Preface

In *A Mathematical Olympiad Primer*, Geoff Smith described the technique of inversion as a ‘dark art’. It is difficult to define precisely what is meant by this phrase, although a suitable definition is ‘an advanced technique, which can offer considerable advantage in solving certain problems’. These ideas are not usually taught in schools, mainstream olympiad textbooks or even IMO training camps. One case example is projective geometry, which does not feature in great detail in either *Plane Euclidean Geometry* or *Crossing the Bridge*, two of the most comprehensive and respected British olympiad geometry books. In this volume, I have attempted to amass an arsenal of the more obscure and interesting techniques for problem solving, together with a plethora of problems (from various sources, including many of the extant mathematical olympiads) for you to practice these techniques in conjunction with your own problem-solving abilities. Indeed, the majority of theorems are left as exercises to the reader, with solutions included at the end of each chapter. Each problem should take between 1 and 90 minutes, depending on the difficulty.

The book is not exclusively aimed at contestants in mathematical olympiads; it is hoped that anyone sufficiently interested would find this an enjoyable and informative read.

All areas of mathematics are interconnected, so some chapters build on ideas explored in earlier chapters. However, in order to make this book intelligible, it was necessary to order them in such a way that no knowledge is required of ideas explored in *later* chapters! Hence, there is what is known as a *partial order* imposed on the book. Subject to this constraint, the material is arranged in such a way that related concepts are as close as possible together; this is complemented by a hierarchical division into chapters and sections.

One concern is that a book of this depth would be too abstract. Wherever possible, both two-dimensional and three-dimensional full-colour diagrams are included to aid one’s intuition.

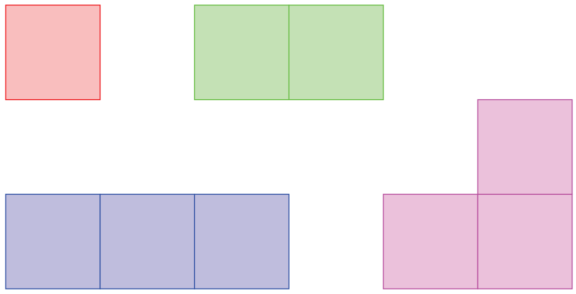
I have assumed that the reader will have at least the cumulative knowledge contained in both *A Mathematical Olympiad Primer* and a typical A-level mathematics syllabus. I also recommend reading either *Plane Euclidean Geometry* or *Crossing the Bridge*, although this is not a prerequisite to understanding the content of this book.

Be fruitful, and multiply.

Adam P. Goucher, 2012

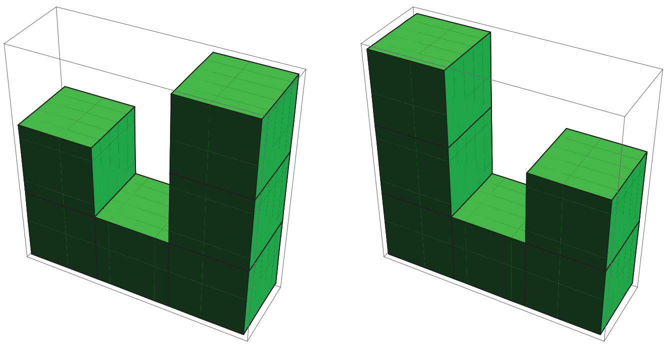
Combinatorics I

Combinatorics is the study of discrete objects. Combinatorial problems are usually simple to define, but can be very difficult to solve. For example, a *polyomino* is a set of unit squares connected edge-to-edge, such that the vertices are positioned at integer coordinates. The four polyominoes with three or fewer squares are shown below:

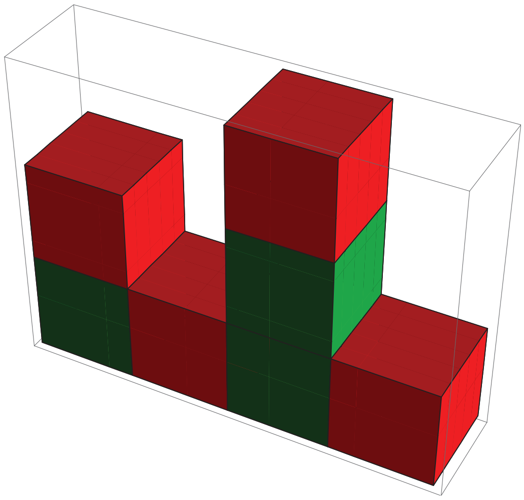


A natural question to ask is how many polyominoes there are of size n. We have already proved by exhaustion that this sequence begins $\{1, 1, 2, \ldots\}$. After a little effort, you will discover that there are five *tetrominoes* (polyominoes of size 4) and twelve *pentominoes* (polyominoes of size 5). Although this is a very simple problem to state, it is very difficult to find a formula for the number of polyominoes of a particular size. Indeed, there is no known formula as of the time of writing, and no-one knows how many polyominoes there are of size 60. Even the conjectured asymptotic formula, $P(n) \sim \frac{c\,\lambda^n}{n}$, is unproved (it is possible that, for instance, $P(n) \sim \frac{c\,\lambda^n}{n^{1.000001}}$ instead).

Counting polyominoes is a hard problem. Variants of this problem are substantially easier. For instance, suppose we restrict ourselves to polyominoes that can be created by stacking cubes in a vertical plane. To make things even easier, we consider rotations and reflections to be distinct, so the following arrangements are counted as two different polyominoes:



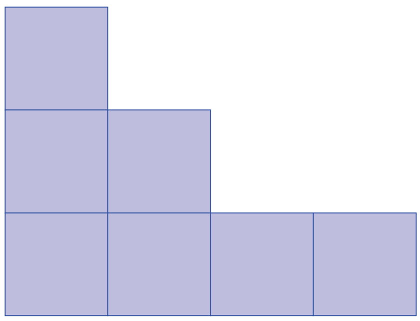
Many seemingly different combinatorial problems can be shown to be equivalent. This question can be converted into an equivalent one by colouring the top cube in each column red, and the remainder green. We then proceed up each column in turn, noting the colour of each cube. The configuration below is associated with the string $G\,R\,R\,G\,G\,R\,R$. Every string must end in R for obvious reasons, so we may as well omit the final R and just consider the string of $n-1$ letters, $G\,R\,R\,G\,G\,R$.



Since each of these polyominoes has a unique string, and vice-versa, we have a *bijection* between the two sets. Counting strings of a particular length is very easy (mathematicians would call this *trivial*); there are 2^{n-1} strings of $n-1$ letters chosen from $\{G, R\}$. Hence, there are 2^{n-1} of these restricted polyominoes. A third way of viewing this problem is to consider it to be an *ordered partition* of n; the above configuration corresponds to the sum $7 = 2 + 1 + 3 + 1$. So, we have solved a third combinatorial problem: there are 2^{n-1} ordered partitions of n identical objects into non-empty subsets.

1. How many ordered partitions are there of n into precisely k subsets?

What if we consider the partitions $2 + 1 + 3 + 1$ and $3 + 1 + 1 + 2$ to be equivalent? In other words, what if order doesn't matter? This problem can be rephrased by forcing the elements of the partition to be arranged in decreasing order of size, *i.e.* $3 + 2 + 1 + 1$. The associated diagram of this partition is known variably as a *Ferrers diagram* or *Young diagram*.

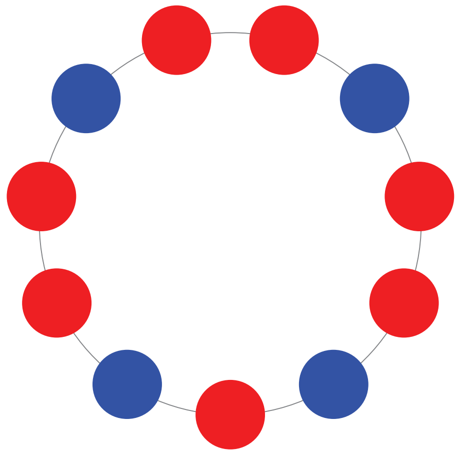


The partition numbers are $\{1, 2, 3, 5, 7, 11, \ldots\}$, as opposed to the ordered partition numbers $\{1, 2, 4, 8, 16, 32, \ldots\}$. Whereas the latter have a very simple formula, the formula for the unordered partition numbers is given by an extremely complicated infinite series by Hardy, Ramanujan and Rademacher:

- $$p(n) = \frac{1}{\pi\sqrt{2}} \sum_{k=1}^{\infty} \sqrt{k} \left(\sum_{m \bmod k;\, \gcd(m,k)=1} e^{\frac{\pi i}{4k}\left(\left(\sum_{n=1}^{k-1} \cot\left(\frac{\pi n}{k}\right) \cot\left(\frac{\pi n m}{k}\right)\right) - 8nm\right)} \right) \frac{d}{dn} \left(\frac{\sinh\left(\frac{\pi}{k}\sqrt{\frac{2}{3}\left(n - \frac{1}{24}\right)}\right)}{\sqrt{n - \frac{1}{24}}} \right)$$

Don't be perturbed by this; the combinatorics explored in this chapter are several orders of magnitude easier than the partition problem. We begin with the problem of colouring p beads on a necklace, where p is a prime number. This leads to an intuitive proof of Fermat's little theorem, and a similarly combinatorial approach yields Wilson's theorem. The idea of symmetry is essential, so we contemplate some group theory as well.

Burnside's lemma



Consider how many ways there are of colouring the 11 beads of this necklace either red or blue. This is an ambiguous question and there are many ways in which it can be answered:

- "There are 2048 ways of colouring the necklace."
- "There are 188 ways of colouring the necklace."
- "There are 126 ways of colouring the necklace."

These answers are all valid, since the question was vague. If rotations and reflections are considered to be distinct, then the first answer is clearly correct (as $2^{11} = 2048$). If rotations are considered to be equivalent, but reflections are distinct, then the second is correct. The third answer applies when both rotations and reflections are equivalent.

It is easy to derive the answer 2048 in the first instance, but the others are somewhat trickier. Probably the best way to count the number of possibilities is to use a result known as *Burnside's lemma*. Firstly, we define what we mean by a symmetry.

■ A *symmetry* is an operation we can perform on an object. Moreover, the set of symmetries must form a group under composition. For example, a group of rotations can be regarded as symmetries. **[Definition of symmetry]**

In the first case of the necklace problem, we only consider the trivial group of one symmetry: the identity. In the second instance, we have the cyclic group of eleven symmetries (ten rotations and the identity). Finally, the third case requires the dihedral group of twenty-two symmetries (eleven reflections, ten rotations and the identity).

A *direct* symmetry can be expressed as a sequence of rigid transformations, such as translations and rotations. For example, the red and blue Rs are related by a direct symmetry (rotation by π through their common barycentre), By comparison, the green R cannot be obtained from the red R by a sequence of rotations and translations, so is related to the red R by an *indirect* symmetry (in this case, a reflection). The composition of two direct or two indirect transformations is a direct transformation; the composition of a direct and indirect transformation is an indirect transformation. This idea can be succinctly represented as a 2×2 *Cayley table*:

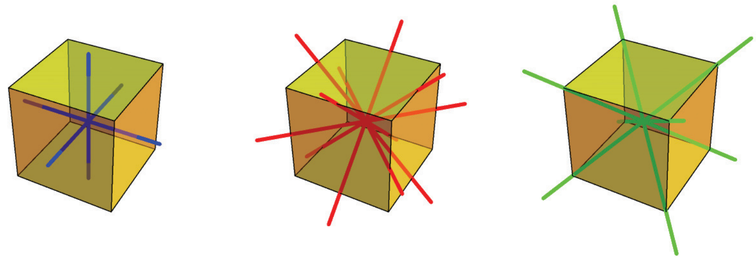
$\circ$	D	I
D	D	I
I	I	D

■ An object is said to be *fixed* by a symmetry if it is unchanged by applying that symmetry. **[Definition of 'fixed']**

For example, the hyperbola $x^2 = y^2 + 1$ is fixed by a rotation of π about the origin, whereas the parabola $y = x^2$ is not.

- The number of distinct objects is equal to the mean number of objects fixed by each symmetry. **[Burnside's lemma]**

For the second case of the necklace problem, there are 11 symmetries. The identity symmetry fixes all 2048 objects, whereas the ten rotations only fix two objects (the monochromatic necklaces). So, Burnside's lemma gives us a total of $\frac{1}{11}(2048 + 10\times2) = 188$ unique necklaces. Similarly, for the third case, we observe that there must be $2^6 = 64$ objects fixed by each of the 11 reflections, so we have $\frac{1}{22}(2048 + 10\times2 + 11\times64) = 126$ unique necklaces. That this gives an integer answer is a useful way to check your arithmetic.

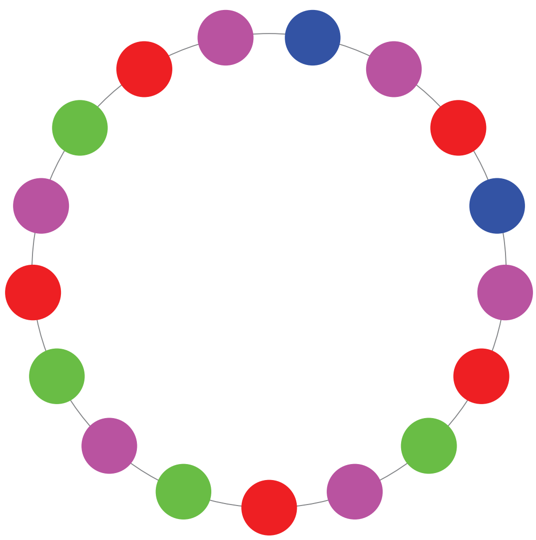


The cube has a group of 24 direct symmetries (and the same number of indirect symmetries). We can classify those 24 direct symmetries into five *conjugacy classes*:

- 1 identity symmetry;
- 6 rotations by $\frac{1}{2}\pi$ about the blue axes;
- 3 rotations by π about the blue axes;
- 6 rotations by π about the red axes;
- 8 rotations by $\frac{2}{3}\pi$ about the green axes.

2. Suppose we colour each face of a cube one of k colours. By considering the number of colourings fixed by each of the above symmetries, deduce the number of distinct colourings of the cube where rotations are considered equivalent.

Fermat's little theorem



We now generalise the previous question to a necklace of p beads (where p is prime) and c different colours.

3. How many distinct ways can a necklace of p beads be coloured with c colours, where p is prime and $c \geq 2$? Rotations are considered to be equivalent, whereas reflections are distinct.

4. Hence show that $c^p \equiv c \pmod{p}$. **[Fermat's little theorem]**

Fermat's little theorem only applies when the modulus is prime. If, instead, the modulus is composite, it is necessary to use a generalisation by Euler. Unlike Fermat's little theorem, Euler's generalisation does not appear to be a consequence of applying Burnside's lemma to necklaces of n beads.

- If a and n are coprime, then $a^{\varphi(n)} \equiv 1 \pmod{n}$, where $\varphi(n)$ is Euler's totient function (the number of positive integers $k \leq n$ which are coprime to n). **[Euler-Fermat]**

Euler's totient function can easily be computed when the prime factorisation of n is known. Specifically, we have the rule $\varphi(a\,b) = \varphi(a)\,\varphi(b)$ if a and b are coprime, and $\varphi(p^n) = (p-1)\,p^{n-1}$.

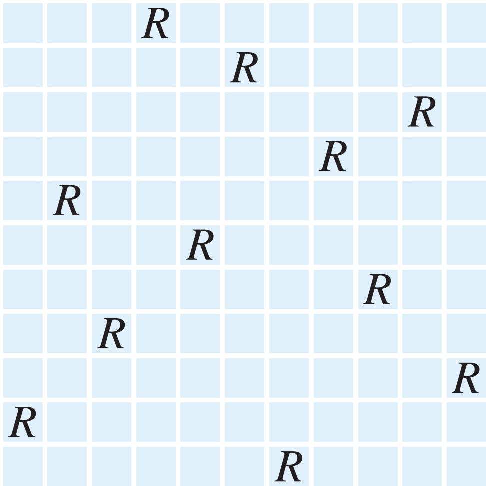
Suppose $N = p\,q$ is a product of two distinct primes, each of which has hundreds of digits. Given N, there is no known algorithm capable of factorising it to find p and q in a reasonable (polynomial) amount of time. This can be used as the basis of a cryptographic system known as *RSA* (after its creators, Rivest, Shamir and Adleman). The idea is that we define a function, $f : \mathbb{Z}_N \to \mathbb{Z}_N$, which the general public has access to. However, we keep the inverse function f^{-1} secret.

5. Suppose that $b = f(a) \equiv a^d \pmod{N}$. Show that $f^{-1}(b) \equiv b^e \pmod{N}$, where $d\,e \equiv 1 \pmod{\varphi(N)}$. **[Basis of RSA]**

In other words, we publish a, d, N (and therefore f) but leave p, q, e secret. As it is impossible to compute e from d without knowledge of p and q, the general public cannot calculate f^{-1}. Hence, they can encrypt an integer, but not decrypt it. As the numbers in $\mathbb{Z}_N$ can have hundreds of digits, it is possible to store a substantial amount of information in one integer. This is typically used to encrypt passwords, safe in the knowledge that there is no known algorithm for rapidly factorising semiprimes.

Interestingly, there is an algorithm called *AKS* which enables a computer (or, more correctly, Turing machine) to determine whether a number is prime in polynomial time (in the number of digits), but actually factorising the number may require exponential time. Additionally, so-called 'quantum computers' are capable of prime factorisation in cubic time, so a sufficiently powerful quantum computer would render RSA useless. Fortunately, this technology is a long way off, and the largest semiprime factorised by Shor's algorithm as of the time of writing is $15 = 5 \times 3$ using a machine with seven quantum bits.

Wilson's Theorem



Suppose we have a $p \times p$ chessboard, where p is prime. We label each square with a coordinate (x, y), where x and y are considered modulo p (in effect, forming a toroidal surface). We then place an arrangement of p non-attacking rooks on the chessboard, *i.e.* one in every row and one in every column. We consider the group of p^2 symmetries (one identity and $p^2 - 1$ translations).

6. Show that there are $p!$ arrangements fixed by the identity symmetry.

7. Show that no arrangements are fixed by any of the $2(p-1)$ horizontal or vertical translations.

8. Show that p arrangements are fixed by each of the $(p-1)^2$ remaining translations.

9. Hence determine the number of unique arrangements, where toroidal translations of the board are considered equivalent.

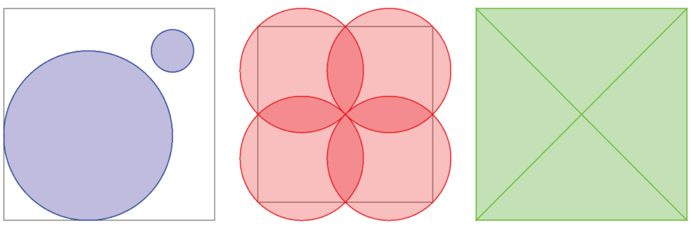
10. Prove that $(p-1)! \equiv -1$ modulo p if p is prime. **[Wilson's theorem]**

If n is composite, then $(n-1)! \equiv 0$ modulo n, except where $n = 4$, in which case $(n-1)! \equiv 2$. Hence, the converse of Wilson's theorem is also true.

Packings, coverings and tilings

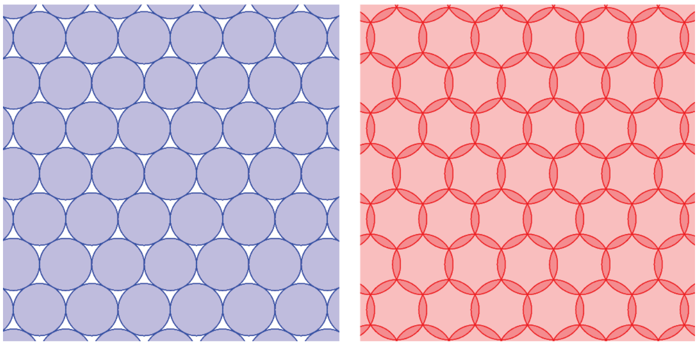
Straddling the boundary between combinatorics and geometry is the idea of *tessellations*, or *tilings*.

Consider a set S of [closed] tiles, each of which is a subset of some region R. If the pairwise intersection of any two tiles of S has zero area, then S is a *packing*. If the union of all tiles in S is the entirety of R, then S is a *covering*. If both of these conditions hold, it is a *tiling*.

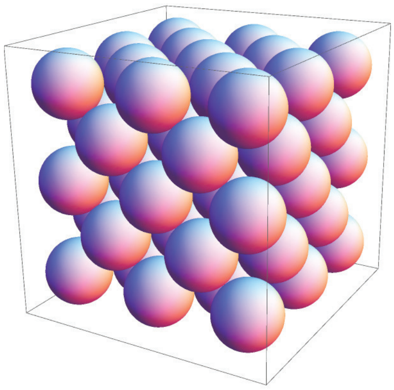


The diagram above highlights the differences. The first diagram is a packing using two blue circles. The second is a covering using four red circles. The third diagram is both a packing and covering, and thus a tiling, using four green isosceles right-angled triangles.

Using circles of unit radius, there are obviously no tilings of the plane. It is of interest to find the packing of the highest density and covering of the lowest density.



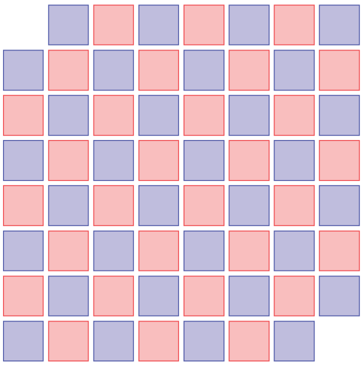
It has been proved that the optimal packings and coverings of the plane using circles of unit radius are obtained by positioning them at the vertices of the regular triangular tiling. Other optimisation problems are solved by the hexagonal lattice, which is why honeybees favour hexagonal honeycombs as opposed to a rectangular Cartesian grid. In higher dimensions, less is known. For three dimensions, the optimal lattice packing of spheres is the *face-centred cubic* lattice $A_3 = \{(x, y, z) \in \mathbb{Z}^3, x + y + z \equiv 0 \pmod 2\}$, whereas the optimal lattice covering is the *body-centred cubic* lattice $A_3{}^* = \{(x, y, z) \in \mathbb{Z}^3, x \equiv y \equiv z \pmod 2\}$.



Each sphere in the face-centred cubic packing is adjacent to twelve other spheres. This suggests another packing problem: what is the maximum number of disjoint unit spheres tangent to a given unit sphere? In two dimensions, the answer is rather trivially six. In three dimensions, Isaac Newton conjectured that the maximum is indeed twelve spheres, whereas David Gregory hypothesised that thirteen could be achieved. It transpires that Newton was correct. The problem has also been solved in 4, 8 and 24 dimensions, again corresponding to the arrangements of spheres in very regular lattice packings (known as D_4, E_8 and Λ_{24}, respectively). Λ_{24} (the *Leech lattice*) has so many interesting properties and profound connections that I cannot hope to list them all here. Nevertheless, its existence is related to string theory, error-correcting codes, the Monster group, and the curious fact that $1^2 + 2^2 + 3^2 + \ldots + 24^2 = 70^2$.

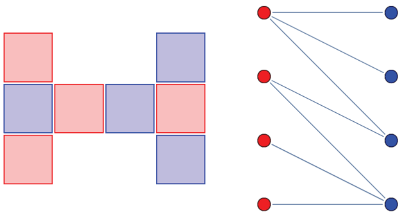
Colouring arguments

To begin with, we ponder tilings of finite, discrete spaces. For example, consider a standard 8×8 chessboard with two opposite corners removed. Is it possible to tile the resulting shape with 31 1×2 dominoes?



If the chessboard is coloured as above, each domino must occupy precisely one blue and one red square. As there are 32 blue and 30 red squares, it is clearly impossible to tile it with 31 dominoes.

The more general problem of determining whether a polyomino-shaped region can be tiled with dominoes can be embedded in graph theory. We represent the squares with vertices, and join vertices corresponding to adjacent squares. Some regions clearly cannot be tiled, even if they have equal quantities of squares of each parity. One such example is the following 'octomino', shown below with an equivalent bipartite graph:



The lowest blue vertex in the graph is connected to three red vertices, two of which are exclusively connected to this blue vertex. It is therefore impossible to place disjoint dominoes to cover both of the corresponding red squares. However, the basic colour-counting argument is insufficient here, as there are four red and four blue squares.

In effect, we want to find a *bipartite matching* between the red and blue vertices of the graph. A necessary and sufficient condition for there to exist an *injection* from the red vertices to the blue vertices is *Hall's marriage theorem.*

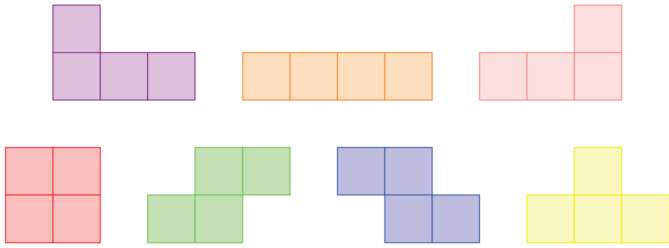
- Let S be the set of red vertices, and T be the set of blue vertices. Consider each subset $S' \subseteq S$, and let $T' \subseteq T$ be the set of vertices directly connected to vertices in S'. Then there exists an injection from the red vertices to the blue vertices if and only if $|S'| \leq |T'|$ for all subsets S'. **[Hall's marriage theorem]**

For a bijection, it is necessary and sufficient that there are equal numbers of red and blue vertices and the above result also holds. Returning to the octomino problem, note that the two red vertices of degree 1 are connected to the same blue vertex, so the marriage condition does not hold.

Verifying the marriage condition can be a time-consuming process, as there are 2^n subsets of red vertices for a bipartite graph with n red and n blue vertices. This is faster than checking every possible bijection, of which there are $n!$. Both of these algorithms are said to take *exponential time*. People are interested in fast, *polynomial-time* algorithms, as they usually can be executed in a reasonable amount of time.

Colouring can solve much more general problems than the domino tiling problem.

11. Determine whether it is possible to tile a 4×7 rectangle with (rotations of) each of the seven tetrominoes (where reflections are considered to be distinct). The seven tetrominoes are shown below:



12. Is it possible to tile a 6×6 rectangle with 15 dominoes and 6 non-attacking rooks? **[Ed Pegg Jr, 2002]**

13. Show that the maximum number of (grid-aligned) $k \times k$ square tiles that can be packed into a $m \times n$ chessboard is given by $\lfloor \frac{m}{k} \rfloor \lfloor \frac{n}{k} \rfloor$.

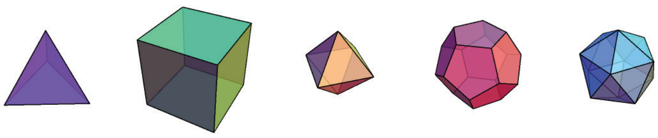
In addition to determining whether or not a region can be tiled, it is occasionally possible to enumerate precisely how many ways in which this can be done. This is typically accomplished using recursion on the size of the region.

14. In how many ways can a $2 \times n$ rectangle be tiled with n dominoes?

This is a simple case of what one would initially imagine to be a completely intractable problem: to count the number of domino tilings of a $m \times n$ rectangle. A remarkable discovery by Kasteleyn enumerates this for any planar graph, and thus how many domino tilings exist for any polyomino. In particular, a $m \times n$ chessboard can be tiled by dominoes in exactly $\prod_{k=1}^{n} \prod_{l=1}^{m} \sqrt[4]{4 \cos^2 \frac{\pi l}{m+1} + 4 \cos^2 \frac{\pi k}{n+1}}$ ways.

Regular solids and tilings

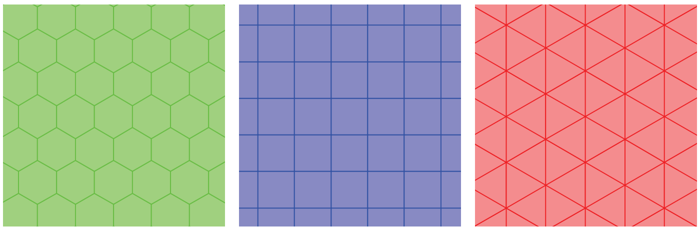
Suppose we attempt to tile a surface with regular n-gons, where k n-gons meet at each vertex. To avoid trivial cases, we assume that both k and n exceed 2. The cases where the *Schläfli symbol* $\{n, k\}$ is either $\{3, 3\}$, $\{4, 3\}$, $\{3, 4\}$, $\{5, 3\}$ and $\{3, 5\}$ result in the five regular solids, namely the tetrahedron, cube, octahedron, dodecahedron and icosahedron.



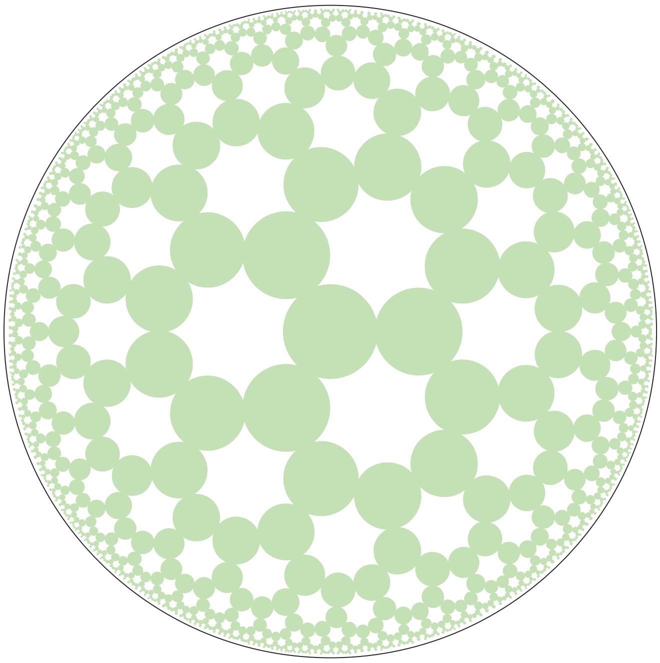
They are also referred to as *Platonic solids*, as Plato believed that all matter was composed (at the atomic level) of minuscule cubes, tetrahedra, octahedra and icosahedra, associating each one with a different classical element. He reserved the dodecahedron for representing the entire universe.

15. Each face of a regular dodecahedron is infected with either *E. coli*, *S. aureus* or *T. rychlik* bacteria. In how many ways is this possible, treating rotations as equivalent? **[Adapted from Google Labs Aptitude Test]**

If $\{n, k\}$ is $\{6, 3\}$, $\{4, 4\}$ or $\{3, 6\}$, we obtain the hexagonal, square and triangular tilings, respectively, of the plane. The Platonic solids can be regarded as analogous tilings of the sphere.

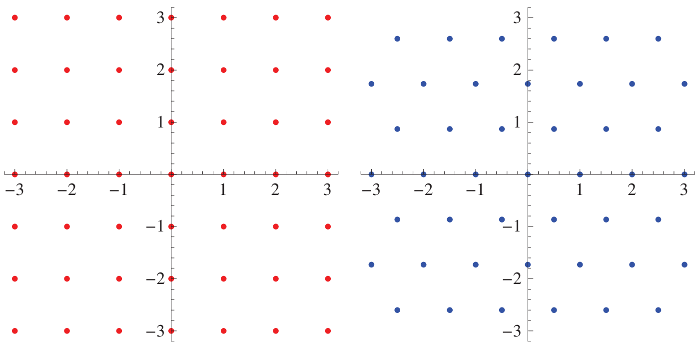


If $\{n, k\}$ is anything other than these eight possibilities, the sum of the angles around each vertex exceeds 2π. This is only possible in the bizarre hyperbolic surfaces described by Bolyai-Lobachevskian geometry.



On the complex plane, numbers of the form $a + b\,i$ ($a, b \in \mathbb{Z}$) form a ring known as the *Gaussian integers*, which are positioned at the vertices of the square tiling. As Euclid's algorithm can be applied to the Gaussian integers, the fundamental theorem of arithmetic still holds: Gaussian integers can be factorised uniquely into a product of *Gaussian primes* (up to multiplication by the *units*, 1, -1, i and $-i$). Not all ordinary primes are Gaussian primes; for example, 2 is not a Gaussian prime, as it can be factorised as $(1 + i)(1 - i)$.

Suppose we have a grasshopper initially positioned at the origin, which can only jump to a Gaussian prime within the disc of radius R centred on its current position. It is an unsolved problem as to whether there is some R for which the grasshopper can visit infinitely many Gaussian primes.



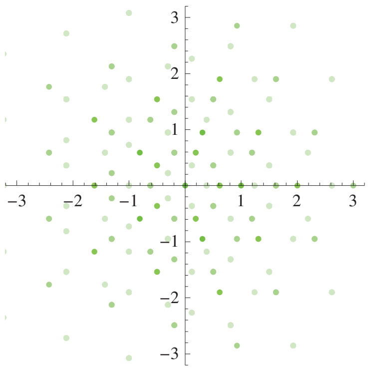
Similarly, numbers of the form $a + b\,\omega$ ($a, b \in \mathbb{Z}$), where ω is a primitive cube root of unity, form the ring of *Eisenstein integers*. They are positioned at the vertices of the triangular tiling. As with the Gaussian integers, the fundamental theorem of arithmetic applies. The units are the sixth roots of unity, namely $\{\pm 1, \pm\omega, \pm\omega^2\}$. It is possible to find the squared distance between two Eisenstein integers a and b by expressing the vector $a - b$ in terms of $\{1, \omega, \omega^2\}$ and calculating $|a-b|^2 = (a-b)(a^* - b^*)$, remembering that $1 + \omega + \omega^2 = 0$ and $\omega^3 = 1$.

16. A set S of 99 points are drawn in the plane, such that no two are within a distance of 2 units. Prove that there exists some subset $T \subset S$ of 15 points, such that no two are within a distance of $\sqrt{7}$ units.

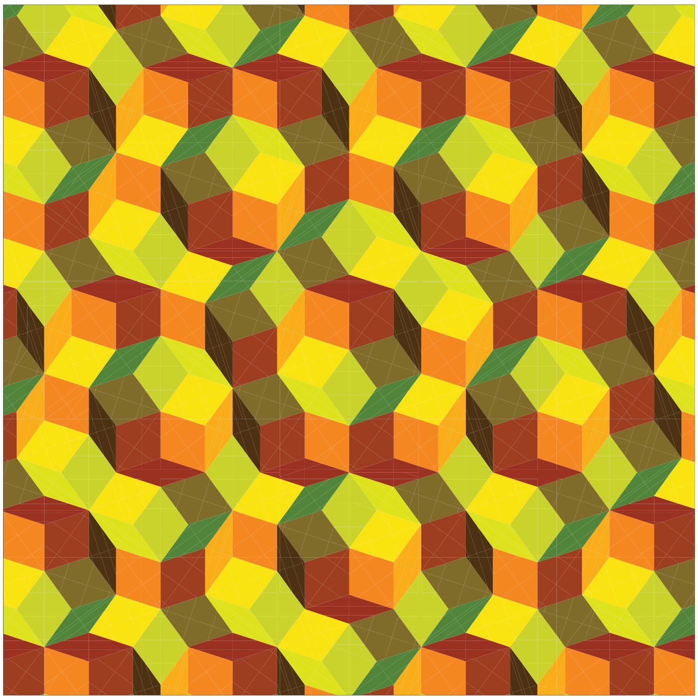
Aperiodic tilings

As we noted, the only regular polygons capable of tiling the Euclidean plane are the triangle, square and hexagon. Pentagons cannot, as three pentagons at each vertex have an interior angle sum of $\frac{9}{5}\pi$, which is slightly less than 2π and causes the pentagons to 'curl up' into a dodecahedron. Similarly, attempting to place four or more pentagons around each vertex results in a hyperbolic tiling, as $\frac{12}{5}\pi > 2\pi$.

More strongly, there is no tiling of the plane which exhibits both translational symmetry and order-5 rotational symmetry. To prove this, we assume without loss of generality that the tiling is fixed by both a translation parallel to the vector $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and a rotation by $\frac{2}{5}\pi$ about the origin. In that case, it is possible to map the origin to any point expressible as the sum of fifth roots of unity.



The points on the real axis expressible in this way are those of the form $a + b\,\phi$, where $a, b \in \mathbb{Z}$ and $\phi = \frac{1}{2}\left(1 + \sqrt{5}\right)$. As ϕ is an irrational number, these points form a *dense subset* of the reals, *i.e.* for every $\varepsilon > 0$, every point x on the real axis is within a distance of ε from a point of the form $a + b\,\phi$. This means that the tiling must be composed of infinitesimally small tiles, which contradicts our notion of discrete tiles.

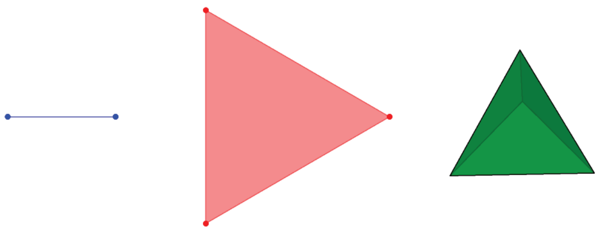


If we dispose of the translational symmetry, we can indeed have tilings with order-5 rotational symmetry. Perhaps the most famous is an aperiodic tiling known as the *Penrose tiling* (above), formed from interlocking 'thin' and 'thick' rhombi in the ratio $1 : \phi$. It is a remarkable fact that every tiling of the plane with these two tiles (and certain matching rules) exhibits this ratio, and is thus aperiodic (since ϕ is irrational). An unsolved problem is whether there is a *single* connected shape (an 'aperiodic monotile'), which can only tile the plane aperiodically. Joshua Socolar and Joan Taylor recently (2010) discovered a disconnected aperiodic monotile based on the hexagonal honeycomb, suggesting that there may indeed be a connected variant waiting to be found.

There is a three-dimensional analogue of the Penrose tiling. It is formed from equilateral parallelepipeds (three-dimensional rhombi) and displays icosahedral symmetry. Crystallographers were very surprised to find naturally occurring crystals with this structure, termed 'quasicrystals'. It was previously believed that solids could only be either periodic crystals or totally irregular.

Invariants

An *invariant* is, as suggested by the name, something that doesn't change. One of the simplest invariants is parity: whether something is even or odd. Integers are one of the most common things to display parity; however, the idea is equally applicable to other things such as permutations. To realise that permutations have a parity, it is necessary to consider them in a more geometrical light.



An *n-simplex* is a regular n-dimensional figure (polytope) with $n + 1$ vertices, which is fixed under any permutation of the vertices. The 1-simplex, 2-simplex and 3-simplex are the line segment, triangle and tetrahedron, respectively, as in the above diagram. Interchanging two of the vertices of a simplex can be regarded as a reflection. For example, reflecting a regular tetrahedron $ABCD$ with circumcentre O in the plane OCD causes the vertices A and B to be swapped.

This suggests two different sets of permutations: the *odd* permutations, which correspond to indirect isometries of $\mathbb{R}^n$; and *even* permutations, which correspond to direct isometries. A k-cycle (cyclic permutation of some subset containing k elements) is an odd permutation if k is even, and *vice-versa*. In particular, 2-cycles (or swaps) are odd permutations.

The set of even permutations of n elements forms a group known as the *alternating group* A_n. This is a subgroup of the group of all permutations, known as the *symmetric group* S_n. Any composition of even permutations is itself an even permutation, which can form a useful invariant. For example, it shows that not all conceivable configurations of a Rubik's cube can be attained by applying legal moves to the initial 'solved' position.

17. Suppose we have a hollow 4×4 square containing 15 unit square tiles and one empty space, into which any adjacent tile can be moved. The fifteen tiles are numbered from 1 to 15. Determine whether it is possible to get from the left-hand configuration to the right-hand configuration in the diagram below. **[Sam Loyd's 15 puzzle]**

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	

$\xrightarrow{?}$

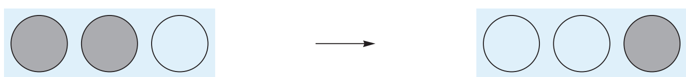
1	2	3	4
5	6	7	8
9	10	11	12
13	15	14	

Instead of an invariant, it is possible to define a value that only changes in one direction, known as a *monovariant*. This is useful for proving that a process (such as a perturbation argument) eventually terminates.

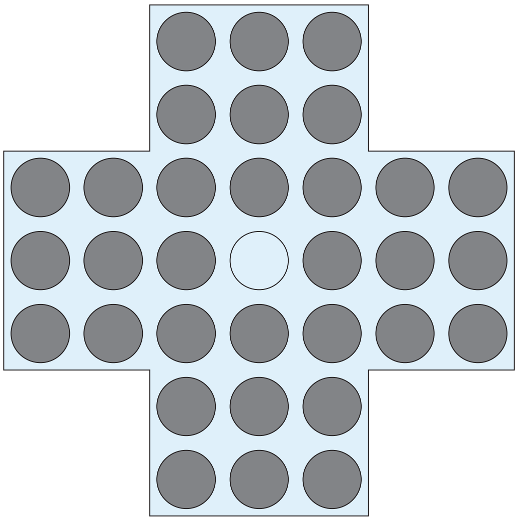
18. There are n red points and n blue points in the plane, no three of which are collinear. Prove that it is possible to pair each red point with a distinct blue point using n non-intersecting line segments. **[EGMO 2012, Friday bulletin]**

Solitaire

Quite a few interesting problems pertain to the game of peg solitaire. We have a (possibly infinite) board, which is a subset of $\mathbb{Z}^2$ containing some (possibly infinite) initial configuration of identical counters. The only allowed move is to jump horizontally or vertically over an occupied square to an unoccupied one; the piece that has been jumped over is removed. This is demonstrated below.



19. Suppose we have a game of solitaire on a bounded board beginning with the configuration of 32 pieces shown below. Show that if we can reach a position where only one piece remains on the board, then we can do so where the piece is in the centre.



20. We begin with an infinite chessboard, and divide the board into two half-planes with a straight horizontal line. All squares below the line are occupied with counters; all squares above the line are unoccupied. Show that it is impossible, after a finite sequence of moves, for a counter to occupy the fifth row above the line. **[Conway's soldiers]**

21. Suppose we have an infinite chessboard with an initial configuration of n^2 pieces occupying n^2 squares that form a square of side length n. For what positive integers n can the game end with only one piece remaining on the board? **[IMO 1993, Question 3]**

Solutions

1. We are enumerating strings containing precisely $k-1$ Rs and $n-k$ Gs. Hence, the number of ordered partitions of n into k subsets is given by the binomial coefficient $\binom{n-1}{k-1} = \frac{(n-1)!}{(k-1)!\,(n-k)!}$.

2. All k^6 colourings of the cube are fixed by the identity. Consider a rotation by $\frac{1}{2}\pi$ about the vertical blue axis. The top and bottom faces can be any colour, whereas the four other faces must all be the same colour. Hence, each of the 6 symmetries in this conjugacy class fix k^3 colourings. By similar reasoning, the 3 rotations by π about the blue axes each fix k^4 colourings. The 6 rotations about the red axes each fix k^3 colourings, whereas the 8 rotations by $\frac{2}{3}\pi$ about the green axes fix only k^2 colourings. Applying Burnside's lemma, the total number is $\frac{1}{24}\left(k^6 + 3\,k^4 + 12\,k^3 + 8\,k^2\right)$.

3. There are p symmetries, namely the identity and $p-1$ rotations. The former fixes all n^p colourings, whereas the latter fixes only the n monochromatic necklaces. Hence, we have $\frac{1}{p}\,(n^p + n(p-1))$ unique necklaces.

4. The result of the previous question is an integer, so $c^p + c(p-1)$ is divisible by p. Hence, $c^p + c\,p - c \equiv 0$. As $c\,p \equiv 0$, this means that $c^p \equiv c \pmod{p}$.

5. Note that $\varphi(N) = \varphi(p)\,\varphi(q) = (p-1)\,(q-1)$. Expressing b in terms of a, we obtain $b^e = a^{d\,e}$. As $a^{\varphi(N)} \equiv 1 \pmod{N}$ by Euler-Fermat, and $d\,e \equiv 1 \pmod{\varphi(N)}$, $a^{d\,e} \equiv a^1 = a \pmod{N}$, so is precisely the inverse function we are looking for.

6. The position of the rooks can be regarded as a bijection mapping rows to columns. There are $p!$ permutations of p elements.

7. Without loss of generality, just consider horizontal translations by $(a, 0)$. If there is a rook in (x, y), there must also be a rook in $(x+a, y)$, contradicting the assumption that the rooks are non-attacking.

8. Consider the rook positioned at the coordinates $(x, 0)$, and let the translation be parallel to vector (a, b). This forces there to be rooks in positions $(x+a, b)$, $(x+2\,a, 2\,b)$, …, $(x-a, -b)$. Hence, the arrangement is determined uniquely by the abscissa of the rook in the 0th row, of which there are p possibilities. Hence, p arrangements are fixed by each of these translations.

9. We have $\frac{1}{p^2}\left(p! + p(p-1)^2\right)$ distinct arrangements by Burnside's lemma.

10. The previous answer must be an integer, so $p! + p(p-1)^2 \equiv 0 \left(\operatorname{mod} p^2\right)$. Dividing throughout by p, we obtain $(p-1)! + (p-1)^2 \equiv 0 \pmod{p}$. We can expand this to yield $(p-1)! + p^2 - 2\,p + 1 \equiv 0 \pmod{p}$. As p^2 and $2\,p$ are divisible by p, we can eliminate those terms, resulting in the statement of Wilson's theorem.

11. Colour the squares black and white, as on a standard chessboard. The T-shaped tetromino must cover three black squares and one white square (or *vice-versa*), whereas each of the other tetrominoes cover precisely two squares of each colour. As the chessboard features equal numbers of black and white squares, this is indeed impossible.

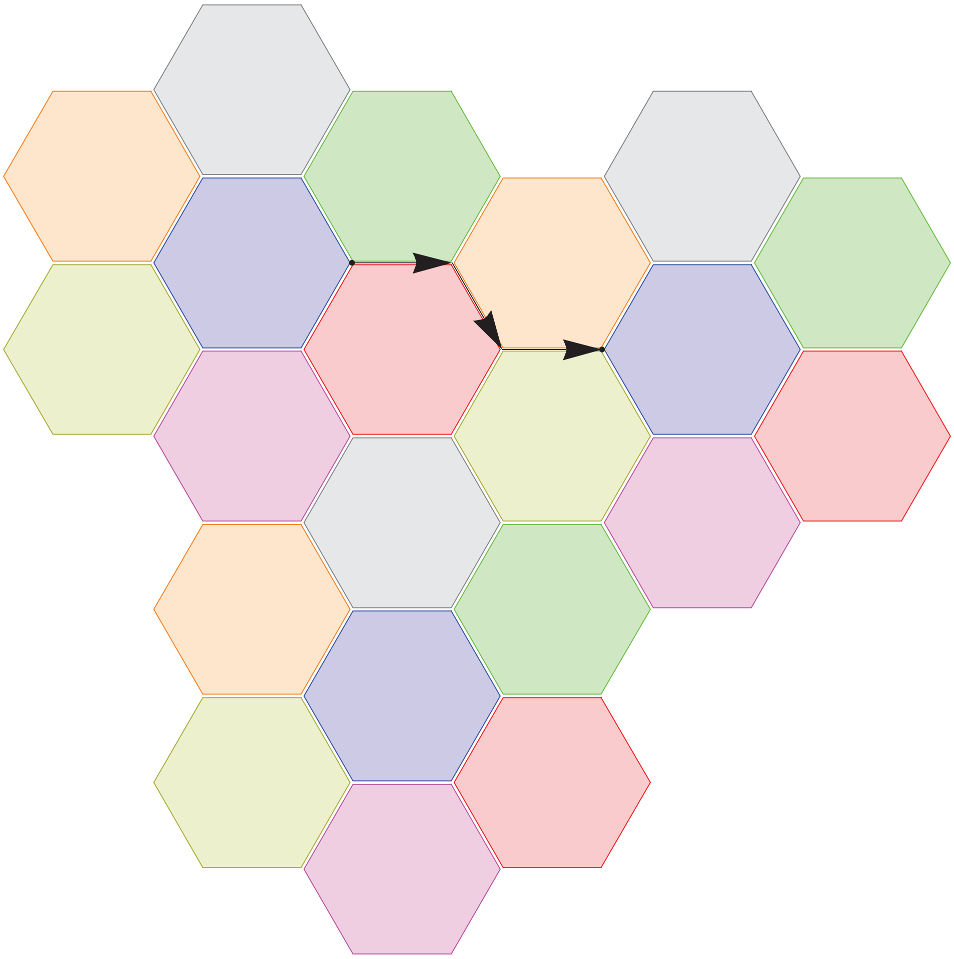
12. Colour the squares black and white, as on a standard chessboard. The six rooks are positioned on squares $(i, \sigma(i))$, where σ is a permutation of $\{1, 2, 3, 4, 5, 6\}$. Select two rooks at positions $(i, \sigma(i))$ and $(j, \sigma(j))$, and move them to $(i, \sigma(j))$ and $(j, \sigma(i))$, respectively. Applying this move does not alter the parity of rooks on white squares. Since we can do this until they lie on the long diagonal of white squares, it is clear that there must have been an even number of rooks on white squares to begin with. However, the constraint that the remaining 30 squares can be tiled by dominoes forces the rooks to occupy three white and three black squares, which contradicts the previous statement. Hence, it is impossible.

13. Represent each square with coordinates (x, y), where $x \in \{1, 2, \ldots, m\}$ and $y \in \{1, 2, \ldots, n\}$. Colour the square blue if $x \equiv y \equiv 0 \pmod{k}$, and white otherwise. Clearly, each tile must conceal precisely one blue square, and there are only $\lfloor \frac{m}{k} \rfloor \lfloor \frac{n}{k} \rfloor$ of them. This bound is attainable.

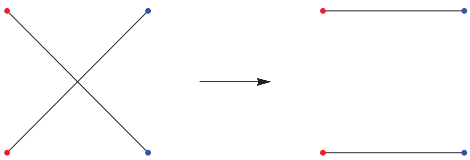
14. Let this number be denoted $f(n)$. Either the rightmost 2×1 rectangle is a (vertical) domino or the rightmost 2×2 rectangle is a pair of horizontal dominoes. Now consider how many ways there are of tiling the remaining area. In the first case, there are $f(n-1)$ possible configurations; in the second, there are $f(n-2)$. This gives us the recurrence relation $f(n) = f(n-1) + f(n-2)$. Together with the obvious fact that $f(1) = 1$ and $f(2) = 2$, this generates the Fibonacci sequence, $f(n) = F(n+1)$.

15. There are 60 symmetries of the regular dodecahedron. The identity symmetry fixes all 3^{12} infections. There are 24 rotations about axes passing through the centres of opposite faces, each of which fix 3^4 infections. The 15 rotations about axes passing through the midpoints of edges each fix 3^6 infections. Finally, the 20 rotations about axes passing through opposite vertices each fix 3^4 infections. By Burnside's lemma, there are $\frac{1}{60}\left(3^{12} + 24 \times 3^4 + 15 \times 3^6 + 20 \times 3^4\right) = 9099$ unique infections of the dodecahedron with three strains of bacteria.

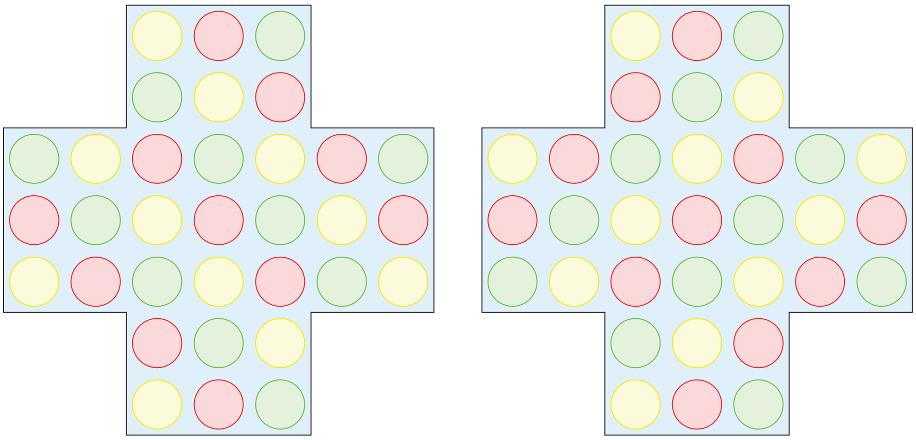
16. Tile the plane with the regular hexagonal tiling, where each hexagon has side length 1. Clearly, no two points in S can occupy the same hexagon. 7-colour the hexagons in a repetitive fashion, such that each hexagon is adjacent to six hexagons of different colours. By the pigeonhole principle, at least 15 of the points must lie in identically-coloured hexagons. It is straightforward to show that no two of those points can be within $\sqrt{7}$ of each other, by considering the closest approach of the vertices of the hexagons and using cube roots of unity to calculate the distance: the arrow shown in the honeycomb below has a complex vector of $2 - \omega$, which has squared length $(2 - \omega)\left(2 - \omega^2\right) = 4 - 2\left(\omega + \omega^2\right) + 1 = 7$.



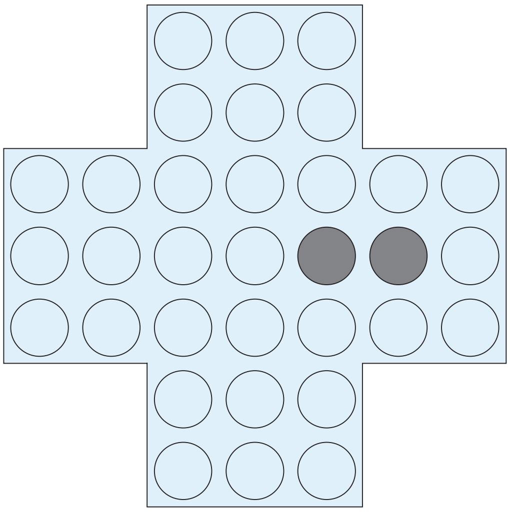
17. Label the empty space with 0, so we can regard this as a permutation of $\{0, 1, \ldots, 15\}$. Consider the parity of $x + y$, where (x, y) is the location of the empty space, together with the parity of the permutation σ. Note that each move flips both parities, thus leaving the total parity of $x + y + \sigma$ unchanged. However, interchanging any two tiles without moving the empty space alters the parity of $x + y + \sigma$, so it is impossible to get from the left configuration to the right configuration.



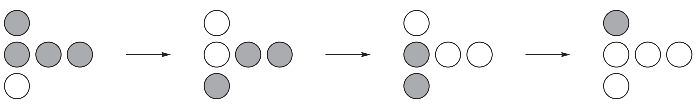
18. Biject them in an arbitrary way using n line segments. If we encounter a configuration of four points joined by two intersecting line segments, as above, then we can replace the line segments with disjoint line segments. Let the monovariant E be the total length of line segments. E strictly decreases at each step (by the triangle inequality), so the process cannot cycle. As there are only finitely many bijections between red and blue points, the process must terminate with n disjoint line segments.



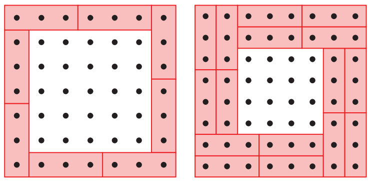
19. Firstly, colour the tile at coordinates (x, y) either red, green or yellow depending on the value of $x + y$ modulo 3, where we consider the central tile to be the origin (coloured red). As the parities of red, green and yellow counters all change simultaneously when a solitaire move is played, the final counter must be on a red square. However, we can also colour the tiles depending on $x - y$ modulo 3, resulting in a perpendicular pattern of colouring as shown above. The only tiles that are red in both colourings are given by $(3i, 3j)$, where i and j are integers. On the bounded board, there are only five such tiles. Backtracking by one move must result in a configuration equivalent to the one shown below, in which case we can trivially jump to the central square.



20. Assume that it is possible to reach a square in the fifth row, in attempt to derive a contradiction. Without loss of generality, we will use $T = (0, 0)$ as the 'target square'. For each square (x, y), we assign a value of $\phi^{-(|x|+|y|)}$, where $|x| + |y|$ is the Manhattan distance between (x, y) and $(0, 0)$, and $\phi = \frac{1+\sqrt{5}}{2}$ is the golden ratio. Let E be the sum of the values of the occupied squares. If a counter on a square of value ϕ^k jumps over one of value ϕ^{k+1}, this results in a single counter on a square of value less than or equal to ϕ^{k+2}. As $\phi^{k+2} = \phi^{k+1} + \phi^k$, the value of E cannot increase. At the beginning of the game, the value of E can be calculated by summing some geometric progressions; it is simple to show that this value equals 1. As the value of the target square is also 1, it is necessary to use all of the counters to reach it. However, that is impossible in a finite amount of time, as there are infinitely many counters.



21. For $n = 1$ and $n = 2$, this is trivial. If we have an arrangement shown above, it is possible to 'delete' three adjacent pieces. This can be used, rather effectively, to reduce a problem from $n = 3k + 4$ to $n = 3k + 2$ by deleting the outermost 'layer' of pieces, as in the diagram below. Similarly, we can reduce a problem from $n = 3k + 5$ to $n = 3k + 1$ by deleting the outermost two layers. By induction, we can solve the problem for all n except for multiples of three. If n is a multiple of three, we colour the tile at coordinates (x, y) either red, green or yellow depending on the value of $x + y$ modulo 3. Let the number of pieces on red, green and yellow tiles be indicated by R, G and Y, respectively. Note that if $(-1)^R = (-1)^G = (-1)^Y$ before a solitaire move, then it will remain true afterwards. This condition is clearly true for a $3k \times 3k$ square of pieces, but false for a single piece. Hence, we cannot reduce the arrangement to a single piece if n is a multiple of 3.

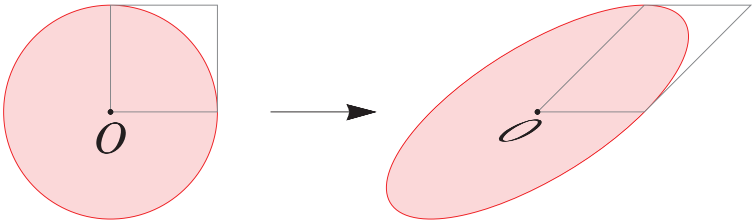


Linear algebra

The second chapter of this book is concerned with vectors, matrices and linear transformations. Determinants are introduced, together with ways in which to calculate them. These concepts are particularly relevant in analytic geometry, where we use them to describe projective transformations.

Linear transformations

Linear transformations are transformations of n-dimensional Euclidean space $\mathbb{R}^n$ expressible as $\underline{x} \to M\,\underline{x}$, where $\underline{x} = (x_1, x_2, \ldots, x_n)$ is the position vector of a point X. M is known as the *transformation matrix*. For example, the linear transformation with matrix $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ is shown below.

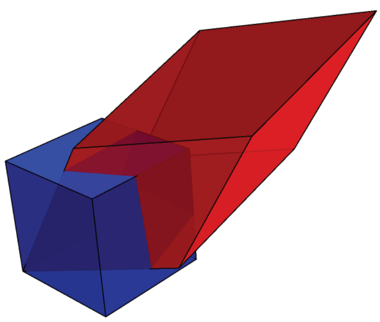


The position of the origin, O, is left unchanged by a linear transformation. Degree-d algebraic curves remain as degree-d algebraic curves; in particular, lines map to lines and conics map to conics. In the shear shown above, a circle is transformed into an ellipse. Parallel lines remain parallel when linear transformations are applied. Finally, the (signed) area of any shape is multiplied by $\det(M)$ when the transformation is applied, where $\det(M)$ is the *determinant* of the transformation matrix. Hence, ratios of areas remain unchanged.

Common linear transformations include rotations (about the origin), reflections (in lines through the origin), dilations (where the origin is the centre of homothety) and stretches (again, preserving the origin). One can combine transformations by multiplying their matrices.

1. Let $A = (1, 0, 0)$, $B = (0, 1, 0)$ and $C = (0, 0, 1)$ be three points in $\mathbb{R}^3$. After applying the transformation with matrix $M = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}$, find the new locations of A, B and C.

Consider the unit cube $[0, 1] \times [0, 1] \times [0, 1]$, where $\times$ denotes Cartesian product. It is transformed into a parallelepiped with volume $V = \det(M)$.



In the diagram above, the blue cube is transformed into the red parallelepiped. The origin (the common vertex of the cube and parallelepiped) remains fixed.

Determinants

The determinant of a square matrix M is a positive real number $\det(M)$ associated with that matrix. It behaves like the norm of a complex number, in that it is multiplicative.

- For two square matrices A and B of equal dimension, $\det(A\,B) = \det(A)\det(B)$. **[Multiplicativity of determinants]**

If a matrix A has an inverse matrix A^{-1} such that $A\,A^{-1} = A^{-1}\,A = I$, then $\det(A)\det\left(A^{-1}\right) = \det(I) = 1$. Hence, it is clear that a matrix with a determinant of zero has no inverse. Indeed, the converse is also true: all square matrices with non-zero determinants possess unique well-defined inverses. If a matrix is one-dimensional, then its determinant is equal to its only element. Otherwise, we compute it recursively.

$$M = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,n} \end{pmatrix}$$

Consider the matrix above. We compute the determinant using the following process:

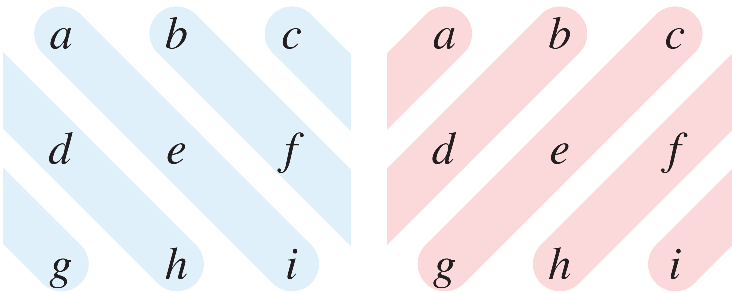
- For some $1 \le i \le n$, consider the element $a_{i,1}$ in the first column of M.
- The $(n-1)$-dimensional matrix M_i is obtained by removing everything in the same row or column as $a_{i,1}$.
- Compute the value $S_i = a_{i,1}\det(M_i)$.
- Then, we have $\det(M) = S_1 - S_2 + S_3 - S_4 + \ldots + (-1)^{n+1} S_n$.

This recursion results in the determinant equating to a sum of $n!$ terms, each of which is a product of n elements of M. After expanding this somewhat complicated recursive definition, we reach a more elegant formulation.

- $\det(M) = \sum_{\text{sym}} (-1)^{P(\sigma)} (a_{1,\sigma(1)}\, a_{2,\sigma(2)} \ldots a_{n,\sigma(n)})$, where the sum is taken over all permutations σ of $\{1, 2, 3, \ldots, n\}$. We define $P(\sigma)$ to be even if σ is an even permutation, and odd otherwise. **[Leibniz formula for determinants]**

2. Express $\det\begin{pmatrix} x & y & z \\ z & x & y \\ y & z & x \end{pmatrix}$ as a polynomial in x, y, z.

You may have noticed that for 3×3 determinants, the even permutations correspond to the three NW-SE 'diagonals' and the odd permutations correspond to the three NE-SW 'diagonals'. The diagonals are considered to wrap around the edges of the matrix as though it were a cylinder. This trick is known as the *Rule of Sarrus*.



$$a\,e\,i + b\,f\,g + c\,d\,h - a\,f\,h - b\,d\,i - c\,e\,g$$

Leibniz's formula requires $n(n!)$ elementary operations, so is rather time-consuming for large matrices, taking exponential time. Instead, it helps to simplify the calculation by performing operations on the matrix.

- Applying elementary operations to the rows or columns of M cause its determinant to behave in a predictable manner:
  - Multiplying any row or column of M by x causes the determinant of M to be multiplied by x;
  - Adding (or subtracting) any multiple of one row to another row does not affect the determinant of M;
  - Swapping any two rows causes $\det(M)$ to be multiplied by -1;
  - The transpose of M has the same determinant as M.

This can also be used to easily factorise the determinants of matrices.

3. Factorise $\det\begin{pmatrix} 1 & 1 & 1 \\ x & y & z \\ x^3 & y^3 & z^3 \end{pmatrix}$ into four linear factors.

Interpolating curves

The determinant of a matrix is zero if and only if one row can be expressed as a linear combination of the others. This is known as *linear dependence*. This enables one to create a curve of some type (*e.g.* a polynomial, circle or conic) interpolating between various points. For example, if we have a sequence of n points (x_i, y_i), then the following curve is a degree-$(n-1)$ polynomial passing through all n points.

- The curve $\det\begin{pmatrix} 1 & y & x & x^2 & x^3 & \cdots & x^{n-1} \\ 1 & y_1 & x_1 & {x_1}^2 & {x_1}^3 & \cdots & {x_1}^{n-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & y_n & x_n & {x_n}^2 & {x_n}^3 & \cdots & {x_n}^{n-1} \end{pmatrix} = 0$ passes through all points (x_i, y_i). **[Lagrange interpolating polynomial]**

This is obvious, as the determinant equals zero if two rows are identical. It is also a degree-$(n-1)$ polynomial, as we can use the recursive determinant formula to express it as $A_1 + A_2\, y + A_3\, x + A_4\, x^2 + A_5\, x^3 + \ldots + A_{n+1}\, x^{n-1}$ and rearrange it. If $A_2 = 0$ then this method will fail, but that only occurs if two points have the same abscissa.

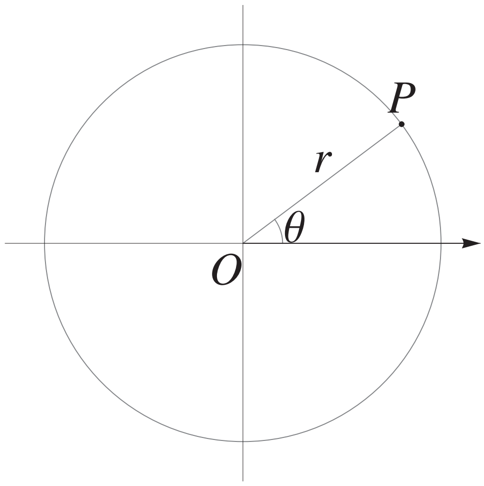
Using this idea, we can create a unique conic passing through any 5 points in general position, a cubic passing through 9 points *et cetera*. If the points are not in general position, then seemingly paradoxical things can occur. This forms the basis of the powerful *Cayley-Bacharach theorem* explored in the projective geometry chapter. The general equation of a conic is $A + B\, x + C\, y + D\, x^2 + E\, y^2 + F\, x\, y = 0$, so we can determine the equation of the conic passing through five given points.

- The conic $\begin{vmatrix} 1 & x & y & x^2 & y^2 & x\, y \\ 1 & x_1 & y_1 & {x_1}^2 & {y_1}^2 & x_1\, y_1 \\ 1 & x_2 & y_2 & {x_2}^2 & {y_2}^2 & x_2\, y_2 \\ 1 & x_3 & y_3 & {x_3}^2 & {y_3}^2 & x_3\, y_3 \\ 1 & x_4 & y_4 & {x_4}^2 & {y_4}^2 & x_4\, y_4 \\ 1 & x_5 & y_5 & {x_5}^2 & {y_5}^2 & x_5\, y_5 \end{vmatrix} = 0$ passes through all points (x_i, y_i). **[Interpolating conic]**

Circles also have a simple characterisation in Cartesian coordinates.

4. Find the equation of the circle passing through the non-collinear points (x_1, y_1), (x_2, y_2) and (x_3, y_3). **[Circumcircle equation]**

The determinant formula is not constrained to Cartesian coordinates; it can be used to find interpolating curves in any coordinate system, such as projective homogeneous coordinates, areal coordinates, complex numbers and even polar coordinates. As we cover the other coordinate systems in greater depth later in the book, it is worth messing around with polar coordinates here.



- The point with *polar coordinates* $P = \langle r,\ \theta\rangle$ in the Euclidean plane is defined such that $O\,P$ has length r and makes an angle of θ with the positive x-axis. In Cartesian coordinates, $P = (r\cos\theta,\ r\sin\theta)$. **[Definition of polar coordinates]**

Although the value of r is uniquely defined, θ is not; adding or subtracting multiples of $2\,\pi$ will describe the same point. This is a consequence of the periodicity of the elementary trigonometric functions.

5. Let $Q = \langle r_1,\ \theta_1\rangle$ be a point on the polar plane. Show that the equation of the circle with centre Q and radius a is given by $r^2 + {r_1}^2 - 2\,r\,r_1\cos(\theta - \theta_1) = a^2$. **[Polar equation of a circle]**

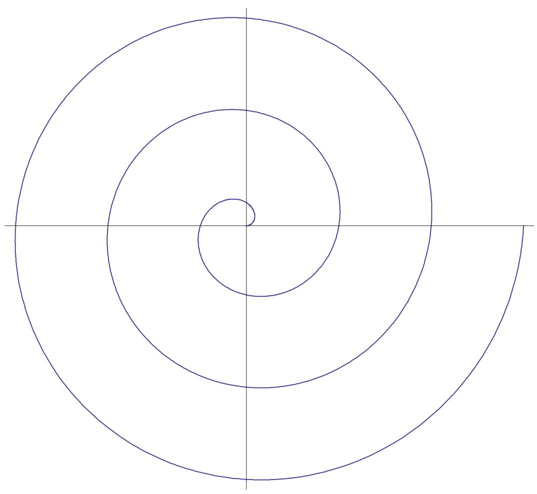
6. Hence show that a circle has general equation $A\,r^2 + B\,r\cos\theta + C\,r\sin\theta + D = 0$. **[General polar equation of a circle]**

It now becomes more obvious why this should work: the general equation for a circle in Cartesian coordinates is $A\left(x^2 + y^2\right) + B\,x + C\,y + D = 0$, and we have $x^2 + y^2 = r^2$, $x = r\cos\theta$ and $y = r\sin\theta$.

7. Find the equation, in polar coordinates, of the circle passing through the non-collinear points $\langle r_1,\ \theta_1\rangle$, $\langle r_2,\ \theta_2\rangle$ and $\langle r_3,\ \theta_3\rangle$. **[Circumcircle equation for polar coordinates]**

If three of the points are collinear, the term in r^2 vanishes and we are left with the equation of a line.

A curve which is particularly amenable to expressing in polar coordinates is the *Archimedean spiral*. If the spiral is centred on the origin, then it has polar equation $r = \frac{h}{2\,\pi}\,(\theta - \phi)$. h is the separation between successive turns of the spiral, and ϕ is the angle at which is emerges from the origin.

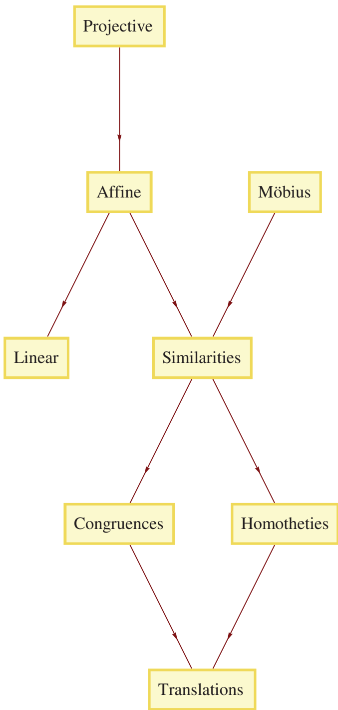


8. Find the equation for an Archimedean spiral of centre O passing through $\langle r_1, \theta_1 \rangle$ and $\langle r_2, \theta_2 \rangle$. **[Interpolating spiral]**

Adding multiples of 2π to either of the angles can alter the number of turns on the spiral and its direction. There is not a unique interpolating spiral with centre O passing through two given points; there are countably infinitely many.

Geometric transformations

So far, we have considered linear transformations. If we compose an arbitrary linear transformation with an arbitrary translation, then we obtain an *affine* transformation. Affine transformations have all the geometric properties of linear transformations, but do not necessarily preserve the origin. They are a special case of *projective* transformations, which are covered in a later chapter.



Affine transformations are projective transformations which preserve the line at infinity. Linear transformations also preserve the origin, whereas similarities preserve (or reverse) the circular points at infinity (thus mapping circles to circles). Congruences are similarities with a determinant of ± 1, whereas homotheties are similarities which preserve the direction of all lines (thus all points on the line at infinity). Translations (and reflections in a point) lie in the intersection of congruences and homotheties.

Do not worry if these terms are unfamiliar to you; they are explained properly in later chapters.

Scalar product

Let $\underline{a} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}$, $\underline{b} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$ and $\underline{c} = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix}$ be three vectors in $\mathbb{R}^3$.

- The dot product (or inner product, or scalar product) $\underline{a}\cdot\underline{b} = a_1\, b_1 + a_2\, b_2 + a_3\, b_3 = |\,\underline{a}\,|\ \ |\,\underline{b}\,|\ \cos\theta$, where θ is the angle between the vectors a and b. **[Definition of dot product]**

The dot product is commutative and distributive, so $\underline{a}\cdot\underline{b} = \underline{b}\cdot\underline{a}$ and $\underline{a}\cdot(\underline{b}+\underline{c}) = \underline{a}\cdot\underline{b} + \underline{a}\cdot\underline{c}$.

9. Prove that, for every triangle $A\,B\,C$, we have $a^2 = b^2 + c^2 - 2\,b\,c\cos A$. **[Law of cosines]**

The dot product generalises to vectors in $\mathbb{R}^n$. This allows us to interchange between trigonometric, geometric and algebraic inequalities.

- The following three statements are all equivalent:
  - $\cos\theta \le 1$, with equality if and only if $\theta = \pi\, n$ for some integer n;
  - $\underline{a}\cdot\underline{b} \le |\,\underline{a}\,|\ \ |\,\underline{b}\,|$, with equality if and only if the vectors have the same direction;
  - $a_1\, b_1 + a_2\, b_2 + \ldots + a_n\, b_n \le \sqrt{{a_1}^2 + {a_2}^2 + \ldots + {a_n}^2}\ \sqrt{{b_1}^2 + {b_2}^2 + \ldots + {b_n}^2}$, with equality if and only if $a_i = \lambda\, b_i$ for some scalar $\lambda \in [0, \infty]$. **[Cauchy-Schwarz inequality]**

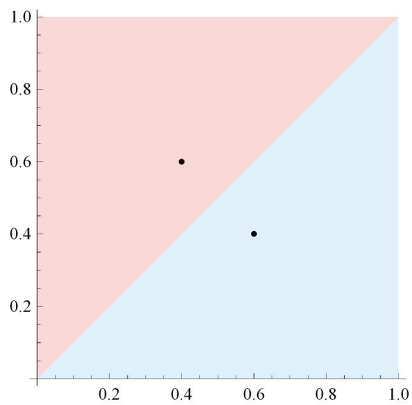
We can generalise the idea of a vector to a more abstract object, and thus extend the Cauchy-Schwarz inequality even further. See *Introduction to Inequalities* (Bradley) for an example of this.

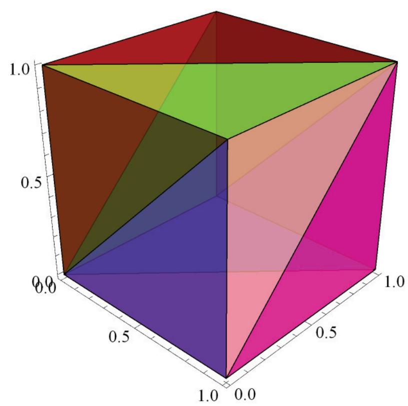
Another application of the dot product in inequalities is a proof of the *rearrangement inequality*. That states that if we have two non-negative sequences of equal length and multiply corresponding terms, the product is greatest when the sequences are sorted in the same order.

- Suppose that $a_1 \ge a_2 \ge \ldots \ge a_n \ge 0$ and $b_1 \ge b_2 \ge \ldots \ge b_n \ge 0$ are two decreasing sequences of non-negative integers. Then $\sum_{i=1}^{n} a_i\, b_i \ge \sum_{i=1}^{n} a_i\, b_{\sigma(i)}$ for any permutation σ. **[Rearrangement inequality]**

Proof:

We can prove this by considering the vectors $\underline{a}$ and $\underline{b}$ in the space $\mathbb{R}^n$. Observe that all $n!$ vectors in $\{\underline{b_\sigma}\}$ (the set of vectors obtained by permuting the elements of $\underline{b}$) are of equal length, so lie on a sphere with centre $\underline{0}$. The dot product $\sum_{i=1}^{n} a_i\, b_{\sigma(i)}$ of the vectors $\underline{a}$ and $\underline{b_\sigma}$ is greatest when the angle between them is smallest, which occurs when $\underline{a}$ and $\underline{b_\sigma}$ are closest (as all vectors in $\{\underline{b_\sigma}\}$ are of equal length). So, this has been converted into the equivalent problem of proving that $\underline{b}$ is the closest vector to $\underline{a}$ in $\{\underline{b_\sigma}\}$. We consider the *Voronoi diagram* of $\mathbb{R}^n$, which is simply a division of space depending on which $\underline{b_\sigma}$ is closest.





The diagrams above illustrate the cases when $n = 2$ or $n = 3$. The Voronoi diagram is created by the set of planes of the form $x_i = x_j$, which each partition space into the regions $x_i > x_j$ and $x_i < x_j$. This means that the regions of the Voronoi diagram are determined by the ordering of the elements; in the case where $n = 3$, we have six tetrahedral regions, namely $x_1 > x_2 > x_3$ and the five other permutations. As the elements of $\underline{a}$ and $\underline{b}$ are ordered in the same way, they must inhabit the same region. Hence, $\underline{b}$ is the closest vector in $\{\underline{b_\sigma}\}$ to $\underline{a}$, and we are finished.

Vector and triple products

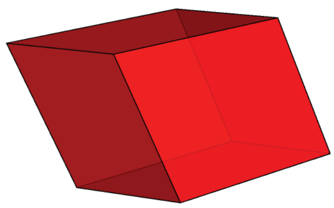
So far, we are able to 'multiply' two vectors in $\mathbb{R}^n$, resulting in a scalar. We can also define a vector (cross) product, which is specific to $\mathbb{R}^3$. (There is also a 7-dimensional version based on the octonion algebra, but that is outside the scope of the book.)

- The cross product (or vector product, or exterior product) $\underline{a} \times \underline{b} = \begin{pmatrix} a_2 b_3 - a_3 b_2 \\ a_3 b_1 - a_1 b_3 \\ a_1 b_2 - a_2 b_1 \end{pmatrix} = \det \begin{pmatrix} \underline{i} & a_1 & b_1 \\ \underline{j} & a_2 & b_2 \\ \underline{k} & a_3 & b_3 \end{pmatrix}$, where $\underline{i}$, $\underline{j}$, $\underline{k}$ are the unit vectors $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$, respectively. **[Definition of cross product]**

The cross product is **anti-**commutative and distributive, so $\underline{a} \times \underline{b} = -\underline{b} \times \underline{a}$ and $\underline{a} \times (\underline{b} + \underline{c}) = \underline{a} \times \underline{b} + \underline{a} \times \underline{c}$. The vector $\underline{a} \times \underline{b}$ is perpendicular to both $\underline{a}$ and $\underline{b}$, and its magnitude is equal to the area of the parallelogram with vertices $\{0, \underline{a}, \underline{b}, \underline{a} + \underline{b}\}$.

Finally, we define the scalar triple product, which is the volume of the parallelepiped with vertices $\{0, \underline{a}, \underline{b}, \underline{c}, \underline{a} + \underline{b}, \underline{b} + \underline{c}, \underline{c} + \underline{a}, \underline{a} + \underline{b} + \underline{c}\}$.

- $\underline{a} \cdot (\underline{b} \times \underline{c}) = (\underline{a} \times \underline{b}) \cdot \underline{c} = \det \begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix}$. **[Scalar triple product]**



Sir William Rowan Hamilton once had an epiphany whilst crossing a bridge, and carved the formula $i^2 = j^2 = k^2 = i\,j\,k = -1$ into one of the stones. This defines an extension to the complex numbers, which has four orthogonal units $(1, i, j, k)$ as opposed to two. A *Hamiltonian quaternion* is a number of the form $w + x\,i + y\,j + z\,k$, where $w, x, y, z \in \mathbb{R}$. Using a slight abuse of notation, this can be written as $w + \begin{pmatrix} x \\ y \\ z \end{pmatrix}$. A scalar added to a vector?! We can multiply two quaternions $p = a + \underline{b}$ and $q = c + \underline{d}$ together to give the quaternion $p\,q = (a\,c - \underline{b} \cdot \underline{d}) + (a\,\underline{d} + c\,\underline{b} + \underline{b} \times \underline{d})$. Multiplication of quaternions is associative and distributive, but not commutative; $p\,q \neq q\,p$ in general. This is inherited from the non-commutativity of the cross product.

The quaternion $w + x\,i + y\,j + z\,k$ has a norm of $\sqrt{w^2 + x^2 + y^2 + z^2}$. As with complex numbers, $|\,p\,|\;|\,q\,| = |\,p\,q\,|$ for any $p, q \in \mathbb{H}$, where $\mathbb{H}$ is the set of all quaternions.

Solutions

1. Using matrix multiplication, we get $A = \begin{pmatrix} a \\ d \\ g \end{pmatrix}$, $B = \begin{pmatrix} b \\ e \\ h \end{pmatrix}$ and $C = \begin{pmatrix} c \\ f \\ i \end{pmatrix}$.

2. $\det\begin{pmatrix} x & y & z \\ z & x & y \\ y & z & x \end{pmatrix} = x^3 + y^3 + z^3 - 3\,x\,y\,z$, as the NW-SE diagonals are x^3, y^3, z^3 and the NE-SW diagonals are each $x\,y\,z$.

3. We deduct the first column from the other two, obtaining $\det\begin{pmatrix} 1 & 0 & 0 \\ x & y - x & z - x \\ x^3 & y^3 - x^3 & z^3 - x^3 \end{pmatrix}$. Applying the recursion formula reduces this to $\det\begin{pmatrix} y - x & z - x \\ y^3 - x^3 & z^3 - x^3 \end{pmatrix}$. We then divide the first column by $y - x$ and multiply the entire determinant by $y - x$, obtaining $(y - x)\det\begin{pmatrix} 1 & z - x \\ y^2 + x^2 + x\,y & z^3 - x^3 \end{pmatrix}$. Applying a similar factorisation to the second column results in $(y - x)(z - x)\det\begin{pmatrix} 1 & 1 \\ x^2 + y^2 + x\,y & x^2 + z^2 + x\,z \end{pmatrix}$. Leibniz's formula can now be used to expand the determinant, giving $(y - x)(z - x)\left(x\,y - x\,z + y^2 - z^2\right)$. The quadratic factorises to $(y - z)(x + y + z)$, so the entire determinant is equal to $(y - x)(z - x)(y - z)(x + y + z)$.

4. $\det\begin{pmatrix} 1 & x & y & x^2 + y^2 \\ 1 & x_1 & y_1 & {x_1}^2 + {y_1}^2 \\ 1 & x_2 & y_2 & {x_2}^2 + {y_2}^2 \\ 1 & x_3 & y_3 & {x_3}^2 + {y_3}^2 \end{pmatrix} = 0$ will suffice, as the general equation for a circle is $A + B\,x + C\,y + D\left(x^2 + y^2\right) = 0$.

5. Let $P = \langle r, \theta\rangle$ be a point on the circle, so $P\,Q = a$. By using the cosine rule, we have $a^2 = r^2 + {r_1}^2 - 2\,r\,r_1\cos(\theta - \theta_1)$.

6. Using the compound angle formula, we get $r^2 - 2\,r\,r_1\cos\theta_1\cos\theta - 2\,r\,r_1\sin\theta_1\sin\theta + {r_1}^2 - a^2 = 0$. By altering θ_1 and r_1, we can change the coefficients of $r\sin\theta$ and $r\cos\theta$ to anything. Similarly, altering a enables us to change the constant term. Multiplying out by a constant scaling factor enables the coefficient of r^2 to be changed. Hence, the general equation is simply $A\,r^2 + B\,r\cos\theta + C\,r\sin\theta + D = 0$.

7. $\det\begin{pmatrix} 1 & r^2 & r\sin\theta & r\cos\theta \\ 1 & {r_1}^2 & r_1\sin\theta & r_1\cos\theta \\ 1 & {r_2}^2 & r_2\sin\theta & r_2\cos\theta \\ 1 & {r_3}^2 & r_3\sin\theta & r_3\cos\theta \end{pmatrix} = 0$.

8. The general spiral has equation $A + B\,r + C\,\theta = 0$, so an interpolating spiral is $\det\begin{pmatrix} 1 & r & \theta \\ 1 & r_1 & \theta_1 \\ 1 & r_2 & \theta_2 \end{pmatrix} = 0$.

9. Consider the triangle $O\,A\,B$. $\left(|\underline{a} - \underline{b}|^2\right) = (\underline{a} - \underline{b})\cdot(\underline{a} - \underline{b}) = \left(|\underline{a}|^2\right) + \left(|\underline{b}|^2\right) - 2\,\underline{a}\cdot\underline{b}$. The last term equates to $-2\,|\underline{a}|\;|\underline{b}|\cos\theta$. This is the cosine rule, as required.

Combinatorics II

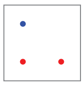
This chapter discusses Ramsey theory, graph theory and topology. The principal principle of Ramsey theory is that 'sufficiently large objects contain arbitrarily large homogeneous objects'. For example, *Ramsey's theorem* in graph theory states that one can find arbitrarily large monochromatic cliques in a sufficiently large complete graph coloured with c colours.

Gallai-Witt theorem

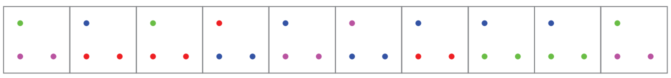
- Suppose we have a d-dimensional hypercube divided into g^d elements, each of which is coloured with one of c colours. If $g \geq G(d, c)$, where G is a function of d and c, then there exists some monochromatic (irregular) d-simplex homothetic to $\{(0, 0, 0, \ldots, 0), (1, 0, 0, \ldots, 0), (0, 1, 0, \ldots, 0), (0, 0, 1, \ldots, 0), \ldots, (0, 0, 0, \ldots, 1)\}$. **[Lemma 1]**

Proof:

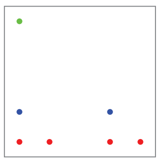
To prove this, we induct on the number of dimensions. For $d = 1$, this is trivially true by the pigeonhole principle: $G(1, c) = c + 1$, as any set of $c + 1$ elements must contain two of the same colour. We can use this as a starting point for proving the case for $d = 2$. Firstly, we can guarantee the existence of things like this, known as $(1, c, 1)$-objects:



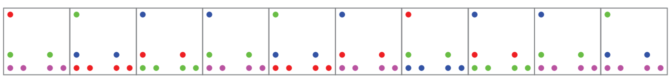
We assume that the top vertex is a different colour to either of the bottom vertices of the triangle, since otherwise we are done. We consider a strip of squares, each of size $G(1, c, 1) = G(1, c)$. They must each contain at least one $(1, c, 1)$-object, like so:



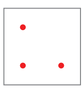
Moreover, each square can only have $c^{G(1,c,1)^2}$ possible states. Consider a row of $G\left(1, c^{G(1,c,1)^2}\right)$ such squares. At least two of them must be identical, so we can guarantee the existence of things like this, known as $(1, c, 2)$-objects:



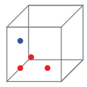
All three rows of points must necessarily have different colours. We define $G(1, c, 2) = G\left(1, c^{G(1,c,1)^2}\right) G(1, c, 1)$ to be an upper bound on the size of a box containing such an object. We then repeat the argument, considering a row of $G\left(1, c^{G(1,c,2)^2}\right)$ boxes of size $G(1, c, 2)$:



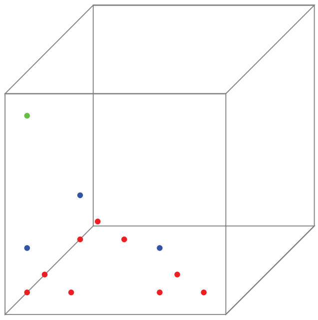
This gives us $G(1, c, f + 1) = G\left(1, c^{G(1,c,f)^2}\right) G(1, c, f)$. Now consider $(1, c, c)$-objects. They must have $c + 1$ rows of points, two of which must be the same colour by the pigeonhole principle. This means that we have a monochromatic isosceles right-angled triangle:



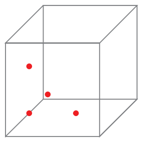
In other words, $G(2, c) = G(1, c, c)$. Now that we have tackled the two-dimensional case, we can begin work on three dimensions. We use the two-dimensional result $G(2, c, 1) = G(2, c)$ as a base case, and perform an identical inductive argument. Firstly, we can guarantee the existence of (2, c, 1)-objects within a box of side length $G(2, c, 1)$.



Now, consider a plane of $G\left(2, c^{G(2,c,f)^3}\right)$ such boxes. They must necessarily contain an isosceles right-angled triangle of identical boxes, by the theorem for $G(2, c)$, so we can guarantee a (2, c, 2)-object.



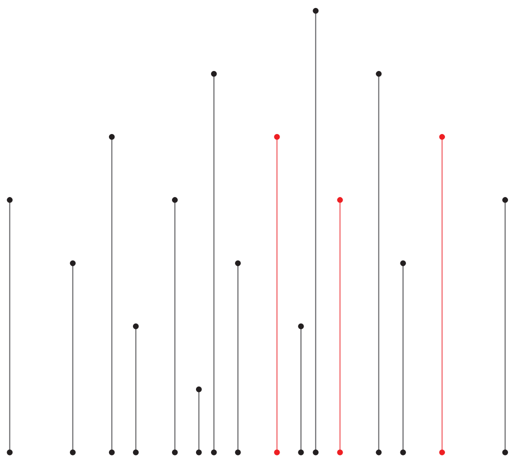
Proceeding in this manner, we can guarantee the existence of a (2, c, c)-object, therefore a monochromatic tetrahedron.



By repeating this argument and inducting on the number of dimensions, we can find upper bounds for $G(d, c)$ for all integers d and c. This concludes the proof of Lemma 1.

- Suppose we colour the elements of $\mathbb{Z}^n$ with c colours. Then, given a set of $d + 1$ vectors $\{0, a_1, a_2, \ldots, a_d\} \subset \mathbb{Z}^n$, we can find an integer $\lambda \in \mathbb{N}$ and vector $v \in \mathbb{Z}^n$ such that the points $\{v, v + \lambda a_1, v + \lambda a_2, \ldots, v + \lambda a_d\}$ are monochromatic. **[Gallai-Witt theorem]**

This theorem also holds if $\mathbb{Z}^n$ is replaced with $\mathbb{Q}^n$ or $\mathbb{R}^n$. To prove the Gallai-Witt theorem, we 'project' Lemma 1 from d dimensions onto a n-dimensional subplane by using a degenerate affine transformation.



For example, the existence of the monochromatic triangle in the diagram above proves the existence of a set of reals homothetic to (0, a, b) on the line below. This argument generalises very easily to prove the Gallai-Witt theorem.

1. Suppose c and n are integers. Prove that there exists an integer $w = W(c, n)$ such that any c-colouring of the integers $\{1, 2, \ldots, w\}$ contains a monochromatic arithmetic progression of length n. **[Van der Waerden's theorem]**

By the pigeonhole principle, at least one of these c subsets has a 'density' $\delta \geq \frac{1}{c}$. A generalised version of van der Waerden's theorem states that if $s \geq S(\delta, n)$, then any subset $A \subset \{1, 2, \ldots, s\}$ with $|A| \geq \delta s$ must contain an arithmetic progression of length n. This is known as *Szemeredi's theorem*. Allowing s to approach infinity, we can apply this to the set of positive integers and obtain the following theorem:

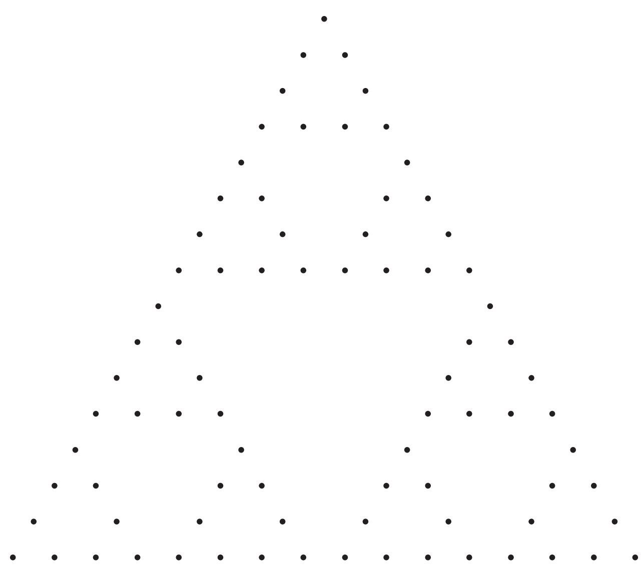
- If some subset of the positive integers has a non-zero asymptotic density, then it contains arbitrarily long arithmetic progressions.

Some other subsets of the positive integers contain arbitrarily long arithmetic progressions. Ben Green and Terence Tao proved that the prime numbers exhibit this property, despite having zero asymptotic density due to the prime number theorem. An unproven conjecture by Paul Erdős is that any set $\{a_1, a_2, a_3, \ldots\} \subset \mathbb{N}$ such that $\frac{1}{a_1} + \frac{1}{a_2} + \frac{1}{a_3} + \ldots$ diverges to infinity contains arbitrarily long arithmetic progressions. Of course, Szemeredi's theorem and the Green-Tao theorem are both special cases of this conjecture, since the prime harmonic series $\frac{1}{2} + \frac{1}{3} + \frac{1}{5} + \ldots$ diverges (albeit very slowly).

Our argument gives very weak upper bounds for the van der Waerden numbers (minimal values of $W(c, n)$). By considering Szemeredi's theorem, Tim Gowers currently has the tightest known upper bound, which is $W(c, n) \leq 2^{2^{c^{2^{n+9}}}}$. The lower bounds are merely exponential, so very little is known about the asymptotics of van der Waerden numbers.

Hales-Jewett theorem

Observe that in no part of our proof of Lemma 1 did we use the entire $\mathbb{Z}^n$. For the two-dimensional case, we only used a set of points corresponding to approximations to a fractal known as the *Sierpinski triangle*. The Sierpinski triangle is generated by beginning with a single point, then repeatedly placing three copies of it at the vertices of an equilateral triangle and scaling by $\frac{1}{2}$. Repeating this process four times, we obtain the order-4 approximation to the Sierpinski triangle, with $3^4 = 81$ points (shown below). The Sierpinski triangle is the limit, when this process is repeated infinitely.



We can associate points in the order-3 Sierpinski triangle with words of length 3 from the alphabet $\{A, B, C\}$, like so:

CCC

CCB CCA

CBC CAC

CBB CBA CAB CAA

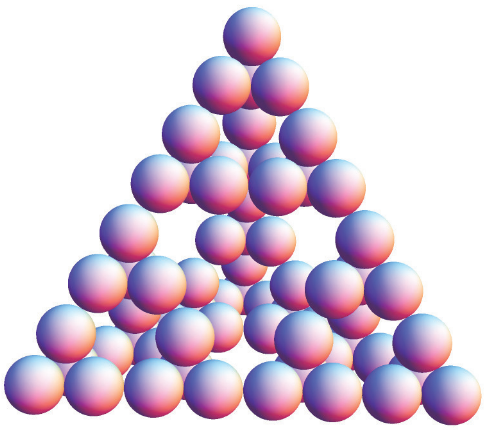
BCC ACC

BCB BCA ACB ACA

BBC BAC ABC AAC

BBB BBA BAB BAA ABB ABA AAB AAA

Suppose we introduce an additional symbol, $*$, which is considered to be a 'variable'. A word containing at least one asterisk is known as a *root*. The root $A\,B\,A * * C$ corresponds to the three words $A\,B\,A\,A\,A\,C$, $A\,B\,A\,B\,B\,C$ and $A\,B\,A\,C\,C\,C$, where $*$ successively takes on each of the three possible values. This set of three words is known as a *combinatorial line*. Note that combinatorial lines correspond to (upright) equilateral triangles of points in the Sierpinski triangle.



More generally, we can associate points in the order-h Sierpinski $(n-1)$-simplex with words from Σ^h, where Σ is an alphabet of n symbols.

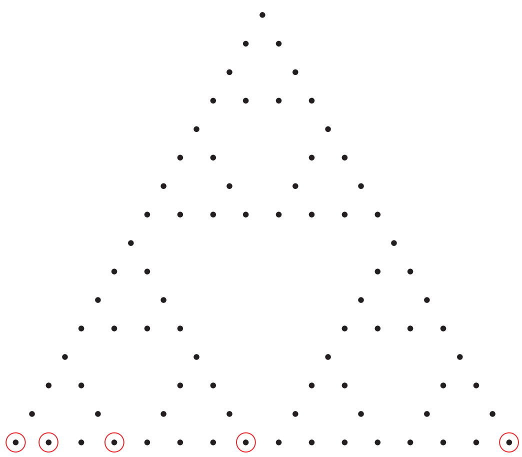
- Let Σ be an alphabet of n symbols. We colour each word of Σ^h with one of c colours. If $h \geq H(n, c)$, then there exists a monochromatic combinatorial line. **[Hales-Jewett theorem]**

So, it is equivalent to the following alternative formulation.

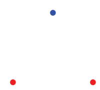
- Suppose we colour the vertices of an order-h Sierpinski $(n-1)$-simplex with c colours. If $h \geq H(n, c)$, there exists a monochromatic (upright) equilateral $(n-1)$-simplex. **[Hales-Jewett theorem]**

Proof:

The proof of Lemma 1 requires some slight refinement before it can be applied to prove Hales-Jewett. If we choose two generic points on the base of the Sierpinski triangle, then we cannot guarantee that there is a third vertex capable of completing the equilateral triangle. For example, $B\,A\,B$ and $A\,B\,A$ do not belong to a combinatorial line. So, we cannot merely apply the pigeonhole principle to the 2^h points on the base of the Sierpinski triangle. Instead, however, we *can* apply it to the $h+1$ points corresponding to 'powers of two'.



If we select any two of the $h + 1$ circled vertices, there is a third point capable of completing the equilateral triangle. Applying the Pigeonhole principle, we can let $h = c$ and there must be two circled vertices of the same colour. This proves the existence of $(1, c, 1)$-objects, *i.e.* equilateral triangles with two vertices of the same colour.

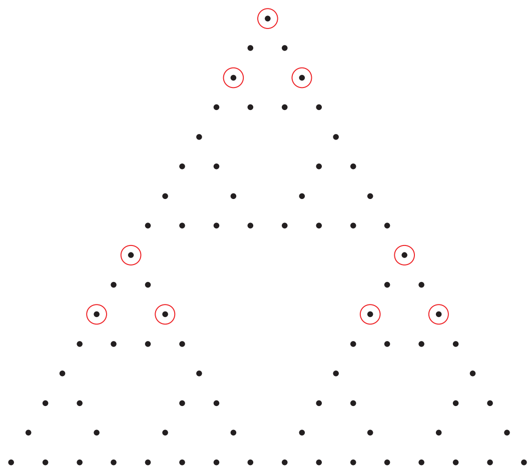


The remainder of the argument is identical to that of Lemma 1, so there is no need to repeat it here.

The Hales-Jewett theorem has a generalisation. If we allow roots with two variable symbols, such as $*$ and $\odot$, then we can create a *combinatorial* 2-*plane* by considering the set of n^2 words formed by replacing each variable with each of the letters in Σ. For example, if $\Sigma = \{A, B\}$, the root $A * B B \odot A *$ would correspond to the combinatorial 2-plane $\{A A B B A A A,\ A B B B A A B,\ A A B B B A A,\ A B B B B A B\}$. More generally, if we have roots with p variable symbols, then we generate a combinatorial p-plane of n^p words.

2. Let Σ be a finite alphabet of n symbols, and colour the words of Σ^j with c colours. Prove that if $j > J(n, c, p)$, then there exists a monochromatic combinatorial p-plane. **[Generalised Hales-Jewett]**

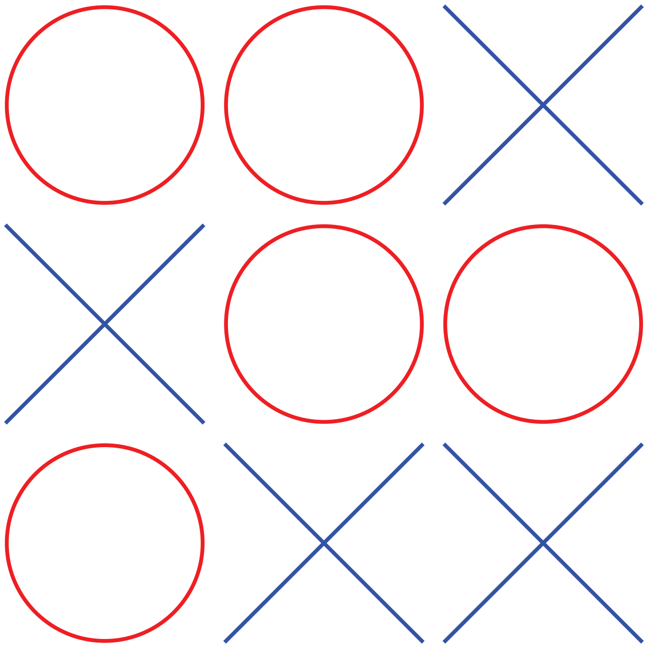
Here is an example of a combinatorial 2-plane on the order-4 Sierpinski triangle:



Just as van der Waerden's theorem has a stronger 'density' version (namely Szemeredi's theorem), so does the Hales-Jewett theorem and its generalisation. Using rather advanced methods, Furstenberg and Katznelson proved the theorem in 1991. More recently, a large collaborative effort (the *Polymath* project) led by Gowers resulted in an elementary combinatorial proof of the theorem, and thus Szemeredi's theorem and its multidimensional extension (the density version of Gallai-Witt).

Noughts and crosses

Effectively, the ordinary Hales-Jewett theorem states that in a c-player, h-dimensional game of tic-tac-toe on a board of size n, where $h \geq H(n, c)$, the game cannot terminate in a draw. Hence, one player has a winning strategy. The second player cannot have a winning strategy, as the first player can play randomly on the first move and emulate the winning strategy of the second player, knowing that owning an extra square cannot possibly be detrimental. This means that the first player can always win if the dimension is sufficiently large.



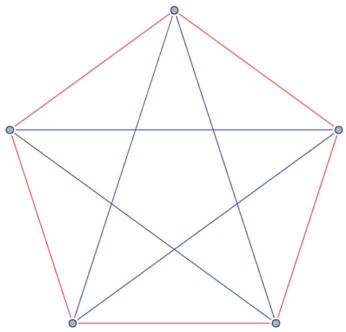
For ordinary 'noughts and crosses' where $c = 2$, $n = 3$ and $h = 2$, it is well known that neither player has a winning strategy. A typical drawing pattern is displayed above. It has been proved that $H(3, 2) = 4$, so it is impossible to draw in a four-dimensional game of tic-tac-toe. This does not necessarily mean that three-dimensional games terminate in a draw; certain diagonal lines are not considered to be combinatorial lines.

3. Does the first player have a winning strategy for $c = 2$, $n = 3$ and $h = 3$? **[Three-dimensional noughts and crosses]**

4. A game is played between two players on a 1 by 2010 grid. Taking it in turns, they place either an S or an O in an empty square. The game ends when three consecutive squares spell out $S\,O\,S$, at which point the player who has just played wins. If the grid fills up without this happening, the game is a draw. Prove that the second player has a winning strategy. **[Advanced Mentoring Scheme, November 2010, Question 2]**

Ramsey's theorem

Colour each of the edges of the complete graph K_n either red or blue. Let $R_2(r, b)$ be the smallest value of n such that there must be either a red K_r or blue K_b contained within the graph. For example, $R_2(3, 3) = 6$, as all colourings of K_6 contain a monochromatic triangle, whereas the following colouring of K_5 does not:



5. Prove that $R_2(r+1, b+1) \le R_2(r+1, b) + R_2(r, b+1)$. **[Bicoloured Ramsey's theorem]**

This argument generalises. If we colour the edges of K_n red, blue and green, where $n \ge R_2(r, g, b)$, then there must be either a red K_r, a green K_g or a blue K_b.

A further generalisation is by considering *hypergraphs* instead of graphs. Edges can be considered to be unordered pairs of vertices; if, instead, we colour unordered sets of k vertices, we obtain a *complete k-hypergraph*. It transpires that Ramsey's theorem generalises to hypergraphs.

- Let $\{C_1, C_2, \ldots, C_c\}$ be a set of c colours. Colour each unordered k-tuple of $\{1, 2, 3, \ldots, n\}$ with one of $\{C_1, C_2, \ldots, C_c\}$. Then, if $n \ge R_k(a_1, a_2, \ldots, a_c)$, there is some $1 \le i \le c$ and some subset of a_i vertices, all k-tuples of which are coloured with C_i. **[Generalised Ramsey's theorem]**

Proof:

We induct on the value of k. For $k = 1$, this reduces to the pigeonhole principle: $R_1(a_1, a_2, \ldots, a_c) = 1 + (a_1 - 1) + (a_2 - 1) + \ldots + (a_c - 1)$. Suppose we are trying to prove the existence of $R_k(a_1, a_2, \ldots, a_c)$. Let $n = 1 + R_{k-1}(R_k(a_1 - 1, a_2, \ldots, a_c), R_k(a_1, a_2 - 1, \ldots, a_c), \ldots, R_k(a_1, a_2, \ldots, a_c - 1))$. Select an arbitrary vertex V. By Ramsey's theorem for $(k-1)$-hypergraphs, we can guarantee that there must be, for some colour C_i, a set of $b = R_k(a_1, a_2, \ldots, a_i - 1, \ldots, a_c)$ vertices $\{W_1, W_2, \ldots, W_b\}$ such that every unordered k-tuple containing V and $k-1$ elements of $\{W_1, W_2, \ldots, W_b\}$ is coloured with C_i. Amongst those b vertices, there must either be a set of a_j vertices, all k-tuples of which are coloured with C_j (in which case we are done), or a set of $a_i - 1$ vertices, all k-tuples of which are coloured with C_i. Consider those $a_i - 1$ vertices together with V. All k-tuples of those a_i vertices are coloured with C_i, so the inductive step is complete. As the base case $R_k(0, a_2, \ldots, a_c)$ is trivial, we are done.

6. Prove that, for every $n \ge 3$, there exists an integer $k = K(n)$ such that every set S of k points in the plane in general position must contain a convex n-gon formed from n points of S. **[Happy ending problem]**

This problem is so named as it lead to the eventual marriage of George Szekeres and Esther Klein. Klein was responsible for discovering that $K(4) = 5$, and the result was generalised by Erdős and Szekeres.

Dilworth's theorem

The case of Ramsey's theorem for two colours and ordinary graphs gives exponential bounds on the number of vertices. With the base case of $R(2, n) = n$, it is evident that $R(n, m)$ is bounded above by the binomial coefficient $\frac{(n+m)!}{n!\,m!}$. If we make additional constraints on how the edges are allowed to be coloured, then we obtain a much stronger (indeed, optimal) bound.

- Suppose we define a relation $\succeq$ on the elements of a set S, such that, for all $a, b, c \in S$:
  - $a \succeq a$; **[Reflexivity]**
  - If $a \succeq b$ and $b \succeq a$ then $a = b$; **[Antisymmetry]**
  - If $a \succeq b$ and $b \succeq c$ then $a \succeq c$; **[Transitivity]**

  Then, $\succeq$ is known as a *partial order* on the elements of S. **[Definition of partial order]**

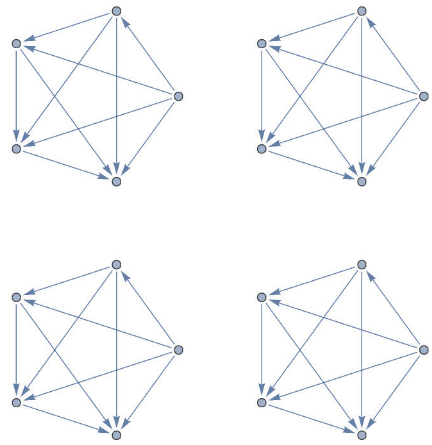
Common examples of partial orders include the relation $a \mid b$ on the set of positive integers and the relation $a \leq b$ on the set of real numbers.

- If neither $a \succeq b$ nor $b \succeq a$, then a and b are said to be *incomparable*. A set $\{c_1, c_2, \ldots, c_n\} \subset S$ such that $c_1 \succeq c_2 \succeq \ldots \succeq c_n$ is known as a *chain* of length n. A set $\{a_1, a_2, \ldots, a_m\} \subset S$ such that a_i and a_j are incomparable for all $i \neq j$ is known as an *antichain* of length m. **[Definition of chains and antichains]**

We can interpret the elements of S as the vertices of a complete graph, joined with a blue edge if the elements are incomparable and a red edge otherwise. Then, Ramsey's theorem guarantees that if $\left|S\right| \geq R(n, m) \leq \frac{(n+m)!}{n!\,m!}$, there must be either a chain of length n or an antichain of length m. However, it is possible to prove much tighter bounds than those applicable to general graphs.

7. Prove that if $|S| \geq (n-1)(m-1)+1$, then there is either a chain of length n or antichain of length m. **[Dilworth's theorem]**

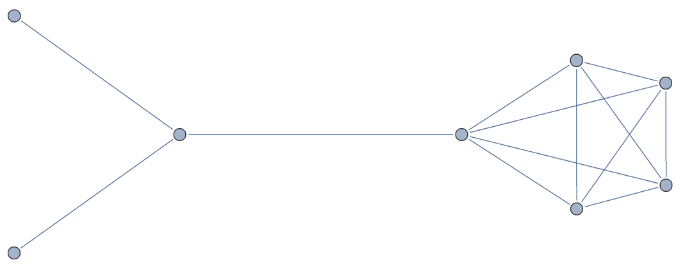
In general, $(n-1)(m-1)+1$ is much smaller than $\frac{(n+m)!}{n!\,m!}$. Note that Dilworth's theorem cannot be improved upon, as it is easy to define sets of $(n-1)(m-1)$ elements where the longest chain is length $n-1$ and the longest antichain is length $m-1$. For example, here is a partially ordered set of 20 elements where there are no chains of length 6 or antichains of length 5.



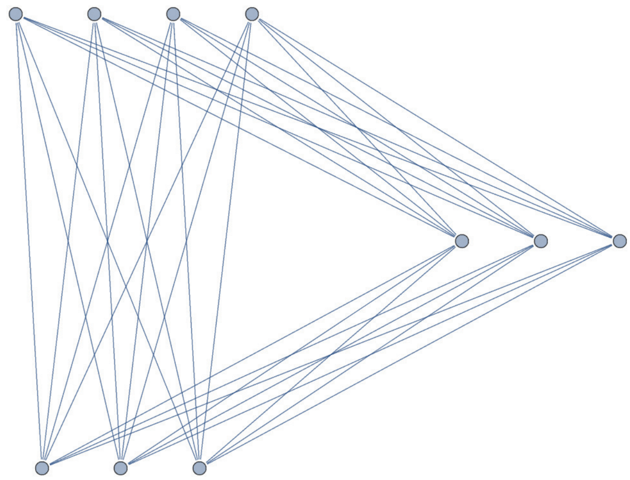
8. Show that a sequence of length $n\,m - n - m$ has either a monotonically increasing subsequence of length n or a monotonically decreasing subsequence of length m. **[Erdős-Szekeres theorem]**

Turán's theorem

Suppose we have a graph G of n vertices. If there exists a set of k vertices $K = \{A_1, A_2, \ldots, A_k\} \subseteq G$ such that every pair of vertices $A_i\,A_j$ is connected by an edge, then K is described as a k*-clique*. For example, the following graph contains a 5-clique:

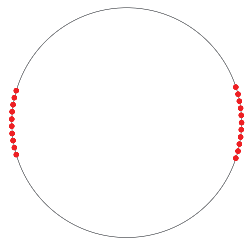


What is the maximum number of edges G can have such that there are no $(r+1)$-cliques? For $r = 1$, there can be no edges, since an edge is a 2-clique. For $r = n-1$, there can be $\binom{n}{2} - 1$ edges, since we can delete a single edge from the complete graph K_n. For other r, it turns out that the maximum number of edges is uniquely achieved by the *Turán graph* $T(n, r)$, which is constructed by partitioning the n vertices into r subsets of almost equal size (differing by at most 1) and joining two vertices if and only if they inhabit different subsets. For instance, the tetrahedron-free graph on 10 vertices with the most edges is shown below:



9. 21 apples are placed on the unit circle. Show that there are at least 100 line segments of length $\leq \sqrt{3}$ with Rosaceae endpoints. **[Ross Atkins, Trinity 2012]**

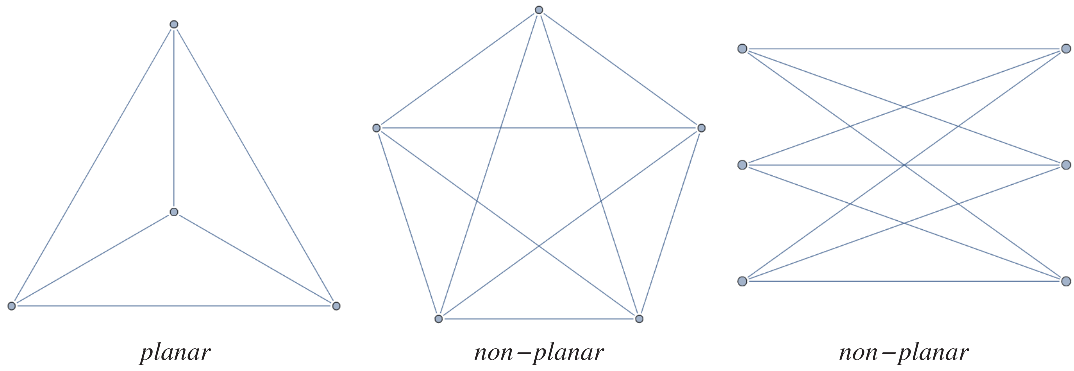
Again, this bound is attained by the configuration equivalent to the Turán graph, by separating the 21 apples into two groups of roughly equal size, situated near diametrically opposite points on the unit circle.



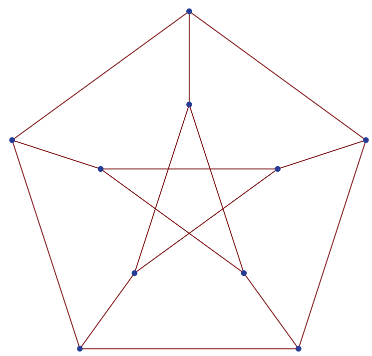
10. Given nine points in space, no four of which are coplanar, find the minimal natural number n such that for any colouring with red or blue of n edges drawn between these nine points there always exists a monochromatic triangle. **[IMO 1992, Question 3]**

Planar graphs

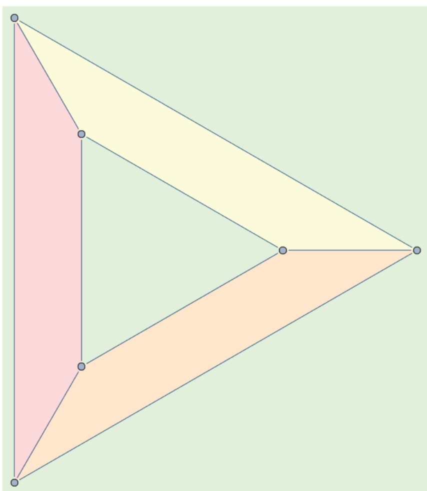
We describe a graph as *planar* if it can be drawn in the plane without any edges crossing. For example, the complete graph K_4 is planar, whereas K_5 and the complete bipartite graph $K_{3,3}$ are not.



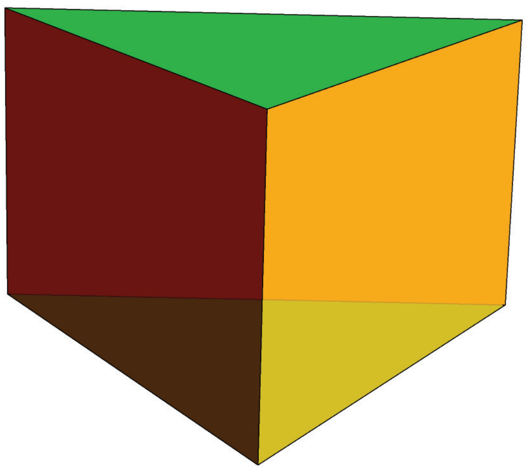
Indeed, a graph is planar if and only if it contains either K_5 or $K_{3,3}$ as a *minor*, *i.e.* can be reduced to one of these graphs by a combination of deleting edges, deleting vertices and contracting edges. For example, the (non-planar) *Petersen graph* below can be reduced to K_5 by contracting the five 'shortest' edges.



A planar graph divides the plane into well-defined regions (or *faces*). The following graph has five faces, four of which are bounded. We have used four colours such that neighbouring faces are different colours; in general, this is possible with any planar graph.



Indeed, it is best to append a point at infinity (this will become familiar to you later when we explore inversion) to convert the plane into a topological sphere. In this case, our planar graph is equivalent to the triangular prism.

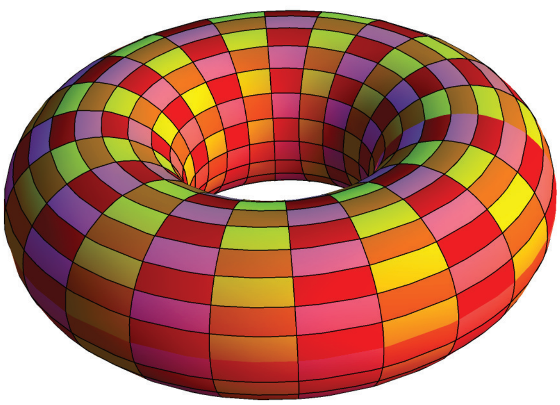


There is a useful invariant applying to graphs drawn on some surface. Suppose we have the following two elementary operations:

- A: inserting a vertex somewhere on an edge;
- B: drawing an edge between two vertices, ensuring the graph remains planar.

We are also allowed their inverse operations, A^{-1} and B^{-1}. The value of $\chi = V + F - E$ (where V, F, E are the numbers of vertices, faces and edges, respectively) is referred to as the *Euler characteristic*, and is unaffected by these elementary operations. As we can obtain all planar graphs (or, equivalently, polyhedra with no holes) from these operations, then the Euler characteristic is constant. It is a trivial exercise to verify for a simple polyhedron (such as the tetrahedron, with $(V, F, E) = (4, 4, 6)$), that the Euler characteristic must be 2.

Assuming no 'funny business' such as faces containing holes, the Euler characteristic is constant for all graphs drawn on a particular surface. Equivalently, it is constant for all polyhedra with a certain number of holes. For example, the Euler characteristic of a torus is 0. Every new hole decreases the Euler characteristic by 2, so by induction we have $\chi = 2 - 2H$, where H is the number of holes. More remarkably, an unbounded surface can be identified simply by its Euler characteristic and *orientability* (whether indirect isometries exist). For instance, the torus and *Klein bottle* are the orientable and unorientable surfaces, respectively, with an Euler characteristic of zero.



11. The *Klein quartic* is a surface topologically equivalent to a multi-holed torus, which can be tiled by 24 heptagons, where three heptagons meet at each vertex. What is its Euler characteristic, and thus how many holes does it have?

This description of the Klein quartic may remind you of the {7, 3} tiling of hyperbolic space. Indeed, it is obtained by 'rolling up' a finite patch of the hyperbolic tiling into a surface in the same way that a chessboard (finite patch of the square tiling {4, 4}) can be rolled up into an ordinary torus as in the diagram. Similarly, the Platonic solids are obtained by rolling up a finite patch (namely all) of a spherical tiling into a sphere. The Klein quartic has a symmetry group of order 336, far exceeding that of the most symmetrical Platonic solids (the icosahedron and dodecahedron, with 120 symmetries).

Solutions

1. This is a trivial corollary of Gallai-Witt in one dimension.

2. As the base case, $J(n, c, 1) = H(n, c)$. For the inductive step, we let $j = J\left(n, c^{n^{J(n,c,k)}}, 1\right)$. This means that there must be a monochromatic line of identical objects, each of which must contain a monochromatic k-plane. In total, this gives us a monochromatic $(k + 1)$-plane. Note that this is a ridiculously fast-growing function.

3. Yes. Place a 'nought' in the central cube. Assume the opponent plays a 'cross' in a cube C. Choose a cube A which is not collinear with C, and place a 'nought' in the cube diametrically opposite to A. This forces your opponent to place a 'cross' in A. As the two 'crosses' are non-collinear, you now have a free move. Place a 'nought' in a position coplanar with your existing two 'noughts'. This creates two partial lines; your opponent can only block one of them.

4. Define an 'unsafe move' to be one that results in the opponent winning on the subsequent move. The only unsafe move for placing an O is $S\ \_\ \_ \rightarrow S\ O\ \_$. Define an 'unsafe square' to be one where placing either an S or an O is an unsafe move. The only unsafe squares are of the form $S\ \_\ \_\ S$, which occur in pairs due to the bilateral symmetry. Hence, there is always an even number of unsafe squares. So, if there is a nonzero number of unsafe squares, the second player has a winning strategy as her opponent must eventually place a letter in an unsafe square, resulting in a win for the second player. To force a win, therefore, she must simply create an arrangement of the form $S\ \_\ \_\ S$. Immediately after the first player moves, the second player places a S sufficiently far from the first move. If the first player tries to block by playing close to the S, simply place another S on the opposite side, resulting in $S\ \_\ \_\ S$. As soon as the first player makes an unsafe move (which he inevitably will), the second player can immediately win.

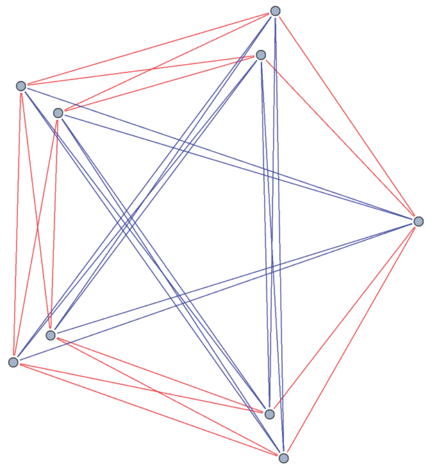
5. Consider a vertex V from the graph. It must have either at least $R(r, b + 1)$ red edges or $R(r + 1, b)$ blue edges connected to it. Assume, without loss of generality, that the former is true. Consider the set S of $R(r, b + 1)$ vertices connected to V via red edges. The subgraph on the vertices of S must then either contain a blue K_{b+1} (in which case we are done) or a red K_r. V is connected to each of the vertices of the K_r by a red edge, resulting in a red K_{r+1}.

6. It is straightforward to show that $K(4) = 5$, *i.e.* that every set of five points contains a convex quadrilateral, by considering all possible diagrams. This acts as a 'base case' to apply Ramsey's theorem for $K(n)$. Colour each 4-tuple of points blue if they are convex, or red if they are non-convex. By Ramsey's theorem for $R_4(5, n)$, there must be either a set of n points that form a convex polygon or a set of 5 points, no 4 of which form a convex polygon. Due to the base case, the latter is impossible, so the former must invariably be true.

7. Let $f : S \rightarrow \mathbb{Z}^+$ be a function mapping each element of S to the length of the longest chain terminating in S. If there are no n-chains, then the values may only range from 1 to $n - 1$. Similarly, if there are no m-antichains, then only $m - 1$ elements are allowed to take each value. So, there are at most $(n - 1)(m - 1)$ elements in S.

8. Consider the sequence $\{a_1, a_2, \ldots, a_{nm-n-m}\}$. For each a_i and a_j, we say that $a_i \geq a_j$ if both $i \geq j$ and $a_i \geq a_j$. Then, Dilworth's theorem guarantees that either a chain (increasing subsequence) or antichain (decreasing subsequence) exists.

9. Let G be the graph on 21 vertices, where two vertices share an edge if and only if they are separated by a distance greater than $\sqrt{3}$. As any three apples must form the vertices of a triangle, and one angle must be at most $\frac{\pi}{3}$, we can use the sine rule to deduce that one of the edges must be less than or equal to $\sqrt{3}$. Hence, G is triangle-free and has at most 110 edges by Turán's theorem. As we have a total of $\binom{21}{2} = 210$ pairs of apples, there are at least $210 - 110 = 100$ line segments of length $\leq \sqrt{3}$.



10. To prove that $n > 32$, consider the Turán graph above on nine vertices containing no 6-cliques. It has 32 edges. We call the five subsets of vertices A_1, A_2, A_3, A_4, A_5, and join vertices $X \in A_i$ and $Y \in A_j$ with a red edge if $(i - j) \equiv \pm 1 \pmod 5$, a blue edge if $(i - j) \equiv \pm 2 \pmod 5$, and no edge if $i = j$. To prove that $n \leq 33$, note that all graphs with 33 edges must contain a 6-clique by Turán's theorem. This 6-clique must contain a monochromatic triangle by Ramsey's theorem.

11. Each vertex is adjacent to three heptagons, and each heptagon has seven vertices. Hence, there must be $24 \times \frac{7}{3} = 56$ vertices. Similarly, each heptagon has seven edges, and each edge is adjacent to two heptagons, so there are 84 edges. The Euler characteristic is $24 + 56 - 84 = -4$, so the Klein quartic is topologically a three-holed torus.

Polynomials

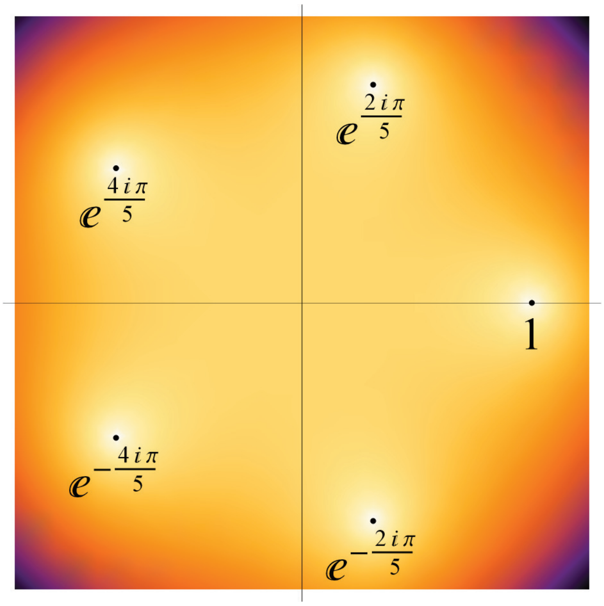
The set of polynomials $\{f(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n\}$, where n is a non-negative integer and $\{a_0, a_1, \ldots, a_n\} \subset S$, is denoted $S[x]$ (pronounced 'S adjoin x'). In this chapter, we will explore the cases where S is the set of complex numbers or real numbers.

Complex polynomials, $\mathbb{C}[x]$

Suppose we have a degree-n polynomial $f(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n$, where $a_0, \ldots, a_n$ are complex constants and a_n is non-zero. According to the *fundamental theorem of algebra*, we can express it as a product of linear factors of the form $x - \alpha_i$, where α_i is a (complex) root of the polynomial.

- If we have a monic polynomial $f(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n$, where $a_n = 1$, then we can express $f(x) = (x - \alpha_1)(x - \alpha_2)\ldots(x - \alpha_n)$, where $\alpha_1, \ldots, \alpha_n$ are (not necessarily distinct) roots of the polynomial. **[Fundamental theorem of algebra]**

For example, the polynomial $x^4 - 1$ can be factorised as $(x - 1)(x + 1)(x + i)(x - i)$, where $i = \sqrt{-1}$ is the imaginary unit. In general, the polynomial $x^n - 1 = (x - 1)(x - \tau)\left(x - \tau^2\right)\ldots\left(x - \tau^{n-1}\right)$, where $\tau = e^{\frac{2\pi i}{n}}$ is a principal nth root of unity. The roots of unity are positioned at the vertices of a regular n-gon with centre 0 and a vertex at 1. The example for $x^5 - 1$ is shown below.



This means that the degree-n curve $y = f(x)$ meets the degree-1 line $y = 0$ in at most n points. There is nothing special about the line $y = 0$, and this also applies to any line. More remarkably, the polynomial curve can be replaced with any algebraic curve (such as the unit circle, $x^2 + y^2 = 1$, which has degree 2). Even more generally, where the line is replaced with another algebraic curve, we have *Bezout's theorem*.

- Suppose P and Q are two curves of degrees m and n, respectively. If they intersect in finitely many points, then they intersect in at most $m\,n$ points. **[Bezout's theorem]**

Equality occurs if we consider intersections on the complex projective plane (instead of the real plane), and count intersections with appropriate multiplicity (*e.g.* twice for tangency, thrice for osculation *et cetera*). The complex projective plane is discussed in later chapters, and it is only necessary at this point to use the weak form of Bezout's theorem.

1. Show that two ellipses intersect in at most four points.

2. Consider the regular n-gon with vertices $A_1, A_2, \ldots, A_n$, where $n \geq 5$. Let P be a variable point on the circumcircle of the n-gon. Show that the value of $f(P) = A_1 P^4 + A_2 P^4 + \ldots + A_n P^4$ remains constant.

Difference of three cubes

A particularly useful polynomial is $x^3 + y^3 + z^3 - 3xyz$. Over the reals, it factorises into $(x+y+z)\left(x^2+y^2+z^2-xy-yz-zx\right)$, and the quadratic term can be further factorised over the complex numbers. This polynomial recurs in many situations, including olympiad problems.

- The polynomial $x^3 + y^3 + z^3 - 3xyz = \det\begin{pmatrix} x & y & z \\ z & x & y \\ y & z & x \end{pmatrix} = (x+y+z)(x+\omega y+\omega^2 z)(x+\omega^2 y+\omega z)$, where ω is a principal cube-root of unity. **[Difference of three cubes]**

This generalises to n variables, instead of merely three. Indeed, the name is derived from the $n = 2$ case, known as the 'difference of two squares', $x^2 - y^2 = \det\begin{pmatrix} x & y \\ y & x \end{pmatrix} = (x+y)(x-y)$.

- $\det\begin{pmatrix} x_0 & x_1 & x_2 & \cdots & x_{n-1} \\ x_{n-1} & x_0 & x_1 & \cdots & x_{n-2} \\ x_{n-2} & x_{n-1} & x_0 & \cdots & x_{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_1 & x_2 & x_3 & \cdots & x_0 \end{pmatrix} = \prod_{r=0}^{n-1}\left(\sum_{r=0}^{n-1}\left(x_r\, e^{2\pi i \frac{k}{n} r}\right)\right).$

The sums in the product on the right-hand side are the terms of the *discrete Fourier transform* of $\{x_0, x_1, \ldots, x_{n-1}\}$. (Continuous) Fourier transforms were originally discovered to explain how the sound of an entire orchestra can be composed of basic sinusoidal waves. Today, this idea is used to analyse electrical circuits. The discrete Fourier transform can be computed quickly using certain algorithms, forming the basis of the fastest known algorithm for multiplying two large integers.

3. Find the minimum value of $x^2 + y^2 + z^2$, where x, y, z are real numbers such that $x^3 + y^3 + z^3 - 3xyz = 1$. **[BMO2 2008, Question 1]**

Cubic equations

For sufficiently small degree, it is possible to solve polynomial equations using *radicals* (nth roots). For the general quadratic equation, the Babylonian technique of 'completing the square' suffices. Solving the cubic equation is a more difficult, multi-step process. If we can immediately find a root α, it is possible to divide by $x - \alpha$ to reduce the equation to a quadratic. Otherwise, more ingenious techniques are required.

4. Suppose we have an equation $a y^3 + b y^2 + c y + d = 0$. Show that this can be converted to an equation of the form $x^3 + px + q = 0$, where x is a linear function of y. **[Reduction to a monic trinomial]**

Hence, it is only necessary to consider the latter case, as all other cubics can be reduced to it.

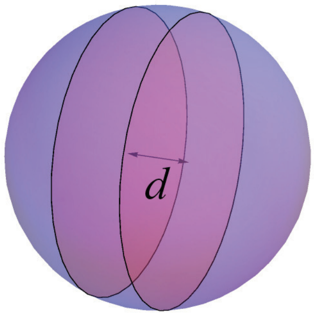
5. Show that $x^3 + px + q = 0$ is equivalent to $x^3 + a^3 + b^3 - 3abx = 0$, where a^3 and b^3 are the roots of the quadratic equation $z^2 - qz - \frac{p^3}{27} = 0$.

6. Hence show that $x^3 + px + q = 0$ has a root

$x = -(a+b) = \sqrt[3]{-\frac{1}{2}q + \sqrt{\frac{1}{4}q^2 + \frac{1}{27}p^3}} + \sqrt[3]{-\frac{1}{2}q - \sqrt{\frac{1}{4}q^2 + \frac{1}{27}p^3}}$. **[General solution to cubic equations]**

The general solution to the cubic equation was the quest of many mediæval mathematicians. After a partial solution by Omar Khayyam, the first complete solution was by Niccolò Tartaglia. However, when the quartic was later solved by Lodovico Ferrari, Tartaglia's solution of the cubic was mistakenly attributed to Gerolamo Cardano, and is thus referred to as *Cardano's formula*. This displeased Tartaglia to a great extent.

Due to the existence of formulae for solving the general quadratic, cubic and quartic equations, people imagined that there might be similar algebraic methods for solving any polynomials using radicals. However, this is not the case. Galois theory demonstrates that it is impossible to solve the general quintic, and of course polynomial equations of higher degree.



7. Two parallel planes cut the sphere of unit radius into three equal volumes. Find a cubic equation in rational coefficients, one root of which is the separation d between the two planes.

Symmetric polynomials

So far, we have mainly considered polynomials in one variable. The difference of three cubes was an exception, as it featured three variables. Indeed, it is what is known as a *symmetric polynomial*, as interchanging any two of the variables leaves the polynomial unchanged. $x^2 - y^2$ is not symmetric, since interchanging x and y negates the value, rather than preserving it. $(x - y)^2$, however, is symmetric.

8. Suppose we have a symmetric polynomial in two variables, x and y. Show that it can be expressed as a polynomial in s and p, where $s = x + y$ and $p = x\,y$.

This is a special case of Newton's theorem of symmetric polynomials:

- Any n-variable symmetric polynomial in $x_1, x_2, \ldots, x_n$ can be expressed as a polynomial in the *elementary symmetric polynomials (ESPs)*, i.e. coefficients of $(x - x_1)(x - x_2)\ldots(x - x_n)$. **[Newton's theorem of symmetric polynomials]**

Proof:

Note that any symmetric polynomial can be multiplied out to yield a sum of terms of the form $k \sum_{\text{sym}} (x_1^{a_1} x_2^{a_2} \ldots x_n^{a_n})$, where the sigma indicates a symmetric sum. We will represent this as $k\,f(a_1, a_2, \ldots, a_n)$.

We assume without loss of generality that $a_1 \geq a_2 \geq \ldots \geq a_n$, and lexicographically order the terms. (Specifically, $f(a_1, a_2, \ldots, a_n)$ precedes $f(b_1, b_2, \ldots, b_n)$ if $a_1 < b_1$, or $a_1 = b_1$ and $a_2 < b_2$, or $a_1 = b_1$ and $a_2 = b_2$ and $a_3 < b_3$, *et* *cetera*.) We then proceed via an inductive argument.

For a given term $k\,f(a_1, a_2, \ldots, a_n)$ of the ordering, assume that all preceding terms can indeed be expressed as polynomials in the ESPs. Suppose that $a_1 = a_2 = \ldots = a_h > a_{h+1}$. We subtract the symmetric polynomial $k\,f(1, 1, \ldots, 1, 0, \ldots, 0)\,f(a_1 - 1, a_2 - 1, \ldots, a_h - 1, a_{h+1}, a_{h+2}, \ldots, a_n)$. Since $f(1, 1, \ldots, 1, 0, \ldots, 0)$ is already an elementary symmetric polynomial, and $f(a_1 - 1, a_2 - 1, \ldots, a_h - 1, a_{h+1}, a_{h+2}, \ldots, a_n)$ is a symmetric polynomial of lower degree than the original expression, the term we have subtracted can be expressed as a polynomial in ESPs. The remainder exclusively contains terms that precede $f(a_1, a_2, \ldots, a_n)$, thus can also be expressed as a polynomial in the ESPs.

9. Given real numbers a, b, c, with $a + b + c = 0$, show that $a^3 + b^3 + c^3 > 0$ if and only if $a^5 + b^5 + c^5 > 0$.
[BMO2 2004, Question 3]

Solutions

1. Ellipses can be represented by quadratic equations in x and y (like all conic sections). As a consequence of Bezout's theorem, they can intersect in no more than $2\times 2 = 4$ points.

2. Consider an arbitrary point Q in general position on the circumcircle of the n-gon, and consider the curve $f(P) = f(Q)$. It is a quartic curve (by definition) and must intersect the circumcircle in $2n$ points (rotations and reflections of Q). Due to Bezout's theorem, a quartic and circle sharing no common component can only intersect in at most 8 points; however, $2n > 8$, so the quartic must contain the circle. Hence, all points on the circle satisfy $f(P) = f(Q)$.

3. $x^3 + y^3 + z^3 - 3xyz = \det\begin{pmatrix} x & y & z \\ z & x & y \\ y & z & x \end{pmatrix}$ is the volume of the equilateral parallelepiped with vertex $O = (0, 0, 0)$ and adjacent vertices $A = (x, y, z)$, $B = (y, z, x)$ and $C = (z, x, y)$. Suppose we fix $x^2 + y^2 + z^2 = r^2$, the square of the side length, instead of fixing the volume. If O, A and B are constrained to lie in a plane, then the volume is maximised when $\overrightarrow{OC}$ is a normal to this plane. Hence, the volume is maximised relative to the side length when the parallelepiped is a cube with volume r^3. As such, the minimum value of r^2 for a parallelepiped with unit volume is 1, when $\{x, y, z\} = \{1, 0, 0\}$.

4. Firstly, we divide by a to obtain $x^3 + \frac{b}{a}x^2 + \frac{c}{a}x + \frac{d}{a} = 0$. Let $y = x + \frac{b}{3a}$. Then, our cubic becomes $\left(y - \frac{b}{3a}\right)^3 + \frac{b}{a}\left(y - \frac{b}{3a}\right)^2 + y - \frac{b}{3a} + \frac{d}{a} = 0$. By using the binomial expansion of the first two terms, the coefficients in y^3 and y^2 are one and zero, respectively.

5. We let $q = a^3 + b^3$ and $p = -3ab$. a^3 and b^3 are the roots of $\left(z - a^3\right)\left(z - b^3\right) = 0$, which expands to $z^2 - \left(a^3 + b^3\right)z + a^3 b^3 = 0$. We already have $a^3 + b^3 = q$, and $a^3 b^3 = (ab)^3 = \left(\frac{-p}{3}\right)^3 = -\frac{p^3}{27}$.

6. $x^3 + a^3 + b^3 - 3abx = (x + a + b)\left(x + \omega a + \omega^2 b\right)\left(x + \omega^2 a + \omega b\right) = 0$ has roots $x = -(a + b)$, $x = -\left(\omega a + \omega^2 b\right)$ and $x = -\left(\omega^2 a + \omega b\right)$. Using the Babylonian formula for solving the quadratic equation gives us the values of a^3 and b^3, whence we can obtain a and b by cube-rooting.

7. The volume enclosed by the two planes is given by $\int_{-\frac{d}{2}}^{\frac{d}{2}} \pi y^2 \, dx = \int_{-\frac{d}{2}}^{\frac{d}{2}} \pi\left(1 - x^2\right) dx = \pi\left(d - \frac{d^3}{12}\right)$. For this to be $\frac{1}{3}$ of the total volume of the sphere, we have $\pi\left(d - \frac{d^3}{12}\right) = \frac{4}{9}\pi$. This simplifies to $d^3 - 12d + \frac{16}{3} = 0$, one root of which must be the separation between the planes.

8. This is a special case of Newton's theorem of symmetric polynomials, which is proved later in the chapter.

9. We have $a^3 + b^3 + c^3 = 3abc$, by the difference of three cubes. Also, $a^2 + b^2 + c^2 = -2ab - 2bc - 2ca$, since $(a + b + c)^2 = 0$. We can express $a^5 + b^5 + c^5$ as $(a + b + c)\left(a^4 + b^4 + c^4\right) - \sum_{\text{sym}} a^4 b$. The first term is zero, so $a^5 + b^5 + c^5 = -\sum_{\text{sym}} a^4 b = a^3 bc + b^3 ca + c^3 ab - \left(a^3 + b^3 + c^3\right)(ab + bc + ca)$. After factorisation, this is then equal to $abc\left(a^2 + b^2 + c^2\right) - 3abc(ab + bc + ca)$, or $\frac{5}{2}abc\left(a^2 + b^2 + c^2\right)$. Except in the trivial case where all variables are zero, $a^2 + b^2 + c^2 > 0$, so $a^3 + b^3 + c^3$ and $a^5 + b^5 + c^5$ are positive real multiples of abc, thus both have the same sign.

Sequences

In this chapter, we are concerned with infinite sequences of either integers or, more generally, real numbers. Although it is no longer one of the main phyla of questions in the IMO (combinatorics, algebra, geometry and number theory), sequences do feature very prominently.

Generating functions

To manipulate sequences, it is useful to be able to represent them algebraically as a power series known as a *generating function*.

- A sequence $\{A_0, A_1, A_2, A_3, \ldots\}$ has the *ordinary generating function* $a_o(x) = A_0 + A_1 x + A_2 x^2 + A_3 x^3 + \ldots$. **[Definition of OGF]**

We can add and multiply ordinary generating functions, which correspond to addition and convolution of their respective sequences.

1. Suppose we have two sequences, $\{A_n\}$ and $\{B_n\}$, with ordinary generating functions $a_o(x)$ and $b_o(x)$, respectively. Let $\{C_n\}$ and $\{D_n\}$ have ordinary generating functions $a_o(x) + b_o(x)$ and $a_o(x)\, b_o(x)$, respectively. Show that $C_n = A_n + B_n$ and $D_n = A_0 B_n + A_1 B_{n-1} + \ldots + A_n B_0$. **[Addition and convolution]**

2. Find a closed form for $1 + x + x^2 + \ldots$, the ordinary generating function of $\{1, 1, 1, \ldots\}$. Hence find an ordinary generating function for the sequence of natural numbers, $\{1, 2, 3, \ldots\}$, and the triangular numbers, $\{1, 3, 6, 10, \ldots\}$.

3. Suppose $\{A_0, A_1, A_2, A_3, \ldots\}$ has ordinary generating function $a_o(x)$. What sequence has ordinary generating function $\frac{d}{dx} a_o(x)$?

4. Hence find the sequence with ordinary generating function $\ln(1 - x)$.

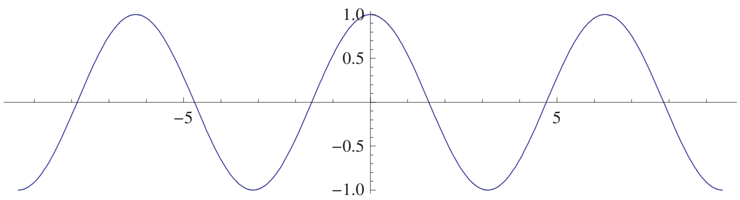
When differentiating $A_0 + A_1 x + A_2 x^2 + A_3 x^3 + \ldots$, we obtain the sequence $A_1 + 2 A_2 x + 3 A_3 x^2 + 4 A_4 x^3 + \ldots$. As an operation on sequences, this is a left shift followed by a (somewhat annoying) multiplication of each term by a different scalar. We can remove this inelegance by defining a more complicated *exponential generating function*, or *EGF*.

- A sequence $\{B_0, B_1, B_2, B_3, \ldots\}$ has the *exponential generating function* $b_e(x) = \frac{B_0}{0!} + \frac{B_1}{1!} x + \frac{B_2}{2!} x^2 + \frac{B_3}{3!} x^3 + \ldots$. **[Definition of EGF]**

If we differentiate it, we obtain the exponential generating function of the sequence $\{B_1, B_2, B_3, \ldots\}$, which is simply the original sequence shifted to the left. The sequence $\{1, 1, 1, \ldots\}$ is invariant when shifted to the left, so its exponential generating function (namely $f(x) = e^x = 1 + x + \frac{1}{2} x^2 + \frac{1}{6} x^3 + \frac{1}{24} x^4 + \ldots$) is invariant under differentiation, and is thus a solution to the differential equation $f(x) = f'(x)$.

5. Show that the exponential generating function of $\{1, 0, -1, 0, 1, 0, -1, 0, \ldots\}$ is a solution of the differential equation $f''(x) + f(x) = 0$.

This type of differential equation is encountered in the mechanics of mass-spring systems. This particular solution is the function $f(x) = \cos x$; the general solution is $f(x) = A \cos x + B \sin x$, corresponding to the sequence $\{A, B, -A, -B, A, B, -A, -B, \ldots\}$. For this reason, a stretched spring with a suspended mass will oscillate periodically in simple harmonic motion.



Some basic sequences and their exponential generating functions are given below.

Sequence	EGF
$\{1, 1, 1, 1, \ldots\}$	e^x
$\{1, -1, 1, -1, 1, \ldots\}$	e^{-x}
$\{1, 0, 1, 0, 1, \ldots\}$	$\cosh(x)$
$\{0, 1, 0, 1, 0, \ldots\}$	$\sinh(x)$
$\{1, 0, -1, 0, 1, 0, -1, 0, \ldots\}$	$\cos(x)$
$\{0, 1, 0, -1, 0, 1, 0, -1, \ldots\}$	$\sin(x)$
$\{1, 2, 4, 8, 16, \ldots\}$	e^{2x}
$\{1, 2, 3, 4, 5, \ldots\}$	$e^x x$

6. Find a sequence $\{F_0, F_1, F_2, \ldots\}$ whose exponential generating function satisfies the differential equation $f''(x) = f'(x) + f(x)$.

In general, the solution to any homogeneous linear differential equation is the exponential generating function of a sequence defined by a *linear recurrence relation*.

Linear recurrence relations

Suppose we have a sequence defined by the *linear recurrence relation* $A_{n+k} = \alpha_0 A_n + \alpha_1 A_{n+1} + \alpha_2 A_{n+2} + \cdots + \alpha_{k-1} A_{n+k-1}$. This is linear and homogeneous, which means that for any two sequences $\{A_n\}$ and $\{B_n\}$ satisfying the equation, so does the sequence $\{\lambda A_n + \mu B_n\}$. It is also determined entirely by the values of $\{A_0, A_1, \ldots, A_{k-1}\}$, so there are k degrees of freedom in the solution set.

If a sequence $\{B_n = x^n\}$ satisfies the recurrence relation, then we have $x^k = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \ldots + \alpha_{k-1} x^{k-1}$. By the fundamental theorem of algebra, we can rearrange and factorise this into k linear terms. Assuming that this polynomial has l distinct roots, $\beta_1, \beta_2, \ldots, \beta_l$, we obtain the general solution $\{A_n = P_1(n) \beta_1{}^n + P_2(n) \beta_2{}^n + \ldots + P_l(n) \beta_l{}^n\}$. P_i is a polynomial of degree $m_i - 1$, where m_i is the multiplicity of the root β_i. Specifically, when all the roots are distinct, all values of $P_i(n)$ are constants. It is possible to verify that each term satisfies the linear recurrence relation, so the general solution is valid. Also, it has k degrees of freedom, so there can be no other solutions.

Probably the simplest non-trivial linear recurrence relation is the *Fibonacci sequence* $\{F_n\}$ with $F_0 = 0$, $F_1 = 1$ and $F_{n+2} = F_{n+1} + F_n$. It was discovered by Leonardo of Pisa when contemplating a problem about the exponential growth of a rabbit population. It has the closed-form expression $F_n = \frac{\phi^n - \psi^n}{\phi - \psi}$, where ϕ and ψ are the positive and negative roots, respectively, of the equation $x^2 = x + 1$.

- $F_n = \frac{\phi^n - \psi^n}{\sqrt{5}}$, where $\phi = \frac{1+\sqrt{5}}{2}$ and $\psi = \frac{1-\sqrt{5}}{2}$. **[Binet's formula]**

This enables one to compute a closed form for the exponential generating function of $\{0, 1, 1, 2, 3, 5, 8, \ldots\}$, and thus find a closed-form solution to the differential equation $f''(x) = f'(x) + f(x)$.

7. Find a closed form for the exponential generating function of $\{0, 1, 1, 2, 3, 5, 8, \ldots\}$.

8. Show that, for all $a, b \in \mathbb{Z}^+$, we have $F_a \mid F_{ab}$.

9. Prove that $F_n{}^2 = F_{n+1} F_{n-1} - (-1)^n$. **[Cassini's identity]**

10. Show that the Fibonacci sequence has the ordinary generating function $f_o(x) = \frac{x}{1-x-x^2}$, and thus find a rational approximation to 0.01010203050813213455.

Every sequence generated by a linear recurrence relation has an ordinary generating function expressible as a ratio between two polynomials.

Consider the sequence $\{A_i \pmod p\}$. As the sequence is generated by the previous k terms, each of which can be one of n values, the sequence must eventually cycle with period $P \le n^k$. The recurrence relation is deterministic in both directions, so it must be completely periodic. Indeed, this bound can be slightly improved upon, as k consecutive zeros would result in a sequence that is identically zero. Hence, $P \le n^k - 1$. Sequences where equality holds are known as *maximal*. For example, the non-zero sequence with recurrence relation $X_{n+3} \equiv X_{n+1} + X_n \pmod 2$ has a period of $2^3 - 1$ and is displayed below (where red and blue discs indicate 0 and 1, respectively).

In a maximal sequence, every non-empty string of k digits appears precisely once in each cycle. Together with the property that the sequence has a long period, linear recurrence sequences are used as simple pseudo-random number generators. A refined algorithm, known as the *Mersenne twister*, has a period of $2^{19937} - 1$ and is almost indistinguishable from random data.

11. Let n be an integer greater than 1. In a circular arrangement of n lamps $L_0, L_1, \ldots, L_{n-1}$, each one of which can be either ON or OFF, we begin with the configuration where all lamps are ON. We carry out a sequence of steps, $\text{Step}_0, \text{Step}_1, \ldots$. If L_{j-1} is ON, then Step_j changes the status of L_j (from ON to OFF or *vice-versa*), but does not change the status of any other lamps; if L_{j-1} is OFF, then Step_j does not change anything at all. Show that:

- There is a positive integer $M(n)$ such that all lamps are ON again after $M(n)$ steps.
- If n has the form 2^k, then all lamps are on after $n^2 - 1$ steps.
- If n has the form $2^k + 1$, then all lamps are on after $n^2 - n + 1$ steps. **[IMO 1993, Question 6]**

Positional systems

When we write numbers in decimal notation, we let '$d_k\, d_{k-1} \ldots d_2\, d_1\, d_0$' represent the non-negative integer $d_k\, 10^k + d_{k-1}\, 10^{k-1} + \ldots + 10\, d_1 + d_0$, where each digit $d_i \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$. Note that this is unique; an integer cannot have two valid representations. This can be generalised to other bases (or *radices*) such as binary and ternary. For example, we can uniquely express any non-negative integer in binary as a sum of distinct powers of two.

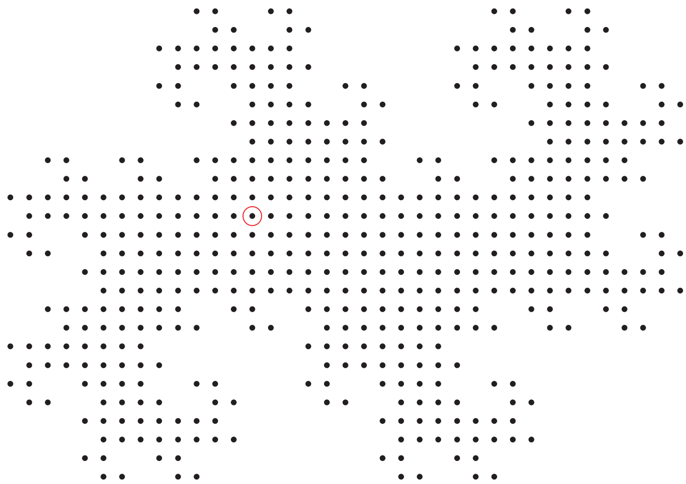
- The n-ary representation '$d_k\, d_{k-1} \ldots d_2\, d_1\, d_0$', where each digit $d_i \in \{0, 1, \ldots, n-1\}$, uniquely represents the integer $\sum_{i=0}^{k} n^i\, d_i$.

A certain type of functional equation, known as a *binary functional equation* or *Monk equation*, is best approached by expressing the numbers in binary.

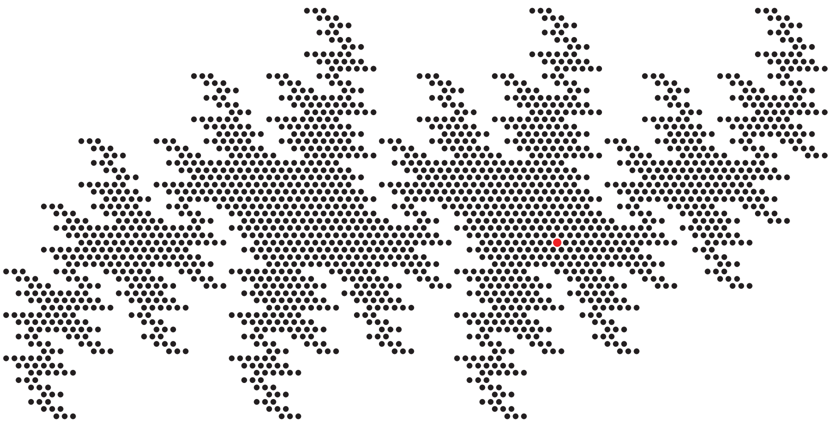
12. A function $f : \mathbb{Z}_{\geq 0} \to \mathbb{R}$ is defined using the initial term $f(0) = 0$ and the recurrence relations $f(2n) = \frac{1}{2} f(n)$ and $f(2n+1) = 1 + f(2n)$. How many integers x exist such that $x = 2^{20} f(x)$?

There are some more interesting variants of positional systems. *Balanced ternary* has $n = 3$ and $d_i \in \{-1, 0, 1\}$, as opposed to the $\{0, 1, 2\}$ of ordinary base-3. There exist unique expressions in balanced ternary for every integer, as opposed to merely non-negative integers. By contrast with the modern binary computers we use today, there was an early computer (Setun), which operated in balanced ternary.

The system with $n = i - 1$ and $d_i \in \{0, 1\}$ is even better, as it can represent any Gaussian integer. The Gaussian integers with representations using at most $k + 1$ digits are the 2^{k+1} points on a space-filling fractal known as the *twindragon curve*. The example for $k = 8$ is shown below, with the origin encircled red.



13. Prove that every Eisenstein integer has a unique representation in the positional system with base $n = \omega - 1$ and digits $\{0, 1, 2\}$.



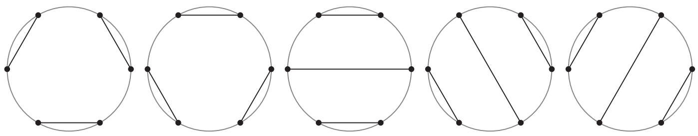
A particularly interesting positional notation is the *Zeckendorf representation*. There is a unique way to express any non-negative integer as the sum of Fibonacci numbers, no two of which are consecutive. For example, $100 = 89 + 8 + 3$.

14. Determine whether there exists a function $f : \mathbb{N} \to \mathbb{N}$ such that $f(f(n)) = f(n) + n$ for all $n \in \mathbb{N}$.

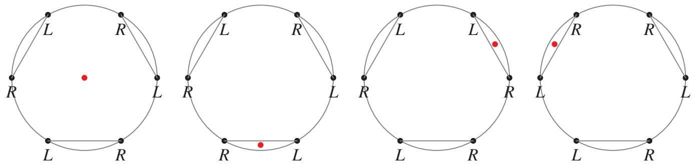
15. A set A of integers is called *sum-full* if $A \subseteq A + A$, *i.e.* each element $a \in A$ is the sum of some pair of (not necessarily distinct) elements $b, c \in A$. A set A of integers is said to be *zero-sum-free* if 0 is the only integer that cannot be expressed as the sum of the elements of a finite non-empty subset of A. Does there exist a set of integers which is both sum-full and zero-sum-free? **[EGMO 2012, Question 4, Dan Schwarz]**

Catalan sequence

We define the nth *Catalan number*, C_n, to be the number of ways of pairing $2\,n$ points on the circumference of a circle with n non-intersecting chords. For example, we have $C_3 = 5$:



Each of these pairings divides the interior of the circle into $n + 1$ regions. We can place an 'observer' in any region. Now label each point with a L or R depending on whether it is connected to a point to the left or right of itself when viewed by the observer. For example, the first pairing can lead to any of the following labellings, depending on where the observer is positioned:



As every labelling of $2\,n$ points with n Ls and n Rs can result in a valid pairing and position of observer, and every pairing and position of observer results in a unique labelling, we have a bijection between the two. As there are C_n pairings and $n + 1$ regions in which to place the observer, there are $(n + 1)\,C_n$ different labellings. However, there are clearly $\binom{2\,n}{n}$ different labellings, so we have $C_n = \frac{1}{n+1}\binom{2\,n}{n}$.

16. Show that the number of valid balanced strings S of $2\,n$ parentheses is given by C_n. (There must be n left parentheses, n right parentheses, and for all $k \le 2\,n$, the number of right parentheses in the first k symbols of S cannot exceed the number of left parentheses.)

17. An insect walks on the integer lattice $\mathbb{Z}^2$, beginning at $(0, 0)$. After $2\,n$ steps, it reaches (n, n).
- How many different paths could the insect have taken?
- Assuming that, for all points (x, y) on the path, $x \ge y$, how many different paths are possible?

18. Prove that $C_{n+1} = C_0\,C_n + C_1\,C_{n-1} + C_2\,C_{n-2} + \ldots + C_n\,C_0$. **[Segner's recurrence relation]**

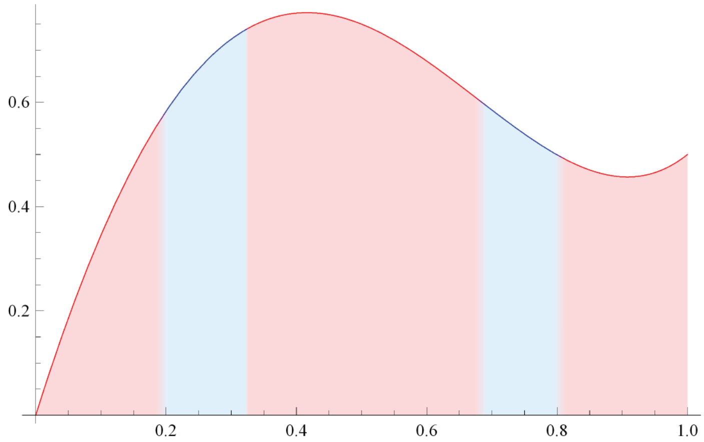
19. Let $c_o(x) = C_0 + C_1\,x + C_2\,x^2 + \ldots$ be the ordinary generating function for the Catalan numbers. Show that $x\,c_o(x)^2 - c_o(x) + 1 = 0$, and thus that $c_o(x) = \frac{1-\sqrt{1-4\,x}}{2\,x}$.

L-systems

Suppose we have a string S consisting entirely of the symbols A and B. We define $f(S)$ by simultaneously replacing every A with $A\,B$ and every B with $B\,A$. For example, starting with a single A, we have $f(A) = A\,B$, $f(f(A)) = A\,B\,B\,A$, $f^3(A) = A\,B\,B\,A\,B\,A\,A\,B$, and so on. The limit of this is an infinite sequence $A\,B\,B\,A\,B\,A\,A\,B\,B\,A\,A\,B\,A\,B\,B\,A\ldots$, known as the *Thue-Morse sequence*. This process of repeatedly applying *substitution rules* to every symbol in a string is known as an *L-system*.

20. A sequence $\{x_i\}$ is defined by $x_1 = 1$ and the recurrences $x_{2\,k} = -x_k$ and $x_{2\,k-1} = (-1)^{k+1}\,x_k$ (for all $k \in \mathbb{Z}^+$). Prove that, for all $n \ge 1$, we have $x_1 + x_2 + \ldots + x_n \ge 0$. **[IMO 2010 shortlist, Question A4]**

21. Prove that we can divide the interval [0, 1] into finitely many intervals, alternately coloured red and blue, such that $\int_{\text{red}} P(x)\,dx = \int_{\text{blue}} P(x)\,dx$ for all polynomials P of degree 2013.



One of the more interesting L-systems is the *golden string*, generated by iterating the system $S \to g(S)$ with substitution rules $X \to X\,Y$ and $Y \to X$ to the initial string X. The first few iterations are $X \to X\,Y \to X\,Y\,X \to X\,Y\,X\,X\,Y \to X\,Y\,X\,X\,Y\,X\,Y\,X \to \ldots$.

22. Show that the number of symbols in $g^n(X)$ is given by F_{n+2}, i.e. the $(n+2)$th term of the Fibonacci sequence.

23. Prove that $g^n(X)$ can be expressed as the concatenation of two palindromic substrings for all $n \in \mathbb{Z}^+$. Moreover, find the lengths of the palindromic substrings for all $n \geq 3$. (A string is described as *palindromic* if it reads the same in both directions. For example, $L\,E\,V\,E\,L$ and $R\,A\,C\,E\,C\,A\,R$ are palindromic strings. The empty string is also considered to be palindromic.)

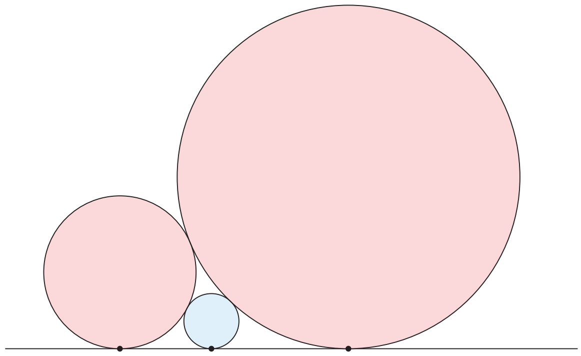
24. Hence prove that, for all $n \in \mathbb{Z}^+$, removing the last two symbols of $g^n(X)$ results in a palindromic string.

The golden string may be regarded as a one-dimensional analogue of the Penrose tiling. Indeed, if you know where to look, you will be able to find the golden string recurring throughout any Penrose tiling. Additionally, the last digit of the Zeckendorf representation of n (for all non-negative integers n) forms the golden string.

Farey sequences

Suppose we have two tangent circles resting on the real line. Circle Γ_1 is positioned at $\frac{p_1}{q_1}$ and has a diameter of $\frac{1}{{q_1}^2}$. Similarly, circle Γ_2 is positioned at $\frac{p_2}{q_2}$ and has a diameter of $\frac{1}{{q_2}^2}$.

25. Prove that $p_1\,q_2 + 1 = p_2\,q_1$.



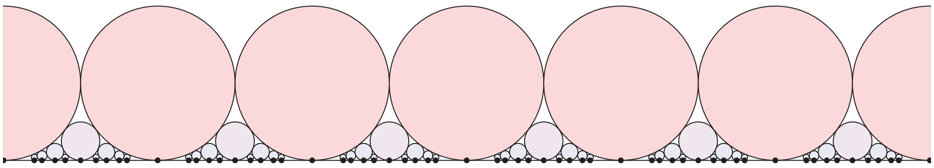
We now position a smaller circle, Γ_3, externally tangent to the two larger circles and the real line.

26. Show that Γ_3 is tangent to the line at $\frac{p_3}{q_3} = \frac{p_1+p_2}{q_1+q_2}$.

This gives a geometrical relationship between the radii of the circles.

- Suppose circles Γ_1 and Γ_2 are tangent to each other, and one of the outer common tangents is l. Let Γ_3 be a third circle tangent internally to Γ_1, Γ_2 and l. Then $\frac{1}{\sqrt{r_3}} = \frac{1}{\sqrt{r_1}} + \frac{1}{\sqrt{r_2}}$, where r_i is the radius of Γ_i. **[Sangaku problem]**

If we begin with a circle of unit diameter for each positive integer, and iterate this process infinitely, we create a pattern known as the *Ford circles*. This has an elegant symmetry associated with modular forms and certain tilings of the hyperbolic plane.



Taking only the circles where $q \le n$ and confining ourselves to the interval [0, 1], we generate a *Farey sequence*, F_n. For example, $F_5 = \left\{0, \frac{1}{5}, \frac{1}{4}, \frac{1}{3}, \frac{2}{5}, \frac{1}{2}, \frac{3}{5}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, 1\right\}$. Each term is the *mediant* of the two neighbouring terms. For instance, $\frac{1}{3}$ is situated between $\frac{1}{4}$ and $\frac{2}{5}$, and $\frac{1}{3} = \frac{1+2}{4+5}$. This enables a Farey sequence to be extrapolated in both directions from two adjacent terms. It turns out that Rademacher's proof of the formula for the partition numbers involves Ford circles and Farey sequences.

Returning to the Sangaku problem, there is a generalisation known as *Descartes' theorem*, named after the philosopher who said *'cogito ergo sum'* and invented Cartesian coordinates. If we have four circles, which are pairwise externally tangent, then there is a quadratic relationship between the reciprocals of the radii.

- $\left(\frac{1}{r_1} + \frac{1}{r_2} + \frac{1}{r_3} + \frac{1}{r_4}\right)^2 = 2\left(\frac{1}{r_1^2} + \frac{1}{r_2^2} + \frac{1}{r_3^2} + \frac{1}{r_4^2}\right)$. **[Descartes' theorem]**

$\frac{1}{r_4}$ is thus the solution to a quadratic equation. If we choose the other root (and multiply by -1 to make it positive), we obtain the reciprocal of the radius of the *circumscribed* circle, rather than the inscribed circle.

Egyptian fractions

The Fibonacci sequence, Catalan sequence and powers of two grow reasonably quickly, namely *exponentially*. By comparison, *Sylvester's sequence* grows even more quickly (*doubly-exponentially*), with the first few terms being {2, 3, 7, 43, 1807, 3 263 443, …}. This is defined with the initial term $s_0 = 2$ together with the recurrence relation $s_n = 1 + s_0\, s_1\, s_2 \dots s_{n-1}$.

27. Prove that the terms in Sylvester's sequence are pairwise coprime.

This is a direct proof that there are infinitely many primes, as no two terms in Sylvester's sequence share a prime factor. Euclid's proof of the infinitude of primes is similar, but with a proof by contradiction instead.

28. Show that $\frac{1}{s_0} + \frac{1}{s_1} + \frac{1}{s_2} + \ldots = 1$.

As decimal expansions, continued fractions and ratios had not been invented, the ancient Egyptians expressed fractions as the sum of reciprocals of distinct positive integers. It is a remarkable fact that *Egyptian fractions* can represent any positive rational number. One algorithm which proves the possibility of this is the following:

- Initially express $\frac{a}{b}$ as $\frac{1}{b} + \frac{1}{b} + \frac{1}{b} + \ldots + \frac{1}{b}$.
- If there are multiple copies of $\frac{1}{n}$, replace one of them with $\frac{1}{n+1} + \frac{1}{n(n+1)}$.
- Repeat the previous step until all unit fractions are distinct.

As arbitrarily large fractions can be generated in this manner, we have a proof that the series $\zeta(1) = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \ldots$ diverges to infinity. It does so rather slowly, with the first n terms tending to $\ln(n) + \gamma$, where γ is the *Euler-Mascheroni constant*. Also, we have yet another proof of the infinitude of primes, because we can factorise $\zeta(1)$ as $\left(1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \ldots\right)\left(1 + \frac{1}{3} + \frac{1}{9} + \ldots\right)\left(1 + \frac{1}{5} + \frac{1}{25} + \ldots\right)\ldots$. Each term is the sum of a geometric series, resulting in the product expansion $\zeta(1) = \frac{2}{1} \cdot \frac{3}{2} \cdot \frac{5}{4} \cdot \frac{7}{6} \cdot \ldots \cdot \frac{p}{p-1} \cdot \ldots$. As each of the terms is finite, but the product is infinite, there must be infinitely many primes.

Another algorithm for generating Egyptian fraction expansions is the *greedy algorithm*, where we choose the largest unused unit fraction less than or equal to the remainder. For example, if we wanted to express $\frac{4}{5}$ as an Egyptian fraction, we would first subtract $\frac{1}{2}$, resulting in $\frac{3}{10}$, followed by $\frac{1}{4}$, resulting in $\frac{1}{20}$, and finally $\frac{1}{20}$, resulting in the expansion $\frac{4}{5} = \frac{1}{2} + \frac{1}{4} + \frac{1}{20}$.

If, instead, we restrict ourselves to choosing the largest *odd* unit fraction at each point, the process may continue forever. For example, applying this algorithm to $\frac{1}{2}$ generates the remainder of Sylvester's sequence. This is similarly the case for all fractions with even denominators. It is an open problem as to whether the process necessarily terminates for all fractions with odd denominators.

Fermat numbers

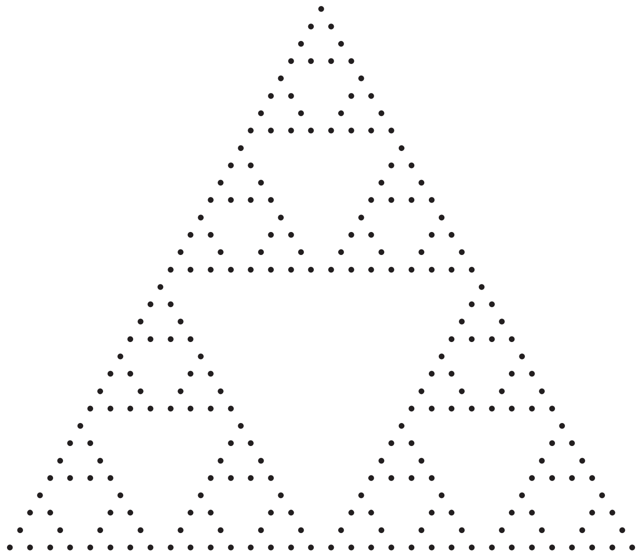
Another doubly-exponential sequence is the sequence of *Fermat numbers* of the form $2^{2^n} + 1$, namely $\{3, 5, 17, 257, 65\,537, \ldots\}$. You may notice that the first five Fermat numbers are prime, known as *Fermat primes*. Fermat conjectured that all Fermat numbers are prime, with the first counter-example found by Euler.

- $2^{2^5} + 1 = 4\,294\,967\,297 = 641 \cdot 6\,700\,417$. **[Euler's factorisation]**

It is unknown whether there are infinitely many Fermat primes. Only the first five are known to be prime; the next 28 have been proved to be composite!

29. Prove that if $2^k + 1$ is prime, then it is a Fermat prime.

A regular n-gon of unit side length is *constructible* if and only if it can be constructed using a compass and straightedge. Equivalently, this means that the Cartesian coordinates of each of the vertices can be expressed as a finite combination of integers together with the operations $\{+, -, \times, \div, \sqrt{\ }\}$. Gauss proved that a regular n-gon is constructible if and only if $n = 2^a\, p_1\, p_2 \ldots p_k$, where each p_i is a distinct Fermat prime. The only known odd values for n are thus products of distinct Fermat primes. Expressed in binary, these form the first 32 rows of Pascal's triangle modulo 2, namely {1, 11, 101, 1111, 10 001, …}. When each 1 is replaced with a dot and 0 is replaced with an empty space, this is the fifth-order approximation to the Sierpinski triangle.



Certain geometrical constructions, such as trisecting the angle, are possible with paper folding but not with Euclidean constructions. Using origami, we can apply the operations $\left\{+, -, \times, \div, \sqrt{\ }, \sqrt[3]{\ }\right\}$ to Cartesian coordinates, and thus reach points unattainable with compass and straightedge alone. In this new system, a regular n-gon is constructible if and only if $n = 2^a\, 3^b\, q_1\, q_2 \ldots q_k$, where each q_i is a distinct *Pierpont prime* (prime expressible in the form $2^n\, 3^m + 1$). Fermat primes are, by definition, a subset of Pierpont primes.

Solutions

1. We have $a_o(x) + b_o(x) = A_0 + B_0 + A_1\,x + B_1\,x + \cdots$. Similarly, $a_o(x)\,b_o(x) = \left(A_0 + A_1\,x + A_2\,x^2 + \ldots\right)\left(B_0 + B_1\,x + B_2\,x^2 + \ldots\right)$. Expanding the brackets results in $A_0\,B_0 + (A_0\,B_1 + A_1\,B_0)\,x + (A_0\,B_2 + A_1\,B_1 + A_2\,B_0)\,x^2 + \ldots$.

2. The geometric series $1 + x + x^2 + \ldots$ is given by $\frac{1}{1-x}$. As $\{1, 2, 3, \ldots\}$ is the convolution of $\{1, 1, 1, \ldots\}$ with itself, its ordinary generating function is $\frac{1}{(1-x)^2}$. The convolution of this with $\{1, 1, 1, \ldots\}$ gives the triangular numbers with ordinary generating function $\frac{1}{(1-x)^3}$.

3. Differentiating each term, we obtain $\frac{d}{dx}\,a_o(x) = A_1 + 2\,A_2\,x + 3\,A_3\,x^2 + 4\,A_4\,x^3 + \ldots$. This is the ordinary generating function of $\{A_1, 2\,A_2, 3\,A_3, \ldots\}$.

4. We already know that $\frac{1}{1-x} = 1 + x + x^2 + x^3 + \ldots$. This can be integrated to yield $c + x + \frac{1}{2}\,x^2 + \frac{1}{3}\,x^3 + \ldots$. As $\ln(1-x)$ evaluates to zero when x is zero, the constant term is zero. Hence, $\ln(1-x)$ is the ordinary generating function of $\left\{0, 1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots\right\}$.

5. Shift the sequence two places to the left and add it to the original sequence. This results in the zero sequence, the EGF of which is the zero function.

6. This has the recurrence relation $F_{n+2} = F_{n+1} + F_n$. A possible solution is the Fibonacci sequence, $\{0, 1, 1, 2, 3, 5, 8, \ldots\}$.

7. $f_e(x) = \frac{1}{\sqrt{5}}\left(1 + \phi\,x + \frac{(\phi\,x)^2}{2!} + \frac{(\phi\,x)^3}{3!} + \ldots\right) - \frac{1}{\sqrt{5}}\left(1 + \psi\,x + \frac{(\psi\,x)^2}{2!} + \frac{(\psi\,x)^3}{3!} + \ldots\right)$. This is equal to $\frac{1}{\sqrt{5}}\left(e^{\phi\,x} - e^{\psi\,x}\right)$.

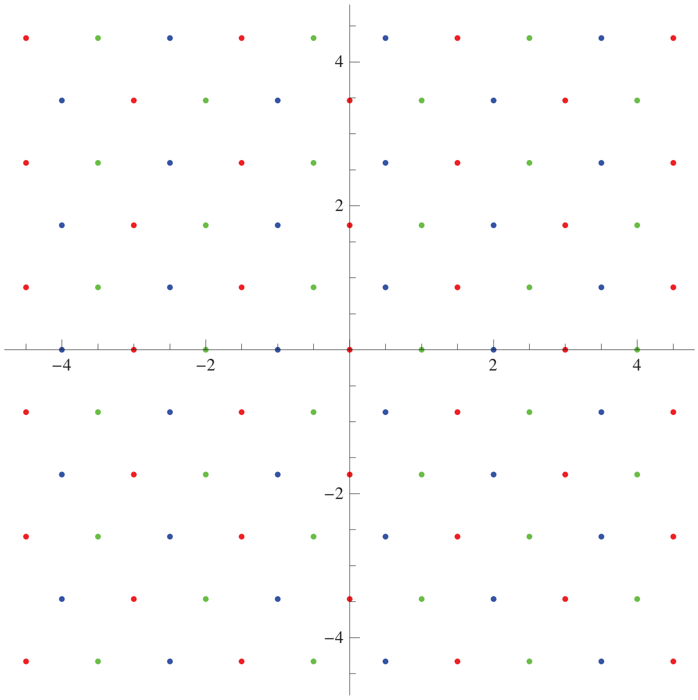
8. The ratio $\frac{F_{a\,b}}{F_a} = \frac{\phi^{a\,b} - \psi^{a\,b}}{\phi^a - \psi^a} = \phi^{a(b-1)} + \phi^{a(b-2)}\,\psi^a + \ldots + \phi^a\,\psi^{a(b-2)} + \psi^{a(b-1)}$ is a symmetric polynomial in ϕ and ψ, so is expressible as a polynomial with integer coefficients in $\phi\,\psi$ and $\phi + \psi$. As the elementary symmetric polynomials are themselves integers, so is the ratio $\frac{F_{a\,b}}{F_a}$.

9. $\frac{(\phi^n - \psi^n)^2}{(\phi-\psi)^2} - \frac{\left(\phi^{n-1} - \psi^{n-1}\right)\left(\phi^{n+1} - \psi^{n+1}\right)}{(\phi-\psi)^2} = \frac{\phi^{2\,n} + \psi^{2\,n} - 2\,(\phi\,\psi)^n}{5} - \frac{\phi^{2\,n} + \psi^{2\,n} + 3\,(\phi\,\psi)^n}{5} = -(\phi\,\psi)^n = -(-1)^n$.

10. $f_o(x) = x + x^2 + 2\,x^3 + 3\,x^4 + 5\,x^5 + \cdots$. It is straightforward to verify from the recurrence relation that $f_0(x) = x + x\,f_o(x) + x^2\,f_o(x)$. Rearranging, we obtain the closed form $f_o(x) = \frac{x}{1-x-x^2}$. Setting $x = 0.01$ gives us the rational approximation $\frac{0.01}{1-0.01-0.0001} = \frac{0.01}{0.9899} = \frac{100}{9899}$.

11. This system is deterministic and finite, so must necessarily eventually become cyclic. As it is reversible as well, it must be completely cyclic; this solves the first part of the question. We can represent the state of the system using the ordinary generating function $L_{n-1} + L_{n-2}\,x + L_{n-3}\,x^2 + L_{n-4}\,x^3 + \ldots + L_0\,x^{n-1}$, where 0 and 1 correspond to OFF and ON, respectively. This is a polynomial $\mathbb{Z}_2[x]$, as each coefficient can be either 0 or 1 and we consider addition modulo 2. If we follow each step with a rotation to the left (such that L_i moves into the position of L_{i-1}), then we only need to alter the state of L_0 depending on L_{n-1}. Refer to this composite operation as Step*. The lamp alteration is equivalent to the operation $P(x) \to P(x) - 1 + x^{n-1}$ (if applicable, or the identity function otherwise), and the rotation is equivalent to $P(x) \to x\,P(x)$. So, Step* performs the polynomial operation $P(x) \to x\,P(x) \left(\bmod x^n + x^{n-1} + 1\right)$; this modulus is the characteristic polynomial associated with this linear feedback shift register. If we apply m steps in succession, then we obtain the polynomial $x^m P(x) \left(\bmod x^n + x^{n-1} + 1\right)$. If m is the period of the system, then $x^m P(x) \equiv P(x)$ and thus $x^m \equiv 1$. For the second part of the question, we get $x^{n^2} \equiv (x^n)^n \equiv \left(x^{n-1} + 1\right)^n$. As $n = 2^k$, all binomial coefficients apart from $\binom{2^k}{0} = 1$ and $\binom{2^k}{2^k} = 1$ are even (consider Pascal's triangle modulo 2, which approximates the Sierpinski triangle), thus equal to zero in the field $\mathbb{Z}_2$. So, $\left(x^{n-1} + 1\right)^n$ expands to give $x^{n^2-n} + 1$. With the identities $x^{n^2} \equiv x^{n^2-n} + 1$ and $x^{n^2} \equiv x^{n^2-n} + x^{n^2-1}$, we obtain $x^{n^2-1} \equiv 1$, and thus $(\text{Step}^*)^{n^2-1}$ is the identity function. For the third part of the question, we get $x^{n^2-1} \equiv \left(x^{n+1}\right)^{n-1} \equiv \left(x^{n+1}\right)^{2^k} \equiv (x^n + x)^{2^k} \equiv x^{n(n-1)} + x^{n-1} \equiv x^{n^2-n} + x^{n-1}$. Hence, $x^{n^2} \equiv x^{n^2-1} + x^{n^2-n} \equiv x^{n-1} + 2\,x^{n^2-n} \equiv x^{n-1}$. Dividing throughout by x^{n-1} results in $x^{n^2-n+1} \equiv 1$, so $(\text{Step}^*)^{n^2-n+1}$ is the identity function.

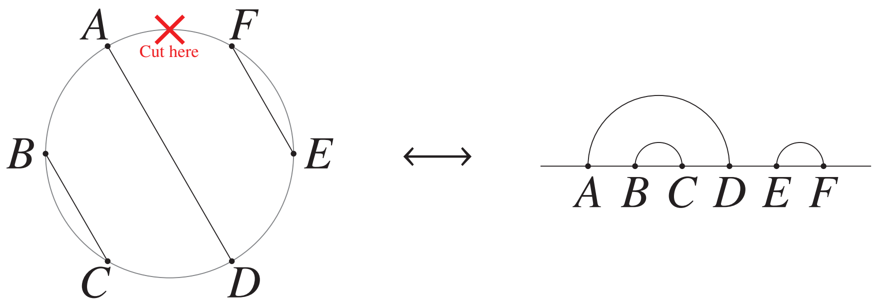
12. The function 'reflects' the binary representation. For example, it maps 10 011 101 to 1.0111001. Multiplying by 2^{20} means that it is subsequently shifted 20 places to the left. The fixed points of this function are the palindromes of length 21, of which there are 2^{11}.



13. Colour the Eisenstein integers red, green and blue, as in the diagram above. If we multiply all Eisenstein integers by $\omega - 1$, we obtain the red Eisenstein integers (a copy of the Eisenstein integers rotated by $\frac{5}{6}\pi$ and scaled by $\sqrt{3}$). The green and blue Eisenstein integers can be obtained by translating the red Eisenstein integers by 1 and 2, respectively. Hence, if no Eisenstein integer has two representations using k digits, then no Eisenstein integer has two representations using $k + 1$ digits. By induction, all representations are unique. However, we have yet to show that all Eisenstein integers can be obtained in this manner. If all Eisenstein integers within the hexagon with vertices $\{1, -\omega^2, \omega, -1, \omega^2, -\omega\}$ can be represented (which can be shown using trial and error), then we can represent all Eisenstein integers within the hexagon with vertices $\{1\,(\omega - 1), -\omega^2(\omega - 1), \omega(\omega - 1), -1\,(\omega - 1), \omega^2(\omega - 1), -\omega(\omega - 1)\}$ by multiplying those Eisenstein integers by $\omega - 1$ and translating by 1 and 2 to fill in the gaps. By induction, we can represent all Eisenstein integers within an arbitrarily large hexagon centred about the origin, and therefore all Eisenstein integers.

14. Yes, consider the function which shifts the Zeckendorf representation of a positive integer one position to the left. Due to the Fibonacci recurrence relation, $f(f(n)) = f(n) + n$.

15. Yes, for example $\{1, -2, 3, -5, 8, -13, 21, \ldots\}$. This is sum-full as each term is the sum of the next two terms. To prove that it is zero-sum-free, consider all numbers expressible as the sum of the first k terms of the sequence. We can prove from a trivial base case and simple inductive argument that this is $\{1 - F_{k+1}, 2 - F_{k+1}, \ldots, -2, -1, 1, 2, \ldots, F_{k+2} - 2, F_{k+2} - 1\}$ for odd k, and the negation thereof for even k. The limiting set is the set of nonzero integers.



16. We can biject between these strings and non-intersecting pairings of points on a circle by cutting the circle at a given point and 'unfolding' it, as demonstrated above.

17. The insect can only move right or up at each step, as otherwise it would take too long to reach (n, n). There must be n moves to the right and n moves up, so there are $\binom{2n}{n}$ possible paths in the first part of the problem. For the second part of the problem, we represent a horizontal move with a left parenthesis and a vertical move with a right parenthesis. This reduces the problem to the previous question, so there are $C_n = \frac{1}{n+1}\binom{2n}{n}$ paths with this constraint.

18. Consider $2n + 2$ points on the circumference of a circle, and label one vertex A. Choose another vertex B such that the number of vertices on each arc AB is even, and join A and B with a chord. Let the number of points right of the chord be $2k$; the number of points left of the chord must be $2(n - k)$. There are C_k non-intersecting pairings of the vertices to the left of the chord, and C_{n-k} non-intersecting pairings of vertices to the right of the chord, giving a total of $C_k C_{n-k}$ pairings. Repeating this for each location of B gives $C_0 C_n + C_1 C_{n-1} + \ldots + C_n C_0$ possible pairings. This must be equal to C_{n+1}, by definition.

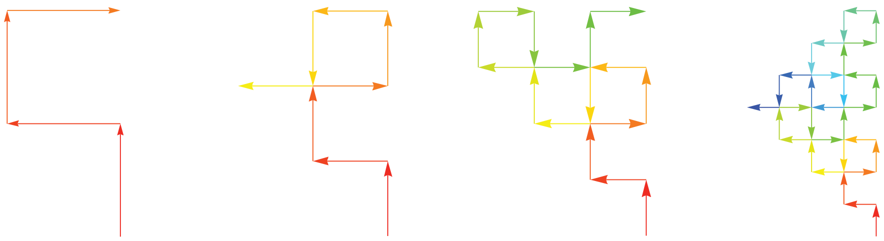
19. $x c_o(x)^2 = x(C_0 + C_1 x + C_2 x^2 + \ldots)^2 = C_0{}^2 x + (C_0 C_1 + C_1 C_0) x^2 + (C_0 C_2 + C_1 C_1 + C_2 C_0) x^3 + \ldots$, which simplifies to $C_1 x + C_2 x^2 + \ldots = c_o(x) - C_0 = c_o(x) - 1$ by Segner's recurrence relation. So, $x c_o(x)^2 - c_o(x) + 1 = 0$, and we can obtain $c_o(x)$ by the Babylonian formula for the roots of a quadratic equation. Specifically, we have $c_o(x) = \frac{1 \pm \sqrt{(-1)^2 - 4x}}{2x}$. As $c_o(x) \to 1$ when $x \to 0$, the correct root is $c_o(x) = \frac{1 - \sqrt{1 - 4x}}{2x}$.

20. We define a sequence $\{y_i\}$ such that:
- $y_i = A$ if k is even and $x_k = 1$;
- $y_i = B$ if k is odd and $x_k = 1$;
- $y_i = C$ if k is even and $x_k = -1$;
- $y_i = D$ if k is odd and $x_k = -1$.

Consider the string Y_l obtained by concatenating the first 2^l elements of $\{y_i\}$. It is straightforward to verify that $Y_0 = B$, and that Y_{l+1} can be obtained from Y_l by applying the L-system with substitution rules: $A \to D\,C$, $B \to B\,C$, $C \to B\,A$, and $D \to D\,A$. The first few terms of $\{Y_i\}$ are $B \to B\,C \to B\,C\,B\,A \to B\,C\,B\,A\,B\,C\,D\,C \to \ldots$. Interpret these strings as sequences of instructions for moving an insect on the integer lattice $\mathbb{Z}^2$:
- A: move east;
- B: move north;
- C: move west;

- *D*: move south;

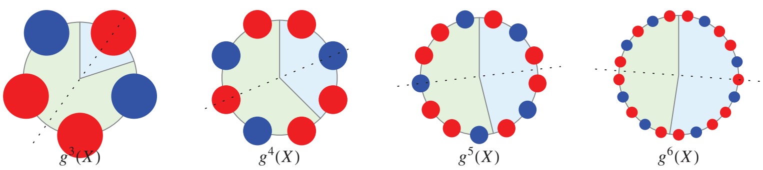
The pattern generated by this is a space-filling curve bounded by the lines $y = -x$ and $x = 0$ (with the first few iterations displayed below). The substitution rules effectively reflect the path in a NNW line through the origin, combined with a dilation of scale factor $\sqrt{2}$. Hence, the insect always remains in the octant bounded by the lines $y = -x$ and $x = 0$. Remaining to the right of $y = -x$ is equivalent to the condition $a_n + b_n \geq c_n + d_n$ (for all $n \in \mathbb{Z}^+$), where $a_n = |\{k : y_k = A\}|$, *et cetera*. This is in turn equivalent to the desired inequality.



21. We can scale the set of intervals without affecting anything, so let's colour $[-1, 1]$ instead. Note that if we have a set of monic polynomials of degrees $0, 1, \ldots, n$, inclusive, which satisfy the condition $(\int_{\text{red}} P(x)\, dx = \int_{\text{blue}} P(x)\, dx)$, then the condition holds for all linear combinations of them (namely all polynomials of degree $\leq n$). If we have a colouring that is symmetric (i.e. x is coloured identically to $-x$ for all $x \in [-1, 1]$), then all odd functions satisfy the condition. Similarly, if we have a colouring that is antisymmetric (i.e. x is coloured oppositely to $-x$ for all $x \in [-1, 1]$), then all even functions satisfy the condition. Begin with the colouring C_0, where the interval $[-1, 0]$ is red and $[0, 1]$ is blue. Clearly, all constant functions (polynomials of degree 0) satisfy the condition. Let C^* denote the complement of C, where an interval is red in C^* if and only if it is blue in C. We can scale C_0 to the interval $[-1, 0]$, and scale $C_0{}^*$ to the interval $[0, 1]$, and append them to form the colouring C_1. As C_0 works for all degree-0 polynomials, so must the scaled copies of C_0 and $C_0{}^*$, and thus also C_1. As C_1 is symmetric, it must also work for the function $y = x$, and therefore all linear polynomials. We define C_2 by performing this operation on C_1, resulting in an antisymmetric colouring which must also satisfy $y = x^2$, and therefore all quadratics. Continuing in this manner, we obtain C_{2013} which works for all degree-2013 polynomials. Note that this colouring is related to the Thue-Morse sequence.

22. Observe that $g(g(X)) = X\,Y\,X$ is the concatenation of $g(X) = X\,Y$ followed by X. Hence, $g^{n+2}(X)$ is the concatenation of $g^{n+1}(X)$ followed by $g^n(X)$, and thus $|g^{n+2}(X)| = |g^{n+1}(X)| + |g^n(X)|$. This is the recurrence relation for the Fibonacci sequence.

23. As opposed to considering strings of symbols, consider instead beads on a necklace (positioned at the kth roots of unity, where k is the length of the string). The substitution rules $X \to X\,Y$ and $Y \to X$ are then equivalent to the alternative rules $X \to \langle\, Y \,\rangle$ and $Y \to \langle\,\rangle$, where we consider $\langle$ and $\rangle$ to be 'half-beads', where $\rangle\langle = X$, followed by a rotation of the entire necklace by one half-bead to restore the correct orientation. As the substitution rules map each symbol to a palindromic string, and the initial necklace has bilateral symmetry, then all subsequent necklaces have bilateral symmetry. We then reflect the 'start' of the necklace in this axis of symmetry to produce another position. Cutting at these two positions will clearly result in two palindromic strings. Using this idea of the half-bead substitution (which preserves the axis of symmetry) followed by a rotation, it is straightforward to show that $g^n(X)$ has palindromic substrings of length $F_{n+1} - 2$ and $F_n + 2$, where F_n is the nth Fibonacci number.



$g^3(X)$ $g^4(X)$ $g^5(X)$ $g^6(X)$

24. We already proved, in the previous exercise, that the first $F_{n+2} - 2$ symbols of $g^{n+1}(X)$ is a palindrome. As we obtain $g^n(X)$ by taking the first F_{n+2} symbols of the infinite golden string, $g^\omega(X)$, the first $F_{n+2} - 2$ symbols of $g^n(X)$ also form a palindrome.

25. Consider the trapezium formed by the centres of the circles and the points of tangency with the real line. The hypotenuse has length $\frac{1}{2}\left(\frac{1}{q_1^2} + \frac{1}{q_2^2}\right)$, the base has length $\frac{p_2}{q_2} - \frac{p_1}{q_1}$ and the difference between the left and right heights is $\pm\frac{1}{2}\left(\frac{1}{q_1^2} - \frac{1}{q_2^2}\right)$. Applying Pythagoras' theorem, we obtain $\frac{1}{4}\left(\frac{1}{q_1^2} + \frac{1}{q_2^2}\right)^2 = \left(\frac{p_2}{q_2} - \frac{p_1}{q_1}\right)^2 + \frac{1}{4}\left(\frac{1}{q_1^2} - \frac{1}{q_2^2}\right)^2$. Expanding, this results in $\left(\frac{p_2}{q_2} - \frac{p_1}{q_1}\right)^2 = \frac{1}{q_1^2\, q_2^2}$, so $\frac{p_2}{q_2} - \frac{p_1}{q_1} = \frac{1}{q_1\, q_2}$. Multiplying by $q_1\, q_2$ gives the identity $p_2\, q_1 - p_1\, q_2 = 1$.

26. We have three equations, namely $p_2\, q_1 - p_1\, q_2 = 1$, $p_3\, q_1 - p_1\, q_3 = 1$, and $p_2\, q_3 - p_3\, q_2 = 1$. Subtracting the third equation from the second gives us $(p_1 + p_2)\, q_3 = (q_1 + q_2)\, p_3$, which rearranges to give $\frac{p_3}{q_3} = \frac{p_1+p_2}{q_1+q_2}$.

27. Let $j > i$. Then $s_j = k\, s_i + 1$, so $s_j \equiv 1 \pmod{s_i}$. By applying Euclid's algorithm, s_i and s_j have a greatest common divisor of 1.

28. By reverse-engineering the definition, we get the recurrence relation $s_{n+1} - 1 = s_n(s_n - 1)$. Assume that $\frac{1}{s_0} + \frac{1}{s_1} + \ldots + \frac{1}{s_k} = \frac{s_{k+1-2}}{s_{k+1}-1}$. Adding the next reciprocal would give us $\frac{s_{k+1}-2}{s_{k+1}-1} + \frac{1}{s_{k+1}} = \frac{s_{k+1}{}^2-s_{k+1}-1}{s_{k+1}{}^2-s_{k+1}} = \frac{s_{k+2}-2}{s_{k+2}-1}$. So, by induction, this tends towards 1.

29. Assume that k is not a power of two, so $k = l\, p$ for some odd prime p and integer l. Then, $\left(2^l + 1\right)\left(1 - 2^l + 2^{2\,l} - \ldots + 2^{(p-1)\,l}\right) = 2^k + 1$, thus proving that $2^k + 1$ is composite.

Inequalities

Sums of squares

Over the complex numbers, every polynomial has at least one root by the fundamental theorem of algebra. Over the reals, however, it is possible to define polynomials that are always greater than (or equal to) zero. These are known as *positive (semi)definite* functions. One such example is $x^2 + y^2 \geq 0$, which is true for all $x, y \in \mathbb{R}$. In general, as squares of real numbers are non-negative, sums of squares are also non-negative. This is the most basic useful inequality.

- If $x_1, x_2, \ldots, x_n \in \mathbb{R}$ and $\alpha_1, \alpha_2, \ldots, \alpha_n > 0$, then $\alpha_1 x_1^2 + \alpha_2 x_2^2 + \ldots + \alpha_n x_n^2 \geq 0$, with equality if and only if $x_1^2 = x_2^2 = \ldots = x_n^2 = 0$. **[Sum of squares inequality]**

Artin proved Hilbert's seventeenth problem, namely that every positive semidefinite polynomial (and, by extension, rational function) can be expressed as the sum of squares of rational functions. Charles Delzell later developed an algorithm to do so. Hence, it is *theoretically* possible to prove any inequality involving rational functions simply by reducing it to the sum of squares inequality. However, this approach is similar in its impracticality to building an automobile using Stone Age tools. Certainly, it is impossible in the 270 minutes allocated in the International Mathematical Olympiad. Nevertheless, we can still tackle *some* basic inequalities in this way, especially if they are expressible as the sums of squares of polynomials.

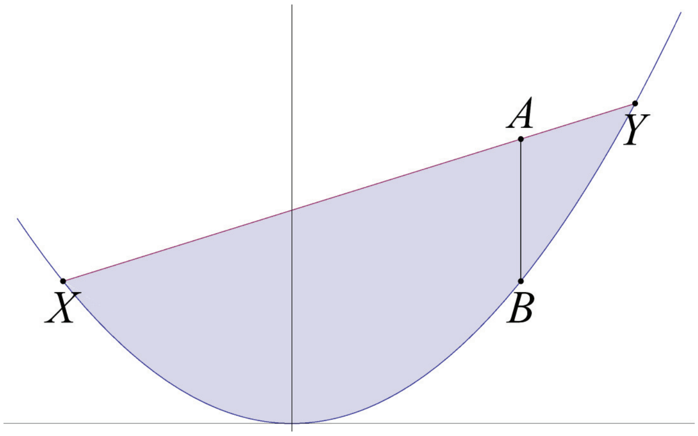
1. Prove that $x^2 + y^2 + z^2 \geq x\,y + y\,z + z\,x$.

Jensen's inequality

According to Ross Atkins, "Jensen's inequality is greater than or equal to all other inequalities". This strongly indicates that it is advisable to assimilate it into one's problem-solving repertoire. It is geometrically very obvious, namely that the barycentre of a convex figure is located inside it. This makes it all the more remarkable that so many useful inequalities, such as the power means inequality, are trivialised by Jensen's inequality.

- A continuous function f is *convex* over an interval (a, b) if, for all $x_1, x_2 \in (a, b)$ and $\alpha_1, \alpha_2 \in [0, 1]$ such that $\alpha_1 + \alpha_2 = 1$, we have $f(\alpha_1 x_1 + \alpha_2 x_2) \leq \alpha_1 f(x_1) + \alpha_2 f(x_2)$. If the reverse inequality holds instead, the function is *concave*. **[Definition of convexity]**

This is most easily represented with the aid of a diagram:



For any two points X and Y on the curve of a convex function, any point A on the line segment XY lies above the curve. The Australian IMO team leader, Ivan Guo, created a mnemonic for remembering the shapes of generic convex and concave functions:

- Ivan: "Con**cave** looks like a **cave**, and con**vex** looks like a **vex**."
- Someone else: "What's a vex?"
- Ivan: "An upside-down cave."

2. Let f be a convex function over the interval (a, b). Let $\{x_1, x_2, \ldots, x_n\} \subset (a, b)$ and $\{\alpha_1, \alpha_2, \ldots, \alpha_n\} \subset [0, 1]$ such that $\alpha_1 + \alpha_2 + \ldots + \alpha_n = 1$. Show that $f(\alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_n x_n) \leq \alpha_1 f(x_1) + \alpha_2 f(x_2) + \ldots + \alpha_n f(x_n)$. **[Weighted Jensen's inequality]**

Observe that the $n = 2$ case of the weighted Jensen inequality is just the definition of convexity. It is often quoted as the slightly less general (but asymptotically equivalent) theorem where $\alpha_1 = \alpha_2 = \ldots = \alpha_n = \frac{1}{n}$.

- Let f be a convex function over the interval (a, b), and let $\{x_1, x_2, \ldots, x_n\} \subset (a, b)$. Then $f\left(\frac{1}{n}(x_1 + x_2 + \ldots + x_n)\right) \leq \frac{1}{n}(f(x_1) + f(x_2) + \ldots + f(x_n))$. **[Jensen's inequality]**

3. If $\{x_1, x_2, \ldots, x_n\}$ are all positive, show that $\frac{1}{n}(x_1 + x_2 + \ldots + x_n) \geq \sqrt[n]{x_1 x_2 \ldots x_n}$. **[AM-GM inequality]**

4. If a and b are two non-zero real numbers such that $a \geq b$, show that $\sqrt[a]{\frac{1}{n}(x_1{}^a + x_2{}^a + \ldots + x_n{}^a)} \geq \sqrt[b]{\frac{1}{n}\left(x_1{}^b + x_2{}^b + \ldots + x_n{}^b\right)}$. **[Power means inequality]**

The *arithmetic mean*, *quadratic mean* and *harmonic mean* arise when a is 1, 2 and −1, respectively. The geometric mean is the limit as $a \to 0$.

Muirhead's inequality

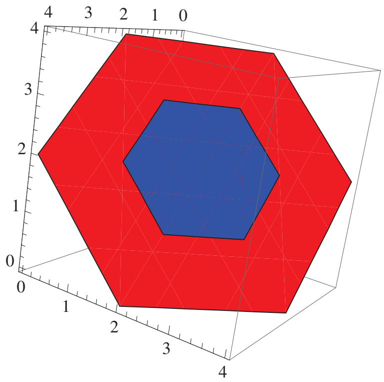
Muirhead's inequality is a powerful generalisation of the AM-GM inequality. Before we can define it, however, it is necessary to introduce the idea of *majorisation*.

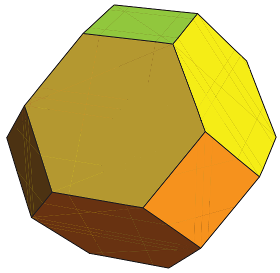
- Let $a_1 + a_2 + \ldots + a_n = 1$ and $b_1 + b_2 + \ldots + b_n = 1$, and all $a_i \in [0, 1]$ and $b_i \in [0, 1]$. Assume further that the sequences are ordered such that $a_1 \geq a_2 \geq \ldots \geq a_n$ and $b_1 \geq b_2 \geq \ldots \geq b_n$. Then $\{a_i\}$ *majorises* $\{b_i\}$ if and only if $a_1 + a_2 + \ldots + a_k \geq b_1 + b_2 + \ldots + b_k$ for all $k \in [1, n]$. **[Definition of majorisation]**

The sequence (4, 0, 0, 0), for example, majorises (1, 1, 1, 1), as they are sorted into descending order and the following inequalities hold:

- $4 \geq 1$;
- $4 + 0 \geq 1 + 1$;
- $4 + 0 + 0 \geq 1 + 1 + 1$;
- $4 + 0 + 0 + 0 = 1 + 1 + 1 + 1$.

Occasionally, the notation $(4, 0, 0, 0) \succcurlyeq (1, 1, 1, 1)$ is used to denote this relationship. Majorisation may appear at first to be a contrived relation, although it has several equivalent and more enlightening formulations. We interpret $\underline{a} = (a_1, a_2, \ldots, a_n)$ as a vector in $\mathbb{R}^n$, and consider the set of $n!$ (not necessarily distinct) vectors obtained by permuting the elements of the vector $\underline{a}$. They all lie in the $(n-1)$-dimensional plane with equation $x_1 + x_2 + \ldots + x_n = a_1 + a_2 + \ldots + a_n$, and form the vertices of a *permutation polytope*. In general (when all elements are distinct), the three-variable case is a hexagon, whereas the four-variable case is a truncated octahedron.





The red and blue hexagons correspond to the sets {4, 2, 0} and {3, 2, 1}, respectively. The condition that the red hexagon contains the blue hexagon is equivalent to $\{4, 2, 0\} \succcurlyeq \{3, 2, 1\}$, which in turn is equivalent to the *Birkhoff-von Neumann theorem*: (3, 2, 1) can be expressed as a weighted average of permutations of (4, 2, 0). More subtly, this also implies that, for all $x, y, z \geq 0$, the polynomial $x^4 y^2 z^0 + z^4 y^2 x^0 + y^4 z^2 x^0 + x^4 z^2 y^0 + z^4 x^2 y^0 + y^4 x^2 z^0$ is greater than or equal to $x^3 y^2 z^1 + z^3 y^2 x^1 + y^3 z^2 x^1 + x^3 z^2 y^1 + z^3 x^2 y^1 + y^3 x^2 z^1$; a fact known as *Muirhead's inequality*.

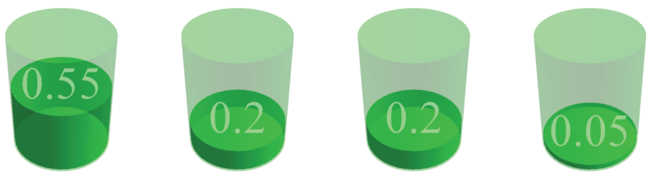
- Let $\{x_1, x_2, \ldots, x_n\}$, $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ and $\{\beta_1, \beta_2, \ldots, \beta_n\}$ be sequences of non-negative real numbers. If $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ majorises $\{\beta_1, \beta_2, \ldots, \beta_n\}$, then $\sum_{\text{sym}} (x_1^{\alpha_1} x_2^{\alpha_2} \ldots x_n^{\alpha_n}) \geq \sum_{\text{sym}} \left(x_1^{\beta_1} x_2^{\beta_2} \ldots x_n^{\beta_n}\right)$. The sigmas denote symmetric sums, *i.e.* sums over all $n!$ permutations of $\{x_1, x_2, \ldots, x_n\}$. **[Muirhead's inequality]**

It is discussed in https://nrich.maths.org/discus/messages/67613/Muirhead-69859.pdf. Geoff Smith described how Muirhead's inequality is not well known amongst members of the IMO jury; occasionally certain inequalities, which were highly amenable to attack by this method, appeared on the IMO as a result of this.

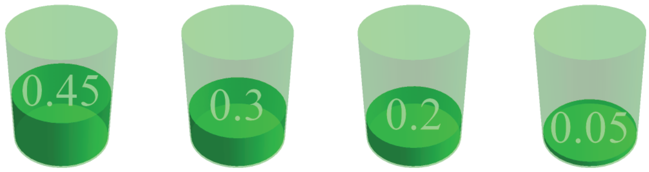
5. Prove that, for all positive real numbers x, y and z, we have $2x^3 + 2y^3 + 2z^3 \geq x^2 y + y^2 x + y^2 z + z^2 y + z^2 x + x^2 z$.

Majorisation as fluid transfer

We have already defined majorisation in terms of decreasing sequences and permutation polytopes. A third interpretation involves containers of fluid. In each configuration below, the total volume of fluid is 1 unit; we will assume this without loss of generality to simplify things.



Suppose we have a sequence of containers of fluid, such that if container X is immediately to the left of container Y, then X contains at least as much fluid as Y. We are allowed to siphon fluid from X to Y as long as this weak inequality is maintained. From the configuration above, we can siphon up to 0.175 units of fluid from the first container to the second one without breaking the weak inequality. In the diagram below, we have transferred 0.1 units.



This is known as a *valid q-move*, where $q = 0.1$ is the amount of fluid transferred. We can continue in this manner. The *fluid transfer lemma* states that we can get from an initial sequence S_0 to (arbitrarily close to) a target sequence S_ω by applying valid q-moves if and only if S_0 majorises S_ω. A more formal definition follows:

- Suppose S_0 and $S_\omega = \{b_1, b_2, \ldots, b_n\}$ are two weakly decreasing sequences of non-negative real numbers, each with unit sum and length n. Let $\epsilon > 0$ be a small real number. Define a *valid q-move* to be an operation $\{a_1, a_2, \ldots, a_k, a_{k+1}, \ldots, a_n\} \to \{a_1, a_2, \ldots, a_k - q, a_{k+1} + q, \ldots, a_n\}$ such that the sequence remains strictly decreasing and still majorises S_ω. Then there exists some $\delta \ll \epsilon$ such that there exists a finite sequence of N valid δ-moves $S_0 \to S_1 \to \ldots \to S_N$ such that each term of S_N differs by the corresponding term of S_ω by at most ϵ if and only if S_0 majorises S_ω. **[Fluid transfer lemma]**

Proof:

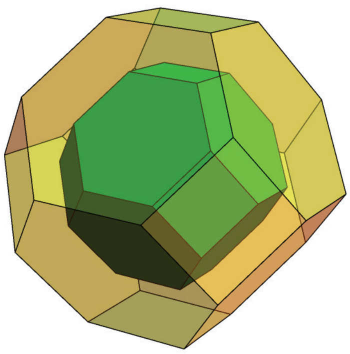
The 'only if' part is much easier, as it is evident that S_0 majorises S_1, which in turn majorises S_2. By induction, S_0 majorises S_N. If S_0 does not majorise S_ω, then one of the weak inequalities must be broken by an amount h. If we let $\epsilon < \frac{h}{n}$, then S_N must be sufficiently close to S_ω to also break one of those inequalities. Hence, S_0 does not majorise S_N, so we have a contradiction.

For the 'if' part, note that there are a finite number of attainable configurations for a given S_0 and δ, and the process cannot cycle, so must eventually terminate. Suppose we perform valid δ-moves arbitrarily until we reach a position S_N where no further valid δ-moves are possible.

By definition, for each pair of adjacent elements (a_i, a_{i+1}) in S_N, it must be the case that either:

- $a_i - a_{i+1} < 2\,\delta$ (in which case applying a δ-move would break the weakly decreasing criterion);
- or $(a_1 + a_2 + \ldots + a_i) - (b_1 + b_2 + \ldots + b_i) < \delta$ (in which case applying a δ-move would break the majorisation criterion).

If the first case applies to all pairs of adjacent elements, we have $a_1 - a_n < 2\,(n-1)\,\delta$. So, each element must be within $2\,(n-1)\,\delta$ of the mean, $\frac{1}{n}$. As S_N majorises S_ω, the same must be true of S_ω. Hence, corresponding elements can differ by no more than $4\,(n-1)\,\delta$, which we can make smaller than ϵ by letting δ be sufficiently small. This leaves the alternative case where there exists some i such that $0 < (a_1 + a_2 + \ldots + a_i) - (b_1 + b_2 + \ldots + b_i) < \delta$. In that case, we can split the problem into two separate problems: one involving the first i elements of the sequences, and the other involving the last $n - i$. (We need not worry that the sum of the first i elements of S_N is slightly greater than that of S_ω, as we can make the difference arbitrarily small. It is not important that things are exact, as long as the largest accumulative error is smaller than ϵ.) By inducting on the number of elements, we prove the fluid transfer lemma.



Returning to the geometric interpretation, this means that we can incrementally move the vertices of the larger polytope inwards (varying two coordinates of any vertex at any one time whilst preserving the full symmetry group) until it becomes arbitrarily close to 'suffocating' the smaller polytope. This is rather intuitive, and implies the Birkhoff-von Neumann theorem.

Energy minimisation lemma

A corollary of this lemma is the *energy minimisation lemma*. The proof relies on concepts from real analysis such as continuity and convergence, which are taught in most undergraduate maths degrees (such as the Cambridge Mathematical Tripos).

- Suppose we have a continuous function $E : \mathbb{R}^n \to \mathbb{R}$, known as the *energy function*. Suppose that applying a valid q-move to $S = \{a_1, a_2, \ldots, a_n\}$ cannot increase the value of $E(S)$. If we have two sequences S_0 and S_ω, such that S_0 majorises S_ω, then $E(S_0) \geq E(S_\omega)$. **[Energy minimisation lemma]**

Effectively, we associate an 'energy function' with the configuration of containers, such that the energy either remains constant or decreases whenever a valid q-move is applied. The energy minimisation lemma states that $E(S_0) \geq E(S_\omega)$ if S_0 majorises S_ω.

Proof:

Due to the fluid transfer lemma, we can apply valid q-moves to S_n to result in a new sequence (S_{n+1}) where each term differs from S_ω by at most $\epsilon = e^{-n}$. Starting from S_0, we produce an infinite sequence of sequences $\{S_0, S_1, \ldots\}$ where each term is an increasingly close approximation to S_ω. More specifically, this sequence of sequences *converges* to S_ω. As E is a continuous function, this means that $\{E(S_0), E(S_1), \ldots\}$ must converge to $E(S_\omega)$. Also, as valid q-moves cannot increase the value of $E(S)$, we have $E(S_0) \geq E(S_1) \geq \ldots$; by the monotone convergence theorem, this means $E(S_\omega)$ is the infimum of these terms, and therefore no larger than any of them. The result then follows.

Generalised Muirhead inequality

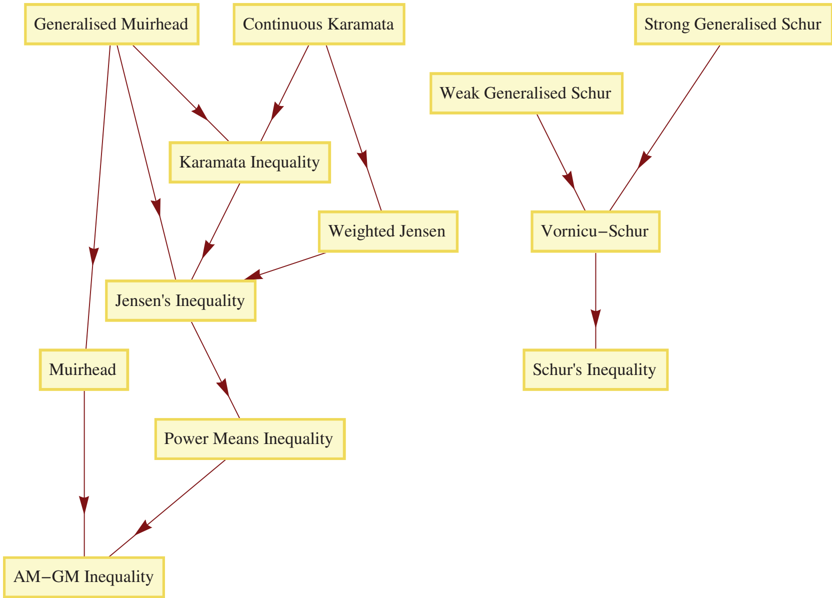
Using the lemmas developed above, it is straightforward to prove the generalised Muirhead inequality.

6. Let $f : (a, b) \to \mathbb{R}$ be a convex continuous function. Let $\alpha_1 \geq \beta_1 \geq \beta_2 \geq \alpha_2 \geq 0$, such that $\alpha_1 + \alpha_2 = \beta_1 + \beta_2 = 1$, and let $\{x_1, x_2\} \subset (a, b)$. Prove that $f(\alpha_1 x_1 + \alpha_2 x_2) + f(\alpha_1 x_2 + \alpha_2 x_1) \geq f(\beta_1 x_1 + \beta_2 x_2) + f(\beta_1 x_2 + \beta_2 x_1)$. **[Generalised Muirhead's inequality for 2 variables]**

7. Let $f : (a, b) \to \mathbb{R}$ be a convex continuous function, and let $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ be a weakly decreasing sequence of non-negative reals with unit sum. Let $\{x_1, x_2, \ldots, x_n\} \subset (a, b)$. Define the function $g(\alpha_1, \alpha_2, \ldots, \alpha_n, x_1, x_2, \ldots, x_n) = \sum_{\text{sym}} f(\alpha_1 x_{\sigma(1)} + \alpha_2 x_{\sigma(2)} + \ldots + \alpha_n x_{\sigma(n)})$, where the sum is taken over all $n!$ permutations σ of $\{1, 2, \ldots, n\}$. Prove that applying a valid q-move to $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ cannot cause g to increase.

8. Let f be a convex continuous function over (a, b) and $\{x_1, x_2, \ldots, x_n\} \subset (a, b)$. Let $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ and $\{\beta_1, \beta_2, \ldots, \beta_n\}$ be weakly decreasing sequences of non-negative reals, each with unit sum. Further, the former sequence majorises the latter. Prove that $\sum_{\text{sym}} f(\alpha_1 x_{\sigma(1)} + \alpha_2 x_{\sigma(2)} + \ldots + \alpha_n x_{\sigma(n)}) \geq \sum_{\text{sym}} f(\beta_1 x_{\sigma(1)} + \beta_2 x_{\sigma(2)} + \ldots + \beta_n x_{\sigma(n)})$, where the sums are taken over all $n!$ permutations σ of $\{1, 2, \ldots, n\}$. **[Generalised Muirhead's inequality]**

We can derive the ordinary Muirhead's inequality by letting $f(x) = e^x$. Similarly, Jensen's inequality follows from using the sequences $\{1, 0, \ldots, 0\} \succcurlyeq \left\{\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right\}$. This idea of inequalities generalising other inequalities gives a hierarchy:

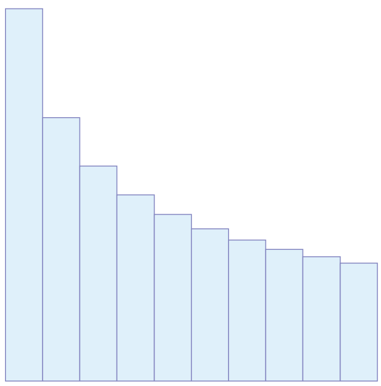


Karamata inequality

Consider the generalised Muirhead inequality. If we let $\{x_1, x_2, \ldots, x_n\} = \{1, 0, 0, \ldots, 0\}$, then we obtain the *Karamata inequality* as a special case.

- Suppose $\{a_i\}$ majorises $\{b_i\}$, and f is a convex function. Then $f(a_1) + \ldots + f(a_n) \geq f(b_1) + \ldots + f(b_n)$. **[Karamata inequality]**

This can be extended in another direction. We can assume without loss of generality that $\sum a_i = \sum b_i = 1$. Effectively, we can consider two new functions, $g'(x) = n\, a_{\lceil x n \rceil}$ and $h'(x) = n\, b_{\lceil x n \rceil}$, which are defined on the open interval $(0, 1)$. As $\{a_i\}$ majorises $\{b_i\}$ and the sequences are sorted in descending order, we have that g' and h' are weakly decreasing and $\int_0^k g'(x)\, dx \geq \int_0^k h'(x)\, dx$ for all $0 < k < 1$. We represent these integrals by $g(x)$ and $h(x)$, respectively. It is clear that $g(0) = h(0) = 0$ and $g(1) = h(1) = 1$.



The graph of $g'(x)$ is a collection of n rectangles of decreasing height. Integrating this to obtain $g(x)$ results in a concave line formed from n straight line segments of decreasing gradient. If we take the limit as n tends towards infinity, the sequences in Karamata's inequality are replaced with arbitrary non-negative decreasing functions, g' and h'.

- Suppose g and h are increasing concave functions with domain $[0, 1]$ such that $g(0) = h(0) = 0$, $g(1) = h(1) = 1$ and $g(k) \geq h(k)$ for all $k \in [0, 1]$. The derivatives of $g(x)$ and $h(x)$ with respect to x are denoted $g'(x)$ and $h'(x)$, respectively. Let f be an arbitrary convex function. Then $\int_0^1 f(g'(x))\,dx \geq \int_0^1 f(h'(x))\,dx$. **[Continuous Karamata inequality]**

Schur's inequality

A useful inequality that can be proved using sums of squares is *Schur's inequality*. Unlike the previous inequalities, which generalise to arbitrarily many variables, this has just three terms.

9. Suppose $a \geq b \geq c$ and $x + z \geq y \geq 0$. Show that $x^2(a-b)(a-c) + y^2(b-c)(b-a) + z^2(c-a)(c-b) \geq 0$. **[Strong 6-variable Schur]**

It is often quoted as the much weaker result shown below.

10. Show also that $x(a-b)(a-c) + y(b-c)(b-a) + z(c-a)(c-b) \geq 0$. **[Weak 6-variable Schur]**

This can be used, with a little work, to form a very powerful inequality.

11. Let $f : \mathbb{R} \to \mathbb{R}^+$ be a function expressible as the sum of non-negative monotonic functions. Let $g : \mathbb{R} \to \mathbb{R}$ and $h : \mathbb{R} \to \mathbb{R}$ be odd and increasing. Show that
$f(a)\, g(h(a-b)\, h(a-c)) + f(b)\, g(h(b-c)\, h(b-a)) + f(c)\, g(h(c-a)\, h(c-b)) \geq 0$. **[Weak generalised Schur]**

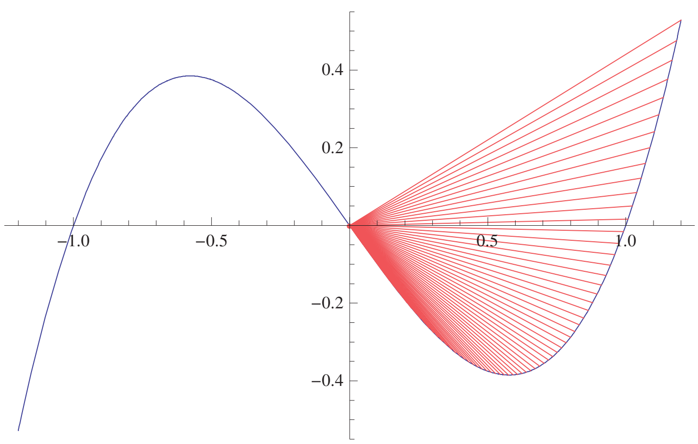
When $h(w) = w^k$ and $g(w) = w$, this is known as the *Vornicu-Schur inequality*. With the additional constraints of $k = 1$ and $f(w) = w^p$, this is simply Schur's inequality.

- If $a, b, c \in \mathbb{R}^+$, then $a^p(a-b)(a-c) + b^p(b-c)(b-a) + c^p(c-a)(c-b) \geq 0$. **[Schur's inequality]**

It is popularly believed that a suitable combination of Muirhead and Schur can conquer any inequality. This is obviously an exaggeration, since neither can prove (for instance) Jensen's inequality. Nevertheless, most symmetric inequalities in three variables submit to such an attack.

12. Prove that $x^6 + y^6 + z^6 + 3\,x^2\,y^2\,z^2 \geq 2\,x^3\,y^3 + 2\,y^3\,z^3 + 2\,z^3\,x^3$.

Nevertheless, we can go further. The strong 6-variable Schur inequality can also be generalised in a similar way to its weaker counterpart. We define a function f to be *positive-illuminable* if $f(\alpha\, x) \leq \alpha\, f(x)$ for all $0 \leq \alpha \leq 1$ and $x \geq 0$. Informally, this means that a light source placed infinitesimally above the origin will be able to illuminate every point on the curve $y = f(x)$, $x \geq 0$ from above. This is demonstrated in the following diagram, where no rays emitted from the origin intersect the curve twice. Positive-illuminability is a weaker condition than convexity.



We are now in a position to state and prove the stronger generalised form of Schur's inequality.

13. Let $f : \mathbb{R} \to \mathbb{R}^+$ be a function expressible as the sum of non-negative monotonic functions. Let $g : \mathbb{R} \to \mathbb{R}$ and $h : \mathbb{R} \to \mathbb{R}$ be odd, increasing and positive-illuminable. Show that $f(a)^2\, g(h(a-b)\, h(a-c)) + f(b)^2\, g(h(b-c)\, h(b-a)) + f(c)^2\, g(h(c-a)\, h(c-b)) \geq 0$. **[Strong generalised Schur]**

Calculus

Although ideas of limiting processes and integration can be traced back to Archimedes, our modern understanding of calculus was developed much later. It was conceived independently, and almost simultaneously, by Sir Isaac Newton and Gottfried Leibniz. As Newton only considered differentiation with respect to time, we currently use Leibniz's (much clearer) notation instead.

In the explorations of various general inequalities, terms such as 'increasing', 'convex' and 'positive-illuminable' appeared. It is possible to express each of these concepts in the environment of calculus. We will represent the first derivative of a function $f(x)$ with $f'(x)$. The second derivative, $f''(x)$, is also of interest.

- A differentiable function f is increasing on an interval I if and only if $f'(x) \geq 0$ for all $x \in I$.

This is intuitive. The derivative measures the rate of increase of a function, which we require to be non-negative. Convex functions have an increasing gradient, so we require the second derivative to be positive.

- A differentiable function f is convex on an interval I if and only if $f''(x) \geq 0$ for all $x \in I$.

The properties 'decreasing' and 'concave' are similarly defined, but with the '≥' operator reversed in direction.

14. Prove that $e^{2x} + e^{2y} \geq 2\, e^{x+y}$ for all $x, y \in \mathbb{R}$.

So far, we have considered calculus in one variable. Nevertheless, it is possible to delve into the realms of *multivariate calculus*. The main approach is to consider the *partial derivative* of a function with respect to a variable. To do this, we allow one variable to vary and force the others to remain constant. For example, $z = y^2 + 2xy$ has the partial derivatives $\frac{\partial z}{\partial x} = 2y$ and $\frac{\partial z}{\partial y} = 2y + 2x$.

If we want to show that the value of a function $z = f(x, y)$ increases as we move parallel to the x-axis, we need to show that $\frac{\partial z}{\partial x}$ is always non-negative. To investigate how it changes as we move parallel to the vector (3, 2), we are interested in $3\frac{\partial z}{\partial x} + 2\frac{\partial z}{\partial y}$.

15. Let x, y and z be positive real numbers. Prove that $4(x+y+z)^3 > 27\left(x^2 y + y^2 z + z^2 x\right)$. **[BMO2 2010, Question 4]**

Warning: A stationary point is a point where all partial derivatives are zero. Be careful, however, as this could be a point of inflection or saddle point instead of a minimum or maximum. Also, calculus does not guarantee that a particular extremum is global; for example, $x^3 - 3x$ has a *local minimum* at $x = 1$, but still takes on arbitrarily low values. You should bear this in mind when attempting to tackle a problem using calculus, especially Lagrange multipliers. If you want to use calculus to locate an extremum of a function, it is invariably a good idea to sketch a graph of the function first. Unfortunately, your two-dimensional paper and three-dimensional imagination are insufficient when there are many variables.

Lagrange multipliers

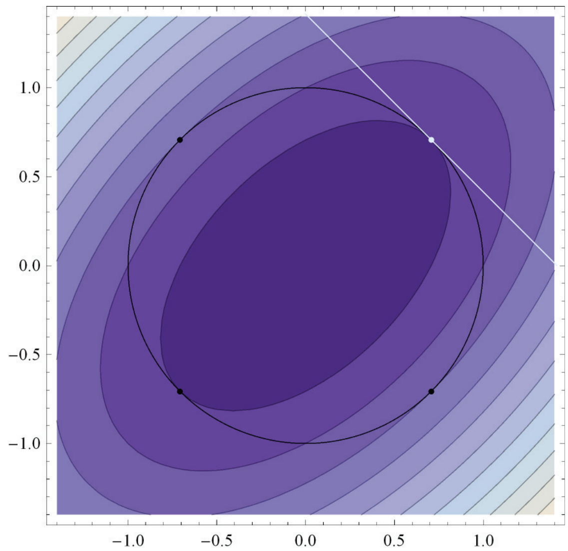
Suppose we have some additional constraints on the variables in an inequality. For example, we encountered a problem where we had to minimise $x^2 + y^2 + z^2$ subject to the constraint that $x^3 + y^3 + z^3 - 3xyz = 1$. One way of incorporating the side constraint is to *homogenise* the inequality. In that example, it would involve making all terms in $x^2 + y^2 + z^2$ of degree zero. In this case, it 'reduces' to the following problem, which is really quite horrible:

- Find the minimum value of $\frac{x^2+y^2+z^2}{(x^3+y^3+z^3-3xyz)^{\frac{2}{3}}}$, where $x, y, z \in \mathbb{R}$.

If we could guess that the minimum value is 1 (which is by no means obvious), then it is equivalent to proving that $\left(x^2 + y^2 + z^2\right)^3 \geq \left(x^3 + y^3 + z^3 - 3xyz\right)^2$. One could attempt to bash this degree-6 polynomial inequality with any combination of Muirhead, Schur and the $u\,v\,w$ method (as we shall do shortly), but it lacks a certain elegance.

A method that is more amenable to incorporating side constraints into problems is the use of *Lagrange multipliers*, which enable the application of calculus. If we want to minimise the value of f (which is a function of some variables) subject to the algebraic constraint $g = 0$ (where g is a function of those variables), then we introduce a new variable, λ. We consider the function $\Lambda = f + \lambda g$, and minimise it by locating its stationary points. We'll start with a simple non-trivial example in two variables:

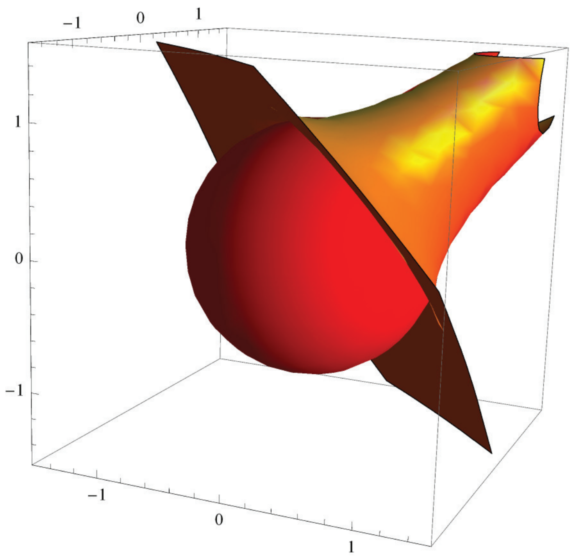
- Find the minimum and maximum of $f = x^2 + y^2 - xy$, subject to the constraint $x^2 + y^2 = 1$.



The contours of f are ellipses of the form $f = x^2 + y^2 - xy = k$, and we want to find the ones that touch the circle $g = x^2 + y^2 - 1 = 0$. Let $\Lambda = f + \lambda g$. Consider a point of tangency, such as that highlighted in the diagram above. We imagine setting a new orthogonal coordinate system centred at this point, with an axis normal to the common tangent. Call this coordinate ω. The partial derivatives $\frac{\partial f}{\partial \omega}$ and $\frac{\partial g}{\partial \omega}$ are both non-zero, whereas the partial derivatives with respect to the other axes are all zero. Hence, if we let $\lambda = -\frac{\frac{\partial f}{\partial \omega}}{\frac{\partial g}{\partial \omega}}$, the partial derivatives of Λ with respect to all of the (new) axes are zero, so the partial derivatives are all zero. In other words, any extremal point of f on the curve $g = 0$ is also a stationary point of Λ. This method only works if $\frac{\partial g}{\partial \omega}$ is non-zero at the extremal points, so it is important to verify this before proceeding with the method of Lagrange multipliers. In this example, g is quadratic and only stationary at the origin, so we can safely apply the method.

- Find the stationary points of $\Lambda = x^2 + y^2 - xy + \lambda\left(x^2 + y^2 - 1\right)$.

Equating $\frac{\partial \Lambda}{\partial x} = 0$ and $\frac{\partial \Lambda}{\partial y} = 0$, we have the equations $2x - y + 2\lambda x = 0$ and $2y - x + 2\lambda y = 0$, which simplify to $2(\lambda + 1) = \frac{y}{x} = \frac{x}{y}$. Hence, $y^2 = x^2$ and thus $x = \pm y$, from whence we obtain all four tangency points: $\left(\pm\frac{1}{\sqrt{2}}, \pm\frac{1}{\sqrt{2}}\right)$. They correspond to the maximum value $f = 3$ and minimum value $f = 1$.



We shall now contemplate the original problem. As shown above, let $f = x^2 + y^2 + z^2$ and $g = x^3 + y^3 + z^3 - 3\,x\,y\,z - 1$. This simplification is more appetising than the previous attempt at homogenising the problem.

- Find the stationary points of $\Lambda = x^2 + y^2 + z^2 + \lambda\left(x^3 + y^3 + z^3 - 3\,x\,y\,z - 1\right)$.

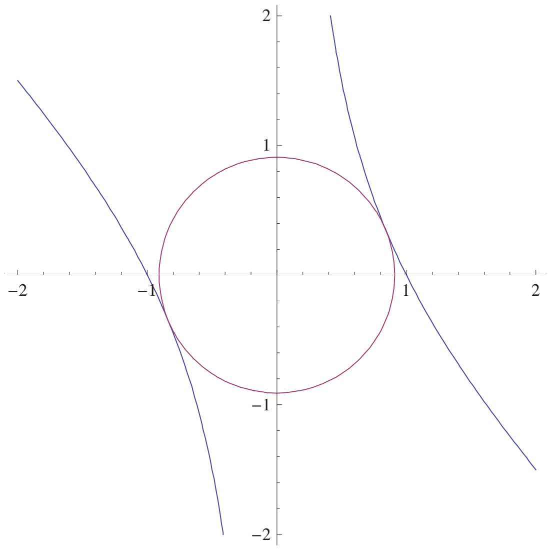
Differentiating it with respect to x gives the partial derivative $\frac{\partial\Lambda}{\partial x} = 2\,x + 3\,\lambda\,x^2 - 3\,\lambda\,y\,z$, which we wish to equate to zero. Similarly, by differentiating with respect to y and z, we obtain two more equations. (The final equation, $\frac{\partial\Lambda}{\partial\lambda} = 0$, is precisely the original side constraint, $x^3 + y^3 + z^3 - 3\,x\,y\,z = 1$.)

More interestingly, we can multiply $2\,x + 3\,\lambda\,x^2 - 3\,\lambda\,y\,z = 0$ by x to result in the cubic equation $3\,\lambda\,x^3 + 2\,x^2 = 3\,\lambda\,x\,y\,z$. Hence, x, y and z are all solutions of the equation $3\,\lambda\,x^3 + 2\,x^2 = k$, where $k = 3\,\lambda\,x\,y\,z$. Either x, y, z are the three distinct roots, or two of them are equal. In the former case, we have $x\,y + y\,z + z\,x = 0$ by Vieta's formulas, resulting in the equation $x^3 + y^3 + z^3 - 3\,x\,y\,z = (x + y + z)^3 = \left(x^2 + y^2 + z^2\right)^{\frac{3}{2}}$, and thus $x^2 + y^2 + z^2 = 1$ and we are done. In the other case, we can assume without loss of generality that $y = z$ and thus eliminate a variable.

- Find the stationary points of $\Lambda = x^2 + 2\,y^2 + \lambda\left(x^3 + 2\,y^3 - 3\,x\,y^2 - 1\right)$.

We obtain $\frac{\partial\Lambda}{\partial y} = 4\,y + 6\,\lambda\,y^2 - 6\,\lambda\,x\,y = 0$, which has solutions $y = 0$ and $\lambda(x - y) = \frac{2}{3}$. The former case clearly results in $(x, y, z) = (1, 0, 0)$, again giving a minimum of $x^2 + y^2 + z^2 = 1$. The other solution is more intricate. By considering the other partial derivative, $\frac{\partial\Lambda}{\partial x} = 2\,x + 3\,\lambda\,x^2 - 3\,\lambda\,y^2 = 2\,x + 3\,\lambda(x + y)\,(x - y) = 0$, we get $2\,x + 2\,(x + y) = 0$. This gives $2\,x = -y$, which can be substituted back into the original equation to give $-27\,x^3 = 1$, or $(x, y, z) = \left(-\frac{1}{3}, \frac{2}{3}, \frac{2}{3}\right)$. This also attains the value of $x^2 + y^2 + z^2 = 1$. We now just need to choose the minimum value of $x^2 + y^2 + z^2$, which is 1.

16. Find the distance from the closest points on the hyperbola $x\,y + x^2 = 1$ to the origin $O = (0, 0)$.



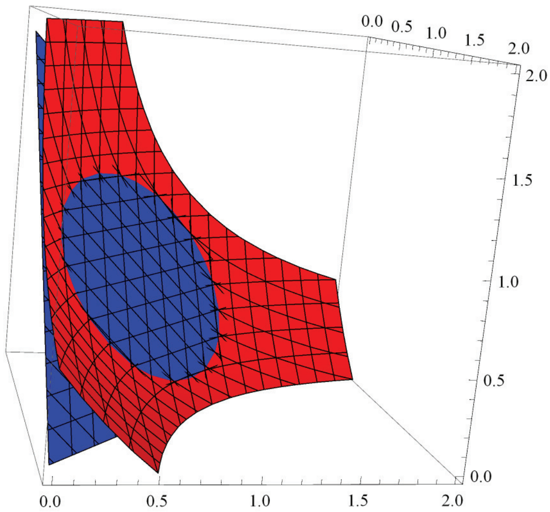
The $u\,v\,w$ method

Symmetric polynomial inequalities in three positive real variables have frequently appeared in olympiads. The '$u\,v\,w$ method' uses the idea of expressing these as polynomials in the ESPs.

- $3\,u = x + y + z$
- $3\,v^2 = x\,y + y\,z + z\,x$
- $w^3 = x\,y\,z$

17. If $x,\ y,\ z \in \mathbb{R}$, prove that $\left(x^2 + y^2 + z^2\right)^3 \geq \left(x^3 + y^3 + z^3 - 3\,x\,y\,z\right)^2$. When does equality occur?

The full power of the $u\,v\,w$ method is realised when we require that $x,\ y,\ z \geq 0$. By the AM-GM inequality, $u \geq v \geq w$ with equality if and only if $x = y = z$. This leads to an approach for tackling all three-variable symmetric polynomial inequalities of reasonably low degree.



The blue plane $x + y + z = 3\,u$ intersects the red two-sheeted hyperboloid $x\,y + y\,z + z\,x = 3\,v^2$ in a conic. It is the intersection of the blue plane with the sphere with equation $x^2 + y^2 + z^2 = 9\,u^2 - 6\,v^2$, so is a circle. If we fix u and v, we can 'move' around the circumference of the circle and examine how w varies. This can be accomplished by the method of Lagrange multipliers.

18. Show that the stationary points of $\Lambda = x\,y\,z + \lambda\left(x^2 + y^2 + z^2 + 6\,v^2 - 9\,u^2\right) + \mu(x + y + z - 3\,u)$ occur only where two of the variables are equal.

If the circle intersects the planes $x = 0$, $y = 0$ and $z = 0$, however, we must also account for the 'boundary case' where one of the variables is zero.

- If we want to find the maximum or minimum values of w^3 for some fixed u and v^2, it suffices to only check the cases where $x = 0$ or $y = z$.

Now let's suppose we are trying to prove a symmetric polynomial inequality where the degree of the greatest term is 8. It can be expressed as the inequality $F\,w^6 + 2\,G\,w^3 + H \geq 0$, where F, G, H are functions of u and v^2. This is a quadratic in w^3, so its extreme values occur when either w^3 is minimised, maximised, or reaches the stationary point. By differentiating the above expression with respect to w^3, this occurs when $F\,w + G = 0$, *i.e.* $w = -\frac{G}{F}$.

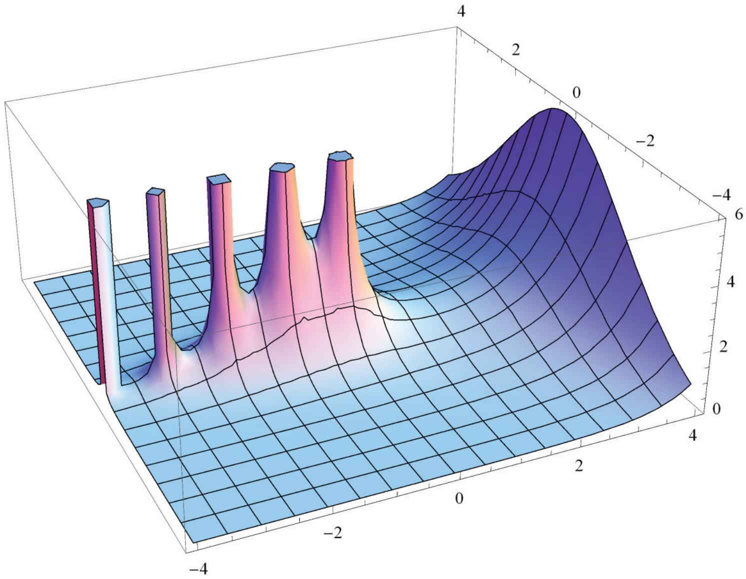
- To prove the inequality $F\,w^6 + 2\,G\,w^3 + H \geq 0$ (which is an arbitrary symmetric polynomial of degree $d \leq 8$ in three variables), where F, G, H are polynomials in u and v^2, it suffices only to check that it holds under each of the following three cases:
  - One variable is zero (without loss of generality, $x = 0$);
  - Two variables are equal (without loss of generality, $y = z$);
  - $F\,w + G = 0$. (only relevant where $d \geq 6$). **[Generalised Tejs' corollary]**

$F\,w + G = 0$ is a degree-$(d - 3)$ symmetric polynomial equation, where d is the degree of the inequality. In some problems, you may be sufficiently fortunate to find that equality can never occur, for instance if $F\,w + G > 0$ in all cases. Since this is a degree-5 inequality, it can be itself verified using Tejs' corollary.

Gamma function

The function $f(x) = 2^x$ can be defined on the positive integers by the product $\underbrace{2 \times 2 \times \ldots \times 2}_{n \text{ times}}$. If we want to extend this function to the reals and complex numbers, we can do so by using the recurrence $f(x + 1) = 2\,f(x)$. This gives an uncountably infinite number of possible contenders. If we insist that the function is continuous, differentiable and is 'logarithmically convex', then there is only one possible function: $f(x) = \exp(x \log(2))$, where $\exp(x) = e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots$ and $\log(x)$ is its inverse.

Euler did the same for the factorial function. The Gamma function is defined by $\Gamma(x) = (x - 1)!$ for $x \in \mathbb{N}$, and more generally over the positive complex numbers with positive real part by the convergent integral $\Gamma(x) = \int_0^\infty e^{-t}\,t^{x-1}\,dt$. For complex numbers with negative real part, we can extrapolate using the recurrence $\Gamma(x + 1) = x\,\Gamma(x)$. For example, it is known that $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$, so $\Gamma\left(-\frac{1}{2}\right) = -2\sqrt{\pi}$ and $\Gamma\left(-\frac{3}{2}\right) = \frac{4}{3}\sqrt{\pi}$.



A plot of $|\Gamma(z)|$ is shown above for complex values of z. Observe that for non-positive integers, the function is undefined.

From the integral definition of the Gamma function, it is straightforward to establish this identity:

- $\frac{1}{A^x} = \frac{1}{\Gamma(x)} \int_0^\infty e^{-At} t^{x-1}\, dt$. **[Identity involving the Gamma function]**

If we want to show that $\frac{a}{A^x} + \frac{b}{B^x} + \frac{c}{C^x} \geq 0$, we can convert it to the equivalent inequality $\frac{1}{\Gamma(x)} \int_0^\infty \left(a\, e^{-At} + b\, e^{-Bt} + c\, e^{-Ct}\right) t^{x-1}\, dt \geq 0$. If x is positive and $a\, e^{-At} + b\, e^{-Bt} + c\, e^{-Ct} \geq 0$, the integrand and integral are therefore also non-negative.

19. Prove that $\sum_{i=1}^{n} \sum_{j=1}^{n} \frac{a_i a_j}{(p_i + p_j)^c} \geq 0$, where $c, p_1, p_2, \ldots, p_n > 0$ and $a_1, a_2, \ldots, a_n \in \mathbb{R}$. **[KöMaL, Problem A493, November 2009]**

An interesting fact concerning the Gamma function is that the volume of a n-dimensional hypersphere of radius r is given by $\frac{\pi^{\frac{n}{2}} r^n}{\Gamma\left(\frac{n}{2}+1\right)}$. One can verify easily that this agrees with known formulae for the line segment, circle and sphere.

20. The E_8 lattice consists of points in $\mathbb{R}^8$ such that the coordinates are either all integers or all half-integers, and the sum of the coordinates is an even integer. Suppose we place (hyper)spheres of radius r, centred at each point in E_8. What is the maximum value of r such that the spheres are disjoint, and what is the density of the resulting sphere packing?

Solutions

1. This follows from the non-negativity of $(x-y)^2+(y-z)^2+(z-x)^2$.

2. By induction on the number of variables, the barycentre $B=(\alpha_1 x_1+\ldots+\alpha_n x_n,\ \alpha_1 f(x_1)+\ldots+\alpha_n f(x_n))$ must lie in the convex hull of the points $\{P_i=(x_i,\ f(x_i))\}$. As every point on the perimeter of the convex hull lies above the curve by the definition of convexity, so too must every point in the interior of the convex hull, including the barycentre.

3. This is the special case of Jensen's inequality where $f(x)=e^x$.

4. We can replace a and b with $\frac{a}{b}$ and 1, respectively, without altering anything, and thus assume without loss of generality that $b=1$. Applying Jensen's inequality to $f(x)=x^a$ gives the desired result.

5. (3, 0, 0) majorises (2, 1, 0), so this follows from Muirhead's inequality.

6. Without loss of generality, re-define the interval so that $\alpha_1=0$, $\alpha_2=1$. We then need to prove that $f(x_2)+f(x_1)\geq f(\beta_1 x_1+\beta_2 x_2)+f(\beta_1 x_2+\beta_2 x_1)$. By the definition of convexity, we have $f(\beta_1 x_1+\beta_2 x_2)\leq \beta_1 f(x_1)+\beta_2 f(x_2)$ and $f(\beta_1 x_2+\beta_2 x_1)\leq \beta_1 f(x_2)+\beta_2 f(x_1)$. Adding these together yields the desired inequality.

7. Suppose we apply a q-move to α_i and α_{i+1}. Consider each sum of the form $f(\alpha_1 x_{\sigma(1)}+\ldots+\alpha_i x_{\sigma(i)}+\alpha_{i+1} x_{\sigma(i+1)}+\ldots+\alpha_n x_{\sigma(n)})+$. Note that this is equal to
$$f(\alpha_1 x_{\sigma(1)}+\ldots+\alpha_i x_{\sigma(i+1)}+\alpha_{i+1} x_{\sigma(i)}+\ldots+\alpha_n x_{\sigma(n)})$$
$g(\alpha_i x_{\sigma(i)}+\alpha_{i+1} x_{\sigma(i+1)})+g(\alpha_i x_{\sigma(i+1)}+\alpha_{i+1} x_{\sigma(i)})$ for some convex function $g(w)=f(w+k)$. The previous theorem tells us that this cannot increase when α_i and α_{i+1} are replaced with β_i and β_{i+1}. Apply this principle to all $\frac{n!}{2}$ pairs of terms.

8. This is a corollary of the energy minimisation lemma and the previous question.

9. Let $a-b=d$ and $b-c=e$. Then the inequality becomes $x^2 d(d+e)-y^2 d e+z^2 e(d+e)\geq 0$. Rearranging, we obtain the equivalent $(x d-z e)^2+\left((x+z)^2-y^2\right) d e\geq 0$. This is clearly true if $(x+z)^2\geq y^2$.

10. $x^2+z^2\geq y^2$ is a weaker condition than $x+z\geq y$, so the result follows from the previous question.

11. Assume without loss of generality that $a\geq b\geq c$, and let $d=a-b$ and $e=b-c$. For any non-negative monotonic function f, we have $f(a)+f(c)\geq f(b)$; hence, this must be true of any sum of non-negative monotonic functions. The problem reduces to showing that $f(a)\, g(h(d)\, h(d+e))+f(c)\, g(h(e)\, h(d+e))\geq f(b)\, g(h(e)\, h(d))$. As g and h are increasing, we have $f(a)\, g(h(d)\, h(d+e))+f(c)\, g(h(e)\, h(d+e))\geq (f(a)+f(c))\, g(h(e)\, h(d))$, which in turn must be greater than $f(b)\, g(h(e)\, h(d))$, as $f(a)+f(c)\geq f(b)$.

12. Obviously, the worst-case scenario is when all variables are positive. Expanding the Schur inequality $x^2\left(x^2-y^2\right)\left(y^2-z^2\right)+y^2\left(y^2-z^2\right)\left(y^2-x^2\right)+z^2\left(z^2-x^2\right)\left(z^2-y^2\right)\geq 0$ gives the variant $x^6+y^6+z^6+3x^2y^2z^2\geq x^4y^2+x^2y^4+y^4z^2+y^2z^4+z^4x^2+z^2x^4$. The other inequality, $x^4y^2+x^2y^4+y^4z^2+y^2z^4+z^4x^2+z^2x^4\geq x^3y^3+y^3z^3+z^3x^3$, is a simple application of Muirhead.

13. Again, assume without loss of generality that $a\geq b\geq c$, and let $d=a-b$ and $e=b-c$. As h is positive-illuminable, $h(m+n)\geq h(m)+h(n)$ for all $m, n\in\mathbb{R}^+$. As g and h are increasing and odd, we have $f(a)^2\, g(h(d)\, h(d+e))+f(c)^2\, g(h(e)\, h(d+e))\geq f(a)^2\, g\left(h(d)^2+h(d)\, h(e)\right)+f(c)^2\, g\left(h(e)^2+h(d)\, h(e)\right)$. Hence, we can reduce this, effectively, to the case where h is the identity function. As g is positive-

illuminable, we can express $g(w) = G(w)^2\, w$ for all $w \in \mathbb{R}^+$, where G is an increasing function. Now, we let $x = f(a)\, G(h(a-b)\, h(a-c))$ and define y and z similarly. We have $x + z \geq y$ by the same argument as in the proof of the weak generalised Schur inequality. The result then follows from the strong 6-variable Schur.

14. We let $f(x) = e^{2x}$. Differentiating this twice gives $4\, e^{2x}$, which is positive-definite. Hence, f is convex and we can apply Jensen's inequality to show that $\frac{1}{2}\,(f(x) + f(y)) \geq f\left(\frac{x+y}{2}\right)$.

15. Let $w = f(x, y, z) = 4\,(x+y+z)^3 - 27\left(x^2\, y + y^2\, z + z^2\, x\right)$. Differentiate with respect to x to give the partial derivative $\frac{\partial w}{\partial x} = 12\,(x+y+z)^2 - 27\left(2\,x\,y + z^2\right)$. The cyclic sum is $\frac{\partial w}{\partial x} + \frac{\partial w}{\partial y} + \frac{\partial w}{\partial z} = 36\,(x+y+z)^2 - 27\left(x^2 + y^2 + z^2 + 2\,x\,y + 2\,y\,z + 2\,z\,x\right) = 9\,(x+y+z)^2$. This is obviously positive, so the function increases as we move parallel to the vector $(1, 1, 1)$. Hence, we have $f(x+h, y+h, z+h) > f(x, y, z)$ for all $h > 0$. Assume without loss of generality that $x \leq y$ and $x \leq z$. Using the previous statement, $f(x, y, z) > f(0, y-x, y-z)$. To prove the strict inequality in general, therefore, we need only prove the weak inequality when one of the variables is zero. We have reduced the problem to showing that $4\,(y+z)^3 \geq 27\, y^2\, z$. This is evident from the factorisation $4\,(y+z)^3 - 27\, y^2\, z = (4\,y + z)\,(y - 2\,z)^2$.

16. We wish to minimise $x^2 + y^2$ subject to the constraint $x\,y + x^2 = 1$. We use Lagrange multipliers to obtain $\Lambda = x^2 + y^2 - \lambda\left(x\,y + x^2 - 1\right)$. We equate each of its partial derivatives, $2\,x - \lambda\, y - 2\,\lambda\, x$ and $2\,y - \lambda\, x$, to zero. The latter gives us the value of λ, namely $\frac{2\,y}{x}$, so we can substitute it into the other equation and obtain $2\,x - 2\,\frac{y^2}{x} - 4\,y = 0$. We can multiply throughout by $\frac{1}{2}\,x$ to give the quadratic $x^2 - y^2 - 2\,x\,y = 0$, or $\left(\frac{x}{y}\right)^2 - 2\left(\frac{x}{y}\right) - 1 = 0$. The Babylonian formula gives us $\frac{x}{y} = 1 \pm \sqrt{2}$. It is sensible to draw a graph of the hyperbola to confirm that the root we are looking for is actually $\frac{x}{y} = 1 + \sqrt{2}$. Hence, $x = \left(1 + \sqrt{2}\right) y$ and $y = \left(\sqrt{2} - 1\right) x$. Substituting this into the equation of the hyperbola gives $x^2 = \frac{1}{\sqrt{2}}$. Similarly, we have $y^2 = \left(\sqrt{2} - 1\right)^2 x^2 = \left(3 - 2\sqrt{2}\right) x^2$, giving $x^2 + y^2 = \frac{\left(4 - 2\sqrt{2}\right)}{\sqrt{2}} = 2\sqrt{2} - 2$. The distance is the square-root of that, namely $\sqrt{2\sqrt{2} - 2}$.

17. The inequality $\left(x^2 + y^2 + z^2\right)^3 \geq \left(x^3 + y^3 + z^3 - 3\,x\,y\,z\right)^2$ can be expressed in the $u\,v\,w$ notation as $\left(9\,u^2 - 6\,v^2\right)^3 \geq \left((3\,u)\left(9\,u^2 - 9\,v^2\right)\right)^2$. We can divide throughout by 3^6, giving the equivalent inequality $\left(u^2 - \frac{2}{3}\,v^2\right)^3 \geq \left(u^3 - v^2\,u\right)^2$, which expands to $u^6 - 2\,u^4\,v^2 + \frac{4}{3}\,u^2\,v^4 - \frac{8}{27}\,v^6 \geq u^6 - 2\,u^4\,v^2 + u^2\,v^4$. Observe that the first two terms on each side of the equation cancel, so we wish to prove that $\frac{4}{3}\,u^2\,v^4 - \frac{8}{27}\,v^6 \geq u^2\,v^4$, or $\frac{1}{3}\,u^2\,v^4 \geq \frac{8}{27}\,v^6$. We can neatly divide by $\frac{1}{27}\,v^4$ (which must be positive, since $9\,v^4 = (x\,y + y\,z + z\,x)^2$), giving $9\,u^2 \geq 8\,v^2$. As u^2 is necessarily positive and greater in magnitude than v^2, this is true. Equality occurs when $v^4 = 0$, *i.e.* $x\,y + y\,z + z\,x = 0$.

18. Firstly, obtain the partial derivative $\frac{\partial \Lambda}{\partial x} = y\,z + 2\,\lambda\,x + \mu = 0$. We can multiply throughout by x to get $2\,\lambda\,x^2 + \mu\,x - w = 0$, where $w = x\,y\,z$. As x, y, z are all roots of this quadratic equation, which only has two roots, at least two must be identical.

19. We note that this is equivalent to $\frac{1}{\Gamma(c)} \int_0^\infty t^{c-1} \sum_{i=1}^n \sum_{j=1}^n a_i\, a_j\, e^{-(p_i + p_j)\,t}\, dt \geq 0$. Observe that this simplifies to $\frac{1}{\Gamma(c)} \int_0^\infty t^{c-1} \left(\sum_{i=1}^n a_i\, a_j\, e^{-p_i\,t}\right)^2 dt \geq 0$, which is necessarily true.

20. The points in the lattice (regarded as vectors) clearly form a group under addition, so we need only calculate the minimum distance from the zero vector to another vector $\underline{a}$. If the coordinates of $\underline{a}$ are all half-integers, the minimum norm (squared length) of $\underline{a}$ is $8\left(\frac{1}{2}\right)^2 = 2$. Similarly, the closest integer point in the lattice is $(1, 1, 0, 0, 0, 0, 0, 0)$, with a norm of 2. Hence, the maximum value of r is $\frac{1}{2}\sqrt{2}$, so the volume of each sphere is $\frac{\pi^4 r^8}{4!} = \frac{\pi^4}{384}$. We now need to determine the number of lattice points per unit volume. The points in $\mathbb{Z}^8$ and $\left(\mathbb{Z} + \frac{1}{2}\right)^8$ each have one point per unit volume, so $\mathbb{Z}^8 \cup \left(\mathbb{Z} + \frac{1}{2}\right)^8$ has two points per unit volume. E_8 comprises half of these points (those with an even sum of coordinates), so it has one point per unit volume. Hence, the sphere packing has a density of $\frac{\pi^4}{384}$.

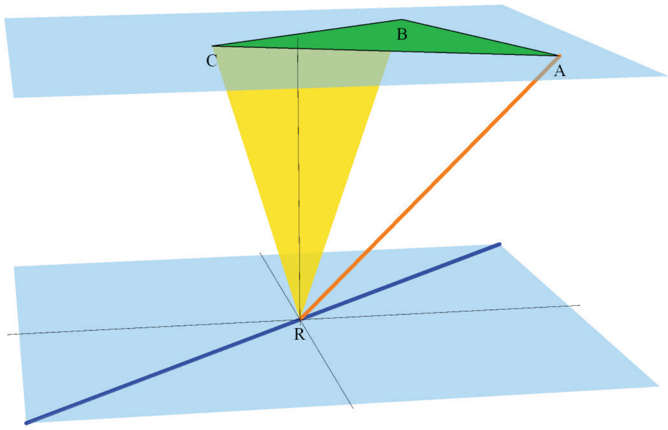
Projective geometry

Projective geometry is an extension of Euclidean geometry, endowed with many nice properties incurred by affixing an extra 'line at infinity'. Certain theorems (such as Desargues' and Pascal's theorems) have projective geometry as their more natural setting, and the wealth of projective transformations can simplify problems in ordinary Euclidean geometry.

The real projective plane

In Euclidean geometry, we assign a coordinate pair (x, y) to each point in the plane. In projective geometry, we augment this with an extra coordinate, so three values are used to represent a point: (x, y, z). Moreover, scalar multiples are considered equivalent; (x, y, z) and $(\lambda x, \lambda y, \lambda z)$ represent the same point. $R = (0, 0, 0)$ is not part of the projective plane, but can be regarded as a 'projector', from which all points, lines, circles *et cetera* emanate.

Since scalar multiples of points are considered equivalent, we can identify points in $\mathbb{RP}^2$ (the real projective plane) with lines through the origin. Projective lines are identified with planes through the origin, and are of the form $a x + b y + c z = 0$. Note that this equation is *homogeneous*: all terms are of first degree. In general, all algebraic curves are represented by homogeneous polynomials in x, y and z.



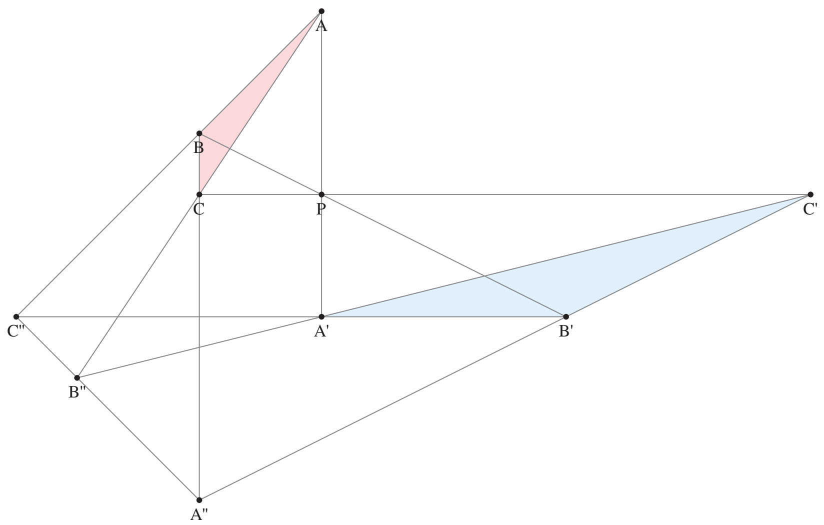
In the diagram above, a triangle $A B C$ is shown in the *reference plane* $(z = 1)$. The point A is 'projected' from R along the orange line, intersecting the reference plane at A. Similarly, the 'plane' containing the yellow triangle represents the line $B C$. Horizontal 'lines' such as the blue one do not intersect the reference plane, so correspond to points 'at infinity'. The horizontal 'plane' (parallel to the reference plane) through R represents the line at infinity $(z = 0)$. Parallel lines on the projective plane can be considered to meet at a point on the line at infinity.

1. Prove that any two distinct lines intersect in precisely one point.

2. Show that the equation of a line through points (x_1, y_1, z_1) and (x_2, y_2, z_2) is given by $\det\begin{pmatrix} x & y & z \\ x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \end{pmatrix} = 0$.

You have shown that any two points share a common line, and any two lines share a common point. This suggests a fundamental interchangeability between lines and points, known as *projective duality*. We shall explore this later.

3. Let $A B C$ and $A' B' C'$ be two triangles. Let $A B$ meet $A' B'$ at C'', and define A'' and B'' similarly. Show that $A A'$, $B B'$ and $C C'$ are concurrent if and only if A'', B'' and C'' are collinear. **[Desargues' theorem]**



If two triangles exhibit this relationship, they are said to be *in perspective*. The point P is the *perspector*, and the line $A'' B'' C''$ is the *perspectrix*. As Desargues' theorem is the projective dual of its converse, you only need to prove the statement in one direction. Surprisingly, it is actually easier to prove Desargues in three dimensions (where the triangles are in different planes); the two-dimensional result then follows by projecting it onto the plane.

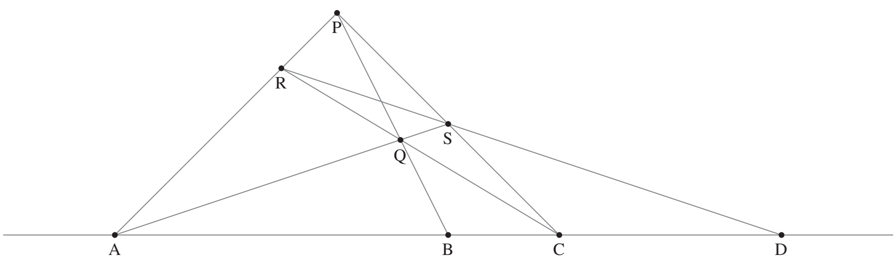
4. Let ABC be a triangle and X be a point inside the triangle. The lines AX, BX and CX meet the circle ABC again at P, Q and R, respectively. Choose a point U on XP which is between X and P. Suppose that the lines through U which are parallel to AB and CA meet XQ and XR at points V and W respectively. Prove that the points R, W, V and Q lie on a circle. **[BMO2 2011, Question 1]**

Cross-ratio and harmonic ranges

You may have encountered homogeneous projective coordinates before in the form of areal (or barycentric) coordinates. Even though qualitative properties such as collinearity and concurrency can be defined in terms of any projective coordinates, to compare distances they must first be *normalised*, *i.e.* projected onto the reference plane.

Firstly, we define a vector, $\underline{n}$, perpendicular to the reference plane. (For Cartesian coordinates, we generally allow $z = 1$ to be the reference plane, so $\underline{n} = (0, 0, 1)$. For areal coordinates, $x + y + z = 1$ is the reference plane, so $\underline{n} = (1, 1, 1)$. In general, if the reference plane is given by $a x + b y + c z = 1$, the vector $\underline{n} = (a, b, c)$. For a given vector $\underline{x}$ in the projective plane, it is normalised by the operation $\underline{x} \to \frac{\underline{x}}{\underline{n} \cdot \underline{x}}$

5. If A, B, C and D are collinear points represented by normalised vectors $\underline{a}$, $\underline{b}$, $\underline{c}$ and $\underline{d}$, respectively, show that $\frac{\overrightarrow{AB}}{\overrightarrow{CD}} = \frac{\underline{a} - \underline{b}}{\underline{c} - \underline{d}}$.



Suppose we have three points, A, B and C, which are collinear. We can select an arbitrary point P, and a further arbitrary point Q lying on the line PB. Let QC meet PA at R, and QA meet PC at S. Finally, let RS meet line

ABC at D.

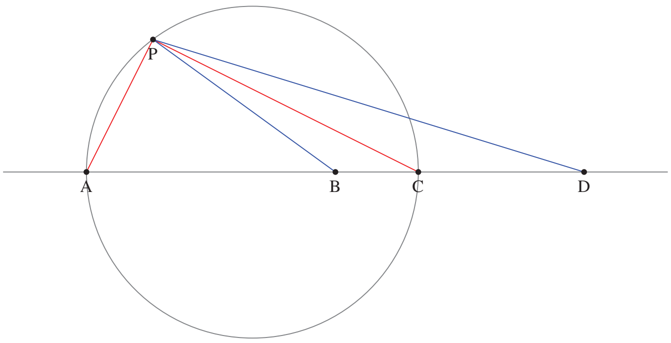
6. Prove that $\frac{\overrightarrow{AB} \cdot \overrightarrow{CD}}{\overrightarrow{BC} \cdot \overrightarrow{DA}} = -1$, irrespectively of the locations of P and Q. *(Hint: this can be done by applying two similar theorems in quick succession.)*

It is a remarkable fact that the location of D does not depend on that of P and Q. B and D are described as *projective harmonic conjugates* with respect to the line segment AC. For any four collinear points, the quantity $(A, C; B, D) = \frac{\overrightarrow{AB} \cdot \overrightarrow{CD}}{\overrightarrow{BC} \cdot \overrightarrow{DA}}$ is known as the *cross-ratio*. If $(A, C; B, D) = -1$, we say that they form a *harmonic range*.

■ $(A, C; B, D) = -1$ is equivalent to A and C being inverse points with respect to the circle on diameter BD, which is in turn equivalent to the circles on diameters AC and BD being orthogonal. **[Equivalent definitions of harmonic range]**

There is another remarkable and useful fact concerning harmonic ranges. Let P be a point not on the line $ABCD$. Then any two of the following four properties implies the other two:

- $\angle APC = \frac{\pi}{2}$;
- PA is an angle bisector of $\angle BPD$;
- PC is the other angle bisector of $\angle BPD$;
- $(A, C; B, D) = -1$.



By Thales' theorem, the first of these conditions is equivalent to P lying on the circle with diameter AC. Hence, if P lies on either of the intersection points of the (orthogonal) circles on diameters AC and BD, the four lines through P divide the plane into eight equal octants of angle $\frac{\pi}{4}$.

Projective transformations

Returning to the idea of representing points in the projective plane as three-element vectors, we can consider the group of operations represented by linear maps $\underline{x} \to M\,\underline{x}$, where M is a non-singular matrix. These are known as *projective transformations*, or *collineations*.

7. Show that applying a projective transformation to a line results in another line, and thus that collinear points remain collinear.

A generalisation is that degree-d algebraic curves are mapped to degree-d algebraic curves by projective transformations. Hence, conics are preserved. Moreover, it is possible to choose a projective transformation carefully to map a given conic to any other conic, with three real degrees of freedom remaining.

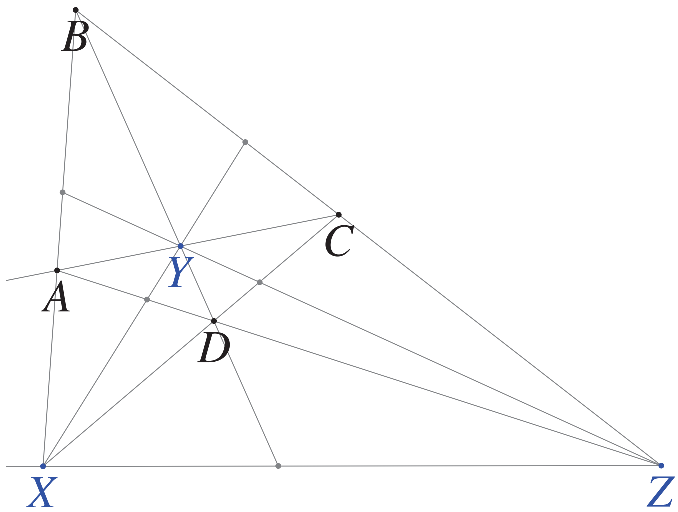
8. Show that applying a projective transformation with matrix M to a point $x = p + \lambda\, q$ results in a point with normalised coordinates $\frac{M\,\underline{p} + \lambda\, M\,\underline{q}}{n \cdot M\,\underline{p} + n \cdot \lambda\, M\,\underline{q}}$.

9. Hence show that the cross-ratio (A, C; B, D) of four collinear points is preserved under projective transformations.

10. For any four points A, B, C and D, no three of which are collinear, show that there exists a unique projective transformation mapping them to (1, 0, 0), (0, 1, 0), (0, 0, 1) and (1, 1, 1), respectively.

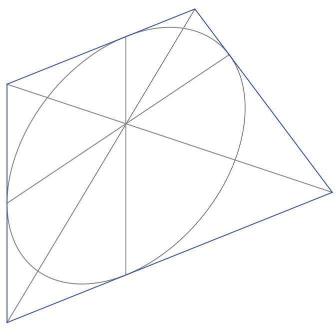
11. Hence show that there exists a unique projective transformation mapping any four points (no three of which are collinear) to any other four points (no three of which are collinear).

The last of these theorems enables one to simplify a projective problem by converting any quadrilateral into a parallelogram. This enables one to find all of the harmonic ranges in the *complete quadrangle* displayed below. Try to spot as many as you can!

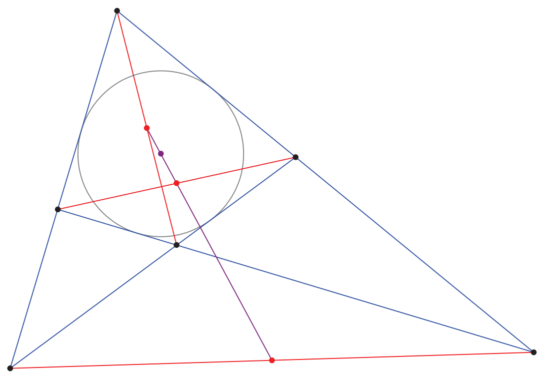


12. The diagonals of the quadrilateral $ABCD$ meet at X. The circumcircles of ABX and CDX intersect again at Y; the circumcircles of BCX and DAX intersect again at Z. The midpoints of the diagonals AC and BD are denoted M and N, respectively. Prove that M, N, X, Y and Z are concyclic. **[Sherry Gong, Trinity 2012]**

Another configuration occurring in many instances is a quadrilateral with an inscribed conic. With a projective transformation, we can convert the quadrilateral into a parallelogram. To simplify matters even further, a (possibly imaginary) affine transformation is capable of turning the conic into a circle. The symmetry of the configuration implies that the diagonals of the quadrilateral are concurrent with the lines joining opposite points of tangency.

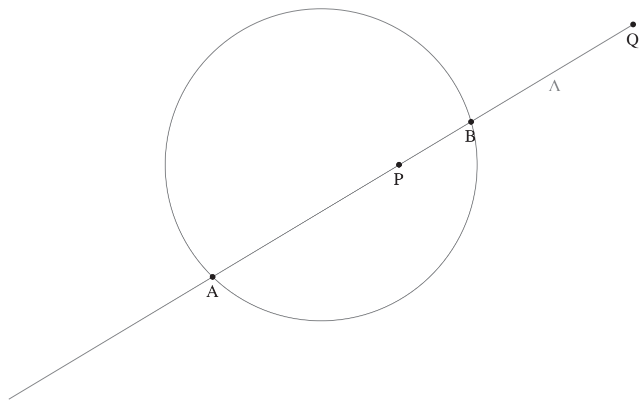


When viewed as a *complete quadrilateral* (four lines in general position intersecting in six points), we have three of these concurrency points and plenty of harmonic ranges! The centre of the inscribed conic also happens to lie on a line passing through the three midpoints of pairs opposite vertices, by *Newton's theorem*. This line is thus the locus of the centres of all possible inscribed conics.



Polar reciprocation and conics

One convenient way to demonstrate projective duality is by creating a bijection between lines and points, such that collinear points map to concurrent lines. There are several ways in which this can be defined, but one of the most elegant is polar reciprocation:



13. Consider a point P in the unit circle. Draw a line Λ through P, intersecting the unit circle at A and B. Let Q be the projective harmonic conjugate of P with respect to AB. Show that the locus of Q as Λ varies is a straight line.

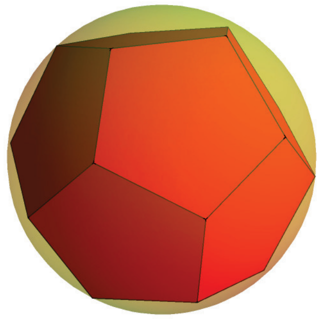
This locus is known as the *polar* of P, and P is its *pole*. Moreover, there is nothing special about the unit circle in projective geometry, so this construction generalises to any conic.

14. Let ABC be a scalene triangle, and let Γ be its nine-point circle. Γ intersects BC at points P and Q; the tangents from Γ at P and Q intersect at A'. Points B' and C' are defined similarly. Prove that the lines AA', BB' and CC' are concurrent. **[NST4 2011, Question 3]**

The easiest route to solving the above problem is to prove the much more general theorem that a triangle and its polar reciprocal triangle (with respect to any conic) are in perspective. This is (like many results in projective geometry!) known as *Chasles' theorem*.

Polar reciprocation generalises to three dimensions, where we can reciprocate about a sphere (or, more generally, quadric surface).

15. What is the polar reciprocal of a dodecahedron with respect to its circumsphere S?



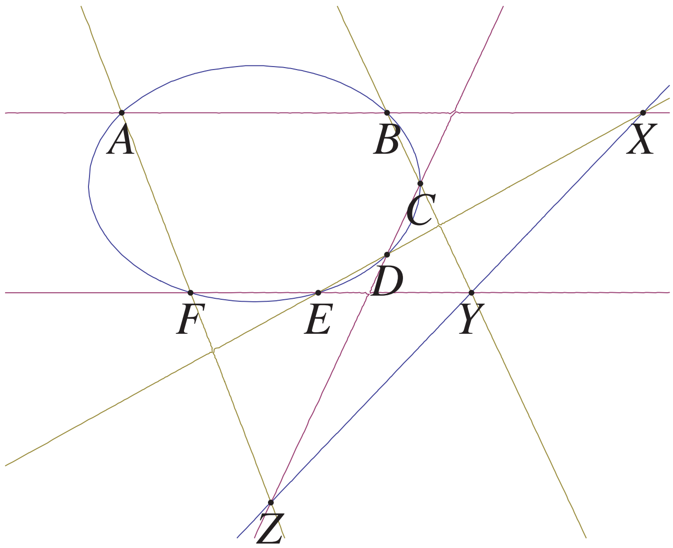
More generally, the dual of the regular polyhedron with Schläfli symbol $\{a, b\}$ has the Schläfli symbol $\{b, a\}$. For higher dimensions, we simply reflect the symbol. The tetrahedron, with the palindromic Schläfli symbol {3, 3}, is thus self-dual, as is the square tiling with Schläfli symbol {4, 4}. More generally, a simplex has Schläfli symbol {3, 3, 3, …, 3, 3} and a hypercubic tessellation has Schläfli symbol {4, 3, 3, …, 3, 4}, both of which are palindromic. For four-dimensional solids, the 4-simplex {3, 3, 3} is not the only self-dual regular polychoron; we also have the '24-cell' with Schläfli symbol {3, 4, 3} (meaning that three octahedral cells are clustered around each edge).

Circular points at infinity

In projective Cartesian coordinates, the equation of a circle is of the form $x^2 + y^2 + b\,x\,z + c\,y\,z + d\,z^2 = 0$. Note that the points $(1, i, 0)$ and $(i, 1, 0)$ satisfy this equation, where i is the imaginary unit. Hence, all circles can be considered to pass through two imaginary 'circular points' on the line at infinity. Indeed, a circle can be *defined* as any conic passing through both circular points. In the n-dimensional complex projective plane ($\mathbb{CP}^n$), we have a $(n-2)$-sphere on the line at infinity contained by all $(n-1)$-spheres.

Apart from the circular points, other imaginary points on the complex projective plane $\left(\mathbb{CP}^2\right)$ are scarcely useful, so there is no need to worry about them.

16. Let $A\,B\,C\,D\,E\,F$ be a hexagon, the vertices of which lie on a conic. Let $A\,B$ and $D\,E$ meet at X; $B\,C$ and $E\,F$ meet at Y; and $C\,D$ and $F\,A$ meet at Z. Prove that X, Y and Z are collinear. **[Pascal's theorem]**



17. Write down the projective dual of Pascal's Theorem. **[Brianchon's theorem]**

18. Let A, C and E lie on a straight line, and B, D and F lie on another straight line. Let AB and DE meet at X; BC and EF meet at Y; and CD and FA meet at Z. Prove that X, Y and Z are collinear. **[Pappus' theorem]**

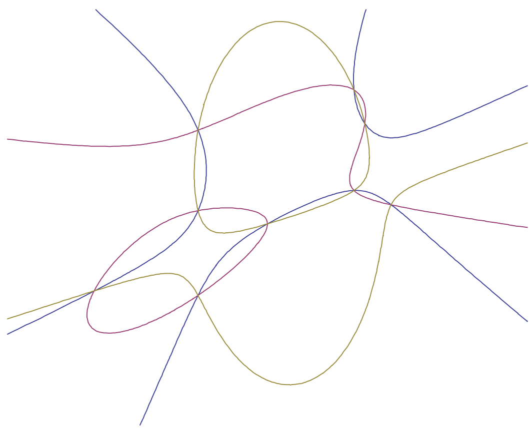
19. Let ABC be a triangle, and let the tangent to its circumcircle at A meet BC at D. Let l be a line meeting AD internally at P, the circumcircle at Q and T, the sides AB and AC internally at R and S respectively, and BC at U. Suppose that $PQRSTU$ lie in that order on l. Show that if $QR = ST$ then $PQ = UT$. **[UK IMO Squad Practice Exam 2011, Question 2]**

20. Suppose Γ_1 and Γ_2 are two disjoint ellipses, with Γ_1 inside Γ_2. If there is at least one triangle with its sides tangent to Γ_1 and vertices on Γ_2, show that there are infinitely many. **[Poncelet's porism]**

Actually, this theorem generalises to polygons with any number of sides. However, it is very difficult to prove with elementary methods.

- Suppose Γ_1 and Γ_2 are two disjoint ellipses, with Γ_1 inside Γ_2. If there is at least one n-gon with its sides tangent to Γ_1 and vertices on Γ_2, then there are infinitely many. **[Poncelet's porism]**

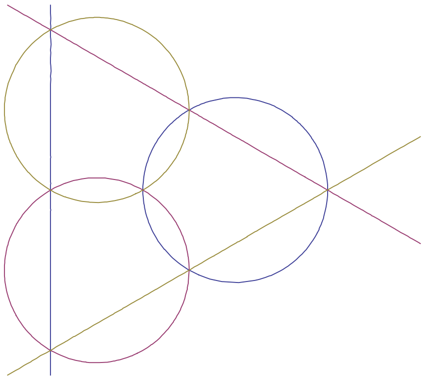
You may have noticed that Pappus' theorem is a special case of Pascal's theorem (and indeed Brianchon's theorem, as it is self-dual) when the conic degenerates into a pair of straight lines. Both of these theorems can be considered to be special cases of the Cayley-Bacharach theorem.



- Three cubic curves each pass through the same eight points, no four of which are collinear and no seven of which are conconic. The three cubics then share a ninth point. **[Degree-3 Cayley-Bacharach theorem]**

21. If two circles intersect, the *radical axis* is the line passing through both intersection points. For three mutually intersecting circles, prove that the three radical axes are concurrent. **[Radical axis theorem]**

22. Three circles MNP, NLP and LMP have a common point P. A point A is chosen on circle MNP (other than M, N or P). AN meets circle NLP at B and AM meets circle LMP at C. Prove that BC passes through L. **[UK MOG 2011, Question 1]**



Cayley-Bacharach generalises to curves of arbitrary degree, but the generalised version is difficult to prove. The quartic version is given below.

- Three quartic curves pass through the same thirteen points, no five of which are collinear, no nine of which are conconic and no twelve of which are concubic. The three quartics then share a further three points. **[Degree-4 Cayley-Bacharach theorem]**

23. Let the eight vertices of an octagon lie on a conic, and alternately colour the edges red and blue. Prove that the remaining eight heterochromatic intersections of (the extensions of) the edges lie on another conic. **[Generalised Pascal's theorem]**

24. Two cyclic quadrilaterals, $ABCD$ and $A'B'C'D'$, share the same circumcircle. The four intersections of corresponding edges (e.g. AB with $A'B'$) are labelled P, Q, R and S. Show that if P, Q and R are collinear, then S also lies on this line. **[Two butterflies theorem]**

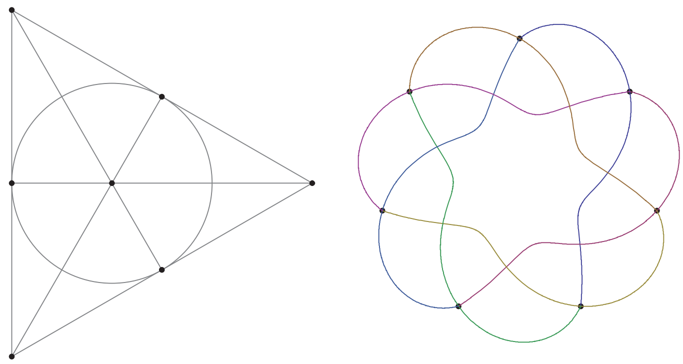
25. Let PQ be the chord of a circle Γ, and let M be the midpoint of PQ. Chords AB and CD are drawn through M. Let AC and BD intersect PQ at R and S, respectively. Prove that M is the midpoint of RS. **[Butterfly theorem]**

26. Let $ABCD$ be a quadrilateral. AC and BD intersect at E. X and Y are two points in the plane, and the line XY intersects AC at F and BD at G. R is the harmonic conjugate of F with respect to AC; S is the harmonic conjugate of G with respect to BD. The conics $ABEXY$ and $CDEXY$ intersect a fourth time at P; the conics $BCEXY$ and $DAEXY$ intersect a fourth time at Q. Prove that $PQRSEXY$ are conconic. **[Sam Cappleman-Lynes, 2012]**

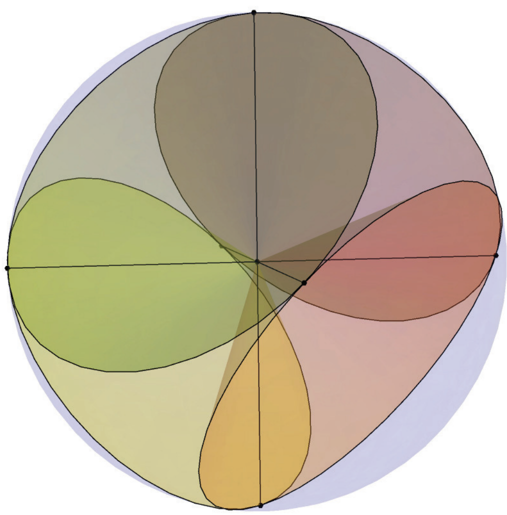
27. A number of line segments ($l_1, l_2, \cdots, l_n$) are drawn in general position on the plane, such that every pair of line segments intersects. A line Λ cuts all of the line segments. For each l_i, the endpoint on the left of Λ is called A_i, and the other endpoint is called B_i. An ant walks along a line segment l_i in the direction $A_i \to B_i$. Whenever it hits B_i, it teleports to A_i. Whenever it meets an intersection point ($l_i \cap l_j$), it moves onto the other line segment l_j and continues moving (in the direction $A_j \to B_j$, still). Prove that there exists an initial position of the ant such that it visits every line segment infinitely often.

Finite projective planes

The applications of projective geometry in olympiad problems involve infinite projective planes, namely $\mathbb{RP}^2$ and $\mathbb{CP}^2$. The construction where we take three coordinates (x, y, z) and consider scalar multiples to be equivalent generalises, enabling one to define a projective geometry over any field. For instance, the finite field of order 2 results in a projective plane with seven points: (0, 0, 1), (0, 1, 0), (0, 1, 1), (1, 0, 0), (1, 0, 1), (1, 1, 0) and (1, 1, 1). This finite projective plane, considered to be the simplest non-trivial geometry, is called the *Fano plane*:



The seven lines are shown (in the left diagram) as six straight lines and one circle; it is impossible to embed it in the real projective plane using only straight lines. Although not obvious from the left diagram, the right diagram demonstrates that all points are equivalent. There are, in fact, no fewer than 168 symmetries, corresponding to the rotation group PSL(2, 7) of Klein's quartic. Fixing a single vertex reduces the number to $\frac{168}{7} = 24$ symmetries, which are apparent in the following embedding in three-space. One point is at the centre of an octahedron formed by the other six points. The three orthogonal axes of the octahedron, together with the circumcircles of alternate faces, form the seven lines of the Fano plane.



Coincidentally, this resembles an embedding (the *Roman surface*) of the real projective plane into $\mathbb{R}^3$.

28. The seven vertices of the Fano plane are each coloured with one of c colours. How many different colourings are possible, taking into account the symmetries?

Solutions

1. Distinct lines correspond to planes through R. They must intersect in a line through R, which corresponds to a point on the projective plane.

2. The volume of the tetrahedron $RABC$, with coordinates $(0, 0, 0)$, (x_1, y_1, z_1), (x_2, y_2, z_2) and (x_3, y_3, z_3), respectively, is given by $\frac{1}{6}\det\begin{pmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{pmatrix}$. If this is zero, the points R, A, B and C must be coplanar. As we can assume without loss of generality that A, B and C lie on the reference plane, and R does not, we can deduce that A, B and C must be collinear. (The converse is obviously also true.) This condition that three points are collinear can be converted into the equation of a line, by allowing C to be a variable point.

3. Assume P exists, and consider the case where the two triangles lie in different planes. The two planes must then intersect on a line, where A'', B'' and C'' must obviously lie. The result for two dimensions is obtained by projecting it onto the plane. Note that the converse is the projective dual, so proving it in one direction is sufficient.

4. Applying Desargues' theorem results in the revelation that WV is parallel to BC. Hence, WXV is similar (indeed, homothetic) to CXB, so the result follows by applying the converse of the intersecting chords theorem to point X.

5. Consider the reference plane and use the Pythagorean distance formula.

6. Applying Ceva's theorem and Menelaus' theorem result in the two equations: $\frac{\overrightarrow{AB}}{\overrightarrow{BC}} \cdot \frac{\overrightarrow{CS}}{\overrightarrow{SP}} \cdot \frac{\overrightarrow{PR}}{\overrightarrow{RA}} = 1$ and $\frac{\overrightarrow{DA}}{\overrightarrow{CD}} \cdot \frac{\overrightarrow{CS}}{\overrightarrow{SP}} \cdot \frac{\overrightarrow{PR}}{\overrightarrow{RA}} = -1$. Dividing the two equations yields the desired result.

7. A projective transformation of the projective plane $\mathbb{P}^2$ is a linear transformation of the Euclidean space $\mathbb{R}^3$ in which it can be considered to reside. As planes are preserved by linear transformations, lines are preserved by projective transformations.

8. Firstly, we have $M\,\underline{x} = M\left(\underline{p} + \lambda\,\underline{q}\right) = M\,\underline{p} + \lambda\,M\,\underline{q}$. However, this must be normalised in the obvious way, yielding $\frac{M\,\underline{p}+\lambda\,M\,\underline{q}}{n\cdot\left(M\,\underline{p}+\lambda\,M\,\underline{q}\right)}$. The dot product is distributive over addition.

9. Consider the four points $\underline{a} = \underline{p} + \alpha\,\underline{q}$, $\underline{b} = \underline{p} + \beta\,\underline{q}$, $\underline{c} = \underline{p} + \gamma\,\underline{q}$ and $\underline{d} = \underline{p} + \delta\,\underline{q}$. Calculating the cross-ratio before and after the projective transformation will result in the same value of $\frac{(\alpha-\beta)(\gamma-\delta)}{(\beta-\gamma)(\delta-\alpha)}$.

10. Instead, we will find the inverse matrix. As we want the three unit vectors to be mapped to a, b and c, we write the (unnormalised) coordinates of each of them in each column of the matrix. To map $(1, 1, 1)$ to $\underline{d}$, we need to multiply each of the three columns by nonzero real constants, which are the solutions to three simultaneous linear equations. For obvious reasons, there is a single such solution.

11. Firstly, transform the four original points to $(1, 0, 0)$, $(0, 1, 0)$, $(0, 0, 1)$ and $(1, 1, 1)$. Now, we can transform these to the four final points. Multiplying the two matrices together results in a single transformation.

12. Inverting about X results in a projective linear configuration, where we have to show that the intersections of opposite sides of quadrilateral $ABCD$ are collinear with the projective harmonic conjugates of P with respect to each of the diagonals. We can projectively transform $ABCD$ to the vertices of a square, rendering the problem trivial.

13. Form two points Q_1 and Q_2 in this manner. Apply a projective transformation taking them to the line at infinity, so P is the midpoint of two chords cutting an ellipse. Applying an affine transformation makes this ellipse into a circle, and P lies on two diameters so must be the centre. Obviously, the polar of P is the line at infinity. Applying the inverse projective transformations to this configuration will result in the original configuration, and the polar will remain a straight line.

14. Instead, we will prove Chasles' theorem (that a triangle and its polar reciprocal are in perspective) here, as the original problem is a special case. Due to Desargues' theorem, we can prove that the corresponding sides of the two triangles intersect at collinear points. Without loss of generality, we can assume two of these points are at infinity, and our objective is to show that the third also lies at infinity. We can then apply a projective transformation to take the two intersections of the conic with the line at infinity (which exist, due to Bezout's theorem) to the circular points at infinity (thus preserving the line at infinity, so this is an affine transformation), so we can assume that the conic is a circle. In this case, we have that the polar of A is parallel to BC, thus OA is perpendicular to BC. The polar of B is parallel to AC, thus OB is perpendicular to AC. Hence, O is the orthocentre of ABC, so OC is perpendicular to AB. This means that the polar of C is parallel to AB, so the triangles are in perspective.

15. The polar reciprocal (or 'dual') is a regular icosahedron with insphere S.

16. By projective transformations, we can assume the conic is a circle and the lines AB and DE are parallel, as are the lines AF and DC. From this, we draw the lines l_1 and l_2 (through O, the centre of the circle) such that B and E are reflections of A and D, respectively, in l_1, and that F and C are the reflections of A and D, respectively, in l_2. Hence, BE is a rotated copy of FC, so they are congruent. This means that BC and FE must indeed be parallel.

17. If a hexagon is circumscribed around a conic, its three main diagonals are concurrent.

18. Consider the degenerate case of Pascal's theorem where the conic is two lines.

19. We can reflect about the perpendicular bisector of RS (also the perpendicular bisector of QT) to transform this into a more projective problem. We will indicated a reflected version of point A with A', et cetera. We want to prove that the tangent at A meets the line $B'C'$ somewhere on the line l. To do this, apply Pascal's theorem to the degenerate hexagon $AAA'C'B'B$. As AA' and BB' intersect at the point at infinity on l, and AB and $A'C'$ intersect at R (also on l), we have that the tangent at A indeed meets $B'C'$ on l. By definition, this must be at point P and point U', so P is the reflection of U in the perpendicular bisector of RS. The problem becomes trivial.

20. In general, the ellipses must intersect four times on the complex projective plane; we can transform two of those points to the circular points at infinity, resulting in two circles. The problem then becomes equivalent to showing that infinitely many triangles share the same incircle (or excircle) and circumcircle. This is a consequence of Euler's formula, $OI^2 = R^2 - 2Rr$, and dimension counting.

21. We can consider the union of one circle and the radical axis of the other two to be a cubic. Repeating for the other two circles yields three cubics which intersect in at least eight points (the two circular points, plus the six pairwise circle intersections) so must intersect in the ninth. By applying Bezout's theorem, this additional intersection point cannot lie on any of the circles, so must lie on all three radical axes.

22. Same argument as in the previous question, but applied to the diagram displayed near the question.

23. Let Q_1 be the union of red lines, Q_2 be the union of blue lines, and Q_3 be the union of the main conic with the conic passing through five of the other eight heterochromatic intersections. By the quartic version of Cayley-Bacharach, Q_3 must pass through the other three intersections.

24. This is essentially the same argument as before, but with the realisation that a conic passing through three collinear points implies that it is a degenerate conic (union of two lines). Four of the eight 'other heterochromatic intersections' must lie on one line, and the other four lie on another line. The result then follows.

25. Reflect the 'butterfly' $A\,C\,B\,D$ in the perpendicular bisector of $P\,Q$ to create another butterfly, $A'\,C'\,B'\,D'$. Colour $A\,B$, $B'\,C'$, $C\,D$ and $D'\,A'$ red, and the remaining four edges blue. Since four of the heterochromatic intersections must be collinear (lying on the mirror line), so must the other four. Hence, $D'\,B'$ and $A\,C$ intersect at R, and the mirror image must be S. (This is a degenerate case of the 'two butterflies theorem'.)

26. Apply a projective transformation to send X and Y to the circular points at infinity. The problem is then reduced to question 12.

27. This is the projective dual of Geoff Smith's 'windmill problem' from IMO 2011. An official solution can be easily obtained from the Internet.

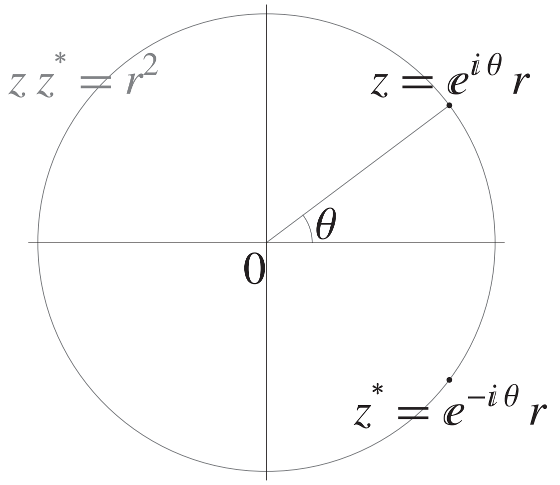
28. There is one identity permutation. If we choose to fix one vertex and consider the three-dimensional embedding, we have eight rotations which permute the remaining six vertices in two 3-cycles. As there were seven vertices to initially choose from, this results in 56 pairs of 3-cycles. We can also rotate by π about any of the six 'diagonal' axes, resulting in another 42 permutations, each comprising a 2-cycle and 4-cycle. A rotation by π about an orthogonal axes gives a pair of 2-cycles and fixes three collinear points; this gives 21 further permutations. There must be a 7-cycle due to the floral embedding of the Fano plane. When translated into the three-dimensional embedding, this becomes totally asymmetric, so we have 24 7-cycles in this conjugacy class. Moreover, reversing their direction gives 24 more 7-cycles, completing the list of 168 symmetries. Applying Burnside's lemma, we get $\frac{1}{168}\left(c^7 + 56\,c^3 + 42\,c^3 + 21\,c^5 + 48\,c\right)$ distinct colourings.

Complex numbers

We assume that you are familiar with complex numbers in algebra, and delve immediately into the use of complex numbers in geometry. We are able to use complex numbers in two-dimensional geometry because the Euclidean plane, $\mathbb{R}^2$, is isomorphic to the complex plane, $\mathbb{C}$. We represent a point with Cartesian coordinates (x, y) with the complex number $z = x + i\, y$, where i is the imaginary unit. Complex numbers are superior to two-dimensional vectors in that rotations are easy to define, as we shall see shortly.

Basic properties of complex numbers

Complex numbers can be added, subtracted and multiplied by real numbers in precisely the same way that vectors can. They have a *magnitude* and *argument*, which correspond to the length and direction of a complex number. Also, for a complex number $z = x + i\, y$, we define its *complex conjugate* to be $z^* = x - i\, y$. As i and $-i$ have definitions analogous to 'left' and 'right', we can interchange all instances of i with $-i$ in an algebraic equation without affecting anything. Hence, $(z\, w)^* = z^*\, w^*$ and $(z + w)^* = z^* + w^*$.



In the above diagram, $\arg(z) = \theta$ is the *argument* of z, and $|z| = r$ is the *magnitude* (or *modulus*) of z. An important identity is that $z\, z^* = r^2$, which enables the (squared) modulus of a complex number to be calculated. Using vector subtraction, this gives us $A\,B^2 = (a - b)\,(a^* - b^*)$ for the squared distance between two points.

If we multiply a complex number with polar form $\langle r_1, \theta_1 \rangle$ with another complex number $\langle r_2, \theta_2 \rangle$, it is easy to verify that we obtain the complex number $\langle r_1\, r_2, \theta_1 + \theta_2 \rangle$. Hence, $\arg(z_1\, z_2) = \arg(z_1) + \arg(z_2)$ and $|z_1\, z_2| = |z_1|\ |z_2|$.

We can define the *inner product* in terms of complex numbers as $z \cdot w = \frac{1}{2}(z\, w^* + w\, z^*)$, which is analogous to the dot product of vectors. Similarly, we define the *exterior product* as $z \times w = \frac{1}{2}\, i(w\, z^* - z\, w^*)$, which resembles the cross product.

1. Prove that the area of $A\,O\,B$ is given by $[A\,O\,B] = \frac{1}{2}(a \times b) = \frac{1}{4}\, i(b\, a^* - a\, b^*)$.

2. Hence prove that the area of triangle $A\,B\,C$ is given by

$$[A\,B\,C] = \frac{1}{2}(b \times a + c \times b + a \times c) = \frac{1}{4}\, i \det\begin{pmatrix} 1 & 1 & 1 \\ a & b & c \\ a^* & b^* & c^* \end{pmatrix}.$$ **[Area of a triangle]**

3. If we have a triangle and erect equilateral triangles on its sides, prove that the centres of those equilateral triangles themselves form an equilateral triangle. **[Napoleon's theorem]**

Angles, circles and concyclicity

4. Show that the directed angle $\angle A B C = \arg(a - b) - \arg(c - b) = \arg\left(\frac{a-b}{c-b}\right)$.

5. Hence deduce that $\angle A B C \equiv \angle A D C \pmod{\pi}$ if and only if $\frac{(a-b)(c-d)}{(b-c)(d-a)}$ is real. **[Real cross ratio ⇔ concyclicity]**

More specifically, if this value is equal to -1, then $A B C D$ is known as a *harmonic* quadrilateral. Harmonic quadrilaterals are covered in the chapter on projective geometry.

6. Show that the equation of a circle with centre P and radius r has the equation $z z^* - p z^* - p^* z + p p^* - r^2 = 0$. **[General form of a circle]**

7. Prove that four points, $A B C D$, are mutually concyclic (or collinear) if and only if $(a - b)(b^* - c^*)(c - d)(d^* - a^*) = (a^* - b^*)(b - c)(c^* - d^*)(d - a)$.

8. Prove that four points, $A B C D$, are mutually concyclic (or collinear) if and only if

$$\det\begin{pmatrix} 1 & 1 & 1 & 1 \\ a & b & c & d \\ a^* & b^* & c^* & d^* \\ a a^* & b b^* & c c^* & d d^* \end{pmatrix} = 0.$$

The observant amongst you may notice that the previous two questions are equivalent quartic expressions. This demonstrates the equivalence between the 'angles in the same segment' and 'equidistant from a common point' conditions for concyclicity. Any quadratic function, which vanishes only on the circumference of a circle, must necessarily be proportional to the power of a point with respect to that circle. This gives us a more general result, which I believe has yet to be published elsewhere:

- For any four points, $A B C D$, no three of which are collinear, we have
$$\det\begin{pmatrix} 1 & 1 & 1 & 1 \\ a & b & c & d \\ a^* & b^* & c^* & d^* \\ a a^* & b b^* & c c^* & d d^* \end{pmatrix} = -4\, i\, [A B C]\ \mathrm{Power}(P,\ A B C).$$ **[Goucher's theorem]**

Interestingly, it was noted that at almost the same time a problem particularly vulnerable to this theorem was proposed at an International Mathematical Olympiad. There are ordinary Euclidean methods of proving this, but they are less inspired and do not explain *why* this result should hold.

9. Suppose we have a non-cyclic quadrilateral, $P_1 P_2 P_3 P_4$. Let O_1 and R_1 be the centre and radius, respectively, of the circumcircle of $P_2 P_3 P_4$, and define O_2, O_3, O_4 and R_2, R_3, R_4 similarly. Show that $\frac{1}{O_1 P_1{}^2 - R_1{}^2} + \frac{1}{O_2 P_2{}^2 - R_2{}^2} + \frac{1}{O_3 P_3{}^2 - R_3{}^2} + \frac{1}{O_4 P_4{}^2 - R_4{}^2} = 0$. **[IMO 2011 shortlist, Question G2]**

Reflections and rotations

A useful property of complex numbers is the ability to express rotations and reflections rather simply. We have concise expressions for reflection about the real axis (complex conjugation), rotation about the origin (multiplication by a unit complex number) and translation (addition of a complex number). Here are the three 'elementary' operations:

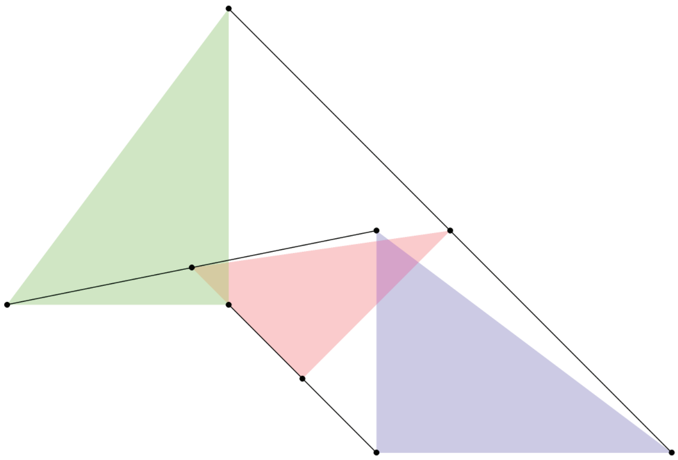
- A translation parallel to the vector $\overrightarrow{OA}$ is represented by $z \to z + a$. **[Translation]**
- An anticlockwise rotation about the origin by the angle θ is represented by $z \to z\,e^{i\theta}$. **[Rotation about the origin]**
- A reflection in the real line is represented by $z \to z^*$. **[Reflection in the real axis]**

These become more useful when one realises that they can be composed to yield any Euclidean transformation.

10. Show that an anticlockwise rotation by θ about the point A is represented by $z \to z\,e^{i\theta} + a\left(1 - e^{i\theta}\right)$.

11. Hence demonstrate that an arbitrary direct congruence is a transformation of the form $z \to a\,z + b$, where $a\,a^* = 1$.

If we relax the condition that $a\,a^* = 1$, we obtain the result that two directly similar figures can be related by a transformation of the form $z \to a\,z + b$. As this is a linear function, we can linearly interpolate between any two directly similar figures to obtain a third directly similar figure. Specifically, if triangles ABC and $A'B'C'$ are directly similar, then the (optionally weighted) midpoints of AA', BB' and CC' form a third similar triangle.

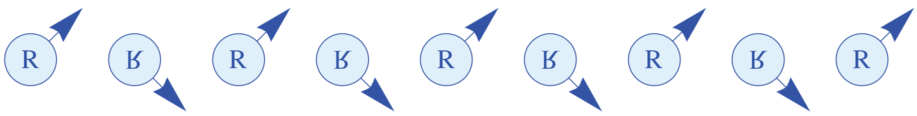


This is known as the *fundamental theorem of directly similar figures*. In the diagram above, the red triangle is the 'arithmetic mean' of the blue and green triangles. If we have five directly similar figures in general position, then any directly similar figure can be expressed as a 'weighted mean' of those five.

12. Show that a reflection in the line $z\,e^{-i\theta} \in \mathbb{R}$ is represented by $z \to z^*\,e^{2i\theta}$.

13. Hence demonstrate that an arbitrary indirect congruence is a transformation of the form $z \to a\,z^* + b$, where $a\,a^* = 1$.

In two dimensions, direct congruences can be either translations or rotations. Indirect congruences can be either reflections or *glide-reflections*. A glide-reflection is a composition of a reflection in a line and a translation parallel to the line.



14. Show that a glide-reflection has no fixed points.

15. Hence demonstrate that a reflection in the line $B\,C$ is represented by $z \to \frac{(b-c)\,(z^*-b^*)}{b^*-c^*} + b$.

16. If $b\,b^* = c\,c^* = R^2$, show that a reflection in the line $B\,C$ is represented by $z \to b + c - \frac{b\,c\,z^*}{R^2}$.

The comparative complexities of the previous two expressions show that the calculations become simpler when we assume that the circumcentre of a triangle $A\,B\,C$ is the origin. This is explored more thoroughly in a later section of this chapter.

17. Let $A\,B\,C$ be a triangle, and P be a point in the plane. Let the reflections of P in $B\,C$, $C\,A$ and $A\,B$ be D, E and F, respectively. Prove that $\frac{[D\,E\,F]}{[A\,B\,C]} = \frac{R^2-O\,P^2}{R^2}$. **[Euler's formula for pedal triangles]**

As a special case of the above, we have the Simson line property:

- Let $A\,B\,C$ be a triangle, and P be a point in the plane. Then the reflections of P in $B\,C$, $C\,A$ and $A\,B$ are collinear if and only if P lies on the circumcircle of $A\,B\,C$. Moreover, the orthocentre H lies on this line. **[Dilated Simson line]**

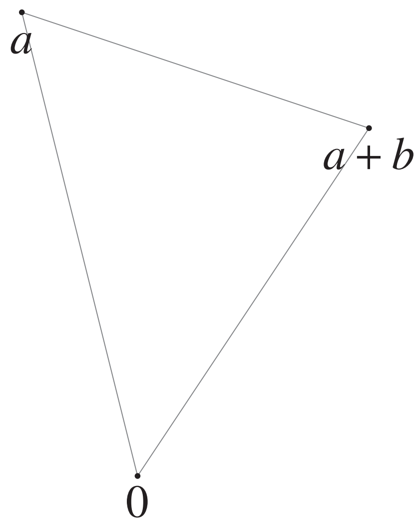
Usually, these results are quoted when D, E and F are the feet of the perpendiculars from P, rather than reflections of P. However, Euler's formula is more elegant with this version, and the standard Simson line in general does not contain H.

Triangle inequality

One of the more rudimentary inequalities governing vectors (and thus complex numbers) is the triangle inequality.

- If a and b are nonzero complex numbers, then $|a+b| \leq |a| + |b|$, with equality if and only if a is a **positive real multiple** of b. **[Triangle inequality]**

This follows immediately from the following configuration, together with the notion that the shortest path between two points is a straight line.



18. Show that $(a-b)\,(c-d) + (a-d)\,(b-c) = (a-c)\,(b-d)$, where $a, b, c, d \in \mathbb{C}$.

19. Hence prove that $|a-b|\;|c-d| + |a-d|\;|b-c| \geq |a-c|\;|b-d|$, with equality if and only if $\frac{(a-b)\,(c-d)}{(b-c)\,(d-a)}$ is a negative real.

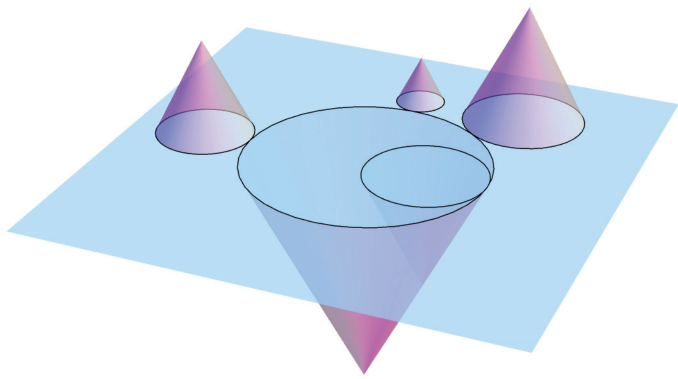
20. Let $ABCD$ be a quadrilateral. Show that $AB \cdot CD + BC \cdot DA \geq AC \cdot BD$, with equality if and only if $ABCD$ is a convex cyclic quadrilateral. **[Ptolemy's inequality]**

Casey's theorem

The equality case of Ptolemy's inequality can be regarded as a special case of Casey's theorem.

- Let Γ be a circle, and φ_1, φ_2, φ_3 and φ_4 be four circles tangent to Γ at P_1, P_2, P_3 and P_4, respectively. The chords $P_1 P_3$ and $P_2 P_4$ intersect inside Γ. For each pair of circles φ_i and φ_j, we let $d(i, j)$ denote the length of the common outer tangents if φ_i and φ_j are both on the same side of Γ, or the length of the common inner tangents if they lie on opposite sides of Γ. Then we have $d(1, 3) \cdot d(2, 4) = d(1, 2) \cdot d(3, 4) + d(2, 3) \cdot d(4, 1)$. **[Casey's theorem]**

The following exercise demonstrates how the latter can be inferred from the former using some basic trigonometry. Firstly, we consider a circle tangent externally to Γ to have positive radius, and a circle tangent internally to Γ to have negative radius. Γ itself is considered to have negative radius.



Erecting cones and 'anticones' on the circles with positive and negative radii, respectively, gives the diagram shown above.

21. If φ_1 and φ_2 are two circles with radii r_1 and r_2 and centres O_1 and O_2, respectively, then show that $d^2 = O_1 O_2^2 - (r_1 - r_2)^2$, where d is the length of the common outer tangents.

The value of d is dependent only on the positions of the centres and difference between the radii. This means we can fix the centres and uniformly increase the radii (using the sign convention described above) of all five circles in the problem by the same amount, without changing the values of $d(i, j)$ or affecting the tangency of the circles. (In the three-dimensional diagram, this is equivalent to moving the horizontal reference plane upwards or downwards.) So, we can assume without loss of generality that Γ is a single point (circle of zero radius) through which each φ_i passes. This greatly simplifies the analysis.

- Let P be a point, and φ_1, φ_2, φ_3 and φ_4 be four circles passing through P with centres O_1, O_2, O_3 and O_4, respectively. The V-shaped line $O_1 P O_3$ separates the plane into two regions; O_2 and O_4 lie in opposite regions. For each pair of circles φ_i and φ_j, we let $d(i, j)$ denote the length of the common outer tangents. Then we have $d(1, 3) \cdot d(2, 4) = d(1, 2) \cdot d(3, 4) + d(2, 3) \cdot d(4, 1)$. **[Simplified Casey's theorem]**

By proving the simplification of Casey's theorem, we will therefore implicitly prove the original theorem.

22. In the above problem, let φ_1 and φ_2 have radii r_1 and r_2, respectively. Show that $d(1, 2)^2 = 2\, r_1\, r_2(1 - \cos\theta)$, where $\theta = O_1 \hat{P} O_2$.

23. Hence show that $d(1, 3) \cdot d(2, 4)$ and the other terms in the simplified Casey's theorem are unaffected when r_1, r_2, r_3 and r_4 are simultaneously replaced with their geometric mean.

24. Hence prove the simplified Casey's theorem.

Casey's theorem, like Ptolemy's theorem, has a converse. If we have four (directed) circles and $d(1, 3) \cdot d(2, 4) = d(1, 2) \cdot d(3, 4) + d(2, 3) \cdot d(4, 1)$, then there exists a fifth circle tangent to all four circles. This is the basis of the shortest known proof of Feuerbach's theorem, demonstrating that there is a circle (the nine-point circle) tangent to the incircle and three excircles of a generic triangle.

Solutions

1. Suppose $\angle AOB = \theta$. Then we have $[AOB] = \frac{1}{2}\left|a\right|\,\left|b\right|\,\sin\theta = \frac{1}{2}(a\times b)$.

2. $[ABC] = [BOA] + [COB] + [AOC] = \frac{1}{2}(b\times a + c\times b + a\times c)$. Using the formula for cross product, this equals $\frac{i}{4}(a\,b^* - b\,a^* + b\,c^* - c\,b^* + c\,a^* - a\,c^*) = \frac{1}{4}\,i\det\begin{pmatrix}1 & 1 & 1\\ a & b & c\\ a^* & b^* & c^*\end{pmatrix}$.

3. Assume, without loss of generality, that the triangle is labelled anticlockwise. Let x, y and z be the centres of the equilateral triangles erected on BC, CA and AB, respectively. We have $x = \frac{1}{3}\left(b + c - \omega\,c - \omega^2\,b\right)$, where $\omega = e^{\frac{2}{3}i\pi}$ is a cube root of unity. This gives us $x - y = \frac{1}{3}\left(a(\omega - 1) + b\left(1 - \omega^2\right) + c(\omega - 1)\right)$. The symmetrical expression means that $(x - y) = \omega(z - x)$, which is a sufficient condition for the triangle to be equilateral.

4. We translate the configuration so that B is the origin, and A and C are represented by complex numbers $a - b$ and $c - b$, respectively. Hence, we have $\angle ABC = \arg(a - b) - \arg(c - b)$, as required. The final part of the proof, namely that this also equals $\arg\left(\frac{a-b}{c-b}\right)$, follows from the prosthaphaeretic property of the arg() function.

5. $\angle ABC \equiv \angle ADC \pmod{\pi} \Leftrightarrow \arg\left(\frac{a-b}{c-b}\right) \equiv \arg\left(\frac{a-d}{c-d}\right) \pmod{\pi} \Leftrightarrow \frac{(a-b)(c-d)}{(b-c)(d-a)} \in \mathbb{R}$.

6. $|z - p| = r \Leftrightarrow (z - p)(z^* - p^*) = r^2 \Leftrightarrow z\,z^* - p\,z^* - p^*\,z + p\,p^* - r^2 = 0$.

7. This is a consequence of the ‘angle in the same segment’ theorem for concyclicity and the result of Question 5. As $\frac{(a-b)(c-d)}{(b-c)(d-a)} \in \mathbb{R}$, it must be equal to its complex conjugate $\frac{(a^*-b^*)(c^*-d^*)}{(b^*-c^*)(d^*-a^*)}$. We then multiply throughout by the common denominator, giving the quartic equation
$(a - b)(b^* - c^*)(c - d)(d^* - a^*) = (a^* - b^*)(b - c)(c^* - d^*)(d - a)$.

8. Suppose D is variable and A, B and C are constants. The equation multiplies out to the form in Question 6, so is the condition that D lies on some circle. As the determinant vanishes whenever two columns are equal, $D = A$, $D = B$ and $D = C$ all satisfy this equation. Hence, it must be the circumcircle of ABC, and we are done.

9. Multiply through by $\frac{1}{4}\,i\det\begin{pmatrix}1 & 1 & 1 & 1\\ a & b & c & d\\ a^* & b^* & c^* & d^*\\ a\,a^* & b\,b^* & c\,c^* & d\,d^*\end{pmatrix}$ and use Goucher’s theorem. This leads to the equivalent statement about signed areas $[ABC] + [CDA] = [BCD] + [DAB]$.

10. To obtain this transformation, we compose a translation by $\overrightarrow{AO}$, rotation by θ anticlockwise about O, and a translation by $\overrightarrow{OA}$. They have the formulae $z \to z - a$, $z \to z\,e^{i\theta}$ and $z \to z + a$, respectively. Composing these in order yields $z \to (z - a)\,e^{i\theta} + a = z\,e^{i\theta} + a\left(1 - e^{i\theta}\right)$.

11. As $z \to a\,z + b$ is a linear transformation, it is closed under composition. Translations and rotations are of this form, ergo every rigid transformation is.

12. Again, we compose a rotation by θ clockwise ($z \to z\, e^{-i\theta}$), a reflection in the real axis ($z \to z^*$) and a rotation by θ anticlockwise ($z \to z\, e^{i\theta}$). The composite transformation has rule $z \to z^*\, e^{2i\theta}$.

13. Composing a function of the form $z \to a\, z^* + b$ with any linear function results in another function of the form $z \to a\, z^* + b$. All indirect congruences can be built from a reflection in the real axis and a rigid transformation, so must also be of this form.

14. Composing a glide-reflection with itself results in a translation, which clearly has no fixed points.

15. The reflection must have the form $z \to p\, z^* + q$, where $p\, p^* = 1$, as we demonstrated earlier. Having B and C as fixed points (as the transformation does indeed) proves that it is not a glide reflection, and must be the reflection in the line $B\,C$.

16. The reasoning is identical to the previous question.

17. Without loss of generality, assume $a\, a^* = b\, b^* = c\, c^* = R^2$. Then $d = b + c - \frac{b\, c\, p^*}{R^2}$, and e and f have similar forms. The determinant (proportional to the area) is given by $\sum_{\text{cyc}} (d\, e^* - e\, d^*)$, where $\sum_{\text{cyc}}$ denotes the cyclic sum interchanging a, b and c. This evaluates to

$$\det\begin{pmatrix} 1 & 1 & 1 \\ d & e & f \\ d^* & e^* & f^* \end{pmatrix} = \sum_{\text{cyc}} \left(a\, b^* - b\, a^* + \frac{b\, a^*\, p\, p^*}{R^2} - \frac{a\, b^*\, p\, p^*}{R^2}\right) = \sum_{\text{cyc}} (a\, b^* - b\, a^*)\left(1 - \frac{p\, p^*}{R^2}\right) = \det\begin{pmatrix} 1 & 1 & 1 \\ a & b & c \\ a^* & b^* & c^* \end{pmatrix}\left(1 - \frac{O\,P^2}{R^2}\right).$$

Or, in other words, $[D\,E\,F] = [A\,B\,C]\left(1 - \frac{O\,P^2}{R^2}\right)$.

18. Both sides of the equation expand to $a\,b - b\,c + c\,d - d\,a$.

19. As the modulus function is multiplicative, this is equivalent to $|(a-b)(c-d)| + |(a-d)(b-c)| \ge |(a-c)(b-d)|$. This is the triangle inequality, so equality only holds when $\frac{(a-b)(c-d)}{(a-d)(b-c)}$ is a positive real, or $\frac{(a-b)(c-d)}{(b-c)(d-a)}$ is a negative real.

20. The real cross-ratio condition forces A, B, C and D to be concyclic. If A and C lie on the same side of $B\,D$, the cross-ratio would be positive; hence, chords $A\,C$ and $B\,D$ must intersect.

21. Assume, without loss of generality, that $r_1 \ge r_2$. Let one of the common outer tangents meet Ω_1 at A and Ω_2 at B. Further, let C be the point on the radius $O_1\,A$ such that $A\,C = r_2$ and $O_1\,C = r_1 - r_2$. As $C\,O_2\,B\,A$ is a rectangle, we have $d = A\,B = C\,O_2$ (not carbon dioxide!). By applying Pythagoras' theorem to the triangle $O_1\,O_2\,C$, we obtain $C\,O_2{}^2 = O_1\,O_2{}^2 - C\,O_1{}^2$, each term of which is equal to $d^2 = O_1\,O_2{}^2 - (r_1 - r_2)^2$.

22. Using the formula from the previous question, we have $d(1, 2)^2 = O_1\,O_2{}^2 - (r_1 - r_2)^2$. The cosine rule gives us $O_1\,O_2{}^2 = r_1{}^2 + r_2{}^2 - 2\, r_1\, r_2 \cos\theta$, and multiplying out yields $(r_1 - r_2)^2 = r_1{}^2 + r_2{}^2 - 2\, r_1\, r_2$. The difference between these expressions is $2\, r_1\, r_2(1 - \cos\theta)$, as required.

23. The previous question results in $(d(1, 3)\cdot d(2, 4))^2 = 4\, r_1\, r_2\, r_3\, r_4(1 - \cos\theta)(1 - \cos\phi)$, where θ and ϕ are defined in the obvious way. This is unaffected when we replace each of r_1, r_2, r_3 and r_4 with $r = \sqrt[4]{r_1\, r_2\, r_3\, r_4}$, as the product remains equal to r^4. Hence, the value of $d(1, 3)\cdot d(2, 4)$ also remains invariant. By symmetry, so do the other terms.

24. We can assume without loss of generality that $r_1 = r_2 = r_3 = r_4 = r$. The distance $d(1, 2) = O_1\,O_2$, *et cetera*. Since O_1, O_2, O_3 and O_4 lie on a circle by Ptolemy's theorem, we are done.

Triangle geometry

In this chapter, we consider the basic properties of a generic triangle ABC, and how the angles and distances between points are related. We explore parametrisations of the triangle using both trigonometry and complex numbers. In the process, we develop an arsenal of identities suitable for attacking both geometrical and trigonometrical problems, noting the interchangeability between the representations.

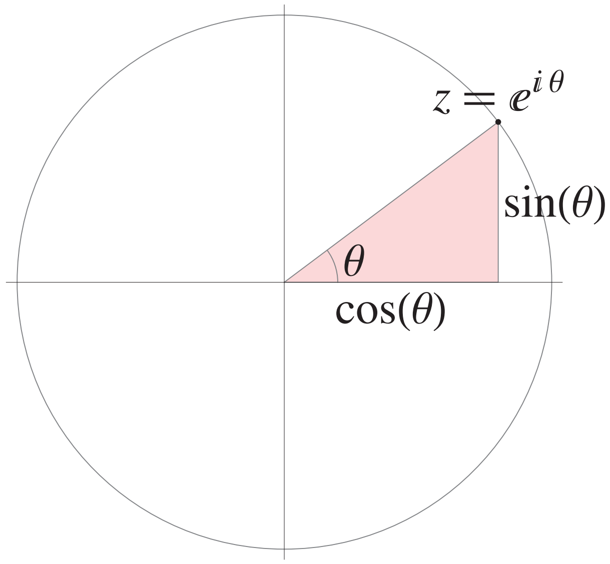
Trigonometry

The elementary trigonometric functions, namely sine and cosine, can be expressed in terms of the exponential function and *vice-versa*.

- $\sin(\theta) = \mathrm{Im}(e^{i\theta}) = \frac{e^{i\theta} - e^{-i\theta}}{2i}$. **[Definition of sine]**

- $\cos(\theta) = \mathrm{Re}(e^{i\theta}) = \frac{e^{i\theta} + e^{-i\theta}}{2}$. **[Definition of cosine]**

We can view this on the Argand plane, where we consider a point on the unit circle with Cartesian coordinates $(\cos\theta,\ \sin\theta)$ and complex representation $\cos\theta + i\sin\theta = e^{i\theta}$. From applying Pythagoras' theorem, we instantly obtain the famous identity $\sin^2\theta + \cos^2\theta = 1$.



This is arguably the most reliable approach to proving trigonometric identities, as it is a simple matter of converting each expression to its exponential counterpart and verifying that both sides of the equation are indeed equal. However, it is preferable to derive a few identities first, as working with lots of exponentials can be laborious. Perhaps the most rudimentary trigonometric identities are the *compound angle formulae*.

1. Prove that $\sin(\theta + \phi) = \sin\theta\cos\phi + \sin\phi\cos\theta$. **[Compound angle formula I]**

2. Similarly, prove that $\cos(\theta + \phi) = \cos\theta\cos\phi - \sin\phi\sin\theta$. **[Compound angle formula II]**

With these, it is no longer necessary to rely on the exponential form for proving identities. Indeed, we can now avoid using complex numbers altogether.

More sophisticated trigonometric functions can be expressed as ratios of sine and cosine.

- $\tan\theta = \frac{\sin\theta}{\cos\theta}$; $\cot\theta = \frac{\cos\theta}{\sin\theta}$; $\sec\theta = \frac{1}{\cos\theta}$; $\mathrm{cosec}\,\theta = \frac{1}{\sin\theta}$. **[Definitions of tangent, cotangent, secant and cosecant]**

This enables us to derive a compound angle formula for the tangent function.

3. Hence prove that $\tan(\theta + \phi) = \frac{\tan\theta + \tan\phi}{1 - \tan\theta\tan\phi}$. **[Compound angle formula III]**

As special cases of the above, where $\theta = \phi$, we obtain the *double-angle formulae*.

- $\sin(2\theta) = 2\sin\theta\cos\theta$. **[Double-angle formula I]**
- $\cos(2\theta) = \cos^2\theta - \sin^2\theta$. **[Double-angle formula II]**

4. Prove further that $\cos(2\theta) = 2\cos^2\theta - 1 = 1 - 2\sin^2\theta$. **[Extended double-angle formula]**

We can rearrange the above formula to obtain $\sin^2\theta$ and $\cos^2\theta$ in terms of $\cos 2\theta$. Hence, we can calculate the value of $\cos\frac{\pi}{6}$ from that of $\cos\frac{\pi}{3}$, for example.

5. Hence deduce that $\cos(\theta + \phi)\cos(\theta - \phi) = \frac{1}{2}(\cos 2\theta + \cos 2\phi) = \cos^2\theta - \sin^2\phi$. **[Prosthaphaeresis]**

The compound angle formulae can be used recursively to derive expressions for three angles.

6. Prove that $\sin(3\theta) = 3\sin\theta - 4\sin^3\theta$. **[Triple-angle formula]**

More generally, we can expand $\sin(\theta + \phi + \psi)$ to obtain $\sin(\theta + \phi)\cos\psi + \cos(\theta + \phi)\sin\psi$, then apply the compound angle formulae again to each term. This results in the following expression:

- $\sin(\theta + \phi + \psi) = \sin\theta\cos\phi\cos\psi + \cos\theta\sin\phi\cos\psi + \cos\theta\cos\phi\sin\psi - \sin\theta\sin\phi\sin\psi$. **[Compound angle formula IV]**

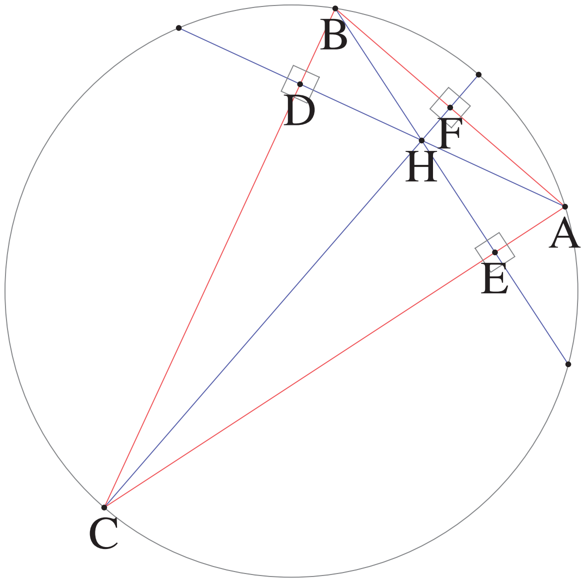
Let A, B and C be the angles of a triangle opposite sides of lengths a, b and c, respectively. R, r and s are the circumradius, inradius and semiperimeter, respectively. We can apply the trigonometric identities explored in the previous section to triangles, remembering that $A + B + C = \pi$, and thus $\sin(A + B) = \sin C$ and $\cos(A + B) = -\cos C$.

7. Prove that $\sin A\sin B\cos C + \sin A\cos B\sin C + \cos A\sin B\sin C - \cos A\cos B\cos C = 1$, and thus $\cot A + \cot B + \cot C - 1 = \operatorname{cosec} A\operatorname{cosec} B\operatorname{cosec} C$.

8. Prove that $\tan A\tan B\tan C = \tan A + \tan B + \tan C$, and thus $\cot A\cot B + \cot B\cot C + \cot C\cot A = 1$.

Altitudes and orthocentre

Consider the triangle ABC together with its orthocentre H. The altitudes meet BC, CA and AB at D, E and F, respectively.



The orthocentric configuration exhibits a plethora of particularly interesting properties:

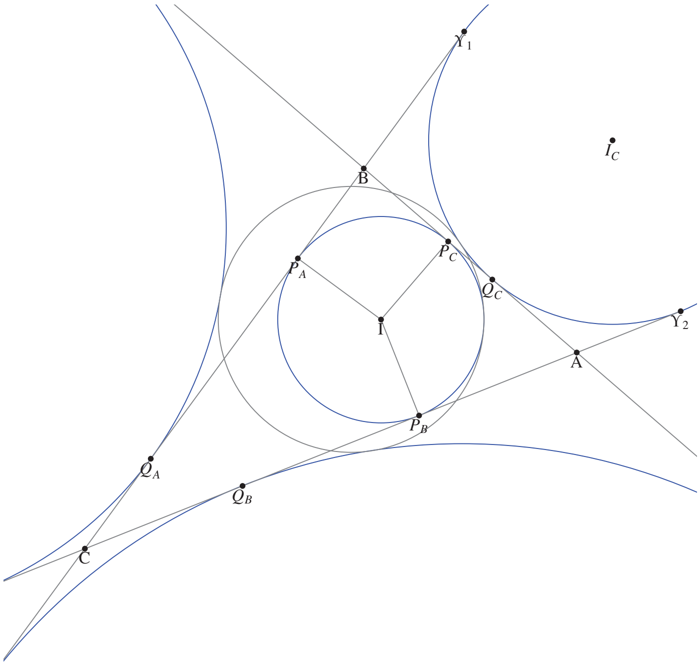
- The reflections of the orthocentre, H, in each of the sides of the triangle land on the circumcircle.
- If three points out of $\{A, B, C, D, E, F, H\}$ are collinear, the remaining four are concyclic, and *vice-versa*.
- If H is the orthocentre of ABC, then A is the orthocentre of HBC, *et cetera*. This is called an *orthocentric quadrangle* or *perpendicularogram*. All four triangles share the same nine-point circle, the centre of which is the barycentre of $\{A, B, C, H\}$.
- Due to cyclic quadrilaterals, we have $\angle DHB = \angle AHE = \angle BCA$.
- Every inter-point distance has a simple expression in terms of the circumradius and trigonometric functions of the vertex angles:

	A	B	C	D	E	F	H
A	0	$2R\sin(C)$	$2R\sin(B)$	$2R\sin(B)\sin(C)$	$2R\cos(A)\sin(C)$	$2R\cos(A)\sin(B)$	$2R\cos(A)$
B		0	$2R\sin(A)$	$2R\cos(B)\sin(C)$	$2R\sin(A)\sin(C)$	$2R\sin(A)\cos(B)$	$2R\cos(B)$
C			0	$2R\sin(B)\cos(C)$	$2R\sin(A)\cos(C)$	$2R\sin(A)\sin(B)$	$2R\cos(C)$
D				0	$2R\sin(C)\cos(C)$	$2R\sin(B)\cos(B)$	$2R\cos(B)\cos(C)$
E					0	$2R\sin(A)\cos(A)$	$2R\cos(A)\cos(C)$
F						0	$2R\cos(A)\cos(B)$
H							0

- Every rectangular hyperbola passing through A, B and C also passes through H. The centre of the hyperbola lies on the nine-point circle of triangle ABC.

Tritangential circles

We consider the four circles tangent to all three sides of a triangle, together with their centres and tangency points. One of these circles is enclosed by the triangle (the *inscribed circle*, or *incircle*, with *incentre* I), and the other three are called *escribed circles*, or *excircles*. Collectively, they are known as *tritangential circles*.



As the two tangents from a single point to a circle are equal, we have that $AP_B = AP_C$. Let l, m and n denote the distances AP_B, BP_C and CP_A, respectively. We have $a = m + n$, $b = n + l$ and $c = l + m$. This enables us to deduce that $AP_B = AP_C = l = s - a$, where $s = \frac{1}{2}(a + b + c)$ is the semiperimeter of the triangle ABC. We know that $C\Upsilon_1 = C\Upsilon_2$, and that $B\Upsilon_1 = BQ_C$, which gives us $CB + BQ_C = Q_CA + AC = s$, from which we can deduce that $BQ_C = s - a$.

■ $AP_C = BQ_C = s - a$. **[Distances to intouch and extouch points]**

This means that the line segments P_CQ_C and AB share a midpoint.

Applying Pythagoras' Theorem to triangle AP_BI enables the distance $AI^2 = AP_B{}^2 + P_BI^2 = (s-a)^2 + r^2$ to be determined. Similarly, we have $AI_C{}^2 = (s-b)^2 + r_C{}^2$ and $AI_A{}^2 = s^2 + r_A{}^2$.

9. Prove that $rs = r_A(s-a) = r_B(s-b) = r_C(s-c) = [ABC]$. **[Area of a triangle]**

10. Show that $\tan\frac{A}{2} = \frac{r}{s-a} = \frac{r_A}{s}$. **[Half-angle formula]**

11. Show that $AI = 4R\sin\frac{B}{2}\sin\frac{C}{2}$, and thus $r = 4R\sin\frac{A}{2}\sin\frac{B}{2}\sin\frac{C}{2}$. **[Inradius formula]**

12. Prove that $s - b = r_C\tan\frac{A}{2}$. **[Complementary half-angle formula]**

With the inradius formula and half-angle formula, we obtain an expression for $s - a$; by symmetry, we also get $s - b$ and $s - c$. The complementary half-angle formula gives us r_A, r_B and r_C, and the bog-standard half-angle formula gives us s. These eight quantities are included in the table below so you can gape in awe at the elegant symmetries between the formulae.

Length	Trigonometrical expression
r	$4R\sin\left(\frac{A}{2}\right)\sin\left(\frac{B}{2}\right)\sin\left(\frac{C}{2}\right)$
$s-a$	$4R\cos\left(\frac{A}{2}\right)\sin\left(\frac{B}{2}\right)\sin\left(\frac{C}{2}\right)$
$s-b$	$4R\sin\left(\frac{A}{2}\right)\cos\left(\frac{B}{2}\right)\sin\left(\frac{C}{2}\right)$
$s-c$	$4R\sin\left(\frac{A}{2}\right)\sin\left(\frac{B}{2}\right)\cos\left(\frac{C}{2}\right)$
r_A	$4R\sin\left(\frac{A}{2}\right)\cos\left(\frac{B}{2}\right)\cos\left(\frac{C}{2}\right)$
r_B	$4R\cos\left(\frac{A}{2}\right)\sin\left(\frac{B}{2}\right)\cos\left(\frac{C}{2}\right)$
r_C	$4R\cos\left(\frac{A}{2}\right)\cos\left(\frac{B}{2}\right)\sin\left(\frac{C}{2}\right)$
s	$4R\cos\left(\frac{A}{2}\right)\cos\left(\frac{B}{2}\right)\cos\left(\frac{C}{2}\right)$

13. Prove that $[ABC] = \sqrt{r\, r_A\, r_B\, r_C} = \sqrt{s(s-a)(s-b)(s-c)} = \frac{abc}{4R}$. **[Beyond Heron's formula]**

14. Hence prove that $\cot A = \frac{b^2+c^2-a^2}{4[ABC]}$.

15. Show further that $[ABC] = \frac{1}{2}\sqrt[3]{a^2 b^2 c^2 \sin A \sin B \sin C}$. **[Gendler's formula]**

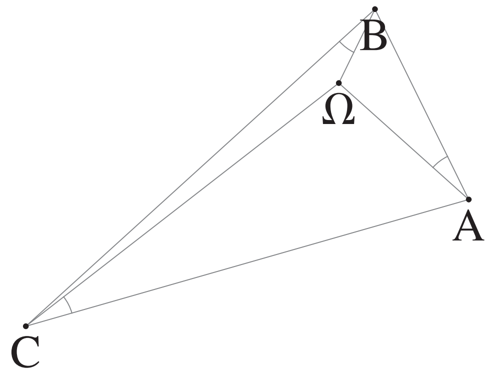
16. Prove that $\sin A + \sin B + \sin C = 4\cos\frac{A}{2}\cos\frac{B}{2}\cos\frac{C}{2} = \frac{s}{R}$.

Here are some miscellaneous properties of the tritangential circles:

- If the lines $C I I_C$ and AB intersect at R, then $(C, R; I, I_C) = -1$ is a harmonic range.
- The lines $A P_A$, $B P_B$ and $C P_C$ are concurrent at the *Gergonne point* Ge, whilst the lines $A Q_A$, $B Q_B$ and $C Q_C$ are concurrent at the *Nagel point* Na. They are isotomic conjugates. The incentre, centroid, *Spieker centre* (incentre of the medial triangle) and Nagel point are collinear such that $3\overrightarrow{IG} = 6\overrightarrow{G\,\mathrm{Sp}} = 2\overrightarrow{\mathrm{Sp}\,\mathrm{Na}}$. This is known as the *Nagel line*, and is not unlike the Euler line.
- Feuerbach's theorem states that the incircle and three excircles are tangent to the nine-point circle.
- The excentral triangle $I_A I_B I_C$ has orthocentre I. Its nine-point circle is the circumcircle of the reference triangle.

Brocard points

The *first Brocard point* Ω is positioned such that $\angle \Omega BC = \angle \Omega CA = \angle \Omega AB = \omega$, where ω is known as the *Brocard angle*. The *second Brocard point* Ω' is its isogonal conjugate, where $\angle \Omega' BA = \angle \Omega' CB = \angle \Omega' AC = \omega$.



17. Prove that $(\cot\omega - \cot A)(\cot\omega - \cot B)(\cot\omega - \cot C) = \operatorname{cosec} A \operatorname{cosec} B \operatorname{cosec} C$.

18. Hence show that $\cot\omega$ is a root of the cubic equation
$x^3 - (\cot A + \cot B + \cot C)x^2 + x - (\cot A + \cot B + \cot C) = 0$.

Let Γ_{AB} be the circle through A and B tangent to BC, and define Γ_{BC} and Γ_{CA} similarly. Ω lies on the intersection of the three circles. The other triple intersections of these three circles are the two circular points at infinity, which correspond to the imaginary roots of the cubic equation.

19. Show that the above equation has only one real root, and thus $\cot\omega = \cot A + \cot B + \cot C$. **[Brocard angle formula]**

Now that we have this expression for $\cot\omega$, we can derive further identities:

20. Prove that $\tan\omega = \frac{\sin A \sin B \sin C}{1+\cos A \cos B \cos C}$.

21. Prove that $\cot\omega = \frac{a^2+b^2+c^2}{4[ABC]}$.

22. Hence show that $\omega \leq \frac{\pi}{6}$, with equality if and only if ABC is equilateral.

23. Let P be a point interior to a scalene triangle ABC. Prove that one of the angles $\angle PAB$, $\angle PBC$ and $\angle PCA$ must be less than $\frac{\pi}{6}$. **[Adapted from IMO 1991, Question 5]**

R, *r* and *s*

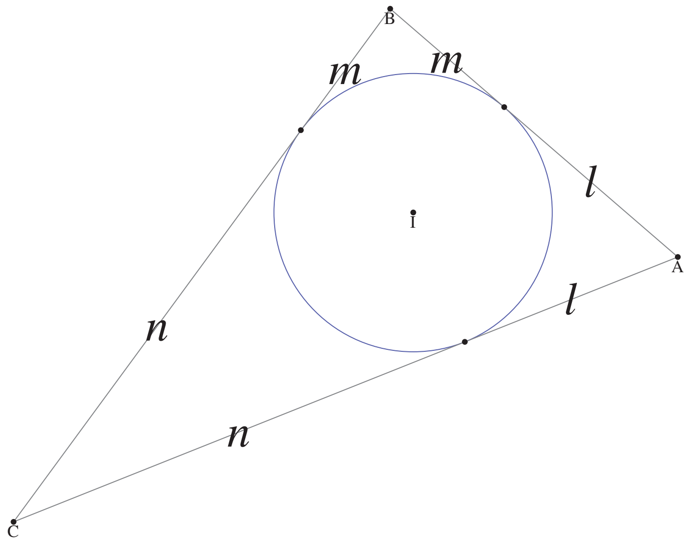
When evaluating (squared) distances between triangle centres, one often finds a symmetric polynomial in the side lengths (a, b and c). It is convenient to convert this into an expression in the circumradius R, inradius r and semiperimeter s.

24. Using the formulae $[ABC] = rs = \frac{abc}{4R} = \sqrt{s(s-a)(s-b)(s-c)}$, derive expressions for $a+b+c$, $ab+bc+ca$ and abc in terms of R, r and s.

Due to Newton's theorem of symmetric polynomials, it is possible to express any symmetric polynomial in the side lengths in terms of these elementary symmetric polynomials, and thus in terms of R, r and s.

Symmetric polynomial	R, r and s
$a+b+c$	$2\,s$
$a\,b+a\,c+b\,c$	$r^2+4\,r\,R+s^2$
$a\,b\,c$	$4\,r\,R\,s$
$a^2+b^2+c^2$	$-2\,r^2-8\,r\,R+2\,s^2$
$a^3+b^3+c^3$	$2\,s\left(-3\,r^2-6\,r\,R+s^2\right)$

When dealing with inequalities in the side lengths of a triangle, it is most convenient to convert it into an inequality in $l=s-a$, $m=s-b$ and $n=s-c$. The triangle inequality is equivalent to l, m and n being positive reals. Symmetric polynomials in l, m and n can similarly be converted into polynomials in R, r and s using this method.



Symmetric polynomial	R, r and s
$l+m+n$	s
$l\,m+l\,n+m\,n$	$r^2+4\,r\,R$
$l\,m\,n$	$r^2\,s$
$l^2+m^2+n^2$	$-2\,r^2-8\,r\,R+s^2$
$l^3+m^3+n^3$	$s^3-12\,r\,R\,s$

25. Prove that, for every triangle, $s^2 \geq 16\,R\,r-5\,r^2$. **[One half of Gerretsen's inequality]**

26. Similarly prove that $s^2 \leq 4\,R^2+4\,R\,r+3\,r^2$. **[Other half of Gerretsen's inequality]**

We can go further and find the tightest bounds possible for s^2 given R and r. As l, m, n are three real roots of the equation $(x-l)\,(x-m)\,(x-n)=0$, it is a necessary and sufficient condition for the cubic $x^3-(l+m+n)\,x^2+(l\,m+m\,n+n\,l)\,x-l\,m\,n=0$ to have three (not necessarily distinct) real roots. Using the formulae for the elementary symmetric polynomials, we require the discriminant of $x^3-s\,x^2+\left(r^2+4\,R\,r\right)x-r^2\,s\,x=0$ to be non-negative. The discriminant is given by $18\,r^2\,s^2\left(r^2+4\,R\,r\right)-4\,r^2\,s^4+s^2\left(r^2+4\,R\,r\right)-4\left(r^2+4\,R\,r\right)^3-27\,r^4\,s^2$, which is a quadratic function in s^2. Solving this inequality gives us the following necessary and sufficient condition on s^2 in terms of R and r:

- $2R^2 + 10Rr - r^2 - 2\sqrt{R(R-2r)^3} \leq s^2 \leq 2R^2 + 10Rr - r^2 + 2\sqrt{R(R-2r)^3}$.

The semiperimeter will fluctuate between these two values as the vertices of the triangle move around the circumcircle in Poncelet's porism.

27. Show that $R \geq 2r$. **[Euler's inequality]**

28. Hence prove that $\sin A + \sin B + \sin C \geq \sin 2A + \sin 2B + \sin 2C$. **[Gendler's inequality]**

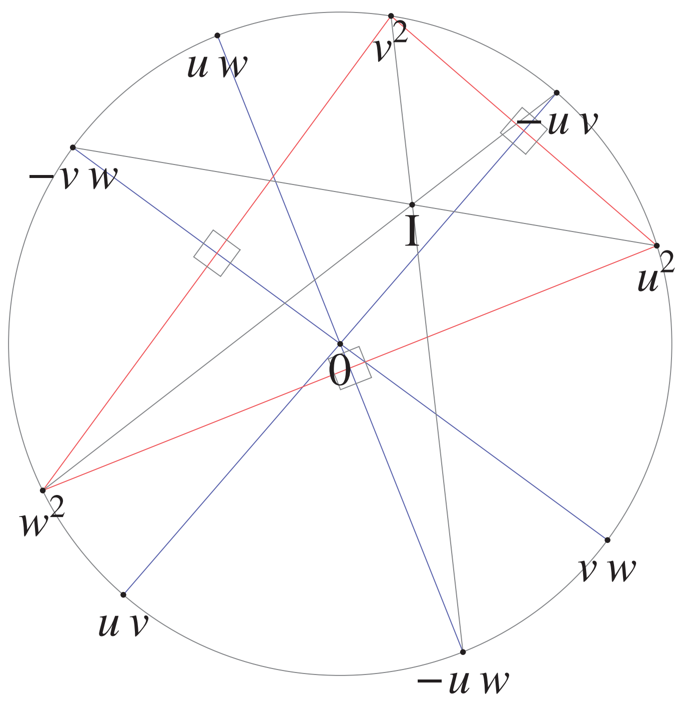
29. Express $\cot \omega$ in terms of R, r and s.

30. Prove that $\cos A + \cos B + \cos C = 1 + 4 \sin \frac{A}{2} \sin \frac{B}{2} \sin \frac{C}{2} = 1 + \frac{r}{R}$.

31. Hence prove that $\cos A - \cos B - \cos C = 1 - 4 \sin \frac{A}{2} \cos \frac{B}{2} \cos \frac{C}{2} = 1 - \frac{r_A}{R}$.

32. Show that $4R = r_A + r_B + r_C - r$.

Complex parametrisation of triangles



Consider the reference triangle ABC. If we represent the vertices with complex numbers u^2, v^2 and w^2, respectively, such that $u u^* = v v^* = w w^* = R$, and choose the signs of u, v and w such that $u v$ lies on the arc AB containing C (and cyclic permutations thereof), then most of the useful aspects of the triangle have simple algebraic expressions:

Quantity	Symbol	Expression
Circumradius	R	$u\,u^* = v\,v^* = w\,w^*$
Inradius	r	$\frac{1}{2}\left(-v\,u^* - u\,v^* - w\,u^* - u\,w^* - w\,v^* - v\,w^*\right) - R$
Exradius opposite A	r_A	$\frac{1}{2}\left(-v\,u^* - u\,v^* - w\,u^* - u\,w^* + w\,v^* + v\,w^*\right) + R$
Semiperimeter	s	$\left(\frac{1}{2}\,i\left(v\,u^* - u\,v^* - w\,u^* + u\,w^* + w\,v^* - v\,w^*\right)\right)$
(Signed) area	[ABC]	$\left(\frac{1}{4}\,i\left(u^2\left(v^*\right)^2 - u^2\left(w^*\right)^2 - v^2\left(u^*\right)^2 + w^2\left(u^*\right)^2 + v^2\left(w^*\right)^2 - w^2\left(v^*\right)^2\right)\right)$
Side length of BC	a	$i\left(w\,v^* - v\,w^*\right)$
Angle exponential	$e^{i\,A}$	$-\frac{w\,v^*}{R}$
Sine	$\sin(A)$	$\frac{i\left(w\,v^* - v\,w^*\right)}{2\,R}$
Cosine	$\cos(A)$	$\frac{-w\,v^* - v\,w^*}{2\,R}$

There are also versions of these formulae expressed in linear factors, which can easily be multiplied and divided:

Quantity	Symbol	Expression
Inradius	r	$-\frac{1}{2}\,R\left(\frac{u}{v} + 1\right)\left(\frac{w}{u} + 1\right)\left(\frac{v}{w} + 1\right)$
Exradius opposite A	r_A	$\frac{1}{2}\,R\left(1 - \frac{u}{v}\right)\left(1 - \frac{w}{u}\right)\left(\frac{v}{w} + 1\right)$
Semiperimeter	s	$\frac{1}{2}\,i\,R\left(1 - \frac{u}{v}\right)\left(1 - \frac{w}{u}\right)\left(1 - \frac{v}{w}\right)$
	$s - a$	$-\frac{1}{2}\,i\,R\left(\frac{u}{v} + 1\right)\left(\frac{w}{u} + 1\right)\left(1 - \frac{v}{w}\right)$
(Signed) area	[ABC]	$\left(-\frac{1}{4}\,i\,R^2\left(\frac{v}{u} - \frac{u}{v}\right)\left(\frac{u}{w} - \frac{w}{u}\right)\left(\frac{w}{v} - \frac{v}{w}\right)\right)$
Side length of BC	a	$i\,R\left(\frac{w}{v} - \frac{v}{w}\right)$
Angle exponential	$e^{i\,A}$	$-\frac{w}{v}$
Sine	$\sin(A)$	$\left(\frac{1}{2}\,i\left(\frac{w}{v} - \frac{v}{w}\right)\right)$
Cosine	$\cos(A)$	$\frac{1}{2}\left(-\frac{v}{w} - \frac{w}{v}\right)$

Many triangle centres have simple quadratic expressions in u, v and w. Others, such as the Feuerbach points, are more complicated:

Point	Symbol	Expression
Vertex	A	u^2
Vertex	B	v^2
Vertex	C	w^2
Circumcentre	O	0
Centroid	G	$\left(\frac{1}{3}\left(u^2+v^2+w^2\right)\right)$
Nine−point centre	T	$\left(\frac{1}{2}\left(u^2+v^2+w^2\right)\right)$
Orthocentre	H	$u^2+v^2+w^2$
Altitude foot on BC	D	$\left(\frac{1}{2}\left(-\frac{v^2\,w^2}{u^2}+u^2+v^2+w^2\right)\right)$
Incentre	I	$-u\,v-u\,w-v\,w$
Intouch point on BC	P_A	$\frac{(u+v)\,(u+w)\,(v+w)}{2\,u}-u\,v-u\,w-v\,w$
Excentre opposite A	I_A	$u\,v+u\,w-v\,w$
Extouch point on BC	Q_A	$\frac{(u-v)\,(w-u)\,(v+w)}{2\,u}+u\,v+u\,w-v\,w$
Nagel point	Na	$(u+v+w)^2$
Spieker centre	Sp	$\frac{1}{2}\left(u^2+u\,v+u\,w+v^2+v\,w+w^2\right)$
Feuerbach point	F	$\frac{1}{2}\left(-\frac{R\,(u+v+w)}{u^*+v^*+w^*}+u^2+v^2+w^2\right)$
Feuerbach point on excircle I_A	F_A	$\frac{1}{2}\left(-\frac{R\,(u-v-w)}{u^*-v^*-w^*}+u^2+v^2+w^2\right)$

With these results, one can express any rational function of the side lengths and basic trigonometric functions as a rational function in u, v and w (and their conjugates).

Firstly, however, it is a fulfilling exercise to derive the expressions in the table above.

33. Show that $u\,v$ is the midpoint of the arc $A\,B$ containing C, and thus that $-u\,v$ is the midpoint of the arc $A\,B$ not containing C.

34. Prove that the circumcircle of $A\,B\,C$ is the nine-point circle of the *excentric triangle* $I_A\,I_B\,I_C$.

35. Hence show that $-u\,v$ is the midpoint of $I\,I_C$, and thus $u\,v$ is the midpoint of $I_A\,I_B$.

36. Hence verify the expressions for I, I_A, I_B and I_C.

With the expressions for the circumcentre and centroid, one can derive expressions for the orthocentre and nine-point centre by using the Euler line. Similarly, the Nagel line enables one to extrapolate expressions for the Spieker centre and Nagel point based on those of the incentre and centroid.

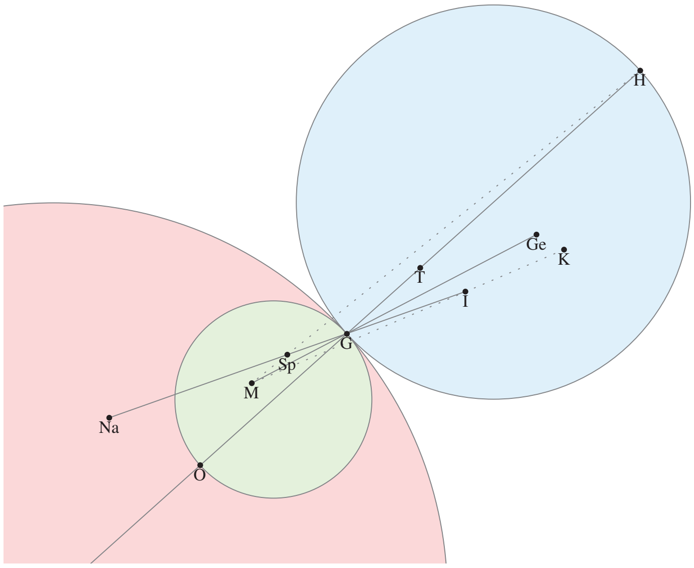
37. If J is the reflection of I in $B\,C$, show that J has representation $v^2+w^2+v\,w+(v+w)\,\frac{v\,w}{u}$, and that $\overrightarrow{I\,J}=\frac{(u+v)\,(v+w)\,(w+u)}{u}$.

38. Hence show that $2\,r=-R\left(\frac{u}{v}+1\right)\left(\frac{v}{w}+1\right)\left(\frac{w}{u}+1\right)$, and thus derive the expression for r in the table.

39. Prove that $OI^2 = R^2 - 2Rr$. **[Euler's identity]**

40. Prove that $IT = \frac{1}{2}R - r$, and thus that the nine-point circle and incircle are tangent. **[Feuerbach's theorem]**

The combination of the two above formulae results in the inequality $OI \geq 2IT$, which means that I lies in the Apollonius disc of diameter GH, known as the *Euler-Apollonius lollipop*. Geoff Smith and Christopher Bradley discovered that the symmedian point and Gergonne point also reside in this disc. As T lies on the line segment GH, it must also inhabit the Euler-Apollonius lollipop.

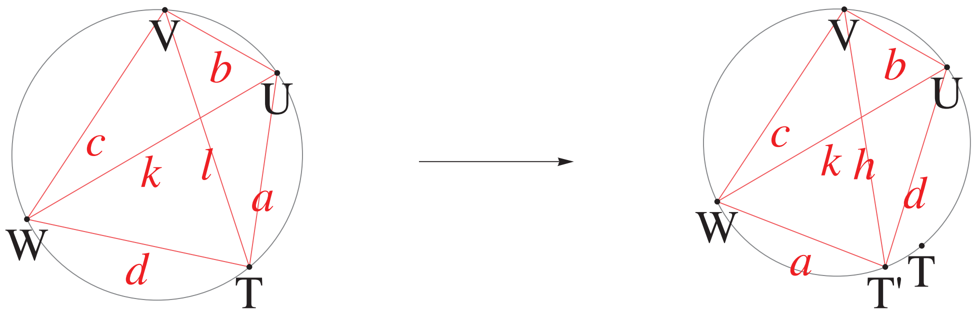


As a consequence of this, together with the Euler line and Nagel line properties, the Spieker centre lies in the disc of diameter OG and therefore outside the Euler-Apollonius lollipop. The lines IK and H Sp intersect at the symmedian point of the excentral triangle, known as the *mittenpunkt* M. M, G, Ge are collinear with ratio $MG : G\text{Ge} = 1 : 2$. The mittenpunkt must therefore reside in the disc of diameter OG.

Indeed, the points shown inside the circles on the diagram above **always** remain in those circles. The red disc containing the Nagel point has diameter GL, where L is the *de Longchamps point* (reflection of H in O).

Cyclic quadrilaterals

Consider an arbitrary cyclic quadrilateral $TUVW$ (labelled anticlockwise). We denote the lengths of edges TU, UV, VW and WU with a, b, c and d, respectively. The lengths of diagonals UW and TV are denoted with k and l, respectively.



If we reflect T in the perpendicular bisector of UW to form T', we obtain a new cyclic quadrilateral $T'UVW$ with the same area and side lengths as $TUVW$, but in a different order. The diagonal UW is unaffected, and remains k. The diagonal $T'V$, however, now has length h, in general distinct from its original length l.

Hence, we can consider $TUVW$ to have *three* diagonal lengths: k and l as well as the invisible diagonal of length h obtainable by interchanging any two adjacent side lengths.

41. Show that $[TUVW] = \frac{(ab+cd)l}{4R} = \frac{hkl}{4R}$. **[Parameshvara's formula]**

Parameshvara's formula is very similar to the formula $[ABC] = \frac{abc}{4R}$. The latter can be regarded as a special case of the former, where two of the vertices of the cyclic quadrilateral are coincident. Similarly, there is a generalisation of Heron's formula applicable to cyclic quadrilaterals:

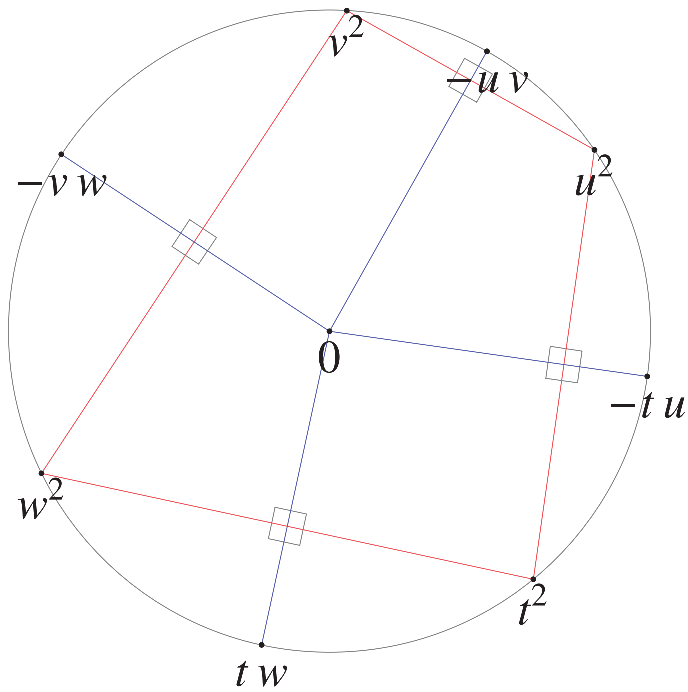
- If a cyclic quadrilateral $TUVW$ has edge lengths a, b, c, d and semiperimeter $s = \frac{1}{2}(a+b+c+d)$, then $[TUVW] = \sqrt{(s-a)(s-b)(s-c)(s-d)}$. **[Brahmagupta's formula]**

Heron's formula is a special case of Brahmagupta's formula, which is in turn a special case of Bretschneider's formula for convex quadrilaterals.

- $[TUVW] = \sqrt{\left((s-a)(s-b)(s-c)(s-d) - abcd\cos^2\left(\frac{T+V}{2}\right)\right)}$. **[Bretschneider's formula]**

The term $abcd\cos^2\left(\frac{T+V}{2}\right) = \frac{1}{4}(ac+bd+kl)(ac+bd-kl)$, where k and l are the diagonal lengths. It is easy to see, by Ptolemy's theorem, that this vanishes when the quadrilateral is cyclic.

By definition, a cyclic quadrilateral is inscribed in a circle, so we can use a related parametrisation to that used for triangles. A cyclic quadrilateral $TUVW$ is represented by complex numbers t^2, u^2, v^2 and w^2, where $tt^* = uu^* = vv^* = ww^* = R$. Unfortunately, the parametrisation is slightly less elegant for cyclic polygons with an even number of sides, as we cannot treat all vertices and edges equivalently.



Notice that the midpoints of the outer arcs of TU, UV and VW are indeed represented by $-tu$, $-uv$ and $-vw$, respectively, as one would imagine. However, due to annoying parity constraints, this forces the midpoint of the outer arc of WT to be **positive** tw. Hence, only the triangles TUV and UVW are correctly parametrised; the others have asymmetric (but equally simple) formulae associated with them. Nevertheless, we can now derive the aforementioned seven lengths and the area via Parameshvara's formula. It is thus straightforward to verify Brahmagupta's formula.

Quantity	Symbol	Expression
Side length of TU	a	$i\,R\left(\frac{u}{t}-\frac{t}{u}\right)$
Side length of UV	b	$i\,R\left(\frac{v}{u}-\frac{u}{v}\right)$
Side length of VW	c	$i\,R\left(\frac{w}{v}-\frac{v}{w}\right)$
Side length of WT	d	$-i\,R\left(\frac{t}{w}-\frac{w}{t}\right)$
Diagonal UW	k	$i\,R\left(\frac{u}{w}-\frac{w}{u}\right)$
Diagonal TV	l	$i\,R\left(\frac{t}{v}-\frac{v}{t}\right)$
Invisible diagonal	h	$i\,R\left(\frac{u}{t}+\frac{v}{w}\right)\left(\frac{t}{u}-\frac{w}{v}\right)$
	$s-a$	$-\frac{1}{2}\,i\,R\left(\frac{t}{v}+1\right)\left(1-\frac{w}{u}\right)\left(\frac{u}{t}+\frac{v}{w}\right)$
	$s-b$	$\frac{1}{2}\,i\,R\left(\frac{t}{v}+1\right)\left(\frac{w}{u}+1\right)\left(\frac{u}{t}-\frac{v}{w}\right)$
	$s-c$	$\frac{1}{2}\,i\,R\left(1-\frac{t}{v}\right)\left(\frac{w}{u}+1\right)\left(\frac{u}{t}+\frac{v}{w}\right)$
	$s-d$	$\frac{1}{2}\,i\,R\left(1-\frac{t}{v}\right)\left(1-\frac{w}{u}\right)\left(\frac{u}{t}-\frac{v}{w}\right)$
(Signed) area	[TUVW]	$-\frac{1}{4}\,i\,R^2\left(\frac{t}{v}-\frac{v}{t}\right)\left(\frac{u}{w}-\frac{w}{u}\right)\left(\frac{u}{t}+\frac{v}{w}\right)\left(\frac{t}{u}-\frac{w}{v}\right)$

42. Let the *maltitude* M_a be the line passing through the midpoint of $T\,U$ and perpendicular to $V\,W$. Define M_b, M_c and M_d similarly. Prove that the four maltitudes are concurrent at a point. **[Anticentre property]**

This concurrency point Q has representation $\frac{1}{2}\left(t^2+u^2+v^2+w^2\right)$, and is known as the *anticentre*. It is obvious from this that the centroid of the four vertices is the midpoint of $O\,Q$.

43. Let the diagonals $T\,V$ and $U\,W$ intersect at P. M and N are the midpoints of $T\,V$ and $U\,W$, respectively. Prove that the anticentre Q is the orthocentre of triangle $M\,N\,P$.

44. Let T' be the orthocentre of $U\,V\,W$, and define U', V' and W' similarly. Prove that $T'\,U'\,V'\,W'$ is congruent to $T\,U\,V\,W$.

45. Let I_T be the incentre of $U\,V\,W$, and define I_U, I_V and I_W similarly. Prove that $I_T\,I_U\,I_V\,I_W$ is a rectangle. **[Japanese theorem for cyclic quadrilaterals]**

46. Let r_T be the inradius of $U\,V\,W$, and define r_U, r_V and r_W similarly. Prove that $r_T+r_V=r_U+r_W$.

47. Suppose we have a cyclic polygon $A_1\,A_2\,\ldots\,A_n$. We draw $n-3$ non-intersecting lines between vertices to dissect the polygon into $n-2$ triangles. Let the sum of the inradii of the triangles be σ. Prove that the value of σ is independent of the choice of lines drawn. **[Japanese theorem for cyclic polygons]**

Solutions

1. We have $\sin\theta\cos\phi+\cos\theta\sin\phi=\frac{\left(e^{i\theta}-e^{-i\theta}\right)\left(e^{i\phi}+e^{-i\phi}\right)+\left(e^{i\theta}+e^{-i\theta}\right)\left(e^{i\phi}-e^{-i\phi}\right)}{4i}=\frac{2\left(e^{i(\theta+\phi)}-e^{-i(\theta+\phi)}\right)}{4i}=\sin(\theta+\phi)$.

2. Similarly, we have $\cos\theta\cos\phi-\sin\theta\sin\phi=\frac{\left(e^{i\theta}+e^{-i\theta}\right)\left(e^{i\phi}+e^{-i\phi}\right)+\left(e^{i\theta}-e^{-i\theta}\right)\left(e^{i\phi}-e^{-i\phi}\right)}{4}=\frac{2\left(e^{i(\theta+\phi)}+e^{-i(\theta+\phi)}\right)}{4}=\cos(\theta+\phi)$.

3. Based on the previous two theorems, $\tan(\theta+\phi)=\frac{\sin(\theta+\phi)}{\cos(\theta+\phi)}=\frac{\sin\theta\cos\phi+\cos\theta\sin\phi}{\cos\theta\cos\phi-\sin\theta\sin\phi}=\frac{\tan\theta+\tan\phi}{1-\tan\theta\tan\phi}$. (The last step is where the numerator and denominator are both divided by $\cos\theta\cos\phi$.)

4. This is a consequence of the double-angle formula for cosine and the identity $\sin^2\theta+\cos^2\theta=1$.

5. $\cos(\theta+\phi)\cos(\theta-\phi)=(\cos\theta\cos\phi-\sin\theta\sin\phi)(\cos\theta\cos\phi+\sin\theta\sin\phi)=\cos^2\theta\cos^2\phi-\sin^2\theta\sin^2\phi$. Using the Pythagorean identity, we obtain $\left(1-\sin^2\theta\right)\cos^2\phi-\left(1-\cos^2\phi\right)\sin^2\theta=\cos^2\phi-\sin^2\theta=\left(\cos^2\phi-1\right)+\left(1-\sin^2\theta\right)=\cos 2\theta+\cos 2\phi$.

6. $\sin 3\theta=\sin 2\theta\cos\theta+\cos 2\theta\sin\theta=2\sin\theta\cos^2\theta+\cos^2\theta\sin\theta-\sin^3\theta=3\left(1-\sin^2\theta\right)\sin\theta-\sin^3\theta$. This clearly expands to $3\sin\theta-4\sin^3\theta$.

7. If we let $\theta=\frac{\pi}{2}-A$, $\phi=\frac{\pi}{2}-B$ and $\psi=\frac{\pi}{2}-C$, then the expression becomes equal to $\sin(\theta+\phi+\psi)$ by the compound angle formula. As $A+B+C=\pi$, it must be the case that $\theta+\phi+\psi=\frac{\pi}{2}$, the sine of which is 1. Dividing by $\sin A\sin B\sin C$ results in the desired equation.

8. $\tan A\tan B\tan C=\tan A\tan B\tan(A+B)=\frac{\tan A+\tan B}{1-\tan A\tan B}\tan A\tan B=\tan A+\tan B+\frac{\tan A+\tan B}{1-\tan A\tan B}=\tan A+\tan B+\tan C$. We can divide by $\tan A\tan B\tan C$ to obtain $\cot A\cot B+\cot B\cot C+\cot C\cot A=1$.

9. The area of triangle IBC is given by $\frac{1}{2}ra$, as r is the height of the triangle when orientated such that a is the base. By symmetry, $[ABC]=\frac{1}{2}ra+\frac{1}{2}rb+\frac{1}{2}rc=rs$. The equivalence of rs to the other terms is evident from considering the similar triangles CIP_A and $CI_C\Upsilon_1$, which provides the identity $\frac{r_c}{s}=\frac{r}{s-c}$.

10. The angle $\angle IAB=\frac{A}{2}$, as I is the intersection of the three angle bisectors. We have $\tan\left(\frac{A}{2}\right)=\frac{IP}{AP}=\frac{r}{s-a}$. The second part of the formula comes from the identity $rs=r_A(s-a)$.

11. Applying the sine rule to triangle AIB gives us $\frac{AI}{\sin\frac{B}{2}}=\frac{AB}{\sin\left(\frac{\pi}{2}+\frac{C}{2}\right)}=\frac{2R\sin C}{\cos\frac{C}{2}}=4R\sin\frac{C}{2}$, with the last step utilising the double-angle formula. Rearranging results in $AI=4R\sin\frac{B}{2}\sin\frac{C}{2}$. The expression for r originates from considering the right-angled triangle AP_CI and applying basic trigonometry.

12. As the internal and external angle bisectors of A are perpendicular, we can quickly deduce that $\angle Q_CAI_C=\frac{\pi}{2}-\frac{A}{2}$ and thus $\angle AI_CQ_C=\frac{A}{2}$. We then have $\tan\frac{A}{2}=\frac{s-b}{r_C}$ from applying basic trigonometry to the right-angled triangle.

13. It is straightforward, from the expressions in the table together with the Sine Rule and double-angle formula, to verify that each term is equal to $\frac{R}{2}\sin A\sin B\sin C$.

14. A combination of the sine rule and cosine rule provides $\cot A = \frac{\cos A}{\sin A} = \frac{2\,R\cos A}{a} = \frac{R(b^2+c^2-a^2)}{a\,b\,c}$. We also have $[A\,B\,C] = \frac{a\,b\,c}{4\,R}$, so $\cot A = \frac{b^2+c^2-a^2}{4[A\,B\,C]}$.

15. $[A\,B\,C] = \frac{a\,b\,c}{4\,R} = \frac{1}{2}\,a\,b\sin C = \frac{1}{2}\,b\,c\sin A = \frac{1}{2}\,c\,a\sin B$. The first expression is equal to the final, penultimate and antepenultimate expressions, so is trivially equal to their geometric mean.

16. $\sin A + \sin B + \sin C = \frac{a}{2\,R} + \frac{b}{2\,R} + \frac{c}{2\,R} = \frac{s}{R} = 4\cos\frac{A}{2}\cos\frac{B}{2}\cos\frac{C}{2}$, using the final expression in the table.

17. Applying the sine rule to triangle $A\,B\,\Omega$, we obtain $\frac{A\,\Omega}{B\,\Omega} = \frac{\sin(B-\omega)}{\sin\omega} = \frac{\sin B\cos\omega - \cos B\sin\omega}{\sin\omega} = (\cot\omega - \cot B)\sin B$. The cyclic product tells us that $(\cot\omega - \cot B)(\cot\omega - \cot C)(\cot\omega - \cot A)\sin A\sin B\sin C = 1$, from which we obtain the desired identity by dividing throughout by $\sin A\sin B\sin C$.

18. Expanding the previous identity results in $\cot^3\omega - (\cot A + \cot B + \cot C)\cot^2\omega + (\cot A\cot B + \cot B\cot C + \cot C\cot A)\cot\omega - \cot A\cot B\cot C = \operatorname{cosec} A\operatorname{cosec} B\operatorname{cosec} C$. Using the identities proved in previous questions, this simplifies to $\cot^3\omega - (\cot A + \cot B + \cot C)\cot^2\omega + \cot\omega - (\cot A + \cot B + \cot C) = 0$.

19. The cubic factorises to $(x - \cot A - \cot B - \cot C)(x - i)(x + i) = 0$. As it is impossible for $\cot\omega$ to be imaginary, it must instead be $\cot A + \cot B + \cot C$.

20. $\tan\omega = \frac{1}{\cot\omega} = \frac{1}{\cot A + \cot B + \cot C} = \frac{\sin A\sin B\sin C}{\sin A\sin B\cos C + \sin A\cos B\sin C + \cos A\sin B\sin C} = \frac{\sin A\sin B\sin C}{1 + \cos A\cos B\cos C}$.

21. $\cot\omega = \cot A + \cot B + \cot C = \frac{a^2+b^2-c^2}{4[A\,B\,C]} + \frac{b^2+c^2-a^2}{4[A\,B\,C]} + \frac{c^2+a^2-b^2}{4[A\,B\,C]} = \frac{a^2+b^2+c^2}{4[A\,B\,C]}$.

22. Proving this is equivalent to showing that $\cot\omega \geq \sqrt{3}$, or $a^2 + b^2 + c^2 \geq 4\sqrt{3}\,[A\,B\,C]$. Square both sides and apply Heron's formula, giving the equivalent inequality $a^4 + b^4 + c^4 + 2\,a^2\,b^2 + 2\,b^2\,c^2 + 2\,c^2\,a^2 \geq 3\left(2\,a^2\,b^2 + 2\,b^2\,c^2 + 2\,c^2\,a^2 - a^4 - b^4 - c^4\right)$. Rearranging and dividing by 4 gives $a^4 + b^4 + c^4 \geq a^2\,b^2 + b^2\,c^2 + c^2\,a^2$, which (by Muirhead's inequality) is true, with equality if and only if $a = b = c$.

23. P must lie in either triangle $A\,\Omega\,B$, $B\,\Omega\,C$ or $C\,\Omega\,A$. Without loss of generality, assume the former. Then, we have $\angle P\,A\,B \leq \angle\Omega\,A\,B < \frac{\pi}{6}$ from the previous question.

24. We clearly have $a + b + c = 2\,s$. As $[A\,B\,C] = r\,s = \frac{a\,b\,c}{4\,R}$, we obtain $a\,b\,c = 4\,R\,r\,s$. By squaring Heron's formula and dividing throughout by s, we get $(s-a)(s-b)(s-c) = r^2\,s$. Polynomial expansion results in $s^3 - (a+b+c)\,s^2 + (a\,b + b\,c + c\,a)\,s - a\,b\,c = r^2\,s$. The other symmetric polynomials can be replaced by the expressions in R, r and s, yielding $s^3 - 2\,s^3 + (a\,b + b\,c + c\,a)\,s - 4\,R\,r\,s = r^2\,s$. Final manipulation and division by s culminates in the expression for the third symmetric polynomial, $a\,b + b\,c + c\,a = s^2 + r^2 + 4\,R\,r$.

25. Schur's inequality provides $l(l-m)(l-n) + m(m-n)(m-l) + n(n-l)(n-m) \geq 0$. We can then expand to obtain $l^3 + m^3 + n^3 + 3\,l\,m\,n \geq l^2\,m + m^2\,l + m^2\,n + n^2\,m + n^2\,l + l^2\,n$. Adding $l^3 + m^3 + n^3$ to each side of the equation yields $2\left(l^3 + m^3 + n^3\right) + 3\,l\,m\,n \geq (l+m+n)\left(l^2 + m^2 + n^2\right)$. Replacing each of these symmetric polynomials with their R, r and s counterparts gives $2\left(s^3 - 12\,R\,r\,s\right) + 3\,r^2\,s \geq s\left(-2\,r^2 - 8\,R\,r + s^2\right)$. Rearranging, we obtain $s^3 - 16\,R\,r\,s + 5\,r^2\,s \geq 0$. Further rearrangement and division by s gives the required inequality, $s^2 + 5\,r^2 \geq 16\,R\,r$. Equality occurs if and only if the triangle is equilateral.

26. We can convert the expression into an inequality in l, m, n. Firstly, we derive the expressions $r^2 = \frac{l\,m\,n}{l+m+n}$,

$4R^2 = \frac{(l+m)^2 (m+n)^2 (n+l)^2}{4 l m n(l+m+n)}$ and $4Rr = lm + mn + nl - r^2$. By multiplying the inequality by $4lmn(l+m+n)$, we obtain the equivalent inequality $4lmn(l+m+n)^3 \le (l+m)^2 (m+n)^2 (n+l)^2 + 4lmn((lm+mn+nl)(l+m+n) + 2lmn)$. It is now helpful to apply the uvw method to express it as $w^6 + \left(2uv^2 - 12u^3\right) w^3 + 9u^2 v^4 \ge 0$. This is a quadratic in w^3, so we only need to check three cases by Tejs' corollary. The third case only occurs when $F w^3 + G = 0$, or $w^3 + uv^2 - 6u^3 = 0$. The expression is negative since $u \ge v \ge w$, so this cannot occur. Hence, we only need to consider when $l = m$ or $n = 0$. In the latter, we have a degenerate triangle comprising three collinear points, and thus $r = 0$, $2R = s$; this satisfies the inequality. In the former case, the triangle is isosceles and the inequality reduces to $4l^2 n(2l+n)^3 \le 4l^2(l+n)^4 + 4l^2 n\left(\left(l^2 + 2ln\right)(2l+n) + 2l^2 n\right)$. Expanding out gives the equivalent inequality $4l^6 - 8l^5 n + 4l^4 n^2 \ge 0$, or $4l^4(l-n)^2 \ge 0$. This is trivially true.

27. With Muirhead's inequality, it is trivial to verify that $(lm + mn + nl)(l+m+n) \ge 9lmn$. Expanding each term gives $\left(r^2 + 4Rr\right) s \ge 9r^2 s$, which simplifies to $4Rrs \ge 8r^2 s$. Dividing throughout by $4rs$ yields the desired inequality.

28. $[OAB] = \frac{1}{2} R^2 \sin 2C$, so we have $[ABC] = \frac{1}{2} R^2(\sin 2A + \sin 2B + \sin 2C)$. Hence, the left-hand side of the inequality is equal to $\frac{s}{R}$, and the right-hand side is equal to $\frac{2[ABC]}{R^2} = \frac{2rs}{R^2}$. So, we need to prove $\frac{s}{R} \ge \frac{2rs}{R^2}$, which simplifies to Euler's inequality, $R \ge 2r$.

29. $\cot\omega = \frac{a^2+b^2+c^2}{4[ABC]} = \frac{s^2-r^2-4Rr}{2rs}$.

30. $\cos A + \cos B + \cos C = \sum_{\text{cyc}} \frac{b^2+c^2-a^2}{2bc} = \sum_{\text{cyc}} \frac{ab^2+ac^2-a^3}{2abc} = \frac{\left(a^2+b^2+c^2\right)(a+b+c)-2\left(a^3+b^3+c^3\right)}{2abc}$. We can now apply the formulae to convert this expression to $\frac{4s\left(s^2-r^2-4Rr\right)-4s\left(s^2-3r^2-6Rr\right)}{8Rrs}$. Some cancellation results in $\frac{2r^2+2Rr}{2Rr}$, which further simplifies to $1 + \frac{r}{R}$. We already have $r = 4R\sin\frac{A}{2}\sin\frac{B}{2}\sin\frac{C}{2}$, so we are done.

31. The previous equation works for all triangles, and thus, due to the analytic continuity of sine and cosine, works for all angles such that $A + B + C = \pi$. If we use the angles $A + 2\pi$, $B - \pi$ and $C - \pi$, we obtain the equation $\cos A - \cos B - \cos C = 1 + \sin\left(\frac{A}{2} + \pi\right)\sin\left(\frac{B}{2} - \frac{\pi}{2}\right)\sin\left(\frac{C}{2} - \frac{\pi}{2}\right)$, the right-hand side of which simplifies to $1 - \sin\frac{A}{2}\cos\frac{B}{2}\cos\frac{C}{2}$.

32. Note that $\cos A + \cos B + \cos C = 1 + \frac{r}{R}$ and $\cos A - \cos B - \cos C = 1 - \frac{r_A}{R}$ by the previous two results. We add the first equation to the cyclic sum of the second equation, yielding $0 = 4 + \frac{r}{R} - \frac{r_A}{R} - \frac{r_B}{R} - \frac{r_C}{R}$. Multiplying throughout by R and rearranging gives $4R = r_A + r_B + r_C - r$.

33. The magnitude of uv is the same as u^2 and v^2, namely R. We also have $2\arg(uv) = \arg\left(u^2 v^2\right) = \arg\left(u^2\right) + \arg\left(v^2\right)$. Hence, it must be the bisector of one of the arcs, namely the arc containing C (as we have defined it as such).

34. The angle $\angle I A I_C$ is a right-angle. By symmetry, I must be the orthocentre of $I_A I_B I_C$, and thus ABC is its orthic triangle. The circumcircle of the orthic triangle is the nine-point circle.

35. $-uv$ lies on the angle bisector $I I_C$, due to the 'equal arcs subtend equal angles' property. As it lies on the nine-point circle of $I I_A I_B I_C$, it must be the Euler point (midpoint of $I I_C$). The nine-point centre of the excentric triangle is O, the circumcentre of the reference triangle. It is also the barycentre of I, I_A, I_B and I_C, so must be the midpoint of the line joining the midpoints of $I I_C$ with $I_A I_B$. This enables us to deduce that uv is indeed the midpoint of $I_A I_B$.

36. In complex coordinates, we have $I_A = (I_A + I_B + I_C) - (I_B + I_C) = \frac{1}{2}(I_A + I_B) - \frac{1}{2}(I_B + I_C) + \frac{1}{2}(I_C + I_A) = u\,v - v\,w + w\,u$, thence it is trivial to find the representations of the other excentres and incentre.

37. Using the formula for the reflection of a point in a chord, we have that J has representation $v^2 + w^2 + \frac{v^2 w^2 (v^* w^* + w^* u^* + u^* v^*)}{R^2} = v^2 + w^2 + v\,w + \frac{v^2 w}{u} + \frac{w^2 v}{u}$. Hence, $\overrightarrow{I\,J}$ has representation $v^2 + w^2 - v\,w - \frac{v^2 w}{u} - \frac{w^2 v}{u} + u\,v + v\,w + w\,u$. Multiplying throughout by u yields the symmetric expression $u^2 v + v^2 u + v^2 w + w^2 v + w^2 u + u^2 w + 2\,u\,v\,w$. This factorises to $(u + v)(v + w)(w + u)$, so the original expression is $\frac{(u+v)(v+w)(w+u)}{u}$.

38. Multiplying the expression for $\overrightarrow{I\,J}$ by its complex conjugate gives $4\,r^2 = \frac{(u+v)(u^*+v^*)(v+w)(v^*+w^*)(w+u)(w^*+u^*)}{R} = R^2\left(\frac{u}{v} + 1\right)\left(1 + \frac{u}{v}\right)\left(\frac{v}{w} + 1\right)\left(1 + \frac{v}{w}\right)\left(\frac{w}{u} + 1\right)\left(1 + \frac{u}{w}\right)$. The square root of this is thus the distance $2\,r$.

39. We have $O\,I^2 = (u\,v + v\,w + w\,u)(u^* v^* + v^* w^* + w^* u^*) = R(3\,R + u\,w^* + w\,u^* + v\,u^* + u\,v^* + w\,v^* + v\,w^*)$. Combined with the expression for r given in the table, this equals $R(R - 2\,r)$, as required.

40. $\overrightarrow{I\,T} = \frac{1}{2}\left(u^2 + v^2 + w^2\right) + u\,v + v\,w + w\,u = \frac{1}{2}(u + v + w)^2$. It is obvious that the modulus is $\frac{1}{2}(u + v + w)(u^* + v^* + w^*) = \frac{1}{2}(3\,R + u\,w^* + w\,u^* + v\,u^* + u\,v^* + w\,v^* + v\,w^*) = \frac{1}{2}\,R - r$.

41. $[T\,U\,V\,W] = [T\,U\,V] + [V\,W\,T] = \frac{a\,b\,l}{4\,R} + \frac{c\,d\,l}{4\,R} = \frac{(a\,b + c\,d)\,l}{4\,R}$. By Ptolemy's theorem on the quadrilateral $T'\,U\,V\,W$ (where T' is the reflection of T in the perpendicular bisector of $U\,W$), we have $a\,b + c\,d = l$, giving us Parameshvara's formula.

42. The maltitude passes through $\frac{1}{2}\left(u^2 + t^2\right)$, and travels parallel to the vector $\frac{1}{2}\left(v^2 + w^2\right)$. It is clear that $\frac{1}{2}\left(t^2 + u^2 + v^2 + w^2\right)$ lies on this maltitude, and thus all four maltitudes by symmetry.

43. If we consider the other two maltitudes (from the midpoint of each diagonal perpendicular to the other diagonal), they must also pass through Q. By definition, they also pass through the orthocentre of $M\,N\,P$, so Q must be this orthocentre.

44. T' has representation $u^2 + v^2 + w^2$. This is the reflection of T in the anticentre Q. Hence, $T'\,U'\,V'\,W'$ is congruent and homothetic to the original with Q as the centre of similitude.

45. Let I_U be the incentre of $V\,W\,T$, *et cetera*. We have $I_u = -v\,w + t\,w + v\,t$, $I_T = -u\,v - v\,w - w\,u$, and $I_W = -t\,u - u\,v - v\,t$. We wish to prove that $\angle\, I_T\,I_U\,I_V = \frac{\pi}{2}$, which is equivalent to $\frac{I_V - I_U}{I_T - I_U} = \frac{(w-t)(u+v)}{(t+u)(w+v)}$ being purely imaginary. We know that t, u, v and w have equal modulus, so must be concyclic. Hence, $\frac{(w+t)(u+v)}{(t+u)(v+w)}$ is real, and we only need to prove that $\frac{w-t}{w+t} = \frac{(w-t)(w^*+t^*)}{(w+t)(w^*+t^*)}$ is imaginary. The numerator is $t^*\,w - w^*\,t$, which is equal to the negative of its conjugate and is therefore imaginary. Similarly, the denominator is equal to its conjugate and therefore real. Hence, we are done and $\angle\, I_T\,I_U\,I_V = \frac{\pi}{2}$; by symmetry, $I_T\,I_U\,I_V\,I_W$ is a rectangle.

46. For any point P in the plane of a rectangle $A\,B\,C\,D$, we have $A\,P^2 + C\,P^2 = B\,P^2 + D\,P^2$. This can be derived from assuming P is the origin and orienting the rectangle parallel to the coordinate axes and using Cartesian coordinates. Applying this to the point O (circumcentre of $T\,U\,V\,W$) and the rectangle of incentres, we obtain $O\,I_T{}^2 + O\,I_V{}^2 = O\,I_U{}^2 + O\,I_W{}^2$. We can determine each of these squared distances from Euler's formula, obtaining the equation $\left(R^2 - 2\,R\,r_T\right) + \left(R^2 - 2\,R\,r_V\right) = \left(R^2 - 2\,R\,r_U\right) + \left(R^2 - 2\,R\,r_W\right)$. Cancelling terms and dividing throughout by R yields the desired equation.

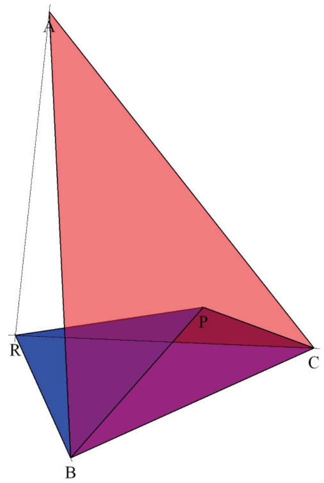
47. This is true for $n = 4$, by the previous question. We prove the general case by induction on n, assuming it is true for all $n \leq k$, and proving it for $n = k + 1$. Note that for every triangulation, there must be at least one triangle with three adjacent vertices. Suppose we have a triangulation where $A_1 A_2 A_3$ is one of the triangles. Then we can 're-triangulate' the k-gon $A_3 A_4 \ldots A_{k+1} A_1$ such that $A_1 A_3 A_4$ is also a triangle. Now, re-triangulate the cyclic quadrilateral $A_1 A_2 A_3 A_4$, so that $A_2 A_3 A_4$ is a triangle. Repeating this process, we can ensure that $A_{i-1} A_i A_{i+1}$ is a triangle for any i (with subscripts considered modulo $k + 1$). Arbitrarily re-triangulating the k-gon $A_{i+1} A_{i+2} \ldots A_{i-2} A_{i-1}$ gives any possible triangulation of the $(k + 1)$-gon. As we did not affect σ during any of the re-triangulations of the k-gons, σ has remained constant throughout the whole process.

Areal coordinates

A large quantity of problems are concerned with a triangle ABC, known as the reference triangle. It is particularly useful, in these instances, to apply a special type of projective homogeneous coordinates, namely *areal coordinates*. The vertices A, B and C are given by the coordinates $(1, 0, 0)$, $(0, 1, 0)$ and $(0, 0, 1)$, respectively. The line at infinity is given by $x + y + z = 0$. This exploits the symmetry of the triangle in a way that Cartesian coordinates do not.

Areas and lines

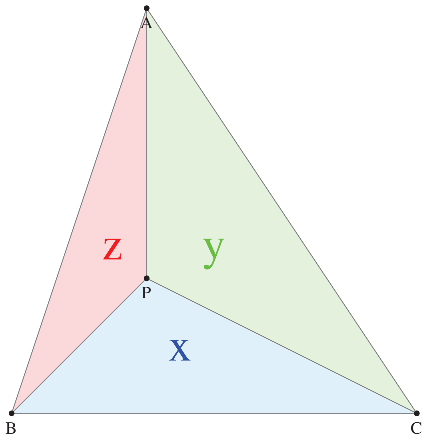
We have already defined areal coordinates as a special case of projective homogeneous coordinates. However, there are several other equivalent definitions explored later, explaining the synonyms 'areal' and 'barycentric'.



We can *normalise* the areal coordinates (x, y, z) in the plane (not on the line at infinity) by assuming that $x + y + z = 1$. To convert unnormalised areals into their normalised counterparts, simply apply the map $(x, y, z) \to \frac{(x,y,z)}{x+y+z}$.

1. If the point P is represented by normalised areal coordinates (x, y, z), prove that $x = \frac{[PBC]}{[ABC]}$. *(Hint: consider the volumes of tetrahedra $RABC$ and $RPBC$.)*

This gives us one definition of areal coordinates, namely the ratio between the areas $[PBC]$, $[PCA]$ and $[PAB]$. If the triangle $[ABC]$ has unit area, then the areas of these triangles are equal to the normalised areal coordinates. This is encapsulated by the following diagram from Tom Lovering's excellent introduction to areal coordinates (available at http://www.bmoc.maths.org/home/areals.pdf).



As areal coordinates can be defined in terms of ratios of areas of triangles (which are unchanged by affine transformations), the areal coordinates of a point remain invariant when an affine transformation is applied.

2. Deduce that the lines BC, CA and AB correspond to the equations $x = 0$, $y = 0$ and $z = 0$, respectively.

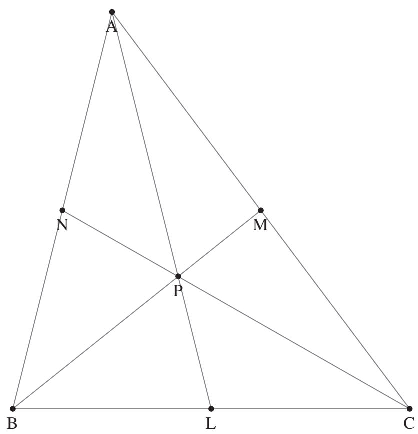
3. Show that the centroid, G, has normalised areal coordinates $\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$.

4. Let the points P, Q and S be represented by normalised areal coordinates (x_1, y_1, z_1), (x_2, y_2, z_2) and (x_3, y_3, z_3), respectively. Show that $\det\begin{pmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{pmatrix} = \frac{[PQS]}{[ABC]}$.

Ceva's theorem and cevians

5. Show that the equation of the line AP, where $P = (x_1, y_1, z_1)$, is given by $y_1 z = z_1 y$. Hence find the coordinates of the intersection point, $L = AP \cap BC$.

The line AP is known as a *cevian* through A, named after Ceva's theorem. This can easily be proved using areal coordinates.



6. Let L, M and N lie on sides BC, CA and AB, respectively, of a triangle ABC. Show that AL, BM and CN are concurrent if and only if $\frac{\overrightarrow{BL}}{\overrightarrow{LC}} \cdot \frac{\overrightarrow{CM}}{\overrightarrow{MA}} \cdot \frac{\overrightarrow{AN}}{\overrightarrow{NB}} = 1$. **[Ceva's theorem]**

7. Suppose we have a point P, and draw the three cevians through it to meet the sides BC, CA and AB at L, M and N, respectively. Reflect L in the perpendicular bisector of BC to obtain L', and define M' and N' similarly. Prove that AL', BM' and CN' are concurrent. **[Existence of isotomic conjugates]**

8. Let L, M and N lie on sides BC, CA and AB, respectively, of a triangle ABC. Show that AL, BM and CN are concurrent if and only if $\frac{\sin\angle LAB}{\sin\angle CAL}\cdot\frac{\sin\angle MBC}{\sin\angle ABM}\cdot\frac{\sin\angle NCA}{\sin\angle BCN}=1$. **[Trigonometric Ceva's theorem]**

9. Suppose we have a point P, and draw the three cevians through it. Reflect the cevian through A in the line AI, and repeat for the other two cevians. Prove that these three new lines are concurrent. **[Existence of isogonal conjugates]**

In unnormalised areal coordinates, the isotomic conjugate of (x, y, z) is given by $(1/x, 1/y, 1/z)$ and the isogonal conjugate of (x, y, z) is $\left(a^2/x, b^2/y, c^2/z\right)$. The *symmedian point* (intersection of the reflections of the medians in the corresponding angle bisectors) is defined as the isogonal conjugate of the centroid, giving it unnormalised areal coordinates $\left(a^2, b^2, c^2\right)$. Here are the unnormalised coordinates of common triangle centres:

Point	*Unnormalised* areal coordinates of point		
	x	y	z
Vertex A	1	0	0
Centroid	1	1	1
Incentre	a	b	c
Excentre opposite A	$-a$	b	c
Nagel point	$s-a$	$s-b$	$s-c$
Gergonne point	r_A	r_B	r_C
Symmedian point	a^2	b^2	c^2
Circumcentre	$\sin(2A)$	$\sin(2B)$	$\sin(2C)$
Orthocentre	$\tan(A)$	$\tan(B)$	$\tan(C)$
Nine−point centre	$\sin(2B)+\sin(2C)$	$\sin(2A)+\sin(2C)$	$\sin(2A)+\sin(2B)$
First Brocard point	$\frac{1}{b^2}$	$\frac{1}{c^2}$	$\frac{1}{a^2}$
Second Brocard point	$\frac{1}{c^2}$	$\frac{1}{a^2}$	$\frac{1}{b^2}$

The circumcentre, orthocentre and nine-point centre also have non-trigonometric forms (expressed in terms of a^2, b^2 and c^2 alone). However, they are even more complicated than the trigonometrical expressions here, so their practical use is unrecommended. If these points are involved, it is better to use the parametrisation involving complex numbers.

By analogy with the first and second Brocard points, the triangle centre with areal coordinates $\left(\frac{1}{a^2}, \frac{1}{b^2}, \frac{1}{c^2}\right)$ is known as the *third Brocard point*. Apart from being the isotomic conjugate of the symmedian point, it is completely boring.

10. Using the formulae for isogonal conjugates, prove that the incentre and excentres indeed have the coordinates shown in the table.

11. Prove that the orthocentre has normalised coordinates $(\cot B\cot C,\ \cot C\cot A,\ \cot A\cot B)$.

12. Show that the circumcentre and orthocentre are isogonal conjugates.

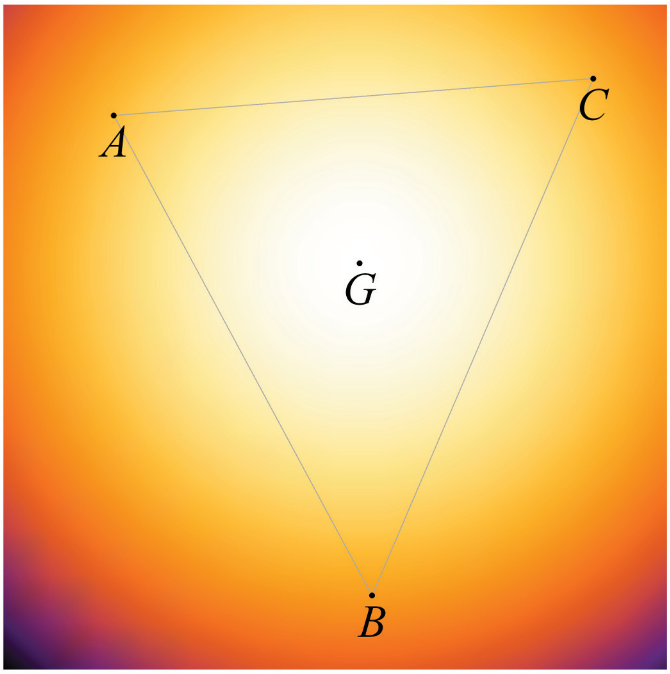
Barycentres and Huygens-Steiner

Another interpretation of the point with coordinates (x, y, z) is the *barycentre* (centre of mass) of the system where masses of x, y and z are placed at the vertices A, B and C, respectively. Hence, areal coordinates are occasionally known as *barycentric coordinates*.

13. Suppose we have a set S of masses in the plane, with total mass $m_1 + m_2 + \cdots + m_n = 1$. The mass m_i is located at the point A_i, and the barycentre is denoted P. For any point Q in the plane, define the *weighted mean square distance* $\mathbb{M}(S, Q) = \sum m_i (A_i Q)^2$. Prove that $P Q^2 = \mathbb{M}(S, Q) - \mathbb{M}(S, P)$. **[Huygens-Steiner theorem]**

This theorem is named after Jakob Steiner and Christiaan Huygens. The latter is famous for inventing the pendulum clock, proposing a wave theory of light, and discovering Titan (the largest of Saturn's many moons) with a telescope he built.

There are equivalent formulations of the Huygens-Steiner theorem in mechanics (the parallel-axis theorem) and statistics ($\sigma^2 = E(X^2) - E(X)^2$).



As a corollary of this theorem, the centroid of a set of points minimises the sum of squared distances to each of those points. This is demonstrated by the density plot of the function $A P^2 + B P^2 + C P^2$, which has a global minimum at $P = G$.

By repeated application of Huygens-Steiner, we can determine the weighted mean square distance between two sets. The ordinary version can be regarded as the case where one of the sets has a single element.

- Suppose we have two sets, S_1 and S_2, each with unit total mass. Every mass $m_i \in S_1$ is located at the point A_i; every mass $n_j \in S_2$ is located at the point B_j. The barycentres of S_1 and S_2 are denoted P_1 and P_2, respectively. We define the *weighted mean square distance* $\mathbb{M}(S_1, S_2) = \sum m_i n_j (A_i B_j)^2$. Then we have $P_1 P_2{}^2 + \mathbb{M}(S_1, P_1) + \mathbb{M}(S_2, P_2) = \mathbb{M}(S_1, S_2)$. **[Generalised Huygens-Steiner theorem]**

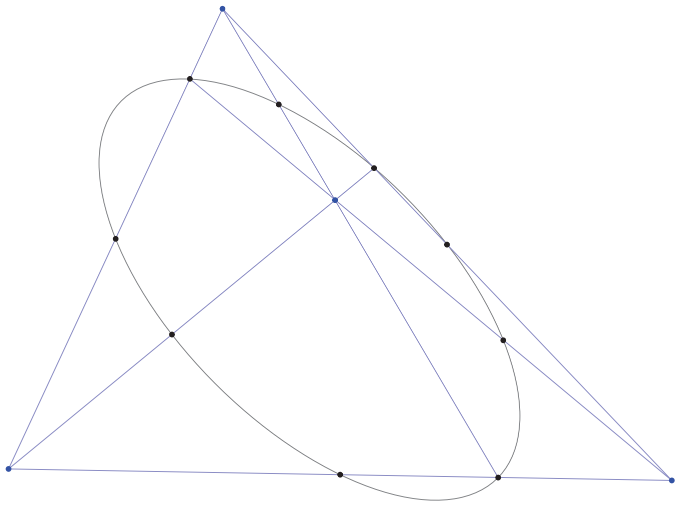
It is particularly relevant to our exploration of barycentric coordinates to consider Huygens-Steiner where $n = 3$ and the masses are positioned at the vertices of the reference triangle $A B C$.

14. If O is the circumcentre of the reference triangle ABC and P has normalised areal coordinates (x, y, z), show that $OP^2 = R^2 - \left(x\,AP^2 + y\,BP^2 + z\,CP^2\right)$.

15. Hence show that $OG^2 = R^2 - \frac{1}{9}\left(a^2 + b^2 + c^2\right)$. **[Circumcentre-centroid distance]**

16. Prove that $OI^2 = R^2 - 2Rr$. **[Euler's formula]**

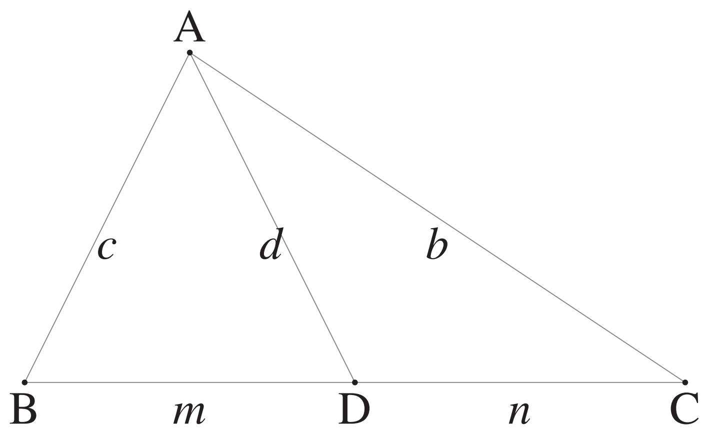
17. For every $n \geq 3$, determine all the configurations of n distinct points $X_1, X_2, \ldots, X_n$ in the plane, with the property that for any pair of distinct points X_i, X_j there exists a permutation σ of the integers $\{1, 2, \ldots, n\}$ such that $d(X_i, X_k) = d\left(X_j, X_{\sigma(k)}\right)$ for all $k \in \{1, 2, \ldots, n\}$, where $d(A, B)$ denotes the distance between A and B. **[RMM 2011, Question 5, Alexander (formerly, at the time he composed the problem, known as Luke) Betts]**



18. A quadrilateral $ABCD$ is drawn in the plane. Show that the midpoints of the four sides, midpoints of the two diagonals, intersections of opposite sides, and intersection of the diagonals all lie on a single conic. Show further that this conic cannot be a parabola. **[Nine-point conic]**

Distance geometry

19. In a triangle ABC (with the side lengths labelled in the usual way), we choose a point D on BC such that $BD = m$, $CD = n$ and $AD = d$. Prove that $man + dad = bmb + cnc$. **[Stewart's theorem]**



Stewart's theorem is particularly attractive, as it is defined solely in terms of distances and nothing else. It can be derived through simple application of the cosine rule; however, the derivation using the Huygens-Steiner theorem remains firmly within the realms of distance geometry. The statement of Stewart's theorem can be remembered with the mnemonic 'a man and his dad put a bomb in the sink'.

If $m = n$, then Stewart's theorem reduces to a special case called *Apollonius' theorem*.

- Suppose we have a triangle $A\,B\,C$, where M is the midpoint of $B\,C$. Then $A\,M^2 = \frac{1}{2}\,b^2 + \frac{1}{2}\,c^2 - \frac{1}{4}\,a^2$. **[Apollonius' theorem]**

A much more impressive theorem in distance geometry is that of the Cayley-Menger determinant. If $A_1\,A_2\,\ldots\,A_n\,A_{n+1}$ is a n-simplex with volume V, then the following identity applies.

- $-(-2)^n\,(n\,!)^2\,V^2 = \det\begin{pmatrix} 0 & 1 & 1 & 1 & \cdots & 1 \\ 1 & 0 & A_1\,A_2{}^2 & A_1\,A_3{}^2 & \cdots & A_1\,A_{n+1}{}^2 \\ 1 & A_2\,A_1{}^2 & 0 & A_2\,A_3{}^2 & \cdots & A_2\,A_{n+1}{}^2 \\ 1 & A_3\,A_1{}^2 & A_3\,A_2{}^2 & 0 & \cdots & A_3\,A_{n+1}{}^2 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & A_{n+1}\,A_1{}^2 & A_{n+1}\,A_2{}^2 & A_{n+1}\,A_3{}^2 & \cdots & 0 \end{pmatrix}$. **[Cayley-Menger determinant]**

For $n = 2$, this is equivalent to Heron's formula. For $n = 3$, this is known as *Tartaglia's formula* (remember that angry guy who solved the cubic equation?) for the volume of a tetrahedron. Equating this to zero gives an equation relating the squared distances between four coplanar points, which can itself be considered to be a generalisation of Stewart's theorem.

20. Prove that $R^2 = \frac{a^2\,b^2\,c^2}{(a+b+c)\,(a+b-c)\,(a-b+c)\,(-a+b+c)}$.

Circles in areal coordinates

21. Let $A\,B\,C$ be the reference triangle, with side lengths $B\,C = a$, $C\,A = b$ and $A\,B = c$. Show that if two points have normalised areal coordinates $P = (x_1,\ y_1,\ z_1)$ and $Q = (x_2,\ y_2,\ z_2)$, then $P\,Q^2 = -a^2\,v\,w - b^2\,w\,u - c^2\,u\,v$, where $u = x_1 - x_2$, $v = y_1 - y_2$ and $w = z_1 - z_2$. **[Areal distance formula]**

By considering a circle to be the locus of points of a particular distance from a given point, we obtain the general formula for a circle.

- A circle has the equation $a^2\,y\,z + b^2\,z\,x + c^2\,x\,y + (x + y + z)\,(A\,x + B\,y + C\,z) = 0$, where A, B and C are constants. **[Equation of a circle]**

The $(x + y + z)$ bracket is included to make the equation homogeneous, so that it is compatible with unnormalised coordinates. From this, we obtain the equation of a circle through three given points.

- The circle through points $P = (x_1, y_1, z_1)$, $Q = (x_2, y_2, z_2)$ and $R = (x_3, y_3, z_3)$ is given by equating the determinant of the following matrix to zero: **[Concyclicity condition]**

$$\begin{pmatrix} x(x+y+z) & x_1(x_1+y_1+z_1) & x_2(x_2+y_2+z_2) & x_3(x_3+y_3+z_3) \\ y(x+y+z) & y_1(x_1+y_1+z_1) & y_2(x_2+y_2+z_2) & y_3(x_3+y_3+z_3) \\ z(x+y+z) & z_1(x_1+y_1+z_1) & z_2(x_2+y_2+z_2) & z_3(x_3+y_3+z_3) \\ a^2\,y\,z+b^2\,z\,x+c^2\,x\,y & a^2\,y_1\,z_1+b^2\,z_1\,x_1+c^2\,x_1\,y_1 & a^2\,y_2\,z_2+b^2\,z_2\,x_2+c^2\,x_2\,y_2 & a^2\,y_3\,z_3+b^2\,z_3\,x_3+c^2\,x_3\,y_3 \end{pmatrix}$$

This is itself a special case of a variant of Goucher's theorem applicable to areal coordinates.

- Let $S = (x_4, y_4, z_4)$, $P = (x_1, y_1, z_1)$, $Q = (x_2, y_2, z_2)$ and $R = (x_3, y_3, z_3)$. Then the determinant of the following matrix:

$$\begin{pmatrix} x_4(x_4+y_4+z_4) & x_1(x_1+y_1+z_1) & x_2(x_2+y_2+z_2) & x_3(x_3+y_3+z_3) \\ y_4(x_4+y_4+z_4) & y_1(x_1+y_1+z_1) & y_2(x_2+y_2+z_2) & y_3(x_3+y_3+z_3) \\ z_4(x_4+y_4+z_4) & z_1(x_1+y_1+z_1) & z_2(x_2+y_2+z_2) & z_3(x_3+y_3+z_3) \\ a^2\,y_4\,z_4+b^2\,z_4\,x_4+c^2\,x_4\,y_4 & a^2\,y_1\,z_1+b^2\,z_1\,x_1+c^2\,x_1\,y_1 & a^2\,y_2\,z_2+b^2\,z_2\,x_2+c^2\,x_2\,y_2 & a^2\,y_3\,z_3+b^2\,z_3\,x_3+c^2\,x_3\,y_3 \end{pmatrix}$$

is equal to $(x_1+y_1+z_1)^2\,(x_2+y_2+z_2)^2\,(x_3+y_3+z_3)^2\,(x_4+y_4+z_4)^2\,\frac{[P\,Q\,R]}{[A\,B\,C]}\,\text{Power}(S,\ P\,Q\,R)$. **[Goucher's theorem for areal coordinates]**

22. Show that the power of a point $P = (x, y, z)$ (in normalised areal coordinates) with respect to the circumcircle of the reference triangle $A\,B\,C$ is given by $\text{Power}(P, A\,B\,C) = -a^2\,y\,z - b^2\,z\,x - c^2\,x\,y$. **[Power with respect to circumcircle]**

We can combine this with Huygens-Steiner to yield the following equation:

- $R^2 - O\,P^2 = a^2\,y\,z + b^2\,z\,x + c^2\,x\,y = x\,A\,P^2 + y\,B\,P^2 + z\,C\,P^2$, where $P = (x, y, z)$ in normalised areals.

This enables us to calculate the distances between the circumcentre and several other points.

23. Hence show that $R^2 - O\,I^2 = 2\,R\,r$. **[Euler's formula]**

24. Prove similarly that $R^2 - O\,{I_A}^2 = -2\,R\,r_A$. **[Excentral analogue of Euler's formula]**

25. Hence prove that $O\,I^2 + O\,{I_A}^2 + O\,{I_B}^2 + O\,{I_C}^2 = 12\,R^2$.

26. Demonstrate also that $R^2 - O\,H^2 = 4[A\,B\,C]\cot A\cot B\cot C = 8\,R^2\cos A\cos B\cos C$. **[Power of the orthocentre]**

As areal coordinates are projective homogeneous coordinates, conics have the general form $A\,x^2 + B\,y^2 + C\,z^2 + D\,y\,z + E\,z\,x + F\,x\,y = 0$, where A, B, C, D, E and F are constants. It is easy to see that a circle is thus a special case of a conic.

27. Let P be chosen randomly in the interior of triangle $A\,B\,C$, such that equal areas have equal probabilities of containing P. Find the probability that $\sqrt{[A\,B\,P]} \geq \sqrt{[B\,C\,P]} + \sqrt{[C\,A\,P]}$. **[Adapted from RMM 2008]**

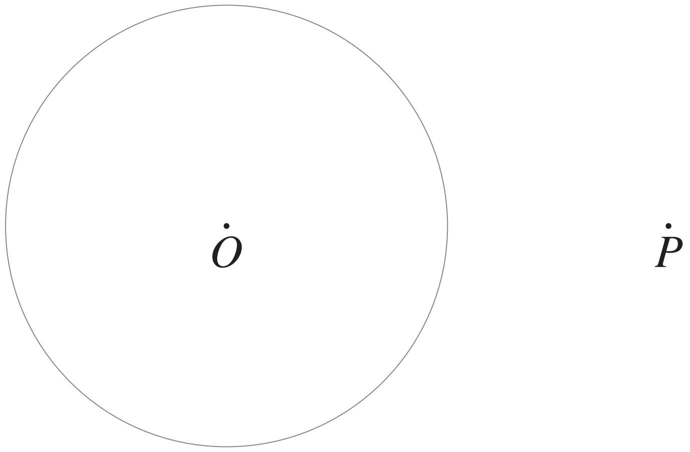
Barycentric combinations of circles

The theory of barycentric combinations of circles is a relatively recent one, emerging from the following problem:

28. Let Γ_1, Γ_2 and Γ_3 each pass through fixed points A and B. Let a line l pass through B and meet the circles again at P_1, P_2 and P_3. Prove that the ratio $P_1\,P_2 : P_2\,P_3$ is independent of l. **[Adapted from APMO 2012, Question 4]**

The original problem was solved in many unique ways by members of the British IMO squad, using techniques such as spiral similarity, vectors, coordinates, inversion, trigonometry and similar triangles.

Let Γ be a circle with centre O and radius r. We uniformly distribute a unit mass around the circumference of Γ. By applying the Huygens-Steiner theorem, we can deduce that $\mathbb{M}(\Gamma, P) = \mathbb{M}(\Gamma, O) + OP^2 = r^2 + OP^2 = 2r^2 + \text{Power}(P, \Gamma)$.



Suppose we have n circles in the plane, $\{\Gamma_1, \ldots, \Gamma_n\}$, each considered to have a mass m_i such that $m_1 + m_2 + \ldots + m_n = 1$. We then define an 'average circle' F such that $\text{Power}(P, \text{F}) = \sum(m_i \text{Power}(P, \Gamma_i))$. This is possible by considering the equation for the power of a point in either Cartesian or areal coordinates.

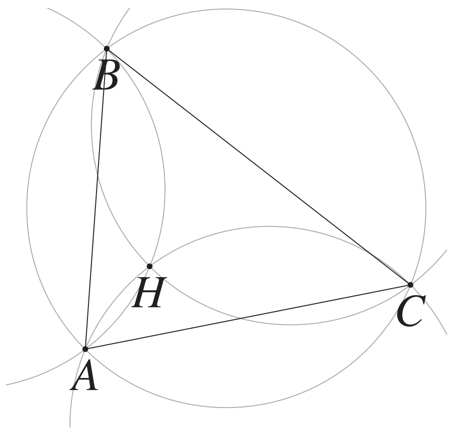
29. Prove that the centre of F is the weighted centroid $G = \sum(m_i O_i)$, where O_i is the centre of Γ_i. **[Barycentric combination of circles]**

It is now possible to determine the radius of F.

- $R^2 = -\text{Power}(G, \text{F}) = \sum m_i(r_i^2 - GO_i^2)$, where the circle Γ_i has centre O_i and radius r_i. **[Radius of barycentric circle]**

Certain barycentric combinations of circles are interesting.

30. Let ABC be a triangle with circumcircle Γ. Γ_A, Γ_B and Γ_C are the reflections of Γ in the sides BC, CA and AB, respectively. Show that the average of the three circles Γ_A, Γ_B and Γ_C is the Euler-Apollonius lollipop.



31. What is the average of the four circles Γ, Γ_A, Γ_B, Γ_C?

If we have four points A, B, C, D which do not form a cyclic quadrilateral, then every circle in the plane can be expressed uniquely as a barycentric combination of the four circles BCD, CDA, DAB, ABC. In other words, the set of circles on the plane is isomorphic to a subset of projective three-space $\mathbb{P}^3$. This idea of giving things other than points coordinates is not a new one; Plücker created a geometry based on the four-dimensional space of lines in $\mathbb{R}^3$.

Solutions

1. By using $V = \frac{1}{3} A\,h$, where A is the base area and h is the perpendicular height (i.e. distance from the origin to the reference plane), we have $[P\,B\,C] = \frac{3}{h}[R\,P\,B\,C] = \frac{1}{2\,h} \det\begin{pmatrix} x & y & z \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \frac{x}{2\,h}$. For the point A, $x = 1$ so this area is equal to $\frac{1}{2\,h}$. Hence, $[P\,B\,C] = x\,[A\,B\,C]$.

2. If P lies on the line $B\,C$, the area x is obviously zero (as $X\,B\,C$ is a straight line). By symmetry, we obtain the equations of the other two lines.

3. For the centroid, we have $[G\,B\,C] = [G\,C\,A] = [G\,A\,B]$. When normalised so these areas sum to unity, they must each equal $\frac{1}{3}$.

4. This is the same argument as in the first question, but with the tetrahedra $S\,P\,Q\,R$ and $A\,B\,C\,R$.

5. We can scale the coordinates of P such that $x_1 = 1$. As the line passes through $\left(1, \frac{y_1}{x_1}, \frac{z_1}{x_1}\right)$ and $(1, 0, 0)$, it must obviously be the equation $y_1\,z = z_1\,y$. Hence, the intersection point has unnormalised coordinates $(0,\ y_1,\ z_1)$.

6. Assume P exists and $P = (x_1,\ y_1,\ z_1)$. By the previous result, we have $\frac{\overrightarrow{B\,L}}{\overrightarrow{L\,C}} = \frac{z_1}{y_1}$. The cyclic product is $\frac{\overrightarrow{B\,L}}{\overrightarrow{L\,C}} \cdot \frac{\overrightarrow{C\,M}}{\overrightarrow{M\,A}} \cdot \frac{\overrightarrow{A\,N}}{\overrightarrow{N\,B}} = \frac{x_1\,y_1\,z_1}{x_1\,y_1\,z_1} = 1$. For the converse result, we know that there must exist precisely one point L on $B\,C$ such that $A\,L$, $B\,M$ and $C\,N$ are concurrent, and it must thus be the case where $\frac{\overrightarrow{B\,L}}{\overrightarrow{L\,C}} \cdot \frac{\overrightarrow{C\,M}}{\overrightarrow{M\,A}} \cdot \frac{\overrightarrow{A\,N}}{\overrightarrow{N\,B}} = 1$.

7. This process effectively ‘flips’ each fraction in Ceva’s theorem, so the product remains equal to unity and thus the cevians are concurrent.

8. By the sine rule, we have $\frac{\sin \angle L\,A\,B}{\sin B} = \frac{\overrightarrow{B\,L}}{L\,A}$ and $\frac{\sin \angle C\,A\,L}{\sin C} = \frac{\overrightarrow{L\,C}}{L\,A}$. Dividing one by the other results in $\frac{\sin \angle L\,A\,B}{\sin \angle C\,A\,L} = \frac{\overrightarrow{B\,L}}{\overrightarrow{L\,C}}$. Substituting cyclic permutations of this into the Ceva equation yields $\frac{\sin \angle L\,A\,B}{\sin \angle C\,A\,L} \cdot \frac{\sin \angle M\,B\,C}{\sin \angle A\,B\,M} \cdot \frac{\sin \angle N\,C\,A}{\sin \angle B\,C\,N} = 1$.

9. This process effectively ‘flips’ each fraction in trigonometric Ceva’s theorem, so the product remains equal to unity and thus the Cevians are concurrent.

10. The incentre and excentres must be their own isogonal conjugates, thus have unnormalised areal coordinates $(\pm a,\ \pm b,\ \pm c)$. The incentre is the only one with a symmetrical expression, $(a,\ b,\ c)$.

11. Using the identity $\tan A \tan B \tan C = \tan A + \tan B + \tan C$, we can divide each term in the expression $(\tan A,\ \tan B,\ \tan C)$ by $\tan A \tan B \tan C$ to obtain $(\cot B \cot C,\ \cot C \cot A,\ \cot A \cot B)$.

12. $B\,O\,A$ is isosceles, so $\angle\,O\,A\,B = \frac{\pi}{2} - C$. As $A\,H$ is perpendicular to $B\,C$, we have $\angle\,C\,A\,O = \frac{\pi}{2} - C$. By symmetry, we are done.

13. In Cartesian coordinates, let mass m_1 be placed at $A_1 = (x_1,\ y_1)$, *et cetera*, and let $\sum m_i = 1$. Let $P = (u,\ v)$ be the barycentre, and $Q = (x,\ y)$ be a variable point. By Pythagoras’ theorem, we have

$A_1 Q^2 = (x - x_1)^2 + (y - y_1)^2 = x^2 + y^2 - 2 x_1 x - 2 y_1 y + c_1$, where c_1 is a constant term that doesn't really matter. Repeat for all points in this manner, and calculate the weighted sum. The weighted mean square distance is given by $\mathbb{M}(S, Q) = x^2 + y^2 - 2 u x - 2 v y + c$, where c is another unimportant constant. But this is just $(x - u)^2 + (y - v)^2 + k$, for some constant k, or $P Q^2 + k$. Substituting $P = Q$ gives $k = \mathbb{M}(S, P)$. *(Normally, one would not use Cartesian coordinates to solve a problem. However, in RMM 2011, I was under the influence of alcohol, so actually successfully performed this derivation.)*

14. Invoking the Huygens-Steiner theorem once again, we obtain $O P^2 = x R^2 + y R^2 + z R^2 - \left(x A P^2 + y B P^2 + z C P^2\right) = R^2 - \left(x A P^2 + y B P^2 + z C P^2\right)$.

15. G has normalised areals $\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$, giving us $O G^2 = R^2 - \frac{1}{3}\left(A G^2 + B G^2 + C G^2\right)$. If D is the midpoint of $B C$, we obtain $A D^2 = \frac{1}{2} b^2 + \frac{1}{2} c^2 - \frac{1}{4} a^2$ from Stewart's theorem. Multiplying by $\frac{4}{9}$ results in $A G^2 = \frac{2}{9} b^2 + \frac{2}{9} c^2 - \frac{1}{9} a^2$, hence the cyclic sum $A G^2 + B G^2 + C G^2 = \frac{1}{3}\left(a^2 + b^2 + c^2\right)$. Substituting this into the expression for $O G^2$ yields the desired formula.

16. I has normalised areals $\left(\frac{a}{a+b+c}, \frac{b}{a+b+c}, \frac{c}{a+b+c}\right)$, giving us $O I^2 = R^2 - \frac{1}{a+b+c}\left(a A I^2 + b B I^2 + c C I^2\right)$. Applying Pythagoras' theorem yields $A I^2 = r^2 + (s - a)^2 = r^2 + s^2 - 2 a s + a^2$. Hence, $a A I^2 + b B I^2 + c C I^2 = (a + b + c) r^2 + (a + b + c) s^2 - 2\left(a^2 + b^2 + c^2\right) s + \left(a^3 + b^3 + c^3\right)$. We can then convert this into an expression in terms of R, r and s, namely $2 s r^2 + 2 s^3 + 4 s\left(r^2 + 4 R r - s^2\right) + 2 s\left(s^2 - 3 r^2 - 6 R r\right) = 4 R r s$. Hence, $O I^2 = R^2 - \frac{4 R r s}{2 s} = R^2 - 2 R r$.

17. The sum of squared distances from each point to the other points is constant. Hence, using the Huygens-Steiner theorem, all points must be concyclic. By considering the closest pairs of points, the points must be the vertices of a regular polygon or truncated regular polygon.

18. Apply an affine transformation to make D the orthocentre of $A B C$. Then, those nine points lie on a conic (the nine-point circle), and T is the barycentre of $A B C D$. Reversing the affine transformation results in a conic passing through those nine points; the centre of the conic is the barycentre of $A B C D$. However, a parabola has no centre, so the conic cannot possibly be a parabola.

19. D has normalised areals $\left(0, \frac{n}{a}, \frac{m}{a}\right)$. Using the Huygens-Steiner theorem, we have $A D^2 = \frac{n}{a} A B^2 + \frac{m}{a} A C^2 - \frac{n}{a} D B^2 - \frac{m}{a} D C^2$, or $d^2 = \frac{c^2 n}{a} + \frac{b^2 m}{a} - \frac{m^2 n}{a} - \frac{n^2 m}{a}$. Multiplying through by a gives us the theorem, $m a n + d a d = b m b + c n c$.

20. By considering the circumcentre and three vertices, $\det\begin{pmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & R^2 & R^2 & R^2 \\ 1 & R^2 & 0 & a^2 & b^2 \\ 1 & R^2 & a^2 & 0 & c^2 \\ 1 & R^2 & b^2 & c^2 & 0 \end{pmatrix} = 0$. We now subtract the last row from the second, third and fourth rows to obtain $\det\begin{pmatrix} 0 & 1 & 1 & 1 & 1 \\ 0 & -R^2 & R^2 - b^2 & R^2 - c^2 & R^2 \\ 0 & 0 & -b^2 & a^2 - c^2 & b^2 \\ 0 & 0 & a^2 - b^2 & -c^2 & c^2 \\ 1 & R^2 & b^2 & c^2 & 0 \end{pmatrix} = 0$. We can now use the recursive formula to reduce this to the 4×4 determinant $\det\begin{pmatrix} 1 & 1 & 1 & 1 \\ -R^2 & R^2 - b^2 & R^2 - c^2 & R^2 \\ 0 & -b^2 & a^2 - c^2 & b^2 \\ 0 & a^2 - b^2 & -c^2 & c^2 \end{pmatrix} = 0$.
Subtract the fourth column from the other three columns, giving the equivalent equation

$\det\begin{pmatrix} 0 & 0 & 0 & 1 \\ -2R^2 & -b^2 & -c^2 & R^2 \\ -b^2 & -2b^2 & a^2-b^2-c^2 & b^2 \\ -c^2 & a^2-b^2-c^2 & -2c^2 & c^2 \end{pmatrix} = 0$. We can then apply the recursive formula to give the 3×3 determinant $\det\begin{pmatrix} 2R^2 & b^2 & c^2 \\ b^2 & 2b^2 & b^2+c^2-a^2 \\ c^2 & b^2+c^2-a^2 & 2c^2 \end{pmatrix} = 0$. It is now convenient to use the Rule of Sarrus to evaluate this directly, resulting in the equation $8R^2b^2c^2 + 2b^2c^2(b^2+c^2-a^2) = 2R^2(b^2+c^2-a^2)^2 + 2b^2c^4 + 2c^2b^4$. Dividing throughout by two and rearranging gives $R^2\left((2bc)^2 - (b^2+c^2-a^2)^2\right) = a^2b^2c^2$. Applying the difference of two squares to the left hand side yields $R^2(a^2-(b-c)^2)\left((b+c)^2-a^2\right) = a^2b^2c^2$. Another couple of applications enables further factorisation to $R^2(a+b+c)(a+b-c)(a-b+c)(-a+b+c) = a^2b^2c^2$.

21. Represent A, B and C with complex numbers l, m and n, respectively, where $l\,l^* + m\,m^* + n\,n^* = R^2$. Then we have $p = x_1\,l + y_1\,m + z_1\,n$ and $q = x_2\,l + y_2\,m + z_2\,n$. Subtracting them results in $p - q = u\,l + v\,m + w\,n$. Multiplying by its complex conjugate gives the squared modulus $(p-q)(p^*-q^*) = (u^2+v^2+w^2)R^2 + u\,v\,l\,m^* + u\,v\,m\,l^* + v\,w\,m\,n^* + v\,w\,n\,m^* + w\,u\,n\,l^* + w\,u\,l\,n^*$. As $u+v+w=0$, $u^2+v^2+w^2 = -(2uv+2vw+2wu)$. Applying this substitution gives $(p-q)(p^*-q^*) = \sum_{\text{cyc}} \left(u\,v(l\,m^* + m\,l^* - 2R^2)\right) = -\sum_{\text{cyc}} (u\,v(l-m)(l^*-m^*))$. The final expression is equal to $-u\,v\,c^2 - v\,w\,a^2 - w\,u\,b^2$, as required.

22. Using Goucher's theorem, we have $\det\begin{pmatrix} x(x+y+z) & 1 & 0 & 0 \\ y(x+y+z) & 0 & 1 & 0 \\ z(x+y+z) & 0 & 0 & 1 \\ a^2yz + b^2zx + c^2xy & 0 & 0 & 0 \end{pmatrix} = \frac{[ABC]}{[ABC]}\,\text{Power}(P, ABC)$. This neatly multiplies out to give $-a^2yz - b^2zx - c^2xy = \text{Power}(P, ABC)$.

23. $OI^2 - R^2$ is minus the power of I with respect to the circumcircle of ABC, so is equal to $\frac{a^2yz+b^2zx+c^2xy}{(x+y+z)^2} = \frac{a^2bc+b^2ca+c^2ab}{(a+b+c)^2} = \frac{abc}{a+b+c}$. We can express this in terms of R, r and s, obtaining $\frac{4Rrs}{2s} = 2Rr$.

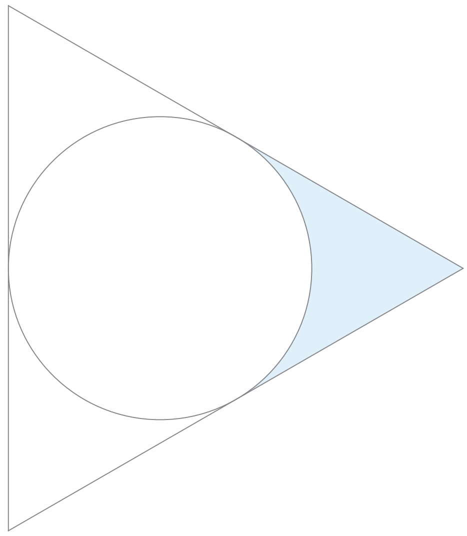
24. $\frac{a^2yz+b^2zx+c^2xy}{(x+y+z)^2} = \frac{a^2bc-b^2ca-c^2ab}{(-a+b+c)^2} = -\frac{abc}{(-a+b+c)} = -\frac{2Rrs}{s-a} = -2Rr_A$.

25. Using Euler's formula on each tritangential circle, this is equal to $4R^2 + 2R(r_A + r_B + r_C - r)$. It was proved in an earlier exercise that $r_A + r_B + r_C - r = 4R$.

26. The orthocentre has unnormalised areal coordinates $(\tan A, \tan B, \tan C)$. Hence, minus the power of H with respect to the circumcircle is $\frac{a^2yz+b^2zx+c^2xy}{(x+y+z)^2} = \frac{a^2\tan B\tan C + b^2\tan C\tan A + c^2\tan A\tan B}{\tan^2 A\tan^2 B\tan^2 C} = \frac{a^2\cot A + b^2\cot B + c^2\cot C}{\tan A\tan B\tan C}$. Remembering that $\cot A = \frac{b^2+c^2-a^2}{4[ABC]}$, the numerator is equal to $\frac{2a^2b^2+2b^2c^2+2c^2a^2-a^4-b^4-c^4}{4[ABC]} = 4[ABC]$. (The last step is from squaring Heron's formula.) This results in $R^2 - OH^2 = 4[ABC]\cot A\cot B\cot C$. For the second part of the problem, we use $[ABC] = \frac{abc}{4R} = 2R^2\sin A\sin B\sin C$. Substituting this into the previous formula gives us $8R^2\cos A\cos B\cos C$, as required.

27. As the question is a homogeneous function in areas, we can apply an affine transformation and consider the case of the equilateral triangle. In areal coordinates, the inequality becomes $\sqrt{c} \geq \sqrt{a} + \sqrt{b}$. The set of points for which $\left(\sqrt{a}+\sqrt{b}+\sqrt{c}\right)\left(-\sqrt{a}+\sqrt{b}+\sqrt{c}\right)\left(\sqrt{a}-\sqrt{b}+\sqrt{c}\right)\left(\sqrt{a}+\sqrt{b}-\sqrt{c}\right)$ is non-negative is the interior of the conic $a^2+b^2+c^2 \leq 2ab+2bc+2ca$. As the conic passes through $\left(\frac{1}{2}, \frac{1}{2}, 0\right)$,

$\left(\frac{1}{6}, \frac{1}{6}, \frac{2}{3}\right)$ and cyclic permutations thereof, it must be the incircle. Using the formula for the area of a circle, we have that the probability that it lands within the conic is $\frac{\pi\sqrt{3}}{9}$. Hence, the probability that it lands outside the conic must be $\frac{9-\pi\sqrt{3}}{9}$. However, this could occur if any of $\sqrt{c} \geq \sqrt{a} + \sqrt{b}$, $\sqrt{a} \geq \sqrt{b} + \sqrt{c}$ or $\sqrt{b} \geq \sqrt{c} + \sqrt{a}$ are true. By symmetry, we actually want one-third of this, namely $\frac{9-\pi\sqrt{3}}{27}$.



28. Invert about B. The three circles are mapped to lines through A' (the inverse of A), and the line l remains invariant. The cross-ratio $(\infty', P_2'; P_1', P_3')$ is independent of l, as we can view A' as a projector. This is the same as the original cross-ratio $(\infty, P_2; P_1, P_3)$, so that must also be independent of l. As one of the points is infinity, the simple ratio $P_1 P_2 : P_2 P_3$ also remains constant.

29. By Huygens-Steiner, $\sum(m_i \mathbb{M}(\Gamma_i, P)) = \sum(m_i \mathbb{M}(\Gamma_i, G)) + P G^2$. Subtracting $2\sum\left(m_i r_i^2\right)$ from each side gives us $\sum(m_i \operatorname{Power}(P, \Gamma_i)) = \sum(m_i \operatorname{Power}(G, \Gamma_i)) + P G^2$, or $\operatorname{Power}(P, \mathrm{F}) = \operatorname{Power}(G, \mathrm{F}) + P G^2$, so G must be the centre of F.

30. The three circles all pass through the orthocentre H, so H must lie on F. $O_A O_B O_C$ is a dilated copy of the medial triangle $L M N$, which has orthocentre O, therefore we can deduce that O is the orthocentre of $O_A O_B O_C$. As H is the circumcentre of $O_A O_B O_C$, the centroid of $O_A O_B O_C$ (and thus centre of F) must be the point Q on the Euler line of $A B C$ halfway between G and H. As F also passes through H, it must necessarily be the Euler-Apollonius lollipop.

31. This is the weighted barycentre of the Euler-Apollonius lollipop and the circumcircle, where the former has mass $\frac{3}{4}$ and the latter has mass $\frac{1}{4}$. Hence, F must have centre T, *i.e.* the centre of the nine-point circle. Let X be the radius of the barycentric circle F. Using the radius formula, we have $X^2 = \frac{3}{4}\left(G Q^2 - T Q^2\right) + \frac{1}{4}\left(R^2 - O T^2\right)$. Let $p = T Q$. By the basic ratios along the Euler line, this is $\frac{3}{4}\left(4 p^2 - p^2\right) + \frac{1}{4}\left(R^2 - 9 p^2\right) = \frac{1}{4} R^2$. Hence, $X = \frac{1}{2} R$ and thus F is the nine-point circle.

The Riemann sphere

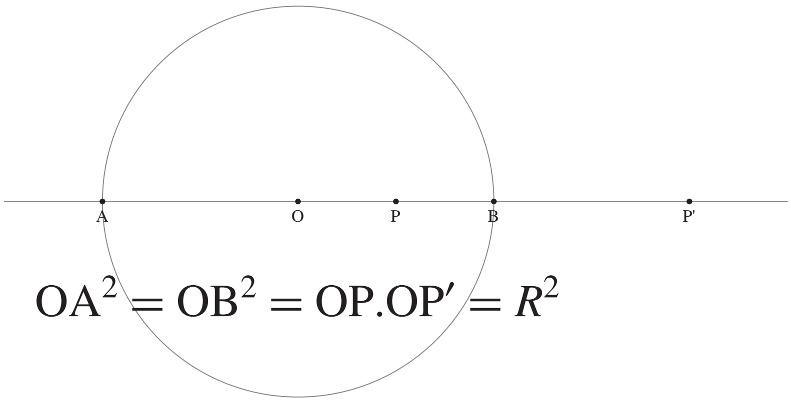
We can augment the complex plane by adding an extra point, ∞, to the plane. This produces the *complex projective line*, $\mathbb{CP}^1$. It can be considered to be a projective space represented by a pair of complex coordinates, (x, y), where scalar (complex) multiples are considered equivalent. Hence, for all points except ∞, we can normalise the coordinates as $\left(\frac{x}{y}, 1\right)$. We identify these points with points on the ordinary complex plane by letting $z = \frac{x}{y}$. If $y = 0$, this results in the point at infinity, ∞. We consider a line to be a *generalised circle* passing through ∞.

Two-dimensional inversion

Inversion is essentially a reflection in a circle. Points outside the circle are interchanged with points inside the circle in an involution of the plane.

- For a circle Γ with centre O, we define the *inverse point* of P to be the point P' on the line OP (at the same side as P from O) such that $OP \cdot OP' = R^2$, where R is the radius of Γ. **[Definition of inversion]**

If O is the origin, the inverse point can be found by the simple transformation $z \to \frac{R^2}{z^*}$.

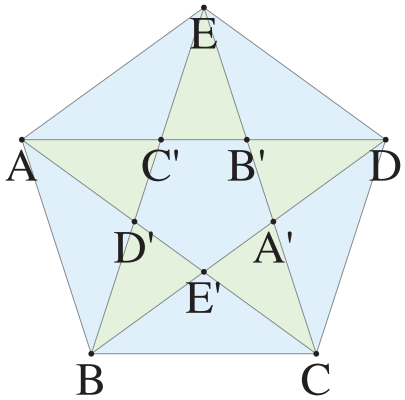


1. Show that O and ∞ are inverse points with respect to any circle centred on O.

2. Demonstrate that $(A, B; P, P')$ is a harmonic range.

3. If P and Q invert to P' and Q', show that P, P', Q and Q' are concyclic.

4. Show that $P'Q' = \frac{PQ \cdot R^2}{PO \cdot QO}$. **[Inversion distance formula]**



5. Let $ABCDE$ be a regular pentagon with side length 1. Let the diagonals BD and CE intersect at A', and define B', C', D' and E' similarly. Show that:

- $BE = \phi$;
- $BD' = \frac{1}{\phi}$;
- $D'C' = \frac{1}{\phi^2}$;
- $BC' = 1$;

where ϕ is the positive root of the equation $x^2 - x - 1 = 0$. ($\phi = \frac{1+\sqrt{5}}{2} \simeq 1.618034$.)

Möbius transformations

We define a *Möbius transformation* to be a transformation of the form $z \to \frac{az+b}{cz+d}$, where $a, b, c, d \in \mathbb{C}$. It is also necessary to include the condition that $ad - bc \neq 0$, to remove degenerate singular non-invertible cases. We can assume without loss of generality, therefore, that $ab - cd = 1$.

6. Show that the composition of any two Möbius transformations is another Möbius transformation.

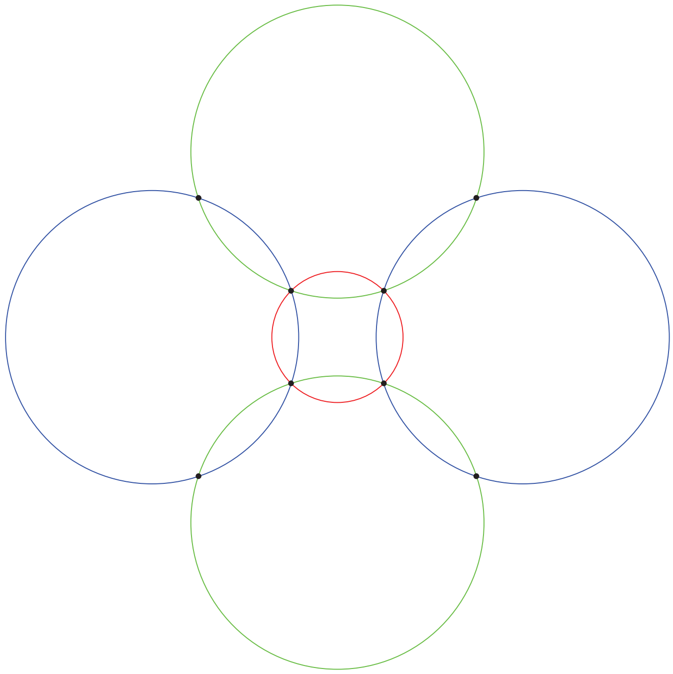
7. Prove that the composition of any inversion followed by any reflection is a Möbius transformation.

8. Show that, for four points $w, x, y, z \in \mathbb{C}$, the value of $\frac{(w-x)(y-z)}{(x-y)(z-w)}$ remains invariant when a Möbius transformation is applied.

Indeed, this follows naturally from the fact that Möbius transformations are projective transformations of the complex projective line. Hence, it is possible to find a unique Möbius transformation mapping any three points to any other three points.

9. Demonstrate that generalised circles remain as generalised circles under any Möbius transformation, and thus under inversion.

Like all non-trivial rational functions of z, Möbius transformations are *conformal maps*, which means angles between curves are (in general) preserved. Inversion reverses the direction of directed angles, but preserves the magnitude. This property can be derived from the fact that generalised circles remain as generalised circles.



10. Suppose $P_1 P_2 P_3 P_4$ is a cyclic quadrilateral. The circle Γ_n passes through P_n and P_{n+1}, with subscripts considered modulo 4. Circles Γ_n and Γ_{n+1} intersect again at Q_{n+1}. Prove that $Q_1 Q_2 Q_3 Q_4$ are either concyclic or collinear. **[Miquel's theorem]**

11. Let ABC be a triangle with a right-angle at C. Let CN be an altitude. A circle Γ is tangent to line segments BN, CN and the circumcircle of ABC. If D is where Γ touches BN, prove that CD bisects angle $\angle BCN$. **[NST2 2011, Question 3]**

Ivan's 25 circles

Consider a cyclic quadrilateral $ABCD$ with circumcentre O. Let AB intersect CD at P. Similarly, we define $Q = BC \bigcap DA$ and $R = AC \bigcap BD$.

12. Prove that P is the pole of QR, and hence that O is the orthocentre of PQR. **[Brocard's theorem]**

Denote the feet of the altitudes from P, Q and R with X, Y and Z, respectively. Then, we obtain six cyclic quadrilaterals from the fact that O is the orthocentre of PQR.

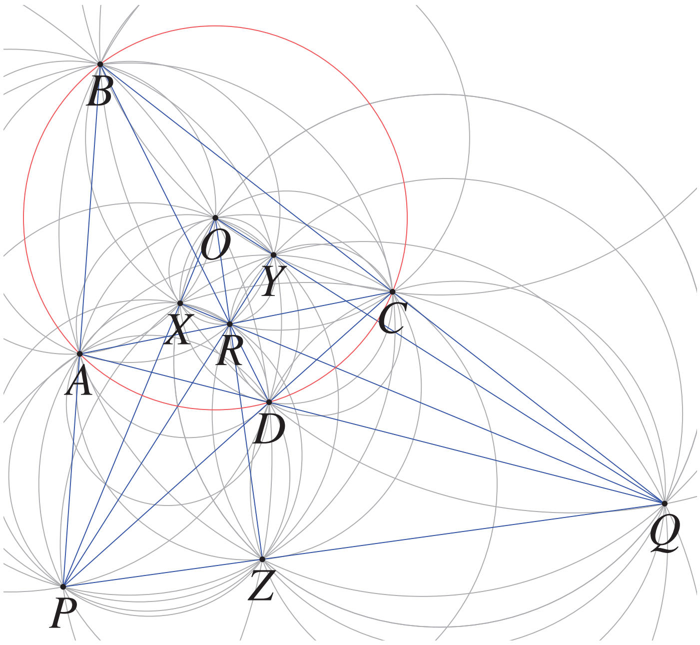
13. Prove that inversion about the circumcircle of $ABCD$ interchanges O with ∞, P with X, Q with Y and R with Z.

14. Hence prove that $OABX$ are concyclic.

By symmetry, this gives us six cyclic quadrilaterals, increasing the total to twelve (excluding $ABCD$). There are still another twelve circles on the diagram to be found.

15. Show also that $BCXR$ are concyclic.

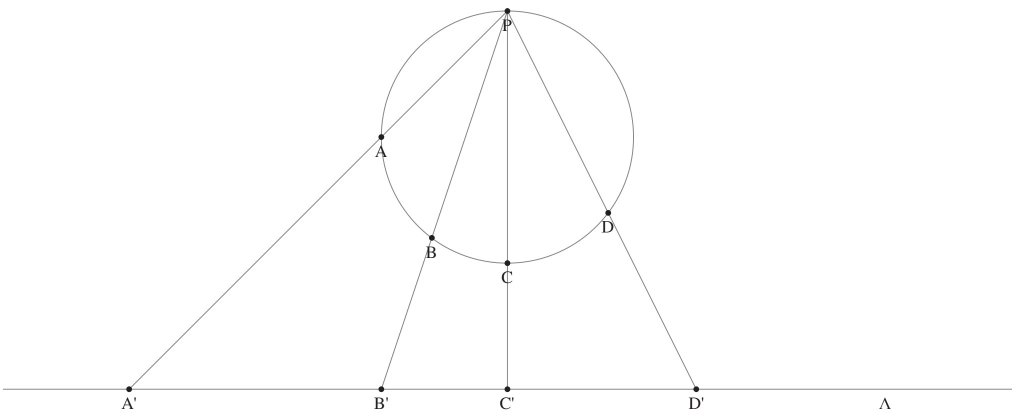
By symmetry, this increases the total of cyclic quadrilaterals to 24 (excluding $ABCD$). These circles were discovered by Ivan Guo.



As well as inverting about O, one can also invert about P, Q or R to permute the vertices.

Harmonic quadrilaterals

16. Let $ABCD$ be a cyclic quadrilateral. Choose a point P on the circumcircle of $ABCD$ and a line Λ outside the circle. The line PA meets Λ at A'; points B', C' and D' are defined similarly. Show that the cross-ratio $(A', C'; B', D')$ does not depend on the locations of P and Λ.



If $(A', C'; B', D') = -1$, then the cyclic quadrilateral $ABCD$ is known as a *harmonic quadrilateral*. Harmonic quadrilaterals have many nice properties:

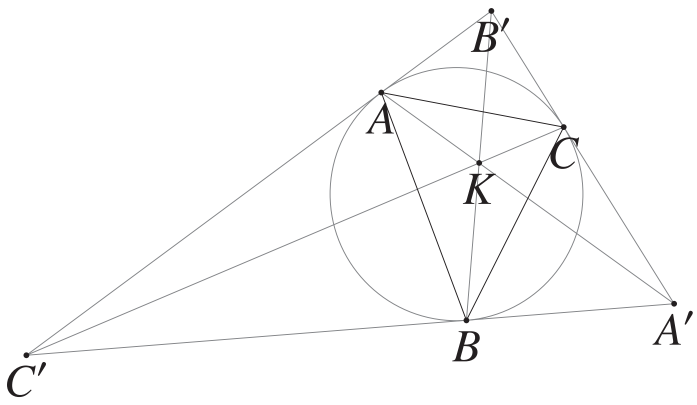
17. Let $ABCD$ be a convex cyclic quadrilateral. Let the tangents to the circumcircle at A and C meet at E; let the tangents to the circumcircle at B and D meet at F. Let the diagonals AC and BD intersect at P. Then show that the following properties are equivalent:

- $ABCD$ is a harmonic quadrilateral;
- $AB \cdot CD = BC \cdot DA = \frac{1}{2} AC \cdot BD$;
- E, B and D are collinear;
- B, D and K are collinear, where K is the symmedian point of ABC;
- $(P, E; B, D) = -1$.

18. If a quadrilateral $ABCD$ is represented by complex numbers a, b, c and d in the Argand plane, show that it is harmonic if and only if $(a-b)(c-d)+(b-c)(d-a)=0$.

19. Deduce that harmonic quadrilaterals/ranges remain harmonic after inversion.

The collinearity of E, B, K and D gives us an elegant construction of the symmedian point: let the tangents to the circumcircle at B and C intersect at A', and define B' and C' similarly. The symmedian point is then the intersection of AA', BB' and CC'.



Generalised spheres

We started by defining Möbius transformations and inversion in the environment of complex numbers. However, complex numbers are restricted to two (real) dimensions, so cannot be used to generalise the ideas to n-dimensional space. Instead, we will need to consider this from a more Euclidean perspective.

When discussing objects in n-dimensional space, we use the following conventions:

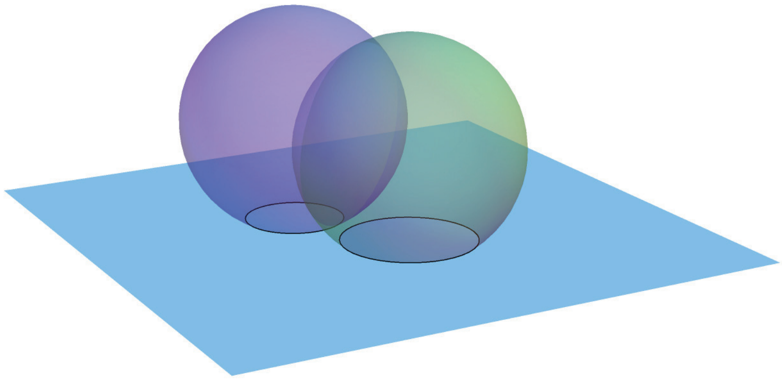
- A *n-ball* comprises all points in $\mathbb{R}^n$ within a distance of R from a point O.
- The surface of a n-ball is a $(n-1)$-*sphere*.
- The set of points in $\mathbb{R}^n$ obeying a single linear equation is a $(n-1)$-*plane*.
- A *generalised* $(n-1)$-*sphere* can be a $(n-1)$-sphere or a $(n-1)$-plane.

20. Show that the intersection of two generalised $(n-1)$-spheres in $\mathbb{R}^n$ is either empty, a single point, or a generalised $(n-2)$-sphere lying in a $(n-1)$-plane of $\mathbb{R}^n$. **[Intersection of generalised spheres]**

21. Let Γ be a $(n-1)$-sphere and P a point in $\mathbb{R}^n$. Lines l_1 and l_2 pass through P. l_1 intersects Γ at A and B; l_2 intersects Γ at C and D. Prove that $PA \cdot PB = PC \cdot PD$. **[*n*-dimensional intersecting chords theorem]**

22. If a point $P \in \mathbb{R}^n$ has equal power with respect to non-concentric $(n-1)$-spheres Γ_1 and Γ_2, then it lies on their *radical axis*. Show that the radical axis of two identical $(n-1)$-spheres is a $(n-1)$-plane.

This forms the basis of an ingenious proof by Géza Kós that the radical axis of any two (non-concentric) circles on the plane Λ is a line. Firstly, erect two spheres of equal radius, Γ_1 and Γ_2, on the circles.



The radical axis of the two circles is the intersection of the radical plane of Γ_1 and Γ_2 with the plane Λ. This construction clearly generalises to two non-concentric $(n-1)$-spheres in $\mathbb{R}^n$, by embedding the situation into $\mathbb{R}^{n+1}$ with equiradial n-spheres.

n-dimensional inversion

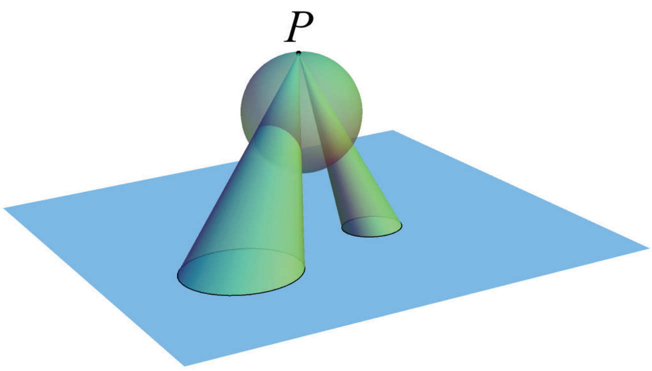
We are now in a suitable position to define n-dimensional inversion and investigate its properties.

- Let Γ be a $(n-1)$-sphere in $\mathbb{R}^n$ with centre O and radius R. For any point $P \in \mathbb{R}^n$, the *inverse point* P' is defined to lie on the line OP at the same side of O such that $OP \cdot OP' = R^2$. **[Definition of *n*-dimensional inversion]**

23. Prove that generalised $(n-1)$-spheres map to generalised $(n-1)$-spheres under inversion.

24. Draw a generalised $(n-2)$-sphere on the surface of some generalised $(n-1)$-sphere in $\mathbb{R}^n$. Prove that, after inversion, this will remain a generalised $(n-2)$-sphere. **[Backward compatibility of inversion]**

The oblique cones in the diagram below intersect both the plane and the sphere in circular cross-sections, as the sphere and plane are inverses with respect to P. This idea of projecting the sphere onto a plane from a point on the sphere is known as *stereographic projection*.



The diagram above enables us to easily define Möbius transformations. Let P be an arbitrary point outside the plane Λ. An arbitrary Möbius transformation of the plane Λ is the composition of:

- A two-dimensional translation and/or homothety of the plane Λ;
- An inversion about the unit sphere centred on P, transforming Λ into a sphere Ω (the Riemann sphere);
- A rotation of the sphere Ω about its centre (P does not move with the sphere);

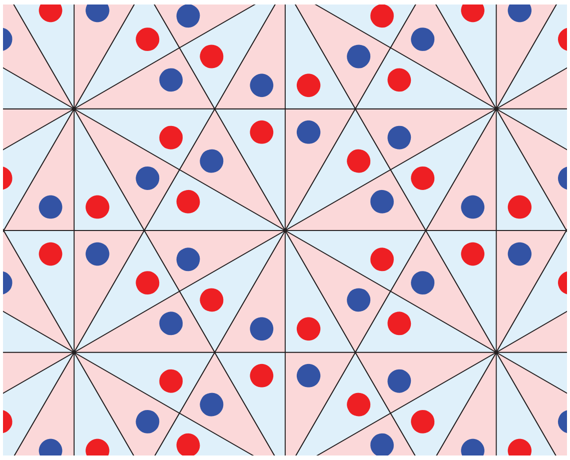
- An inversion about the unit sphere centred on P, transforming Ω back into Λ.

There are six degrees of freedom in this transformation, so they must represent all Möbius transformations, and nothing else. This definition of a Möbius transformation clearly generalises to $\mathbb{R}^n$ (with $\frac{1}{2}(n+1)(n+2)$ degrees of freedom), whereas the complex number definition does not.

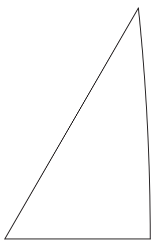
25. Suppose two smooth curves are drawn in $\mathbb{R}^n$, which intersect at a point. Prove that the angle of intersection is preserved (or, more accurately, reversed) after inversion. **[Anti-conformal map]**

Kaleidoscopes

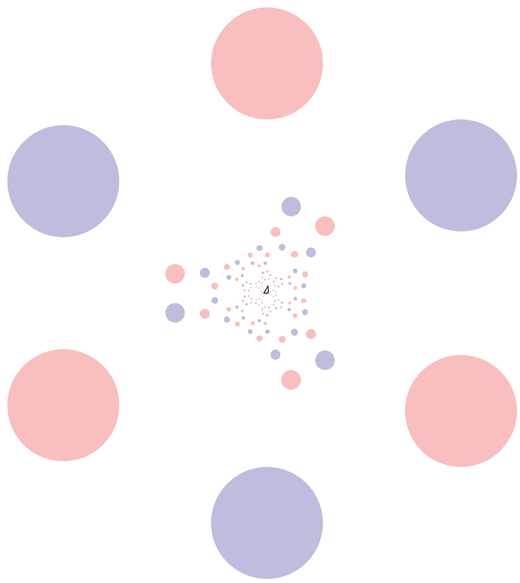
You may have encountered kaleidoscopes, where mirrors are used to form lots of repeated copies of a 'fundamental region'. For example, we can generate patterns with the same symmetries as the hexagonal tiling by using a triangle of mirrors with internal angles of $\frac{1}{2}\pi$, $\frac{1}{3}\pi$ and $\frac{1}{6}\pi$. Mathematicians regard this as a group of symmetries, *generated* by the three reflections. If we call the reflections α, β, γ, then we have the relations $\alpha^2 = \beta^2 = \gamma^2 = I$, where I is the identity element; a reflection is its own inverse. Our three rotations, $\alpha\beta$, $\beta\gamma$ and $\gamma\alpha$, together generate half of the symmetry group, namely the group of direct congruences. We also have $(\alpha\beta)^2 = (\beta\gamma)^3 = (\gamma\alpha)^6 = I$, as applying a rotation of 2π is equivalent to the identity.



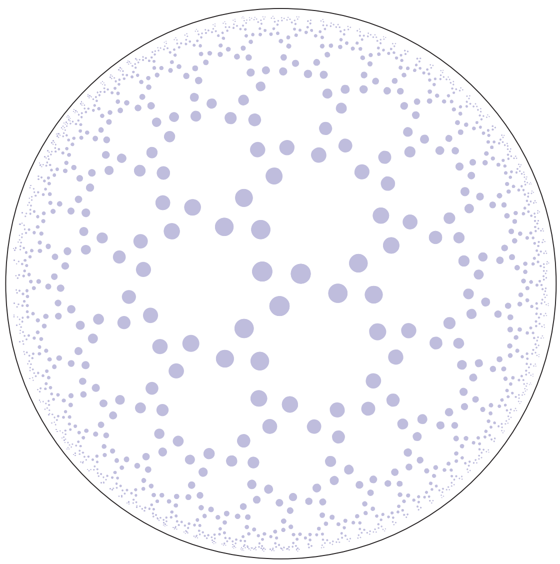
In Euclidean geometry, the interior angles of a triangle necessarily sum to π. However, if we allow circular arcs instead of straight lines, this condition can be relaxed. For example, we can have a curvilinear triangle with interior angles of $\frac{1}{2}\pi$, $\frac{1}{3}\pi$, $\frac{1}{5}\pi$:



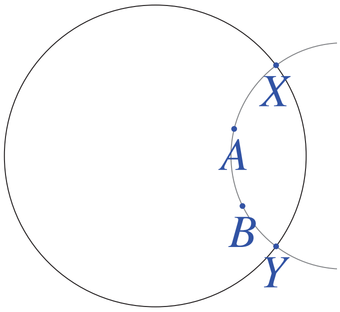
Reflection in the circular arc is simply inversion, and reflections in the straight lines are ordinary reflections. In this manner, we generate a spherical reflection group, namely the symmetry group of a dodecahedron. The compositions $\alpha\beta$, $\beta\gamma$ and $\gamma\alpha$ are no longer necessarily rotations, but are instead Möbius transformations. For this group, $(\alpha\beta)^2 = (\beta\gamma)^3 = (\gamma\alpha)^5$.



The above picture of icosahedral symmetry may look distorted, not least because we have flattened the Riemann sphere into a plane by stereographic projection. Finally, we can produce hyperbolic tilings by ensuring the interior angles of the fundamental triangle sum to less than π (by using concave arcs); this visualisation of the hyperbolic plane is known as the *Poincaré disc model*.



We can define 'distances' and 'angles' in hyperbolic geometry. As hyperbolic space must be invariant under any Möbius transformation mapping the unit circle to itself, distances and angles must also be preserved. We already know that complex (or cyclic) cross-ratio is invariant under Möbius transformations. The (directed) hyperbolic distance between two points, A and B, is given by the logarithm of the cyclic cross-ratio $(A, B; X, Y)$, where X and Y are the two intersections of the circle $A\,B\,B'\,A'$ (where $'$ indicates inversion in the unit circle) with the unit circle. Note that $A\,B\,B'\,A'$ is orthogonal to the unit circle.



■ The hyperbolic distance between two points is given by $AB = |\ln(A, B; X, Y)|$. **[Definition of hyperbolic distance]**

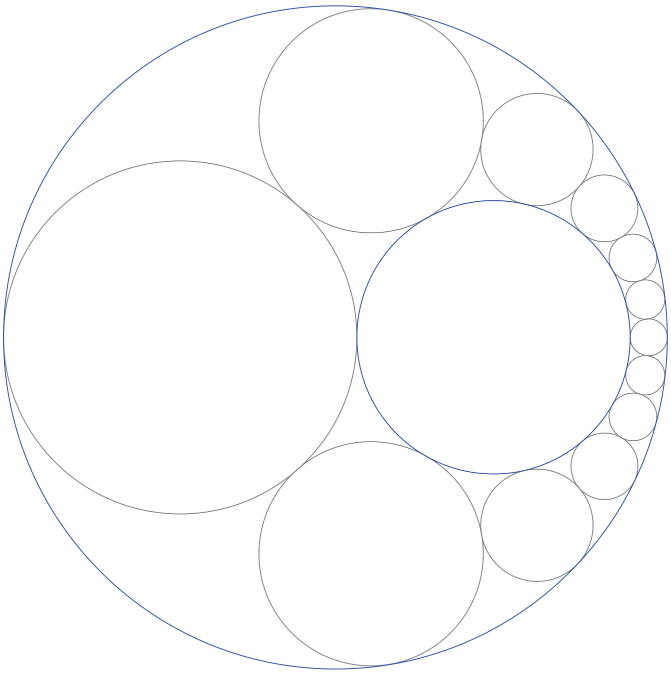
A hyperbolic line, *i.e.* the shortest path between two points A and B, is the arc AB of the circle $ABB'A'$. Hence, the triangle inequality applies: for three points A, B, C, $AB + BC \geq AC$, with equality if and only if they are collinear and in the correct order. As angles are preserved under Möbius transformations, the angles in the Poincaré disc model are the same as those in the hyperbolic plane.

■ The hyperbolic angle between two hyperbolic lines is identical to the ordinary angle between the corresponding circular arcs on the Poincaré disc model. **[Definition of hyperbolic angle]**

With these principles, it is possible to explore the rich world of Bolyai-Lobachevskian geometry. Four of Euclid's postulates (basic assumptions from which all of geometry can be derived) hold in hyperbolic geometry, whereas the fifth postulate does not. The fifth postulate is equivalent to the interior angles of a triangle summing to π.

Steiner's porism and Soddy's hexlet

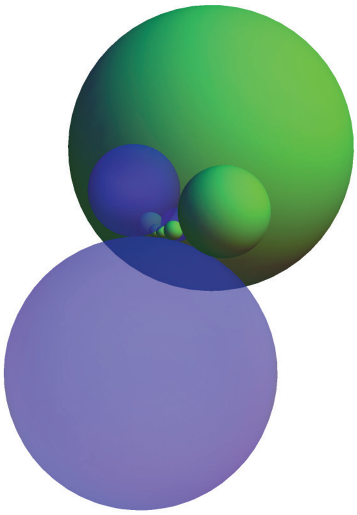
We have already encountered Poncelet's porism, which states that if there is one n-gon inscribed in one ellipse and circumscribed about another, then there are infinitely many. Another example of a porism (a term which people struggle to define) is *Steiner's porism*, again named after Jakob Steiner.



26. Let Γ_1 and Γ_2 be two circles, such that Γ_1 is contained within Γ_2. A set of $n \geq 3$ circles $\{C_1, C_2, \ldots, C_n\}$ is known as a *Steiner chain* of length n if each C_i is tangent externally to C_{i+1}, C_{i-1} (where subscripts are considered modulo n) and Γ_1, and is tangent internally to Γ_2. Show that if there exists one Steiner chain of length n for two given circles, then there exist infinitely many. **[Steiner's porism]**

There is an analogous three-dimensional porism, which is less general but more interesting, called *Soddy's hexlet*. This also appeared as a Japanese Sangaku problem.

27. Let Γ_1, Γ_2 and Γ_3 be three mutually tangent spheres. A set of $n \geq 3$ spheres $\{S_1, S_2, \ldots, S_n\}$ is known as a *Soddy chain* of length n if each S_i is tangent externally to S_{i+1}, S_{i-1} (where subscripts are considered modulo n), Γ_1, Γ_2 and Γ_3. Show that infinitely many Soddy chains exist of length 6, and no Soddy chains exist for $n \neq 6$. **[Soddy's hexlet]**



The six green spheres $\{S_1, \ldots, S_6\}$ are tangent to a quartic doughnut-shaped surface known as a *Dupin cyclide*, which is an inverted torus.

Solutions

1. Assume, without loss of generality, that O is the origin of the complex plane. Then, we have $0 \to \frac{1}{0^*} = \frac{1}{0} = \infty$.

2. We can compute the linear cross-ratio $\frac{\overrightarrow{AP}}{\overrightarrow{PB}} \cdot \frac{\overrightarrow{BP'}}{\overrightarrow{P'A}} = \frac{(R+p)\left(\frac{R^2}{p}-R\right)}{(R-p)\left(-\frac{R^2}{p}-R\right)} = \frac{\frac{R^3}{p}-Rp}{-\frac{R^3}{p}+Rp} = -1$.

3. This is a trivial corollary of the intersecting chords theorem.

4. $P'Q'O$ is similar to QPO, so we have $\frac{P'Q'}{P'O} = \frac{PQ}{QO}$. By using the identity $PO \cdot P'O = R^2$, we obtain the desired formula.

5. Let the length of the diagonals be denoted by x. By Ptolemy's theorem on $BCDE$, we have $1 + x = x^2$, so $x = \phi$. Note that D' and C are inverse points with respect to the unit circle centred on A. Hence, we obtain $BD' = \frac{BC}{AB \cdot AC} = \frac{1}{\phi}$; $BC' = \frac{BD}{AB \cdot AD} = \frac{\phi}{\phi} = 1$; $D'C' = \frac{CD}{AC \cdot AD} = \frac{1}{\phi^2}$.

6. Möbius transformations are projective transformations of the complex projective line, thus the composition is simply the product of their matrices.

7. Translations composed with rotations and dilations are Möbius transformations, as they have the form $z \to az + b$. Hence, we need only consider the case where the inversion is in the unit circle, and thus has the form $z \to \frac{1}{z^*}$. Composing this with the general form of an indirect similarity, $z \to a z^* + b$, results in the transformation $z \to \frac{a}{z} + b$, which is clearly a Möbius transformation.

8. Möbius transformations can be regarded as a projective transformations of the complex projective line, so must necessarily preserve complex cross-ratio.

9. The condition for four points to be concyclic or collinear is that the complex cross-ratio, $\frac{(a-b)(c-d)}{(b-c)(d-a)}$, is real. As it is preserved under Möbius transformations, so must the property that four points are concyclic or collinear. Obviously, a reflection also preserves this property, so inversions (compositions of reflections and Möbius transformations) must also do so.

10. Inverting about P_1 reduces the problem to the pivot theorem.

11. By Thales' theorem, AB is a diameter of the circumcircle, so is orthogonal to it. Invert about C. $A'B'$ and $N'\infty'$ are diameters of the same circle (the inverse of line AB), so $A'N'B'\infty'$ forms a rectangle. The inverse of Γ is tangent to the diagonals and circumcircle of the rectangle $A'N'B'\infty'$, so lies on one of the lines of symmetry of the rectangle. The point D' must thus be the midpoint of the minor arc $B'N'$, as D lies between B and N in the original diagram. As arcs $B'C'$ and $C'N'$ are congruent, they must subtend equal angles at ∞'. As angles are preserved under inversion, we are finished.

12. Let QR intersect AB and CD at E and F, respectively. From applying a projective transformation to convert $ABCD$ into a square, we can deduce that $(P, E; A, B)$ and $(P, F; C, D)$ are harmonic ranges. This is sufficient for QR to be the polar of P. Hence, OP is perpendicular to QR. By symmetry, O must be the orthocentre of PQR.

13. O is the centre of the circle of inversion, so is exchanged with the point at infinity. X is the intersection of OP with the polar of P, so they must be inverse points. By symmetry, we obtain the other two pairs of inverse points.

14. Inverting about the circle $ABCD$ maps O and X to ∞ and P, respectively. ABP is a straight line and thus passes through ∞, so the original four points were concyclic.

15. Note that lines BC, OY and XR concur at Q. Applying the converse of the radical axis theorem to circles $OYBC$ and $ROXY$, we obtain the concyclicity of $XRBC$.

16. It obviously does not depend on Λ, as we can view P as a projector, and moving Λ is simply applying one-dimensional projective transformations to the line $A'B'C'D'$. Hence, the cross-ratio is dependent only upon the angles APB, BPC and CPD. Due to basic circle theorems, these are invariant as P moves on the circumcircle of $ABCD$.

17. Inversion about P and considering similar triangles derives the first of these results. To show that the collinearity of E, B and D implies that the quadrilateral is harmonic, use the intersecting chords theorem to show that AED is similar to BEA, and that DEC is similar to CEB. We then have $DA/AB = DE/AE = DE/CE = CD/BC$, thus the products of opposite sides are equal. The converse follows trivially. Showing that P is in harmonic range with E, B and D stems from the fact that P lies on AC, which is the polar of E. To demonstrate that K lies on BD, it is sufficient to show that P is the foot of the symmedian from B to AC. This is equivalent to the statement that $AP/AB^2 = CP/CB^2$. To prove this, we can exploit similar triangles to show that $AP/AB = DP/CD = (PA \cdot PC/AB)/(BC \cdot DA/AB)$, whence it follows that $AP\big/\left(AB^2\right) = (PA \cdot PC)/(PB \cdot BC \cdot DA) = (PA \cdot PC)/(PB \cdot BA \cdot DC) = CP\big/\left(CB^2\right)$. This shows that P has the required areal coordinates to be the foot of the symmedian.

18. Consider when the complex cross-ratio $(a-b)(c-d)/(b-c)(d-a) = -1$. From here, we use polar coordinates, resulting in $\arg(a-b)/\arg(c-b) = \arg(a-d)/\arg(c-d)$ and $|a-b|\ |c-d| = |b-c|\ |d-a|$. These are equivalent to $\angle ABC = \angle ADC$ and $AB \cdot CD = BC \cdot DA$, respectively. Clearly, the first of these is the ‘angles in the same segment’ criterion for concyclicity, and the latter is condition for a cyclic quadrilateral to be harmonic.

19. A composition of an inversion and a reflection is a Möbius transformation, *i.e.* a map of the form $z \to (az+b)/(cz+d)$. This can be regarded as a projective transformation of the complex projective line, so must necessarily preserve complex cross-ratio.

20. For two $(n-1)$-planes, the intersection is either empty (if they are parallel) or the $(n-2)$-plane formed by solving their algebraic equations simultaneously. For two $(n-1)$-spheres, consider all 2-planes passing through both centres. If the circles are disjoint, so are the original spheres. If the circles are tangent, the original spheres share a single point. If the circles intersect in two points, then let the radical axis meet the line of centres at P. Clearly, all intersection points of the original two spheres must lie on the $(n-1)$-plane through P perpendicular to the line of centres, and must all be equidistant from P. For the case of one $(n-1)$-sphere and one $(n-1)$-plane, we can reflect the sphere in the plane and reduce it to the previous case.

21. Consider the plane containing l_1 and l_2, and apply the two-dimensional intersecting chords theorem to this configuration.

22. The power of a point is equal to $OP^2 - R^2$. When the spheres are of equal radii, the equation for the radical plane becomes $OP_1{}^2 = OP_2{}^2$, which is the locus of all points equidistant from the centres of the two spheres. This must be the plane of reflective symmetry of the configuration.

23. Draw the line l through O orthogonal to the generalised $(n-1)$-sphere, intersecting it at A and B (one of which may be at infinity if the generalised sphere is a plane). Let the inverse points be A' and B', which also lie on l. Every plane containing l must intersect the generalised $(n-1)$-sphere in a generalised circle of diameter AB; this inverts to a generalised circle of diameter $A'B'$ using ordinary two-dimensional inversion. The union of all such generalised circles is the generalised $(n-1)$-sphere of diameter $A'B'$.

24. Consider the generalised $(n-2)$-sphere to be the intersection of two $(n-1)$-spheres. The result follows from the fact that the intersection is either empty (impossible), a single point (impossible) or a generalised $(n-2)$-sphere.

25. We can reduce this to the 3-dimensional problem, as we need only consider the 3-plane containing the tangents to the two curves, l_1 and l_2, and the centre of inversion, O. Assume they intersect at P with an angle α. After inversion, the plane $l_1 l_2$ becomes a generalised 2-sphere, and generalised circles l_1 and l_2 intersect at P' with an angle β. If the generalised 2-sphere is a sphere, we invert about the point diametrically opposite to P', preserving P' whilst mapping l_1 and l_2 back to lines. Clearly, the lines now intersect at an angle $-\beta$. This orientation-preserving generalised-circle-preserving map from the Riemann sphere to itself must necessarily be a Möbius transformation, and thus preserve angles. Hence, $\alpha = -\beta$, and we are done.

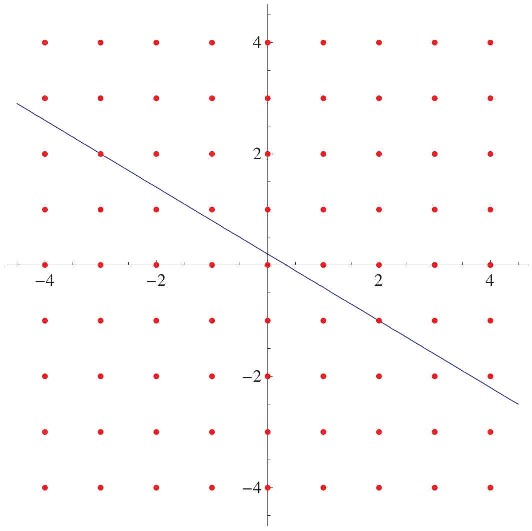
26. Let the line of centres l intersect Γ_1 at B and C, and Γ_2 at A and D, such that $ABCD$ are in that order along l. Note that the two circles intersect l orthogonally. Let $A'B'$ and $C'D'$ be two unit lengths on the same line l', and separate them such that $(A', D'; B', C') = (A, D; B, C)$. Apply a Möbius transformation to map A to A', B to B' and C to C'. As the cross-ratios are equal, D must necessarily map to D'. The two circles are now both orthogonal to l', so the centres both lie on this line. As $A'D'$ and $B'C'$ share a midpoint, the centres must coincide and the circles are concentric. We can now 'rotate' the Steiner chain within the annulus by an arbitrary angle, similar to a ball bearing. Applying the inverse Möbius transformation will restore Γ_1 and Γ_2.

27. Invert about the tangency point of Γ_1 and Γ_2, resulting in two parallel planes sandwiching the sphere Γ_3. All spheres tangent to Γ_1 and Γ_2 must have diameter equal to the separation of the planes. Hence, their centres must be coplanar with Γ_3 and have equal radius. Hence, we can consider this to be a two-dimensional problem of packing a closed loop of n circles of unit radius around a circle of unit radius. As the centres of Γ_3, S_i and S_{i+1} must form an equilateral triangle, this forces the angles to be $\frac{\pi}{3}$ and thus there must be a regular hexagon of six spheres. There are infinitely many orientations in which this can be done.

Diophantine equations

So far, we have considered solutions to equations over the real and complex numbers. This chapter instead focuses on solutions over the integers, natural and rational numbers. There is no algorithm for solving a generic Diophantine equation, which is why they can be very difficult to solve. In fact, many open problems such as the Riemann hypothesis can be embedded in questions about Turing machines, which can in turn be converted to hideously complicated Diophantine equations. Even proving the non-existence of positive integer solutions to the innocuous-looking equation $x^n + y^n = z^n$ $(n \geq 3)$ (Fermat's last theorem) occupied mathematicians for three centuries before finally being settled by Andrew Wiles.

We will adopt a geometric approach to the problem, locating points in $\mathbb{Z}^n$, $\mathbb{N}^n$ and $\mathbb{Q}^n$ lying on some particular curve. The simplest curve is a straight line (or plane, or hyperplane), corresponding to a *linear Diophantine equation*. Linear Diophantine equations are easy. Let's consider the example $3x + 5y = 1$.

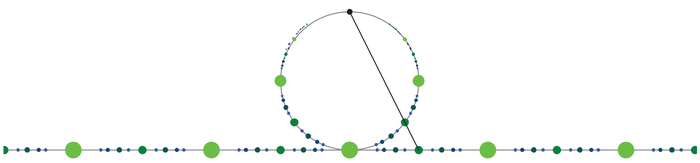


We can see that this has integer solutions, such as $(x, y) = (2, -1)$. In general, any linear Diophantine equation with coprime coefficients has infinitely many solutions in the integers, which can be found using the Chinese remainder theorem. If the coefficients are not coprime, such as $6x - 4y = 5$, there may be (as in this case) no solutions. If there are no solutions, a simple proof exists using modular arithmetic.

- A point (x, y) is *rational* if and only if both x and y are rational.

1. Let Γ be a conic with rational coefficients, and let A be a rational point on Γ. If B is another point on Γ, show that B is rational if and only if AB has rational gradient.

This theorem enables us to find all integer solutions to $a^2 + b^2 = c^2$, known as *Pythagorean triples* as they correspond to right-angled triangles with integer side lengths such as the famous (3, 4, 5) triangle used as a set square by the Ancient Egyptians. We can apply the geometrical concept of stereographic projection.



2. Show that all rational points on the unit circle can be obtained by inverting the rational line, as shown above.

3. Find all rational solutions to $x^2 + y^2 + z^2 = 1$.

4. Hence find all integer solutions to $a^2 + b^2 = c^2$, where $\gcd(a, b, c) = 1$. **[Elementary Pythagorean triples]**

5. Show that there exists an infinite set S of points, no three of which are collinear, such that the distance between any two points is rational.

Vieta jumping

Suppose we are attempting to determine whether or not a certain quadratic equation in two variables has integer solutions. This can be visualised as a conic section, which is either an ellipse, parabola or hyperbola. As an ellipse is finite, we need only check a few pairs of values for integer solutions. To deal with the parabola, we can change the coordinate system by applying an affine transformation to convert it into a simpler equation. For the hyperbola, however, it is necessary to adopt a more sophisticated technique such as *Vieta jumping*.

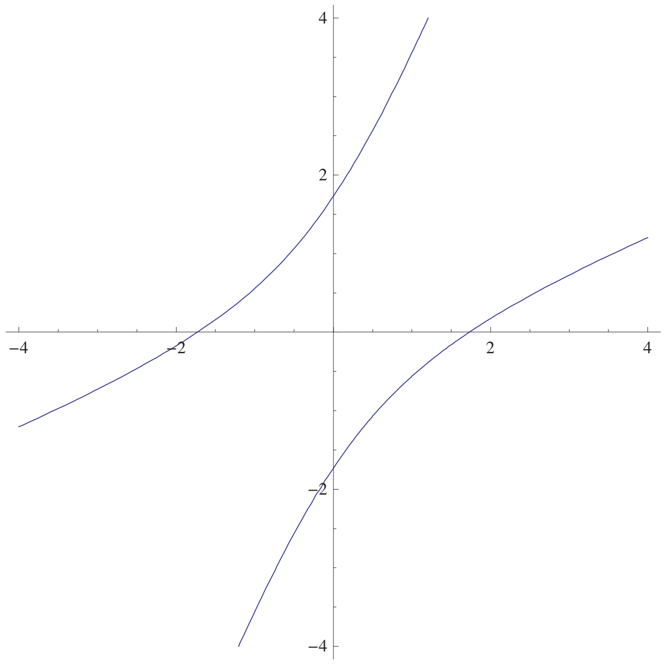
For example, consider the equation $a^2 + b^2 = k(a\,b + 1)$, where k is a fixed positive integer.

6. Find all integer solutions to the equation $a^2 + b^2 = k(a\,b + 1)$ for the cases where $k = 1$ and $k = 2$.

We will now consider the cases where $k \geq 3$.

7. Show that there are no integer solutions to $a^2 + b^2 = k(a\,b + 1)$ where $k \geq 3$ and one of a, b is negative and the other is positive.

8. Sketch the curve Γ described by the equation $x^2 + y^2 = k(x\,y + 1)$, for $k \geq 3$.



9. Suppose that $P = (a, b)$ is a positive integer solution, and draw a vertical line through P. Show that it meets Γ again at another non-negative integer point, $Q = (a, c)$. Also, if $b > a$, then show that $c < b$.

This is the principle behind Vieta jumping. We start with some (hypothetical) solution, then use it as the basis to construct a smaller solution until we reach a contradiction. Fermat employed this process of infinite descent to prove that there are no solutions in the positive integers to $x^4 + y^4 = z^4$. Euler later refined the approach to apply to the equation $x^3 + y^3 = z^3$.

10. If $a^2 + b^2 = k(a\,b + 1)$ (for some $k \geq 3$) has a solution in the integers, then show that there is a solution where $b = 0$.

11. Let a and b be positive integers. Prove that if $\frac{a^2+b^2}{a\,b+1}$ is a positive integer, then it is a perfect square. **[IMO 1988, Question 6]**

12. Let a and b be positive integers. Prove that if $\frac{a^2+b^2+1}{a\,b}$ is a positive integer, then it equals 3.

Pell equations

Vieta jumping is primarily useful for proving the non-existence of solutions to hyperbolic equations. What if, instead, we want to find an infinite family of solutions? The idea is to create new solutions from old by some form of recurrence relation. Firstly, consider equations of the form $x^2 = n\,y^2 + 1$, where n is not a perfect square. These are known as *Pell equations*, even though Pell had absolutely nothing to do with them. Basically, someone wanted to solve these equations, so told Euler that Pell was working on them. Consequently, Euler solved the equations, but attributed them to Pell.

13. Consider the complex numbers $z = a + b\sqrt{-n}$ and $w = c + d\sqrt{-n}$. Prove that $\left(a^2 + n\,b^2\right)\left(c^2 + n\,d^2\right) = (a\,c - b\,d\,n)^2 + n(a\,d + b\,c)^2$. **[Brahmagupta's identity]**

Brahmagupta's identity can also be verified algebraically, so is true even when n is negative. This gives us the related identity $\left(a^2 - n\,b^2\right)\left(c^2 - n\,d^2\right) = (a\,c + b\,d\,n)^2 - n(a\,d + b\,c)^2$.

14. Show that we can generate infinitely many solutions (in positive integers) to $x^2 = n\,y^2 + 1$ (where $\sqrt{n} \notin \mathbb{N}$) from one known solution.

15. Hence prove that there are infinitely many square triangular numbers.

As x and y can be very large, we obtain very good rational approximations $\frac{x}{y} \simeq \sqrt{n}$ in this manner. These rational approximations can also be generated by analysing the *continued fraction* expansion of $\sqrt{n}$, i.e. the unique expression of $\sqrt{n}$ of the following form:

$$x = \cfrac{1}{\cfrac{1}{\cfrac{1}{a_4 + \ldots} + a_3} + a_2} + a_1$$

We write this as $[a_1; a_2, a_3, a_4, \ldots]$. The sequences of all quadratic irrationals (solutions to quadratic equations in integer coefficients) are eventually periodic, and the converse also holds. For example, the sequence $[1; 1, 1, 1, 1, \ldots]$ corresponds to $\phi = \frac{1+\sqrt{5}}{2}$, and truncating the sequence produces the approximations $\frac{F_{n+1}}{F_n}$, where F_n is the nth Fibonacci number. Every Pell equation has infinitely many solutions. This is not true in general of the *negative Pell equation*, $x^2 = n\,y^2 - 1$ (where $\sqrt{n} \notin \mathbb{N}$); the negative Pell equation has solutions if the continued fraction has an odd period length.

The continued fraction for π is somewhat chaotic: $[3; 7, 15, 1, 292, \ldots]$. Truncating before the 292 yields the approximation $[3; 7, 15, 1] = [3; 7, 16] = \frac{355}{113}$, which agrees with the actual value of π to six decimal places. e has a very regular continued fraction expansion of $[2; 1, 2, 1, 1, 4, 1, 1, 6, 1, 1, 8, 1, \ldots]$.

Sums of squares

Brahmagupta's identity, in the case where $n = 1$, provides the identity $\left(a^2 + b^2\right)\left(c^2 + d^2\right) = (a\,c - b\,d)^2 + (a\,d + b\,c)^2$ as a special case. In other words, if S is the set of numbers that can be expressed as the sum of two squares (of integers, or, more generally, of rationals), then S is closed under multiplication. Using Hamiltonian quaternions, we can produce an analogous formula for four squares:

- $(a^2 + b^2 + c^2 + d^2)(e^2 + f^2 + g^2 + h^2) =$
  $(a\,e - b\,f - c\,g - d\,h)^2 + (a\,f + b\,e + c\,h - d\,g)^2 + (a\,g - b\,h + c\,e + d\,f)^2 + (a\,h + b\,g - c\,f + d\,e)^2$
  [Euler's four-square identity]

Using even more bizarre eight-dimensional numbers called octonions, there is a similar identity for eight squares. However, there are no identities beyond this, as doubling the number of dimensions causes the numbers to lose their useful properties. Quaternions are non-commutative, i.e. $x\,y \neq y\,x$ in general. Octonions are even worse, since they also lose associativity: $x(y\,z) \neq (x\,y)\,z$. Beyond this, the numbers have no useful properties remaining, and the 2^n-square identity breaks down.

It is interesting to see when an integer can be expressed as the sum of n squares of integers. Clearly, if N can be expressed as the sum of squares of n integers, then it can also be expressed as the sum of squares of $n + 1$ integers, as we can set one of those equal to zero. This gives a nested hierarchy:

- A positive integer N can be expressed as the sum of **one** square if and only if it is a perfect square, i.e. $N = a^2$ for some $a \in \mathbb{N}$. **[Trivial one-square theorem]**

- A positive integer N can be expressed as the sum of **two** squares if and only if it can be expressed as $N = a^2\, 2^k\,(p_1\, p_2 \ldots p_n)$, where each p_i is a prime congruent to 1 (modulo 4) and a and k are non-negative integers. **[Fermat's Christmas theorem]**

- A positive integer N can be expressed as the sum of **three** squares if and only if it is **not** of the form $4^k\,(8\,m + 7)$. **[Legendre's three-square theorem]**

- Any positive integer can be expressed as the sum of **four** squares. **[Lagrange's four-square theorem]**

16. Generalise each of the above theorems to determine when a rational number is expressible as the sum of one, two, three or four squares of rationals.

Sam Cappleman-Lynes technique

Consider the Diophantine equation $x^3 + y^6 = z^7$. If we have a solution to the equation $a^3 + b^6 = c$ (which is trivial), then we can multiply all terms by c^6 to obtain $a^3\,c^6 + b^6\,c^6 = c^7$, which is a solution to the original equation. This enables us to create infinitely many distinct solutions in this manner. This idea, known as the *Sam Cappleman-Lynes technique*, is applicable in many problems.

17. Show that there are infinitely many solutions to $x^4 + y^6 = z^{10}$ in the positive integers.

More generally, if a, b, c are pairwise coprime, then $x^{a\,k} + y^{b\,k} = z^{c\,k}$ has infinitely many solutions in the positive integers when $k = 1$ or $k = 2$. A consequence of Fermat's last theorem is that there are no solutions for $k \geq 3$.

18. Let A be the set of all integers of the form $a^2 + 13\,b^2$, where a and b are integers and b is non-zero. Prove that there are infinitely many pairs of integers x, y such that $x^{13} + y^{13} \in A$ but $x + y \notin A$. **[Mongolian TST]**

19. Determine whether there exists a set S of 2012 positive integers such that the sum of elements in each subset of S is a non-trivial power of an integer. **[IMO 1992 shortlist]**

A large proportion of the problems solved using the Sam Cappleman-Lynes technique reduce to this (very general) theorem, which is proved first by applying linear algebra to the simultaneous equations followed by a similar approach to the last question.

- Suppose we have a system of m equations in n variables $(x_1, x_2, \ldots, x_n)$ of the following form, where $m < n$:
  - $r_{(1,1)} x_1^{a_1} + r_{(1,2)} x_2^{a_2} + \ldots + r_{(1,n)} x_n^{a_n} = 0$;
  - $r_{(2,1)} x_1^{a_1} + r_{(2,2)} x_2^{a_2} + \ldots + r_{(2,n)} x_n^{a_n} = 0$;
  - …
  - $r_{(m,1)} x_1^{a_1} + r_{(m,2)} x_2^{a_2} + \ldots + r_{(m,n)} x_n^{a_n} = 0$.

  Suppose that the following conditions are also true:
  - All of the coefficients, $r_{(i,j)}$, are rational (not necessarily all non-zero);
  - All of the exponents, a_i, are positive integers;
  - $\det\begin{pmatrix} r_{(1,1)} & r_{(1,2)} & \cdots & r_{(1,m)} \\ r_{(2,1)} & r_{(2,2)} & \cdots & r_{(2,m)} \\ \vdots & \vdots & \ddots & \vdots \\ r_{(m,1)} & r_{(m,2)} & \cdots & r_{(m,m)} \end{pmatrix} \neq 0$;
  - For all $i \leq m$ and $i < j \leq n$, a_i is coprime with a_j;
  - There is at least one solution in the positive *real* numbers.

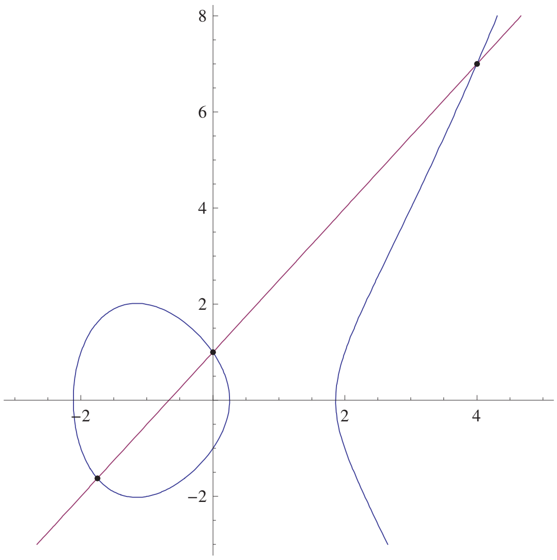
  In that case, there are infinitely many solutions in the positive integers. **[Generalised Sam Cappleman-Lynes theorem]**

The invertibility of the matrix enables us to apply elementary row operations to reduce it to a diagonal matrix in a process known as *Gauss-Jordan manipulation*. We then have m equations of the form $x_i^{a_i} = q_{(i,m+1)} x_{m+1}^{a_{m+1}} + q_{(i,m+2)} x_{m+2}^{a_{m+2}} + \ldots + q_{(i,n)} x_n^{a_n}$ (one equation for each $1 \leq i \leq m$), and a real solution to the equations. This is a linear equation in $x_i^{a_i}$ (for all $1 \leq i \leq n$), so can be represented by a $(n-m)$-dimensional hyperplane in n-dimensional space (which passes through the origin, like a projective $(n-m-1)$-hyperplane). The single real solution means that this hyperplane intersects the positive quadrant, so we can set initial rational values for each of the *free variables* $\{x_{m+1}, x_{m+2}, \ldots, x_n\}$. This forces every $x_i^{a_i}$ to be positive and rational. We then apply the Sam Cappleman-Lynes technique to the equations, relying on the coprimality of various exponents. After all variables have been rationalised, we multiply out by a large power of the lowest common multiple of the denominators of the variables, turning them into integer solutions.

Elliptic curves

20. Let Γ be a non-singular cubic curve with integer coefficients. Suppose we have a line Λ which meets Γ at three points, A, B and C. Prove that if A and B are rational, then C is also rational.

For example, the curve $y^2 = x^3 - 4x + 1$ has rational points $(0, 1)$ and $(4, 7)$. We then draw the line through the two points, namely $y = \frac{3}{2}x + 1$. This intersects the curve at a third point, which is the solution of the cubic equation $\left(\frac{3}{2}x + 1\right)^2 = x^3 - 4x + 1$, which simplifies to $x^3 - \frac{9}{4}x^2 - 7x = 0$. We can factorise this, as we already know two of the roots, obtaining $x(x-4)\left(x + \frac{7}{4}\right) = 0$. This gives us a third rational point on the elliptic curve, namely $\left(-\frac{7}{4}, -\frac{13}{8}\right)$.



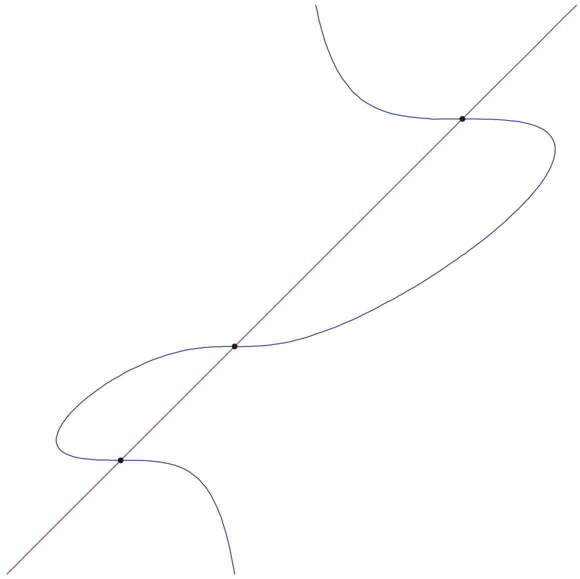
One of the great unsolved problems in mathematics, the Birch and Swinnerton-Dyer conjecture, is concerned with counting the number of integer points on an elliptic curve modulo p. Elliptic curves also featured prominently in Andrew Wiles' proof of the Tanayama-Shimura conjecture and Fermat's last theorem. Specifically, Wiles showed that every elliptic curve could be associated with a 'modular form', a complex function with the same hyperbolic symmetries as the Ford circles. Even the simplest, most elementary proof of Poncelet's porism involves elliptic curves.

We define a binary operation $+$ on the points on the cubic curve, such that $A + B + C = 0$ for any three collinear points $A, B, C \in \Gamma$. In other words, $C = -(A + B)$. Let $O = 0$ be one of the points of inflection, so the line passing through C and O meets Γ again at $C' = A + B$. It is trivially obvious that this operation is commutative, but proving associativity is a little more difficult.

21. Let X, Y, Z be three non-collinear points on Γ. Show that $(X + Y) + Z = X + (Y + Z)$, where addition is defined as in the last paragraph. **[Associativity of elliptic curve operation]**

After proving associativity, parentheses can be omitted from expressions without ambiguity. For example, we can refer to the last expression simply as $X + Y + Z$. Addition of elements forms a group operation. We can then multiply by an integer n, by defining $n X = X + X + \ldots + X$. It is difficult to compute n from the points $n X$ and X, so can form the basis of a cryptosystem similar to RSA using the group $\{\Gamma, +\}$ instead of $\{\mathbb{Z}_{pq}, \times\}$. Elliptic curve cryptography is considered to be more secure than RSA. Again, it is susceptible to attacks from quantum computers.

22. Let A, B, C be three collinear points on Γ. If A and B are both points of inflection, then show that C is also a point of inflection.



In the above diagram, a line is drawn through the three real points of inflection of a cubic curve. A general cubic curve on the complex projective plane has nine points of inflection lying on twelve lines in what is known as the *Hesse configuration*. It is a remarkable fact that this cannot be embedded in the *real* projective plane due to the *Sylvester-Gallai theorem*: if there is a finite set S of points such that no line contains exactly two points, then all points are collinear. Hence, a cubic curve has at most three real points of inflection.

Solutions

1. Clearly, if AB has irrational gradient, then A is the only rational point on Γ. If AB has rational gradient, then we can express it as a linear equation in rational coefficients. By solving this simultaneously with the equation for Γ, we obtain a quadratic equation in rational coefficients for x (the abscissa). As one root of this (the coordinates of A) is rational, then the other root must also be. Repeating this process for y (the ordinate), it is evident that $B \in \mathbb{Q}^2$.

2. Inverting about a rational point A on the unit circle transforms the circle into a line by stereographic projection. All lines with rational slopes through A clearly correspond to the entire set of rational points on the real line.

3. This is the three-dimensional analogue of the problem. We want to transform a point $\overrightarrow{OA}$ on the horizontal plane into a vector $\overrightarrow{OB}$ on the unit sphere by inverting about a sphere with centre $C = (0, 0, 1)$ and radius $\sqrt{2}$. We have that $\overrightarrow{CA}$ and $\overrightarrow{CB}$ are parallel, and the product of their lengths is 2. Hence, this gives us the formula $\overrightarrow{CB} = \frac{2\overrightarrow{CA}}{|\overrightarrow{CA}|^2}$. If we let A have coordinates $\left(\frac{p}{q}, \frac{r}{q}, 0\right)$, where p, q and r are coprime, then we obtain $\overrightarrow{CB} = \frac{2\left(\frac{p}{q}, \frac{r}{q}, -1\right)}{\left(\frac{p}{q}\right)^2 + \left(\frac{r}{q}\right)^2 + 1}$. This simplifies to $\overrightarrow{CB} = \frac{2(pq, rq, -q^2)}{p^2+q^2+r^2}$. Now, $\overrightarrow{OB} = \overrightarrow{OC} + \overrightarrow{BC} = \frac{(2pq, 2rq, p^2+r^2-q^2)}{p^2+q^2+r^2}$, giving us the complete set of solutions; $x = \frac{2pq}{p^2+q^2+r^2}$, $y = \frac{2rq}{p^2+q^2+r^2}$, $z = \frac{p^2-q^2+r^2}{p^2+q^2+r^2}$, where $p, q, r \in \mathbb{Z}$, $q > 0$ and $\gcd(p, q, r) = 1$.

4. a and b cannot both be odd; assume without loss of generality that a is even. Hence, b is odd as otherwise a, b and c would all be even. Using the previous question (and setting $r = 0$), the solutions to $\left(\frac{a}{b}\right)^2 + \left(\frac{a}{c}\right)^2 = 1$ are $\frac{a}{c} = \frac{2pq}{p^2+q^2}$ and $\frac{b}{c} = \frac{p^2-q^2}{p^2+q^2}$. If p and q are both odd then c is even, so one of p and q must be even. Hence, $p^2 - q^2$ and $p^2 + q^2$ are coprime (by applying Euclid's algorithm to obtain $2p^2$ and $2q^2$, which only share 2 as a common factor). This gives us the irreducible general solution; $(a, b, c) = \left(2pq, p^2 - q^2, p^2 + q^2\right)$, where p and q are coprime integers.

5. Let AB be the base of a semicircle with unit radius. A point C on the curved edge of the semicircle belongs to S if and only if both AC and BC are rational. (The previous question guarantees infinitely many choices of C.) If we have two such points, C_1 and C_2, (such that A, C_1, C_2, B appear in that order) then $C_1C_2 \cdot AB = C_1B \cdot C_2A - C_1A \cdot C_2B$ (by Ptolemy's theorem), and thus C_1C_2 is rational.

6. For $k = 2$, the equation $a^2 + b^2 = 2ab + 2$ simplifies to the $(a - b)^2 = 2$, which clearly has no integer solutions. For $k = 1$, $a^2 + b^2 = 1 + ab$ is an ellipse. As $a^2 + b^2 \geq 2ab$, then $ab \leq 1$. Also, $ab \geq -1$, since otherwise $a^2 + b^2$ would be negative. It is a simple matter to check all combinations of $a, b \in \{-1, 0, 1\}$ and deduce that the solutions $(0, 1)$, $(1, 0)$, $(0, -1)$, $(-1, 0)$, $(1, 1)$ and $(-1, -1)$ are the only solutions.

7. If one of a, b is negative and the other is positive, then $ab < 0$. Hence, $ab \leq -1$ and thus $k(ab + 1) \leq 0$. However, $a^2 + b^2 = k(ab + 1)$ is strictly positive.

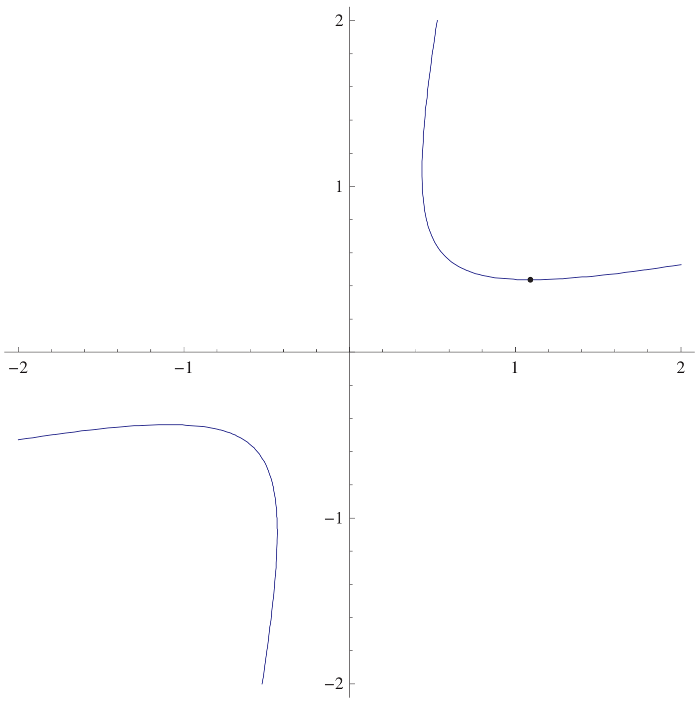
8. If we apply the substitution $z = x + y$, $w = x - y$ then we obtain a hyperbola in standard position. Hence, the original curve is a hyperbola with centre $(0, 0)$ and an axis of symmetry $y = x$, which does not intersect the hyperbola on the real plane.

9. As $a^2 + b^2 = k(ab + 1)$, b is a root of the quadratic $z^2 - kaz + \left(a^2 - 1\right) = 0$. By Vieta's formulas, we have $b + c = ka$, and thus $c = ka - b$. This is obviously an integer point. As $b > a$, P is on the upper branch of the hyperbola, so Q must be on the lower branch and therefore $c < b$.

10. Assume that $P = (a, b)$ is a positive integer solution and that $b > a$. We can generate a smaller solution in the non-negative integers by choosing the other intersection point $Q = (a, c)$ of the line $x = a$ and curve Γ. By reflecting in the line $y = x$, we obtain a solution $P' = (c, a)$, where $a > c$. This process will generate continually smaller solutions until one of the coordinates is zero.

11. Let $\frac{a^2+b^2}{a\,b+1} = k$. If $k = 2$, then there are trivially no solutions. Otherwise, if $k \geq 3$ and we have an integer solution (a, b), we can generate a solution where $b = 0$. So, there is a solution in the positive integers if and only if there is a value k such that $\frac{a^2}{1} = k$, which only occurs when k is a perfect square.

12. Consider the equation $a^2 + b^2 + 1 = a\,b\,k$. This can be algebraically manipulated to $(a - b)^2 + 1 = (k - 2)\,a\,b$; hence, it is obvious that $k \geq 3$. We now fix the value of k. As there are no solutions for $a = 0$ or $b = 0$, and every solution (a, b) yields alternative solutions $(-a, -b)$ and (b, a), the curve must be a hyperbola with centre $(0, 0)$ and diagonal lines of symmetry. Moreover, the hyperbola is contained entirely within the first and third quadrants. Using Vieta jumping, we can get from a solution (a, b) to $(b\,k - a, b)$. If (a, b) lies to the right of the stationary point of the hyperbola, then this will generate a smaller (in terms of $a + b$) solution. We can then reflect in the line $a = b$ and repeat the process until (a, b) lies to the left of this stationary point and below (or on) the line $a = b$. By considering the discriminant of $a^2 - b\,k\,a + b^2 + 1 = 0$, the stationary point can be located as $\left(\frac{k}{\sqrt{k^2-4}}, \frac{2}{\sqrt{k^2-4}}\right)$. The only integer point lying to the left of this such that $a \geq b$ is $(1, 1)$, which clearly is only a solution when $k = 3$. The graph below is the hyperbola for $k = 5$.



13. This is simply the equation $|z|^2\,|w|^2 = |z\,w|^2$.

14. If we have two solutions (not necessarily distinct), (a, b) and (c, d), then $a^2 - n\,b^2 = 1$ and $c^2 - n\,d^2 = 1$. We can multiply them together and use Euler's identity to obtain $(a\,c + b\,d\,n)^2 - n(a\,d + b\,c)^2 = 1$, which is a strictly larger solution to the equation. Repeating the process from some initial solution (b, a), we expand $\left(a^2 - n\,b^2\right)^k$ using Euler's identity recursively to obtain infinitely many solutions, one for each $k \in \mathbb{N}$.

15. For every triangular number T, $8\,T + 1$ is square and the converse also holds. Hence, we want to solve the Pell equation $x^2 = 8\,y^2 + 1$. A preliminary solution is that 36 is both square and triangular, or $17^2 = 8 \cdot 6^2 + 1$, from which we generate an infinitude.

16. Suppose we have a rational $\frac{a}{b}$, where a and b are coprime. Squaring it results in $\frac{a^2}{b^2}$, where a^2 and b^2 are still coprime. Hence, squares of rationals (or sums of one square) have both a square numerator and denominator when expressed in lowest terms. Suppose $\frac{a}{b} = \frac{c^2}{d^2} + \frac{e^2}{f^2} = \frac{c^2\,f^2 + e^2\,d^2}{d^2\,f^2}$ is expressible as the sum of two rational squares. Then, it is expressible as a square of a rational multiplied by the sum of squares of two integers, therefore of the form $N = t^2\,2^k\,(p_1\,p_2\,\ldots\,p_n)$, where each p_i is a prime congruent to 1 (modulo 4), k is an integer and t is a non-negative rational. For sums of three squares of rationals, we can express it as a square of a rational multiplied by the sum of squares of three integers. Hence, it is something that **cannot** be expressed as $\frac{4^k\,(8\,m+7)}{t^2}$, where t is an odd integer, k is an integer and m is a non-negative integer. Any rational can be expressed as the sum of four squares of rationals.

17. We begin with a solution to $a^2 + b^2 = c^2$, i.e. a Pythagorean triple. Multiplying by a^2 gives us $a^4 + b^2\,a^2 = c^2\,a^2$. Then multiply by $b^4\,a^4$, giving us $b^4\,a^8 + b^6\,a^6 = c^2\,b^4\,a^4$. Finally, multiply by $c^{48}\,b^{96}\,a^{96}$, resulting in $c^{48}\,b^{100}\,a^{104} + c^{48}\,b^{102}\,a^{102} = c^{50}\,b^{100}\,a^{100}$. This is clearly a solution to the equation $x^4 + y^6 = z^{10}$.

18. If $x \equiv 1 \pmod 4$ and $y \equiv 2 \pmod 4$, then clearly $x + y \notin A$. All numbers of the form $13\,b^2$ are in A, so if we can find infinitely many solutions satisfying $x^{13} + y^{13} = 13\,b^2$ and the modulo-4 congruences then we are done. Start with a solution to $w^{13} + z^{13} = 13\,u$, where $w \equiv 1$ and $z \equiv 2 \pmod 4$. We note that u must be congruent to 1 (mod 4), as $w^{13} \equiv 1$ and $z^{13} \equiv 0$. We multiply by u^{13} to give a solution $(u\,w)^{13} + (u\,z)^{13} = 13\,(u^7)^2$, without changing the congruence class of any of the variables. We can start with $w = 13 + 52\,k$ and $z = 26$, generating a solution for every $k \in \mathbb{N}$. As these solutions can become arbitrarily large, there must be infinitely many.

19. For each $1 \le j \le 2^{2012} - 1$, consider the Diophantine equation of the form $\sum a_i^2 = b_j^{p_j}$, where a_i $(1 \le i \le 2012)$ is included in the sum if and only if the ith binary digit of j is 1, and p_j is the jth odd prime. We then apply the Sam Cappleman-Lynes technique to all equations simultaneously starting from the $2^{2011} - 1$ equations of the form $\sum x_i^2 = y_j$. We initially set $x_i = i$, and then multiply all equations by an appropriate power of $y_j^{2\,p_1\,p_2\,\cdots\,p_{j-1}}$. This preserves the perfect-power property of all previous equations, and we can do this until the jth equation is also converted to the desired form. Repeating for all $2^{2012} - 1$ equations gives us a solution to the original Diophantine equations. Then just take $S = \{a_1^2, a_2^2, \ldots, a_{2^{2012}-1}^2\}$.

20. The line passes through two rational points, so must have rational gradient. Hence, we can express y as a linear function of x with rational coefficients. Let a, b, c be the abscissae of A, B, C, respectively. The intersection of the cubic curve and the line is a cubic equation with rational coefficients and roots a, b, c. Due to Vieta's formulas, $a + b + c$ is one of the rational coefficients. Hence, the rationality of a and b implies that of c.

21. Let Φ be the union of the lines $\{Y, -(X+Y), X\}$, $\{-(Y+Z), O, Y+Z\}$ and $\{Z, X+Y, -((X+Y)+Z)\}$. Similarly, let Ψ be the union of the lines $\{Y, -(Y+Z), Z\}$, $\{-(X+Y), O, X+Y\}$ and $\{X, Y+Z, -(X+(Y+Z))\}$. The three cubic curves Γ, Φ and Ψ intersect in eight points, namely $\{X, Y, Z, -(X+Y), -(Y+Z), X+Y, Y+Z, 0\}$, so must intersect in the ninth point by Cayley-Bacharach. Hence, $-((X+Y)+Z) = -(X+(Y+Z))$, and thus $(X+Y)+Z = X+(Y+Z)$.

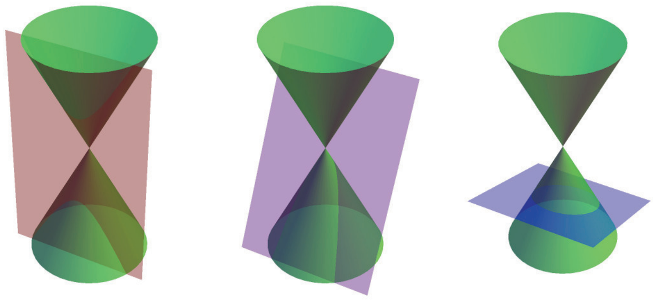
22. The tangent lines passing through A and B do so with multiplicity 3, so we have $3\,A = 3\,B = 0$. We then have $3\,C = -3\,(A+B) = 0$, so C is also a point of inflection.

Conic sections

Conics have recurred throughout this book in both geometric and algebraic settings. Hence, I have decided to dedicate the final chapter to them. As the only conics appearing on IMO geometry problems are invariably circles, the results proved in this chapter are largely irrelevant. Nevertheless, the material is sufficiently interesting to be worthy of inclusion.

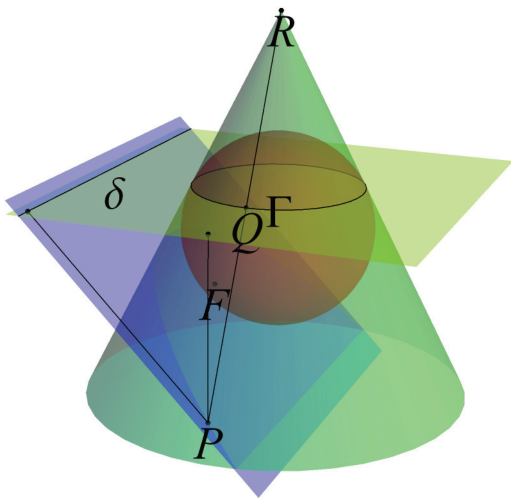
Sections of cones

With the obvious exception of the circle, the conics were first discovered by the Greek mathematician Menaechmus who contemplated slicing a right circular cone C with a flat plane Λ. Indeed, the term 'conic' is an abbreviation of *conic section*.



It is more natural to consider C as the double cone with equation $x^2 + y^2 = z^2$. If Λ cuts both cones, then the conic section is a *hyperbola*. If it cuts only one cone in a closed curve, it is an *ellipse*. The intermediate case, where the plane is inclined at exactly the same slope as the cone, results in a *parabola*.

Observe that $x^2 + y^2 = z^2$ is the equation of a projective circle; this explains why all conic sections are equivalent under projective transformations.



We define a Dandelin sphere Ω to be a sphere tangent to both C (at a circle Γ) and Λ (at a point F, namely the *focus*). The plane containing Γ intersects Λ at a line δ, known as the *directrix*.

1. Prove that the directrix is the polar of the focus.

For an arbitrary point P on the conic, we let PR meet Γ at Q.

2. Prove that $PQ = PF$.

3. Let A be the foot of the perpendicular from P to the plane containing Γ. Let D be the foot of the perpendicular from P to δ. Show that $\frac{PQ}{PD}$ is independent of the location of P.

By combining the two previous theorems, we establish the *focus-directrix property* of a conic section.

- For every point P on a conic section, the ratio $\frac{PF}{P\delta} = \varepsilon$ remains constant. ε is known as the *eccentricity* of the conic. **[Focus-directrix property]**

The type of conic section can be determined by its eccentricity.

4. Show that $\varepsilon < 1$ for an ellipse, $\varepsilon = 1$ for a parabola and $\varepsilon > 1$ for a hyperbola.

Conics on a plane

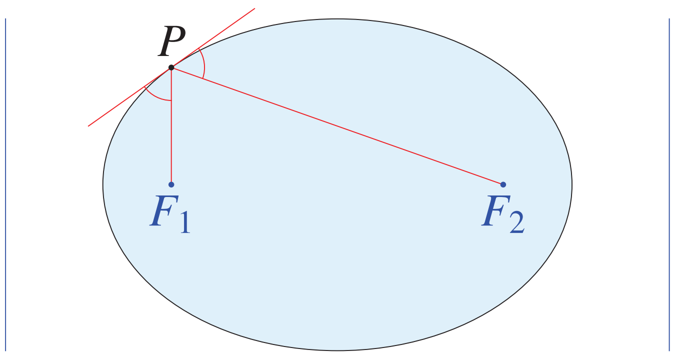
The focus-directrix property enables us to give conic sections a Cartesian treatment. By allowing the directrix to be the x-axis and scaling the conic so that the focus is at (0, 1), the equation of a conic becomes $x^2 + (y-1)^2 = \varepsilon^2 y^2$. We can see that a conic section is a quadratic curve (although this is obvious from the projective definition). More remarkably, the converse is also true: all non-degenerate quadratic curves are conic sections.

5. Prove that, if $\varepsilon \neq 1$, the conic has two lines of reflectional symmetry.

Hence, for ellipses and hyperbolae, we can reflect the focus and directrix in the line of symmetry to obtain an alternative focus and directrix. Returning to the Dandelin spheres, the other focus corresponds to placing the sphere below the plane instead of above it. A parabola can be regarded as an ellipse/hyperbola with a focus on the line at infinity.

The Cartesian equation also makes it evident that ellipses are indeed 'squashed (affine transformed) circles'. We can translate the ellipse so that the lines of symmetry are coordinate axes, giving us the equation for an ellipse.

- $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$ is the general equation for an ellipse. If we reverse the sign of $\frac{y^2}{b^2}$, we obtain a hyperbola instead. **[Cartesian equations for ellipses and hyperbolae]**

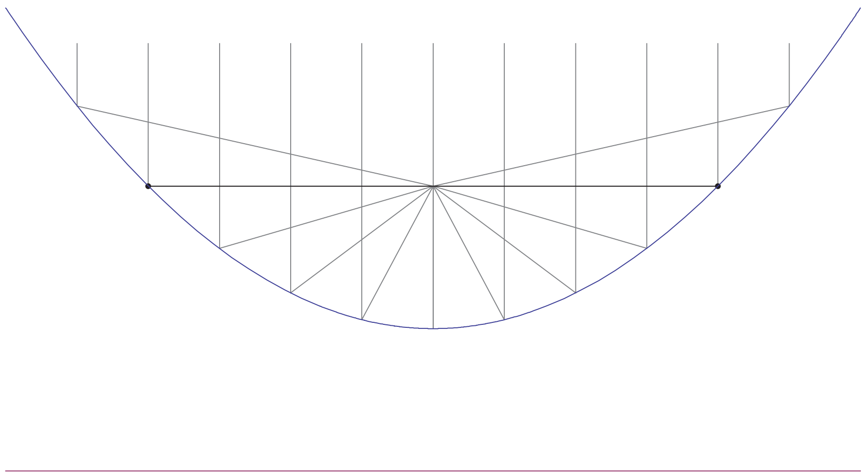


Consider an ellipse, with its two foci (F_1 and F_2) and directrices. Let P be a variable point on the ellipse.

6. Prove that $PF_1 + PF_2$ is constant.

7. Show that the angles between the tangent at P and the lines PF_1 and PF_2 are equal. **[Reflector property of the ellipse]**

A parabola can be considered to be the limit of ellipses with one focus fixed and the other tending towards infinity. This gives us the reflector property of the parabola, which states that a pencil of rays originating from the focus is reflected to a pencil of parallel lines perpendicular to the directrix. This was known to Archimedes, and formed the basis of a mechanism for igniting the sails of enemy ships by reflecting sunlight from polished metal shields. Nowadays, it is used in Newtonian telescopes for focusing light from infinity.



Let the focus be the origin, and the directrix be represented with $x = d$ in Cartesian coordinates. Consider the polar coordinates $\langle r, \theta \rangle = (r\cos\theta,\ r\sin\theta)$.

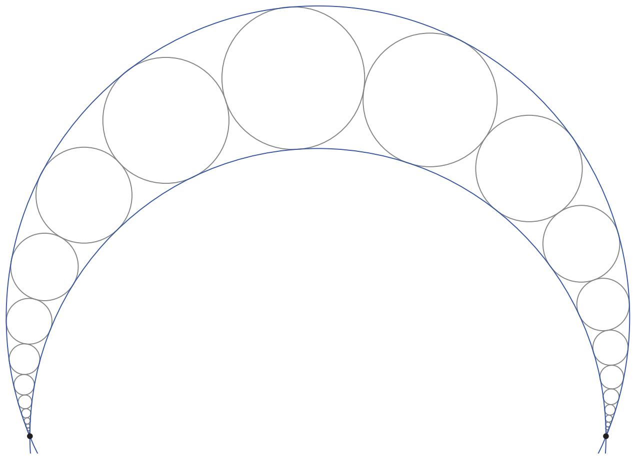
8. Prove that the equation for a conic in polar coordinates is $r = \frac{l}{1+\varepsilon\cos\theta}$, where $l = \varepsilon d$. **[Polar equation of a conic]**

This parametrisation of a conic will prove useful when verifying Kepler's laws of planetary motion in the next section.

The length $l = \varepsilon d$ is known as the *semi-latus rectum*, as it is half of the length of the line segment parallel to the directrix passing through the focus and meeting the conic twice. The latus rectum is shown in black on the diagram of the parabolic reflector.

The return of Steiner's porism

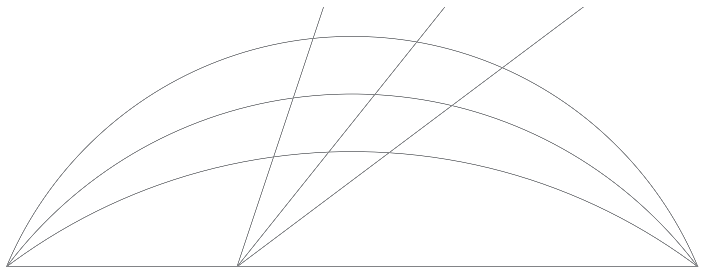
In Steiner's porism, we considered the family of circles internally tangent to Γ_2 and externally tangent to Γ_1. Let's suppose these two circles now intersect. Instead of a finite Steiner chain, we obtain an infinite set of circles bounded by Γ_1 and Γ_2.



9. Show that the centres of the circles lie on an ellipse, the foci of which are the centres of Γ_1 and Γ_2.

10. Hence demonstrate that the radius of the variable circle is proportional to the distance between its centre and the radical axis of Γ_1 and Γ_2.

11. Three circular arcs, γ_1, γ_2 and γ_3, connect the points A and C. These arcs lie in the same half-plane defined by the line AC in such a way that γ_2 lies between γ_1 and γ_3. Three rays, h_1, h_2 and h_3, emanate from a point B on the line AC, resulting in a grid of four curvilinear quadrilaterals as shown in the diagram below. Prove that if one can inscribe a circle in each of three of the curvilinear quadrilaterals, then a circle can be inscribed in the fourth. **[IMO 2010 shortlist, Question G7, Géza Kós]**



Géza created many more problems on this theme, all of which are amenable to embedding in three-dimensional space. One equivalent problem is where the situation is on the surface of a sphere, which can be transformed into the original problem through stereographic projection. Similarly, there are variants in hyperbolic space.

Kepler's laws of planetary motion

The vast majority of the content of this book is exclusively in the realms of pure mathematics. Nevertheless, conic sections naturally occur as the paths traced by objects in gravitational fields. The elliptical orbits of planets were first proposed by Johannes Kepler, as a refinement of earlier (mostly Greek) ideas that celestial bodies travel in perfect circles. Isaac Newton later inferred his law of gravitation from Kepler's laws; the derivation is not too difficult, although it relies heavily upon differential calculus.

Consider an object of negligible mass moving around a fixed object O of large mass due to gravitational attraction. An example of this is the Earth orbiting the Sun. We aim to show that the path must be a conic section, by showing that the polar equation of a conic satisfies Newton's laws of gravitation. Suppose an object is initially at P and moves to Q (very close to P). Let $A = [POQ]$ be the area 'swept out' by the object, and consider the derivative $\frac{dA}{dt}$, known as the *areal velocity*.

12. If an object moves in a straight line at constant velocity, show that $\frac{dA}{dt}$ is constant.

13. If the acceleration of an object is entirely radial (towards or away from O) at all times, then show that $\frac{dA}{dt}$ again remains constant. **[Conservation of angular momentum]**

Indeed, the converse is also true: if areal velocity is conserved, then acceleration is entirely radial. By integrating $\frac{dA}{dt}$ with respect to time, we obtain Kepler's second law.

- A planet P in orbit around the Sun O sweeps out equal areas in equal intervals of time. **[Kepler's second law]**

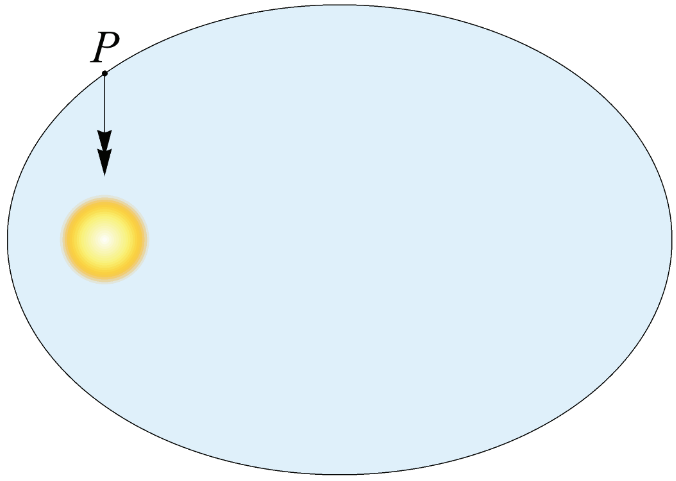
If $P = \langle r, \theta \rangle$, then $\frac{dA}{dt} = \frac{1}{2} r^2 \frac{d\theta}{dt}$. So, the value of $r^2 \frac{d\theta}{dt}$ must remain constant. Let's refer to this value (twice the areal velocity) as k.

In ordinary circular motion, the radial acceleration is given by $-r\left(\frac{d\theta}{dt}\right)^2$. Hence, in the general case, radial acceleration equals $a = \frac{d^2 r}{dt^2} - r\left(\frac{d\theta}{dt}\right)^2$.

14. Show that, if the planet P follows the path of a conic with focus O and semi-latus rectum l, then $a = -\frac{k^2}{r^2 l}$. **[Newton's inverse square law]**

Conversely, if we assume the inverse square law $a = -\frac{GM}{r^2}$, then we can choose a conic with centre O, passing through P in the appropriate direction, and with a latus rectum of $l = \frac{k^2}{GM}$. As Newton's law of universal gravitation is deterministic, the conic must be the **unique** solution. Hence, the converse is also true: all planets obeying the inverse square law travel in conic orbits. If the orbit is cyclic (and therefore closed), then it must be an ellipse.

- A planet P describes an ellipse, one focus of which is the Sun O. **[Kepler's first law]**



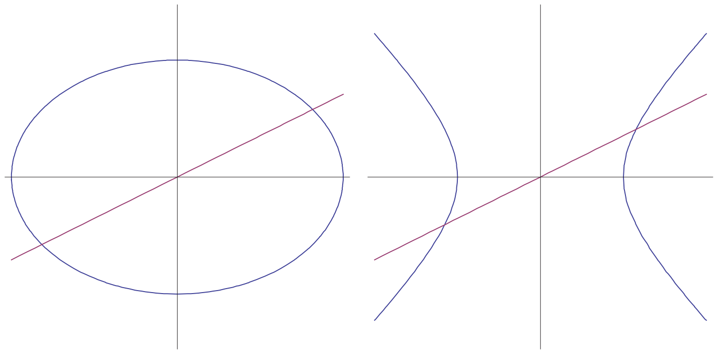
By considering the equation $r = \frac{l}{1+\varepsilon\cos\theta}$, the value of r is minimised at $\frac{l}{1+\varepsilon}$ (the *perihelion*) and maximised at $\frac{l}{1-\varepsilon}$

(the *aphelion*). When the eccentricity is zero, the orbit is circular.

In general, two bodies experiencing gravitational attraction will orbit each other in coplanar conic orbits, where the barycentre of the system (assumed to be stationary) is their common focus. For three or more bodies, the equations cannot be solved algebraically, and the system behaves chaotically (arbitrarily small initial perturbations lead to arbitrarily large effects). Indeed, it has been shown to be undecidable, so no computer or Turing machine is capable of calculating the movements with perfect precision.

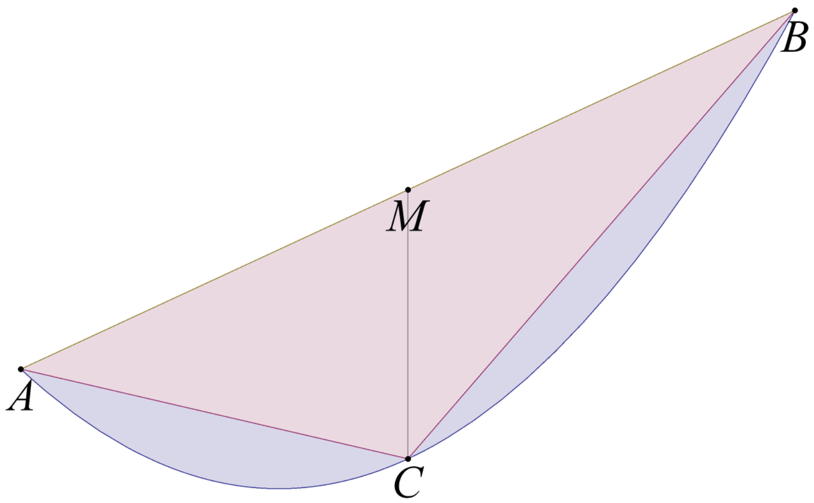
Areas of conics

If we take the pole of the line at infinity, we obtain the *centre* of a conic. For parabolae, this point is at infinity, therefore does not lie on the affine plane. For ellipses and hyperbolae, however, the centre lies on the plane and can be taken as the origin. It is then possible to apply a rotation about the origin to place the conic in *standard position*.



The ellipse has Cartesian equation $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$. The line $y = x \tan\theta$ meets the curve at $P = (a\cos\theta,\ b\sin\theta)$. The area bounded by the curve, the line OP and the positive x-axis is given by $\frac{1}{2} a b \theta$. In particular, when $\theta = 2\pi$, the total area of the ellipse is $\pi a b$.

Similarly, the hyperbola has Cartesian equation $\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$. The line $y = x \tanh\phi$ meets the curve at $P = (a\cosh\phi,\ b\sinh\phi)$. The area bounded by the curve, the line OP and the positive x-axis is given by $\frac{1}{2} a b \phi$. The hyperbolic functions are defined in a similar way to the trigonometric functions, with $\cosh\phi = \frac{1}{2}\left(e^{\phi} + e^{-\phi}\right)$ and $\sinh\phi = \frac{1}{2}\left(e^{\phi} - e^{-\phi}\right)$.



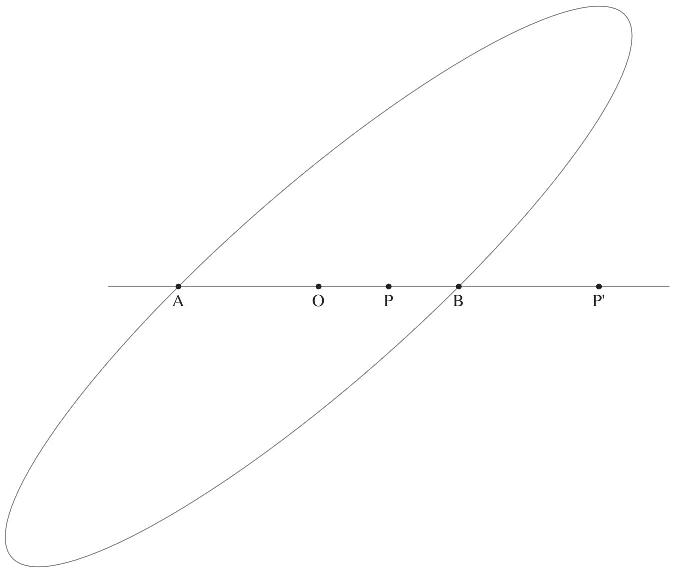
The area of a parabolic segment is much easier to calculate, as it can be obtained by integration of the equation of the parabola, $y = \frac{x^2}{4l}$, with respect to x. Archimedes instead used completely Euclidean methods in his *Quadrature* *of the Parabola*, determining the area recursively by adding together the area of triangle ABC with the areas of the parabolic segments below AC and BC. M is the midpoint of AB, and C is the intersection of the parabola with the perpendicular from M to the directrix. The triangle constructed by repeating this process with AC instead of AB has one eighth of the area of the original triangle. By summing an infinite geometric series, the total area is equal to $[ABC]\left(1+\frac{2}{8}+\frac{4}{64}+\frac{8}{512}+\ldots\right)=\frac{4}{3}[ABC]$.

Unlike areas, the arc lengths of ellipses are not easy to compute. The circumference of an ellipse is $4\,a\,E(\varepsilon)$, where E is the *complete elliptic integral of the second kind*. With the exception of the circle, where $E(0)=\frac{\pi}{2}$ and the circumference is $2\,\pi\,r$, $E(\varepsilon)$ cannot be expressed in terms of basic functions.

Inversion in arbitrary conics

In the chapter about the Riemann sphere, we considered inversion in a circle. It is, however, possible to invert about any conic section. We have a seven-parameter set of inversions we can apply, as the centre of inversion and conic can be chosen independently.

- Let Γ be a non-degenerate conic, and O be a point not on Γ. For any point P other than O, we draw the line l through O and P, and let it meet Γ at A and B. We then define P' to be the projective harmonic conjugate of P with respect to AB. **[Inversion in a conic]**



Equivalently, we can define P' as the intersection of the polar of P with the line OP.

To investigate the properties of conic inversion, apply a projective transformation to make Γ a circle and O its centre. Then, the basic theorems applying to ordinary inversion translate into projective versions.

- O inverts to an entire line Ω, namely the polar of O, and *vice-versa*. (If O is the centre of the conic, then this is the line at infinity.) We treat this line as a single point, so the projective plane becomes topologically equivalent to a sphere. Allow Ω to intersect Γ at the points I and J.

In this perverse world of conic inversion, Ω behaves like the point-line at infinity and I and J are analogous to the circular points. This enables us to convert theorems in circle inversion to their conic counterparts.

- Straight lines passing through O remain invariant under inversion.

- Conics containing O, I and J invert to straight lines not passing through O, and *vice-versa*.

- Conics containing I and J (but not O) invert to other conics containing I and J (but not O).

In circle inversion, angles between curves remain constant (or, more precisely, are reversed). In conic inversion, this must be converted into a projective statement.

15. If P and Q invert to P' and Q', respectively, then show that $P\,Q\,P'\,Q'\,I\,J$ are conconic.

16. Let curves C and D intersect at P. Let C' and D' be the inverse curves with respect to Γ, and let P' be the inverse of P. The tangent to C at P and the tangent to D' at P' intersect at R. Similarly, the tangent to D at P and the tangent to C' at P' intersect at S. Show that $P\,P'\,R\,S\,I\,J$ are conconic. **[Preservation of generalised angle]**

Solutions

1. Apply a projective transformation to take δ to infinity, then apply an affine transformation to return Ω to being a sphere. The cone is tangent to Ω, so remains a right circular cone. The plane containing Γ becomes parallel to the plane Λ (i.e. horizontal). By symmetry, F is now the centre of the conic (which is a circle), and therefore the pole of the line at infinity δ.

2. They are both tangents from P to Ω, therefore of equal length.

3. The angle $\angle APQ$ in the right-angled triangle is constant (equal to half the angle at the vertex of the cone), so $\frac{PA}{PQ}$ is independent of the location of P. Similarly, $\frac{PD}{PA}$ is also constant, by considering the right-angled triangle PAD.

4. Observe that $\varepsilon = \frac{PQ}{PD} = \frac{\cos \angle HPD}{\cos \angle HPQ}$. When $\varepsilon = 1$, the plane Λ is inclined at the same slope as the cone, thus creating a parabola. When $\varepsilon < 1$, the plane is shallower than this, so the conic is an ellipse. Conversely, when the eccentricity exceeds 1, we have a hyperbola.

5. The equation of the conic is $x^2 + \left(1 - \varepsilon^2\right) y^2 - 2y + 1 = 0$. We can complete the square, resulting in $x^2 + \left(1 - \varepsilon^2\right)\left(y - \frac{1}{1-\varepsilon^2}\right)^2 - \frac{\varepsilon^2}{1-\varepsilon^2}$. This is symmetric about the lines $x = 0$ and $y = \frac{1}{1-\varepsilon^2}$.

6. Let the feet of the perpendiculars from P to the directrices be D_1 and D_2. We have $PF_1 + PF_2 = \varepsilon(PD_1 + PD_2) = \varepsilon(D_1 D_2)$.

7. Let Q be another point on the ellipse, very close to P. Let $F_1 P = u$, $F_1 Q = u + \delta$ and $F_2 Q = v$. By the previous theorem, $F_2 P = v + \delta$. Let $PQ = \gamma$. We apply the cosine rule, to get $\cos \angle F_1 PQ = \frac{u^2+\gamma^2-(u+\delta)^2}{2u\gamma} = \frac{\gamma^2-2u\delta-\delta^2}{2u\gamma}$. As Q approaches P, the γ^2 and δ^2 terms become negligible, and this cosine equates to $-\frac{\delta}{\gamma}$. This is the same as $\cos \angle F_2 QP$, so the rays $F_1 P$ and $F_2 P$ describe equal angles with the normal to the curve. This establishes the reflector property.

8. From the focus-directrix property, the equation for the conic is $r = \varepsilon(d - x) = \varepsilon(d - r\cos\theta)$. Rearranging this equation gives us the formula for r.

9. Let the variable circle have centre P and radius r, and let Γ_i have radius R_i and centre O_i. We have $PO_1 = r + R_1$ and $PO_2 = R_2 - r$. Adding the equations gives $PO_1 + PO_2 = R_1 + R_2$, which is constant. Hence, the locus of centres is an ellipse with foci O_1 and O_2.

10. $r = PO_1 - R_1$ is a linear function of the distance to the focus, which is a linear function of the distance to the directrix, which is a linear function of the distance to the radical axis (which is parallel to the directrix). As the radius of the variable circle tends to zero as it approaches the radical axis, this linear function must have a constant term of 0. Hence, the radius of the circle is proportional to the distance between its centre and the radical axis.

11. Let the three circles be C_1, C_2 and C_3, such that (without loss of generality) C_2 and C_3 are tangent to both h_1 and h_2, and C_1 and C_2 are tangent to both γ_1 and γ_2. Let a fourth circle C_4 be tangent to γ_2 and γ_3, and let its centre lie on the line BC_1. Let d_i denote the perpendicular distance between the centre of C_i and the line AC, and let r_i denote its radius. We want to show that $\frac{r_4}{r_1} = \frac{d_4}{d_1}$, as this means that the circles are homothetic with centre of homothety B, so C_4 is tangent to both h_2 and h_3. Using this idea of homothety, we have $\frac{r_3}{r_2} = \frac{d_3}{d_2}$. Similarly, the previous exercise gives us $\frac{r_4}{r_3} = \frac{d_4}{d_3}$ and $\frac{r_2}{r_1} = \frac{d_2}{d_1}$. We can multiply these three equations

to prove that $\frac{r_4}{r_1} = \frac{d_4}{d_1}$, so C_4 is indeed inscribed in the remaining curvilinear quadrilateral. For further discussion of the problem, see http://www.imo-official.org/problems/IMO2010SL.pdf. This includes a three-dimensional interpretation of the problem where cones of equal gradient are erected on the circles.

12. Let d be the perpendicular distance from the locus of motion to O. Then $[O\,P\,Q] = \frac{1}{2}\,v\,d\,t$, so $\frac{dA}{dt} = \frac{1}{2}\,v\,d$ is constant.

13. If P continues in a straight line at its present velocity, let the new position after time t be denoted Q. If P instead is accelerated towards or away from O, then its new position is denoted Q'. As the acceleration is in the direction of O, and t is very small (technically, the limit as $t \to 0$), $O\,P$ is parallel to $Q\,Q'$. So, $[O\,Q\,P] = [O\,Q'\,P]$ and thus $\frac{dA}{dt}$ is unaffected by the acceleration. So, it must remain constant, as in the previous scenario.

14. We begin with the polar form of a conic, $r = \frac{l}{1+\varepsilon\cos\theta}$. We then rearrange to obtain $\frac{l}{r} = 1 + \varepsilon\cos\theta$, and differentiate both sides with respect to time. This gives us $-\frac{l}{r^2}\,\frac{dr}{dt} = -\varepsilon\sin\theta\,\frac{d\theta}{dt}$. Multiplying both sides by $-\frac{r^2}{l}$ gives us $\frac{dr}{dt} = \frac{k\,\varepsilon}{l}\sin\theta$. Proceeding to differentiate again results in $\frac{d^2r}{dt^2} = \frac{k\,\varepsilon}{l}\cos\theta\,\frac{d\theta}{dt} = k\left(\frac{1}{r} - \frac{1}{l}\right)\frac{d\theta}{dt}$, where the last stage involves substituting the equation of the conic back into the equation. Acceleration is then $a = k\left(\frac{1}{r} - \frac{1}{l}\right)\frac{d\theta}{dt} - r\left(\frac{d\theta}{dt}\right)^2 = k\left(\frac{1}{r} - \frac{1}{l}\right)\frac{k}{r^2} - \frac{k^2}{r^3} = -\frac{k^2}{r^2\,l}$.

15. Project I and J to the circular points at infinity, so O is the centre of the circle Γ. $P\,Q\,P'\,Q'$ are concyclic, thus $P\,Q\,P'\,Q'\,I\,J$ are conconic.

16. Again, project I and J to the circular points at infinity, so O is the centre of the circle Γ. Then, this statement equates to $P\,P'\,R\,S$ being a cyclic quadrilateral, which is obvious from the fact that angles are preserved in circle inversion.

Glossary

When writing the book, I have assumed you are familiar with terminology such as 'orthocentre' and 'geometric mean'. As this may not necessarily be the case, some common terms are explained here.

- **abscissa:** the x-coordinate of a point on the plane. Compare with *ordinate*.
- **altitude:** a line from a vertex of a triangle, which is perpendicular to the opposite side. The three altitudes intersect at the *orthocentre*.
- **AM-GM inequality:** for n non-negative real numbers, the *arithmetic mean* is greater than or equal to the *geometric mean*, with equality if and only if all variables are equal.
- **Apollonius' theorem:** in a triangle ABC, where M is the centre of BC, we have $AM^2 = \frac{1}{2}b^2 + \frac{1}{2}c^2 - \frac{1}{4}a^2$.
- **areal coordinates:** a system of projective homogeneous coordinates where each point is considered to be the weighted barycentre of three variable masses, each of which is positioned at a vertex of a fixed 'reference triangle'.
- **Argand plane:** the idea of representing the real and imaginary parts of a complex number as the Cartesian coordinates of a point on the Euclidean plane.
- **arithmetic mean:** for n variables $\{x_1, \ldots, x_n\}$, the arithmetic mean is $\frac{1}{n}(x_1 + \ldots + x_n)$.
- **barycentre:** the centre of mass of a set of masses positioned at points (on the plane).
- **barycentric coordinates:** a synonym of *areal coordinates*.
- **Bezout's theorem:** two algebraic curves of degrees m and n intersect in precisely mn points on the complex projective plane, when counted with the appropriate multiplicity.
- **Brahmagupta's formula:** for a cyclic quadrilateral with side lengths a, b, c, d and semiperimeter s, the area is given by $\sqrt{(s-a)(s-b)(s-c)(s-d)}$. This is a generalisation of *Heron's formula*.
- **Brianchon's theorem:** if a hexagon $ABCDEF$ is circumscribed about a circle (or, more generally, a *conic*), its three major diagonals (AD, BE and CF) are concurrent.
- **Cardano's formula:** the general solution to a cubic equation.
- **Catalan sequence:** a sequence of integers that counts the number of valid strings of $2n$ parentheses.
- **Cauchy-Schwarz inequality:** for two vectors u and v, $u \cdot v \leq |u|\ |v|$, with equality if and only if u and v have the same direction.
- **Cayley-Bacharach theorem:** if two cubics intersect in nine points and a third cubic passes through eight of those points, then it also passes through the ninth.
- **Cayley-Menger determinant:** a formula for the square of the volume of a simplex in terms of the squares of the side lengths.
- **centroid:** the intersection of the three *medians* of a triangle. More generally, it is synonymous with *barycentre*.
- **Ceva's theorem:** if D, E and F are points on the (possibly extended) sides BC, CA and AB, respectively, then AD, BE and CF are concurrent if and only if $\frac{\overrightarrow{BD}}{\overrightarrow{DC}} \cdot \frac{\overrightarrow{CE}}{\overrightarrow{EA}} \cdot \frac{\overrightarrow{AF}}{\overrightarrow{FB}} = 1$.
- **circular points at infinity:** a pair of points on the complex projective plane through which all circles pass.
- **circumcentre:** the point O equidistant from the three vertices of a triangle.
- **circumradius:** the radius R of the circumscribing circle of a triangle or cyclic polygon.
- **collinear:** points lying on the same straight line.
- **concentric:** objects sharing the same centre. This is usually applied to circles, but is equally applicable to conics.
- **conconic:** points lying on the same conic section.

- **concurrent:** three (or more) lines are said to be concurrent if they all intersect at a single point or are all mutually parallel.
- **concyclic:** points lying on the same circle.
- **conic:** a curve in the plane described by a quadratic equation in Cartesian coordinates.
- **coplanar:** points (or curves) lying on the same flat plane.
- **coprime:** two integers with a greatest common divisor of 1.
- **cosine rule:** for a generic triangle $A\,B\,C$, $a^2 = b^2 + c^2 - 2\,b\,c\cos A$.
- **cross-ratio:** for four collinear points, the ratio $\frac{\overrightarrow{AB}\cdot\overrightarrow{CD}}{\overrightarrow{BC}\cdot\overrightarrow{DA}}$. If the cross-ratio is -1, the points form a *harmonic range*.
- **cube roots of unity:** the three roots of the polynomial $z^3 - 1$. We often use ω to represent the 'north-west' complex cube root of unity $\frac{\sqrt{3}}{2}\,i - \frac{1}{2}$.
- **Desargues' theorem:** two triangles are in perspective about a point if and only if they are in perspective about a line.
- **difference of two squares:** the polynomial $a^2 - b^2 = (a - b)(a + b)$.
- **difference of three cubes:** the polynomial $a^3 + b^3 + c^3 - 3\,a\,b\,c = (a + b + c)\left(a + b\,\omega + c\,\omega^2\right)\left(a + b\,\omega^2 + c\,\omega\right)$, where ω is a primitive *cube root of unity*.
- **Euclid's algorithm:** the greatest common divisor of a and b can be obtained by subtracting the smaller from the larger and repeating until one of the numbers is zero. For example, $(26, 10) \to (16, 10) \to (6, 10) \to (6, 4) \to (2, 4) \to (2, 2) \to (2, 0)$, so the greatest common divisor of 26 and 10 is 2.
- **Euler-Apollonius lollipop:** the disc on diameter $G\,H$, which contains the *incentre*, *symmedian point* and *Gergonne point*.
- **Euler-Fermat theorem:** if a and n are coprime, then $a^{\varphi(n)} \equiv 1 \pmod{n}$, where $\varphi(n)$ is the number of positive integers $\leq n$ which are coprime to n.
- **Euler line:** the circumcentre, centroid, nine-point centre and orthocentre are collinear in the ratio $O\,G : G\,T : T\,H = 2 : 1 : 3$.
- **Euler's inequality:** $O\,I^2 = R^2 - 2\,R\,r$, where the *circumcircle* has centre O and radius R, and the *incircle* has centre I and radius r.
- **excentre:** the centre of an *excircle*.
- **excircle:** one of three circles (other than the *incircle*) tangent to (the extensions of) the three sides of a triangle.
- **Feuerbach's theorem:** the *nine-point circle* is tangent to the *incircle* and three *excircles*.
- **Fibonacci sequence:** the sequence defined with $F_0 = 0$, $F_1 = 1$ and $F_{n+2} = F_{n+1} + F_n$. If you extrapolate it backwards, you obtain the 'nega-Fibonacci numbers'.
- **fundamental theorem of algebra:** a degree-d polynomial can be factorised into d linear factors over the complex numbers.
- **fundamental theorem of arithmetic:** every positive integer has a unique prime factorisation.
- **geometric mean:** for n variables $\{x_1, \ldots, x_n\}$, the geometric mean is $\sqrt[n]{x_1\,x_2\ldots x_n}$.
- **Gergonne point:** the intersection of the lines joining each vertex of a triangle to the point of tangency of the incircle with the opposite side.
- **glide-reflection:** the composition of a reflection in a line and a translation parallel to the line.
- **harmonic mean:** for n variables $\{x_1, \ldots, x_n\}$, the harmonic mean is $\frac{n}{\frac{1}{x_1}+\frac{1}{x_2}+\ldots+\frac{1}{x_n}}$.
- **harmonic quadrilateral:** a cyclic quadrilateral where the products of opposite side lengths are equal.

- **harmonic range:** a set of collinear points with a cross-ratio of -1.
- **Heron's formula:** if a triangle has side lengths a, b, c and semiperimeter s, the area is given by $\sqrt{s(s-a)(s-b)(s-c)}$. It is a special case of *Brahmagupta's formula*.
- **heterochromatic:** differently-coloured.
- **homothety:** a synonym of enlargement, homothecy, scaling, dilation or dilatation.
- **incentre:** the centre of the *incircle* of a triangle (or, more generally, inscribable polygon).
- **incircle:** the circle tangent to the three sides of a triangle and contained within it.
- **inradius:** the radius r of the *incircle*.
- **intersecting chords theorem:** if there is a point P in the plane of a circle Γ, and a line l passing through P and meeting Γ at A and B, then the value of $PA \cdot PB$ is independent of l and equal to the *power of the point P*.
- **median:** a straight line joining a vertex of a triangle to the midpoint of its opposite side.
- **Menelaus' theorem:** if D, E and F are points on the (possibly extended) sides BC, CA and AB, respectively, then D, E and F are collinear if and only if $\frac{\overrightarrow{BD}}{\overrightarrow{DC}} \cdot \frac{\overrightarrow{CE}}{\overrightarrow{EA}} \cdot \frac{\overrightarrow{AF}}{\overrightarrow{FB}} = -1$.
- **monic polynomial:** a polynomial of degree n where the coefficient of x^n is 1. Every polynomial is a scalar multiple of a monic polynomial.
- **monochromatic:** everything is the same colour.
- **Nagel point:** the intersection of the lines joining each vertex of a triangle to the point of tangency of the opposite excircle with its corresponding side.
- **nine-point circle:** the circle passing through the midpoints of the sides, the feet of the *altitudes* and the midpoints of AH, BH and CH, where H is the *orthocentre*.
- ***n*th roots of unity:** the n roots of the complex polynomial $z^n - 1$. If it cannot be expressed as a mth root of unity for some $m < n$, then it is known as 'primitive'. The *monic polynomial* whose roots are the $\varphi(n)$ primitive nth roots of unity is known as a 'cyclotomic polynomial'.
- **ordinate:** the y-coordinate of a point on the plane. Compare with abscissa.
- **orthocentre:** the intersection point H of the three *altitudes* of a triangle.
- **Pappus' theorem:** the special case of *Pascal's theorem* when the conic is a pair of straight lines.
- **parallelepiped:** a three-dimensional version of a parallelogram, obtained by applying a generic affine transformation to a cube. The n-dimensional generalisation is called a parallelotope.
- **Pascal's theorem:** if a hexagon is inscribed in a circle (or, more generally, a *conic*), the three pairwise intersections of opposite sides are collinear.
- **power of a point:** for a point P in the plane of a circle with centre O and radius R, the value of $OP^2 - R^2$ is known as its 'power'.
- **projective plane:** an extension of the Euclidean plane where parallel lines are considered to meet on a line at infinity.
- **Ptolemy's inequality:** if A, B, C and D are four points in space, the inequality $AB \cdot CD + BC \cdot DA \geq AC \cdot BD$ holds, with equality if and only if $ABCD$ is a (non-self-intersecting) cyclic quadrilateral.
- **Pythagoras' theorem:** for a right-angled triangle ABC, where $C = \frac{\pi}{2}$, the identity $a^2 + b^2 = c^2$ applies. It is a special case of the *cosine rule*.
- **quadratic mean (RMS):** for n variables $\{x_1, \ldots, x_n\}$, the quadratic mean is $\sqrt{\frac{1}{n}\left(x_1{}^2 + \ldots + x_n{}^2\right)}$.
- **radical axis:** the locus of points of equal power with respect to two circles Γ_1 and Γ_2. This is necessarily a straight line.
- **semiperimeter:** half of the perimeter of a polygon.

- **semiprime:** the product of two distinct primes, *e.g.* $23 \times 89 = 2047$.
- **sine rule:** For every triangle ABC, $\frac{a}{\sin A} = \frac{b}{\sin B} = \frac{c}{\sin C} = 2R$, where R is the *circumradius*.
- **Stewart's theorem:** If D is a point on the line BC, then $BD \cdot DC \cdot BC + AD^2 \cdot BC = AC^2 \cdot BD + AB^2 \cdot DC$.
- **symmedian:** the reflection of a *median* of a triangle in the corresponding interior angle bisector.
- **symmedian point:** the intersection of the three *symmedians* of a triangle. It has unnormalised areal coordinates $\left(a^2, b^2, c^2\right)$.
- **triangle inequality:** each side of a triangle is smaller than the sum of the other two sides. In terms of vectors, this is $|\underline{a} + \underline{b}| \leq |\underline{a}| + |\underline{b}|$.

Further reading

If you enjoyed this book, which I hope you did, then you may find these other books, papers and websites of interest.

Books

***The IMO Compendium*, by Djukić, Janković, Matić and Petrović:** This book is divided into a list of useful theorems, a massive repository of shortlisted IMO problems, and solutions to the aforementioned problems. If you're training for international competitions, this will give you surplus experience.

***A Mathematical Olympiad Primer*, by Geoff Smith:** If you have no past experience of mathematical olympiads, this is the place to begin.

***Plane Euclidean Geometry*, by Bradley and Gardiner:** This is a rigorous exploration of Euclidean geometry, including more basic techniques such as angle chasing and similar triangles in addition to vectors, Cartesian coordinates and trigonometry.

***The Elements*, by Euclid:** This constructive approach to geometry begins with five postulates from which everything else is proved. It is possible to define 'non-Euclidean geometries', such as spherical and hyperbolic geometry, by rejecting a subset of these postulates.

***Introduction to Number Theory and Inequalities*, by Christopher Bradley:** Affectionately known as 'INTI', this book covers these topics in extensive detail. At the expense of losing a catchy acronym, it has since been separated into two disjoint books, unsurprisingly called *'Introduction to Inequalities'* and *'Introduction to Number Theory'*.

***The Symmetries of Things*, by Conway, Burgiel and Goodman-Strauss:** In addition to the familiar Platonic solids and regular tilings, there is a cornucopia of objects with fascinating symmetry. This book features a systematic exploration of different symmetry groups in two, three and four dimensions.

***An Introduction to Diophantine Equations*, by Titu Andreescu and Dorin Andrica:** The title of the book is rather self-explanatory.

***Complex Numbers from A to... Z*, also by Andreescu and Andrica:** This features an exploration of the Argand plane, including a substantial amount of triangle geometry.

***The Algebra of Geometry*, by Christopher Bradley:** Coordinate methods, such as Cartesian, areal and projective coordinates, can be employed to solve geometry problems with varying degrees of success. If you want to learn more about them, this book is ideal.

Online resources

Complex Projective 4-Space (http://cp4space.wordpress.com): My website, which is updated periodically with mathematical miscellany.

Wolfram MathWorld (http://mathworld.wolfram.com): This is an online encyclopedia containing definitions and information about practically every mathematical concept discovered.

***Areal Co-ordinate Methods in Euclidean Geometry*, by Tom Lovering** (http://www.bmoc.maths.org/home/areals.pdf): This is more succinct than Bradley's book, but definitely worth reading.

***Curves in cages: an algebro-geometric zoo*, by Gabriel Katz** (http://arxiv.org/abs/math/0508076): A special case of the generalised Cayley-Bacharach occurs when two of the curves are unions of lines. This leads to some interesting generalisations of Pascal's theorem, complete with proofs.

***Online Encyclopedia of Integer Sequences*, by N. J. A. Sloane** (http://oeis.org/): A continually expanding collection of over 200 000 sequences of integers. Practically every conceivable integer sequence (well, 200 000 out of $2^{\aleph_0}$) is featured somewhere within this vast repository.

***Encyclopedia of Triangle Centres*, by Clark Kimberling** (http://faculty.evansville.edu/ck6/encyclopedia/ETC.html): If you close this book and gaze at the front cover, you will see a configuration of triangle centres. In reality, there are uncountably infinitely many possible triangle centres, over five thousand of which are featured on this website. Moreover, there is a method of searching for triangle centres based on *trilinear coordinates*.

Virtual Geoff Smith (http://people.bath.ac.uk/masgcs/): This features some papers by Geoff and Bradley, including the combinatorics of snail venom. Exciting!

***UK IMO Register*, by Joseph Myers** (http://imo-register.org.uk/): A hall of fame of everyone who has represented this sceptred isle in any of the five main mathematical olympiads (RMM, IMO, BalkMO, EGMO and CGMO). Also, you can browse reports of competitions, such as Richard Freeland's excellent and witty report of RMM 2011.

***The UVW method*, by Tejs** (http://ohkawa.cc.it-hiroshima.ac.jp/AoPS.pdf/The%20 uvw %20 method.pdf): The first paper describing the use of the *u v w* method of solving trivariate symmetric polynomial inequalities.

***What is... a Dimer*, by Richard Kenyon and Andrei Okounkov** (http://www.ams.org/notices/200503/what-is.pdf): We touched upon bipartite matchings and domino tilings in the first chapter, but this is a more detailed analysis.

Discussion forums

Ask nRich (http://nrich.maths.org/discus/messages/board-topics.html): A friendly atmosphere where everyone is nice and polite towards each other, helping to solve problems in an idyllically co-operative way.

MathLinks (http://mathlinks.ro): A fierce battleground where unwitting bypassers are sadistically subjected to supposedly 'trivial' problems. The atmosphere differs greatly from *nRich*, being described as 'the difference between carnivores and herbivores'.

Acknowledgements

At the risk of resembling a speech at the Oscars, I would like to acknowledge the generations of mathematicians responsible for discovering the theorems that appear here, ranging from Thales (twenty-six centuries ago) to my colleagues at the recent IMO training camps. In particular, the commitment of all those who delivered fascinating and informative lectures during the aforementioned camps should be credited here. Many other publications by the United Kingdom Mathematics Trust have influenced this work, written mainly by Christopher Bradley and Geoff Smith. Thanks also go to Louise Beach for advocating the use of the duck-egg blue colour scheme for highlighting theorems due to its soothing effect. Finally, I would especially like to thank everyone who shaped the book with their praise, comments and corrections, including my parents, Joseph Myers, James Aaronson, James Pogmore, Sam Cappleman-Lynes, Márton Mester and the legendary Suzanna Eames.