos C$, $\underline{AP} = b\sin C$, and in the right angled triangle $\triangle APB$ we have:

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

If P lies on $\underline{CB}$ extended beyond B, then $\underline{PC} = b\cos C$, $\underline{AP} = b\sin C$ as before, and in the right angled triangle $\triangle APB$ we have:

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

(ii) If P lies on $\underline{BC}$ extended beyond C, then $\underline{PC} = b\cos\angle ACP = -b\cos C$, $\underline{AP} = b\sin\angle ACP = b\sin C$, and in the right angled triangle $\triangle APB$ we have:

$$\begin{aligned} c^2 &= (b\sin C)^2 + (a + \underline{PC})^2 \\ &= (b\sin C)^2 + (a - b\cos C)^2 \\ &= a^2 + b^2(\sin^2 C + \cos^2 C) - 2ab\cos C \\ &= a^2 + b^2 - 2ab\cos C. \end{aligned}$$

193. There are many ways of doing this – once one knows the Sine Rule and Cosine Rule. If we let $\angle BDC = y$ and $\angle ACD = z$, then one route leads to the identity $\cos(z - 10^\circ) = 2\sin 10^\circ \cdot \sin z$, from which it follows that $z = 80^\circ$, $y = 20^\circ$.

194.

(a) The angle between two faces, or two planes, is the angle one sees "end-on" – as one looks along the line of intersection of the two planes. This is equal to the angle between two perpendiculars to the line of intersection – one in each plane.
If M is the midpoint of $\underline{BC}$, then $\triangle ABC$ is isosceles with apex A, so the median $\underline{AM}$ is perpendicular to $\underline{BC}$; similarly $\triangle DBC$ is isosceles with apex D, so the median $\underline{DM}$ is perpendicular to $\underline{BC}$.
$\triangle MAD$ is isosceles with apex M, $\underline{MA} = \underline{MD} = \sqrt{3}$, $\underline{AD} = 2$, so we can use the Cosine Rule to conclude that $2^2 = 3 + 3 - 2\cdot 3\cdot\cos(\angle AMD)$, whence $\cos(\angle AMD) = \frac{1}{3}$.

(b) $\cos(\angle AMD) = \frac{1}{3}$, and $\frac{1}{3} < \frac{1}{2}$, so $\angle AMD > 60^\circ$; hence we cannot fit six regular tetrahedra together so as to share an edge.
Since $\angle AMD$ is acute, $\angle AMD < 90^\circ$, so we can certainly fit four regular tetrahedra together with lots of room to spare.
We can now appeal to "trigonometric tables", or a calculator, to see that

$$\arccos\left(\frac{1}{3}\right) < 1.24,$$

that is 1.24 *radians*, so five tetrahedra use up less than 6.2 radians – which is less than 2π. Hence we can fit five regular tetrahedra together around a common edge with room to spare (but not enough to fit a sixth).

195.

(a) The angle between the faces ABC and FBC is equal to the angle between two perpendiculars to the common edge $\underline{BC}$. Since the two triangles are isosceles with the common base $\underline{BC}$, it suffices to find the angle between the two medians $\underline{AM}$ and $\underline{FM}$.

In Problem **188** we saw that $BCDE$ is a square, with sides of length 2. If we switch attention from the opposite pair of vertices A, F to the pair C, E, then the same proof shows that $ABFD$ is a square with sides of length 2. Hence the diagonal $\underline{AF} = 2\sqrt{2}$.

Now apply the Cosine Rule to $\triangle AMF$ to conclude that:

$$\left(2\sqrt{2}\right)^2 = 3 + 3 - 2 \cdot 3 \cdot \cos(\angle AMF),$$

so it follows that $\cos(\angle AMF) = -\frac{1}{3}$.

(b) $\cos(\angle AMF) = -\frac{1}{3} < 0$, so $\angle AMF > 90°$; hence we cannot fit four regular octahedra together so as to share a common edge. Moreover, $-\frac{1}{2} < -\frac{1}{3}$, so $\angle AMD < 120°$; hence we can fit three octahedra together to share an edge with room to spare (but not enough room to fit a fourth).

196. The angle $\angle AMD = \arccos\left(\frac{1}{3}\right)$ in Problem **194** is acute, and the angle $\angle AMF = \arccos\left(-\frac{1}{3}\right)$ in Problem **195** is obtuse. Hence these angles are supplementary; so the regular tetrahedron fits *exactly* into the wedge-shaped hole between the face ABC of the regular octahedron and the table.

197.

(i) $\underline{AB} = \sqrt{2} = \underline{BC} = \underline{CA} = \underline{AW} = \underline{BW} = \underline{CW}$. Hence the four faces ABC, BCW, CWA, WAB are all equilateral triangles, so the solid is a regular tetrahedron (or, more correctly, the surface of the solid is a regular tetrahedron).

(ii) $\underline{AC} = \sqrt{2} = \underline{CD} = \underline{DA} = \underline{AX} = \underline{CX} = \underline{DX}$. Hence the four faces ACD, CDX, DXA, XAC are equilateral triangles, so the solid is a regular tetrahedron.

(iii) $\underline{AD} = \sqrt{2} = \underline{DE} = \underline{EA} = \underline{AY} = \underline{DY} = \underline{EY}$. Hence the four faces ADE, DEY, EYA, YAD are equilateral triangles, so the solid is a regular tetrahedron.

(iv) $\underline{AE} = \sqrt{2} = \underline{EB} = \underline{BA} = \underline{AZ} = \underline{EZ} = \underline{BZ}$. Hence the four faces AEB, EBZ, BZA, ZAE are equilateral triangles, so the solid is a regular tetrahedron.

(v) We get another four regular tetrahedra – such as $FBCP$, where $P = (1, 1, -1)$.

(vi) $ABCDEF$ is a regular octahedron (or, more correctly, the surface of the solid is a regular octahedron).

Note: The six vertices $(\pm1,0,0)$, $(0,\pm1,0)$, $(0,0,\pm1)$ span a regular octahedron, with a regular tetrahedron fitting exactly on each face. The resulting compound star-shaped figure is called the *stellated octahedron.* Johannes Kepler (1571–1630) made an extensive study of polyhedra and this figure is sometimes referred to as Kepler's 'stella octangula'. It is worth making in order to appreciate the way it appears to consist of two interlocking tetrahedra.

198. Let the unlabeled vertex of the pentagonal face $ABW-V$ be P. In the pentagon $ABWPV$ the edge $\underline{AB}$ is parallel to the diagonal $\underline{VW}$. Hence $ABWV$ is an isosceles trapezium. The sides VA and WB (produced) meet at S in the plane of the pentagon $ABWPV$.

$\triangle SAB$ has equal base angles, so $\underline{SA} = \underline{SB}$.

Hence $\underline{SV} = \underline{SW}$.

Similarly $BC \parallel WX$, and WB and XC meet at some point S' on the line WB.

Now $\triangle S'BC \equiv \triangle SAB$, so $\underline{S'B} = \underline{SA} = \underline{SB}$. Hence $S' = S$, and the lines VA, WB, XC, YD, ZE all meet at S.

Since $VW \parallel AB$, we know that $\triangle SAB \sim \triangle SVW$, with scale factor

$$\tau : 1 = \underline{VW} : \underline{AB} = \underline{SV} : \underline{SA}.$$

If $\underline{SA} = x$, then $x+1 : x = \tau : 1$, so $x = \tau$.

Let M be the midpoint of $\underline{AB}$ and let O denote the circumcentre of the regular pentagon $ABCDE$.

$\triangle OAB$ is isosceles, so OM is perpendicular to AB, and the required dihedral angle between the two pentagonal faces is $\angle OMP$.

It turns out to be better to find not the dihedral angle $\angle OMP$, but its supplement: namely the angle $\angle SMO$.

From $\triangle OAM$ we see that

$$\underline{OA} = \frac{1}{2\sin 36^\circ},$$

and we know that $\underline{SA} = \tau = 2\cos 36^\circ$. One can then check that $\underline{OS} = \cot 36^\circ$.

Similarly, in $\triangle OAM$ we have

$$\underline{OM} = \frac{\cot 36^\circ}{2}.$$

Hence in $\triangle OMS$ we have

$$\tan(\angle SMO) = 2.$$

Hence the required dihedral angle is equal to $\pi - \arctan 2 \approx 116.56^\circ$.

199.

(a) (i) Let M be the midpoint of $\underline{AC}$. The angle between the two faces is equal to $\angle BMD$.

In $\triangle BMD$, we have $\underline{BM} = \underline{DM} = \sqrt{3}$, $\underline{BD} = 2\tau = 1 + \sqrt{5}$. So the Cosine Rule in $\triangle BMD$ gives:

$$6 + 2\sqrt{5} = 3 + 3 - 2 \cdot 3 \cdot \cos(\angle BMD),$$

so $\cos(\angle BMD) = -\frac{\sqrt{5}}{3}$,

$$\angle BMD = \arccos\left(-\frac{\sqrt{5}}{3}\right) = \pi - \arccos\left(\frac{\sqrt{5}}{3}\right) \approx 138.19^\circ.$$

(ii) $\frac{\sqrt{5}}{3} > \frac{1}{2}$; hence $\angle BMD > \frac{2\pi}{3}$, so we can fit two copies along a common edge, but not three.

(b)(i) Now let M denote the midpoint of $\underline{BC}$, and suppose that OM extended beyond M meets the circumcircle at V. Then $\underline{BV}$ is an edge of the regular 10-gon inscribed in the circumcircle. The circumradius $\underline{OB}$ of $BCDEF$ (which is equal to the edge length of the inscribed regular hexagon) is

$$\underline{OB} = \frac{1}{\sin 36^\circ};$$

and $\angle MBV = 18^\circ$, so

$$\underline{BV} = \frac{1}{\cos 18^\circ}.$$

It is easiest to use the converse of Pythagoras' Theorem, and to write everything first in terms of $\cos 36^\circ$, then (since $\tau = 2\cos 36^\circ$, so $\cos 36^\circ = \frac{1+\sqrt{5}}{4}$) write everything in terms of $\sqrt{5}$.

If we use $\sin^2 36^\circ = 1 - \cos^2 36^\circ$, and $2\cos^2 18^\circ - 1 = \cos 36^\circ = \frac{1+\sqrt{5}}{4}$, then

$$\begin{aligned}
\underline{BV}^2 + \underline{OB}^2 &= \left(\frac{1}{\cos 18^\circ}\right)^2 + \left(\frac{1}{\sin 36^\circ}\right)^2 \\
&= \frac{8}{5+\sqrt{5}} + \frac{8}{5-\sqrt{5}} \\
&= \frac{80}{20} = 2^2 \\
&= \underline{AB}^2.
\end{aligned}$$

Hence, in the right angled triangle $\triangle AOB$, we must have $\underline{OA} = \underline{BV}$ as claimed.

(ii) Miraculously no more work is needed. Let the vertex at the 'south pole' be L, and let the pentagon formed by its five neighbours be $GHIJK$. It helps if we can refer to the circumcircle of $BCDEF$ as the 'tropic of Cancer', and to the circumcircle of $GHIJK$ as the 'tropic of Capricorn'.

The pentagon $GHIJK$ is parallel to $BCDEF$, but the vertices of the southern pentagon have been rotated through $\frac{\pi}{5}$ relative to $BCDEF$, so that G (say) lies on the circumcircle of the pentagon $GHIJK$, but sits directly below the midpoint of the minor arc $\underline{BC}$. Let X denote the point on the 'tropic of Capricorn' which is directly beneath B. Then $\triangle BXG$ is a right angled triangle with $\underline{BG} = \underline{AB} = 2$, and $\underline{XG} = \underline{BV}$ is equal to the edge length of a regular 10-gon inscribed in the circumcircle of $GHIJK$. Hence, by the calculation in (i), $\underline{BX}$ is equal to the edge length of the regular hexagon inscribed in the same circle – which is also equal to the circumradius.

200. A necessary condition for copies of a regular polyhedron to "tile 3D (without gaps or overlaps)" is that an integral number of copies should fit together around an edge. That is, the dihedral angle of the polyhedron should be an exact submultiple of 2π. Only the cube satisfies this necessary condition.

Moreover, if we take as vertices the points (p, q, r) with integer coordinates p, q, r, and as our regular polyhedra all possible translations of the standard unit cube having opposite corners at $(0, 0, 0)$ and $(1, 1, 1)$, then we see that it is possible to tile 3D using just cubes.

201. The four diameters form the four edges of a square $ABCD$ of edge length 2.

The protruding semicircular segments on the left and right can be cut off and inserted to exactly fill the semicircular indentations above and below.

Hence the composite shape has area exactly equal to $2^2 = 4$ square units.

202.

(a) If the regular n-gon is $ACEG\cdots$ and the regular $2n$-gon is $ABCDEFG\cdots$, then $\underline{AC} = s_n = s$ and $\underline{AB} = s_{2n} = t$. If M is the midpoint of $\underline{AC}$, $\underline{AM} = \frac{s}{2}$ and

$$\underline{MB} = 1 - \sqrt{(1 - \left(\frac{s}{2}\right)^2}.$$

$$\begin{aligned} \therefore \qquad t^2 &= \underline{AB}^2 \\ &= \left(\frac{s}{2}\right)^2 + \left[1 - \sqrt{1 - \left(\frac{s}{2}\right)^2}\right]^2 \\ &= 1 + 1 - 2\sqrt{1 - \left(\frac{s}{2}\right)^2} \\ &= 2 - \sqrt{4 - s^2}. \end{aligned} \qquad (1)$$

(b) Put $s = s_2 = 2$ in (1) to get $t = s_4 = \sqrt{2}$. Then put $s = s_4 = \sqrt{2}$ in (1) to get

$$t = s_8 = \sqrt{2 - \sqrt{2}}.$$

(c) Put $t = s_6 = 1$ in (1) to get $s = s_3 = \sqrt{3}$. Then put $s = s_6 = 1$ to get

$$t = s_{12} = \sqrt{2 - \sqrt{3}}.$$

(d) Put $s = s_5 = \frac{\sqrt{10-2\sqrt{5}}}{2}$ to get

$$t = s_{10} = \sqrt{\frac{3-\sqrt{5}}{2}}.$$

203.

(a)(i) $n = 3$: $p_3 = 3\sqrt{3} \times r$
$n = 4$: $p_4 = 4\sqrt{2} \times r$
$n = 5$: $p_5 = \frac{5\sqrt{10-2\sqrt{5}}}{2} \times r$
$n = 6$: $p_6 = 6 \times r$
$n = 8$: $p_8 = 8\sqrt{2-\sqrt{2}} \times r$
$n = 10$: $p_{10} = 5\sqrt{6-2\sqrt{5}} \times r$
$n = 12$: $p_{12} = 12\sqrt{2-\sqrt{3}} \times r$.

(ii)

$$\begin{aligned} c_3 = 5.19\cdots \quad &< \quad c_4 = 5.65\cdots \\ &< \quad c_5 = 5.87\cdots \\ &< \quad c_6 = 6 \\ &< \quad c_8 = 6.12\cdots \\ &< \quad c_{10} = 6.18\cdots \\ &< \quad c_{12} = 6.21\cdots. \end{aligned}$$

(b)(i) $n = 3$: $P_3 = 6\sqrt{3} \times r$
$n = 4$: $P_4 = 8 \times r$
$n = 5$: $P_5 = 10\sqrt{5-2\sqrt{5}} \times r$
$n = 6$: $P_6 = 4\sqrt{3} \times r$
$n = 8$: $P_8 = 8\left(2\sqrt{2}-2\right) \times r$
$n = 10$: $P_{10} = 4\sqrt{25-10\sqrt{5}} \times r$
$n = 12$: $P_{12} = 12\left(4-2\sqrt{3}\right) \times r$.

(ii)

$$\begin{aligned} C_3 = 10.39\cdots \quad &> \quad C_4 = 8 \\ &> \quad C_5 = 7.26\cdots \\ &> \quad C_6 = 6.92\cdots \\ &> \quad C_8 = 6.62\cdots \\ &> \quad C_{10} = 6.49\cdots \\ &> \quad C_{12} = 6.43\cdots. \end{aligned}$$

(c) Let O be the centre of the circle of radius r. Let A, B lie on the circle with $\angle AOB = 30^\circ$.

Let M be the midpoint of $\underline{AB}$ – so that $\triangle OAB$ is isosceles, with apex O and height $h = \underline{OM}$.

Let A' lie on $\underline{OA}$ produced, and let B' lie on $\underline{OB}$ produced, such that $\underline{OA'} = \underline{OB'}$ and $A'B'$ is tangent to the circle.

Then $\triangle OAB \sim \triangle OA'B'$ with $\triangle OA'B'$ larger than $\triangle OAB$, so the scale factor $\frac{1}{h} = 2\sqrt{2-\sqrt{3}} > 1$.

Hence $P_{12} = 2\sqrt{2-\sqrt{3}} \times p_{12} > p_{12}$, so $C_{12} > c_{12}$.

204. (i) πr (ii) $\frac{\pi}{2}r$ (iii) θr

205.

(a)(i) **Note:** This could be a long slog. However we have done much of the work before: when the radius is 1, the most of the required areas were calculated exactly back in Problem **3** and Problem **19**.

Alternatively, the area of each of the n sectors is equal to $\frac{1}{2}\sin\theta$, where θ is the angle subtended at the centre of the circle, and the exact values of the required trig functions were also calculated back in Chapter 1.

$$\begin{aligned}
a_3 &= \frac{3\sqrt{3}}{4} \times r^2 \\
a_4 &= 2 \times r^2 \\
a_5 &= \frac{5}{8}\sqrt{10+2\sqrt{5}} \times r^2 \\
a_6 &= \frac{3\sqrt{3}}{2} \times r^2 \\
a_8 &= 2\sqrt{2} \times r^2 \\
a_{10} &= \frac{5}{4}\sqrt{10-2\sqrt{5}} \times r^2 \\
a_{12} &= 3 \times r^2
\end{aligned}$$

(ii)

$$\begin{aligned}
d_3 = 1.29\cdots \quad &< \quad d_4 = 2 \\
&< \quad d_5 = 2.37\cdots \\
&< \quad d_6 = 2.59\cdots \\
&< \quad d_8 = 2.82\cdots \\
&< \quad d_{10} = 2.93\cdots \\
&< \quad d_{12} = 3.
\end{aligned}$$

(b)(i) **Note:** This could also be a long slog. However we have done much of the work before. But notice that, when the radius is 1, the area of each of the n sectors is equal to half the edge length times the height ($= 1$); so if $r = 1$, then the total area is numerically equal to "half the perimeter P_n".

$$\begin{aligned}
A_3 &= 3\sqrt{3} \times r^2 \\
A_4 &= 4 \times r^2 \\
A_5 &= 5\sqrt{5 - 2\sqrt{5}} \times r^2 \\
A_6 &= 2\sqrt{3} \times r^2 \\
A_8 &= 8\left(\sqrt{2} - 1\right) \times r^2 \\
A_{10} &= 2\sqrt{25 - 10\sqrt{5}} \times r^2 \\
A_{12} &= 12\left(2 - \sqrt{3}\right) \times r^2
\end{aligned}$$

(ii)

$$\begin{aligned}
D_3 = 5.19\cdots \quad &> \quad D_4 = 4 \\
&> \quad D_5 = 3.63\cdots \\
&> \quad D_6 = 3.46\cdots \\
&> \quad D_8 = 3.31\cdots \\
&> \quad D_{10} = 3.24\cdots \\
&> \quad D_{12} = 3.21\cdots .
\end{aligned}$$

(c) Let O be the centre of the circle of radius r. Let A, B lie on the circle with $\angle AOB = 30^\circ$.

Let M be the midpoint of $\underline{AB}$ – so that $\triangle OAB$ is isosceles, with apex O and height $h = \underline{OM}$.

Let A' lie on $\underline{OA}$ produced, and let B' lie on $\underline{OB}$ produced, such that $\underline{OA'} = \underline{OB'}$ and $A'B'$ is tangent to the circle.

Then $\triangle OAB \sim \triangle OA'B'$ with $\triangle OAB$ contained in $\triangle OA'B'$, so the scale factor $\frac{1}{h} = 2\sqrt{2 - \sqrt{3}} > 1$. Hence

$$4(2 - \sqrt{3}) \cdot a_{12} = A_{12} > a_{12},$$

so $D_{12} > d_{12}$.

206. The rearranged shape (shown in Figure 9) is an "almost rectangle", where $\underline{OA}$ forms an "almost height" and $\underline{OA} = r$. Half of the $2n$ circular arcs such as AB form the upper "width", and the other half form the "lower width", so each of these "almost widths" is equal to half the perimeter of the circle – namely πr.

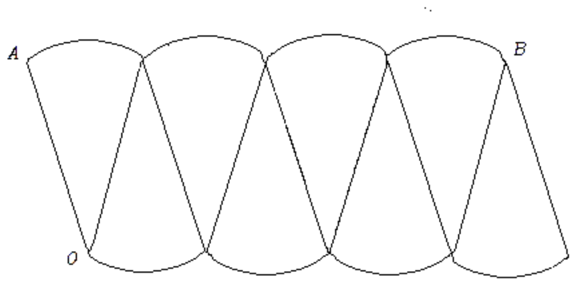


Figure 9: Rectification of a circle.

Hence the area of the rearranged shape is very close to

$$r \times \pi r = \pi r^2.$$

207.

(a) Cut along a generator, open up and lay the surface flat to obtain: a $2\pi r$ by h rectangle and two circular discs of radius r.
Hence the total surface area is

$$2\pi r^2 + 2\pi rh = \boldsymbol{2\pi r(r+h)}.$$

(b) The lateral surface consists of n rectangles, each with dimensions s_n by h (where s_n is the edge length of the regular n-gon), and hence has area $P_n h = 2\Pi_n rh$. Adding in the two end discs (each with area $\frac{1}{2}P_n r$) then gives total surface area $\boldsymbol{2\Pi_n r(r+h)}$.

208.

(a) (i) $\frac{1}{2}\pi r^2$; (ii) $\frac{1}{4}\pi r^2$; (iii) $\frac{\theta}{2}r^2$.

(b) The sector has two radii of total length $2r$. Hence the circular must have length $(\pi - 2)r$, and so subtends an angle $\pi - 2$ at the centre, so has area $\frac{\pi-2}{2}r^2$.

209.

(a) Focus first on the sloping surface. If we cut along a "generator" (a straight line segment joining the apex to a point on the circumference of the base), the surface opens up and lays flat to form a sector of a circle of radius l. The outside arc of this sector has length $2\pi r$, so the sector angle at the centre is equal to $\frac{r}{l} \cdot 2\pi$, and hence its area is $\frac{r}{l} \cdot \pi l^2 = \pi rl$.
Adding the area of the base gives the total surface area of the cone as

$$\pi r(r+l) = \boldsymbol{\frac{1}{2} \cdot 2\pi r(r+l)}.$$

(b) Let M be the midpoint of the edge $\underline{BC}$ and let l denote the 'slant height' $\underline{AM}$. Then the area of the n sloping faces is equal to $\frac{1}{2}P_n \cdot l$, while the area of the base is equal to $\frac{1}{2}P_n \cdot r$.

Hence the surface area is precisely $\boldsymbol{\frac{1}{2}P_n(r+l)}$.

210.

(a) If $\underline{AB}$ is an edge of the inscribed regular n-gon $ABCD\cdots$, and O is the circumcentre, then $\angle AOB = \frac{2\pi}{n}$, so $\underline{AB} = 2r\sin\frac{\pi}{n}$. Hence the required ratio is equal to $\frac{\sin\frac{\pi}{n}}{\frac{\pi}{n}}$, which tends to 1 as n tends to ∞.

(b) If $\underline{AB}$ is an edge of the circumscribed regular n-gon $ABCD\cdots$, and O is the circumcentre, then $\angle AOB = \frac{2\pi}{n}$, so $\underline{AB} = 2r\tan\frac{\pi}{n}$. Hence the required ratio is equal to $\frac{\tan\frac{\pi}{n}}{\frac{\pi}{n}}$, which tends to 1 as n tends to ∞.

211.

(a) If $\underline{AB}$ is an edge of the inscribed regular n-gon $ABCD\cdots$, and O is the circumcentre, then $\angle AOB = \frac{2\pi}{n}$, so area$(\triangle OAB) = \frac{1}{2}r^2\sin\frac{2\pi}{n}$. Hence the required ratio is equal to $\frac{\sin\frac{2\pi}{n}}{\frac{2\pi}{n}}$, which tends to 1 as n tends to ∞.

(b) If $\underline{AB}$ is an edge of the circumscribed regular n-gon $ABCD\cdots$, and O is the circumcentre, then $\angle AOB = \frac{2\pi}{n}$, so area$(\triangle OAB) = r^2\tan\frac{\pi}{n}$. Hence the required ratio is equal to $\frac{\tan\frac{\pi}{n}}{\frac{\pi}{n}}$, which tends to 1 as n tends to ∞.

212.

(a) area(P_2) = area(P_1) + area$(P_2 - P_1)$ > area(P_1).

(b) This general result is clearly related to the considerations of the previous section. But it is not clear whether we can really expect to prove it with the tools available. So it has been included partly in the hope that readers might come to appreciate the difficulties inherent in proving such an "obvious" result.

In the end, any attempt to prove it seems to underline the need to use "proof by induction" – for example, on the number of edges of the inner polygon. This method is not formally treated until Chapter 6, but is needed here.

- Suppose the inner polygon $P_1 = ABC$ has just $n = 3$ edges, and has perimeter p_1.

Draw the line through A parallel to BC, and let it meet the (boundary of the) polygon P_2 at the points U and V. The triangle inequality (Problem **146**(c)) guarantees that the length $\underline{UV}$ is less than or equal to the length of the compound path from U to V along the perimeter of the polygon P_2 (keeping on the opposite side of the line UV from B and C). So, if we cut off the part of P_2 on the side of the line UV opposite to B and C, we obtain a new outer convex polygon P_3, which contains P_1, and whose perimeter is no larger than that of P_2.

Now draw the line through B parallel to AC, and let it meet the boundary of P_3 at points W, X. If we cut off the part of P_3 on the side of the line WX opposite to A and C, we obtain a new outer convex polygon P_4, which contains P_1, and whose perimeter is no larger than that of P_3.

If we now do the same by drawing a line through C parallel to AB, and cut off the appropriate part of P_4, we obtain a final outer convex polygon P_5, which contains the polygon P_1, and whose perimeter is no larger than that of P_4 – and hence no larger than that of the original outer polygon P_2.

All three vertices A, B, C of P_1 now lie on the boundary of the outer polygon P_5, so the triangle inequality guarantees that AB is no larger than the length of the compound path along the boundary of P_5 from A to B (staying on the opposite side of the line AB from C). Similarly BC is no larger than the length of the compound path along the boundary of P_5 from B to C; and CA is no larger than the length of the compound path along the boundary of P_5 from C to A.

Hence the perimeter p_1 of the triangle P_1 is no larger than the perimeter of the outer polygon P_5, whose perimeter was no larger than the perimeter of the original outer polygon P_2. Hence the result holds when the inner polygon is a triangle.

- Now suppose that the result has been proved when the inner polygon is a k-gon, for some $k \geqslant 3$, and suppose we are presented with a pair of polygons P_1, P_2 where the inner polygon $P_1 = ABCD \cdots$ is a convex $(k+1)$-gon.

Draw the line m through C parallel to BD. Let this line meet the outer polygon P_2 at U and V. Cut off the part of P_2 on the opposite side of the the line UCV to B and D, leaving a new outer convex polygon P with perimeter no greater than that of P_2. We prove that the perimeter of polygon P_1 is less than that of polygon P – and hence less than that of P_2. Equivalently, we may assume that UCV is an edge of P_2.

Translate the line m parallel to itself, from m to BD and beyond, until it reaches a position of final contact with the polygon P_1, passing through the vertex X (and possibly a whole edge $\underline{XY}$) of the inner polygon P_1. Let this final contact line parallel to m be m'.

Since P_1 is convex and $k \geqslant 3$, we know that X is different from B and from D. As before, we may assume that m' is an edge of the outer polygon P_2. Cut both P_2 and P_1 along the line CX to obtain two smaller configurations, each of which consists of an inner convex polygon inside an outer convex polygon, but in which

– each of the inner polygons has at most k edges, and

– in each of the smaller configurations, the inner and outer polygons both share the edge $\underline{CX}$.

Then (by induction on the number of edges of the inner polygon) the perimeter of each inner polygon is no larger than the perimeter of the corresponding outer polygon; so for each inner polygon, the partial perimeter running from C to X (omitting the edge $\underline{CX}$) is no larger than the partial perimeter of the corresponding outer polygon running from C to X (omitting the edge $\underline{CX}$). So when we put the two parts back together again, we see that the perimeter of P_1 is no larger than the perimeter of P_2.

Hence the result holds when P_1 is a triangle; and if the result holds whenever the inner polygon has $k \geqslant 3$ edges, it also holds whenever the inner polygon has $(k+1)$ edges.

It follows that the result holds whatever the number of edges of the inner polygon may be. QED

213.

(a) Join $\underline{PQ}$. Then the lines $y = b$ and $x = d$ meet at R to form the right angled triangle PQR.

Pythagoras' Theorem then implies that $(d-a)^2 + (e-b)^2 = \underline{PQ}^2$.

(b) Join $\underline{PQ}$. The points $P = (a, b, c)$ and $R = (d, e, c)$ lie in the plane $z = c$. If we work exclusively in this plane, then part (a) shows that

$$\underline{PR}^2 = (d-a)^2 + (e-b)^2.$$

$\underline{QR} = |f - c|$, and $\triangle PRQ$ has a right angle at R. Hence

$$\underline{PQ}^2 = \underline{PR}^2 + \underline{RQ}^2 = (d-a)^2 + (e-b)^2 + (f-c)^2.$$

214.

(a) $\text{area}(\triangle ABC) = \frac{bc}{2}$, $\text{area}(\triangle ACD) = \frac{cd}{2}$, $\text{area}(\triangle ABD) = \frac{bd}{2}$.

(b) [This can be a long algebraic slog. And the answer can take very different looking forms depending on how one proceeds. Moreover, most of the resulting expressions are not very pretty, and are likely to incorporate errors.

One way to avoid this slog is to appeal to the fact that the modulus of the vector product $\boldsymbol{DB} \times \boldsymbol{DC}$ is equal to the area of the parallelogram spanned by $\boldsymbol{DB}$ and $\boldsymbol{DC}$ – and so is twice the area of $\triangle BCD$.]

(c) However part (b) is approached, it is in fact true that

$$\text{area}(\triangle BCD)^2 = \left(\frac{bc}{2}\right)^2 + \left(\frac{cd}{2}\right)^2 + \left(\frac{bd}{2}\right)^2,$$

so that

$$\text{area}(\triangle ABC)^2 + \text{area}(\triangle ACD)^2 + \text{area}(\triangle ABD)^2 = \text{area}(\triangle BCD)^2.$$

215.

(a) b and c are indeed lengths of arcs of great circles on the unit sphere: that is, arcs of circles of radius 1 (centred at the centre of the sphere). However, back in Chapter 1 we used the 'length' of such circular arcs to define the **angle** (in radians) subtended by the arc at the centre. So b and c are also *angles* (in radians).

(b) The only standard functions of angles are the familiar trig functions ($\sin, \cos$).

(c) If $c = 0$, then the output should specify that $a = b$, so c should have no effect on the output. This suggests that we might expect a formula that involves "adding $\sin c$" or "multiplying by $\cos c$".

Similarly when $b = 0$, the output should give $a = c$, so we might expect a formula that involved "adding $\sin b$" or "multiplying by $\cos b$".

In general, we should expect a formula in which b and c appear interchangeably (since the input pair (b, c) could equally well be replaced by the input pair (c, b) and should give the same output value of a).

(d)(i) If $\underline{BC}$ runs along the equator and $\angle B = \angle C = \frac{\pi}{2}$, then $\underline{BA}$ and $\underline{CA}$ run along circles of longitude, so A must be at the North pole. Since A is a right angle, it follows that $a = b = c = \frac{\pi}{2}$. (This tends to rule out the idea that the formula might involve "adding $\sin b$ and adding $\sin c$".)

(ii) Suppose that $\angle B = \frac{\pi}{2}$. Since we can imagine $\underline{AB}$ along the equator, and since there is a right angle at A, it follows that $\underline{AC}$ and $\underline{BC}$ both lie along circles of longitude, and so meet at the North pole. Hence C will be at the North pole, so $a = b = \frac{\pi}{2}$.

The inputs to any spherical version of Pythagoras' Theorem are then $b = \frac{\pi}{2}$, and c. And c is not constrained, so every possible input value of c must lead to the same output $a = \frac{\pi}{2}$. This tends to suggest that the formula involves some multiple of a product combining "$\cos b$" with some function of c. And since the inputs "b" and "c" must appear symmetrically, we might reasonably expect some multiple of "$\cos b \cdot \cos c$".

216.

(a) $\triangle OAB'$ has a right angle at A with $\angle AOB' = \angle AOB = c$. Hence $\underline{AB'} = \tan c$. Similarly $\underline{AC'} = \tan b$.
Hence $\underline{B'C'}^2 = \tan^2 b + \tan^2 c$.

(b) In $\triangle OAB'$ we see that $\underline{OB'} = \sec c$. Similarly $\underline{OC'} = \sec b$. We can now apply the Cosine Rule to $\triangle OB'C'$ to obtain:

$$\begin{aligned}\tan^2 b + \tan^2 c &= \sec^2 b + \sec^2 c - 2\sec b \cdot \sec c \cdot \cos(\angle B'OC') \\ &= \sec^2 b + \sec^2 c - 2\sec b \cdot \sec c \cdot \cos a.\end{aligned}$$

Hence $\cos a = \cos b \cdot \cos c$. QED

217. Construct the plane tangent to the sphere at A. Extend $\underline{OB}$ to meet this plane at B', and extend $\underline{OC}$ to meet the plane at C'.

$\triangle OAB'$ has a right angle at A with $\angle AOB' = \angle AOB = c$. Hence $\underline{AB'} = \tan c$. Similarly $\underline{AC'} = \tan b$.

Hence $\underline{B'C'}^2 = \tan^2 b + \tan^2 c - 2 \cdot \tan b \cdot \tan c \cdot \cos A$.

In $\triangle OAB'$ we see that $\underline{OB'} = \sec c$. Similarly $\underline{OC'} = \sec b$.

We can now apply the Cosine Rule to $\triangle OB'C'$ to obtain:

$$\begin{aligned}\tan^2 b + \tan^2 c - 2 \cdot \tan b \cdot \tan c \cdot \cos A \\ &= \sec^2 b + \sec^2 c - 2 \sec b \cdot \sec c \cdot \cos(\angle B'OC') \\ &= \sec^2 b + \sec^2 c - 2 \sec b \cdot \sec c \cdot \cos a.\end{aligned}$$

Hence $\cos a = \cos b \cdot \cos c + \sin b \cdot \sin c \cdot \cos A$. QED

218. We show that $\frac{\sin b}{\sin a} = \sin B$ (where B denotes the angle $\angle ABC$ at the vertex B).

Construct the plane $\boldsymbol{T}$ which is tangent to the sphere at B. Let O be the centre of the sphere; let $\underline{OA}$ produced meet the plane $\boldsymbol{T}$ at A'', and let $\underline{OC}$ produced meet the plane $\boldsymbol{T}$ at C''.

Imagine $\underline{BA}$ positioned along the equator; then $\underline{BA''}$ is horizontal; $\underline{AC}$ lies on a circle of longitude, so $A''C''$ is vertical. Hence $\angle C''BA'' = \angle B$, and $\angle BA''C''$ is a right angle; so $\sin B = \frac{\underline{A''C''}}{\underline{BC''}}$.

$\triangle OA''C''$ is right angled at A''; and $\angle A''OC'' = b$; so $\sin b = \frac{\underline{A''C''}}{\underline{OC''}}$.

$\triangle OBC''$ is right angled at B; and $\angle BOC'' = a$; so $\sin a = \frac{\underline{BC''}}{\underline{OC''}}$.

Hence $\frac{\sin b}{\sin a} = \sin B$ depends only on the angle at B, so $\frac{\sin b}{\sin a} = \frac{\sin b'}{\sin a'} = \sin B$.

219. We show that
$$\frac{\sin a}{\sin A} = \frac{\sin b}{\sin B}.$$
Position the triangle (or rather "rotate the sphere") so that $\underline{AB}$ runs along the equator, with $\underline{AC}$ leading into the northern hemisphere.

(i) If $\angle A$ is a right angle, then
$$\frac{\sin b}{\sin a} = \sin B = \frac{\sin B}{\sin A}$$
by Problem **218**. Hence
$$\frac{\sin a}{\sin A} = \frac{\sin b}{\sin B}.$$
The same is true if $\angle B$ is a right angle.

(ii) If $\angle A$ and $\angle B$ are both less than a right angle, one can draw the circle of longitude from C to some point X on $\underline{AB}$. One can then apply Problem **218** to the two triangles $\triangle CXA$ and $\triangle CXB$ (each with a right angle at X). Let x denote the length of the $\underline{CX}$. Then $\frac{\sin x}{\sin b} = \sin A$, and $\frac{\sin x}{\sin a} = \sin B$.

Hence $\sin b \cdot \sin A = \sin x = \sin a \cdot \sin B$, so

$$\frac{\sin a}{\sin A} = \frac{\sin b}{\sin B}.$$

(iii) If $\angle A$ (say) is greater than a right angle, the circle of longitude from C meets the line BA extended beyond A at a point X (say). If we let $\underline{CX} = x$, then one can argue similarly using the triangles $\triangle CXA$ and $\triangle CXB$ to get $\frac{\sin x}{\sin a} = \sin B$, and $\frac{\sin x}{\sin b} = \sin A$, whence

$$\frac{\sin a}{\sin A} = \frac{\sin b}{\sin B}.$$

220. Let $X = (x, y)$ be an arbitrary point of the locus.

(i) Then the distance from X to m is equal to y, so setting this equal to $\underline{XF}$ gives the equation:

$$y^2 = x^2 + (y - 2a)^2,$$

or

$$x^2 = 4a(y - a).$$

(ii) If we change coordinates and choose the line $y = a$ as a new x-axis, then the equation becomes $x^2 = 4aY$. The curve is then tangent to the new x-axis at the (new) origin, and is symmetrical about the y-axis.

221.

(a) Choose the line AB as x-axis, and the perpendicular bisector of $\underline{AB}$ as the y-axis. Then $A = (-3, 0)$ and $B = (3, 0)$. The point $X = (x, y)$ is a point of the unknown locus precisely when

$$(x+3)^2 + y^2 = \underline{XA}^2 = (2 \cdot \underline{XB})^2 = 2^2((x-3)^2 + y^2)$$

that is, when

$$(x-5)^2 + y^2 = 4^2.$$

This is the equation of a circle with centre $(5, 0)$ and radius $r = 4$.

(b) Choose the line AB as x-axis, and the perpendicular bisector of $\underline{AB}$ as the y-axis.
If $f = 1$, the locus is the perpendicular bisector of $\underline{AB}$.
We may assume that $f > 1$ (since if $f < 1$, then $\underline{BX} : \underline{AX} = f^{-1} : 1$ and $f^{-1} > 1$, so we may simply swap the labelling of A and B).
Now $A = (-b, 0)$ and $B = (b, 0)$, and the point $X = (x, y)$ is a point of the unknown locus precisely when

$$(x+b)^2 + y^2 = \underline{XA}^2 = (f \cdot \underline{XB})^2 = f^2\left[(x-b)^2 + y^2\right]$$

that is, when

$$x^2(f^2-1) - 2bx(f^2+1) + y^2(f^2-1) + b^2(f^2-1) = 0$$

$$\left(x - \frac{b(f^2+1)}{f^2-1}\right)^2 + y^2 = \left(\frac{2fb}{f^2-1}\right)^2.$$

This is the equation of a circle with centre $\left(\frac{b(f^2+1)}{f^2-1}, 0\right)$ and radius $r = \frac{2fb}{f^2-1}$.

222.

(a) Choose the line AB as x-axis, and the perpendicular bisector of $\underline{AB}$ as the y-axis.
Then $A = (-c, 0)$ and $B = (c, 0)$. The point $X = (x, y)$ is a point of the unknown locus precisely when

$$2a = \underline{AX} + \underline{BX} = \sqrt{(x+c)^2 + y^2} + \sqrt{(x-c)^2 + y^2},$$

that is, when

$$\begin{aligned} 2a - \sqrt{(x+c)^2+y^2} &= \sqrt{(x-c)^2+y^2}. \\ \therefore\ 4a^2 - 4a\sqrt{(x+c)^2+y^2} + [(x+c)^2+y^2] &= [(x-c)^2+y^2] \\ \therefore\quad a^2 + cx &= a\sqrt{(x+c)^2+y^2} \\ \therefore\quad (a^2-c^2)x^2 + a^2y^2 &= a^2(a^2-c^2) \end{aligned}$$

Setting $\frac{c}{a} = e$ then yields the equation for the locus in the form:

$$\frac{x^2}{a^2} + \frac{y^2}{a^2(1-e^2)} = 1.$$

Note: In the derivation of the equation we squared both sides (twice). This may introduce spurious solutions. So we should check that every solution (x, y) of the final equation satisfies the original condition.

(b) The real number $e < 1$ is given, so we may set the distance from F to m be $\frac{a}{e}\left(1-e^2\right)$. Choose the line through F and perpendicular to m as x-axis. To start with, we choose the line m as y-axis and adjust later if necessary.
Hence F has coordinates $\left(\frac{a}{e}\left(1-e^2\right), 0\right)$, and the point $X = (x, y)$ is a point of the unknown locus precisely when

$$\left(x - \frac{a}{e}\left(1-e^2\right)\right)^2 + y^2 = (ex)^2,$$

which can be rearranged as

$$\left(1-e^2\right)x^2 - 2\frac{a}{e}(1-e^2)x + \left(\frac{a}{e}\right)^2(1-e^2)^2 + y^2 = 0,$$

and further as

$$\begin{aligned}(1-e^2)\left[x^2-2\frac{a}{e}x+\left(\frac{a}{e}\right)^2\right]+y^2 &= \left(\frac{a}{e}\right)^2(1-e^2)-\left(\frac{a}{e}\right)^2(1-e^2)^2\\ &= \left(\frac{a}{e}\right)^2(e^2-e^4)\\ &= a^2(1-e^2)\\ \therefore \left(x-\frac{a}{e}\right)^2+\frac{y^2}{1-e^2} &= a^2.\end{aligned}$$

If we now move the y-axis to the line $x = \frac{a}{e}$ the equation takes the simpler form:

$$\frac{x^2}{a^2}+\frac{y^2}{a^2(1-e^2)}=1.$$

(c) This was done in the derivations in the solutions to parts (a) and (b).

223.

(a) The triangle inequality shows that, if $\underline{AX} > \underline{BX}$, then $\underline{AB}+\underline{BX} \geqslant \underline{AX}$; hence the locus is non-empty only when $a \leqslant c$. If $a = c$, then X must lie on the line AB, but not between A and B, so the locus consists of the two half-lines on AB outside $\underline{AB}$. Hence we may assume that $a < c$.

Choose the line AB as x-axis, and the perpendicular bisector of $\underline{AB}$ as y-axis. Then $A = (-c, 0)$ and $B = (c, 0)$. The point $X = (x, y)$ is a point of the unknown locus precisely when

$$2a = |\underline{AX}-\underline{BX}| = \left|\sqrt{(x+c)^2+y^2}-\sqrt{(x-c)^2+y^2}\right|.$$

If $\underline{AX} > \underline{BX}$, we can drop the modulus signs and calculate as in Problem **222**.

$$\begin{aligned}2a+\sqrt{(x-c)^2+y^2} &= \sqrt{(x+c)^2+y^2}.\\ \therefore\ 4a^2+4a\sqrt{(x-c)^2+y^2}+(x-c)^2+y^2 &= (x+c)^2+y^2\\ \therefore\qquad a^2-cx &= -a\sqrt{(x-c)^2+y^2}\\ \therefore\qquad (c^2-a^2)x^2-a^2y^2 &= a^2(c^2-a^2)\end{aligned}$$

Setting $\frac{c}{a} = e\ (> 1)$, then yields the equation for the locus in the form:

$$\frac{x^2}{a^2}-\frac{y^2}{a^2(e^2-1)}=1.$$

Note: In the derivation of the equation we squared both sides (twice). This may introduce spurious solutions. So we should check that every solution (x, y) of the final equation satisfies the original condition.In fact, the squaring process introducesadditional solutions in the form of a second branch of the locus, corresponding precisely to points X where $\underline{AX} < \underline{BX}$.

(b) The real number $e > 1$ is given, so we may set the distance from F to m be $\frac{a}{e}(e^2 - 1)$. Choose the line through F and perpendicular to m as x-axis. To start with, we choose the line m as y-axis and adjust later if necessary.

Hence F has coordinates $\left(\frac{a}{e}(e^2 - 1), 0\right)$, and the point $X = (x, y)$ is a point of the unknown locus precisely when

$$\begin{aligned}
\left(x - \frac{a}{e}(e^2 - 1)\right)^2 + y^2 &= (ex)^2 \\
\therefore \quad (e^2 - 1)x^2 + \frac{2a}{e}(e^2 - 1)x - y^2 &= \left(\frac{a}{e}\right)^2 (e^2 - 1)^2 \\
\therefore \quad (e^2 - 1)\left[x^2 + \frac{2a}{e}x + \left(\frac{a}{e}\right)^2\right] - y^2 &= \left(\frac{a}{e}\right)^2 (e^2 - 1)^2 + \left(\frac{a}{e}\right)^2 (e^2 - 1) \\
&= \left(\frac{a}{e}\right)^2 (e^4 - e^2) \\
\therefore \quad (e^2 - 1)\left[x^2 + \frac{2a}{e}x + \left(\frac{a}{e}\right)^2\right] - y^2 &= a^2(e^2 - 1) \\
\therefore \quad \left(x + \frac{a}{e}\right)^2 - \frac{y^2}{e^2 - 1} &= a^2
\end{aligned}$$

If we now move the y-axis to the line $x = -\frac{a}{e}$ the equation takes the simpler form:

$$\frac{x^2}{a^2} - \frac{y^2}{a^2(e^2 - 1)} = 1.$$

(c) This was done in derivations in the solutions to parts (a) and (b).

224.

(a) When $z = k$ is a constant, the equation reduces to that of a circle

$$x^2 + y^2 = (rk)^2$$

in the plane $z = k$. When the cutting plane is the xy-plane "$z = 0$", the circle has radius 0, so is a single point.

(b) (i) A vertical plane through the apex cuts the cone in a pair of generators crossing at the apex.

(ii) If the cutting plane through the apex is less steep than a generator, then it cuts the cone only at the apex.

If the cutting plane through the apex is parallel to a generator, then it cuts the cone in a generator – a single line (the next paragraph indicates that this line may be better thought of as a pair of "coincident" lines).

What happens when the cutting plane through the apex is steeper than a generator may not be intuitively clear. One way to make sense of this is to treat the cross-section as the set of solutions of two simultaneous equations –

one for the cone, and the other for the plane (say $y = nz$, with $n < r$). This leads to the equation

$$x^2 = (r^2 - n^2)z^2, \quad y = nz,$$

with solution set

$$x = \pm z\sqrt{r^2 - n^2}, \quad xy = nz$$

which specifies a pair of lines crossing at the apex.

A slightly easier way to visualize the cross-section in this case is to let the apex of the double cone be A, and to let X be any other point of the cross-section. Then the line AX is a generator of the cone, so lies on the cone's surface. But A and X also lie in the cutting plane – so the whole line AX must lie in the cutting plane. Hence the cross-section contains the whole line AX.

(c) (i) If the cutting plane passes through the apex and is parallel to a generator of the cone, then we saw in (b) that the cross-section is simply the generator itself.

(ii) Thus we assume that the cutting plane does not pass through the apex, and may assume that it cuts the bottom half of the cone. If V is the point nearest the apex where the cutting plane meets the cone, then the cross-section curve starts at V and becomes wider as we go down the cone. Because the plane is parallel to a generator, the plane never cuts the "other side" of the bottom half of the cone, so the cross section never "closes up" – but continues to open up wider and wider as we go further and further down the bottom half of the cone.

Let $\boldsymbol{S}$ be the sphere, which is inscribed in the cone above the cutting plane, and which is tangent to the cutting plane at F. Let $\boldsymbol{C}$ be the circle of contact between $\boldsymbol{S}$ and the cone. Let A be the apex of the cone, and let the apex angle of the cone be equal to 2θ. Let X be an arbitrary point of the cross-section.

To illustrate the general method, consider first the special case where $X = V$ is the "highest" point of the cross-section. The line segment $\underline{VA}$ is tangent to the sphere $\boldsymbol{S}$, so crosses the circle $\boldsymbol{C}$ at some point M. Now $\underline{VF}$ lies in the cutting plane, so is also tangent to the sphere $\boldsymbol{S}$ at the point F. Any two tangents to a sphere from the same exterior point are equal, so it follows that $\underline{VM} = \underline{VF}$. Moreover, $\underline{VM}$ is exactly equal to the distance from V to the line m (since

- $*$ firstly the line m lies in the horizontal plane through $\boldsymbol{C}$ and so is on the same horizontal level as M, and
- $*$ secondly the shortest line VM^* from V to m, runs straight up the cutting plane, which is parallel to a generator – so the angle $\angle MVM^* = 2\theta$, whence $\underline{VM^*} = \underline{VM} = \underline{VF}$).

Now let X be an arbitrary point of the cross-sectional curve, and use a similar argument. First the line XA is always a generator of the cone, so is tangent to the sphere $\boldsymbol{S}$, and crosses the circle $\boldsymbol{C}$ at some point Y. Moreover, XF is also tangent to the sphere. Hence $\underline{XY} = \underline{XF}$. It remains to see that $\underline{XY}$ is equal to the distance $\underline{XY^*}$ from X to the closest point Y^* on the line m – since

∗ firstly the two points Y and Y^* both lie on the same horizontal level (namely the horizontal plane through the circle $\boldsymbol{C}$), and

∗ secondly both make the same angle θ with the vertical.

Hence the cross-sectional curve is a parabola with focus F and directrix m.

(d) (i) If the cutting plane is less steep than a generator, the cross-section is a closed curve. If V and W are the highest and lowest points of intersection of the plane with the cone, then the cross-section is clearly symmetrical under reflection in the line VW. Intuitively it is tempting to think that the lower end near W should be 'fatter' than the upper part of the curve (giving an egg-shaped cross-section). This turns out to be false, and the correct version was known to the ancient Greeks, though the error was repeated in many careful drawings from the 14th and 15th centuries (e.g. Albrecht Dürer (1471–1528)).

(ii) The derivation is very similar to that in part (c), and we leave the reader to reconstruct it.

An alternative approach is to insert a second sphere $\boldsymbol{S}'$ below the cutting plane, and inflate it until it makes contact with the cone along a circle $\boldsymbol{C}'$ while at the same time touching the cutting plane at a point F'. If X is an arbitrary point of the cross-sectional curve, then XA is tangent to both spheres, so meets the circle $\boldsymbol{C}$ at some point Y and meets the circle $\boldsymbol{C}'$ at some point Y'. Then Y, X, Y' are collinear. Moreover, $\underline{XY} = \underline{XF}$ (since both are tangents to the sphere $\boldsymbol{S}$ from the point X), and $\underline{XY'} = \underline{XF'}$ (since both are tangents to the sphere $\boldsymbol{S}'$ from the point X), so

$$\underline{XF} + \underline{XF'} = \underline{XY} + \underline{XY'} = \underline{YY'}$$

But $\underline{YY'}$ is equal to the slant height of the cone between the two fixed circles $\boldsymbol{C}$ and $\boldsymbol{C}'$, and so is equal to a constant k. Hence, the focus-focus specification in Problem **222** shows that the cross-section is an ellipse.

(e) (i) If the cutting plane is steeper than a generator, the cross-section cuts both the bottom half and the top half of the cone to give two separate parts of the cross-section. Neither part "closes up", so each part opens up more and more widely.

If V is the highest point of the cross-section on the lower half of the cone, and W is the lowest point of the cross-section on the upper half of the cone, then it seems clear that the cross-section is symmetrical under reflection in the line VW. But it is quite unclear that the two halves of the cross-section are exactly congruent (though again it was known to the ancient Greeks).

(ii) The formal derivation is very similar to that in part (c), and we leave the reader to reconstruct it.

An alternative approach is to copy the alternative in (d), and to insert a second sphere $\boldsymbol{S}'$ in the upper part of the cone, on the same side of the cutting plane as the apex, inflate it until it makes contact with the cone along a circle $\boldsymbol{C}'$ while at the same time touching the cutting plane at a point F'. If X is an arbitrary point of the cross-sectional curve, then XA is tangent to both spheres, so meets the circle $\boldsymbol{C}$ at some point Y and meets the circle

$\boldsymbol{C}'$ at some point Y'. Then Y, X, Y' are collinear. Moreover, X, F, and F' all lie on the cutting plane. Now $\underline{XY} = \underline{XF}$ (since both are tangents to the sphere $\boldsymbol{S}$ from the point X), and $\underline{XY'} = \underline{XF'}$ (since both are tangents to the sphere $\boldsymbol{S}'$ from the point X). If X is on the upper half of the cone, then

$$\underline{XF} - \underline{XF'} = \underline{XY} - \underline{XY'} = \underline{YY'}.$$

But $\underline{YY'}$ is equal to the slant height of the cone between the two fixed circles $\boldsymbol{C}$ and $\boldsymbol{C}'$, and so is equal to a constant k. Hence, the focus-focus specification in Problem **223** shows that the cross-section is a hyperbola.

225.

(a) (i) 2^1; (ii) $1 = 2^0$

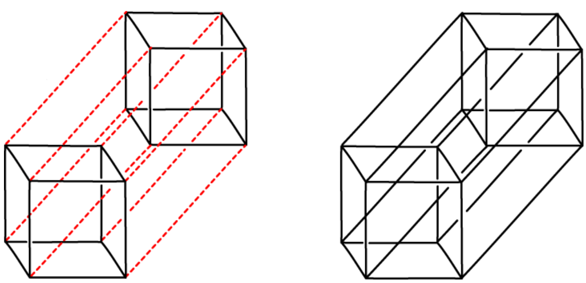
(b) (i) 2^2; (ii) $1 = 2^0$; (iii) $4 = 2 \times 2^0 + 2^1$

(c) (i) 2^3; (ii) $1 = 2^0$; (iii) $12 = 2 \times 4 + 2^2$; (iv) $6 = 2 \times 2^0 + 4$

(d) (i) 2^4; (ii) $1 = 2^0$; (iii) $32 = 2 \times 12 + 2^3$; (iv) $24 = 2 \times 6 + 12$; (v) $8 = 2 \times 2^0 + 6$

226.

(c) (i) If you look carefully at the diagram shown here you should be able to see not only the upper and lower 3D-cubes, but also the four other 3D-cubes formed by joining each 2D-cube in the upper 3D-cube to the corresponding 2D-cube in the lower 3D-cube.



Note: Once we have the 3D-cube expressed in coordinates, we can specify precisely which planes produce which cross-sections in Problem **38**. The plane $x+y+z=1$ passes through the three neighbours of $(0,0,0)$ and creates an equilateral triangular cross-section. Any plane of the form $z=c$ (where c is a constant between 0 and 1) produces a square cross-section. And the plane $x+y+z=\frac{3}{2}$ is the perpendicular bisector of the line joining $(0,0,0)$ to $(1,1,1)$, and creates a regular hexagon as cross-section.

227.

(a) View the coordinates as (x,y,z). Start at the origin $(0,0,0)$ and travel round 3 edges of the lower 2D-cube "$z=0$" to the point $(0,1,0)$. Copy this path of length 3 on the upper 2D-cube "$z = 1$" (from $(0,0,1)$ to $(0,1,1)$. Then join $(0,0,0)$ to $(0,0,1)$ and join $(0,1,0)$ to $(0,1,1)$. The result

$(0,0,0)$, $(1,0,0)$, $(1,1,0)$, $(0,1,0)$, $(0,1,1)$, $(1,1,1)$, $(1,0,1)$, $(0,0,1)$ (and back to $(0,0,0)$)

has the property that exactly one coordinate changes when we move from each vertex to the next. This is an example of a *Gray code of length* 3.

Note: How many such paths/circuits are there in the 3D-cube? We can certainly count those of the kind described here. Each such circuit has a "direction": the 12 edges of the 3D-cube lie in one of 3 "directions", and each such circuit contains all four edges in one of these 3 directions. Moreover this set of four edges can be completed to a circuit in exactly 2 ways. So there are 3×2 such circuits. In 3D this accounts for all such circuits. But in higher dimensions the numbers begin to explode (in the 4D-cube there are 1344 such circuits).

(b) View the coordinates as (w, x, y, z). Start at the origin $(0,0,0,0)$ and travel round the 8 vertices of the lower 3D-cube "$z = 0$" to the point $(0,0,1,0)$. Then copy this path on the upper 3D-cube "$z = 1$" from $(0,0,0,1)$ to $(0,0,1,1)$. Finally join $(0,0,0,0)$ to $(0,0,0,1)$ and join $(0,0,1,0)$ to $(0,0,1,1)$. The result

$(0,0,0.0)$, $(1,0,0,0)$, $(1,1,0,0)$, $(0,1,0,0)$, $(0,1,1,0)$, $(1,1,1,0)$, $(1,0,1,0)$, $(0,0,1,0)$ $(0,0,1,1)$, $(1,0,1,1)$, $(1,1,1,1)$, $(0,1,1,1)$, $(0,1,0,1)$, $(1,1,0,1)$, $(1,0,0,1)$, $(0,0,0,1)$ (and back to $(0,0,0,0)$)

has the property that exactly one coordinate changes when we move from each vertex to the next. This is an example of a *Gray code of length* 4.

Note: The general construction in dimension $n + 1$ depends on the previous construction in dimension n, so makes use of *mathematical induction* (see Problem **262** in Chapter 6).

VI. Infinity: recursion, induction, infinite descent

Mathematical induction
– i.e. proof by recurrence –
is ... imposed on us,
because it is ... the affirmation
of a property of the mind itself.
Henri Poincaré (1854–1912)

Allez en avant, et la foi vous viendra.
(Press on ahead, and understanding will follow.)
Jean le Rond d'Alembert (1717–1783)

Mathematics has been called "the science of infinity". However, infinity is a slippery notion, and many of the techniques which are characteristic of modern mathematics were developed precisely to tame this slipperiness. This chapter introduces some of the relevant ideas and techniques.

There are aspects of the story of infinity in mathematics which we shall not address. For example, astronomers who study the night sky and the movements of the planets and stars soon note its immensity, and its apparently 'fractal' nature – where increasing the detail or magnification reveals more or less the same level of complexity on different scales. And it is hard then to avoid the question of whether the stellar universe is finite or infinite.

In the *mental* universe of mathematics, once counting, and the process of halving, become routinely iterative processes, questions about infinity and infinitesimals are almost inevitable. However, mathematics recognises the conceptual gulf between the finite and the infinite (or infinitesimal), and rejects the lazy use of "infinity" as a metaphor for what is simply "very large". Large finite numbers are still numbers; and long finite sums are conceptually quite different from sums that "go on for ever". Indeed, in the third century BC, Archimedes (c. 287–212 BC) wrote a small booklet called *The Sand Reckoner*, dedicated to King Gelon, in which he introduced the arithmetic of powers (even though the ancient Greeks had no convenient notation for writing such numbers), in order to demonstrate that – contrary to what some people had claimed – the number of grains of sand in the known universe must be finite (he derived an upper bound of approximately 8×10^{63}).

The influence wielded by ideas of infinity on mathematics has been profound, even if we now view some of these ideas as flights of fancy –

- from Zeno of Elea (c. 495 BC – c. 430 BC), who presented his paradoxes to highlight the dangers inherent in reasoning sloppily with infinity,
- through Giordano Bruno (1548–1600), who declared that there were infinitely many inhabited universes, and who was burned at the stake when he refused to retract this and other "heresies",
- to Georg Cantor (1845–1918) whose groundbreaking work in developing a true "mathematics of infinity" was inextricably linked to his religious beliefs.

In contrast, we focus here on the delights of the mathematics, and in particular on how an initial doorway into "ideas of infinity" can be forged from careful reasoning with finite entities. Readers who would like to explore what we pass over in silence could do worse than to start with the essay on "infinity" in the MacTutor History of Mathematics archive:

http://www-history.mcs.st-and.ac.uk/HistTopics/Infinity.html.

The simplest infinite processes begin with *recursion* – a process where we repeat exactly the same operation over and over again (in principle, continuing for ever). For example, we may start with 0, and repeat the operation "add 1", to generate the sequence:

$$0, 1, 2, 3, 4, 5, 6, 7, \ldots.$$

Or we may start with $2^0 = 1$ and repeat the operation "multiply by 2", to generate:

$$1, 2, 4, 8, 16, 32, 64, 128, \ldots.$$

Or we may start with $1.000000\cdots$, and repeat the steps involved in "dividing by 7" to generate the infinite decimal for $\frac{1}{7}$:

$$\frac{1}{7} = 0.142857142857142857\cdots.$$

We can then vary this idea of "recursion" by allowing each operation to be "essentially" (rather than exactly) the same, as when we define *triangular* numbers by "adding n" at the n^{th} stage to generate the sequence:

$$0, 1, 3, 6, 10, 15, 21, 28, \ldots.$$

In other words, the sequence of *triangular numbers* is defined by a *recurrence relation*:

> $T_0 = 0$; and
> when $n \geqslant 1$, $T_n = T_{n-1} + n$.

We can vary this idea further by allowing more complicated recurrence relations – such as that which defines the *Fibonacci numbers*:

> $F_0 = 0$, $F_1 = 1$; and
> when $n \geqslant 1$, $F_{n+1} = F_n + F_{n-1}$.

All of these "images of infinity" hark back to the familiar counting numbers.

- We know how the counting numbers begin (with 0, or with 1); and
- we know that we can "add 1" over and over again to get ever larger counting numbers.

The intuition that this process is, in principle, *never-ending* (so is never actually completed), yet somehow manages to count all positive integers, is what Poincaré called a "property of the mind itself": that is, the idea that we can define an **infinite** sequence, or process, or chain of deductions (involving digits, or numbers, or objects, or statements, or truths) by

- specifying how it *begins*, and by then
- specifying in a uniform way "how to construct *the next term*", or "how to perform *the next step*".

This idea is what lies behind "**proof** by mathematical induction", where we prove that some assertion $\mathbf{P}(n)$ holds **for all** $\boldsymbol{n} \geqslant \mathbf{1}$ – so demonstrating infinitely many separate statements at a single blow. The validity of this method of proof depends on a fundamental property of the *positive integers*, or of the counting sequence

$$\text{“}1, 2, 3, 4, 5, \ldots\text{”},$$

namely:

> **The Principle of Mathematical Induction**: If a subset S of the positive integers
>
> - contains the integer "1",
>
> and has the property that
>
> - whenever an integer k is in the set S, then the next integer $k+1$ is always in S too,
>
> then S contains **all** positive integers.

6.1. Proof by mathematical induction I

When students first meet "proof by induction", it is often explained in a way that leaves them feeling distinctly uneasy, because it appears to break the fundamental taboo:

> never assume what you are trying to prove.

This tends to leave the beginner in the position described by d'Alembert's quote at the start of the chapter: they may "press on" in the hope that "understanding will follow", but a doubt often remains. So we encourage readers who have already met proof by induction to take a step back, and to try to understand afresh how it really works. This may require you to study the solutions (Section 6.10), and to be prepared to learn to write out proofs more carefully than, and rather differently from, what you are used to.

When we wish to prove a general result which involves a parameter n, where n can be **any positive integer**, we are really trying to prove *infinitely many results all at once*. If we tried to prove such a collection of results in turn, "one at a time", not only would we never finish, we would *scarcely get started* (since completing the first million, or billion, cases leaves just as much undone as at the start). So our only hope is:

- to think of the sequence of assertions in a uniform way, as consisting of infinitely many different, but similar-looking, statements $\mathbf{P}(n)$, one for each n separately (with each statement depending on a particular n); and
- to recognise that the **overall** result to be proved is not just a single statement $\mathbf{P}(n)$, but the *compound statement* that "$\mathbf{P}(n)$ is true, **for all** $\boldsymbol{n \geqslant 1}$".

Once the result to be proved has been formulated in this way, we can

- use bare hands to check that the very first statement is true (usually $\mathbf{P}(1)$); and
- try to find some way of demonstrating that,
  - as soon as we know that "$\mathbf{P}(k)$ is true, for some (particular, but unspecified) $k \geqslant 1$",
  - we can prove in a uniform way that the next result $\mathbf{P}(k+1)$ is then automatically true.

Having implemented the first of the two induction steps, we know that $\mathbf{P}(1)$ is true.

The second bullet point above then comes into play and assures us that (since we know that $\mathbf{P}(1)$ is true), $\mathbf{P}(2)$ must be true.

And once we know that $\mathbf{P}(2)$ is true, the second bullet point assures us that $\mathbf{P}(3)$ is also true.

And once we know that $\mathbf{P}(3)$ is true, the second bullet point assures us that $\mathbf{P}(4)$ is also true.

And so on *for ever*.

We can then conclude that the *whole sequence* of infinitely many statements are all true – namely that:

> "**every** statement $\mathbf{P}(n)$ is true",

or that

> "$\mathbf{P}(n)$ is true, **for all** $\boldsymbol{n} \geqslant \mathbf{1}$."

In other words, if we define S to be the set of positive integers n for which the statement $\mathbf{P}(n)$ is true, then S contains the element "1", and whenever k is in S, so is $k+1$; hence, by the *Principle of Mathematical Induction* we know that S contains all positive integers.

At this stage we should acknowledge an important didactical (rather than mathematical) ploy in our recommended layout here. It is important to underline the distinction between

(i) the *individual* statements $\mathbf{P}(n)$ which are the separate ingredients in the overall statement to be proved, namely:

> "$\mathbf{P}(\boldsymbol{n})$ is true, **for all** $\boldsymbol{n} \geqslant \mathbf{1}$",

where infinitely many individual statements have been compressed into a single compound statement, and

(ii) the *induction step*, where we

- assume some *particular* $\mathbf{P}(k)$ is known to be true, and
- show that the next statement $\mathbf{P}(k+1)$ is then automatically true.

To underline this distinction we consistently use a different "dummy variable" (namely "k") in the latter case. This distinction is a psychological ploy rather than a logical necessity. However, we recommend that readers should imitate this distinction (at least initially).

6.2. 'Mathematical induction' and 'scientific induction'

The idea of a "list that goes on for ever" arose in the sequence of powers of 4 back in Problem **16**, where we asked

Do the two sequences arising from successive powers of 4:

- the *leading* digits:

$$4, 2, 6, 2, 1, 4, 2, 6, 2, 1, 4, \ldots,$$

  and

- the *units* digits:

$$4, 6, 4, 6, 4, 6, 4, 6, \ldots,$$

really "repeat for ever" as they seem to?

This example illustrates the most basic misconception that sometimes arises concerning *mathematical induction* – namely to confuse it with the kind of pattern spotting that is often called 'scientific induction'.

In science (as in everyday life), we routinely infer that something that is observed to occur repeatedly, apparently without exception (such as the sun rising every morning; or the Pole star seeming to be fixed in the night sky) may be taken as a "fact". This kind of "scientific induction" makes perfect sense when trying to understand the world around us – even though the inference is not warranted in a strictly logical sense.

Proof by mathematical induction is quite different. Admittedly, it often requires intelligent guesswork at a preliminary stage to make a guess that allows us to formulate precisely what it is that we should be trying to prove. But this initial guess is separate from the proof, which remains a strictly deductive construction. For example,

> the fact that "1", "1 + 3", "1 + 3 + 5", "1 + 3 + 5 + 7", etc. all appear to be successive squares gives us an idea that perhaps the identity
>
> $\mathbf{P}(n)$: $1 + 3 + 5 + \cdots + (2n - 1) = n^2$
>
> is true, **for all $\boldsymbol{n} \geqslant \mathbf{1}$**.

This guess is needed before we can start the proof by mathematical induction. **But the process of guessing is not part of the proof.** And until we construct the "proof by induction" (Problem **231**), we cannot be sure that our guess is correct.

The danger of confusing 'mathematical induction' and 'scientific induction' may be highlighted to some extent if we consider the *proof* in Problem **76** above that "we can always construct ever larger prime numbers", and contrast it with an observation (see Problem **228** below) that is often used in its place – even by authors who should know better.

In Problem **76** we gave a strict construction by *mathematical induction*:

- we first showed how to begin (with $p_1 = 2$ say);
- then we showed how, given any finite list of distinct prime numbers $p_1, p_2, p_3, \ldots, p_k$, it is always possible to construct a **new** prime p_{k+1} (as the *smallest* prime number dividing $N_k = p_1 \times p_2 \times p_3 \times \cdots \times p_k + 1$).

This construction was very carefully worded, so as to be logically correct.

In contrast, one often finds lessons, books and websites that present the essential idea in the above proof, but "simplify" it into a form that encourages anti-mathematical "pattern-spotting" which is all-too-easily misconstrued. For example, some books present the sequence

$$(2;)\ 2+1 = \mathbf{3};\ 2\times 3+1 = \mathbf{7};\ 2\times 3\times 5+1 = \mathbf{31};\ 2\times 3\times 5\times 7+1 = \mathbf{211};\ldots$$

as a way of generating more and more primes.

Problem 228

(a) Are 3, 7, 31, 211 all prime?

(b) Is $2 \times 3 \times 5 \times 7 \times 11 + 1$ prime?

(c) Is $2 \times 3 \times 5 \times 7 \times 11 \times 13 + 1$ prime? △

We have already met two excellent historical examples of the dangers of plausible pattern-spotting in connection with Problem **118**. There you proved that:

> "if $2^n - 1$ is prime, then n must be prime."

You then showed that $2^2 - 1$, $2^3 - 1$, $2^5 - 1$, $2^7 - 1$ are all prime, but that $2^{11} - 1 = 2047 = 23 \times 89$ is **not**. This underlines the need to avoid jumping to (possibly false) conclusions, and never to confuse a statement with its converse.

In the same problem you showed that:

> "if $a^b + 1$ is to be prime and $a \neq 1$, then a must be even, and b must be a power of 2."

You then looked at the simplest family of such candidate primes namely the sequence of Fermat numbers f_n:

$$f_0 = 2^1 + 1 = 3, f_1 = 2^2 + 1 = 5, f_2 = 2^4 + 1 = 17, f_3 = 2^8 + 1 = 257, f_4 = 2^{16} + 1.$$

It turned out that, although f_0, f_1, f_2, f_3, f_4 are all prime, and although Fermat (1601–1665) claimed (in a letter to Marin Mersenne (1588–1648)) that **all** Fermat numbers f_n are prime, we have yet to discover a sixth Fermat prime!

There are times when a mathematician may appear to guess a general result on the basis of what looks like very modest evidence (such as noticing that it appears to be true in a few small cases). Such "informed guesses" are almost always rooted in other experience, or in some unnoticed feature of the particular situation, or in some striking analogy: that is, an apparent pattern strikes a chord for reasons that go way beyond the mere numbers. However those with less experience need to realise that apparent patterns or trends are often no more than numerical accidents.

Pell's equation (John Pell (1611–1685)) provides some dramatic examples.

- If we evaluate the expression "$n^2 + 1$" for $n = 1, 2, 3, \ldots$, we may notice that the outputs $2, 5, 10, 17, 26, \ldots$ never give a perfect square. And this is to be expected, since the next square after n^2 is
$$(n+1)^2 = n^2 + 2n + 1,$$
and this is always greater than $n^2 + 1$.
- However, if we evaluate "$991n^2+1$" for $n = 1, 2, 3, \ldots$, we may observe that the outputs never seem to include a perfect square. But this time there is no obvious reason why this should be so – so we may anticipate that this is simply an accident of "small" numbers. And we should hesitate to change our view, even though this accident goes on happening for a very, very, very long time: the smallest value of n for which $991n^2 + 1$ gives rise to a perfect square is apparently
$$n = 12\,055\,735\,790\,331\,359\,447\,442\,538\,767.$$

6.3. Proof by mathematical induction II

Even where there is no confusion between *mathematical induction* and 'scientific induction', students often fail to appreciate the essence of "proof by induction". Before illustrating this, we repeat the basic structure of such a proof.

A mathematical result, or formula, often involves a parameter n, where n can be any positive integer. In such cases, what is presented as a single result, or formula, is a short way of writing an *infinite family of results*. The proof that such a result is correct therefore requires us to prove *infinitely many results at once*. We repeat that our only hope of achieving such a mind-boggling feat is

- to formulate the stated result for each value of n separately: that is, as a statement $\mathbf{P}(n)$ which depends on n;

- then to use bare hands to check the "beginning" - namely that the simplest case (usually $\mathbf{P}(1)$) is true;
- finally to find some way of demonstrating that,
  - as soon as we know that $\mathbf{P}(k)$ is true, for some (unknown) $k \geqslant 1$,
  - we can prove that the next result $\mathbf{P}(k+1)$ is then automatically true.

We can then conclude that

"**every** statement $\mathbf{P}(n)$ is true",

or that

"$\mathbf{P}(n)$ is true, **for all** $n \geqslant 1$".

Problem 229 Prove the statement:

"$5^{2n+2} - 24n - 25$ is divisible by 576, for all $n \geqslant 1$". △

When trying to construct proofs in private, one is free to write anything that helps as 'rough work'. But the intended thrust of Problem **229** is two-fold:

- to introduce the habit of distinguishing clearly between
  (i) the statement $\mathbf{P}(n)$ for a particular n, and
  (ii) the statement to be proved – namely "$\mathbf{P}(n)$ is true, for all $n \geqslant 1$"; and
- to draw attention to the "induction step" (i.e. the third bullet point above), where
  (i) we assume that some unspecified $\mathbf{P}(k)$ is known to be true, and
  (ii) seek to prove that the next statement $\mathbf{P}(k+1)$ must then be true.

The central lesson in completing the "induction step" is to recognize that:

to prove that $\mathbf{P}(k+1)$ is true,
one has to start by looking at what $\mathbf{P}(k+1)$ says.

In Problem **229** $\mathbf{P}(k+1)$ says that

"$5^{2(k+1)+2} - 24(k+1) - 25$ is divisible by 576".

Hence one has to start the induction step with the relevant expression

$$5^{2(k+1)+2} - 24(k+1) - 25,$$

and look for some way to rearrange this into a form where one can use $\mathbf{P}(k)$ (which we assume is already known to be true, and so are free to use).

It is in general a false strategy to work the other way round – by "starting with $\mathbf{P}(k)$, and then fiddling with it to try to get $\mathbf{P}(k+1)$". (This strategy can be made to work in the simplest cases; but it does not work in general, and so is a bad habit to get into.) So the induction step should **always** start with the hypothesis of $\mathbf{P}(k+1)$.

The next problem invites you to prove the formula for the sum of the angles in any polygon. The result is well-known; yet we are fairly sure that the reader will never have seen a correct proof. So the intention here is for you to recognise the basic character of the result, to identify the flaws in what you may until now have accepted as a proof, and to try to find some way of producing a general proof.

Problem 230 Prove by induction the statement:

> "for each $n \geqslant 3$, the angles of any n-gon in the plane have sum equal to $(n-2)\pi$ radians." $\triangle$

The formulation certainly involves a parameter $n \geqslant 3$; so you clearly need to begin by formulating the statement $\mathbf{P}(n)$. For the proof to have a chance of working, finding the right formulation involves a modest twist! So if you get stuck, it may be worth checking the first couple of lines of the solution.

No matter how $\mathbf{P}(n)$ is formulated, you should certainly know how to prove the statement $\mathbf{P}(3)$ (essentially the formula for the sum of the angles in a triangle). But it is far from obvious how to prove the "induction step":

> "if we know that $\mathbf{P}(k)$ is true for some particular $k \geqslant 1$, then $\mathbf{P}(k+1)$ must also be true".

When tackling the induction step, we certainly cannot start with $\mathbf{P}(k)$! The statement $\mathbf{P}(k+1)$ says something about polygons with $k+1$ sides: and there is no way to obtain a typical $(k+1)$-gon by fiddling with some statement about polygons with k sides. (If you start with a k-gon, you can of course draw a triangle on one side to get a $(k+1)$-gon; but this is a very special construction, and there is no easy way of knowing whether all $(k+1)$-gons can be obtained in this way from some k-gon.) The whole thrust of mathematical induction is that we must always start the induction step by thinking about the *hypothesis* of $\mathbf{P}(k+1)$ – that is in this case, by considering an arbitrary

$(k+1)$-gon and then finding some guaranteed way of "reducing" it in order to make use of $\mathbf{P}(k)$.

The next two problems invite you to prove some classical algebraic identities. Most of these may be familiar. The challenge here is to think carefully about the way you lay out your induction proof, to learn from the examples above, and (later) to learn from the detailed solutions provided.

Problem 231 Prove by induction the statement:

"$1+3+5+\cdots+(2n-1) = n^2$ holds, for all $n \geqslant 1$". △

The summation in Problem **231** was known to the ancient Greeks. The mystical *Pythagorean* tradition (which flourished in the centuries after Pythagoras) explored the character of integers through the 'spatial figures' which they formed. For example, if we arrange each successive integer as a new line of dots in the plane, then the sum "$1+2+3+\cdots+n$" can be seen to represent a *triangular* number. Similarly, if we arrange each odd number $2k-1$ in the sum "$1+3+5+\cdots+(2n-1)$" as a "k-by-k reverse L-shape", or *gnomon* (a word which we still use to refer to the L-shaped piece that casts the shadow on a sundial), then the accumulated L-shapes build up an n by n square of dots – the "1" being the dot in the top left hand corner, the "3" being the reverse L-shape of 3 dots which make this initial "1" into a 2 by 2 square, the "5" being the reverse L-shape of 5 dots which makes this 2 by 2 square into a 3 by 3 square, and so on. Hence the sum "$1+3+5+\cdots+(2n-1)$" can be seen to represent a *square* number.

There is much to be said for such geometrical illustrations; but there is no escape from the fact that they hide behind an *ellipsis* (the three dots which we inserted in the sum between "5" and "$2n-1$", which were then summarised when arranging the reverse L-shapes by ending with the words "and so on"). Proof by mathematical induction, and its application in Problem **231**, constitute a formal way of avoiding both the appeal to pictures, and the hidden *ellipsis*.

Problem 232 The sequence

$$2, 5, 13, 35, \ldots$$

is defined by its first two terms $u_0 = 2$, $u_1 = 5$, and by the recurrence relation:

$$u_{n+2} = 5u_{n+1} - 6u_n.$$

(a) Guess a closed formula for the n^{th} term u_n.

(b) Prove by induction that your guess in (a) is correct for all $n \geqslant 0$. △

Problem 233 The sequence of Fibonacci numbers

$$0, 1, 1, 2, 3, 5, 8, 13, \ldots$$

is defined by its first two terms $F_0 = 0$, $F_1 = 1$, and by the recurrence relation:

$$F_{n+2} = F_{n+1} + F_n \quad \text{when } n \geqslant 0.$$

Prove by induction that, for all $n \geqslant 0$,

$$F_n = \frac{\alpha^n - \beta^n}{\sqrt{5}}, \quad \text{where } \alpha = \frac{1+\sqrt{5}}{2} \quad \text{and} \quad \beta = \frac{1-\sqrt{5}}{2}. \qquad \triangle$$

Problem 234 Prove by induction that

$$5^{2n+1} \cdot 2^{n+2} + 3^{n+2} \cdot 2^{2n+1}$$

is divisible by 19, for all $n \geqslant 0$. △

Problem 235 Use mathematical induction to prove that each of these identities holds, for all $n \geqslant 1$:

(a) $1 + 2 + 3 + \cdots + n = \frac{n(n+1)}{2}$

(b) $\frac{1}{1\cdot 2} + \frac{1}{2\cdot 3} + \frac{1}{3\cdot 4} + \cdots + \frac{1}{n(n+1)} = 1 - \frac{1}{n+1}$

(c) $1 + q + q^2 + q^3 + \cdots + q^{n-1} = \frac{1}{1-q} - \frac{q^n}{1-q}$

(d) $0 \cdot 0! + 1 \cdot 1! + 2 \cdot 2! + \cdots + (n-1) \cdot (n-1)! = n! - 1$

(e) $(\cos\theta + i\sin\theta)^n = \cos n\theta + i\sin n\theta$. △

Problem 236 Prove by induction the statement:

"$(1 + 2 + 3 + \cdots + n)^2 = 1^3 + 2^3 + 3^3 + \cdots + n^3$, for all $n \geqslant 1$". △

We now know that, for all $n \geqslant 1$:

$$1 + 1 + 1 + \cdots + 1 \ \ (n \text{ terms}) = n.$$

And if we sum these "outputs" (that is, the first n natural numbers), we get the n^{th} triangular number:

$$1 + 2 + 3 + \cdots + n = \frac{n(n+1)}{2} = T_n.$$

The next problem invites you to find the sum of these "outputs": that is, to find the sum of the first n triangular numbers.

Problem 237

(a) Experiment and guess a formula for the sum of the first n triangular numbers:

$$T_1 + T_2 + T_3 + \cdots + T_n = 1 + 3 + 6 + \cdots + \frac{n(n+1)}{2}.$$

(b) Prove by induction that your guessed formula is correct for all $n \geqslant 1$. △

We now know closed formulae for

"$1 + 2 + 3 + \cdots + n$"

and for

"$1 \cdot 2 + 2 \cdot 3 + 3 \cdot 4 + \cdots + (n-1)n$".

The next problem hints firstly that these identities are part of something more general, and secondly that these results allow us to find identities for the sum of the first n squares:

$$1^2 + 2^2 + 3^2 + \cdots + n^2$$

for the first n cubes:

$$1^3 + 2^3 + 3^3 + \cdots + n^3$$

and so on.

Problem 238

(a) Note that

$$1{\cdot}2+2{\cdot}3+3{\cdot}4+\cdots+n(n+1) = 1{\cdot}(1+1)+2{\cdot}(2+1)+3{\cdot}(3+1)+\cdots+n{\cdot}(n+1).$$

Use this and the result of Problem **237** to derive a formula for the sum:

$$1^2 + 2^2 + 3^2 + \cdots + n^2.$$

(b) Guess and prove a formula for the sum

$$1 \cdot 2 \cdot 3 + 2 \cdot 3 \cdot 4 + 3 \cdot 4 \cdot 5 + \cdots + (n-2)(n-1)n.$$

Use this to derive a closed formula for the sum:

$$1^3 + 2^3 + 3^3 + \cdots + n^3.$$

△

It may take a bit of effort to digest the statement in the next problem. It extends the idea behind the "method of undetermined coefficients" that is discussed in **Note 2** to the solution of Problem **237**(a).

Problem 239

(a) Given $n+1$ distinct real numbers

$$a_0, a_1, a_2, \ldots, a_n,$$

find all possible polynomials of degree n which satisfy

$$f(a_0) = f(a_1) = f(a_2) = \cdots = f(a_{n-1}) = 0,\ f(a_n) = b$$

for some specified number b.

(b) For each $n \geqslant 1$, prove the following statement:

Given two labelled sets of $n+1$ real numbers

$$a_0, a_1, a_2, \ldots, a_n,$$

and

$$b_0, b_1, b_2, \ldots, b_n,$$

where the a_i are all distinct (but the b_i need not be), there exists a polynomial f_n of degree n, such that

$$f_n(a_0) = b_0,\ f_n(a_1) = b_1,\ f_n(a_2) = b_2,\ \ldots,\ f_n(a_n) = b_n. \qquad \triangle$$

We end this subsection with a mixed bag of three rather different induction problems. In the first problem the induction step involves a simple construction of a kind we will meet later.

Problem 240 A country has only 3 cent and 4 cent coins.

(a) Experiment to decide what seems to be the largest amount, N cents, that **cannot** be paid directly (without receiving change).

(b) Prove by induction that "n cents can be paid directly for each $n > N$".

$\triangle$

Problem 241

(a) Solve the equation $z + \frac{1}{z} = 1$. Calculate z^2, and check that $z^2 + \frac{1}{z^2}$ is also an integer.

(b) Solve the equation $z + \frac{1}{z} = 2$. Calculate z^2, and check that $z^2 + \frac{1}{z^2}$ is also an integer.

(c) Solve the equation $z + \frac{1}{z} = 3$. Calculate z^2, and check that $z^2 + \frac{1}{z^2}$ is also an integer.

(d) Solve the equation $z + \frac{1}{z} = k$, where k is an integer. Calculate z^2, and check that $z^2 + \frac{1}{z^2}$ is also an integer.

(e) Prove that if a number z has the property that $z + \frac{1}{z}$ is an integer, then $z^n + \frac{1}{z^n}$ is also an integer for each $n \geqslant 1$. $\triangle$

Problem 242 Let p be any prime number. Use induction to prove:

"$n^p - n$ is divisible by p for all $n \geqslant 1$". $\triangle$

6.4. Infinite geometric series

Elementary mathematics is predominantly about equations and identities. But it is often impossible to capture important mathematical relations in the form of exact equations. This is one reason why inequalities become more central as we progress; another reason is because inequalities allow us to make precise statements about certain infinite processes.

One of the simplest infinite process arises in the formula for the "sum" of an infinite *geometric series*:

$$1 + r + r^2 + r^3 + \cdots + r^n + \cdots \quad \text{(for ever)}.$$

Despite the use of the familiar-looking "+" signs, this can be no ordinary addition. Ordinary addition is defined for **two** summands; and by repeating the process, we can add **three** summands (thanks in part to the associative law of addition). We can then add four, or **any finite number** of summands. But this does not allow us to "add" infinitely many terms as in the above sum. To get round this we combine ordinary addition (of finitely many terms) and simple inequalities to find a new way of giving a meaning to the above "endless sum". In Problem **116** you used the factorisation

$$r^{n+1} - 1 = (r-1)(1 + r + r^2 + r^3 + \cdots + r^n)$$

to derive the closed formula:

$$1 + r + r^2 + r^3 + \cdots + r^n = \frac{1 - r^{n+1}}{1 - r}.$$

This formula for the sum of a finite geometric series can be rewritten in the form

$$1 + r + r^2 + r^3 + \cdots + r^n = \frac{1}{1-r} - \frac{r^{n+1}}{1-r}.$$

At first sight, this may not look like a clever move! However, it separates the part that is independent of n from the part on the RHS that depends on n; and it allows us to see how the second part behaves as n gets large:

> when $|r| < 1$, successive powers of r get smaller and smaller and converge rapidly towards 0,

so the above form of the identity may be interpreted as having the form:

$$1 + r + r^2 + r^3 + \cdots + r^n = \frac{1}{1-r} - (\text{an "error term"}).$$

Moreover if $|r| < 1$, then the "error term" converges towards 0 as $n \to \infty$.

In particular, if $1 > r > 0$, the error term is always positive, so we have proved, for all $n \geqslant 1$:

$$1 + r + r^2 + r^3 + \cdots + r^n < \frac{1}{1-r}$$

and

> the difference between the two sides tends rapidly to 0 as $n \to \infty$.

We then make the natural (but bold) step to interpret this, when $|r| < 1$, as offering a new **definition** which explains precisely what is meant by the **endless** sum

$$1 + r + r^2 + r^3 + \cdots \ (\textbf{for ever}),$$

declaring that, when $|r| < 1$,

$$1 + r + r^2 + r^3 + \cdots \ (\textbf{for ever}) = \frac{1}{1-r}.$$

More generally, if we multiply every term by a, we see that

$$a + ar + ar^2 + ar^3 + \cdots \ (\text{for ever}) = \frac{a}{1-r}.$$

Problem 243 Interpret the recurring decimal $0.037037037\cdots$ (for ever) as an infinite geometric series, and hence find its value as a fraction. △

Problem 244 Interpret the following endless processes as infinite geometric series.

(a) A square cake is cut into four quarters, with two perpendicular cuts through the centre, parallel to the sides. Three people receive one quarter each – leaving a smaller square piece of cake. This smaller piece is then cut in the same way into four quarters, and each person receives one (even smaller) piece – leaving an even smaller residual square piece, which is then cut in the same way. And so on for ever. What fraction of the original cake does each person receive as a result of this endless process?

(b) I give you a whole cake. Half a minute later, you give me half the cake back. One quarter of a minute later, I return one quarter of the cake to you. One eighth of a minute later you return one eighth of the cake to me. And so on. Adding the successive time intervals, we see that

$$\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\cdots\ \text{(for ever)}\ = 1,$$

so the whole process is completed in exactly 1 minute. How much of the cake do I have at the end, and how much do you have? △

Problem 245 When John von Neumann (1903–1957) was seriously ill in hospital, a visitor tried (rather insensitively) to distract him with the following elementary mathematics problem.

> Have you heard the one about the two trains and the fly? Two trains are on a collision course on the same track, each travelling at 30 km/h. A super-fly starts on Train A when the trains are 120 km apart, and flies at a constant speed of 50 km/h – from Train A to Train B, then back to Train A, and so on. Eventually the two trains collide and the fly is squashed. How far did the fly travel before this sad outcome? △

6.5. Some classical inequalities

The fact that our formula for the sum of a geometric series gives us an *exact* sum is very unusual – and hence very precious. For almost all other infinite series – no matter how natural, or beautiful, they may seem – you can be fairly sure that there is no obvious exact formula for the value of the sum. Hence in those cases where we happen to know the exact value, you may infer that it took the best efforts of some of the finest mathematical minds to discover what we know.

One way in which we can make a little progress in estimating the value of an infinite series is to obtain an *inequality* by comparing the given sum with a geometric series.

Problem 246

(a)(i) Explain why

$$\frac{1}{3^2} < \frac{1}{2^2},$$

so

$$\frac{1}{2^2} + \frac{1}{3^2} < \frac{2}{2^2} = \frac{1}{2}.$$

(ii) Explain why $\frac{1}{5^2}, \frac{1}{6^2}, \frac{1}{7^2}$ are all $< \frac{1}{4^2}$, so

$$\frac{1}{4^2} + \frac{1}{5^2} + \frac{1}{6^2} + \frac{1}{7^2} < \frac{4}{4^2} = \frac{1}{4}.$$

(b) Use part (a) to prove that

$$\frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \cdots + \frac{1}{n^2} < 2, \quad \text{for all} \quad n \geqslant 1.$$

(c) Conclude that the endless sum

$$\frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \cdots + \frac{1}{n^2} + \cdots \text{ (for ever)}$$

has a definite value, and that this value lies somewhere between $\frac{17}{12}$ and 2.

△

The next problem presents a rather different way of deriving a similar equality. Once the relevant inequality has been guessed, or given (see Problem **247**(a) and (b)), the proof by mathematical induction is often relatively straightforward. And after a little thought about Problem **246**, it should be clear that much of the inaccuracy in the general inequality arises from the rather poor approximations made for the first few terms (when $n = 1$, when $n = 2$, when $n = 3$, etc.); hence by keeping the first few terms as they are, and only approximating for $n \geqslant 2$, or $n \geqslant 3$, or $n \geqslant 4$, we can often prove a sharper result.

Problem 247

(a) Prove by induction that

$$\frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \cdots + \frac{1}{n^2} \leqslant 2 - \frac{1}{n}, \quad \text{for all} \quad n \geqslant 1.$$

(b) Prove by induction that

$$\frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \cdots + \frac{1}{n^2} < 1.68 - \frac{1}{n}, \text{ for all } n \geqslant 4. \qquad \triangle$$

The infinite sum

$$\frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \cdots + \frac{1}{n^2} + \cdots \text{ (for ever)}$$

is a historical classic, and has many instructive stories to tell. Recall that, in Problems **54**, **62**, **63**, **236**, **237**, **238** you found closed formulae for the sums

$$1 + 2 + 3 + \cdots + n$$

$$1^2 + 2^2 + 3^2 + \cdots + n^2$$

$$1^3 + 2^3 + 3^3 + \cdots + n^3$$

and for the sums

$$1 \times 2 + 2 \times 3 + 3 \times 4 + \cdots + (n-1)n$$

$$1 \times 2 \times 3 + 2 \times 3 \times 4 + 3 \times 4 \times 5 + \cdots + (n-2)(n-1)n.$$

Each of these expressions has a "natural" feel to it, and invites us to believe that there must be an equally natural compact answer representing the sum. In Problem **235** you took this idea one step further by finding a beautiful closed expression for the sum

$$\frac{1}{1 \cdot 2} + \frac{1}{2 \cdot 3} + \frac{1}{3 \cdot 4} + \cdots + \frac{1}{n(n+1)} = 1 - \frac{1}{n+1}$$

When we began to consider *infinite* series, we found the elegant closed formula

$$1 + r + r^2 + r^3 + \cdots + r^n = \frac{1}{1-r} - \frac{r^{n+1}}{1-r}.$$

We then observed that the final term on the RHS could be viewed as an "error term", indicating the amount by which the LHS **differs from** $\frac{1}{1-r}$, and noticed that, for any given value of r between -1 and $+1$, this error term "tends towards 0 as the power n increases". We interpreted this as indicating that one could assign a value to the endless sum

$$1 + r + r^2 + r^3 + \cdots \text{ (for ever)} = \frac{1}{1-r}.$$

In the same way, in the elegant closed formula

$$\frac{1}{1 \cdot 2} + \frac{1}{2 \cdot 3} + \frac{1}{3 \cdot 4} + \cdots + \frac{1}{n(n+1)} = 1 - \frac{1}{n+1}$$

the final term on the RHS indicates the amount by which the finite sum on the **LHS differs from 1**; and since this "error term" tends towards 0 as n increases, we may assign a value to the endless sum

$$\frac{1}{1\cdot 2}+\frac{1}{2\cdot 3}+\frac{1}{3\cdot 4}+\cdots \text{ (for ever)} = 1.$$

It is therefore natural to ask whether other infinite series, such as

$$\frac{1}{1^2}+\frac{1}{2^2}+\frac{1}{3^2}+\cdots+\frac{1}{n^2}+\cdots \text{ (for ever)}$$

may also be assigned some natural finite value. And since the series is purely numerical (without any variable parameters, such as the "r" in the geometric series formula), this answer should be a strictly *numerical* answer. And it should be **exact** – though all we have managed to prove so far (in Problems **246** and **247**) is that this numerical answer lies somewhere between $\frac{17}{12}$ and 1.68.

This question arose naturally in the middle of the seventeenth century, when mathematicians were beginning to explore all sorts of infinite series (or "sums that go on for ever"). With a little more work in the spirit of Problems **246** and **247** one could find a much more accurate approximate value. But what is wanted is an *exact* expression, not an unenlightening decimal approximation. This aspiration has a serious mathematical basis, and is not just some purist preference for elegance. The actual decimal value is very close to

$$1.649934\cdots.$$

But this conveys no structural information. One is left with no hint as to **why** the sum has this value. In contrast, the eventual form of the exact expression suggests connections whose significance remains of interest to this day.

The greatest minds of the seventeenth and early eighteenth century tried to find an exact value for the infinite sum – and failed. The problem became known as the *Basel problem* (after Jakob Bernoulli (1654–1705) who popularised the problem in 1689 – one of several members of the Bernoulli family who were all associated with the University in Basel). The problem was finally solved in 1735 – in truly breathtaking style – by the young Leonhard Euler (1707–1783) (who was at the time also in Basel). The answer

$$\frac{\pi^2}{6}$$

illustrates the final sentence of the preceding paragraph in unexpected ways, which we are still trying to understand.

In the next problem you are invited to apply similar ideas to an even more important series. Part (a) provides a relatively crude first analysis. Part (b) attacks the same question; but it does so using algebra and induction (rather than the formula for the sum of a geometric series) in a way that is then further refined in part (c).

Problem 248

(a)(i) Choose a suitable r and prove that

$$\frac{1}{1!} + \frac{1}{2!} + \cdots + \frac{1}{n!} < 1 + r + r^2 + \cdots + r^{n-1} < 2.$$

(ii) Conclude that

$$\frac{1}{0!} + \frac{1}{1!} + \frac{1}{2!} + \cdots + \frac{1}{n!} < 3, \quad \text{for every } n \geqslant 0,$$

and hence that the endless sum

$$\frac{1}{0!} + \frac{1}{1!} + \frac{1}{2!} + \cdots + \frac{1}{n!} + \cdots \text{ (for ever)}$$

can be assigned a value "e" satisfying $2 < e \leqslant 3$.

(b)(i) Prove by induction that

$$\frac{1}{0!} + \frac{1}{1!} + \frac{1}{2!} + \cdots + \frac{1}{n!} \leqslant 3 - \frac{1}{n.n!}, \quad \text{for all } \ n \geqslant 1.$$

(ii) Use part (i) to conclude that the endless sum

$$\frac{1}{0!} + \frac{1}{1!} + \frac{1}{2!} + \cdots + \frac{1}{n!} + \cdots \text{ (for ever)}$$

can be assigned a definite value "e", and that this value lies somewhere between 2.5 and 3.

(c) (It may help to read the **Note** at the start of the solution to part (c) before attempting parts (c), (d).)

(i) Prove by induction that

$$\frac{1}{0!} + \frac{1}{1!} + \frac{1}{2!} + \cdots + \frac{1}{n!} \leqslant 2.75 - \frac{1}{n.n!}, \quad \text{for all } \ n \geqslant 2.$$

(ii) Use part (i) to conclude that the endless sum

$$\frac{1}{0!} + \frac{1}{1!} + \frac{1}{2!} + \cdots + \frac{1}{n!} + \cdots \text{ (for ever)}$$

can be assigned a definite value "e", and that this value lies somewhere between 2.6 and 2.75.

(d)(i) Prove by induction that

$$\frac{1}{0!}+\frac{1}{1!}+\frac{1}{2!}+\cdots+\frac{1}{n!}\leqslant 2.722\cdots\text{(for ever)}-\frac{1}{n.n!},\quad\text{for all}\quad n\geqslant 3.$$

(ii) Use part (i) to conclude that the endless sum

$$\frac{1}{0!}+\frac{1}{1!}+\frac{1}{2!}+\cdots+\frac{1}{n!}+\cdots\ \text{(for ever)}$$

can be assigned a definite value "e", and that this value lies somewhere between 2.708 and 2.7222$\cdots$ (for ever). △

We end this section with one more inequality in the spirit of this section, and two rather different inequalities whose significance will become clear later.

Problem 249 Prove by induction that

$$\tfrac{1}{\sqrt{1}}+\tfrac{1}{\sqrt{2}}+\tfrac{1}{\sqrt{3}}+\cdots+\tfrac{1}{\sqrt{n}}\geqslant\sqrt{n},\ \text{for all } n\geqslant 1.$$ △

Problem 250 Let a, b be real numbers such that $a\neq b$, and $a+b>0$. Prove by induction that

$$2^{n-1}(a^n+b^n)\geqslant(a+b)^n,\ \text{for all } n\geqslant 1.$$ △

Problem 251 Let x be any real number $\geqslant -1$. Prove by induction that

$$(1+x)^n\geqslant 1+nx,\ \text{for all } n\geqslant 1.$$ △

6.6. The harmonic series

The great foundation of mathematics is
***the principle of contradiction**, or **of identity**,*
that is to say that a statement
cannot be true and false at the same time,
*and that thus A **is** A**, and cannot be not** A.*
And this single principle is enough to prove
the whole of arithmetic and the whole of geometry,
that is to say all mathematical principles.
Gottfried W. Leibniz (1646–1716)

We have seen how some infinite series, or sums that go on for ever, can be assigned a finite value for their sum:

$$1 + r + r^2 + r^3 + \cdots \text{ (for ever)} = \frac{1}{1-r}$$

$$\frac{1}{1\cdot 2} + \frac{1}{2\cdot 3} + \frac{1}{3\cdot 4} + \cdots + \frac{1}{n(n+1)} + \cdots \text{ (for ever)} = 1$$

$$\frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \cdots + \frac{1}{n^2} + \cdots \text{ (for ever)} = \frac{\pi^2}{6}$$

$$\frac{1}{0!} + \frac{1}{1!} + \frac{1}{2!} + \cdots + \frac{1}{n!} + \cdots \text{ (for ever)} = e.$$

We say that these series *converge* (meaning that they can be assigned a finite value).

This section is concerned with another very natural series, the so-called *harmonic series*

$$\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} + \cdots \text{ (for ever)}.$$

It is not entirely clear why this is called the harmonic series. The natural overtones that arise in connection with plucking a stretched string (as with a guitar or a harp) have wavelengths that are $\frac{1}{2}$ the basic wavelength, or $\frac{1}{3}$ of the basic wavelength, and so on. It is also true that, just as each term of an arithmetic series is the *arithmetic mean* of its two neighbours, and each term of a geometric series is the *geometric mean* of its two neighbours, so each term of the harmonic series after the first is equal to the *harmonic mean* (see Problems **85**, **89**) of its two neighbours:

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

Unlike the first two series above, there is no obvious closed formula for the finite sum

$$s_n = \frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}.$$

Certainly the sequence of successive sums

$$s_1 = 1,\ s_2 = \frac{3}{2},\ s_3 = \frac{11}{6},\ s_4 = \frac{25}{12},\ s_5 = \frac{137}{60}, \ldots$$

does not suggest any general pattern.

Problem 252 Suppose we denote by S the "value" of the endless sum

$$\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} + \cdots \text{ (for ever)}$$

(i) Write out the endless sum corresponding to "$\frac{1}{2}S$".

(ii) Remove the terms of this endless sum from the endless sum S, to obtain another endless sum corresponding to "$S - \frac{1}{2}S$" $= \frac{1}{2}S$.

(iii) Compare the first term of the series in (i) (namely $\frac{1}{2}$) with the first term in the series in (ii) (namely 1); compare the second term in the series in (i) with the second term in the series in (ii); and so on. What do you notice?
△

The Leibniz quotation above emphasizes that the reliability of mathematics stems from a single principle – namely the refusal to tolerate a contradiction. We have already made explicit use of this principle from time to time (see, for example, the solution to Problem **125**). The message is simple: whenever we hit a contradiction, we know that we have "gone wrong" – either by making an error in calculation or logic, or by beginning with a false assumption. In Problem **252** the observations you were expected to make are paradoxical: you obtained two different series, which both correspond to "$\frac{1}{2}S$", but every term in one series is larger than the corresponding term in the other! What one can conclude may not be entirely clear. But it is certainly clear that something is wrong: we have somehow created a contradiction. The three steps ((i), (ii), (iii)) appear to be relatively sensible. But the final observation "$\frac{1}{2}S < \frac{1}{2}S$" (since $\frac{1}{2} < 1$, $\frac{1}{4} < \frac{1}{3}$, etc.) makes no sense. And the only obvious assumption we have made is to assume that the endless sum

$$\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} + \cdots \text{ (for ever)}$$

can be assigned a value "S", which can then be manipulated *as though it were a number.*

The conclusion would seem to be that, whether or not the endless sum has a meaning, it cannot be assigned a value in this way. We say that the series *diverges.* Each finite sum

$$\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}$$

has a value, and these values "grow more and more slowly" as n increases:

- the first term immediately makes the sum $= 1$
- it takes 4 terms to obtain a sum > 2;
- it takes 11 terms to obtain a sum > 3; and

- it takes 12 367 terms before the series reaches a sum > 10.

However, this slow growth is not enough to guarantee that the corresponding endless sum corresponds to a finite numerical value.

The danger signals should already have been apparent in Problem **249**, where you proved that

$$\frac{1}{\sqrt{1}} + \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{3}} + \cdots + \frac{1}{\sqrt{n}} \geqslant \sqrt{n}$$

The n^{th} term $\frac{1}{\sqrt{n}}$ tends to 0 as n increases; so the finite sums grow ever more slowly as n increases. However, the LHS can be made larger than any integer K simply by taking K^2 terms. Hence there is no way to assign a finite value to the endless sum

$$\frac{1}{\sqrt{1}} + \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{3}} + \cdots + \frac{1}{\sqrt{n}} + \cdots \text{ (for ever).}$$

Problem 253

(a)(i) Explain why

$$\frac{1}{2} + \frac{1}{3} < 1.$$

(ii) Explain why

$$\frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \frac{1}{7} < 1.$$

(iii) Extend parts (i) and (ii) to prove that

$$\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{2^n - 1} < n, \quad \text{for all} \quad n \geqslant 2.$$

(iv) Finally use the fact that, when $n \geqslant 3$,

$$\frac{1}{2^n} < \frac{1}{2} - \frac{1}{3}$$

to modify the proof in (iii) slightly, and hence show that

$$\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{2^n} < n, \quad \text{for all} \quad n \geqslant 3.$$

(b)(i) Explain why

$$\frac{1}{3} + \frac{1}{4} > \frac{1}{2}.$$

(ii) Explain why
$$\frac{1}{5}+\frac{1}{6}+\frac{1}{7}+\frac{1}{8}>\frac{1}{2}.$$

(iii) Extend parts (i) and (ii) to prove that
$$\frac{1}{1}+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{2^n}>1+\frac{n}{2},\quad \text{for all}\quad n\geqslant 2.$$

(c) Combine parts (a) and (b) to show that, for all $n \geqslant 2$, we have the two inequalities
$$1+\frac{n}{2}<\frac{1}{1}+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{2^n}<n.$$
Conclude that the endless sum
$$\frac{1}{1}+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{n}+\cdots\quad \text{(for ever)}$$
cannot be assigned a finite value. △

The result in Problem **253**(c) has an unexpected consequence.

Problem 254 Imagine that you have an unlimited supply of identical rectangular strips of length 2. (Identical empty plastic CD cases can serve as a useful illustration, provided one focuses on their rectangular side profile, rather than the almost square frontal cross-section.) The goal is to construct a ‘stack’ in such a way as to stick out as far as possible beyond a table edge. One strip balances exactly at its midpoint, so can protrude a total distance of 1 without tipping over.

(a) Arrange a stack of n strips of length 2, one on top of the other, with the bottom strip protruding distance $\frac{1}{n}$ beyond the edge of the table, the second strip from the bottom protruding $\frac{1}{n-1}$ beyond the leading edge of the bottom strip, the third strip from the bottom protruding $\frac{1}{n-2}$ beyond the leading edge of the strip below it, and so on until the $(n-1)^{\text{th}}$ strip from the bottom protrudes distance $\frac{1}{2}$ beyond the leading edge of the strip below it, and the top strip protrudes distance 1 beyond the leading edge of the strip below it (see Figure 10). Prove that a stack of n identical strips arranged in this way will just avoid tipping over the table edge.

(b) Conclude that we can choose n so that an arrangement of n strips can (in theory) protrude as far beyond the edge of the table as we wish – without tipping. △

The next problem illustrates, in the context of the harmonic series, what is in fact a completely general phenomenon: an endless sum of *steadily decreasing*

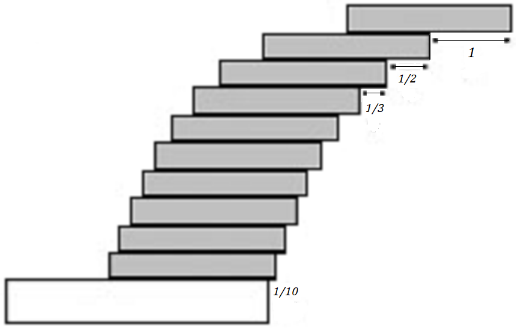


Figure 10: Overhanging strips, $n = 10$.

positive terms may converge or diverge; but **provided** the terms themselves converge to 0, then the the corresponding "alternating sum" – where the same terms are combined but with alternately positive and negative signs – **always** converges.

Problem 255

(a) Let

$$s_n = \frac{1}{1} - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \cdots \pm \frac{1}{n}$$

(where the final operation is "+" if n is odd and "−" if n is even).

(i) Prove that

$$s_{2n-2} < s_{2n} < s_{2m+1} < s_{2m-1},$$

for all $m, n \geqslant 1$.

(ii) Conclude that the endless alternating sum

$$\frac{1}{1} - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \cdots \text{ (for ever)}$$

can be assigned a value s that lies somewhere between $s_6 = \frac{37}{60}$ and $s_5 = \frac{47}{60}$.

(b) Let

$$a_1, a_2, a_3, \ldots$$

be an endless, decreasing sequence of positive terms (that is, $a_{n+1} < a_n$ for all $n \geqslant 1$). Suppose that the sequence of terms a_n converges to 0 as $n \to \infty$.

(i) Let

$$s_n = a_1 - a_2 + a_3 - a_4 + a_5 - \cdots \pm a_n$$

(where the final operation is "+" if n is odd and "−" if n is even). Prove that

$$s_{2n-2} < s_{2n} < s_{2m+1} < s_{2m-1}, \quad \text{for all } \ m, n \geqslant 1.$$

(ii) Conclude that the endless alternating sum

$$a_1 - a_2 + a_3 - a_4 + a_5 - \cdots \ \text{(for ever)}$$

can be assigned a value s that lies somewhere between $s_2 = a_1 - a_2$ and $s_3 = a_1 - a_2 + a_3$. △

Just as with the series

$$\frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \cdots + \frac{1}{n^2} + \cdots \ \text{(for ever)}$$

$$\frac{1}{0!} + \frac{1}{1!} + \frac{1}{2!} + \cdots + \frac{1}{n!} + \cdots \ \text{(for ever)},$$

we can show relatively easily that

$$\frac{1}{1} - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \cdots \ \text{(for ever)}$$

can be assigned a value s. It is far less clear whether this value has a familiar name! (It is in fact equal to the natural logarithm of 2: "$\log_e 2$".) A similarly intriguing series is the alternating series of odd terms from the harmonic series:

$$\frac{1}{1} - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \cdots \ \text{(for ever)}$$

You should be able to show that this endless series can be assigned a value somewhere between $s_2 = \frac{2}{3}$ and $s_3 = \frac{13}{15}$; but you are most unlikely to guess that its value is equal to $\frac{\pi}{4}$. This was first discovered in 1674 by Leibniz (1646–1716). One way to obtain the result is using the integral of $(1+x^2)^{-1}$ from 0 to 1: on the one hand the integral is equal to $\arctan x$ evaluated when $x = 1$ (that is, $\frac{\pi}{4}$); on the other hand, we can expand the integrand as a power series $1 - x^2 + x^4 - x^6 + \cdots$, integrate term by term, and prove that the resulting series converges when $x = 1$. (It does indeed converge, though it does so very, very slowly.)

The fact that the alternating harmonic series has the value $\log_e 2$ seems to have been first shown by Euler (1707–1783), using the power series expansion for $\log(1 + x)$.

6.7. Induction in geometry, combinatorics and number theory

We turn next to a mixed collection of problems designed to highlight a range of applications.

Problem 256 Let $f_1 = 2$, $f_{k+1} = f_k(f_k + 1)$. Prove by induction that f_k has at least k distinct prime factors. △

Problem 257

(a) Prove by induction that n points on a straight line divide the line into $n + 1$ parts.

(b)(i) By experimenting with small values of n, guess a formula R_n for the maximum number of regions which can be created in the plane by an array of n straight lines.

(ii) Prove by induction that n straight lines in the plane divide the plane into at most R_n regions.

(c)(i) By experimenting with small values of n, guess a formula S_n for the maximum number of regions which can be created in 3-dimensions by an array of n planes.

(ii) Prove by induction that n planes in 3-dimensions divide space into at most S_n regions. △

Problem 258 Given a square, prove that, for each $n \geqslant 6$, the initial square can be cut into n squares (of possibly different sizes). △

Problem 259 A tree is a connected graph, or network, consisting of vertices and edges, but with no cycles (or circuits). Prove that a tree with n vertices has exactly $n - 1$ edges. △

The next problem concerns *spherical polyhedra.* A spherical polyhedron is a polyhedral surface in 3-dimensions, which can be inflated to form a sphere (where we assume that the faces and edges can stretch as required). For example, a cube is a spherical polyhedron; but the surface of a picture frame is not. A spherical polyhedron has

- *faces* (flat 2-dimensional polygons, which can be stretched to take the form of a disc),

- *edges* (1-dimensional line segments, where exactly two faces meet), and
- *vertices* (0-dimensional points, where several faces and edges meet, in such a way that they form a single cycle around the vertex).

Each face must clearly have at least 3 edges; and there must be at least three edges and three faces meeting at each vertex.

If a spherical polyhedron has V vertices, E edges, and F faces, then the numbers V, E, F satisfy Euler's formula

$$V - E + F = 2.$$

For example, a cube has $V = 8$ vertices, $E = 12$ edges, and $F = 6$ faces, and $8 - 12 + 6 = 2$.

Problem 260

(a)(i) Describe a spherical polyhedron with exactly 6 edges.

(ii) Describe a spherical polyhedron with exactly 8 edges.

(b) Prove that no spherical polyhedron can have exactly 7 edges.

(c) Prove that for every $n \geqslant 8$, there exists a spherical polyhedron with n edges. △

Problem 261 A map is a (finite) collection of regions in the plane, each with a boundary, or border, that is 'polygonal' in the sense that it consists of a single sequence of distinct vertices and – possibly curved – edges, that separates the plane into two parts, one of which is the polygonal region itself. A map can be properly coloured if each region can be assigned a colour so that each pair of neighbouring regions (sharing an edge) always receive different colours. Prove that the regions of such a map can be properly coloured with just two colours if and only if an even number of edges meet at each vertex. △

Problem 262 (Gray codes) There are 2^n sequences of length n consisting of 0s and 1s. Prove that, for each $n \geqslant 2$, these sequences can be arranged in a cyclic list such that any two neighbouring sequences (including the last and the first) differ in exactly one coordinate position. △

Problem 263 (Calkin-Wilf tree) The binary tree in the plane has a distinguished vertex as 'root', and is constructed inductively. The root is joined to two new vertices; and each new vertex is then joined to two further new vertices – with the construction process continuing for ever (Figure 11).

Label the vertices of the binary tree with positive fractions as follows:

- the root is given the label $\frac{1}{1}$
- whenever we know the label $\frac{i}{j}$ of a 'parent' vertex, we label its 'left descendant' as $\frac{i}{i+j}$, and its 'right descendant' $\frac{i+j}{j}$.

(a) Prove that every positive rational $\frac{r}{s}$ occurs once and only once as a label, and that it occurs in its lowest terms.

(b) Prove that the labels are left-right symmetric in the sense that labels in corresponding left and right positions are reciprocals of each other. △

Problem 264 A collection of n intervals on the x-axis is such that every pair of intervals have a point in common. Prove that all n intervals must then have at least one point in common. △

6.8. Two problems

Problem 265 Several identical tanks of water sit on a horizontal base. Each pair of tanks is connected with a pipe at ground level controlled by a valve, or tap. When a valve is opened, the water level in the two connected tanks becomes equal (to the average, or mean, of the initial levels). Suppose we start with tank T which contains the least amount of water. The aim is to open and close valves in a sequence that will lead to the final water level in tank T being as high as possible. In what order should we make these connections? △

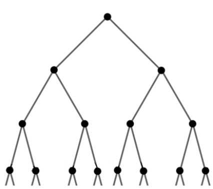


Figure 11: A (rooted) binary tree.

Problem 266 I have two flasks. One is 'empty', but still contains a residue of a dangerous chemical; the other contains a fixed amount of solvent that can be used to wash away the remaining traces of the dangerous chemical. What is the best way to use the fixed quantity of solvent? Should I use it all at once to wash out the first flask? Or should I first wash out the flask using just half of the solvent, and then repeat with the other half? Or is there a better way of using the available solvent to remove as much as possible of the dangerous chemical? △

6.9. Infinite descent

In this final section we touch upon an important variation on mathematical induction. This variation is well-illustrated by the next (probably familiar) problem.

Problem 267 Write out for yourself the following standard proof that $\sqrt{2}$ is irrational.

(i) Suppose to the contrary that $\sqrt{2}$ is rational. Then $\sqrt{2} = \frac{m}{n}$ for some positive integers m, n. Prove first that m must be even.

(ii) Write $m = 2m'$, where m' is also an integer. Show that n must also be even.

(iii) How does this lead to a contradiction? △

Problem **267** has the classic form of a proof which reaches a contradiction by *infinite descent*.

1. We start with a claim which we wish to prove is true. Often when we do not know how to begin, it makes sense to ask what would happen if the claim were *false*. This then guarantees that there must be some *counterexample*, which satisfies the given hypothesis, but which fails to satisfy the asserted conclusion.

2. *Infinite descent* becomes an **option** whenever each such counterexample gives rise to some *positive integer* parameter n (such as the denominator in Problem **267**(i)).

3. *Infinite descent* becomes a **reality**, if one can prove that the existence of the initial counterexample leads to a construction that produces a counterexample *with a smaller value* n' *of the parameter* n, since repeating this step then gives rise to an endlessly decreasing sequence

$$n > n' > n'' > \cdots > 0$$

of positive integers, which is impossible (since such a chain can have length at most n).

4. Hence the initial assumption that the claim was false must itself be false – so the claim must be true (as required).

Proof by "infinite descent" is an invaluable tool. But it is important to realise that the method is essentially a variation on proof by mathematical induction. As a first step in this direction it is worth reinterpreting Problem **267** as an induction proof.

Problem 268 Let $\mathbf{P}(n)$ be the statement:

"$\sqrt{2}$ cannot be written as a fraction with positive denominator $\leqslant n$".

(i) Explain why $\mathbf{P}(1)$ is true.

(ii) Suppose that $\mathbf{P}(k)$ is true for some $k \geqslant 1$. Use the proof in Problem **267** to show that $\mathbf{P}(k+1)$ must then be true as well.

(iii) Conclude that $\mathbf{P}(n)$ is true for all $n \geqslant 1$, whence $\sqrt{2}$ must be irrational.△

Problem **268** shows that, in the particular case of Problem **267** one can translate the standard proof that "$\sqrt{2}$ is irrational" into a proof by induction. But much more is true. The contradiction arising in step 3. above is an application of an important principle, namely

> **The Least Element Principle**: Every non-empty set S of positive integers has a smallest element.

The Least Element Principle is equivalent to *The Principle of Mathematical Induction* which we stated at the beginning of the chapter:

> **The Principle of Mathematical Induction:** If a subset S of the positive integers
>
> - contains the integer "1", and has the property that
> - whenever an integer k is in the set S, then the next integer $k+1$ is always in S too,
>
> then S contains **all** positive integers.

Problem 269

(a) Assume the *Least Element Principle.* Suppose a subset T of the positive integers contains the integer "1", and that whenever k is in the set T, then $k+1$ is also in the set T. Let S be the set of all positive integers which are **not** in the set T. Conclude that S must be empty, and hence that T contains all positive integers.

(b) Assume the *Principle of Mathematical Induction.* Let T be a non-empty set of positive integers, and suppose that T does **not** have a smallest element. Let S be the set of all positive integers which do **not** belong to the set T. Prove that "1" must belong to S, and that whenever the positive integer k belongs to S, then so does $k+1$. Derive the contradiction that T must be empty, contrary to assumption. Conclude that T must in fact have a smallest element. △

To round off this final chapter you are invited to devise a rather different proof of the irrationality of $\sqrt{2}$.

Problem 270 This sequence of constructions presumes that we know – for example, by Pythagoras' Theorem – that, in any square $OPQR$, the ratio

$$\text{"diagonal : side"} = \underline{OQ} : \underline{OP} = \sqrt{2} : 1.$$

Let $OPQR$ be a square. Let the circle with centre Q and passing through P meet $\underline{OQ}$ at P'. Construct the perpendicular to $\underline{OQ}$ at P', and let this meet $\underline{OR}$ at Q'.

(i) Explain why $\underline{OP'} = \underline{P'Q'}$. Construct the point R' to complete the square $OP'Q'R'$.

(ii) Join $\underline{QQ'}$. Explain why $\underline{P'Q'} = \underline{RQ'}$.

(iii) Suppose $\underline{OQ} : \underline{OP} = m : n$. Conclude that $\underline{OQ'} : \underline{OP'} = 2n - m : m - n$.

(iv) Prove that $n < m < 2n$, and hence that $0 < m - n < n$, $0 < 2n - m$.

(v) Explain how, if m, n can be chosen to be positive integers, the above sequence of steps sets up an "infinite descent" – which is impossible. △

6.10. Chapter 6: Comments and solutions

Note: It is important to separate the underlying idea of "induction" from the formal way we have chosen to present proofs. As ever in mathematics, the ideas are what matter most. But the process of struggling with (and slowly coming to understand why we need) the formal structure behind the written proofs is part of the way the ideas are tamed and organised.

Readers should not be intimidated by the physical extent of the solutions to this chapter. As explained in the main text it is important for all readers to review the way they approach induction proofs: so we have erred in favour of completeness – knowing that as each reader becomes more confident, s/he will increasingly compress, or abbreviate, some of the steps.

228.

(a) Yes.

(b) Yes.

$$2 \times 3 \times 5 \times 7 \times 11 + 1 = 2311,$$

and $\sqrt{2311} = 48.07\ldots$, so we only need to check prime factors up to 47.

(c) No.

$$2 \times 3 \times 5 \times 7 \times 11 \times 13 + 1 = 30\,031,$$

and $\sqrt{30\,031} = 173.29\ldots$ so we might have to check all 40 possible prime factors up to 173. However, we only have to start at 17 [Why?], and checking with a calculator is very quick. In fact 30 031 factorises rather prettily as 59×509.

229. Note: The statement in the problem includes the *quantifier* "**for all** $n \geqslant 1$".

What is to be proved is the compound statement

"$\mathbf{P}(n)$ is true **for all** $n \geqslant 1$".

In contrast, each individual statement $\mathbf{P}(n)$ refers to a single value of n.

It is essential to be clear when you are dealing with the *compound* statement, and when you are referring to some *particular* statement $\mathbf{P}(1)$, or $\mathbf{P}(n)$, or $\mathbf{P}(k)$.

Let $\mathbf{P}(n)$ be the statement:

"$5^{2n+2} - 24n - 25$ is divisible by 576".

- $\mathbf{P}(1)$ is the statement:

  "$5^4 - 24 \times 1 - 25$ is divisible by 576".

  That is:

  "$625 - 49 = 576$ is divisible by 576",

  which is evidently true.

- Now suppose that we know $\mathbf{P}(k)$ is true for some $k \geqslant 1$. We must show that $\mathbf{P}(k+1)$ is then also true.

  To prove that $\mathbf{P}(k+1)$ is true, we have to consider the statement $\mathbf{P}(k+1)$. It is no use starting with $\mathbf{P}(k)$. However, since we know that $\mathbf{P}(k)$ is true, we are free to use it at any stage if it turns out to be useful.

  To prove that $\mathbf{P}(k+1)$ is true, we have to show that

  "$5^{2(k+1)+2} - 24(k+1) - 25$ is divisible by 576".

  So we must start with $5^{2(k+1)+2} - 24(k+1) - 25$ and try to "simplify" it (knowing that "simplify" in this case means "rewrite it in a way that involves $5^{2k+2} - 24k - 25$").

  $$\begin{aligned} 5^{2(k+1)+2} - 24(k+1) - 25 & \\ &= [5^{2k+4}] - 24k - 24 - 25 \\ &= [5^2(5^{2k+2} - 24k - 25) + 5^2 \cdot (24k) + 5^2 \cdot 25] \\ &\qquad - 24k - (24+25) \\ &= 5^2(5^{2k+2} - 24k - 25) + [(5^2-1) \times (24k)] \\ &\qquad + [5^2 \times 25 - 24 - 25] \\ &= 5^2(5^{2k+2} - 24k - 25) + 24^2k + [5^2 \times 25 - 24 - 25] \\ &= 5^2(5^{2k+2} - 24k - 25) + 24^2k + [5^4 - 24 - 25]. \end{aligned}$$

  The first term on the RHS is a multiple of $(5^{2k+2} - 24k - 25)$, so is divisible by 576 (since we know that $\mathbf{P}(k)$ is true); the second term on the RHS is a multiple of $24^2 = 576$; and the third term on the RHS is the expression arising in $\mathbf{P}(1)$, which we saw is equal to 576.
  $\therefore$ the whole RHS is divisible by 576
  $\therefore$ the LHS is divisible by 576, so $\mathbf{P}(k+1)$ is true.

Hence

- $\mathbf{P}(1)$ is true; and
- whenever $\mathbf{P}(k)$ is true for some $k \geqslant 1$, we have proved that $\mathbf{P}(k+1)$ must be true.

$\therefore$ $\mathbf{P}(n)$ is true for all $n \geqslant 1$. QED

230. Let $\mathbf{P}(n)$ be the statement:

> "the angles of any p-gon, for any value of p with $3 \leqslant p \leqslant n$, have sum exactly $(p-2)\pi$ radians".

1. $\mathbf{P}(3)$ is the statement:

   "the angles of any triangle have sum π radians".

   This is a known fact: given triangle $\triangle ABC$, draw the line XAY through A parallel to BC, with X on the same side of AC as B and Y on the same side of AB as C. Then $\angle XAB = \angle CBA$ and $\angle YAC = \angle BCA$ (alternate angles), so

$$\angle B + \angle A + \angle C = \angle XAB + \angle A + \angle YAC = \angle XAY = \pi.$$

2. Now we suppose that $\mathbf{P}(k)$ is known to be true for some $k \geqslant 3$. We must show that $\mathbf{P}(k+1)$ is then necessarily true.

To prove that $\mathbf{P}(k+1)$ is true, we have to consider the statement $\mathbf{P}(k+1)$: that is,

> "the angles of a p-gon, for any value of p with $3 \leqslant p \leqslant k+1$, have sum exactly $(p-2)\pi$ radians".

This can be reworded by splitting it into two parts:

> "the angles of any p-gon for $3 \leqslant p \leqslant k$ have sum exactly $(p-2)\pi$ radians;"

and

> "the angles of any $(k+1)$-gon have sum exactly $((k+1)-2)\pi$ radians".

The first part of this revised version is precisely the statement $\mathbf{P}(k)$, which we suppose is known to be true. So the crux of the matter is to prove the second part – namely that the angles of any $(k+1)$-gon have sum $(k-1)\pi$.

Let $A_0A_1A_2 \cdots A_k$ be any $(k+1)$-gon.

[**Note:** The usual move at this point is to say "draw the chord A_kA_1 to cut the polygon into the triangle $A_kA_1A_0$ (with angle sum π (by $\mathbf{P}(3)$), and the k-gon $A_1A_2 \cdots A_k$ (with angle sum $(k-2)\pi$ (by $\mathbf{P}(k)$), whence we can add to see that $A_0A_1A_2 \cdots A_k$ has angle sum $((k+1)-2)\pi$. However, this only works

- if the triangle $A_kA_1A_0$ "sticks out" rather than in, and
- if no other vertices or edges of the $(k+1)$-gon encroach into the part that is cut off – which can only be guaranteed if the polygon is "convex".

So what is usually presented as a "proof" does not work *in general.*

If we want to prove the general result – for polygons of all shapes – we have to get round this unwarranted assumption. Experiment may persuade you that "there is always some vertex that sticks out and which can be safely "cut off"; but it is not at all clear how to prove this fact (we know of no simple proof). So we have to find another way.]

Consider the vertex A_1, and its two neighbours A_0 and A_2.

Imagine each half-line, which starts at A_1, and which sets off into the *interior* of the polygon. Because the polygon is finite, each such half-line defines a line segment $\underline{A_1X}$, where X is the next point of the polygon which the half line hits (that is, X is one of the vertices A_m, or a point on one of the sides $\underline{A_mA_{m+1}}$).

Consider the locus of all such points X as the half line swings from $\underline{A_1A_0}$ (produced) to $\underline{A_1A_2}$ (produced). There are two possibilities: either

(a) these points X all belong to a single side of the polygon; or

(b) they don't.

(a) In the first case none of the vertices or sides of the polygon $A_0A_1A_2\cdots A_k$ intrude into the interior of the triangle $A_0A_1A_2$, so the chord $\underline{A_0A_2}$ separates the $(k+1)$-gon $A_0A_1A_2\cdots A_k$ into a triangle $A_0A_1A_2$ and a k-gon $A_0A_2A_3\cdots A_k$. The angle sum of $A_0A_1A_2\cdots A_k$ is then equal to the sum of (i) the angle sum of the triangle $A_0A_1A_2$ and (ii) the angle sum of $A_0A_2A_3\cdots A_k$ – which are equal to π and $(k-2)\pi$ respectively (by $\mathbf{P}(k)$). So the angle sum of the $(k+1)$-gon $A_0A_2A_3\cdots A_k$ is equal to $((k+1)-2)\pi$ as required.

(b) In the second case, as the half-line A_1X rotates from A_1A_0 to A_1A_2, the point X must at some instant switch from lying on one side of the polygon to lying on another side; at the very instant where X switches sides, $X = A_m$ must be a vertex of the polygon.

Because of the way the point X was chosen, the chord $\underline{A_1X} = \underline{A_1A_m}$ does not meet any other point of the $(k+1)$-gon $A_0A_1A_2\cdots A_k$, and so splits the $(k+1)$-gon into an m-gon $A_1A_2A_3\cdots A_m$ (with angle sum $(m-2)\pi$ by $\mathbf{P}(k)$) and a $(k-m+3)$-gon $A_mA_{m+1}A_{m+2}\cdots A_kA_0A_1$ (with angle sum $(k-m+1)\pi$ by $\mathbf{P}(k)$). So the $(k+1)$-gon $A_0A_1A_2\cdots A_k$ has angle sum $((k+1)-2)\pi$ as required.

Hence $\mathbf{P}(k+1)$ is true.

$\therefore$ $\mathbf{P}(n)$ is true for all $n \geqslant 3$. QED

231. Let $\mathbf{P}(n)$ be the statement

$$1+3+5+\cdots+(2n-1) = n^2.$$

- LHS of $\mathbf{P}(1) = 1$; RHS of $\mathbf{P}(1) = 1^2$. Since these two are equal, $\mathbf{P}(1)$ is true.
- Suppose that $\mathbf{P}(k)$ is true for some particular (unspecified) $k \geqslant 1$; that is, we know that, for this particular k,

  $$1+3+5+\cdots+(2k-1) = k^2.$$

  We wish to prove that $\mathbf{P}(k+1)$ must then be true.

  Now $\mathbf{P}(k+1)$ is an equation, so we start with the LHS of $\mathbf{P}(k+1)$ and try to simplify it in an appropriate way to get the RHS of $\mathbf{P}(k+1)$:

  $$\begin{aligned}\text{LHS of } \mathbf{P}(k+1) &= 1+3+5+\cdots+(2(k+1)-1)\\ &= (1+3+5+\cdots+(2k-1))+(2k+1).\end{aligned}$$

  If we now use $\mathbf{P}(k)$, which we are supposing to be true, then the first bracket is equal to k^2, so this sum is equal to:

  $$\begin{aligned}&= k^2+(2k+1)\\ &= (k+1)^2\\ &= \text{RHS of } \mathbf{P}(k+1).\end{aligned}$$

  Hence $\mathbf{P}(k+1)$ holds.

Combining these two bullet points then shows that "$\mathbf{P}(n)$ holds, for all $n \geqslant 1$".

QED

232.

(a) The only way to learn is by trying and failing; then trying again and failing slightly better! So don't give up too quickly. It is natural to try to relate each term to the one before. You may then notice that each term is slightly less than 3 times the previous term.

(b) **Note:** The recurrence relation for u_n involves the two previous terms. So when we assume that $\mathbf{P}(k)$ is known to be true and try to prove $\mathbf{P}(k+1)$, the recurrence relation for u_{k+1} will involve u_k and u_{k-1}, so $\mathbf{P}(n)$ needs to be formulated to ensure that we can use closed expressions for both these terms. For the same reason, the induction proof has to start by showing that both $\mathbf{P}(0)$ and $\mathbf{P}(1)$ are true.

Let $\mathbf{P}(n)$ be the statement:

"$u_m = 2^m + 3^m$ for all m, $0 \leqslant m \leqslant n$".

- LHS of $\mathbf{P}(0) = u_0 = 2$; RHS of $\mathbf{P}(0) = 2^0 + 3^0 = 1 + 1$. Since these two are equal, $\mathbf{P}(0)$ is true.

  $\mathbf{P}(1)$ combines $\mathbf{P}(0)$, and the equality of $u_1 = 5$ and $2^1 + 3^1$; since these two are equal, $\mathbf{P}(1)$ is true.

- Suppose that $\mathbf{P}(k)$ is true for some particular (unspecified) $k \geqslant 1$; that is, we know that, for this particular k,

  "$u_m = 2^m + 3^m$ for all m, $0 \leqslant m \leqslant k$."

  We wish to prove that $\mathbf{P}(k+1)$ must then be true.

  Now $\mathbf{P}(k+1)$ requires us to prove that

  "$u_m = 2^m + 3^m$ for all m, $0 \leqslant m \leqslant k+1$."

  Most of this is guaranteed by $\mathbf{P}(k)$, which we assume to be true. It remains for us to check that the equality holds for u_{k+1}. We know that

  $$u_{k+1} = 5u_k - 6u_{k-1}.$$

  And we may use $\mathbf{P}(k)$, which we are supposing to be true, to conclude that:

  $$\begin{aligned} u_{k+1} &= 5\left(2^k + 3^k\right) - 6\left(2^{k-1} + 3^{k-1}\right) \\ &= (10-6)2^{k-1} + (15-6)3^{k-1} \\ &= 2^{k+1} + 3^{k+1}. \end{aligned}$$

  Hence $\mathbf{P}(k+1)$ holds.

Combining these two bullet points then shows that "$\mathbf{P}(n)$ holds, for all $n \geqslant 0$".

QED

233. Let $\mathbf{P}(n)$ be the statement:

"$F_m = \frac{\alpha^m - \beta^m}{\sqrt{5}}$ for all m, $0 \leqslant m \leqslant n$",

where $\alpha = \frac{1+\sqrt{5}}{2}$ and $\beta = \frac{1-\sqrt{5}}{2}$.

- LHS of $\mathbf{P}(0) = F_0 = 0$; RHS of $\mathbf{P}(0) = \frac{1-1}{\sqrt{5}} = 0$. Since these two are equal, $\mathbf{P}(0)$ is true.
  LHS of $\mathbf{P}(1) = F_1 = 1$; RHS of $\mathbf{P}(1) = \frac{\alpha-\beta}{\sqrt{5}} = 1$. Since these two are equal, $\mathbf{P}(1)$ is true.
- Suppose that $\mathbf{P}(k)$ is true for some particular (unspecified) $k \geqslant 1$; that is, we know that, for this particular k,

  "$F_m = \frac{\alpha^m - \beta^m}{\sqrt{5}}$ for all m, $0 \leqslant m \leqslant k$."

  We wish to prove that $\mathbf{P}(k+1)$ must then be true.
  Now $\mathbf{P}(k+1)$ requires us to prove that

  "$F_m = \frac{\alpha^m - \beta^m}{\sqrt{5}}$ for all m, $0 \leqslant m \leqslant k+1$."

  Most of this is guaranteed by $\mathbf{P}(k)$, which we assume to be true. It remains to check this for F_{k+1}. We know that

  $$F_{k+1} = F_k + F_{k-1}.$$

  And we may use $\mathbf{P}(k)$, which we are supposing to be true to conclude that:

  $$\begin{aligned} F_{k+1} &= \frac{\alpha^k - \beta^k}{\sqrt{5}} + \frac{\alpha^{k-1} - \beta^{k-1}}{\sqrt{5}} \\ &= \frac{\alpha^k + \alpha^{k-1}}{\sqrt{5}} - \frac{\beta^k + \beta^{k-1}}{\sqrt{5}} \\ &= \frac{\alpha^{k+1} - \beta^{k+1}}{\sqrt{5}} \end{aligned}$$

  (since α and β are roots of the equation $x^2 - x - 1 = 0$)
  Hence $\mathbf{P}(k+1)$ holds.

Combining these two bullet points then shows that "$\mathbf{P}(n)$ holds, for all $n \geqslant 1$".
QED

Note: You may understand the above solution and yet wonder how such a formula could be discovered. The answer is fairly simple. There is a general theory about linear recurrence relations which guarantees that the set of all solutions of a second order recurrence (that is, a recurrence in which each term depends on the two previous terms) is "two dimensional" (that is, it is just like the familiar 2D plane, where every vector (p, q) is a combination of the two "base vectors" $(1, 0)$ and $(0, 1)$:

$$(p, q) = p(1, 0) + q(0, 1)).$$

Once we know this, it remains:

- to find two *special* solutions (like the vectors $(1,0)$ and $(0,1)$ in the plane), which we do here by looking for sequences of the form "$1, x, x^2, x^3, \ldots$" that satisfy the recurrence, which implies that $1 + x = x^2$, so $x = \alpha$, or $x = \beta$;
- then to choose a linear combination $F_m = p\alpha^m + q\beta^m$ of these two power solutions to give the correct first two terms: $0 = F_0 = p + q$, $1 = F_1 = p\alpha + q\beta$, so $p = \frac{1}{\sqrt{5}}$, $q = -\frac{1}{\sqrt{5}}$.

234. Let $\mathbf{P}(n)$ be the statement:

$$\text{“}5^{2n+1} \cdot 2^{n+2} + 3^{n+2} \cdot 2^{2n+1} \text{ is divisible by 19”.}$$

- $\mathbf{P}(0)$ is the statement: "$5 \times 4 + 9 \times 2$ is divisible by 19", which is true.
- Now suppose that we know that $\mathbf{P}(k)$ is true for some $k \geqslant 0$. We must show that $\mathbf{P}(k+1)$ is then also true.

  To prove that $\mathbf{P}(k+1)$ is true, we have to show that

  $$\text{“}5^{2k+3} \cdot 2^{k+3} + 3^{k+3} \cdot 2^{2k+3} \text{ is divisible by 19”.}$$

  $$\begin{aligned}
  5^{2k+3} \cdot 2^{k+3} + 3^{k+3} \cdot 2^{2k+3} &= 5^2 \cdot 2 \left(5^{2k+1} \cdot 2^{k+2} + 3^{k+2} \cdot 2^{2k+1}\right) \\
  &\quad - 5^2 \cdot 2 \cdot 3^{k+2} \cdot 2^{2k+1} + 3^{k+3} \cdot 2^{2k+3} \\
  &= 5^2 \cdot 2 \cdot \left(5^{2k+1} \cdot 2^{k+2} + 3^{k+2} \cdot 2^{2k+1}\right) \\
  &\quad - \left(5^2 - 3 \cdot 2\right) 3^{k+2} \cdot 2^{2k+2}.
  \end{aligned}$$

  The bracket in the first term on the RHS is divisible by 19 (by $\mathbf{P}(k)$), and the bracket in the second term is equal to 19. Hence both terms on the RHS are divisible by 19, so the RHS is divisible by 19. Therefore the LHS is also divisible by 19, so $\mathbf{P}(k+1)$ is true.

$\therefore$ $\mathbf{P}(n)$ is true for all $n \geqslant 0$. QED

235.

Note: The proofs of identities such as those in Problem **235**, which are given in many introductory texts, ignore the lessons of the previous two problems. In particular,

- they often fail to distinguish between
  - the single statement $\mathbf{P}(n)$ for a particular n, and
  - the "quantified" result to be proved ("for all $n \geqslant 1$"),

and

- they proceed in the 'wrong' direction (e.g. starting with the identity $\mathbf{P}(n)$ and "adding the same to both sides").

This latter strategy is psychologically and didactically misleading – even though it can be justified logically when proving very simple identities. In these very simple cases, each statement $\mathbf{P}(n)$ to be proved is unusual in that it refers to **exactly one configuration, or equation, for each n**. And since there is exactly one configuration for each n, the configuration or identity for $k+1$ can often be obtained by fiddling with the configuration for k. In contrast, in Problem **230**, for each value of n, there is a bewildering variety of possible polygons with n vertices, ranging from *regular* polygons to the most convoluted, re-entrant shapes: the statement $\mathbf{P}(n)$ makes an assertion about **all** such configurations, and there is no way of knowing whether we can obtain all such configurations for $k+1$ in a uniform way by fiddling with some configuration for k.

Readers should try to write each proof in the intended spirit, and to learn from the solutions – since this style has been chosen to highlight what mathematical induction is really about, and it is this approach that is needed in serious applications.

(a) Let $\mathbf{P}(n)$ be the statement:

$$\text{“}1+2+3+\cdots+n=\tfrac{n(n+1)}{2}\text{”}.$$

- LHS of $\mathbf{P}(1)=1$; RHS of $\mathbf{P}(1)=\frac{1\cdot(1+1)}{2}=1$. Since these two are equal, $\mathbf{P}(1)$ is true.
- Suppose that $\mathbf{P}(k)$ is true for some particular (unspecified) $k\geqslant 1$; that is, we know that, for this particular k,

  $$\text{“}1+2+3+\cdots+k=\tfrac{k(k+1)}{2}\text{”}.$$

  We wish to prove that $\mathbf{P}(k+1)$ must then be true.

  Now $\mathbf{P}(k+1)$ is an equation, so we start with the LHS of $\mathbf{P}(k+1)$ and try to simplify it in an appropriate way to get the RHS of $\mathbf{P}(k+1)$:

  $$\begin{aligned}\text{LHS of }\mathbf{P}(k+1) &= 1+2+3+\cdots+k+(k+1)\\ &= (1+2+3+\cdots+k)+(k+1).\end{aligned}$$

  If we now use $\mathbf{P}(k)$, which we are supposing to be true, then the first bracket is equal to $\frac{k(k+1)}{2}$, so the complete sum is equal to:

  $$\begin{aligned}&= \frac{k(k+1)}{2}+(k+1)\\ &= \frac{(k+1)(k+2)}{2}\\ &= \text{RHS of }\mathbf{P}(k+1).\end{aligned}$$

  Hence $\mathbf{P}(k+1)$ holds.

If we combine these two bullet points, then we have proved that “$\mathbf{P}(n)$ holds for all $n\geqslant 1$”. QED

(b) Let $\mathbf{P}(n)$ be the statement:

$$\text{“}\tfrac{1}{1\cdot 2}+\tfrac{1}{2\cdot 3}+\tfrac{1}{3\cdot 4}+\cdots+\tfrac{1}{n\cdot(n+1)}=1-\tfrac{1}{n+1}\text{”}.$$

- LHS of $\mathbf{P}(1) = \frac{1}{1\cdot 2} = \frac{1}{2}$; RHS of $\mathbf{P}(1) = 1 - \frac{1}{2} = \frac{1}{2}$. Since these two are equal, $\mathbf{P}(1)$ is true.
- Suppose that $\mathbf{P}(k)$ is true for some particular (unspecified) $k \geqslant 1$; that is, we know that, for this particular k,

  "$\frac{1}{1\cdot 2} + \frac{1}{2\cdot 3} + \frac{1}{3\cdot 4} + \cdots + \frac{1}{k\cdot(k+1)} = 1 - \frac{1}{k+1}$".

We wish to prove that $\mathbf{P}(k+1)$ must then be true.

Now $\mathbf{P}(k+1)$ is an equation, so we start with the LHS of $\mathbf{P}(k+1)$ and try to simplify it in an appropriate way to get the RHS of $\mathbf{P}(k+1)$:

$$\begin{aligned}
\text{LHS of } \mathbf{P}(k+1) &= \frac{1}{1\cdot 2} + \frac{1}{2\cdot 3} + \frac{1}{3\cdot 4} + \cdots + \frac{1}{(k+1)(k+2)} \\
&= \left[\frac{1}{1\cdot 2} + \frac{1}{2\cdot 3} + \frac{1}{3\cdot 4} + \frac{1}{k(k+1)}\right] \\
&\qquad + \frac{1}{(k+1)(k+2)}.
\end{aligned}$$

If we now use $\mathbf{P}(k)$, which we assume is true, then the first bracket is equal to $1 - \frac{1}{k+1}$, so the complete sum is equal to:

$$\begin{aligned}
&= \left[1 - \frac{1}{k+1}\right] + \frac{1}{(k+1)(k+2)} \\
&= 1 - \left[\frac{1}{k+1} - \frac{1}{(k+1)(k+2)}\right] \\
&= 1 - \frac{1}{k+2} \\
&= \text{RHS of } \mathbf{P}(k+1).
\end{aligned}$$

Hence $\mathbf{P}(k+1)$ holds.

If we combine these two bullet points, we have proved that "$\mathbf{P}(n)$ holds for all $n \geqslant 1$". QED

(c) **Note:** If $q = 1$, then the LHS is equal to n, but the RHS makes no sense. So we assume $q \neq 1$.

Let $\mathbf{P}(n)$ be the statement:

"$1 + q + q^2 + q^3 + \cdots + q^{n-1} = \frac{1}{1-q} - \frac{q^n}{1-q}$".

- LHS of $\mathbf{P}(1) = 1$; RHS of $\mathbf{P}(1) = \frac{1}{1-q} - \frac{q}{1-q} = 1$. Since these two are equal, $\mathbf{P}(1)$ is true.
- Suppose that $\mathbf{P}(k)$ is true for some particular (unspecified) $k \geqslant 1$; that is, we know that, for this particular k,

  "$1 + q + q^2 + q^3 + \cdots + q^{k-1} = \frac{1}{1-q} - \frac{q^k}{1-q}$".

We wish to prove that $\mathbf{P}(k+1)$ must then be true.

Now $\mathbf{P}(k+1)$ is an equation, so we start with the LHS of $\mathbf{P}(k+1)$ and try to simplify it in an appropriate way to get the RHS of $\mathbf{P}(k+1)$:

$$\begin{aligned}\text{LHS of } \mathbf{P}(k+1) &= 1+q+q^2+q^3+\cdots+q^k \\ &= \left(1+q+q^2+q^3+\cdots+q^{k-1}\right)+q^k.\end{aligned}$$

If we now use $\mathbf{P}(k)$, which we assume is true, then the first bracket is equal to

$$\frac{1}{1-q}-\frac{q^k}{1-q}$$

so the complete sum is equal to:

$$\begin{aligned} &= \frac{1}{1-q}-\left[\frac{q^k}{1-q}-q^k\right] \\ &= \frac{1}{1-q}-\frac{q^{k+1}}{1-q} \\ &= \text{RHS of } \mathbf{P}(k+1).\end{aligned}$$

Hence $\mathbf{P}(k+1)$ holds.

If we combine these two bullet points, we have proved that "$\mathbf{P}(n)$ holds for all $n \geqslant 1$". QED

(d) **Note:** The statement to be proved starts with a term involving "0!". The definition

$$n! = 1 \times 2 \times 3 \times \cdots \times n$$

does not immediately tell us how to interpret "0!". The correct interpretation emerges from the fact that several different thoughts all point in the same direction.

(i) When $n > 0$, then to get from $n!$ to $(n+1)!$ we multiply by $(n+1)$. If we extend this to $n = 0$, then "to get from 0! to 1!, we have to multiply by 1" – which suggests that $0! = 1$.

(ii) When $n > 0$, $n!$ counts the number of *permutations* of n symbols, or the number of different linear orders of n objects (i.e. how many different ways they can be arranged in a line). If we extend this to $n = 0$, we see that there is just one way to arrange 0 objects (namely, sit tight and do nothing).

(iii) The definition of $n!$ as a *product* suggests that 0! involves a "product with no terms" at all. Now when we "add no terms together" it makes sense to interpret the result as "= 0" (perhaps because if this "sum of no terms" were added to some other sum, it would make no difference). In the same spirit, the *product* of no terms should be taken to be "= 1" (since if this empty product were included at the end of some other product, it would make no difference to the result).

Let $\mathbf{P}(n)$ be the statement:

"$0 \cdot 0! + 1 \cdot 1! + 2 \cdot 2! + \cdots + (n-1) \cdot (n-1)! = n! - 1$".

- LHS of $\mathbf{P}(1) = 0 \cdot 0! = 0$; RHS of $\mathbf{P}(1) = 1! - 1 = 0$. Since these two are equal, $\mathbf{P}(1)$ is true.
- Suppose that $\mathbf{P}(k)$ is true for some particular (unspecified) $k \geqslant 1$; that is, we know that, for this particular k,

  "$0 \cdot 0! + 1 \cdot 1! + 2 \cdot 2! + \cdots + (k-1) \cdot (k-1)! = k! - 1$".

We wish to prove that $\mathbf{P}(k+1)$ must then be true.

Now $\mathbf{P}(k+1)$ is an equation, so we start with the LHS of $\mathbf{P}(k+1)$ and try to simplify it in an appropriate way to get the RHS of $\mathbf{P}(k+1)$:

$$\begin{aligned}\text{LHS of } \mathbf{P}(k+1) &= 0 \cdot 0! + 1 \cdot 1! + 2 \cdot 2! + \cdots + k \cdot k! \\ &= [0 \cdot 0! + 1 \cdot 1! + 2 \cdot 2! + \cdots + (k-1) \cdot (k-1)!] + k.k!.\end{aligned}$$

If we now use $\mathbf{P}(k)$, which we assume is true, then the first bracket is equal to $k! - 1$, so the complete sum is equal to:

$$\begin{aligned}&= (k! - 1) + k \cdot k! \\ &= (k+1) \cdot k! - 1 \\ &= (k+1)! - 1 = \text{RHS of } \mathbf{P}(k+1).\end{aligned}$$

Hence $\mathbf{P}(k+1)$ holds.

If we combine these two bullet points, we have proved that "$\mathbf{P}(n)$ holds for all $n \geqslant 1$". QED

(e) Let $\mathbf{P}(n)$ be the statement:

"$(\cos\theta + i\sin\theta)^n = \cos n\theta + i\sin n\theta$ "

- LHS of $\mathbf{P}(1) = (\cos\theta + i\sin\theta)^1$; RHS of $\mathbf{P}(1) = \cos\theta + i\sin\theta$. Since these two are equal, $\mathbf{P}(1)$ is true.
- Suppose that $\mathbf{P}(k)$ is true for some particular (unspecified) $k \geqslant 1$; that is, we know that, for this particular k,

  "$(\cos\theta + i\sin\theta)^k = \cos k\theta + i\sin k\theta$".

We wish to prove that $\mathbf{P}(k+1)$ must then be true.

Now $\mathbf{P}(k+1)$ is an equation, so we start with the LHS of $\mathbf{P}(k+1)$ and try to simplify it in an appropriate way to get the RHS of $\mathbf{P}(k+1)$:

$$\begin{aligned}\text{LHS of } \mathbf{P}(k+1) &= (\cos\theta + i\sin\theta)^{k+1} \\ &= (\cos\theta + i\sin\theta)^k(\cos\theta + i\sin\theta).\end{aligned}$$

If we now use $\mathbf{P}(k)$, which we assume is true, then the first bracket is equal to $(\cos k\theta + i\sin k\theta)$, so the complete expression is equal to:

$$\begin{aligned}&= (\cos k\theta + i\sin k\theta)(\cos\theta + i\sin\theta) \\ &= [\cos k\theta \cdot \cos\theta - \sin k\theta \cdot \sin\theta] + i[\cos k\theta \cdot \sin\theta + \sin k\theta \cdot \cos\theta] \\ &= \cos(k+1)\theta + i\sin(k+1)\theta = \text{RHS of } \mathbf{P}(k+1).\end{aligned}$$

Hence $\mathbf{P}(k+1)$ holds.

If we combine these two bullet points, we have proved that "$\mathbf{P}(n)$ holds for all $n \geqslant 1$". QED

236. Let $\mathbf{P}(n)$ be the statement:

$$\text{“}(1+2+3+\cdots+n)^2 = 1^3+2^3+3^3+\cdots+n^3\text{ ”}.$$

- LHS of $\mathbf{P}(1) = 1^2$; RHS of $\mathbf{P}(1) = 1^3$. Since these two are equal, $\mathbf{P}(1)$ is true.
- Suppose that $\mathbf{P}(k)$ is true for some particular (unspecified) $k \geqslant 1$; that is, we know that, for this particular k,

$$\text{“}(1+2+3+\cdots+k)^2 = 1^3+2^3+3^3+\cdots+k^3\text{”}.$$

We wish to prove that $\mathbf{P}(k+1)$ must then be true.

Now $\mathbf{P}(k+1)$ is an equation, so we start with one side of $\mathbf{P}(k+1)$ and try to simplify it in an appropriate way to get the other side of $\mathbf{P}(k+1)$. In this instance, the RHS of $\mathbf{P}(k+1)$ is the most promising starting point (because we know a formula for the k^{th} triangular number, and so can see how to simplify it):

$$\begin{aligned}\text{RHS of } \mathbf{P}(k+1) &= 1^3+2^3+3^3+\cdots+k^3+(k+1)^3\\ &= \left[1^3+2^3+3^3+\cdots+k^3\right]+(k+1)^3.\end{aligned}$$

If we now use $\mathbf{P}(k)$, which we assume is true, then the first bracket is equal to

$$(1+2+3+\cdots+k)^2,$$

so the complete RHS is:

$$\begin{aligned} &= (1+2+3+\cdots+k)^2+(k+1)^3\\ &= \left[\frac{k(k+1)}{2}\right]^2+(k+1)^3\\ &= \frac{1}{4}(k+1)^2\left[k^2+4k+4\right]\\ &= \left[\frac{(k+1)(k+2)}{2}\right]^2\\ &= (1+2+3+\cdots+(k+1))^2\\ &= \text{LHS of } \mathbf{P}(k+1).\end{aligned}$$

Hence $\mathbf{P}(k+1)$ holds.

If we combine these two bullet points, we have proved that "$\mathbf{P}(n)$ holds for all $n \geqslant 1$". QED

Note: A slightly different way of organizing the proof can sometimes be useful. Denote the two sides of the equation in the statement $\mathbf{P}(n)$ by $f(n)$ and $g(n)$ respectively. Then

- $f(1) = 1^2 = 1^3 = g(1)$; and

- simple algebra allows one to check that, for each $k \geqslant 1$,

$$f(k+1) - f(k) = (k+1)^3 = g(k+1) - g(k).$$

It then follows (by induction) that $f(n) = g(n)$ for all $n \geqslant 1$.

237.

(a) $T_1 = 1$, $T_1 + T_2 = 1 + 3 = 4$, $T_1 + T_2 + T_3 = 1 + 3 + 6 = 10$. These may not be very suggestive. But

$$T_1 + T_2 + T_3 + T_4 = 20 = 5 \times 4,$$

$$T_1 + T_2 + T_3 + T_4 + T_5 = 35 = 5 \times 7,$$

and

$$T_1 + T_2 + T_3 + T_4 + T_5 + T_6 = 56 = 7 \times 8$$

may eventually lead one to guess that

$$T_1 + T_2 + T_3 + \cdots + T_n = \frac{n(n+1)(n+2)}{6}.$$

Note 1: This will certainly be easier to guess if you remember what you found in Problem **17** and Problem **63**.

Note 2: There is another way to help in such guessing. Suppose you notice that

– adding values for $k = 1$ up to $k = n$ of a polynomial of degree 0 (such as $p(x) = 1$) gives an answer that is a "polynomial of degree 1",

$$1 + 1 + \cdots + 1 = n,$$

and

– adding values for $k = 1$ up to $k = n$ of a polynomial of degree 1 (such as $p(x) = x$) gives an answer that is a "polynomial of degree 2",

$$1 + 2 + 3 + \cdots + n = \frac{n(n+1)}{2}.$$

Then you might guess that the sum

$$T_1 + T_2 + T_3 + \cdots + T_n$$

will give an answer that is a polynomial of degree 3 in n. Suppose that

$$T_1 + T_2 + T_3 + \cdots + T_n = An^3 + Bn^2 + Cn + D.$$

We can then use small values of n to set up equations which must be satisfied by A, B, C, D and solve them to find A, B, C, D:

– when $n = 0$, we get $D = 0$;

– when $n = 1$, we get $A + B + C = 1$;

– when $n = 2$, we get $8A + 4B + 2C = 4$;

– when $n = 3$, we get $27A + 9B + 3C = 10$.

This method assumes that we know that the answer is a polynomial and that we know its degree: it is called "the method of undetermined coefficients".

There are various ways of improving the basic method (such as writing the polynomial $An^3 + Bn^2 + Cn + D$ in the form

$$Pn(n-1)(n-2) + Qn(n-1) + Rn + S,$$

which tailors it to the values $n = 0, 1, 2, 3$ that one intends to substitute).

(b) Let $\mathbf{P}(n)$ be the statement:

"$T_1 + T_2 + T_3 + \cdots + T_n = \frac{n(n+1)(n+2)}{6}$".

- LHS of $\mathbf{P}(1) = T_1 = 1$; RHS of $\mathbf{P}(1) = \frac{1\times2\times3}{6} = 1$. Since these two are equal, $\mathbf{P}(1)$ is true.
- Suppose that $\mathbf{P}(k)$ is true for some particular (unspecified) $k \geqslant 1$; that is, we know that, for this particular k,

$$T_1 + T_2 + T_3 + \cdots + T_k = \frac{k(k+1)(k+2)}{6}.$$

We wish to prove that $\mathbf{P}(k+1)$ must then be true.
Now $\mathbf{P}(k+1)$ is an equation, so we start with the LHS of $\mathbf{P}(k+1)$ and try to simplify it in an appropriate way to get the RHS of $\mathbf{P}(k+1)$:

$$\begin{aligned}\text{LHS of } \mathbf{P}(k+1) &= T_1 + T_2 + T_3 + \cdots + T_k + T_{k+1} \\ &= [T_1 + T_2 + T_3 + \cdots + T_k] + T_{k+1}.\end{aligned}$$

If we now use $\mathbf{P}(k)$, which we assume is true, then the first bracket is equal to

$$\frac{k(k+1)(k+2)}{6}.$$

so the complete sum is equal to:

$$\begin{aligned} &= \frac{k(k+1)(k+2)}{6} + \frac{(k+1)(k+2)}{2} \\ &= \frac{(k+1)(k+2)(k+3)}{6} \\ &= \text{RHS of } \mathbf{P}(k+1).\end{aligned}$$

Hence $\mathbf{P}(k+1)$ holds.

If we combine these two bullet points, we have proved that "$\mathbf{P}(n)$ holds for all $n \geqslant 1$". QED

Note: The triangular numbers $T_1, T_2, T_3, \ldots, T_k, \ldots T_n$ are also equal to the binomial coefficients $\binom{k+1}{2}$. And the sum of these binomial coefficients is another binomial coefficient $\binom{n+2}{3}$, so the result in Problem **237** can be written as:

$$\binom{2}{2} + \binom{3}{2} + \binom{4}{2} + \cdots + \binom{n+1}{2} = \binom{n+2}{3}.$$

You might like to interpret Problem **237** in the language of binomial coefficients, and prove it by repeated use of the basic Pascal triangle relation (Pascal (1623–1662)):

$$\binom{k}{r} + \binom{k}{r+1} = \binom{k+1}{r+1}.$$

Start by rewriting

$$\binom{n+2}{3} = \binom{n+1}{2} + \binom{n+1}{3}.$$

238.

(a) We know from Problem **237**(b) that

$$1 \cdot 2 + 2 \cdot 3 + 3 \cdot 4 + \cdots + n(n+1) = \frac{n(n+1)(n+2)}{3}.$$

Also

$$\begin{aligned} 1 \cdot 2 + 2 \cdot 3 + 3 \cdot 4 + \cdots + n(n+1) &= 1 \cdot (1+1) + 2 \cdot (2+1) + 3 \cdot (3+1) \\ &\qquad + \cdots + n \cdot (n+1) \\ &= (1^2+1) + (2^2+2) + (3^2+3) \\ &\qquad + \cdots + (n^2+n) \\ &= \left(1^2 + 2^2 + 3^2 + \cdots + n^2\right) \\ &\qquad + (1+2+3+\cdots+n). \end{aligned}$$

Therefore

$$\begin{aligned} 1^2 + 2^2 + 3^2 + \cdots + n^2 &= \frac{n(n+1)(n+2)}{3} - \frac{n(n+1)}{2} \\ &= \frac{n(n+1)(2n+1)}{6}. \end{aligned}$$

(b) **Guess:**

$$1 \cdot 2 \cdot 3 + 2 \cdot 3 \cdot 4 + 3 \cdot 4 \cdot 5 + \cdots + n(n+1)(n+2) = \frac{n(n+1)(n+2)(n+3)}{4}.$$

The proof by induction is entirely routine, and is left for the reader.

$$\begin{aligned}
1 \cdot 2 \cdot 3 + 2 \cdot 3 \cdot 4 + \cdots + n(n+1)(n+2) &= 1 \cdot (1+1)(1+2) + 2 \cdot (2+1)(2+2) \\
&\qquad + \cdots + n \cdot (n+1)(n+2) \\
&= (1^3 + 3 \cdot 1^2 + 2 \cdot 1) \\
&\qquad + (2^3 + 3 \cdot 2^2 + 2 \cdot 2) \\
&\qquad\qquad + \cdots + (n^3 + 3n^2 + 2n) \\
&= (1^3 + 2^3 + \cdots + n^3) \\
&\qquad + 3(1^2 + 2^2 + \cdots + n^2) \\
&\qquad\qquad + 2(1 + 2 + \cdots + n).
\end{aligned}$$

Therefore

$$\begin{aligned}
1^3 + 2^3 + 3^3 + \cdots + n^3 &= \frac{n(n+1)(n+2)(n+3)}{4} \\
&\qquad - 3\left[\frac{n(n+1)(2n+1)}{6}\right] \\
&\qquad - n(n+1) \\
&= \left[\frac{n(n+1)}{2}\right]^2.
\end{aligned}$$

239.

(a) Let $f(x)$ be any such polynomial. If $f(a_k) = 0$, then we know (by the *Remainder Theorem*) that $f(x)$ has $(x - a_k)$ as a factor. Since the a_k are all distinct, and $f(a_k) = 0$ for each k, $0 \leqslant k \leqslant n-1$, we have

$$f(x) = (x - a_0)(x - a_1)(x - a_2) \cdots (x - a_{n-1}) \cdot g(x)$$

for some polynomial $g(x)$. And since we are told that $f(x)$ has degree n, $g(x)$ has degree 0, so is a constant. Hence every such polynomial of degree n has the form

$$C \cdot (x - a_0)(x - a_1)(x - a_2) \cdots (x - a_{n-1}).$$

Since $f(a_n) = b$, we can substitute to find C:

$$C = \frac{b}{(a_n - a_0)(a_n - a_1)(a_n - a_2) \cdots (a_n - a_{n-1})}.$$

(b) Let $\mathbf{P}(n)$ be the statement:

> "Given any two labelled sets of $n+1$ real numbers $a_0, a_1, a_2, \ldots, a_n$, and $b_0, b_1, b_2, \ldots, b_n$, where the a_i are all distinct (but the b_i need not be), there exists a polynomial f_n of degree n, such that $f_n(a_0) = b_0$, $f_n(a_1) = b_1$, $f_n(a_2) = b_2$, $\ldots$, $f_n(a_n) = b_n$."

- When $n = 0$, we may choose $f_0(x) = b_0$ to be the constant polynomial. Hence $\mathbf{P}(0)$ is true.

- Suppose that $\mathbf{P}(k)$ is true for some particular (unspecified) $k \geqslant 0$; that is, we know that, for this particular k:

  "Given any two labelled sets of $k+1$ real numbers $a_0, a_1, a_2, \ldots, a_k$, and $b_0, b_1, b_2, \ldots, b_k$, where the a_i are all distinct (but the b_i need not be), there exists a polynomial f_k of degree k, such that $f_k(a_0) = b_0$, $f_k(a_1) = b_1$, $f_k(a_2) = b_2, \ldots, f_k(a_k) = b_k$."

We wish to prove that $\mathbf{P}(k+1)$ must then be true.

Now $\mathbf{P}(k+1)$ is the statement:

"Given any two labelled sets of $(k+1)+1$ real numbers $a_0, a_1, \ldots, a_{k+1}$, and $b_0, b_1, b_2, \ldots, b_{k+1}$, where the a_i are all distinct (but the b_i need not be), there exists a polynomial f_{k+1} of degree $k+1$, such that

$$f_{k+1}(a_0) = b_0,\ f_{k+1}(a_1) = b_1,\ f_{k+1}(a_2) = b_2,\ \ldots,\ f_{k+1}(a_{k+1}) = b_{k+1}."$$

So to prove that $\mathbf{P}(k+1)$ holds, we must start by considering

any two labelled sets of $(k+1)+1$ real numbers

$$a_0, a_1, a_2, \ldots, a_{k+1}, \text{ and } b_0, b_1, b_2, \ldots, b_{k+1},$$

where the a_i are all distinct (but the b_i need not be).

We must then somehow construct a polynomial function f_{k+1} of degree $k+1$ with the required property.

Because we are supposing that $\mathbf{P}(k)$ is known to be true, we can focus on the first $k+1$ numbers in each of the two lists – $a_0, a_1, a_2, \ldots, a_k$, and $b_0, b_1, b_2, \ldots, b_k$ – and we can then be sure that there is a polynomial f_k of degree k such that

$$f_k(a_0) = b_0, f_k(a_1) = b_1, f_k(a_2) = b_2, \ldots, f_k(a_k) = b_k.$$

The next step is slightly indirect: we make use of the polynomial f_{k+1} *which we are still trying to construct*, and focus on the polynomial

$$f(x) = f_{k+1}(x) - f_k(x)$$

satisfying

$$f(a_0) = f(a_1) = \cdots = f(a_k) = 0, f(a_{k+1}) = b_{k+1} - f_k(a_{k+1}) = b \quad \text{(say)}.$$

Part (a) guarantees the existence of such a polynomial $f(x)$ of degree $k+1$ and tells us exactly what this polynomial function $f(x)$ is equal to. Hence we can construct the required polynomial $f_{k+1}(x)$ by setting it equal to $f(x) + f_k(x)$, which proves that $\mathbf{P}(k+1)$ is true.

If we combine these two bullet points, we have proved that "$\mathbf{P}(n)$ holds for all $n \geqslant 1$". QED

240.

(a) 5 cents cannot be made; $6 = 3+3$; $7 = 3+4$; $8 = 4+4$; $9 = 3+3+3$; etc.

Guess: Every amount $> N = 5$ can be paid directly (without receiving change).

(b) Let $\mathbf{P}(n)$ be the statement:

"n cents can be paid directly (without change) using 3 cent and 4 cent coins".

- $\mathbf{P}(6)$ is the statement: "6 cents can be paid directly". And $6 = 3 + 3$, so $\mathbf{P}(6)$ is true.
- Now suppose that we know $\mathbf{P}(k)$ is true for some $k \geqslant 6$. We must show that $\mathbf{P}(k+1)$ must then be true.

To prove $\mathbf{P}(k+1)$ we consider the statement $\mathbf{P}(k+1)$:

"$k + 1$ cents can be paid directly".

We know that $\mathbf{P}(k)$ is true, so we know that "k cents can be paid directly".

- If a direct method of paying k cents involves at least one 3 cent coin, then we can replace one 3 cent coin by a 4 cent coin to produce a way of paying $k + 1$ cents.

Hence we only need to worry about a situation in which the only way to pay k cents directly involves no 3 cent coins at all – that is, paying k cents uses only 4 cent coins. But then there must be at least two 4 cent coins (since $k \geqslant 6$), and we can replace two 4 cent coins by three 3 cent coins to produce a way of paying $k + 1$ cents directly.

Hence

- $\mathbf{P}(6)$ is true; and
- whenever $\mathbf{P}(k)$ is true for some $k \geqslant 6$, we know that $\mathbf{P}(k+1)$ is also true.

$\therefore$ $\mathbf{P}(n)$ is true for all $n \geqslant 6$. QED

241.

(a) $z^2 - z + 1 = 0$, so $z = \frac{1 \pm \sqrt{-3}}{2}$ (these are the two primitive sixth roots of unity).
$\therefore z^2 = \frac{-1 \pm \sqrt{-3}}{2}$ (the two primitive cube toots of unity), and

$$z^2 + \frac{1}{z^2} = -1.$$

(b) $z^2 - 2z + 1 = 0$, so $z = 1$ (repeated root). $\therefore z^2 = 1$ and $z^2 + \frac{1}{z^2} = 2$.

(c), (d) $z^2 - 3z + 1 = 0$, so $z = \frac{3 \pm \sqrt{5}}{2}$.
As soon as one starts calculating z^2 and $\frac{1}{z^2}$, it becomes clear that it is time to p-a-u-s-e and think.

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

so whenever $z + \frac{1}{z} = k$ is an integer,

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

is also an integer.

(e) Let $\mathbf{P}(n)$ be the statement:

"if z has the property that $z + \frac{1}{z}$ is an integer, then $z^m + \frac{1}{z^m}$ is also an integer for all m, $0 \leqslant m \leqslant n$".

- $\mathbf{P}(0)$ and $\mathbf{P}(1)$ are clearly both true; and $\mathbf{P}(2)$ was proved in part (d) above.
- Suppose that $\mathbf{P}(k)$ is true for some particular (unspecified) $k \geqslant 2$; that is, we know that, for this particular k:

  "if z has the property that $z + \frac{1}{z}$ is an integer, then $z^m + \frac{1}{z^m}$ is also an integer for all m, $0 \leqslant m \leqslant k$".

We wish to prove that $\mathbf{P}(k+1)$ must then be true.

If $z + \frac{1}{z}$ is an integer, then, by $\mathbf{P}(k)$,

"$z^m + \frac{1}{z^m}$ is also an integer for all m, $0 \leqslant m \leqslant k$".

So to prove that $\mathbf{P}(k+1)$ holds, we only need to show that

"$z^{k+1} + \frac{1}{z^{k+1}}$ is an integer".

By the Binomial Theorem:

$$\begin{aligned}\left(z + \frac{1}{z}\right)^{k+1} &= \left(z^{k+1} + \frac{1}{z^{k+1}}\right) + \binom{k+1}{1}\left(z^{k-1} + \frac{1}{z^{k-1}}\right) \\ &\quad + \binom{k+1}{2}\left(z^{k-3} + \frac{1}{z^{k-3}}\right) + \cdots\end{aligned}$$

The LHS is an integer (since $z + \frac{1}{z}$ is an integer), and (by $\mathbf{P}(k)$) every term on the RHS is an integer except possibly the first. Hence the first term is the difference of two integers, so must also be an integer.

Hence $\mathbf{P}(k+1)$ is true.

If we combine these two bullet points, we have proved that "$\mathbf{P}(n)$ holds for all $n \geqslant 1$". QED

Note: If $k+1 = 2m$ is even, the expansion of $\left(z + \frac{1}{z}\right)^{k+1}$ has an odd number of terms, so the RHS of the above re-grouped expansion ends with the term $\binom{2m}{m} \cdot z^m \cdot \left(\frac{1}{z}\right)^m$, which is also an integer.

242.

Note: In the solution to Problem **241** we included the condition on z as part of the statement $\mathbf{P}(n)$.

In Problem **242** the result to be proved has a similar background hypothesis – "Let p be a prime number". It may make the induction clearer if, as in the statement of the Problem, this hypothesis is stated *before* starting the induction proof.

Let p be any prime number. We let $\mathbf{P}(n)$ be the statement:

"$n^p - n$ is divisible by p ".

- $\mathbf{P}(1)$ is true (since $1^p - 1 = 0 = 0 \times p$, which is divisible by p).
- Suppose that $\mathbf{P}(k)$ is true for some particular (unspecified) $k \geqslant 1$; that is, we know that, for this particular k:

  "$k^p - k$ is divisible by p".

We wish to prove $\mathbf{P}(k+1)$ – that is,

"$(k+1)^p - (k+1)$ is divisible by p"

must then be true. Using the Binomial Theorem again we see that

$$\begin{aligned}(k+1)^p - (k+1) &= \left[k^p + \binom{p}{p-1}k^{p-1} + \binom{p}{p-2}k^{p-2} + \cdots + \binom{p}{1}k + 1\right] \\ &\qquad -(k+1) \\ &= (k^p - k) + \left[\binom{p}{p-1}k^{p-1} + \binom{p}{p-2}k^{p-2} + \cdots + \binom{p}{1}k\right].\end{aligned}$$

By $\mathbf{P}(k)$, the first bracket on the RHS is divisible by p; and in each of the other terms each of the binomial coefficients $\binom{p}{r}$, $0 < r < p$,

- is an integer, and
- has a factor "p" in the numerator and no such factor in the denominator.

Hence each term in the second bracket is a multiple of p. So the RHS (and hence the LHS) is divisible by p.

Hence $\mathbf{P}(k+1)$ is true.

If we combine these two bullet points, we have proved that "$\mathbf{P}(n)$ holds for all $n \geqslant 1$". QED

243.

$$0.037037037\ldots \text{ (for ever)} = \frac{37}{1000} + \frac{37}{1\,000\,000} + \frac{37}{1\,000\,000\,000} + \cdots \text{ (for ever)}.$$

This is a geometric series with first term $a = \frac{37}{1000}$ and common ratio $r = \frac{1}{1000}$, and so has sum

$$\frac{a}{1-r} = \frac{37}{999} = \frac{1}{27}.$$

244.

(a) Each person receives in total:

$$\frac{1}{4} + \left(\frac{1}{4}\right)^2 + \left(\frac{1}{4}\right)^3 + \left(\frac{1}{4}\right)^4 + \cdots \text{ (for ever)} = \frac{1}{3}$$

(here $a = \frac{1}{4} = r$).

(b) You have

$$1 - \frac{1}{2} + \frac{1}{4} - \frac{1}{8} + \cdots \text{ (for ever) } = \frac{2}{3}$$

(here $a = 1$, $r = -\frac{1}{2}$); I have

$$\frac{1}{2} - \frac{1}{4} + \frac{1}{8} - \cdots \text{ (for ever) } = \frac{1}{3}$$

(here $a = \frac{1}{2}$, $r = -\frac{1}{2}$).

245. The trains are 120 km apart, and the fly travels at 50 km/h towards Train B, which is initially 120 km away and travelling at 30 km/h.

The relative speed of the fly and Train B is 80 km/h, so it takes $\frac{3}{2}$ hours before they meet. In this time Train A and Train B have each travelled 45 km, so they are now 30 km apart. The fly then turns right round and flies back to Train A.

The relative speed of the fly and Train A is then also 80 km/h, so it takes just $\frac{3}{8}$ hours (or 22.5 minutes) for the fly to return to Train A. Train A and Train B have each travelled $\frac{45}{4}$ km in this time, so they are now $\frac{30}{4}$ km apart. The fly then turns round and flies straight back to Train B.

Train B is $\frac{30}{4}$ km away and the relative speed of the fly and Train B is again 80 km/h, so the journey takes $\frac{3}{32}$ hours (or 5.625 minutes).

Continuing in this way, we see that the fly takes

$$\frac{3}{2} + \frac{3}{8} + \frac{3}{32} + \frac{3}{128} + \cdots \text{ (for ever) } = 2 \text{ hours.}$$

Hence the fly travels 100 km before being squashed.

Note: The two trains are approaching each other at 60 km/h, so they crash in exactly 2 hours – during which time the fly travels 100 km.

246.

(a)(i) $3^2 > 2^2$; therefore

$$\frac{1}{3^2} < \frac{1}{2^2},$$

so

$$\frac{1}{2^2} + \frac{1}{3^2} < \frac{2}{2^2} = \frac{1}{2}.$$

(ii) $7^2 > 6^2 > 5^2 > 4^2$; therefore

$$\frac{1}{7^2} < \frac{1}{6^2} < \frac{1}{5^2} < \frac{1}{4^2},$$

so

$$\frac{1}{4^2} + \frac{1}{5^2} + \frac{1}{6^2} + \frac{1}{7^2} < \frac{4}{4^2} = \frac{1}{4}.$$

(b) The argument in part (a) gives an upper bound for each bracketed expression in the sum

$$\left(\frac{1}{1^2}\right)+\left(\frac{1}{2^2}+\frac{1}{3^2}\right)+\left(\frac{1}{4^2}+\frac{1}{5^2}+\frac{1}{6^2}+\frac{1}{7^2}\right)+\left(\frac{1}{8^2}+\cdots+\frac{1}{15^2}\right)+\cdots$$

Replacing each bracket by its upper bound, we see that the sum is

$$\begin{aligned} &< \frac{1}{1^2}+\frac{2}{2^2}+\frac{4}{4^2}+\frac{8}{8^2}+\cdots \\ &= 1+\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\cdots \text{ (for ever)} \\ &= 2. \end{aligned}$$

(c) The finite partial sums

$$S_n = \frac{1}{1^2}+\frac{1}{2^2}+\frac{1}{3^2}+\cdots+\frac{1}{n^2}$$

– **increase steadily** as we take more and more terms, and
– every partial sum S_n is **less than 2**.

It is clear that these partial sums form a sequence

$$1 = S_1 < S_2 < S_3 < \cdots < S_n < S_{n+1} < \cdots < 2.$$

It follows that there is some (unknown) number $S \leqslant 2$ to which the partial sums converge as $n \to \infty$, and we take this (unknown) exact value S to be the exact value of the endless sum

$$\frac{1}{1^2}+\frac{1}{2^2}+\frac{1}{3^2}+\cdots+\frac{1}{n^2}+\cdots \text{ (for ever)}$$

To see, for example, that $S > \frac{17}{12}$, notice that

$$\begin{aligned} S &> S_4 \\ &= \frac{1}{1^2}+\frac{1}{2^2}+\frac{1}{3^2}+\frac{1}{4^2} \\ &= 1+\frac{1}{4}+\frac{1}{9}+\frac{1}{16} \\ &> \frac{17}{12}. \end{aligned}$$

Note 1: The claim that

> "an increasing sequence of partial sums S_n, all less than 2, must converge to some number $S \leqslant 2$"

is a fundamental property of the real numbers – called *completeness*.

Note 2: Just as one can obtain better and better lower bounds for S – like "$\frac{17}{12} < S$", so one can improve the upper bound "$S < 2$". For example, if in part (b) we avoid replacing the third term $\frac{1}{9}$ by $\frac{1}{4}$, we get a better upper bound "$S < \frac{67}{36}$", which is $\frac{5}{36}$ less than 2.

247.

(a) Let $\mathbf{P}(n)$ be the statement:

"$\frac{1}{1^2}+\frac{1}{2^2}+\frac{1}{3^2}+\cdots+\frac{1}{n^2}\leqslant 2-\frac{1}{n}$".

- Then LHS of $\mathbf{P}(1)=\frac{1}{1^2}=1$, and RHS of $\mathbf{P}(1)=2-1=1$. Hence $\mathbf{P}(1)$ is true.
- Suppose we know that $\mathbf{P}(k)$ is true for some $k\geqslant 1$. We want to prove that $\mathbf{P}(k+1)$ holds.

$$\begin{aligned}
\text{LHS of } \mathbf{P}(k+1) &= \frac{1}{1^2}+\frac{1}{2^2}+\frac{1}{3^2}+\cdots+\frac{1}{k^2}+\frac{1}{(k+1)^2}\\
&= \left[\frac{1}{1^2}+\frac{1}{2^2}+\frac{1}{3^2}+\cdots+\frac{1}{k^2}\right]+\frac{1}{(k+1)^2}\\
&\leqslant \left[2-\frac{1}{k}\right]+\frac{1}{(k+1)^2}\\
&= 2-\left[\frac{1}{k}-\frac{1}{(k+1)^2}\right]\\
&< 2-\frac{1}{k+1}.
\end{aligned}$$

Hence $\mathbf{P}(k+1)$ holds.
$\therefore$ $\mathbf{P}(1)$ holds; and whenever $\mathbf{P}(k)$ is known to be true, $\mathbf{P}(k+1)$ is also true.
$\therefore$ $\mathbf{P}(n)$ is true, for all $n\geqslant 1$. QED

(b) Let $\mathbf{P}(n)$ be the statement:

"$\frac{1}{1^2}+\frac{1}{2^2}+\frac{1}{3^2}+\cdots+\frac{1}{n^2}<1.68-\frac{1}{n}$".

- Then
$$\text{LHS of } \mathbf{P}(4)=\frac{1}{1^2}+\frac{1}{2^2}+\frac{1}{3^2}+\frac{1}{4^2}=1.423611111\cdots,$$
and RHS of $\mathbf{P}(4)=1.43$. Hence $\mathbf{P}(4)$ is true.
- Suppose we know that $\mathbf{P}(k)$ is true for some $k\geqslant 4$. We want to prove that $\mathbf{P}(k+1)$ holds.

$$\begin{aligned}
\text{LHS of } \mathbf{P}(k+1) &= \frac{1}{1^2}+\frac{1}{2^2}+\frac{1}{3^2}+\cdots+\frac{1}{k^2}+\frac{1}{(k+1)^2}\\
&= \left(\frac{1}{1^2}+\frac{1}{2^2}+\frac{1}{3^2}+\cdots+\frac{1}{k^2}\right)+\frac{1}{(k+1)^2}\\
&< \left[1.68-\frac{1}{k}\right]+\frac{1}{(k+1)^2}\\
&= 1.68-\left[\frac{1}{k}-\frac{1}{(k+1)^2}\right]\\
&< 1.68-\frac{1}{k+1}.
\end{aligned}$$

Hence $\mathbf{P}(k+1)$ holds.
$\therefore$ $\mathbf{P}(1)$ holds; and whenever $\mathbf{P}(k)$ is known to be true, $\mathbf{P}(k+1)$ is also true.

$\therefore$ $\mathbf{P}(n)$ is true, for all $n \geqslant 1$. QED

248.

(a)(i) $n! = n \times (n-1) \times (n-2) \times \cdots \times 3 \times 2 \times 1 \geqslant 2 \times 2 \times 2 \times \cdots \times 2 \times 1 = 2^{n-1}$ whenever $n \geqslant 1$.

$\therefore$ $\frac{1}{n!} \leqslant \left(\frac{1}{2}\right)^{n-1}$ for all $n \geqslant 1$.

$\therefore$ $\frac{1}{0!} + \frac{1}{1!} + \frac{1}{2!} + \cdots + \frac{1}{n!} \leqslant 1 + \left[1 + \frac{1}{2} + \left(\frac{1}{2}\right)^2 + \cdots + \left(\frac{1}{2}\right)^{n-1}\right] < 3$ for all $n \geqslant 0$.

(ii) As we go on adding more and more terms, each finite sum

$$\frac{1}{0!} + \frac{1}{1!} + \frac{1}{2!} + \cdots + \frac{1}{n!}$$

is bigger than the previous sum. Since every finite sum

$$\frac{1}{0!} + \frac{1}{1!} + \frac{1}{2!} + \cdots + \frac{1}{n!} < 3,$$

the sums increase, but never reach 3, so they accumulate closer and closer to a value "e" $\leqslant 3$. Moreover, this value "e" is certainly larger than the sum of the first two terms $\frac{1}{0!} + \frac{1}{1!} = 2$, so $2 < e \leqslant 3$.

(b)(i) Let $\mathbf{P}(n)$ be the statement:

"$\frac{1}{0!} + \frac{1}{1!} + \frac{1}{2!} + \cdots + \frac{1}{n!} \leqslant 3 - \frac{1}{n \cdot n!}$".

- LHS of $\mathbf{P}(1) = 2 =$ RHS of $\mathbf{P}(1)$. Hence $\mathbf{P}(1)$ is true.
- Suppose we know that $\mathbf{P}(k)$ is true for some $k \geqslant 1$. We want to prove that $\mathbf{P}(k+1)$ holds.

$$\begin{aligned}
\text{LHS of } \mathbf{P}(k+1) &= \frac{1}{0!} + \frac{1}{1!} + \frac{1}{2!} + \cdots + \frac{1}{(k+1)!} \\
&= \left[\frac{1}{0!} + \frac{1}{1!} + \frac{1}{2!} + \cdots + \frac{1}{k!}\right] + \frac{1}{(k+1)!} \\
&\leqslant 3 - \frac{1}{k \cdot k!} + \frac{1}{(k+1)!} \\
&= 3 - \frac{1}{k(k+1)!} \\
&< 3 - \frac{1}{(k+1) \cdot (k+1)!}
\end{aligned}$$

Hence $\mathbf{P}(k+1)$ holds.

$\therefore$ $\mathbf{P}(1)$ holds; and whenever $\mathbf{P}(k)$ is known to be true, $\mathbf{P}(k+1)$ is also true.

$\therefore$ $\mathbf{P}(n)$ is true, for all $n \geqslant 1$. QED

(ii) [The reasoning here uses the constant "3" while ignoring the refinement "$3 - \frac{1}{n.n!}$", and so sounds exactly like part (a)(ii).] As we add more terms, each finite sum

$$\frac{1}{0!} + \frac{1}{1!} + \frac{1}{2!} + \cdots + \frac{1}{n!}$$

is bigger than the previous sum. Since every finite sum

$$\frac{1}{0!} + \frac{1}{1!} + \frac{1}{2!} + \cdots + \frac{1}{n!} < 3,$$

the partial sums increase, but never reach 3; so they accumulate closer and closer to a value "e" $\leqslant 3$. Moreover, this value "e" is certainly larger than the sum of the first three terms $\frac{1}{0!} + \frac{1}{1!} + \frac{1}{2!} = 2.5$, so $2.5 < e \leqslant 3$.

(c) **Note:** Examine carefully the role played by the number "3" in the above induction proof in (b)(ii). It is needed precisely to validate the statement $\mathbf{P}(1)$: since $\frac{1}{0!} + \frac{1}{1!} = 3 - \frac{1}{1\times 1!}$". But the number "3" plays no active part in the second induction step, and could be replaced by any other number we choose.

The exact value "e" of the infinite series is not really affected by what happens when $n = 1$. Suppose we ask: "What number C_2 should replace "3" if we only want to prove that

$$\frac{1}{0!} + \frac{1}{1!} + \frac{1}{2!} + \cdots + \frac{1}{n!} \leqslant C_2 - \frac{1}{n \cdot n!}, \text{ for all } n \geqslant 2?$$

The only part of the induction proof where C_2 then matters is when we try to check that $\mathbf{P}(2)$ holds; so we must choose the smallest possible C_2 to satisfy

$$\frac{1}{0!} + \frac{1}{1!} + \frac{1}{2!} \leqslant C_2 - \frac{1}{2.2!}:$$

that is, $C_2 = 2.75$.

(i) Let $\mathbf{P}(n)$ be the statement:
"$\frac{1}{0!} + \frac{1}{1!} + \frac{1}{2!} + \cdots + \frac{1}{n!} \leqslant 2.75 - \frac{1}{n \cdot n!}$".

- LHS of $\mathbf{P}(2) = 2.5$; RHS of $\mathbf{P}(2) = 2.75 - \frac{1}{4}$. Hence $\mathbf{P}(2)$ is true.
- Suppose we know that $\mathbf{P}(k)$ is true for some $k \leqslant 2$. We want to prove that $\mathbf{P}(k+1)$ holds.

$$\begin{aligned}
\text{LHS of } \mathbf{P}(k+1) &= \frac{1}{0!} + \frac{1}{1!} + \frac{1}{2!} + \cdots + \frac{1}{(k+1)!} \\
&= \left[\frac{1}{0!} + \frac{1}{1!} + \frac{1}{2!} + \cdots + \frac{1}{k!}\right] + \frac{1}{(k+1)!} \\
&\leqslant 2.75 - \frac{1}{k \cdot k!} + \frac{1}{(k+1)!} \\
&= 2.75 - \frac{1}{k(k+1)!} \\
&< 2.75 - \frac{1}{(k+1)(k+1)!}
\end{aligned}$$

Hence $\mathbf{P}(k+1)$ holds.
$\therefore$ $\mathbf{P}(2)$ holds; and whenever $\mathbf{P}(k)$ is known to be true, $\mathbf{P}(k+1)$ is also true.
$\therefore$ $\mathbf{P}(n)$ is true, for all $n \geqslant 2$. QED

(ii) As we add more terms, each finite sum

$$\frac{1}{0!}+\frac{1}{1!}+\frac{1}{2!}+\cdots+\frac{1}{n!}$$

is bigger than the previous sum.
Since every finite sum

$$\frac{1}{0!}+\frac{1}{1!}+\frac{1}{2!}+\cdots+\frac{1}{n!}<2.75,$$

the finite sums increase, but never reach 2.75, so they accumulate closer and closer to a value "e" $\leqslant 2.75$. Moreover, this value "e" is certainly larger than the sum of the first four terms

$$\frac{1}{0!}+\frac{1}{1!}+\frac{1}{2!}+\frac{1}{3!}>2.66,$$

so $2.66 < e \leqslant 2.75$.

(d)(i) Let $\mathbf{P}(n)$ be the statement:
"$\frac{1}{0!}+\frac{1}{1!}+\frac{1}{2!}+\cdots+\frac{1}{n!}\leqslant 2.7222\cdots(\text{for ever})-\frac{1}{n.n!}$".

- LHS of $\mathbf{P}(3)=\frac{1}{0!}+\frac{1}{1!}+\frac{1}{2!}+\frac{1}{3!}=2.666\cdots(\text{for ever})$;
  RHS of $\mathbf{P}(3)=2.7222\cdots(\text{for ever})\ -\frac{1}{18}=2.666\cdots(\text{for ever})$.
  Hence $\mathbf{P}(3)$ is true.
- Suppose we know that $\mathbf{P}(k)$ is true for some $k\geqslant 3$. We want to prove that $\mathbf{P}(k+1)$ holds.

$$\begin{aligned}
\text{LHS of }\mathbf{P}(k+1) &= \frac{1}{0!}+\frac{1}{1!}+\frac{1}{2!}+\cdots+\frac{1}{(k+1)!}\\
&= \left[\frac{1}{0!}+\frac{1}{1!}+\frac{1}{2!}+\cdots+\frac{1}{k!}\right]+\frac{1}{(k+1)!}\\
&\leqslant 2.7222\cdots(\text{for ever})-\frac{1}{k\cdot k!}+\frac{1}{(k+1)!}\\
&= 2.7222\cdots(\text{for ever})-\frac{1}{k(k+1)!}\\
&< 2.7222\cdots(\text{for ever})-\frac{1}{(k+1)(k+1)!}
\end{aligned}$$

Hence $\mathbf{P}(k+1)$ holds.
$\therefore$ $\mathbf{P}(3)$ holds; and whenever $\mathbf{P}(k)$ is known to be true, $\mathbf{P}(k+1)$ is also true.
$\therefore$ $\mathbf{P}(n)$ is true, for all $n\geqslant 3$. QED

(ii) As we add more terms, each finite sum

$$\frac{1}{0!}+\frac{1}{1!}+\frac{1}{2!}+\cdots+\frac{1}{n!}$$

is bigger than the previous sum.

Since every finite sum

$$\frac{1}{0!} + \frac{1}{1!} + \frac{1}{2!} + \cdots + \frac{1}{n!} < 2.7222\cdots \text{ (for ever)},$$

the finite sums increase, but never reach $2.7222\cdots$ (for ever), so they accumulate closer and closer to a value "e" $\leqslant 2.7222\cdots$ (for ever). Moreover, this value "e" is certainly larger than the sum of the first five terms

$$\frac{1}{0!} + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} > 2.708,$$

so $2.708 < e \leqslant 2.7222\cdots$ (for ever).

Note: This process of refinement can continue indefinitely. But we only have to go one further step to pin down the value of "e" with surprising accuracy. The next step uses the same proof to show that

$$\text{“}\tfrac{1}{0!} + \tfrac{1}{1!} + \tfrac{1}{2!} + \cdots + \tfrac{1}{n!} \leqslant 2.7185 - \tfrac{1}{n \cdot n!}, \text{ for all } n \geqslant 4\text{”},$$

and to conclude that the endless sum

$$\frac{1}{0!} + \frac{1}{1!} + \frac{1}{2!} + \cdots + \frac{1}{n!} + \cdots \text{ (for ever)}$$

has a definite value "e" that lies somewhere between 2.716 and 2.71875.

We could then repeat the same proof to show that

$$\frac{1}{0!} + \frac{1}{1!} + \frac{1}{2!} + \cdots + \frac{1}{n!} \leqslant 2.718333\cdots \text{(for ever)} - \frac{1}{n \cdot n!}, \text{ for all } n \geqslant 5,$$

and use the lower bound $2.7177\ldots$ from the first seven terms to conclude that the endless sum

$$\frac{1}{0!} + \frac{1}{1!} + \frac{1}{2!} + \cdots + \frac{1}{n!} + \cdots \text{ (for ever)}$$

has a definite value "e" that lies somewhere between 2.7177 and $2.718333\cdots$ (for ever). And so on.

249. Let $\mathbf{P}(n)$ be the statement:

$$\text{“}\tfrac{1}{\sqrt{1}} + \tfrac{1}{\sqrt{2}} + \tfrac{1}{\sqrt{3}} + \cdots + \tfrac{1}{\sqrt{n}} \geqslant \sqrt{n}\text{”}.$$

- LHS of $\mathbf{P}(1) = 1 =$ RHS of $\mathbf{P}(1)$. Hence $\mathbf{P}(1)$ is true.

- Suppose we know that $\mathbf{P}(k)$ is true for some $k \geqslant 1$. We want to prove that $\mathbf{P}(k+1)$ holds.

$$\begin{aligned}
\text{LHS of } \mathbf{P}(k+1) &= \frac{1}{\sqrt{1}} + \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{3}} + \cdots + \frac{1}{\sqrt{k+1}} \\
&= \left(\frac{1}{\sqrt{1}} + \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{3}} + \cdots + \frac{1}{\sqrt{k}}\right) + \frac{1}{\sqrt{(k+1)}} \\
&\geqslant \sqrt{k} + \frac{1}{\sqrt{k+1}} \\
&\geqslant \sqrt{k+1} \quad \left(\text{since } \frac{1}{\sqrt{k+1}} \geqslant \frac{1}{\sqrt{k+1}+\sqrt{k}}\right).
\end{aligned}$$

Hence $\mathbf{P}(k+1)$ holds.

$\therefore$ $\mathbf{P}(1)$ holds; and whenever $\mathbf{P}(k)$ is known to be true, $\mathbf{P}(k+1)$ is also true.
$\therefore$ $\mathbf{P}(n)$ is true, for all $n \geqslant 1$. QED

250. Let a, b be real numbers such that $a \neq b$, and $a+b > 0$.

One of a, b is then the greater, and we may suppose this is a – so that $a > b$. If $a > b > 0$, then $a^n > b^n > 0$ for all n; if $b < 0$, then $a+b > 0$ implies that $a = |a| > |b|$, so $a^n > b^n$ for all n.

Let $\mathbf{P}(n)$ be the statement:

"$\frac{a^n+b^n}{2} \geqslant \left(\frac{a+b}{2}\right)^n$".

- LHS of $\mathbf{P}(1) = \frac{a+b}{2}$ = RHS of $\mathbf{P}(1)$. Hence $\mathbf{P}(1)$ is true.
- Suppose we know that $\mathbf{P}(k)$ is true for some $k \geqslant 1$. We want to prove that $\mathbf{P}(k+1)$ holds.

$$\begin{aligned}
\text{RHS of } \mathbf{P}(k+1) &= \left(\frac{a+b}{2}\right)^{k+1} \\
&= \frac{a+b}{2} \cdot \left(\frac{a+b}{2}\right)^{k} \\
&\leqslant \frac{a+b}{2} \cdot \frac{a^k+b^k}{2} \quad (\text{by } \mathbf{P}(k)) \\
&= \frac{a^{k+1}+b^{k+1}}{4} + \frac{ab^k+ba^k}{4} \\
&< \frac{a^{k+1}+b^{k+1}}{2} \quad (\text{since } (a^k-b^k)(a-b) > 0).
\end{aligned}$$

Hence $\mathbf{P}(k+1)$ holds.
$\therefore$ $\mathbf{P}(1)$ holds; and whenever $\mathbf{P}(k)$ is known to be true, $\mathbf{P}(k+1)$ is also true.
$\therefore$ $\mathbf{P}(n)$ is true, for all $n \geqslant 1$. QED

251. Let x be any real number $\geqslant -1$.

If $x = -1$, then $(1+x)^n = 0 \geqslant 1 - n = 1 + nx$, for all $n \geqslant 1$.

Thus we may assume that $x > -1$, so $1 + x > 0$.

Let $\mathbf{P}(n)$ be the statement: "$(1+x)^n \geqslant 1 + nx$".

- LHS of $\mathbf{P}(1) = 1 + x =$ RHS of $\mathbf{P}(1)$. Hence $\mathbf{P}(1)$ is true.
- Suppose we know that $\mathbf{P}(k)$ is true for some $k \geqslant 1$. We want to prove that $\mathbf{P}(k+1)$ holds.

$$\begin{aligned}
\text{LHS of } \mathbf{P}(k+1) &= (1+x)^{k+1} \\
&= (1+x)\cdot(1+x)^k \\
&\geqslant (1+x)\cdot(1+kx) \quad (\text{by } \mathbf{P}(k), \quad \text{since } 1+x>0) \\
&= 1 + (k+1)x + kx^2 \\
&\geqslant 1 + (k+1)x
\end{aligned}$$

Hence $\mathbf{P}(k+1)$ holds.
$\therefore$ $\mathbf{P}(1)$ holds; and whenever $\mathbf{P}(k)$ is known to be true, $\mathbf{P}(k+1)$ is also true.
$\therefore$ $\mathbf{P}(n)$ is true, for all $n \geqslant 1$. QED

252. The problem is discussed after the statement of Problem **252** in the main text.

253.

(a)(i) $3 > 2$, so $\frac{1}{3} < \frac{1}{2}$.
$\therefore \frac{1}{2} + \frac{1}{3} < \frac{1}{2} + \frac{1}{2} = 1$.

(ii) $5, 6, 7 > 4$; hence $\frac{1}{5}, \frac{1}{6}, \frac{1}{7}$ are all $< \frac{1}{4}$.
$\therefore \frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \frac{1}{7} < \frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = 1$.

(iii) Let $\mathbf{P}(n)$ be the statement:
"$\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{2^n - 1} < n$".
Then

- $\mathbf{P}(2)$ is true by (i), since

$$\frac{1}{1} + \frac{1}{2} + \frac{1}{3} < 1 + \left(\frac{1}{2} + \frac{1}{2}\right) = 2.$$

- Suppose that $\mathbf{P}(k)$ is true for some $k \geqslant 2$.

$$\begin{aligned}
\text{LHS of } \mathbf{P}(k+1) &= \left[\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{2^k - 1}\right] \\
&\quad + \left[\frac{1}{2^k} + \cdots + \frac{1}{2^{k+1} - 1}\right].
\end{aligned}$$

The first bracket is $< k$ (by $\mathbf{P}(k)$); and each of the 2^k terms in the second bracket is $\leqslant \frac{1}{2^k}$, so the whole bracket is $\leqslant 1$.
Hence the LHS of $\mathbf{P}(k+1) < k+1$, so $\mathbf{P}(k+1)$ is true.

Hence $\mathbf{P}(n)$ is true for all $n \geqslant 2$.

(iv) At the very first stage (part (i)) we replaced $\frac{1}{2} + \frac{1}{3}$ by $\frac{1}{2} + \frac{1}{2} = 1$. Hence when $n \geqslant 2$, we know that the two sides of $\mathbf{P}(n)$ differ by at least $\frac{1}{2} - \frac{1}{3}$. This difference is greater than $\frac{1}{2^n}$ when $n \geqslant 3$, so (iv) follows.

(b)(i) $3 < 4$, so $\frac{1}{3} > \frac{1}{4}$.
$\therefore \frac{1}{3} + \frac{1}{4} > \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$.

(ii) $5, 6, 7 < 8$; hence $\frac{1}{5}, \frac{1}{6}, \frac{1}{7}$ are all $> \frac{1}{8}$.
$\therefore \frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8} > \frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{1}{2}$.

(iii) Let $\mathbf{P}(n)$ be the statement:
"$\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{2^n} > 1 + \frac{n}{2}$".

Then

- $\mathbf{P}(2)$ is true by (i), since

$$\begin{aligned} \frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} &= 1 + \frac{1}{2} + \left(\frac{1}{3} + \frac{1}{4}\right) \\ &> 1 + \frac{1}{2} + \left(\frac{1}{4} + \frac{1}{4}\right) \\ &= 1 + 2 \times \frac{1}{2}. \end{aligned}$$

- Suppose that $\mathbf{P}(k)$ is true for some $k \geqslant 2$.
LHS of $\mathbf{P}(k+1) = \left[\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{2^k}\right] + \left[\frac{1}{2^k+1} + \cdots + \frac{1}{2^{k+1}}\right]$.
The first bracket is $> 1 + \frac{k}{2}$ (by $\mathbf{P}(k)$);
and each of the 2^k terms in the second bracket is $\geqslant \frac{1}{2^{k+1}}$, so the whole bracket is $\geqslant \frac{1}{2}$.
Hence the LHS of $\mathbf{P}(k+1) > 1 + \frac{k+1}{2}$, so $\mathbf{P}(k+1)$ is true.

Hence $\mathbf{P}(n)$ is true for all $n \geqslant 2$.

254.

(a) We use induction. Let $\mathbf{P}(n)$ be the statement:

"n identical rectangular strips of length 2 balance *exactly* on the edge of a table if the successive protrusion distances (first beyond the edge of the table, then beyond the leading edge of the strip immediately below, and so on) are the terms

$$\frac{1}{n}, \frac{1}{n-1}, \frac{1}{n-2}, \ldots, \frac{1}{3}, \frac{1}{2}, \frac{1}{1}$$

of the finite harmonic series in reverse order."

- When $n = 1$, a single strip which protrudes distance 1 beyond the edge of the table has its centre of gravity exactly over the edge of the table. Hence $\mathbf{P}(1)$ is true.
- Suppose that we know that $\mathbf{P}(k)$ is true for some $k \geqslant 1$.
  Let $k+1$ identical strips be arranged as described in the statement $\mathbf{P}(k+1)$. The statement $\mathbf{P}(k)$ guarantees that the top k strips would exactly balance if the leading edge of the bottom strip were in fact the edge of the table; hence the combined centre of gravity of the top k strips is positioned exactly over the leading edge of the bottom strip.
  Let M be the mass of each strip; since the leading edge of the bottom strip is distance $\frac{1}{k+1}$ beyond the edge of the table, the top k strips produce a combined moment about the edge of the table equal to $kM \times \frac{1}{k+1}$.
  The centre of gravity of the bottom strip is distance $1 - \frac{1}{k+1} = \frac{k}{k+1}$ from the edge of the table in the opposite direction; hence it contributes a moment about the edge of the table equal to $M \times \left(-\frac{k}{k+1}\right)$.

  $\therefore$ the total moment of the whole stack about the edge of the table is equal to zero, so the centre of gravity of the combined stack of $k+1$ strips lies exactly over the edge of the table. Hence $\mathbf{P}(k+1)$ is true.

Hence $\mathbf{P}(n)$ is true for all $n \geqslant 1$.

(b) Problem **253**(b)(iii) now guarantees that a stack of 2^n strips can protrude a distance $> 1 + \frac{n}{2}$ beyond the edge of the table.

Note: Ivars Petersen's *Mathematical Tourist* blog contains an entry in 2009

http://mathtourist.blogspot.com/2009/01/overhang.html

which explores how one can obtain large overhangs with fewer strips if one is allowed to use strips to *counterbalance* those that protrude beyond the edge of the table.

255.

(a)(i) Let $\mathbf{P}(n)$ be the statement:

"$s_{2p-2} < s_{2p} < s_{2q+1} < s_{2q-1}$ for all p, q such that $1 \leqslant p, q \leqslant n$".

- $\mathbf{P}(1)$ is true (since s_0 is the empty sum, so

$$0 = s_0 < s_2 = \frac{1}{2} < s_3 = \frac{5}{6} < s_1 = 1.$$

- Suppose that $\mathbf{P}(k)$ is true for some $k \geqslant 1$. Then most of the inequalities in the statement $\mathbf{P}(k+1)$ are part of the statement $\mathbf{P}(k)$; the only outstanding inequalities which remain to be proved are:

$$s_{2k} < s_{2k+2} < s_{2k+3} < s_{2k+1}.$$

which are true, since

$$s_{2k+3} = s_{2k+2} + \frac{1}{2k+3} = s_{2k+1} - \frac{1}{2k+2} + \frac{1}{2k+3}$$

and

$$s_{2k+2} = s_{2k} + \frac{1}{2k+1} - \frac{1}{2k+2}.$$

Hence $\mathbf{P}(k+1)$ is true.

Hence $\mathbf{P}(n)$ is true for all $n \geqslant 2$.

(ii) The "even sums" $s_0, s_2, s_4, \ldots$ are increasing, but all are less than $s_1 = 1$, so they approach some value $s_{\text{even}} \leqslant 1$.
The "odd sums" $s_1, s_3, s_5, \ldots$ are decreasing, but all are greater than $s_2 = \frac{1}{2}$, so they approach some value $s_{\text{odd}} \geqslant \frac{1}{2}$.
The "even sums" $s_0, s_2, s_4, \ldots$ are increasing, but all are less than $s_5 = \frac{47}{60}$, so they approach some value $s_{\text{even}} \leqslant \frac{47}{60}$.
The "odd sums" $s_1, s_3, s_5, \ldots$ are decreasing, but all are greater than $s_6 = \frac{37}{60}$, so they approach some value $s_{\text{odd}} \geqslant \frac{37}{60}$.
Moreover, the difference between successive sums is $\frac{1}{n}$, and this tends towards zero, so the difference between each "odd sum" and the next "even sum" tends to zero, so $s_{\text{even}} = s_{\text{odd}}$.

(b) The proof from part (a) carries over word for word, with "$\frac{1}{k}$" replaced at each stage by "a_k".

256. Let $\mathbf{P}(n)$ be the statement:

"f_k has at least k distinct prime factors".

- $f_1 = 2$ has exactly 1 prime factor, so $\mathbf{P}(1)$ is true.
- Suppose that $\mathbf{P}(k)$ is true for some $k \geqslant 1$.
  Then $f_{k+1} = f_k(f_k + 1)$. The first factor f_k has at least k distinct prime factors.
  And the second factor $f_k + 1 > f_k > 1$, so has at least one prime factor.
  Moreover $HCF(f_k, f_k + 1) = 1$, so the second bracket has no factor in common with f_k.
  Hence f_{k+1} has at least $k+1$ distinct prime factors, so $\mathbf{P}(k+1)$ is true.

Hence $\mathbf{P}(n)$ is true for all $n \geqslant 1$.

Note: This problem [suggested by Serkan Dogan] gives a different proof of the result (Problem **76**(d)) that the list of prime numbers goes on for ever.

257.

(a) Let $\mathbf{P}(n)$ be the statement: "n distinct points on a straight line divide the line into $n+1$ intervals".

- 0 points leave the line in pristine condition – namely a single interval – so $\mathbf{P}(0)$ is true.

- Suppose that $\mathbf{P}(k)$ is true for some $k \geqslant 0$.
  Consider an arbitrary straight line divided by $k+1$ points $A_0, A_1, \ldots, A_k$.
  Then the k points $A_1, \ldots, A_k$ divide the line into $k+1$ intervals (by $\mathbf{P}(k)$).
  The additional point A_0 is distinct from $A_1, \ldots, A_k$ and so must lie inside one of these $k+1$ intervals, and divides it in two – giving $(k+1)+1 = k+2$ intervals altogether.
  Hence $\mathbf{P}(k+1)$ is true.

Hence $\mathbf{P}(n)$ is true for all $n \geqslant 0$.

(b) (i) We want a function R satisfying

$$R_0 = 1,\ R_1 = 2,\ R_2 = 4,\ R_3 = 7.$$

If we notice that in part (a)

$$n + 1 = 1 + \binom{n}{1},$$

then we might guess that

$$R_n = 1 + \binom{n}{1} + \binom{n}{2}.$$

(ii) Let $\mathbf{P}(n)$ be the statement:

"n distinct straight lines in the plane divide the plane into at most

$$f(n) = 1 + \binom{n}{1} + \binom{n}{2}$$

regions".

- 0 lines leave the plane in pristine condition – namely a single region – so $\mathbf{P}(0)$ is true, provided that

$$1 + \binom{0}{1} + \binom{0}{2} = 1.$$

- Suppose that $\mathbf{P}(k)$ is true for some $k \geqslant 0$.
  Consider the plane divided by $k+1$ straight lines $m_0, m_1, \ldots, m_k$.
  Then the k lines $m_1, \ldots, m_k$ divide the plane into at most

$$R_k = 1 + \binom{k}{1} + \binom{k}{2}$$

  regions (by $\mathbf{P}(k)$).
  The additional line m_0 is distinct from $m_1, \ldots, m_k$ and so meets each of these lines in at most a single point – giving at most k points on the line m_0. These points divide m_0 into at most $k+1$ intervals, and each of these intervals corresponds to a cut-line, where the line m_0 cuts one of the regions created

by the lines $m_1, m_2, \ldots, m_k$ into **two** pieces – giving at most

$$\begin{aligned} R_k + (k+1) &= 1 + \binom{k}{1} + \binom{k}{2} + k + 1 \\ &= 1 + \binom{k+1}{1} + \binom{k+1}{2} \\ &= R_{k+1} \end{aligned}$$

regions altogether.
Hence $\mathbf{P}(k+1)$ is true.

Hence $\mathbf{P}(n)$ is true for all $n \geqslant 0$.

(c) (i) We want a function S satisfying

$$S_0 = 1,\ S_1 = 2,\ S_2 = 4,\ S_3 = 8,\ S_4 = 15, \ldots$$

After thinking about the differences between successive terms in part (b), we might guess that

$$S_n = \binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \binom{n}{3}.$$

(ii) Let $\mathbf{P}(n)$ be the statement:

"n distinct planes in 3-space divide space into at most

$$S_n = \binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \binom{n}{3}$$

regions".

- 0 planes leave our 3D space in pristine condition – namely a single region – so $\mathbf{P}(0)$ is true – provided that

$$\binom{0}{0} + \binom{0}{1} + \binom{0}{2} + \binom{0}{3} = 1.$$

- Suppose that $\mathbf{P}(k)$ is true for some $k \geqslant 0$.
  Consider 3D divided by $k+1$ planes $m_0, m_1, \ldots, m_k$.
  Then the k planes $m_1, \ldots, m_k$ divide 3D into at most

$$S_k = \binom{k}{0} + \binom{k}{1} + \binom{k}{2} + \binom{k}{3}$$

  regions (by $\mathbf{P}(k)$).
  The additional plane m_0 is distinct from $m_1, \ldots, m_k$ and so meets each of these planes in (at most) a line – giving rise to at most k lines on the plane

m_0. This arrangement of lines on the plane m_0 divides m_0 into at most

$$R_k = 1 + \binom{k}{1} + \binom{k}{2}$$

regions, and each of these regions on the plane m_0 is the "cut" where the plane m_0 cuts an existing region into **two** pieces – giving rise to at most

$$\begin{aligned} S_k + R_k &= \left[\binom{k}{0} + \binom{k}{1} + \binom{k}{2} + \binom{k}{3}\right] + \left[1 + \binom{k}{1} + \binom{k}{2}\right] \\ &= \binom{k+1}{0} + \binom{k+1}{1} + \binom{k+1}{2} + \binom{k+1}{3} \\ &= S_{k+1} \end{aligned}$$

regions altogether. (There is no need for any algebra here: one only needs to use the Pascal triangle condition.)
Hence $\mathbf{P}(k+1)$ is true whenever $\mathbf{P}(k)$ is true.

Hence $\mathbf{P}(n)$ is true for all $n \geqslant 0$.

258. Notice that, given a dissection of a square into k squares, we can always cut one square into four quarters (by lines through the centre, and parallel to the sides), and so create a dissection with $k+3$ squares.

Let $\mathbf{P}(n)$ be the statement:

> "Any given square can be cut into m (not necessarily congruent) smaller squares, for each m, $6 \leqslant m \leqslant n$".

- Let $n = 6$. Given any square of side s (say). We may cut a square of side $\frac{2s}{3}$ from one corner, leaving an L-shaped strip of width $\frac{s}{3}$, which we can then cut into 5 smaller squares, each of side $\frac{s}{3}$. Hence $\mathbf{P}(6)$ is true.

  Let $n = 7$. Given any square, we can divide the square first into four quarters; then divide one of these smaller squares into four quarters to obtain a dissection into 7 smaller squares. Hence $\mathbf{P}(7)$ is true.

  Let $n = 8$. Given a square of side s (say). We may cut a square of side $\frac{3s}{4}$ from one corner, leaving an L-shaped strip of width $\frac{s}{4}$, which we can then cut into 7 smaller squares, each of side $\frac{s}{4}$. Hence $\mathbf{P}(8)$ is true.

- Suppose that $\mathbf{P}(k)$ is true for some $k \geqslant 8$.

  Then $k-2 \geqslant 6$, so any given square can be dissected into $k-2$ smaller squares (by $\mathbf{P}(k)$). Taking this dissection and dividing one of the smaller squares into four quarters gives a dissection of the initial square into $k-2+3$ squares. Hence $\mathbf{P}(k+1)$ is true.

Hence $\mathbf{P}(n)$ is true for all $n \geqslant 6$.

259. Let $\mathbf{P}(n)$ be the statement:

"Any tree with n vertices has exactly $n-1$ edges".

- A tree with 1 vertex is simply a vertex with 0 edges (since any edge would have to be a loop, and would then create a cycle). Hence $\mathbf{P}(1)$ is true.
- Suppose that $\mathbf{P}(k)$ is true for some $k \geqslant 1$.

Consider an arbitrary tree T with $k+1$ vertices.

[**Idea**: We need to find some way of reducing T to a tree T' with k vertices. This suggests "removing an end vertex". So we must first prove that "any tree T has an end vertex".]

Definition The number of edges incident with a given vertex v is called the *valency* of v.

Lemma Let S be a finite tree with $s > 1$ vertices. Then S has a vertex of valency 1 – that is an "end vertex".

Proof Choose any vertex v_0. Then v_0 must be connected to the rest of the tree, so v_0 has valency at least 1.

If v_0 is an "end vertex", then stop; if not, then choose a vertex v_1 which is adjacent to v_0.

If v_1 is an "end vertex", then stop; if not, choose a vertex $v_2 \neq v_0$ which is adjacent to v_1.

If v_2 is an "end vertex", then stop; if not, choose a vertex $v_3 \neq v_1$ which is adjacent to v_2.

Continue in this way.

All of the vertices $v_0, v_1, v_2, v_3, \ldots$ must be different (since any repeat $v_m = v_n$ with $m < n$ would define a cycle

$$v_m, v_{m+1}, v_{m+2}, \ldots, v_{n-1}, v_n = v_m$$

in the tree S). Since we know that the tree is finite (having precisely s vertices), the process must terminate at some stage. The final vertex v_e is then an "end vertex", of valency 1. QED

If we apply the Lemma to our arbitrary tree T with $k+1$ vertices, we can choose an "end vertex" v and remove both it and the edge e incident with it to obtain a tree T' having k vertices. By $\mathbf{P}(k)$ we know that T' has exactly $k-1$ edges, so when we reinstate the edge e, we see that T has exactly $(k-1)+1$ edges, so $\mathbf{P}(k+1)$ is true.

Hence $\mathbf{P}(n)$ is true for all $n \geqslant 1$.

260.

Note: All the polyhedra described in this solution are "spherical" by virtue of having their vertices located on the unit sphere.

(a) (i) A regular tetrahedron.

(ii) A square based pyramid with its apex at the North pole.

(b) If a spherical polyhedron has V vertices, E edges, and F faces, then

$$V - E + F = 2.$$

Now each edge has exactly two end vertices, so $2E$ counts the exact number of ordered pairs (v, e), where e is an edge, and v is a vertex "incident with e".

On the other hand, in a spherical polyhedron, each vertex v has valency at least 3; so each vertex v occurs in at least 3 pairs (v, e) of this kind. Hence $2E \geqslant 3V$.

In the same way, each edge e lies on the boundary of exactly 2 faces, so $2E$ counts the exact number of ordered pairs (f, e), where e is an edge of the face f.

On the other hand, in a spherical polyhedron, each face f has at least 3 edges; so each face f occurs in at least 3 pairs (f, e) of this kind. Hence $2E \geqslant 3F$.

If $E = 7$, then $14 \geqslant 3V$, and $14 \geqslant 3F$; now V and F are integers, so $V \leqslant 4$ and $F \leqslant 4$. Hence $V + F \leqslant 8$. However $V + F = E + 2 = 9$. This contradiction shows that no such polyhedron exists. QED

(c) We show by induction how to construct certain "spherical" polyhedra, with at most one non-triangular face. Let $\mathbf{P}(n)$ be the statement:

> "There exists a spherical polyhedron with at most one non-triangular face, and with e edges for each e, $8 \leqslant e \leqslant n$".

- We know that there exists a such a spherical polyhedron with $n = 6$ edges – namely the regular tetrahedron (with four faces, which are **all** equilateral triangles).

We know there is no such polyhedron with $n = 7$ edges (by part (b)).

When $n = 8$, there is no spherical polyhedron with $n = 8$ edges and **all** faces triangular (since we would then have $16 = 2E = 3F$, as in part (b)). However, there exists a spherical polyhedron with $n = 8$ edges and just one non-triangular face – namely the square based pyramid with its apex at the North pole.

When $n = 9$, we can join three points on the equator to the North and South poles to produce a triangular bi-pyramid (the dual of a triangular prism), with **all** faces triangular, and with $n = 9$ edges.

When $n = 10$, there is no spherical polyhedron with $n = 10$ edges and with all faces triangles (since we would then have to have $20 = 2E = 3F$, as in part (b)); but there exists a spherical polyhedron with $n = 10$ edges and just one face which is not an equilateral triangle – namely the pentagonal based pyramid with its apex at the North pole.

This provides us with a starting point for the inductive construction. In particular $\mathbf{P}(8)$, $\mathbf{P}(9)$, and $\mathbf{P}(10)$ are all true.

- Suppose that $\mathbf{P}(k)$ is true for some $k \geqslant 10$. The only part of the statement $\mathbf{P}(k + 1)$ that remains to be demonstrated is the existence of a suitable polyhedron with $k + 1$ edges.

  Since $k \geqslant 10$, we know that $k - 2 \geqslant 8$, so (by $\mathbf{P}(k)$) there exists a polyhedron with all its vertices on the unit sphere, with at most one non-triangular face, and with $e = k - 2$ edges. Take this polyhedron and remove a triangular face ABC. Now add a new vertex X on the sphere, internal to the spherical triangle ABC, and add the edges XA, XB, XC and the three triangular faces XAB, XBC, XCA, to produce a spherical polyhedron with $e = (k-2)+3 = k+1$ edges, and with at most one non-triangular face. Hence $\mathbf{P}(k+1)$ is true.

Hence $\mathbf{P}(n)$ is true for all $n \geqslant 8$.

261. To prove a result that is given in the form of an "if and only if" statement, we have to prove two things: "if", and "only if".

We begin by proving the "only if" part:

> "a map can be properly coloured with two colours only if every vertex has even valency".

Let M be a map that can be properly coloured with two colours. Let v be any vertex of the map M.

The edges $e_1, e_2, e_3, \ldots$ incident with v form parts of the boundaries of the sequence of regions around the vertex v (with e_1, e_2 bordering one region; e_2, e_3 bordering the next; and so on). Since we are assuming that the regions of the map M can be "properly coloured" with two colours, the succession of regions around the vertex v can be properly coloured with just two colours. Hence the colours of the regions around the vertex v must alternate (say black-white-black- $\ldots$). And since the map is finite, this sequence must return to the start – so the number of such regions at the vertex v (and hence the number of edges incident with v – that is, the valency of v) must be even.

We now prove the "if" part:

> "a map can be properly coloured with two colours if every vertex has even valency".

Suppose that we have a map M in which each vertex has even valency. We must prove that any such map M can be properly coloured using just two colours.

Let $\mathbf{P}(n)$ be the statement:

> "any map with m edges, in which each vertex has even valency, can be properly coloured with two colours whenever m satisfies $1 \leqslant m \leqslant n$,".

- If $n = 1$, a map in which every vertex has even valency, and which has just one edge e, must consist of a single vertex v, with e as a loop from v to v (so v has valency 2, since the edge e is incident with v twice). This creates a map with two regions – the "island" inside the loop, and the "sea" outside; so we can colour the "island" black and the "sea" white. Hence $\mathbf{P}(1)$ is true.
- Suppose that $\mathbf{P}(k)$ is true for some $k \geqslant 1$.

  Most of the contents of the statement $\mathbf{P}(k+1)$ are already guaranteed by $\mathbf{P}(k)$. To prove that $\mathbf{P}(k+1)$ is true, all that remains to be proved is that

  > any map with exactly $k+1$ edges, in which every vertex has even valency, can be properly coloured using just two colours.

Consider an arbitrary map M with $k+1$ edges, in which each vertex has even valency.

[**Idea:** We need to find some way of reducing the map M to a map M' with $\leqslant k$ edges, in which every vertex still has even valency.]

Since $k \geqslant 1$, the map M has at least 2 edges. Choose any edge e of M, with (say) vertices u_1, u_2 as its endpoints, and with regions R, S on either side of e.

Suppose first that $u_1 = u_2$, so the boundary of the region R (say) consists only of the edge e. Hence e is a loop, and S is the only region neighbouring R. The edge e contributes 2 to the valency of u_1; so if we delete the edge e, we obtain a map M' in which every vertex again has even valency, in which the regions R and S have been amalgamated into a region S'. Since M' has just k edges, M' can be properly coloured with just two colours. If we now reinstate the edge e and the region R, we can give S the same colour as S' (in the proper colouring of M') and give R the opposite colour to S' to obtain a proper colouring of the map M with just two colours.

Hence we may assume that $u_1 \neq u_2$, so that e is not the complete boundary of R. We may then slowly shrink the edge e to a point – eventually fusing the old vertices u_1, u_2 together to form a new vertex u', where two new regions R', S' meet. The result is then a new map M', in which all other vertices are unchanged (and so have even valency), and in which

$$\text{valency}(u') = (\text{valency}(u_1) - 1) + (\text{valency}(u_2) - 1)$$

which is also even.

Hence every vertex of the new map M' has even valency. Moreover, M' has at most k edges, so (by $\mathbf{P}(k)$) we know that the map M' can be properly coloured with just two colours. And in this colouring of M', there are an odd number of colour changes as one goes from R' to S' through the other regions that meet around the old vertex u_1 of M, so S' receives the opposite colour to R'. The guaranteed proper two-colouring of M' therefore extends back to give a proper two-colouring of the original map M. Hence $\mathbf{P}(k+1)$ is true.

Hence $\mathbf{P}(n)$ is true for all $n \geqslant 1$.

262. Let $\mathbf{P}(n)$ be the statement:

> "The 2^n sequences of length n consisting of 0s and 1s can be arranged in a cyclic list such that any two neighbouring sequences (including the last and the first) differ in exactly one coordinate position."

- When $n = 2$, the required cycle is obvious:

$$00 \to 10 \to 11 \to 01 \quad (\to 00).$$

So $\mathbf{P}(2)$ is true.

- The general construction is perhaps best illustrated by first showing how $\mathbf{P}(2)$ leads to $\mathbf{P}(3)$.

The above cycle for sequences of length 2 gives rise to two disjoint cycles for sequences of length 3:

– first by adding a third coordinate “0”:

$$000 \to 100 \to 110 \to 010 \quad (\to 000)$$

– then by adding a third coordinate “1”:

$$001 \to 101 \to 111 \to 011 \quad (\to 001).$$

Now eliminate the final join in each cycle ($010 \to 000$ and $011 \to 001$) and instead link the two cycles together by first reversing the order of the first cycle, and then inserting the joins $000 \to 001$ and $011 \to 010$ to form a single cycle.

In general, suppose that $\mathbf{P}(k)$ is true for some $k \geqslant 1$. Then we construct a single cycle for the 2^{k+1} sequences of length $k+1$ as follows:

> Take the cycle of the 2^k sequences of length k guaranteed by $\mathbf{P}(k)$, and form two disjoint cycles of length 2^k
>
> - first by adding a final coordinate “0“
> - then by adding a final coordinate “1”.
>
> Then link the two cycles into a single cycle of length 2^{k+1}, by eliminating the final step
>
> $$v_1v_2\cdots v_k0 \to 00\cdots 00$$
>
> in the first cycle, and
>
> $$v_1v_2\cdots v_k1 \to 00\cdots 01$$
>
> in the second cycle, reversing the first cycle, and inserting the joins
>
> $$00\cdots 00 \to 00\cdots 01 \text{ and } v_1v_2\cdots v_k1 \to v_1v_2\cdots v_k0$$
>
> to produce a single cycle of the required kind joining all 2^{k+1} sequences of length $k+1$. Hence $\mathbf{P}(k+1)$ is true.

Hence $\mathbf{P}(n)$ is true for all $n \geqslant 2$.

Note: The significance of what we call *Gray codes* was highlighted in a 1953 patent by the engineer Frank Gray (1887–1969) – where they were called *reflected binary codes* (since the crucial step in their construction above involves taking two copies of the previous cycles, reversing one of the cycles, and then producing half of the required cycle by traversing the first copy before returning backwards along the second copy). Their most basic use is to re-encode the usual binary counting sequence

$$1 \to 10 \to 11 \to 100 \to 101 \to 110 \to 111 \to 1000 \to 1001 \to 1010 \to \cdots,$$

where a single step can lead to the need to change arbitrarily many binary digits (e.g. the step from 3 = “11” to 4 = “100” changes 3 digits, and the step from 7 = “111” to 8 =“1000” changes 4 digits, etc.) – a requirement that is inefficient in terms of electronic “switching”, and which increases the probability of errors. In contrast, the *Gray code* sequence changes a single binary digit at each step. However, the physical energy which is saved through reducing the amount of "switching" in the circuitry corresponds to an increase in the need for unfamiliar mathematical formulae, which re-interpret each vector in the *Gray code* as the ordinary integer in the counting sequence to which it corresponds.

263.

(a) The whole construction is inductive (each label derives from an earlier label). So let $\mathbf{P}(n)$ be the statement:

> "if $HCF(r,s) = 1$ and $2 \leqslant r+s \leqslant n$, then the positive rational $\frac{r}{s}$ occurs once and only once as a label, and it occurs in its lowest terms".

- By construction the root is given the label $\frac{1}{1}$, so $\frac{1}{1}$ occurs. And it cannot occur again, since the numerator and denominator of each parent vertex are both positive, neither i nor j can ever be 0. Hence $\mathbf{P}(2)$ is true.

Notice that the basic construction:

> "if $\frac{i}{j}$ is a 'parent' vertex, then we label its 'left descendant' as $\frac{i}{i+j}$, and its 'right descendant' $\frac{i+j}{j}$"

guarantees that, since we start by labelling the root with the positive rational $\frac{1}{1}$, **all** subsequent 'descendants' are **positive**.
Moreover, if any 'descendant' were suddenly to appear not "in lowest terms", then either

– $HCF(i, i+j) > 1$, in which case $HCF(i, i+j) = HCF(i, j)$, so $HCF(i, j) > 1$ at *the previous stage*; or
– $HCF(i+j, j) > 1$, in which case $HCF(i+j, j) = HCF(i, j)$, so $HCF(i, j) > 1$ at *the previous stage*.

Since we begin by labelling the root $\frac{1}{1}$, where $HCF(1, 1) = 1$, it follows that **all** subsequent labels are positive rationals in **lowest terms**.

- Suppose that $\mathbf{P}(k)$ is true from some $k \geqslant 2$.
  Most parts of the statement $\mathbf{P}(k+1)$ are guaranteed by $\mathbf{P}(k)$. To show that $\mathbf{P}(k+1)$ is true, it remains to consider cases where $HCF(r,s) = 1$ and $r + s = k + 1$ ($\geqslant 3$). Either (i) $r > s$, or (ii) $s > r$.

(i) Suppose that $r > s$. Then $\frac{r}{s}$ arises in this (fully cancelled) form only as a direct (right) descendant of $\frac{r-s}{s}$. So $\frac{r}{s}$ occurs. Moreover, every label occurs in its lowest terms, so $\frac{r}{s}$ cannot occur again.
(ii) Suppose that $s > r$. Then $\frac{r}{s}$ arises in this (fully cancelled) form only as a direct (left) descendant of $\frac{r}{s-r}$. So $\frac{r}{s}$ occurs. Moreover, every label occurs in its lowest terms, so $\frac{r}{s}$ cannot occur again.
Hence $\mathbf{P}(k+1)$ is true.

Hence $\mathbf{P}(n)$ is true for all $n \geqslant 2$.

(b) The fact that the labels are left-right symmetric is also an inductive phenomenon. We note that the one fully "left-right symmetric" label, namely $\frac{1}{1}$, occurs in the only fully "left-right symmetric" position – namely at the root.

All other labels occur in reciprocal pairs: $\frac{r}{s}$ and $\frac{s}{r}$, where we may assume that $r > s$. The fact that these occur as labels of "left-right symmetric" vertices derives from the fact that

$\frac{r}{s}$ is the '**right** descendant' of $\frac{r-s}{s}$ and

$\frac{s}{r}$ is the '**left** descendant' of $\frac{s}{r-s}$.

So if we know that the earlier reciprocal pair reciprocal pair $\frac{r-s}{s}$ and $\frac{s}{r-s}$ occur as labels of symmetrically positioned vertices, then it follows that the same is true of the descendant reciprocal pair $\frac{r}{s}$ and $\frac{s}{r}$. We leave the reader to write out the proof by induction – for example, using the statement

> "$\mathbf{P}(n)$: if $r, s > 0$, and $2 \leqslant r + s \leqslant n$, then the reciprocal pair $\frac{r}{s}$, $\frac{s}{r}$ occur as labels of vertices at the same level below the root, with the two labelled vertices being mirror images of each other about the vertical mirror through the root vertex."

264. The intervals in this problem may be of any kind (including finite or infinite). Each interval has two "endpoints", which are either ordinary real numbers, or $\pm\infty$ (signifying that the interval goes off to infinity in one or both directions).

Let $\mathbf{P}(n)$ be the statement:

> "if a collection of n intervals on the x-axis has the property that any two intervals overlap in an interval (of possibly zero length – i.e. a point), then the intersection of all intervals in the collection is a non-empty interval".

When $n = 2$, the hypothesis of $\mathbf{P}(2)$ is the same as the conclusion. So $\mathbf{P}(2)$ is true.

Suppose that $\mathbf{P}(k)$ is true for some $k \geqslant 2$. We seek to prove that $\mathbf{P}(k+1)$ is true.

So consider a collection of $k+1$ intervals with the property that any two intervals in the collection intersect in a non-empty interval. If this collection includes one interval that is listed more than once, then the required conclusion follows from $\mathbf{P}(k)$. So we may assume that the intervals in our collection are all different.

Among the $k+1$ intervals, consider first those with the largest right hand endpoint. If there is only one such interval, denote it by I_0; if there is more than one interval with the same largest right hand endpoint, let I_0 be the interval among those with the largest right hand endpoint that has the largest left hand endpoint. In either case, put I_0 aside for the moment, leaving a collection S of k intervals with the required property.

By $\mathbf{P}(k)$ we know that the intervals in the collection S intersect in a non-empty interval I, with left hand endpoint a and right hand endpoint b (say).

We have to show that the intersection $I \cap I_0$ is non-empty.

The proof that follows works if the endpoint b is included in the interval I. The slight adjustment needed if b is not included in the interval I is left to the reader.

Since the right hand endpoint of I_0 is the largest possible, and since points between a and b belong to all the intervals of S, we can be sure that the right hand endpoint of I_0 is $\geqslant b$.

Moreover, for each point $x > b$, we know that there must be some interval I_x in the collection S which does not stretch as far to the right as x. Since, by hypothesis, the intersection $I_0 \cap I_x$ is non-empty, the left hand endpoint of I_0 lies to the left of every such point x, so I_0 must overlap the interval I, whence it follows that $I \cap I_0$ is a non-empty interval as required.

Hence $\mathbf{P}(k+1)$ is true.

Hence $\mathbf{P}(n)$ is true for all $n \geqslant 2$.

265. If one experiments a little, it should become clear

- that if tank T_2 contains more than tank T_1, then linking tank T to T_2 leads to a better immediate outcome (i.e. a larger amount in tank T) than linking T to T_1;
- that if, at some stage in the linking sequence, tank T contains an interim amount of a litres, and is about to link successively to a tank containing b litres, and then to the tank containing c litres, this ordered pair of changes alters the amount in the tank T to $\frac{a+b+2c}{4}$ litres; so for a better final outcome we should always choose the sequence so that $b < c$;
- once the tap linking tank T to another tank has been opened, so that the two levels become equal, there is no benefit from opening the tap linking these two tanks ever again, so tank T should be linked with each other tank at most once.

These three observations essentially determine the answer – namely that tank T should be joined to the other tanks *in increasing order of their initial contents.*

For a proof by induction, let $\mathbf{P}(n)$ be the statement:

> "given n tanks containing $a_1, a_2, a_3, \ldots, a_n$ litres respectively, where
>
> $$a_1 < a_2 < a_3 < \cdots < a_n,$$
>
> if T is the tank containing the smallest amount a_1 litres, then the optimal sequence for linking the other $n-1$ tanks to tank T (optimal in the sense that it transfers the maximum amount of water to tank T) is the sequence that links T successively to the other tanks in increasing order of their initial contents".

- When $n = 2$, there is only one possible sequence, which is the one described, so $\mathbf{P}(2)$ is true.
- Suppose that $\mathbf{P}(k)$ is true for some $k \geqslant 2$, and consider an unknown collection of $k+1$ tanks containing $a_1, a_2, a_3, \ldots, a_{k+1}$ litres respectively, where

$$a_1 < a_2 < a_3 < \cdots < a_{k+1},$$

and where T is the tank which initially contains a_1 litres.

Suppose that in the optimal sequence of k successive joins to the other k tanks (that is, that transfers the largest possible amount of water to tank T), the succession of joins is to join T first to tank T_2, then to tank T_3, and so on up to tank T_{k+1} (where tank T_m is not necessarily the tank containing a_m litres). There are now two possibilities: either

(i) T_{k+1} is the tank containing a_{k+1} litres, or

(ii) T_{k+1} contains less than a_{k+1} litres.

(i) Suppose tank T_{k+1} is the tank containing a_{k+1} litres. Then the first $k-1$ joins involve the k tanks containing $a_1, a_2, a_3, \ldots, a_k$ litres. And we know (by the very first bullet point above) that, in order to maximize the final amount in tank T, the amount in tank T after linking it to tank T_k must be "as large as it could possibly be". Hence, by $\mathbf{P}(k)$, the first $k-1$ joins link T successively to the other tanks in increasing order of their contents – before finally linking to the tank containing a_{k+1} litres. Hence the conclusion of $\mathbf{P}(k+1)$ holds.

(ii) Now suppose that tank T_{k+1} contains $a_m < a_{k+1}$ litres.

By the very first bullet point, in order to guarantee the optimal overall outcome of the final linking with tank T_{k+1} the amount in tank T after it has been linked to tank T_k must be "as large as it can possibly be" (given the amounts in the tanks $T, T_2, T_3, \ldots, T_k$). Hence statement $\mathbf{P}(k)$ applies to the initial sequence of $k-1$ joins (of T to T_2, then to T_3, and so on up to T_k), and guarantees that these tanks must be in increasing order of their initial contents. In particular, the last tank in this sequence, T_k, must be the one containing a_{k+1} litres. But if we denote by a litres the amount in tank T just before it links with tank T_k (containing a_{k+1} litres), then the last two linkings, with $b = a_{k+1}$ and $c = a_m$ contradict the second bullet point at the start of this solution. Hence case (ii) cannot occur.

Hence $\mathbf{P}(k+1)$ is true.

Hence $\mathbf{P}(n)$ is true for all $n \geqslant 2$.

266.

Note: Like all practical problems, this one requires an element of initial "modelling" in order to make the situation amenable to mathematical analysis.

'Residue' clings to surfaces; so the total amount of 'residue' will depend on

(a) the viscosity of the chemical (how 'thick', or 'sticky' it is), and

(b) the total surface area of the inside of the flask.

Since we are given no information about quantities, we may fix the amount of residue remaining in the 'empty' flask at "1 unit", and the amount of solvent in the other flask as "s units".

If we add the solvent, we get a combined amount of $1+s$ units of solution – which we may assume (after suitable shaking) to be homogeneous, with the chemical concentration reduced to "1 part in $1+s$".

The first modelling challenge arises when we try to make mathematical sense of what remains at each stage after we empty the flask. The internal surface area of the flask, to which any diluted residue may adhere, is fixed. If we make the mistake of thinking of the original chemical as "thick and sticky" and the solvent as "thin", then the viscosity of the diluted residue will change relative to the original, and will do so in ways that we cannot possibly know. Hence the only reasonable assumption (which may or may not be valid in a particular instance) is to *assume* that the viscosity of the original chemical is roughly the same as the viscosity of the chemical-solvent mixture. This then suggests that, on emptying the diluted mixture, roughly the same amount (1 unit) of diluted mixture will remain adhering to the walls of the flask. So we will be left with 1 unit of residue, with a concentration of $\frac{1}{1+s}$. In particular, if $s = 1$, using all the solvent at once reduces the amount of toxic chemical residue to $\frac{1}{2}$ unit (with the other $\frac{1}{2}$ unit consisting of solvent).

But what if we use only half of the solvent first, empty the flask, and then use the other half? Adding $\frac{s}{2}$ units of solvent (and shaking thoroughly) produces $1 + \frac{s}{2}$ units of homogeneous mixture, with a concentration of 1 part per $1 + \frac{s}{2}$. When we empty the flask, we expect roughly 1 unit of residue with this concentration – so just $\frac{2}{2+s}$ units of the chemical, with $\frac{s}{2+s}$ units of solvent.

If we then add the other $\frac{s}{2}$ units of solvent, this produces $1 + \frac{s}{2}$ units of mixture with a concentration of 1 part per $\left(1 + \frac{s}{2}\right)^2$. When we empty the flask, we expect roughly 1 unit of residue with concentration 1 part per $\left(1 + \frac{s}{2}\right)^2$. In particular, if $s = 1$, this strategy reduces the amount of toxic chemical in the 1 unit of residue to $\frac{4}{9}$ units. Since $\frac{4}{9} < \frac{1}{2}$ this two-stage strategy seems more effective than the previous "all at once" strategy.

Suppose that we use a four-stage strategy – using first one quarter of the solvent, then another quarter, and so on. We then land up with roughly 1 unit of residue with concentration 1 part per $\left(1 + \frac{s}{4}\right)^4$. In particular, if $s = 1$, we land up with the amount of toxic chemical in the 1 unit of residue equal to $\frac{256}{625}$ units, and $\frac{256}{625} < \frac{4}{9}$.

More generally, if we use $\left(\frac{1}{n}\right)$th of the solvent, n times, the final amount of toxic chemical in the 1 unit of residue is equal to $\left(1 + \frac{s}{n}\right)^{-n}$. And as n gets larger and larger, this expression gets closer and closer to e^{-s}. In particular, when $s = 1$, this strategy leaves a final amount of chemical in the 1 unit of residue approximately equal to $\frac{1}{e} = 0.367879\cdots$.

Note: The situation here is similar to that faced by a washing machine designer, who wishes to remove traces of detergent from items that have been washed, without using unlimited amounts of water. The idea of having a "fixed amount of solvent" corresponds to the goal of "water efficient" rinsing. However, the washing machine cycle, or programme, clearly cannot repeat the rinsing indefinitely (as would be required in the limiting case above).

267.

(i) If $\sqrt{2} = \frac{m}{n}$, then

$$2n^2 = m^2 \qquad (*)$$

Hence m^2 is even.

It follows that m must be even.

Note: It is a fact that, if $m = 2k$ is even, then $m^2 = 4k^2$ is also even. But this is **completely irrelevant** here. In order to conclude that "m must be even", we have to prove:

Claim m cannot be odd.
Proof Suppose m is odd.
$\therefore m = 2k + 1$ for some integer k.
But then

$$m^2 = (2k+1)^2 = 4k^2 + 4k + 1$$

would be odd, contrary to "m^2 must be even".
Hence m cannot be odd. QED

(ii) Since m is even, we may write $m = 2m'$ for some integer m'.
Equation $(*)$ in (i) above then becomes $n^2 = 2(m')^2$, so n^2 is even. Hence, as in the **Note** above, n must be even, so we can write $n = 2n'$ for some integer n'.

(iii) If $\sqrt{2} = \frac{m}{n}$, then $m = 2m'$, and $n = 2n'$ are both even.
$\therefore \sqrt{2} = \frac{m}{n} = \frac{2m'}{2n'} = \frac{m'}{n'}$.
In the same way, it follows that m' and n' are both even, so we may write $m' = 2m''$, $n' = 2n''$ for some integers m'', n''.
Continuing in this way then produces an endless decreasing sequence of positive denominators

$$n > n' > n'' > \cdots > 0.$$

contrary to the fact that such a sequence can have length at most $n - 1$ (or in fact, at most $1 + \log_2 n$).

268.

(i) If $a < b$ and $c > 0$, then $ac < bc$.
If $0 < \sqrt{2} \leqslant 1$, then (multiplying by $\sqrt{2}$) it follows that $2 \leqslant \sqrt{2} \leqslant 1$, which is false. Hence $\sqrt{2} > 1$.
We now know that $1 < \sqrt{2}$, so multiplying by $\sqrt{2}$ gives $\sqrt{2} < 2$. Hence $1 < \sqrt{2} < 2$. In particular, $\sqrt{2}$ cannot be written as a fraction with denominator 1, so $\mathbf{P}(1)$ is true.

(ii) Suppose $\mathbf{P}(k)$ is true for some $k \geqslant 1$. Most of the statement $\mathbf{P}(k+1)$ is implied by $\mathbf{P}(k)$: all that remains to be proved is that $\sqrt{2}$ cannot be written as a fraction with denominator $n = k + 1$.
Suppose $\sqrt{2} = \frac{m}{n}$, where $n = k + 1$ and m are positive integers.
Then $m = 2m'$ and $n = 2n'$ are both even (as in Problem **267**).
So $\sqrt{2} = \frac{m'}{n'}$ with $n' \leqslant k$, contrary to $\mathbf{P}(k)$. Hence $\mathbf{P}(k+1)$ holds.
$\therefore \mathbf{P}(n)$ is true for all $n \geqslant 1$.

269.

(a) Suppose that S is not empty. Then by the *Least Element Principle* the set S must contain a least element k: that is, a smallest integer k which is not in the set T. Then $k \neq 1$ (since we are told that T contains the integer 1). Hence $k > 1$.

Therefore $k-1$ is a positive integer which is smaller than k. So $k-1$ is not an element of S, and hence must be an element of T. But if $k-1$ is an element of T, then we are told that $(k-1)+1$ must also be a member of T. This contradiction shows that S must be empty, so T contains all positive integers.

(b) Suppose that T does not have a smallest element. Clearly 1 does not belong to the set T (or it would be the smallest element of T). Hence 1 must be an element of the set S.

Now suppose that, for some $k \geqslant 1$, all positive integers $1, 2, 3, \ldots, k$ are elements of S, but $k+1$ is not an element of S. Then $k+1$ would be an element of T, and would be the smallest element of T, which is not possible. Hence S has the property that

> "whenever $k \geqslant 1$ is an element of S, we can be sure that $k+1$ is also an element of S."

The *Principle of Mathematical Induction* then guarantees that these two observations (that 1 is an element of S, and that whenever k is an element of S, so is $k+1$) imply that S contains all the positive integers, so that the set T must be empty – contrary to assumption.

Hence T must have a smallest element.

270.

(i) Triangle $OP'Q'$ is a right angled triangle with $\angle P'OQ' = 45°$. Hence the two base angles (at O and at Q') are equal, so the triangle is isosceles: $\underline{P'Q'} = \underline{P'O}$.

(ii) Triangles $QQ'P'$ and $QQ'R$ are congruent by RHS (since they share the hypotenuse $\underline{QQ'}$, and have equal sides ($\underline{QP'} = \underline{QP} = \underline{QR}$). Hence $\underline{Q'P'} = \underline{Q'R}$.

(iii) If $\underline{OQ} : \underline{OP} = m : n$, we may choose a unit so that $\underline{OQ}$ has length m units and $\underline{OP}$ has length n units. Then

$$\underline{OP'} = \underline{OQ} - \underline{QP'} = \underline{OQ} - \underline{QP} = m - n,$$

and

$$\underline{OQ'} = \underline{OR} - \underline{Q'R} = \underline{OR} - \underline{Q'P'} = \underline{OR} - \underline{OP'} = n - (m - n).$$

(iv) $\underline{OP} < \underline{OQ} < \underline{OP} + \underline{PQ}$, so $n < m < n + n$. Hence $0 < m - n < n$, and $0 < 2n - m$.

(v) In the square $OP'Q'R'$, the ratio "diagonal : side" $= \underline{OQ'} : \underline{OP'} = \sqrt{2} : 1$.

If the ratio $\underline{OQ} : \underline{OP} = m : n$, with m, n positive integers, then the construction here replaces the positive integers m, n by smaller positive integers $2n - m$ and $m-n$, with $m-n < n$. And the process can be repeated indefinitely to generate an endless sequence of decreasing positive integers

$$n > m - n > (2n - m) - (m - n) = 3n - 2m > \cdots > 0.$$

Zarathustra's last, most vital lesson:
"Now do without me."
George Steiner (1929–)

Index

This book need not end here...

At Open Book Publishers, we are changing the nature of the traditional academic book. The title you have just read will not be left on a library shelf, but will be accessed online by hundreds of readers each month across the globe. We make all our books free to read online so that students, researchers and members of the public who cant afford a printed edition can still have access to the same ideas as you.

Our digital publishing model also allows us to produce online supplementary material, including extra chapters, reviews, links and other digital resources. Find *The Essence of Mathematics Through Elementary Problems* on our website to access its online extras. Please check this page regularly for ongoing updates, and join the conversation by leaving your own comments:

http://www.openbookpublishers.com/isbn/978178376996

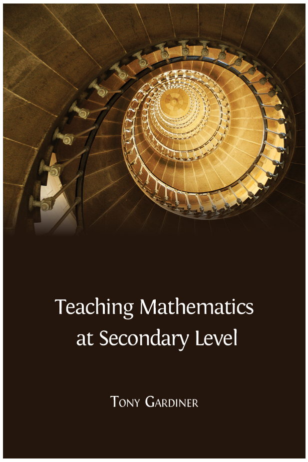
If you enjoyed this book, and feel that research like this should be available to all readers, regardless of their income, please think about donating to us. Our company is run entirely by academics, and our publishing decisions are based on intellectual merit and public value rather than on commercial viability. We do not operate for profit and all donations, as with all other revenue we generate, will be used to finance new Open Access publications.

For further information about what we do, how to donate to OBP, additional digital material related to our titles or to order our books, please visit our website: http://www.openbookpublishers.com

OpenBook Publishers
Knowledge is for sharing

You may also be interested in...

http://www.openbookpublishers.com/product/340



Teaching Mathematics is nothing less than a mathematical manifesto. Arising in response to a limited National Curriculum, and engaged with secondary schooling for those aged 11–14 (Key Stage 3) in particular, this handbook for teachers will help them broaden and enrich their students' mathematical education. It avoids specifying how to teach, and focuses instead on the central principles and concepts that need to be borne in mind by all teachers and textbook authors—but which are little appreciated in the UK at present. This study is aimed at anyone who would like to think more deeply about the discipline of 'elementary mathematics', in England and Wales and anywhere else. By analysing and supplementing the current curriculum, *Teaching Mathematics* provides food for thought for all those involved in school mathematics, whether as aspiring teachers or as experienced professionals. It challenges us all to reflect upon what it is that makes secondary school mathematics educationally, culturally, and socially important.

The Essence of Mathematics
Through Elementary Problems

ALEXANDRE BOROVIK AND TONY GARDINER

It is increasingly clear that the shapes of *reality* – whether of the natural world, or of the built environment – are in some profound sense *mathematical*. So it would benefit students and educated adults to understand what makes mathematics itself 'tick', and why its shapes, patterns and formulae provide us with precisely the language we need to make sense of the world around us. The second part of this challenge may require some specialist experience, but the authors of this book concentrate on the first part, and explore the extent to which elementary mathematics allows us all to understand something of the nature of mathematics *from the inside*.

The Essence of Mathematics consists of a sequence of 270 problems – with commentary and full solutions. The reader is assumed to have a reasonable grasp of school mathematics. More importantly, s/he should want to understand something of mathematics *beyond the classroom*, and be willing to engage with (and to reflect upon) challenging problems that highlight the essence of the discipline.

The book consists of six chapters of increasing sophistication (Mental Skills; Arithmetic; Word Problems; Algebra; Geometry; Infinity), with interleaved commentary. The content will appeal to students considering further study of mathematics at university, teachers of mathematics at age 14-18, and anyone who wants to see what this kind of elementary content has to tell us about how mathematics *really* works.

As with all Open Book publications, this entire book is available to read for free on the publisher's website. Printed and digital editions, together with supplementary digital material, can also be found at www.openbookpublishers.com

Cover image: Photo by Samuel Zeller on Unsplash at https://unsplash.com/photos/j0g8taxHZa0
Cover design: Anna Gatti.

ebook
ebook and OA editions also available

OPEN ACCESS

OpenBook Publishers

ISBN 978-1-78374-699-6
90000
9 781783 746996
www.openbookpublishers.com