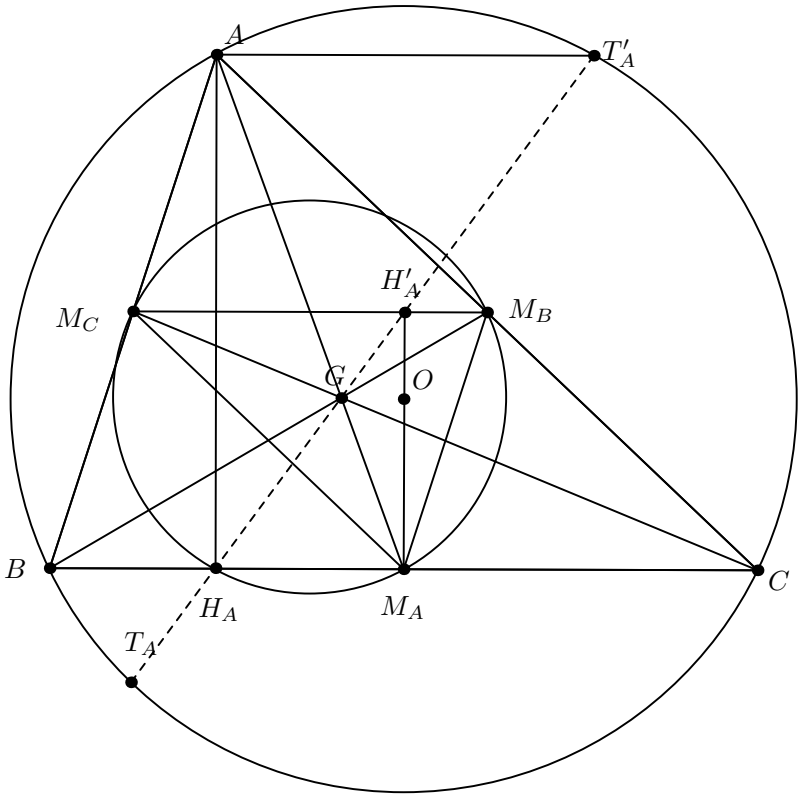
Basic ideas for tackling problems in mathematical olympiads

Krish Nigam *

February, 2021



The nine point circle - points H_A, M_A, M_B and M_C are concylcic (four of the nine points).

*Email: kn221@ic.ac.uk

Contents

Introduction

Throughout this informal guide, I intend to provide the reader with the sufficient fundamental theory required for olympiad mathematics and other related competitions, alongside motivating worked examples.

I believe that, with enough determination, anyone who has not been exposed to much of competition mathematics can become comfortable with the type of problems and do well, and that is my objective with this guide.

Most mathematical olympiad problems can be solved without the use of advanced theorems.

That is the beauty of mathematical olympiads, and quite often we can explain a solution to someone with very elementary mathematical knowledge. Having said that, it is undoubtedly handy to know some theory, and this way we might find ourselves able to write down solutions more easily and the case of actively learning theorems and proofs gives us a more confident understanding of where mathematical ideas come from and how they connect together.

Theory

In this section, I supply the reader with the most fundamental tools necessary, as far as 'knowledge' is concerned, in order to tackle problems in mathematical olympiads and competitions. The theory in this book, however, is not complete to the extent of even some national level competitions, and instead is meant as a starting point to more elementary competitions. Consequently, there may be certain other results which have not been covered here.

In light of all this, this book is aimed at those with very little background with mathematical olympiads who intend to learn some of the basic grounding material. This is most suited to the UK education system and in particular the *UKMT* competitions *BMO1* and *BMO2*, though I have done my best to keep the theory as broad and versatile as possible, so as to be applicable to competitions from all countries.

Learning all the theory that follows will put the reader in good stead, however it should be noted that all olympiad questions also require a significant level of ingenuity which will come after a great deal of practice.

Proofs and derivations for theorems are not always provided so I would encourage the reader to look these up if they are struggling with a theorem or are unfamiliar to a theorem. I would also encourage the reader to have a go at the example problems themselves first before viewing the solution.

The most relevant topics to olympiad mathematics can be broken down into four categories: algebra, combinatorics, (Euclidean) geometry and number theory.[1]

[1]Some like to exclude functional equations as its own area, however I have included this in algebra.

0.1 Techniques for Proof

Almost all olympiad problems will involve proving a statement or finding a solution and proving that it is the only possible one that holds under certain conditions, for instance. There are many methods of proof, and occasionally we would simply just have to follow the logical mathematical steps which should be sufficient to prove something directly - *proof by deduction*. This is especially useful with inequalities.

Or, we may have to be slightly more careful and consider separately all the relevant cases - *proof by exhaustion*. For example, investigating what happens to some type of number when divided by some positive integer.

Those are both fairly standard and logical ways of approaching a proof. Below, however, I cover the two slightly more 'fancy' ways of proving something and both are extremely crucial to have in our arsenal.

0.1.1 Proof by Contradiction

Proof by contradiction is a clever and elegant way of proving a statement, particularly useful in number theory. Essentially, we assume that the statement we are trying to prove is false and show that this leads to some mathematical contradiction, which in turn implies that the original statement must have been true. I demonstrate a common example.

Example Prove that $\sqrt{2}$ is irrational.

Solution Suppose the contrary - that $\sqrt{2}$ is rational. If it were rational, it can be expressed as a fraction $\frac{p}{q}$, where p and q are *coprime* (*i.e.* p and q share no common factor so $\frac{p}{q}$ is simplified to lowest terms). Since p and q are coprime, *both of them cannot be even.*

$$\sqrt{2} = \frac{p}{q} \implies 2q^2 = p^2$$

Since the left hand side is even, the right hand side must also be even. Therefore, p^2 must be even, which implies that p must be even. But, if p is even we can express it as $p = 2k$ where k is some integer. So we have $p^2 = (2k)^2 = 4k^2$.

$$\therefore\ 2q^2 = 4k^2 \implies q^2 = 2k^2$$

From the above, we see that q must be even as the right hand side is even. But that's a contradiction! It was stated that both p and q cannot be even. Assuming that $\sqrt{2}$ is rational led to a contradiction, meaning that $\sqrt{2}$ cannot be rational. Hence we've proved that $\sqrt{2}$ must be irrational. $\square$

0.1.2 Proof by Contrapositive

We can use a *proof by contrapositive* when proving a statement of the form 'if a then b'. This is logically equivalent to proving the *contrapositive*: 'if not b then not a'.

Example Prove that if n^2 is a multiple of 3 then n must be a multiple of 3.

Solution We will try and prove the statement by considering the contrapositive: when n is not a multiple of 3. Therefore, n can take two possible forms: $n = 3k + 1$ or $n = 3k + 2$ for some integer k.

First, let's consider $n = 3k + 1$

$$n^2 = (3k + 1)^2 = 9k^2 + 6k + 1 = 3(3k^2 + 2k) + 1$$

But $3(3k^2 + 2k) + 1$ is not divisible by 3, so n^2 is not divisible by 3. Let's next consider $n = 3k + 2$

$$n^2 = (3k + 2)^2 = 9k^2 + 12k + 4 = 3(3k^2 + 4k + 1) + 1$$

But $3(3k^2 + 4k + 1) + 1$ is not divisible by 3, so n^2 is not divisible by 3.

We have shown that if n is not a multiple of 3 then n^2 is not a multiple of 3 (*i.e.* the contrapositive statement is true). So, if n^2 is a multiple of 3, n must also be a multiple of 3. $\square$

0.1.3 Proof by Induction

Arguably, *proof by induction* is the most useful form of proof when it comes to proving facts in the olympiad world. It is especially useful when we spot a pattern and we 'guess' a formula, and usually the best way to show this formula holds for all values is to proceed with induction.

A *proof by induction* is structured by first showing the statement is true for some initial value - this is known as the *base case*. Then we assume the statement is true for some value of k - this is known as the *inductive hypothesis*. After this, we use our assumption to show that the statement is true for $k+1$ - this is known as the *inductive step*. And this would conclude the proof.

For the intuition behind why such a proof works, let's suppose we want to prove a statement $P(n)$ is true for all positive integers n. Then, we would show that $P(1)$ is true, and after assume $P(k)$ is true for $k \geq 1$ and show that this implies that $P(k+1)$ is also true. If $P(1)$ is true then $P(2)$ is also true since $P(k)$ true implies $P(k+1)$ true. But if $P(2)$ is true then $P(3)$ is also true and if $P(3)$ is true so is $P(4)$ and so on, so the original statement $P(n)$ is true for all positive integers.

To help the reader understand the concept a little more, I will present an analogy here. Suppose that each integer is a domino (all placed in order). We know that if a domino falls, the next one falls too. The first domino (representing 1) falls, causing 2 to fall, then 3 to fall, then 4 to fall, etc. As you can see, it seems obvious that all dominos must eventually fall, even though there are infinitely many of them.

Example Show that, for every positive integer n, the number $3^{3n+4} + 7^{2n+1}$ is a multiple of 11.

Solution Let $f(n) = 3^{3n+4} + 7^{2n+1}$. *Base case*: $n = 1$, $3^{3(1)+4} + 7^{2(1)+1} = 343 + 2187 = 2530 = 11 \times 230$. So, $f(1)$ is a multiple of 11. *Inductive hypothesis*: assume $f(k) = 3^{3k+4} + 7^{2k+1}$ is a multiple of 11 for some positive integer k. *Inductive step*: now consider $f(k+1)$.

$$\begin{aligned} f(k+1) &= 3^{3(k+1)+4} + 7^{2(k+1)+1} \\ &= 3^{3k+7} + 7^{2k+3} \\ &= 27\left(3^{3k+4}\right) + 49\left(7^{2k+1}\right) \\ &= 27\left(3^{3k+4} + 7^{2k+1}\right) + 22\left(7^{2k+1}\right) \\ &= 27f(k) + 22\left(7^{2k+1}\right) \end{aligned}$$

Since $f(k)$ is a multiple of 11 by our hypothesis and $22(7^{2k+1})$ is also a multiple of 11, $f(k+1)$ is a multiple of 11. Hence, by the principle of

mathematical induction, $f(n)$ is a multiple of 11 for all positive integers. $\square$

Example Prove $3^n < n!$ for every positive integer n greater than 6. *$n!$ is n factorial which means $n \times (n-1) \times (n-2) \times ... \times 2 \times 1$ e.g. $3! = 3 \times 2 \times 1 = 6$.*

Solution Let $P(n)$ be $3^n < n!$. *Base case*: $n = 7$, $3^7 = 2187 < 5040 = 7!$. So, $P(7)$ is true. *Inductive hypothesis*: let $P(k)$ be true, so we assume $3^k < k!$ where $k > 6$. *Inductive step*: now consider $P(k+1)$

$$3^{k+1} = 3 \cdot 3^k < 3 \cdot k! \text{ (by our assumption)}$$

Since $k > 6 \implies k + 1 > 7$, so $k + 1 > 3$.

$$3^{k+1} < 3 \cdot k! < (k+1) \cdot k! = (k+1)!$$
$$\therefore\ 3^{k+1} < (k+1)!$$

By the principle of mathematical induction, $P(n)$ is true for all positive integers n greater than 6. $\square$

Strong Induction

Strong induction is a slight extension to 'normal' induction. In strong induction, the only difference is that instead of assuming a statement holds true for some **one** value of k to prove it is true for $k + 1$, we assume that the statement holds true for all values $1, 2, 3, ..., k$ to prove it is true for $k + 1$.

Example Show that every positive integer n can be written as a sum of distinct powers of two.

Solution *Base case*: for $n = 1$ note that $2^0 = 1$, hence our proposition holds for $n = 1$. *Inductive hypothesis*: assume that our proposition holds for every m in $1 \leq m \leq k$, so every positive integer m in the interval can be written as a sum of distinct powers of two. *Inductive step*: now, let's consider what happens with $k + 1$. $k + 1$ can either be even or odd, so we have two separate cases to consider.

If $k + 1$ is even, then observe that $\frac{k+1}{2}$ must be an integer. Now as $1 \leq \frac{k+1}{2} \leq k$ we know by our inductive hypothesis that $\frac{k+1}{2}$ can be written as a

sum of distinct powers of 2. But then multiplying $\frac{k+1}{2}$ by 2 gives

$$\frac{k+1}{2} \cdot 2 = k+1$$

and since each distinct power of 2 in the sum of $\frac{k+1}{2}$ is multiplied by a factor of 2, each power of 2 is increased by 1 and thus remains distinct.

If $k+1$ is instead odd, then we know k is even. Furthermore, by our inductive hypothesis, we know that k can be written as a sum of distinct powers of 2. But if k is even, k does not contain a $2^0 = 1$ in its sum of distinct powers of 2. To show this, note that k, by definition of an even integer, can be expressed as $k = 2m$ for some positive integer m. And since we can view multiplication as repeated addition we have

$$k = 2m = \underbrace{2+2+2...+2}_{m \text{ times}}$$

If the sum were to contain a $2^0 = 1$ we clearly would need to express k as $2m+1$ which would make it odd not even. Hence, we see that

$$k+1 = k+2^0$$

and if $k+1$ is odd, it can be written as a sum of distinct powers of two. It follows by strong induction that for all positive integers n, we can express n as a sum of distinct powers of two. $\square$

This is actually a great fact because it is essentially saying that we can express all (base 10) numbers in binary.

0.1.4 If and only if

If you are ask to prove a statement of the form 'show that p is true if and only if q is true', then only showing that if q is true implies that p is true is not sufficient. We must also show the converse - that if p is true implies that q is true. In mathematics the notation $p \implies q$ means p implies q. The notation $p \iff q$ means p if and only if q; in other words p implies q **and** q implies p.

To say 'a only if b' means that a can only ever be true when b is true. That is, b is necessary for a to be true. And, to say 'a if and only if b' means that a is true if b is true, and b is true if a is true. That is, a is *necessary and sufficient* for b.

$a \implies b$	a *only if* b	b *if* a
$a \impliedby b$	a *if* b	b *only if* a
$a \iff b$	a *if* and *only if* b	

An example may be if we take $a =$ 'lemon', $b =$ 'yellow'. Then $a \implies b$, but it is not necessarily the case that $b \implies a$. So, this is not a necessary and sufficient condition (or an 'if and only if case').

0.2 Approaching a Problem

It is usually not best to dive straight into an olympiad problem blindly after reading it. Instead, it is much better to take the time to understand the problem very thoroughly (no ink is ever wasted) and to gain an overall intuition behind what is going on. Testing out simple cases of the problem help, or trying to spot or guess a pattern for the problem.

I would like to demonstrate this notion of diving straight into a problem with one of my favourite, yet simplistic, problems.

Example If $x^2 - 3x + 1 = 0$, what is the value of $x^2 + \frac{1}{x^2}$?

Solution It is very tempting, particularly in the pressured exam situation, to dive straight into this problem and find the value of x with the quadratic formula and substitute it into the expression. However, the value of x does not come out as a nice number at all, and especially with a non-calculator exam this way will prove to be rather laborious.

Here I present a much neater solution which simply requires a little bit of pre-thought. Note that $x \neq 0$, since $0^2 - 3(0) + 1 = 1 \neq 0$. So we can divide by x - that's the key step.

$$\begin{aligned} x - 3 + \frac{1}{x} &= 0 \\ x + \frac{1}{x} &= 3 \\ \left(x + \frac{1}{x}\right)^2 &= 9 \\ x^2 + 2 + \frac{1}{x^2} &= 9 \\ \therefore \boxed{x^2 + \frac{1}{x^2} = 7} & \end{aligned}$$

Chapter 1

Algebra

1.1 Polynomials

We define a polynomial $P_n(x)$ of degree n as:

$$P_n(x) = a_n x^n + a_{n-1}x^{n-1} + ... + a_1 x + a_0 \tag{1.1.1}$$

where $a_i \in \mathbb{R}$ and $a_n \neq 0$.

We should note that polynomials of odd degree have an infinite range.

1.1.1 Remainder Theorem

This is a topic usually covered very early on in A-Level in the UK, and I think is important to gain a better intuition behind polynomials.

Given some polynomial $P(x)$, let us compute the value of $P(a)$ where a is any real number. Let the value of $P(a) = r$. The theorem states that if our polynomial $P(x)$ was divided by $(x - a)$ then the remainder after division is r. Of course if r is 0, then $(x - a)$ must be a factor of $P(x)$ - this is most useful.

1.1.2 Factor Theorem

Given some polynomial $P(x)$, if $P(a) = 0$ for some $a \in \mathbb{R}$, then $(x - a)$ divides $P(x)$.

Example Prove the useful lemma that for polynomial $P(x)$ with integer coefficients, and any two integers a, b,

$$a - b \mid P(a) - P(b)$$

Solution By the factor theorem, if $(a-b)$ is a factor of $P(a)-P(b)$ then if we make the substitution $a = b$, $P(a) - P(b)$ should be 0. In fact, we get $P(b) - P(b) = 0$, hence $(a - b)$ is a factor of $P(a) - P(b)$. Now, since a, b are integers and our polynomial $P(x)$ has integer coefficients, then $(a - b)$ completely divides $P(a) - P(b)$. $\square$

Another, perhaps more careful, way we can do this problem is by expressing P as some general polynomial so $P(x) = a_n x^n + a_{n-1}x^{n-1} + ... + a_1 x + a_0$ where $a_n, ..., a_1, a_0$ are integer coefficients. So,

$$\begin{aligned} P(a) - P(b) &= \left(a_n \cdot a^n + a_{n-1} \cdot a^{n-1} + ... + a_1 \cdot a + a_0\right) - \\ &\qquad \left(a_n \cdot b^n + a_{n-1} \cdot b^{n-1} + ... + a_1 \cdot b + a_0\right) \\ &= a_n \left(a^n - b^n\right) + a_{n-1}\left(a^{n-1} - b^{n-1}\right) + ... + a_1 \left(a - b\right) \end{aligned}$$

Each of the terms in the right hand side above are divisible by $(a - b)$ (see (1.1.4)) hence $(a - b)$ divides $P(a) - P(b)$. $\square$

1.1.3 Binomial Theorem

This is an efficient way to expand brackets that have two terms to any integer power n. I think it is useful to be aware of this, particularly for speed, and it can be specifically useful in number theory.

$$\begin{aligned} (x+y)^n \equiv \binom{n}{0}x^n + \binom{n}{1}x^{n-1}y + \binom{n}{2}x^{n-2}y + ... \\ ... + \binom{n}{n-1}xy^{n-1} + \binom{n}{n}y^n \end{aligned} \tag{1.1.2}$$

where $\binom{n}{k} = {}^nC_k = \frac{n!}{k!(n-k)!}$ and represents a binomial coefficient. The binomial coefficient is explained more thoroughly in the combinatorics section. Below is a fun, yet tough, example, which doesn't require binomial theorem explicitly but only a good understanding of multiplying brackets.

Example What is the coefficient of x^9 in the expansion of

$$(1+x)(1+x^2)(1+x^3)...(1+x^{100})$$

Solution We have 100 factors, each which are the sum of 1 and a positive power of x. When the product is expanded, each term results from picking a 1 from some factors and a positive power of x from others. For example, if the positive powers of x you pick in one case are x^3, x^{21} and x^{35} then the term is $x^{3+21+35} = x^{59}$ - the ones don't matter since $1 \times x = x$.

For the coefficient of x^9 therefore, the problem reduces to how many different ways can we make the term x^9 from our expansion. Thus we consider, in how many ways can we write 9 as a sum of distinct positive integers.

$$\begin{array}{lll} 9 = 9 & 9 = 1 + 8 & 9 = 6 + 2 + 1 \\ & 9 = 2 + 7 & 9 = 5 + 3 + 1 \\ & 9 = 3 + 6 & 9 = 4 + 3 + 2 \\ & 9 = 4 + 5 & \end{array}$$

If we have the sum of more than three terms making 9, then not all of them will be distinct. So from the list above we have 8 different ways of making x^9. Hence, the coefficient of x^9 is $\boxed{8}$.

1.1.4 Useful Identities

Here are some very helpful identities; it is worth knowing *all* of these since they are indispensable to making life easier when it comes to the algebra bash.

The difference of two squares is the most basic and, arguably, most crucial identity to know.

$$x^2 - y^2 \equiv (x+y)(x-y) \tag{1.1.3}$$

In the next factorisation, we can also use a difference of two squares, however most of the time it is nicer to use the following.

$$x^n - y^n \equiv (x-y)(x^{n-1} + x^{n-2}y + ... + xy^{n-2} + y^{n-1}) \tag{1.1.4}$$

Substituting $-y$ for y gives another identity when n is odd.

$$x^n + y^n \equiv (x+y)(x^{n-1} - x^{n-2}y + ... - xy^{n-2} + y^{n-1}) \tag{1.1.5}$$

Often the next factorisation is very useful particularly in problems where we must find integer solutions.

$$xy + x + y + 1 \equiv (x + 1)(y + 1) \tag{1.1.6}$$
$$xy - x - y + 1 \equiv (x - 1)(y - 1) \tag{1.1.7}$$

In the one below, notice how all the + change to × and vice versa on each side.

$$xyz + (x + y)(y + z)(z + x) \equiv (x + y + z)(xy + yz + zx) \tag{1.1.8}$$

This one is less obvious, but useful nonetheless.

$$x^3 + y^3 + z^3 - 3xyz \equiv (x + y + z)(x^2 + y^2 + z^2 - xy - yz - zx) \tag{1.1.9}$$

The next is known as the *Sophie Germain Identity*, and it displays how the sums of two squares can be factorised in a rather neat way.

$$\begin{aligned} x^4 + 4y^4 &= x^4 + 4x^2y^2 + 4y^4 - 4x^2y^2 \\ &= (x^2 + 2y^2)^2 - (2xy)^2 \\ &= (x^2 + 2xy + 2y^2)(x^2 - 2xy + 2y^2) \end{aligned} \tag{1.1.10}$$

Another great fact is that the set $\mathcal{S}$, which consists of all integers that can be expressed as the sum of two squares, is closed[1] under multiplication. This is due to the following identity.

$$(a^2 + b^2)(c^2 + d^2) \equiv (ac + bd)^2 + (ad - bc)^2 \tag{1.1.11}$$

Series

It is worth knowing the following common standard results:

$$\sum_{i=1}^{n} i = 1 + 2 + 3 + \cdots + n = \frac{n(n+1)}{2} \tag{1.1.12}$$
$$\sum_{i=1}^{n} i^2 = 1^2 + 2^2 + 3^2 + \cdots + n^2 = \frac{n(n+1)(2n+1)}{6} \tag{1.1.13}$$
$$\sum_{i=1}^{n} i^3 = 1^3 + 2^3 + 3^3 + \cdots + n^3 = \frac{n^2(n+1)^2}{4} \tag{1.1.14}$$

Notice the neat idea that $\sum i^3 = (\sum i)^2$. We can also have *arithmetic* and *geometric* series:

[1] This just means that when you multiply two numbers that can be expressed as the sum of two perfect squares, you get another number which is also the sum of two perfect squares.

- In an arithmetic series each term differs by a constant difference d, giving the sequence
  $a, a+d, a+2d, \ldots, a+(n-1)d$. The sum of these terms is $S_n = \frac{a}{2}(2a+(n-1)d)$.
- In a geometric series each term differs by a constant multiple r, giving the sequence
  $a, ar, ar^2, \ldots, ar^{n-1}$. The sum of these terms is $S_n = \frac{a(1-r^n)}{1-r}$. Sometimes, with geometric sequences, we can compute the sum of infinitely many terms should this *converge* (*e.g.* $1 + \frac{1}{2} + \frac{1}{4} + \ldots$). This is given by $S_\infty = \frac{a}{1-r}$.

1.1.5 Vieta's Formulas

Let's start with the simple case of a cubic. Consider $P(x) = Ax^3 + Bx^2 + Cx + D$. Let the roots of the cubic be α, β, γ. Note this means $P(\alpha) = P(\beta) = P(\gamma) = 0$. So we can now say,

$$\begin{aligned} Ax^3 + Bx^2 + Cx + D &= A(x-\alpha)(x-\beta)(x-\gamma) \\ &= Ax^3 - A(\alpha+\beta+\gamma)x^2 + A(\alpha\beta+\beta\gamma+\gamma\alpha)x - A\alpha\beta\gamma \end{aligned}$$

Comparing coefficients, we can obtain the following relationships:

$$\alpha + \beta + \gamma = \frac{-B}{A} \tag{1.1.15}$$

$$\alpha\beta + \beta\gamma + \gamma\alpha = \frac{C}{A} \tag{1.1.16}$$

$$\alpha\beta\gamma = \frac{-D}{A} \tag{1.1.17}$$

We can extrapolate this information and generalise the pattern for any polynomial of degree n.

Consider $P_n(x) = a_n x^n + a_{n-1}x^{n-1} + ... + a_1 x + a_0$. Let $\alpha_1, \alpha_2, ..., \alpha_n$ be roots of $P_n(x)$. Then:

$$\begin{aligned} \sum \alpha_i &= -\frac{a_{n-1}}{a_n} \\ \sum \alpha_i\alpha_j &= \frac{a_{n-2}}{a_n} \\ &\vdots \end{aligned}$$

$$\alpha_1\alpha_2 \times ... \times \alpha_n = (-1)^n \frac{a_0}{a_n} \tag{1.1.18}$$

Notice how we have the sum of roots, then the pair-wise sum of roots, then the triplet-wise sum of roots, and so on until we get the product of roots. It is important to note that Vieta's formulas apply to *all* roots, whether real or complex. So, when a problem asks for real roots, we may be unable to apply Vieta's formulas directly.

Example Given the equation $x^2 - 5x + 9 = 0$ has two solutions α and β, find $\alpha^2 + \beta^2$.

Solution By Vieta's formulas, $\alpha + \beta = -\frac{-5}{1} = 5$ and $\alpha\beta = \frac{9}{1} = 9$. So we have

$$\alpha^2 + \beta^2 = (\alpha + \beta)^2 - 2\alpha\beta = 5^2 - 2 \times 9 = 25 - 18 = \boxed{7}$$

1.2 Inequalities

Inequalities are perhaps the most significant section of any olympiad algebra. Here, we cover the most fundamental ones required.[2]

1.2.1 Core Inequalities

Sums of squares

One should be aware the squares are always *non-negative* (for real numbers) - *i.e.* greater than or equal to 0. The following derives from the fact that $(x - y)^2 \geq 0$.

$$x^2 + y^2 \geq 2xy \tag{1.2.1}$$

Example Prove that $x^2 + y^2 + z^2 \geq xy + yz + zx$

Solution

$$\begin{aligned}(x-y)^2 \geq 0 &\implies x^2 + y^2 \geq 2xy \\ (y-z)^2 \geq 0 &\implies y^2 + z^2 \geq 2yz \\ (z-x)^2 \geq 0 &\implies z^2 + x^2 \geq 2zx\end{aligned}$$

$$\begin{aligned}2x^2 + 2y^2 + 2z^2 &\geq 2xy + 2yz + 2zx \\ \implies x^2 + y^2 + z^2 &\geq xy + yz + zx \quad \square\end{aligned}$$

[2]For the interested reader, others worth adding to your arsenal (though more applicable for IMO) include *Muirhead's* and *Schur's inequalities*, as well as *Jensen's* which is easier to grasp than the other two.

AM-GM

The *arithmetic mean - geometric mean inequality*, also known as AM-GM, states that the arithmetic mean is greater than or equal to the geometric mean of a set of non-negative real numbers.

For non-negative real numbers $x_1, x_2, ..., x_n$

$$\frac{x_1 + x_2 + ... + x_n}{n} \geq \sqrt[n]{x_1 x_2 ... x_n} \tag{1.2.2}$$

Often, the most trivial case is used and can be shown from (1.2.1).

$$\frac{x+y}{2} \geq \sqrt{xy}$$

Example For positive real numbers a, b, c where $a + b + c = 6$, prove that

$$ab^2c^3 \leq 108$$

Solution Write b as $\frac{b}{2} + \frac{b}{2}$ and c as $\frac{c}{3} + \frac{c}{3} + \frac{c}{3}$. By AM-GM

$$\frac{a + \frac{b}{2} + \frac{b}{2} + \frac{c}{3} + \frac{c}{3} + \frac{c}{3}}{6} \geq \left(a \cdot \frac{b}{2} \cdot \frac{b}{2} \cdot \frac{c}{3} \cdot \frac{c}{3} \cdot \frac{c}{3}\right)^{\frac{1}{6}}$$

$$\frac{a+b+c}{6} \geq \left(\frac{ab^2c^3}{4 \times 27}\right)^{\frac{1}{6}}$$

$$1 \geq \left(\frac{ab^2c^3}{108}\right)^{\frac{1}{6}}$$

$$\therefore \ ab^2c^3 \leq 108 \ \square$$

Rearrangement

Suppose we have a permutation of the set $\{a_1, a_2, a_3, a_4, a_5\}$ (that is, a_1, a_2, a_3, a_4, a_5 in any order) and a permutation of the set $\{b_1, b_2, b_3, b_4, b_5\}$. What are the maximum and minimum possible values of

$$a_1b_1 + a_2b_2 + ...a_5b_5 \ ?$$

This leads to the *rearrangement inequality*. If our sets are sorted with $a_1 \geq a_2 \geq ... \geq a_5$ and $b_1 \geq b_2 \geq ... \geq b_5$, then the maximum value we can attain is

$$a_1b_1 + a_2b_2 + a_3b_3 + a_4b_4 + a_5b_5 \tag{1.2.3}$$

and the minimum value we can attain is

$$a_1b_5 + a_2b_4 + a_3b_3 + a_4b_2 + a_5b_1 \tag{1.2.4}$$

Example a, b and c are positive real numbers. Show that

$$a^3 + b^3 + c^3 \geq a^2b + b^2c + c^2a$$

Solution By writing $a^3 + b^3 + c^3$ as $a \times a^2 + b \times b^2 + c \times c^2$, we can see that both sides are of the form $a_1b_1 + a_2b_2 + a_3b_3$, just ordered differently.

a, b and c are positive so the lists a, b, c and a^2, b^2, c^2 are ordered in the same way.

Let $\{a_1, a_2, a_3\}$ be a permutation of $\{a^2, b^2, c^2\}$ and $\{b_1, b_2, b_3\}$ be a permutation of $\{a, b, c\}$. By the rearrangement inequality, $a_1b_1 + a_2b_2 + a_3b_3$ takes its largest value when a_1, a_2, a_3 is ordered in the same way as b_1, b_2, b_3. So this largest value is bigger than any other pairing of the two sets. So:

$$a^3+b^3+c^3 = a^2 \times a+b^2 \times b+c^2 \times c \geq a^2 \times b+b^2 \times c+c^2 \times a = a^2b+b^2c+c^2a$$

□

1.2.2 Further Inequalities

Cauchy-Schwarz
For those that are comfortable with vectors, you will be aware for two vectors $\mathbf{a}$ and $\mathbf{b}$, $\mathbf{a} \cdot \mathbf{b} = |\mathbf{a}|\,|\mathbf{b}| \cos\theta \leq |\mathbf{a}|\,|\mathbf{b}|$. This is essentially what the *Cauchy-Schwarz inequality* states.[3]

The *Cauchy-Schwarz inequality* is extremely versatile in general mathematics and can be used in vectors, integration and matrices. However, the following form is most common for olympiad algebra.

For real numbers $a_1, a_2, ..., a_n$ and $b_1, b_2, ..., b_n$

$$(a_1^2 + a_2^2 + ... + a_n^2)(b_1^2 + b_2^2 + ... + b_n^2) \geq (a_1b_1 + a_2b_2 + ... + a_nb_n)^2 \quad (1.2.5)$$

which is equivalent to

$$\left(\sum_{i=1}^{n} a_i^2\right)\left(\sum_{i=1}^{n} b_i^2\right) \geq \left(\sum_{i=1}^{n} a_ib_i\right)^2 \quad (1.2.6)$$

[3]On a sidenote, I love this inequality so much that I wrote a mini handout for it which can be found on `kn7811.com/cauchy_schwarz.pdf` .

Equality holds if the sequences are proportional. That is, if $\frac{a_1}{b_1} = \frac{a_2}{b_2} = ... = \frac{a_n}{b_n}$. There is a very useful form of this inequality known as the *Engel Form* where we make the substitution $a_k = \frac{a_k^2}{\sqrt{b_k}}$ and $b_k = \sqrt{b_k}$ into (1.2.5).

$$\frac{a_1^2}{b_1} + \frac{a_2^2}{b_2} + ... + \frac{a_n^2}{b_n} \geq \frac{(a_1 + a_2 + ... + a_n)^2}{b_1 + b_2 + ... + b_n} \tag{1.2.7}$$

Weighted AM-GM
The *weighted AM-GM inequality* is an extension to *AM-GM* , and it is not completely obvious either.

If $a_1, a_2, ..., a_n$ are non-negative real numbers, and $\lambda_1, \lambda_2, ..., \lambda_n$ are non-negative real numbers (the 'weights') which sum to w, then it follows:

$$\frac{\lambda_1 a_1 + \lambda_2 a_2 + ... + \lambda_n a_n}{w} \geq \sqrt[w]{a_1^{\lambda_1} a_2^{\lambda_2} ... a_n^{\lambda_n}} \tag{1.2.8}$$

The common case of $w = 1$ is usually used.

Example Let a, b be positive real numbers such that $a + b = 1$. Show that

$$a^a b^b + a^b b^a \leq 1$$

Solution Let the 'weights' be a, b, so by weighted AM-GM we have

$$\frac{a \cdot a + b \cdot b}{a + b} \geq \left(a^a b^b\right)^{\frac{1}{a+b}}$$

But since $a + b = 1 \implies a^2 + b^2 \geq a^a b^b$. We can do a similar thing with weights b, a and by weighted AM-GM we have

$$\frac{b \cdot a + a \cdot b}{a + b} \geq \left(a^b b^a\right)^{\frac{1}{b+a}}$$

$\implies 2ab \geq a^b b^a$. Adding the two final inequalities

$$\begin{aligned} a^2 + b^2 + 2ab &\geq a^a b^b + a^b b^a \\ (a + b)^2 &\geq a^a b^b + a^b b^a \\ \therefore\ a^a b^b + a^b b^a &\leq 1 \quad \square \end{aligned}$$

1.3 Extraneous Solutions

In algebra, we can sometimes introduce erroneous solutions from a step, so it is worth stressing that we must check our solutions into the original equation at the end. This is most common when we do the step of 'squaring both sides', however we may also find this problem arising when working with the modulus function or piecewise functions such as floor functions.

Example Solve $\sqrt{x+3} = x - 3$.

Solution

$$\begin{aligned} \sqrt{x+3} &= x-3 \\ x+3 &= (x-3)^2 \\ x^2 - 7x + 6 &= 0 \\ (x-1)(x-6) &= 0 \end{aligned}$$

So $x = 1$ or $x = 6$. But $x = 1$ is not a valid solution. So we reject that and the only valid solution is $x = 6$. This is because the square root function only outputs non-negative numbers so we must satisfy $x - 3 \geq 0 \implies x \geq 3$. Below is a slightly more challenging and relevant example.

Example Solve $\sqrt{10 + \sqrt{x^3 + 100}} = 10 - \sqrt{x^3 + 100}$

Solution Let's make the substitution $a = \sqrt{x^3 + 100}$

$$\begin{aligned} \sqrt{10+a} &= 10 - a \\ 10 + a &= 100 - 20a + a^2 \\ a^2 - 21a + 90 &= 0 \\ (a-6)(a-15) &= 0 \end{aligned}$$

So, $a = 6$ or $a = 15 \implies \sqrt{x^3 + 100} = 6, 15$.

$$\begin{aligned} x^3 + 100 &= 36,225 \\ x^3 &= -64,125 \end{aligned}$$

So, $x = -4$ or $x = 5$. But, we can't conclude here! We **must** check these solutions actually satisfy the original equation. For $x = -4$, we have

$$\begin{aligned}\sqrt{10 + \sqrt{(-4)^3 + 100}} &\stackrel{?}{=} 10 - \sqrt{(-4)^3 + 100} \\ \sqrt{10 + 6} &\stackrel{?}{=} 10 - 6 \\ \sqrt{16} &\stackrel{?}{=} 4\end{aligned}$$

which is in fact true so $x = -4$ is a solution. Now let's check $x = 5$

$$\begin{aligned}\sqrt{10 + \sqrt{(5)^3 + 100}} &\stackrel{?}{=} 10 - \sqrt{(5)^3 + 100} \\ \sqrt{10 + 15} &\stackrel{?}{=} 10 - 15 \\ \sqrt{25} &\stackrel{?}{=} -5\end{aligned}$$

which is not true so we reject $x = 5$, which is an extraneous root. Hence, the only possible solution is $\boxed{x = -4}$.

I hope that this demonstrates that we must be very careful when presented with such problems.

1.4 Functional Equations

Outside of olympiads, it is not extremely common to find many problems involving functional equations. Without some techniques, these sort of problems can seem very challenging. Generally, we may be able to find functions which satisfy the original conditions, however we have to be rigorous in our argument and we would normally have to find **all** such functions so we must make sure that we exhaust all cases. A golden rule of functional equations:

Choose substitutions to make as much cancel as possible - but not everything.

I will demonstrate some tricks and techniques we can use to solve functional equations, but it is, as usual, a case of logically piecing together the appropriate tools and using some ingenuity.

1.4.1 Substitutions

Here are some helpful substitutions we can make:

1. Substitute 0 or other constants which can help make parts of the equation constant.
2. Substitute $-x$ in place of x. This can lead us to determine if there is any nice symmetry in the function.
3. If we have two variables at play, substitute $y = x$.
4. If we have functions within functions then it is helpful to remove the function inside. For instance, in the case $f(x - f(x))$ setting $x = f(x)$ would be a good idea.
5. It can sometimes be handy to multiply an equation by a constant or x and then subtract equations *e.g.* $f(x) - xf(x)$ which could take us a step further to finding f.

Example Find all functions $f : \mathbb{R} \to \mathbb{R}$ which satisfy $f(x)f(y) = f(x + y) + xy$ for all real numbers x, y.

Solution We will consider the equation $f(x)f(y) = f(x + y) + xy$, and proceed with various substitutions.

Let $x = 0 \implies f(0)f(y) = f(y)$, so there are two cases: either $f(0) = f(y) = 0$, or $f(0) = 1$.

If $f(0) = f(y) = 0$, then $f(x) \cdot 0 = f(x + y) + xy \implies f(x + y) = -xy$. It makes sense to let $y = -x$ here, which gives $f(0) = x^2$ so $x = 0$. But, because y and x are *symmetrical*[4], this gives us the trivial solution that $x = y = 0$ which is unhelpful.

Now, let's take the case $f(0) = 1$. To fully utilise this, we can let $x = 1$ and $y = -1$, because then our $f(0)$ can cancel, and we get a 0 on the RHS which is very useful.

$$f(1)f(-1) = f(0) - 1 \implies f(1)f(-1) = 0$$

Therefore, either $f(1) = 0$ and/or $f(-1) = 0$.

In the case $f(1) = 0$, letting $y = 1$ gives $f(x)f(1) = f(x + 1) + x \implies f(x + 1) = -x$. Now, make the transformation $x \to x - 1$ and we obtain $\boxed{f(x) = 1 - x}$.

[4]This means that if we replace x with y and y with x, the original equation remains unchanged.

In the other case $f(-1) = 0$, letting $y = -1$ gives $f(x)f(-1) = f(x-1) - x \implies f(x-1) = x$. Similarly, letting $x \to x+1$ gives the final solution $\boxed{f(x) = x+1}$.

1.4.2 Other Advice

1. Always try to take an educated guess of what the function could be, so as to have an idea of where we're aiming, and this can also be very helpful for finding good substitutions.
2. For functions defined on the natural numbers, induction is a good way forward.
3. For functions which are restricted to polynomials, considering the degree of the polynomial is always necessary! So, degree of LHS = degree of RHS.

Example Find all polynomials $P(x)$ with real coefficients satisfying $P(P(P(x))) - 3P(x) = -2x$ for all x.

Solution Let's give our polynomial $P(x)$ degree n. Thus, we must have the degree of LHS = degree of RHS. The degree of $P(P(P(x)))$ will be $n \times n \times n = n^3$. So, the degree of the LHS would be n^3 as $n^3 \geq n$ for all integers $n \geq 0$.

However, we must also take into account the case for when $3P(x) = P(P(P(x)))$, which would give a degree of 0. This would be when $P(x) = 3^{\frac{1}{3}}x$, but we can reject this case as the degree of the RHS is 1, and not 0.

So, for degree of LHS = degree of RHS, we have $n^3 = 1 \implies n = 1$, as $n \geq 0$. This means we can express $P(x)$ as $ax + b$. From here on, it is just an algebra bash.

$$\begin{aligned} P(P(P(x))) &= P(P(ax+b)) \\ &= P(a(ax+b)+b) \\ &= a\left[a(ax+b)+b\right]+b \end{aligned}$$

$$\begin{aligned} &\therefore\ a\left[a(ax+b)+b\right]+b-3ax-3b = -2x \\ &\implies \left(a^3-3a\right)x + (a^2b+ab-2b) = -2x \end{aligned}$$

Now, we can *compare coefficients* to determine a and b.

$$a^3 - 3a = -2 \implies (a-1)^2(a+2) = 0 \implies a = 1 \text{ or } a = -2$$
$$a^2b + ab - 2b = 0 \implies b = \text{ anything when } a = 1 \text{ or } a = -2$$

So, $P(x) = \boxed{x + k}$ or $P(x) = \boxed{-2x + k}$ for any $k \in \mathbb{R}$.

Chapter 2

Combinatorics

In a sense, it is usually the combinatorics problems which require the most thought and problem solving, since they do not rely on recalling many theorems or ideas at all. However, I think that the following ideas are most useful and provide assistance when formally writing out a solution to a combinatorics problem.

2.1 Pigeonhole Principle

The *pigeonhole principle* (PHP), also known as the *Dirichlet principle*, is a very obvious and intuitive fact yet it is extremely powerful and can be used to provide explanations for not so trivial facts, as shown in the examples.

Naive Case If more than n pigeons are placed into n pigeonholes, then at least one pigeonhole must contain more than one pigeon.

General Case If more than kn objects are placed into n boxes, then at least one box must contain more than k objects.

Example Prove that at a party with at least two people, there are two people who know the same number of people.

Solution Let n be the number of people at the party. The maximum number of people a person can know is $n-1$ (know all of them) and the minimum they can know is 0 (know no one). The number of people someone can know therefore is some number in the set $\{0, 1, 2, ..., n-1\}$.

However, the key insight is that there cannot be someone who knows everyone and someone who knows no one. So while there are n people they can all either know some number of people from the set $\{0, 1, 2, ..., n-2\}$ or from the set $\{1, 2, 3, ..., n-1\}$. Both these sets have $n-1$ values and there are n people. Let the pigeons be the people and the pigeonholes be the possible number of people each person can know. There are n pigeons but $n-1$ pigeonholes, so by the pigeonhole principle at least one pigeonhole contains more than one pigeon *i.e.* there at least two people who know the same number of people at the party. □

Example There are 5 distinct points randomly placed on the surface of a sphere. Prove that at least 4 of the points lie on the same hemisphere (inclusive hemisphere).

Solution Pick any two distinct points on the surface of the sphere. These two points define a great circle (look this up if you do not know what it looks like). Now, there are two hemispheres with this great circle as boundary, and each of the other three points lie in either hemisphere. By the pigeonhole principle, at least two of those three points lie in the same hemisphere, and that hemisphere thus contains at least four of the five given points. □

2.2 Counting

It is extremely common for problems in combinatorics that we have to count very carefully and logically. Generally, these problems can be easy to trip up on and so we must be very meticulous in counting the cases to avoid double counting or missing a case, especially when numbers are big, and to be organised by using things like factorials or choose notation where necessary. Although, be aware that much of the time merely simple logic and thought is required.

2.2.1 Binomial Coefficients

Binomial coefficients are a family of positive integers that occur as coefficients in the binomial theorem. Binomial coefficients are represented as

$$\binom{n}{r} = {}^nC_r = \frac{n!}{r!(n-r)!} \tag{2.2.1}$$

where $k! = k \times (k-1) \times (k-2) \times ... \times 2 \times 1$. Binomial coefficients form each number in Pascal's triangle. We are mostly interested in binomial coefficients as they allow us to quickly compute the number of ways we can choose r objects from n objects, where the order we choose the objects in does not matter. For example if we have 16 balls and we want to choose 3 of them, then the number of different ways of choosing 3 balls is ${}^{16}C_3 = 560$.

Example If we have 5 red balls and 7 blue balls, what is the probability of picking exactly 2 red and 2 blue balls.

Solution There are ${}^5C_2$ ways of picking 2 red balls and there are ${}^7C_2$ ways of picking 2 blue balls. There are also ${}^{12}C_4$ of picking any 4 balls in general. So the desired probability is

$$\frac{{}^5C_2 \times {}^7C_2}{{}^{12}C_4} = \boxed{\frac{14}{33}}$$

Sometimes, we may have to choose objects where the order **does** matter - these are known as *permutations* rather than *combinations* previously. For example, if we have n objects and we want to pick r objects from them such that the order matters (*i.e.* $\{1,2,3\}$ is not the same as $\{2,1,3\}$), then the number of ways to do this is

$${}^nP_r = \frac{n!}{(n-r)!} \tag{2.2.2}$$

Example How many four letter passwords can be formed if the characters allowed to use without repetition are $0, 1, 2, 3, ..., 9$ and $A, B, C, ..., Z$?

Solution We have $10 + 26 = 36$ choices of characters to choose from. So we have to arrange 4 objects out of 36 available objects. The number of ways of doing this is equal to

$${}^{36}P_4 = \frac{36!}{(36-4)!} = \boxed{1413720}$$

We may also have to **arrange** objects where the order does matter, instead of having to simply choose them. If we have n different objects then there are $n!$ ways of arranging them. If not all the objects are different then we must also account for that. Suppose instead there are n objects with 3

of them which are the same and rest different, then there are $\frac{n!}{3!}$ ways of arranging them. In general, with n total objects and a identical objects, b different identical objects and c different identical objects, there are $\frac{n!}{a!\times b!\times c!}$ different ways of arranging our n objects.

Example How many ways are there of arranging the letters in the word 'mathematics'?

Solution Notice that there are 11 letters, with letters m, a and t appearing twice. So the number of ways to arrange the letters is

$$\frac{11!}{2!\times 2!\times 2!} = \boxed{4989600}$$

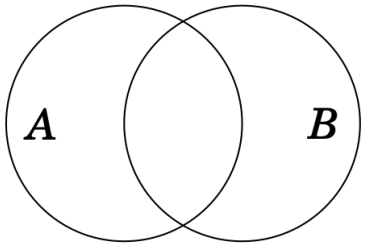
It is also worth bearing in mind how binomial coefficients can lead to sequences of other common numbers, in particular *triangular numbers*. The nth triangular number $T(n)$ is defined by

$$T(n) = \binom{n+1}{2} = \frac{n(n+1)}{2} \tag{2.2.3}$$

Triangular numbers are a common sequence, especially in counting arguments, and should be recognised immediately.

2.2.2 Principle of Inclusion and Exclusion

The *principle of inclusion and exclusion* (PIE) is a counting technique which computes the total number of elements in different sets (which may overlap). This way, PIE guarantees that all elements are counted exactly once and prevents double counting. It is easiest to consider the case of two sets - a venn diagram can be used for visualisation.



Set A contains some elements and set B contains some elements. But, there are some elements which appear in both set A and set B. So, in

order to count the total number of elements in either A or B, we have the following formula

$$|A \cup B| = |A| + |B| - |A \cap B| \tag{2.2.4}$$

Note that for some set S, $|S|$ denotes the number of elements in S (also known as the *cardinality* of S). Here, the formula is telling us that the total number of elements in both sets A and B is the total number of elements in A plus the total number of elements in B minus the total number of elements in both A and B - notice how this prevents double counting.

PIE gives us a similar, but slightly more involved, formula for counting the total number of elements in three sets A, B and C.

$$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C| \tag{2.2.5}$$

The pattern continues for greater number of sets, with alternating plus/minus signs, but we are unlikely to be dealing with more then three sets. At first, it can be difficult to see directly where this is applicable so I include an example which seems like a simple question but is easy to overcount or undercount elements, so PIE works quite nicely.

Example Of the numbers $1, 2, 3, ..., 6000$, how many are not multiples of 2, 3 or 5?

Solution Effectively, this question is asking us how many numbers from $1, ..., 6000$ **are** multiples of 2, 3 or 5, and since we know that there are 6000 numbers in total, we can subtract these to find how many numbers are not multiples of 2, 3 or 5 - this makes life slightly easier for us.

Let X be the set of numbers in $1, ..., 6000$ which are divisible by 2; Y be the set of numbers which are divisible by 3; Z be the set of numbers which are divisible by 5. Notice how some numbers are multiples of both 2 and 3 or all of 2, 3 and 5 for example, so we use PIE to avoid double counting numbers. We want the total number of elements which are multiples of 2, 3 or 5. So we want $|X \cup Y \cup Z|$.

$$|X \cup Y \cup Z| = |X| + |Y| + |Z| - |X \cap Y| - |X \cap Z| - |Y \cap Z| + |X \cap Y \cap Z|$$

One should know that if a number is a multiple of both a and b then it is a multiple of ab. So our equation tells us the total number of elements which

are multiples of 2, 3 or 5 is the number of multiples of 2 plus the number of multiples of 3 plus the number of multiples of 5 minus the number of multiples of 6 minus the number of multiples of 10 minus the number of multiples of 15 plus the number of multiples of 30. This gives

$$|X \cup Y \cup Z| = 3000 + 2000 + 1200 - 1000 - 600 - 400 + 200 = 4400$$

So there are $6000 - 4400 = \boxed{1600}$ numbers which are not divisible by 2, 3 or 5.

2.3 Recurrence Relations

Recurrence relations are used to reduce complex problems into an iterative process, based on simpler versions of the problem. A common example case where we may want to use a recurrence relation is to define the Fibonacci sequence.[1] If we let F_n be the nth Fibonacci number, then we have $F_0 = F_1 = 1$ and $F_{n+1} = F_n + F_{n-1}$. That is, each term is the sum of the two previous terms. So, $F_0 = 1, F_1 = 1, F_2 = 2, F_3 = 3, F_4 = 5, F_5 = 8$, and so on.

2.3.1 Solving Recurrence Relations

Let's say we have a simple recursion such as $a_0 = 1$ and $a_n = 2 \cdot a_{n-1}$. This has an explicit formula, more commonly known as a *closed form*, which is $a_n = 2^n$. More generally, if we had $a_0 = 1$ and $a_n = k \cdot a_{n-1}$, then the solution is $a_n = k^n$.

As it turns out, when we have a recurrence relation like this, which relates previous terms in a *linear* way, then the solution will always be geometric (*i.e.* a sum of some constants each raised to the power of n).

In general, for a *linear recurrence* of the form

$$x_n = c_1 x_{n-1} + c_2 x_{n-2} + \cdots + c_k x_{n-k} \tag{2.3.1}$$

we can plug in our geometric form $x_n = ar^n$, which gives

$$ar^n = ac_1 r^{n-1} + ac_2 r^{n-2} + \cdots + ac_k r^{n-k} \implies \underline{r^k = c_1 r^{k-1} - c_2 r^{k-2} - \cdots - c_k} \tag{2.3.2}$$

[1] It is not completely straightforward to find an explicit nth term expression for Fibonacci numbers.

This is known as the *characteristic polynomial*. We can find the roots of the characteristic polynomial, which is of degree k, in order to find the closed form which is of the form

$$x_n = a_1 (r_1)^n + a_2 (r_2)^n + \cdots + a_k (r_k)^n \tag{2.3.3}$$

where $r_1, \ldots, r_k$ are the roots of the characteristic polynomial and $a_1, \ldots, a_k$ are constants which can be found by plugging numbers in.

Example Find a closed form for the nth Fibonacci number F_n.

Solution We should be aware that Fibonacci numbers are defined by the following recurrence, beginning with $F_0 = 0$, $F_1 = 1$.

$$F_{n+1} = F_n + F_{n-1}$$

Since our recurrence is linear, our solution will be geometric and we get

$$ar^{n+1} = ar^n + ar^{n-1} \implies ar^{n-1}\left(r^2 - r - 1\right) = 0$$

This gives us the characteristic polynomial $r^2 - r - 1 = 0$. Now, solutions to this are $r = \frac{1 \pm \sqrt{5}}{2}$. So our overall closed form will involve a sum of these two roots.

$$\therefore \; x_n = A \cdot \left(\frac{1+\sqrt{5}}{2}\right)^n + B \cdot \left(\frac{1-\sqrt{5}}{2}\right)^n$$

To find A and B, let's simply use the values of $F_0 = 0$ and $F_1 = 1$. This gives the following simultaneous equations.

$$\begin{aligned} A + B &= 0 \\ \left(\frac{1+\sqrt{5}}{2}\right) A + \left(\frac{1-\sqrt{5}}{2}\right) B &= 1 \end{aligned}$$

This gives $A = \frac{1}{\sqrt{5}}$ and $B = -\frac{1}{\sqrt{5}}$. Hence, $\boxed{F_n = \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^n - \frac{1}{\sqrt{5}}\left(\frac{1-\sqrt{5}}{2}\right)^n}$

Aside from using the characteristic polynomial, there is a neat trick we can also employ if we have a recurrence in the following form:

$$x_n = x_{n-1} + f(n) \tag{2.3.4}$$

If you rewrite the recurrence as $x_n - x_{n+1} = f(n)$, then summing up all the equations ranging from 1 to n will give us $a_n - a_0 = \sum_{k=1}^{n} f(k)$ as all the

terms in the middle are subtracted. If we can determine $\sum_{k=1}^{n} f(k)$, then this idea is very helpful. So, we obtain closed form:

$$a_n = a_0 + \sum_{k=1}^{n} f(k) \tag{2.3.5}$$

2.3.2 Catalan Numbers

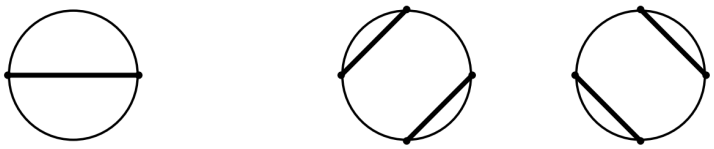
Catalan numbers are a particular family of numbers, which crop up in several problems that lead to the same recurrence relation. The nth Catalan number $C(n)$ is defined by

$$C_0 = 1 \quad \text{and} \quad C_{n+1} = \sum_{i=0}^{n} C_i \cdot C_{n-1} \tag{2.3.6}$$

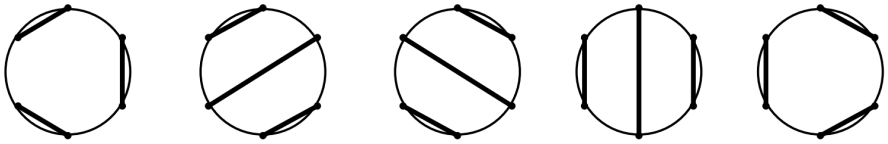
This makes more sense after seeing them in application, but for example, we can take C_4,

$$C_4 = C_0C_3 + C_1C_2 + C_2C_1 + C_3C_0 = 1 \times 5 + 1 \times 2 + 2 \times 1 + 5 \times 1 = 14$$

For example, consider $2n$ people sat in a circle. Then, the number of ways that n pairs of people engage in handshakes so that no arms cross is given by C_n. The diagrams below illustrates this. For $n = 1$, there is only $C_1 = 1$ way (left) and for $n = 2$, there are $C_2 = 2$ ways (right).



For $n = 3$, there are $C_3 = 5$ ways.



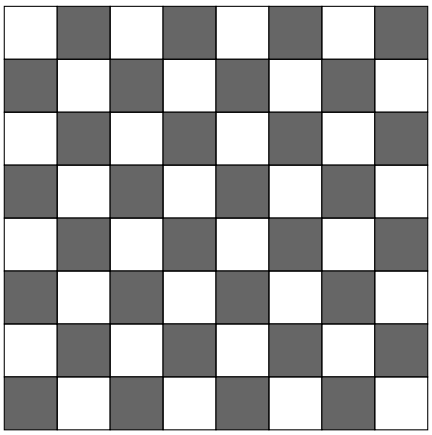
2.4 Tiling and Colouring

Tiling and *colouring* are techniques used often when we are dealing with problems involving some $a \times b$ board (*e.g.* a chessboard). We would typically colour the squares on our board and exploit these colours to prove some statements.

The clearest way to demonstrate colouring is through the following classic example.

Example If we remove two diagonally opposite corners from an 8×8 chess board, can the remaining 62 cells be exactly tiled with dominoes? *[Note: a domino takes up* 2×1 *or* 1×2 *space.]*

Solution A standard 8×8 chessboard comes with alternating coloured cells, and we will make use of this idea.



It should be clear that no matter where we place a domino, it will always cover one grey cell and one white cell - since the domino must take up the space of two adjacent squares.

However, if we remove two diagonally opposite corners from a chessboard, we remove two cells which are the same colour. As a result, we get 32 cells of one colour and 30 of the other colour, so there is no way of placing dominoes to exactly cover our chessboard. □

The same argument shows that if **any** two small squares of the same colour are removed, then the remaining board cannot be tiled with dominoes.

2.4.1 Alternative Colourings

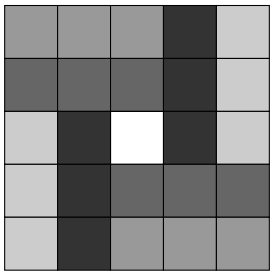
Although it is useful to colour a board alternating between black and white squares, we may sometimes need to employ a different colouring strategy depending on the problem.

To demonstrate an interesting case, let's consider a 5×5 board. Let's also

say that we have 8 straight trominoes (tiles which take up 3×1 or 1×3) space. Clearly, the 8 trominoes can only fill up a maximum of 24 of the 25 squares, so one square will always be left uncovered.

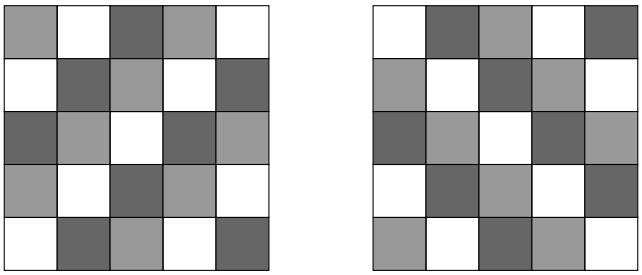
Suppose that we cover our 5×5 board with all 8 trominoes, then what are the possible candidates for the uncovered square?

If we experiment with this idea, it seems that the centre square is always being left uncovered. It's difficult to find any other cell. Here's a possible configuration:



In fact, it will always be the centre square which is uncovered; to prove this we can look for colourings which have useful properties with trominoes.

A tromino covers 3 squares, so it makes sense to use a colouring with three colours so that each tromino can cover one square of each colour.



As shown above, it is helpful to go for a diagonal pattern. I have gone for two similar boards with opposite-directional diagonal patterns, which will help us in our proof. From this pattern, we see that there are 9 white squares but only 8 grey and dark grey squares. So, there is an extra 1 white square.

Since the trominoes can only cover exactly one of each colour when placed on these boards, there will always be one white square left uncovered.

So, in the left board, the trominoes cover all the squares leaving one white one. And, in the right board they cover all the squares also leaving one white one. The only white sqare common to both boards is the central one, and hence it is always the central square which remains uncovered. □

This idea of using two similar board colourings (as opposed to just one) can be especially helpful in proving certain facts. Deciding on a colouring should be done carefully and logically too.

Chapter 3

Geometry

Olympiad geometry tends to require a great deal of imagination and ingenuity. A lot of the time, these questions involve spotting where we can apply a theorem and then apply the theorem and deduce further properties of the setup.

3.1 Constructing Diagrams

A **clear**, **big** diagram is the key to success in a geometry question. The diagram is our visualisation of the setup, so a clumped diagram is of no help at all. An accurate diagram allows us to make good guesses about certain properties of the setup, which helps enormously when it comes to solving geometry problems. Sometimes it is helpful to slightly offset the diagram a little bit to try and avoid assumptions. So, if we had to construct a triangle, it is best to avoid it making it look isosceles or right-angled.

I would advise doing the diagram in pencil (so to help add/rub out any points which we may find eases/distracts the problem), and I would thoroughly recommend using dashed lines, particularly in cases where we want to extend a line or arc, make reflections, or circumscribe objects.

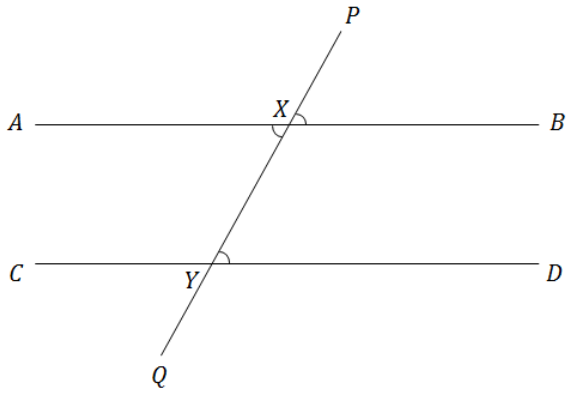
A *rule of thumb*: it is not usually necessary to draw the centre of the circle in most geometry problems; this is what we have circle theorems for. We may end up in circular arguments and find ourselves ending up proving one of the circle theorems, as they are all derived by using the properties of the center and radii.

3.2 Elementary Theorems

Here are some very helpful, basic theorems which the reader should initially become comfortable and fluent with employing in their solutions. Some things are assumed such as basic ideas of quadrilaterals, including the idea of parallelogram diagonals bisecting each other.

3.2.1 Lines

We will use the following diagram for reference.



Vertically Opposite Angles

Vertically opposite angles are angles which are equal and opposite to each other created when two **straight** lines intersect. In problems it is sometimes very helpful to use the converse. For instance, if $\angle PXB = \angle YXA$ then PXY and AXB are straight lines. Although such facts can seem trivial, if it is not explicitly stated in the problem we must mention in their proof that the three points lie on a straight line by the converse of vertically opposite angles to avoid being penalised.

Parallel Lines

Again, similar ideas apply for parallel lines. Parallel lines are two lines which do not intersect at any point. If AB is parallel to CD then $\angle YXA = \angle XYB$ (*alternate angles*) and $\angle BXY = \angle DYQ$ (*corresponding angles*). Also be prepared to use the converse; for instance if you are asked to prove two lines are parallel, it may be easier to consider the angles.

3.2.2 Polygons

We are mostly concerned with *regular*, *convex* polygons. These include pentagon, hexagon, heptagon, octagon etc. and we can generalise such a polygon with n sides as an n-gon. Of course, we should be able to see that as $n \to \infty$ our n-gon tends towards a circle. The key results for polygons that we need to be aware of are that the sum of the exterior angles is always $360°$, and that the size of each interior angle is given by

$$\frac{180(n-2)}{n}$$

where n is the number of sides. It is often also useful to rearrange this to find the number of sides.

Another idea worth remembering is the area of a hexagon

$$A = \frac{3\sqrt{3}}{2}s^2 \tag{3.2.1}$$

where s is the side length of the hexagon. This formula comes about from splitting the hexagon into six equilateral triangles and using '$\frac{1}{2}ab\sin\theta$'. We can generalise this for a regular polygon of n sides

$$A = \frac{n}{4\tan\left(\frac{180}{n}\right)}s^2 \tag{3.2.2}$$

3.2.3 Triangles

These three sided shapes are the building blocks of olympiad geometry problems. Note that three vertices of a triangle define a circle uniquely. Also, know the formulas for the area of a triangle - both '$\frac{1}{2}bh$' and '$\frac{1}{2}ab\sin\theta$'.

Right-Angled Triangles
The hypotenuse is the side opposite the right angle in a right angled triangle. We must know *Pythagoras' theorem* which states for a right angled triangle with hypotenuse of length z and other sides of length x, y

$$x^2 + y^2 = z^2 \tag{3.2.3}$$

It is worth having in mind the first few *Pythagorean triples* $(3, 4, 5)$; $(5, 12, 13)$; $(7, 24, 25)$. Of course, we can multiple each of these by a scalar to form new triples *e.g.* $(6, 8, 10)$. We should also be familiar with $\sin, \cos$ and $\tan$ as the ratios of opposite and hypotenuse, adjacent and hypotenuse and opposite and adjacent respectively.

Triangle Inequality
This idea is extremely obvious - essentially the sum of the lengths of any two sides in the triangle is larger than or equal to the length of the third side. The *triangle inequality* states in triangle ABC

$$\begin{aligned} AB + BC &\geq AC \\ BC + AC &\geq AB \\ AC + AB &\geq BC \end{aligned} \tag{3.2.4}$$

Equality is achieved only when the triangle is *degenerate*.

FULL Sine Rule
The **full** *sine rule* states that in a triangle with side lengths a, b and c and the angle opposite the sides are A, B and C respectively then

$$\frac{a}{\sin A} = \frac{b}{\sin B} = \frac{c}{\sin C} = 2R \tag{3.2.5}$$

where R is the radius of the *circumcircle* (that is, the circle through each vertex of the triangle). This fact which includes the link to the circumradius is not always taught with the sine rule, but when solving olympiad geometry problems it can be very useful to know.

Cosine Rule
The *cosine rule* states that in a triangle with side lengths a, b and c and the angle opposite the sides are A, B and C respectively then

$$\begin{aligned} c^2 &= a^2 + b^2 - 2ab\cos C \\ b^2 &= a^2 + c^2 - 2ac\cos B \\ a^2 &= b^2 + c^2 - 2bc\cos A \end{aligned} \tag{3.2.6}$$

We may also have to rearrange this to find the size of angles. Notice how we use cosine rule when we have two sides and the angle between them, and we want to find the length of the remaining side.

Example If the sides of a triangle have lengths 2, 3, and 4, what is the radius of the circle circumscribing the triangle?

Solution Let the angle opposite side length 2 be θ. Then by cosine rule

$$2^2 = 3^2 + 4^2 - 2 \cdot 3 \cdot 4\cos\theta \implies \cos\theta = \frac{3^2 + 4^2 - 2^2}{2 \cdot 3 \cdot 4} = \frac{7}{8}$$

So, $\sin\theta = \frac{\sqrt{15}}{8}$. By the full sine rule

$$R = \frac{1}{2}\frac{a}{\sin\theta} = \frac{1}{2}\cdot\frac{2}{\frac{\sqrt{15}}{8}} = \boxed{\frac{8}{\sqrt{15}}}$$

3.2.4 Similar and Congruent Triangles

This topic is so crucial that it's been made into its own subsection, rather than in the 'triangles' subsection. If you don't spot one of these then you make your life much harder when solving geometry problems.

Two triangles are *similar* if they have the same interior angles - so that one is a scaled version of the other. This means that the ratios of the respective sides of the triangles are constant and equal to each other. If we have a triangle with sides of length a, b and c then a similar triangle to that would be one with sides of length ka, kb and kc where $k \in \mathbb{R}$.

When a problem asks us to prove a fact about the product or ratio of two or more lengths, then similar triangles are usually at play - but make sure to prove that triangles are similar in the first instance by considering their interior angles.

Two triangles are *congruent* if they are exactly the same triangle. This means they have the same three sides and exactly the same three angles. There are some ways to identify congruence:

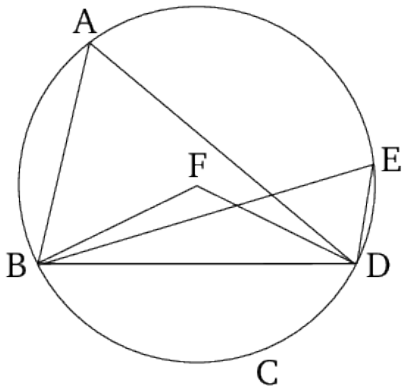
- **SSS**: if we have two triangles with all three sides equal.
- **SAS**: if we have two triangles where we know two sides and the angle between them are equal.
- **ASA**: if we have two triangles where we know two angles and the included side are equal.
- **AAS**: if we have two triangles where we know two angles and the non-included side are equal.
- **RHS**: if we have two right angled triangles with the same length of hypotenuse and the same length for one other of the two sides.

If we can identify any of the conditions above for two triangles, then we can say the two triangles are congruent. Note that **SSA**, which specifies two

sides and a non-included angle, and **AAA**, which specifies three angles, are not sufficient themselves to prove congruence.

3.2.5 Circle Theorems

We will use the following diagram for reference.



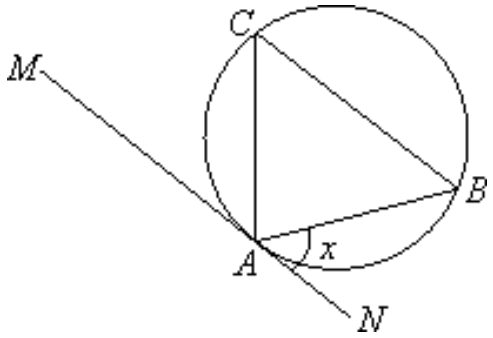
Angle at centre is twice angle at circumference
Consider quadrilateral $ABFD$. F is the centre of the circle. The angle subtended at the circumference is twice the angle subtended at the center so $\angle DAB = 2\times\angle DFB$. There is an important special case of this theorem, where angles in a semicircle are 90°.

Angles in the Same Segment
Angles in the *same segment* are equal, so $\angle DAB = \angle DEB$. We can reason this is true since both these angles have the same angle at the center, $\angle DFB$.

Cyclic Quadrilaterals
If we have four points which can be circumscribed, then these four points form a *cyclic quadrilateral*. Opposite angles in a cyclic quadrilateral add up to 180°. Consider quadrilateral $ABDE$, then $\angle DEA + \angle ABD = 180°$ and $\angle BDE + \angle EAB = 180°$.

When solving a problem, if we identify a cyclic quadrilateral it is worth circumscribing lightly since we can deduce other facts such as using the angles in the same segment theorem.

Alternate Segment Theorem
The *alternate segment theorem* is a favourite to employ. Whenever we see a tangent, we should always try and use this theorem wherever possible - it can be the key to getting started with many geometry problems.



The alternate segment theorem states that the angle between a chord and a tangent through one of the endpoints of the chord is equal to the angle in the alternate segment. By the alternate segment theorem, $\angle BCA = x$.

3.2.6 Length Ratios

Aside from using similar triangles, we can also involve the ratios of side lengths in a geometry problem, using the two following theorems.[1] This turns out to be very convenient in making progress with many problems.

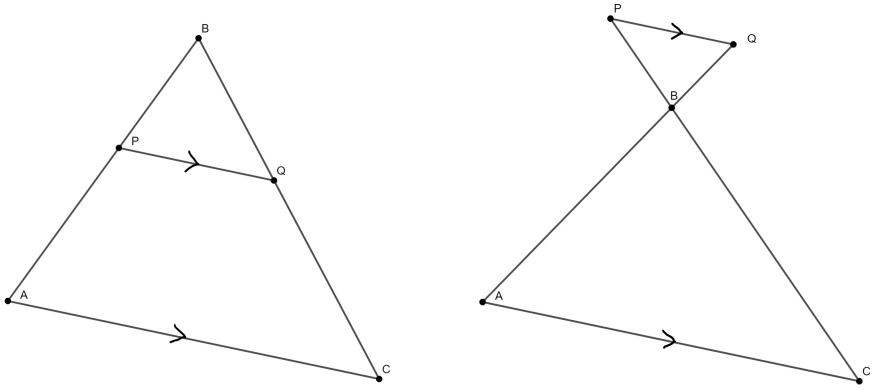
Intercept Theorem
A crucial fact that we should be aware of is, $\frac{a}{b} = \frac{c}{d} \iff \frac{a-c}{b-d}$. Actually, this is not extremely obvious, and nor is it well known, but it is very helpful. Of course, we must have that $a, b, c, d \neq 0$ and $b \neq d$. The proof is very simple:

$$\frac{a}{b} = \frac{c}{d} \implies ad = bc$$

$$\frac{a}{b} = \frac{a(b-d)}{b(b-d)} = \frac{ab-ad}{b(b-d)} = \frac{ab-bc}{b(b-d)} = \frac{a-c}{b-d} \quad \square$$

[1] There are many others which also relate the ratio of lengths together, such as power of a point and Ceva's - see Further Theorems.

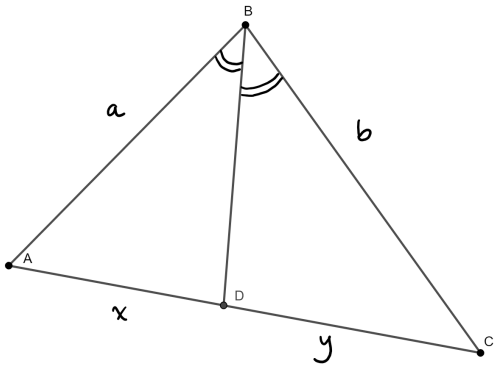
We can use this idea in the following setup:



Clearly, if we have a parallel line intersecting a figure as above, we have similar triangles. However, we can go further than this. Of course, $\frac{BP}{BA} = \frac{BQ}{BC}$, by similar triangles. But if we use the fact above, then we also get the *intercept theorem*:

$$\frac{BP}{PA} = \frac{BQ}{QC} \tag{3.2.7}$$

Angle Bisector Theorem



If we have an angle bisector in our setup, then the following relation holds.

$$\frac{a}{b} = \frac{x}{y} \tag{3.2.8}$$

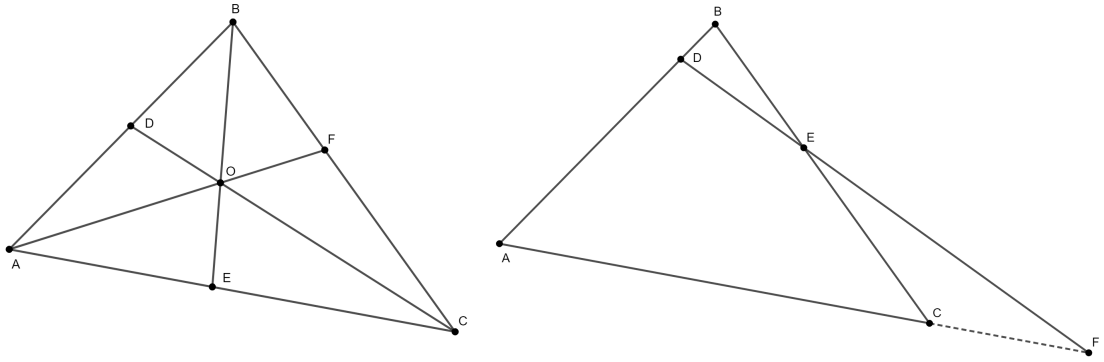
Conversely, if the relation holds, then we have an angle bisector at play. We can esily prove this with the sine rule and using the identity $\sin\theta = \sin(180 - \theta)$.

It is almost always inevitable to apply this idea whenever we have angle bisectors at play in the problem.

3.3 Further Theorems

3.3.1 Concurrency and Collinearity

Concurrency refers to the idea that three or more *lines* intersect at the same point. *Collinearity* refers to the idea that three or more *points* lie on the same line.



Ceva's Theorem
This theorem addresses *concurrency*. By looking at the figure above on the left side, then if we drop lines from each vertex of a triangle to the opposite sides, these lines are concurrent if and only if:

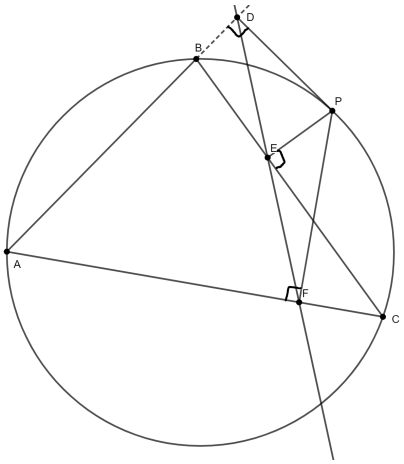
$$\frac{AE}{EC} \times \frac{CF}{FB} \times \frac{BD}{DA} = 1 \tag{3.3.1}$$

We can think of this statement, which seems like a lot going on, as taking each line segment going around the triangle's perimeter.

Menelaus' Theorem
This theorem addresses *collinearity*. By looking at the figure above on the right side, then if we have points on the lines of each side of a triangle, these points are collinear if and only if:

$$\frac{AD}{DB} \times \frac{BE}{EC} \times \frac{CF}{FA} = 1 \tag{3.3.2}$$

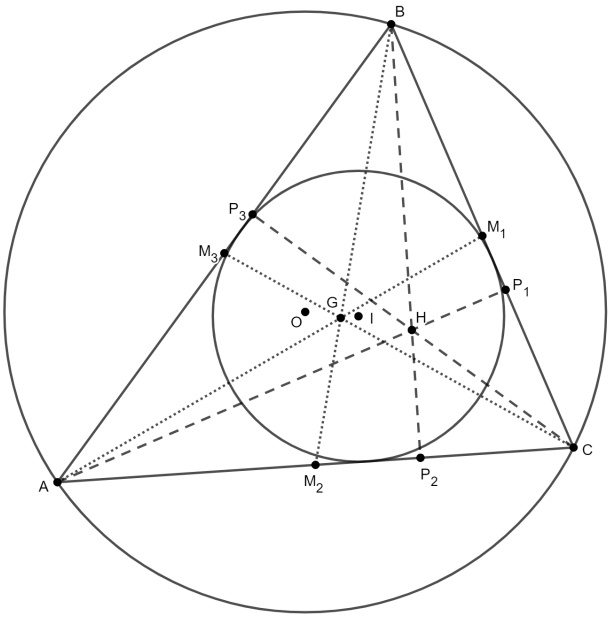
Simson Line
A very interesting and useful result. It can be proved with Menelaus and trig, or we could use a simpler similar triangles argument.



If we have a point P on the circumcircle of triangle ABC, and we drop perpendiculars from P to each of the sides of ABC, then these points on the sides of ABC are collinear. Hence, above we have that point D, E, F are collinear.

3.3.2 Triangle Centres

There are four main triangle centres we need to be aware of. We will use the following diagram for reference.



Points P_1, P_2, P_3 are formed by dropping perpendiculars (or *altitudes*) from the opposite vertices. Points M_1, M_2, M_3 are the midpoints of the sides.

Centroid
The *centroid* is the intersection of the three medians of the three vertices, so the point G above formed from the intersections of line segments AM_1, BM_2, CM_3. We can actually prove that these three line segments are concurrent in the first place using Ceva's theorem.

Orthocentre
The *orthocentre* is the intersection of the triangle's three altitudes, so the point H above formed from the intersections of line segments AP_1, BP_2, CP_3. Again, we can prove the existence of this concurrent point H using Ceva's theorem.

Circumcentre
The *circumcentre* is the centre of a triangle's circumscribed circle. That is, the circle which passes through all three vertices of the triangle. It is shown by the point O above.

Incentre
The *incentre* is the centre of a triangle's inscribed circle. That is, the circle which is tangent to all three sides of the triangle. It is shown by the point I above. Quite often when solving problems, it is helpful to draw the lines from the incentre to the tangential contact points on the sides (of course, these create right angles).

Without having to draw the incircle itself, we can construct the incentre by drawing angle bisectors from each of the vertices of our triangle. These angle bisectors are *concurrent* and meet at the *incentre*.

For both the *orthocentre* and the *circumcentre*, it is worth noting the following:

- If a triangle is obtuse (so it involves an angle bigger than 90°), then the orthocentre and circumcentre will be strictly outside the triangle.
- If a triangle is acute angled, these points will be inside the triangle.
- If a triangle is right angled, both the points will be on the triangle itself.

Note that unlike the other two, the *centroid* and *incentre* of a triangle are always inside the triangle.

Moreover, points O, G, H are collinear. So, the *circumcentre*, *centroid* and *orthocentre* are collinear - this is known as the *Euler line*. We can prove the existence of the Euler line using Menelaus' theorem.

Euler's Theorem

Euler's theorem tells us the distance d between the *centre* of the incircle and the *centre* of the circumcircle of any triangle. If the circumcircle of a triangle has radius R and the incircle has radius r, then the distance d between the centres is given by:

$$d^2 = R^2 - 2Rr \tag{3.3.3}$$

An interesting follow up result is that since all squares are non-negative,

$$R^2 - 2Rr = R(R-2r) \geq 0 \implies \underline{R \geq 2r} \tag{3.3.4}$$

Heron's Formula

Heron's formula gives us a way to compute the area of a triangle if we know only its side lengths. For a triangle with sides a, b, c, the area A is given by:

$$A = \sqrt{s\,(s-a)\,(s-b)\,(s-c)} \tag{3.3.5}$$

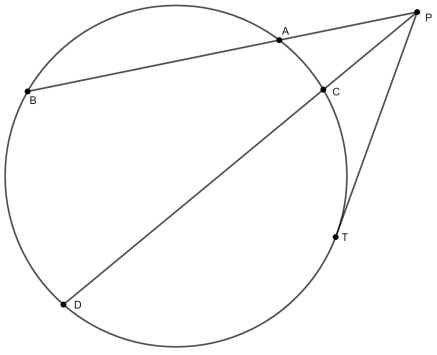
where $s = \frac{a+b+c}{2}$ and is known as the *semi-perimeter*.

This formula is not particularly useful in problems, although it is a handy idea to be aware of and one that is very simple to learn.

3.3.3 More Circle Theorems

Power of a Point

If we have the following setup with a circle, and a point outside the circle:



A, B, C, D are points of intersection with circle, and line PT is tangent to the circle.

$\triangle PAD$ and $\triangle PCB$ are indirectly similar. It follows that

$$PA \times PB = PC \times CD \tag{3.3.6}$$

Also, $\triangle PAT$ is indirectly similar to $\triangle PTB$ (using the alternate segment theorem), so

$$PA \times PB = PT^2 \tag{3.3.7}$$

It is well worth memorising these results to save time of chasing similar triangles.

Ptolemy's Theorem
With *Ptolemy's Theorem*, we concern ourselves with the side and diagonal lengths of a cyclic quadrilateral. For a cyclic quadrilateral $ABCD$,

$$AC \times BD = (AB \times CD) + (AD \times BC) \tag{3.3.8}$$

An easy way to remember this (and see what is actually going on) is:

The sum of the product of the opposite sides is equal to the product of the diagonal lengths.

3.4 Trigonometry

Trigonometry is not extremely important for olympiad problems, since they tend to be susceptible to other methods anyway. However, I thought it would be worth including, since it is imperative to know it inside out for 'pre-university' mathematics including for university entrance tests. Often trigonometry can help as another approach for geometry questions too where we need to make use of angles and things like sine rule.

3.4.1 Trigonometric Identities

These are the most crucial identities to be able to recall fluently.

$$\sin^2 A + \cos^2 A = 1 \tag{3.4.1}$$

$$\tan^2 A + 1 = \sec^2 A \tag{3.4.2}$$

$$1 + \cot^2 A = \csc^2 A \tag{3.4.3}$$

$$\sin(A \pm B) = \sin A \cos B \pm \sin B \cos A \tag{3.4.4}$$

$$\cos(A \mp B) = \cos A \cos B \mp \sin A \sin B \tag{3.4.5}$$

$$\tan(A \pm B) = \frac{\tan A \pm \tan B}{1 \mp \tan A \tan B} \tag{3.4.6}$$

If we let $A = B$, we get the double-angle formulae.

The factor formulae are also very important since they convert sums of trigonometric functions into products and vice versa.

$$\sin(A + B) + \sin(A - B) = 2 \sin A \cos B \tag{3.4.7}$$

$$\sin(A + B) - \sin(A - B) = 2 \cos A \sin B \tag{3.4.8}$$

$$\cos(A + B) + \cos(A - B) = 2 \cos A \cos B \tag{3.4.9}$$

$$\cos(A + B) - \cos(A - B) = -2 \sin A \sin B \tag{3.4.10}$$

	$-A$	$90 - A$	$180 - A$
sin	-sin$A$	cos$A$	sin$A$
cos	cos$A$	sin$A$	-cos$A$
tan	-tan$A$	cot$A$	-tan$A$

Below, I provide the reader with an example which is only meant to be fun; it has very little resemblance to the style of a serious olympiad problem other than some basic problem solving and geometry techniques.

Example If the angles A, B and C of a triangle are in an arithmetic progression and if a, b and c denote the lengths of the sides opposite to A, B and C respectively, then what is the value of the expression

$$\frac{a}{c} \sin 2C + \frac{c}{a} \sin 2A$$

Solution By the sine double-angle formula

$$\begin{aligned}\frac{a}{c}\sin 2C + \frac{c}{a}\sin 2A &= \frac{a}{c}2\sin C\cos C + \frac{c}{a}2\sin A\cos A \\ &= 2a\cos C\cdot\frac{\sin C}{c} + 2c\cos A\cdot\frac{\sin A}{a}\end{aligned}$$

But, by the sine rule, $\frac{\sin A}{a} = \frac{\sin C}{c}$. Substituting this into our expression above gives

$$\begin{aligned}\frac{a}{c}\sin 2C + \frac{c}{a}\sin 2A &= 2\left(\sin A\cos C + \sin C\cos A\right) \\ &= 2\sin(A+C)\end{aligned}$$

Now, we know $A+B+C = 180^\circ$ due to sum of interior angles in a triangle. Also, since the angles are in arithmetic progression

$$\begin{aligned}A &= A \\ B &= A + d \\ C &= A + 2d\end{aligned}$$

So, $A+B+C = 3A+3d = 180^\circ \implies A+d = 60^\circ$. But $A+d = B$, hence $B = 60^\circ$. Also, $A+C = 180^\circ - B = 120^\circ$.

$$\therefore\ \frac{a}{c}\sin 2C + \frac{c}{a}\sin 2A = 2\sin(A+C) = 2\sin(120^\circ) = \boxed{\sqrt{3}}$$

3.5 Loci

Olympiad loci questions in general are often the most challenging, as far as geometry is concerned. The key idea to have when tackling these questions is to:

Look for quantities or objects that are constant or fixed.

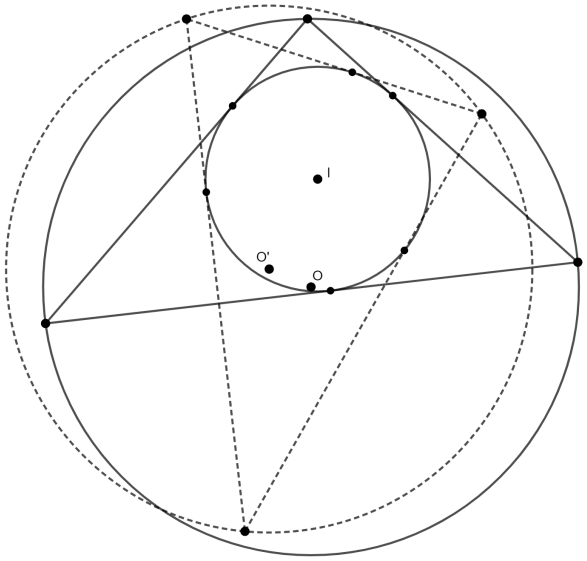
By definition, a *locus* is the set of all points, which satisfy one or more specified conditions. Often, the set of points can trace out a nice figure or curve. A neat example is that the locus of points formed by the focus of a parabola rolling along a straight line is a 'cosh curve'.[2]

The best way with loci problems is through plenty of practice and use basic geometry theorems, but also try and play around with what's going on, try special cases and, most importantly, try to find a quantity which is fixed.

[2]The hyperbolic cosine function defined by $\frac{e^x+e^{-x}}{2}$.

Example A variable triangle has a fixed incircle. Given that its circumradius is constant, find the locus of the circumcentre.

Solution At first glance, this problem seems quite intimidating. However, it is worth drawing a diagram to get a feel for what's going on and then to identify quantities which are constant no matter how we shift the setup.



I have drawn a solid triangle and its circumcircle with centre O and a dashed triangle and its circumcircle with centre O'. As per the question, both triangles have the same fixed incircle with incentre I and the same circumradius.

Of course, from the question, the incentre I, the circumradius R and the inradius r are fixed for any of our triangles. By, Euler's theorem, we know that

$$OI^2 = R^2 - 2Rr$$

where OI denotes the distance between the circumcentre and incentre for any of our triangles. Since $R^2 - 2Rr$ is clearly constant, this means that OI must also be constant.

In other words, the distance between the circumcentre and the incentre for all our satisfied triangles must be constant. Now, we are after the locus of the circumcentre. As the incentre is fixed, we have that the locus of the circumcentre O must be a circle centred at I with radius $\sqrt{R^2 - 2Rr}$. □

Chapter 4

Number Theory

Number theory is the study of the integers and their properties. Problems usually concern themselves with finding integer solutions, dealing with primes, divisibility or showing a number can be written in some form (such as the sum of two squares).

Note that it is extremely common to prove results in number theory by *contradiction*.

It is also worth being clear about the different *sets* of numbers we can have.[1]

- $\mathbb{R}$ is the set of all **real** numbers, including numbers from integers to irrational numbers like π and everything else in between.
- $\mathbb{Q}$ is the set of all **rational** numbers, which are numbers that can be expressed as a fraction of two whole numbers. If a number $x \notin \mathbb{Q}$, then x is **irrational** and cannot be expressed as a fraction of two whole numbers *e.g.* $\sqrt{2}$.
- $\mathbb{Z}$ is the set of all **integers**, which are whole numbers both positive and negative.
- $\mathbb{N} = \mathbb{Z}^+$ is the set of all **positive integers** so $1, 2, 3, \ldots$ *etc.*, and we have two names for the same thing here, since occasionally $\mathbb{N}$ includes 0, but this will be specified and is the exception rather than the norm.

[1] I am deliberately excluding complex numbers.

4.1 Divisibility

Suppose that m and n are integers. We say that m *divides* n, if there is an integer d such that $n = md$. In this case, we also say that m is a *divisor* or *factor* of n, and we can write $m|n$.

4.1.1 Numbers in base 10

We can define a number N with digits a_n, $a_{n-1}, \ldots, a_1$, a_0 as

$$N = 10^n \cdot a_n + 10^{n-1} \cdot a_{n-1} + \cdots + 10^1 \cdot a_1 + 10^0 a_0 \tag{4.1.1}$$

For example, $1729 = 10^3 \cdot 3 + 10^2 \cdot 7 + 10^1 \cdot 2 + 10^0 \cdot 9$. This concept is actually quite helpful, and can be generalise to numbers in other *bases*. We typically deal with base 10, however if we replaced the powers of 10 above with some other number k, then we can move into base k.

Example What are the last two digits of 9^9 and 9^{9^9}?

Solution The most useful observation to make in this problem is that $9 = 10 - 1$. This will be extremely helpful, since we have a 10 and if we can sort out multiples of 10, then we will be able to examine the units and tens digits of 9^9 and 9^{9^9}. We will also make use of the binomial theorem. First let's consider 9^9.

$$\begin{aligned} 9^9 = (10-1)^9 &= \binom{9}{0}10^9 + \binom{9}{1}10^8 \cdot (-1)^1 + \binom{9}{2}10^7 \cdot (-1)^2 + \ldots \\ &\qquad + \binom{9}{8}10^1 \cdot (-1)^8 + \binom{9}{9}(-1)^9 \\ &= 10^9 - 9 \cdot 10^8 + 36 \cdot 10^7 + \cdots + 9 \cdot 10 - 1 \\ &= 10^9 - 9 \cdot 10^8 + 36 \cdot 10^7 + \cdots + 89 \end{aligned}$$

Now, here we have to be careful. Notice, that every term before the last 89 will have a factor of 10^2 *i.e.* 100. So the number 9^9 will be of the form $100k + 89$. Also, clearly 9^9 is positive, so the value of the integer k will also be positive (despite the alternating signs we see from the binomial expansion). Hence the last two digits of 9^9 is $\boxed{89}$.

For 9^{9^9} we will do a similar thing. We have established that 9^9 can be written as $100k + 89$ for some integer k. But, to make our life easier,

$100k + 89 = 10(10k + 8) + 9$ and let's substitute $10k + 8$ for K.

$$\begin{aligned} \therefore\ 9^{9^9} = 9^{10K+9} &= 9^9 \times 9^{10K} \\ &= 9^9 \times \left((10-1)^{10}\right)^K \\ &= 9^9 \times \left(10^{10} - 10 \cdot 10^9 + \cdots - 10 \cdot 10^1 + 1\right)^K \end{aligned}$$

Examining the right hand bracket above, we can pull out factors of 100, so

$$\left(10^{10} - 10 \cdot 10^9 + \cdots - 10 \cdot 10^1 + 1\right)^K = (100m + 1)^K$$

for some integer m. Now, if we try to expand out $(100m + 1)^K$, it should be clear that all terms will have a factor of 100 in them apart from the final term of $1^K = 1$. So we can say further that $(100m + 1)^K = 100n + 1$ for some integer n.

$$\therefore\ 9^{9^9} = 9^9 \times (100n+1) = (100k+89) \cdot (100n+1) = 100(100kn+89n+k)+89$$

Hence, 9^{9^9} also has last digit of $\boxed{89}$.

In fact, all numbers $9^{9^{\cdot^{\cdot^{\cdot^9}}}}$ will have last two digits of 89.

4.1.2 Basic Divisibility Rules

These should be trivial, but below are rules for how we can quickly check a number is divisible by any small numbers (1-12).

1. Any integer is divisible by 1.
2. The last digits is even (0, 2, 4, 6, 8).
3. The sum of the digits is divisible by 3. You can repeat this rule as many times.
4. The last two digits are divisible by 4.
5. The last digits is 0 or 5.
6. It is divisible by both 2 and 3.
7. Double the last digit and subtract it from the number made by the other digits. The result is divisible by 7. You can repeat this rule as many times.
8. The last three digits are divisible by 8, or you can halve the number three times and it is still an integer.

9. The sum of the digits is divisible by 9. You can repeat this rule as many times.

10. The number ends in 0.

11. Subtract and add the digits in an alternating pattern, *e.g.* take first digit subtract next one add next one..., and the result is divisible by 11.

12. It is divisible by both 3 and 4.

Primes

A positive integer with exactly two positive divisors is said to be *prime*. Notice that by this definition 1 is not prime. The first few prime numbers are:

$$2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, \ldots$$

The *fundamental theorem of arithmetic* states that any integer greater than 1 is either a prime number or can be written as a unique product of prime numbers. This can be proved by contradiction.

The opposite of prime numbers are known as *composite* numbers, which are made up of a unique product of primes. Hence, it is clear by the fundamental theorem of arithmetic that every integer greater than 1 is either prime or composite.

Example Prove that there are infinitely many prime numbers.

Solution Suppose, for contradiction, that there are only finitely many primes, $2, 3, \ldots, p_n$, and p_n is the largest prime. Now let's create a number N by multiplying all these primes together and adding 1.

$$N = (2 \times 3 \times 5 \times \cdots \times p_n) + 1$$

Clearly, the number N has remainder 1 when divided by any prime number in our finite list of primes. Hence, N has no prime factors, so N must be prime. But, notice that N is larger than all other primes, so there cannot exist a largest prime p_n, and thus we have arrived at a contradiction. Therefore, there must be infinitely many prime numbers. $\square$

4.1.3 Consequences for Divisors

We will begin by defining a function Ω, which can take an input of two integers m, n and $\Omega(m, n)$ outputs the set of integers which divide both m and n. In other words, the set of *common divisors* of m and n.

$$\Omega(m, n) = \{x \in \mathbb{Z} : x|m, x|n\} \tag{4.1.2}$$

[2]For example $\Omega(2, 4) = \{-2, -1, 1, 2\}$, although it will make our lives easier if we restrict ourselves to *positive* divisors, since if a divides b then automatically $-a$ also divides b. So, we can make another function Ω^+, which restricts Ω to only positive common divisors. So, $\Omega^+(2, 4) = (1, 2)$.

Greatest Common Divisor

The *greatest common divisor*, gcd, of two numbers m, n is given by:

$$\gcd(m, n) = \max\{\Omega(m, n)\} = \max\{\Omega^+(m, n)\} \tag{4.1.3}$$

Note that $\gcd(0, 0)$ does not exist since $\Omega(0, 0) = \{\dots, -1, 0, 1, 2, \dots\}$ and so there is no maximum element.

Integers m, n are said to be *coprime* if and only if $\gcd(m, n) = 1$.

The gcd function lends itself to some useful properties:

- $\gcd(m, n) = \gcd(-m, n) = \gcd(m, -n) = \gcd(-m, -n)$.
- For a non-negative integer k, $\gcd(k \cdot m, k \cdot n) = k \cdot \gcd(m, n)$.
- $\gcd(m, n, p) = \gcd(\gcd(m, n), p) = \gcd(m, \gcd(n, p)) = \gcd(n, \gcd(m, p))$.
- For any integer k, $\gcd(m + kn, n) = \gcd(m, n)$.

Lowest Common Multiple

We define the *lowest common multiple* of two integers m, n as follows:

$$\operatorname{lcm}(m, n) = \frac{|m \times n|}{\gcd(m, n)} \tag{4.1.4}$$

The lcm gives the smallest integer, which is a multiple of both m and n.

[2]Note that the colon ':' reads 'such that'.

Number of factors
Every integer N can be expressed as:

$$N = a^\alpha \times b^\beta \times c^\gamma \times \dots$$

where $a, b, c, \dots$ are distinct primes and $\alpha, \beta, \gamma, ... \in \mathbb{N}$.

If N is a power of only one prime then $N = p^\alpha$. Therefore, N has $\alpha + 1$ factors $(1, p, \dots, p^{\alpha-1}, p^\alpha)$.

This generalises for $N = a^\alpha \times b^\beta \times c^\gamma \times \dots$, and it follows that the *number of factors* of N is

$$(\alpha + 1)(\beta + 1)(\gamma + 1) \dots \tag{4.1.5}$$

This result for the number of factors is extremely useful for maths problems generally. I would highly recommend the reader to learn this result and try and reason why it works.

Sum of factors
If we take our number expressed as a product of distinct primes from above, $N = a^\alpha \times b^\beta \times c^\gamma \times \dots$, then it also follows that the *sum of factors* of N is

$$\left(1 + a + a^2 + \dots + a^\alpha\right)\left(1 + b + b^2 + \dots + b^\beta\right)\left(1 + c + c^2 + \dots + c^\gamma\right) \dots \tag{4.1.6}$$

or if you prefer,

$$\prod_{i=1}^{k} \left(\frac{p_i^{a_i+1} - 1}{p_i - 1} \right) \qquad (\text{for } N = p_1^{a_1} p_2^{a_2} \dots p_k^{a_k}) \tag{4.1.7}$$

For instance, it is trivial that the sum of the factors of $N = p^\alpha$ is $1 + p + p^2 + \dots + p^\alpha$.

4.1.4 Euclidean Algorithm

The *Euclidean algorithm* enables us to compute *iteratively* the gcd of two numbers. It makes use of the fourth bullet point in the greatest common divisor section above.

Given m and n, we can find $\gcd(m, n)$ as follows

$$\begin{aligned}
&1.\ \text{Write } a = q_1 b + r_1, \text{ where } r_1 < b \\
&2.\ \text{Write } b = q_2 r_1 + r_2, \text{ where } r_2 < r_1 \\
&3.\ \text{Write } r_1 = q_3 r_2 + r_3, \text{ where } r_3 < r_2 \\
&\qquad\vdots
\end{aligned}$$

Eventually, we will arrive at $r_k = 0 \implies r_{k-1}$ is the $\gcd(m, n)$.

Effectively, what we are doing, in say step 1, is finding the largest multiple of b that's less than a and the remainder is r_1. Throughout, we use the fact that $\gcd(m + kn, n) = \gcd(m, n)$ to constantly reduce our expression.

Example What is $\gcd(306, 657)$?

Solution By the Euclidean algorithm,

$$\begin{aligned}
657 &= 2 \times 306 + 45 \\
306 &= 6 \times 45 + 36 \\
45 &= 1 \times 36 + 9 \\
36 &= 4 \times 9\ (+0) \qquad \therefore\ \gcd(306, 657) = \boxed{9}
\end{aligned}$$

Example Prove that the fraction $\frac{21n+4}{14n+3}$ is irreducible for every natural number n.

Solution If the fraction is irreducible, then the numerator and denominator must have no common factors. Thus, this would mean that they are coprime. So, it suffices to show that $21n + 4$ and $14n + 3$ are coprime. We will proceed with the Euclidean algorithm. $\gcd(21n + 4, 14n + 3)$:

$$\begin{aligned}
(21n + 4) &= 1 \times (14n + 3) + (7n + 1) \\
(14n + 3) &= 2 \times (7n + 1) + 1 \\
(7n + 1) &= (7n + 1) \times 1\ (+0)
\end{aligned}$$

Therefore, $\gcd(21n+4, 14n+3) = 1$ for any n, so the fraction is irreducible. $\square$

4.2 Modular Arithmetic

In *modular arithmetic*, we concern ourselves with the *remainder* when certain numbers or expressions when divided by some other number. This idea of examing a remainder is particularly useful. We have implicity considered modulo 10 and 100 in the example above with the stacked powers of 9.

For instance, with a simple example, we know all odd numbers can come in the form either $4k+1$ or $4k+3$ for integer k. Clearly, $4k+1$ gives remainder 1 upon division by 4 and $4k+3$ gives remainder 3 upon division by 4. So we can say:

$$4k+1 \equiv 1 \pmod 4, \; 4k+3 \equiv 3 \pmod 4$$

This means that odd numbers can either be 1 (mod 4) or 3 (mod 4). However, also notice that $4k+3$ can be written as $4k-1$ for some different k and hence $3 \pmod 4 \equiv -1 \pmod 4$. Similarly, if we consider (mod 2), then odd numbers are equal to 1 (mod 2) and even numbers 0 (mod 2).

Here are some crucial rules to bear in mind when dealing with *modular arithmetic*.

Addition

- If $a+b=c$, then $a+b \equiv c \pmod n$
- If $a \equiv b \pmod n$, then $a+k \equiv b+k \pmod n$ for any integer k
- If $a \equiv b \pmod n$ and $c \equiv d \pmod n$, then $a+c \equiv b+d \pmod n$

Multiplication

- If $a \cdot b = c$, then $a \cdot b \equiv c \pmod n$
- If $a \equiv b \pmod n$, then $ka \equiv kb \pmod n$ for any integer k
- If $a \equiv b \pmod n$, then $ka \equiv kb \pmod{kn}$ for any real number k
- If $a \equiv b \pmod n$ and $c \equiv d \pmod n$, then $a \cdot c \equiv b \cdot d \pmod n$

4.2.1 Fermat's Little Theorem

Fermat's little theorem is a fundamental idea in number theory, which helps compute powers of integers modulo prime numbers.

The result is called Fermat's 'little theorem' so as to distinguish it from Fermat's Last Theorem. Here are three versions of the result (all of which are the same thing).

1. Suppose that p is a prime number and x is an integer which is not divisible by p, then p divides $x^{p-1} - 1$.
2. Suppose that p is a prime number and x is an integer such that $x \neq 0 \pmod{p}$, then $x^{p-1} \equiv 1 \pmod{p}$.
3. Suppose that p is a prime number and x is an integer, then p divides $x^p - x$.

4.2.2 Chinese Remainder Theorem

The *Chinese remainder theorem* is a theorem which helps us solve problems such as: 'find all integers that leave remainder 1 when divided by 2, 3 and 5'.

A neat way of initially viewing *congruences* and the Chinese remainder theorem is by looking at *arithmetic sequences*.

Let's suppose we have the following two sequences of integers:

$$\ldots, -12, -7, -2, 3, 8, 13, 18, \ldots \tag{4.2.1}$$

$$\ldots, -15, -1, 13, 27, 41, \ldots \tag{4.2.2}$$

In sequence (4.2.1), each term differs by 5 and in sequence (4.2.2), each term differs by 14. If we consider the *intersection* of these two sequences, that is the new sequence formed by the terms which appear in both of the above sequences, then we obtain the following arithmetic sequence:

$$\ldots, -127, -57, 13, 83, 153, \ldots \tag{4.2.3}$$

Notice that each term in sequence (4.2.3) differs by $70 = 5 \times 14$. The one thing to be aware of in this, however, is that $\gcd(5, 14) = 1$ so 5 and 14 are *coprime*. If you omit this condition that the common differences have to

be coprime, then it is not always possible that the two sequences intersect at all.

We do not necessarily need to restrict ourselves to two sequences either; this will work with any number of sequences, provided that they have coprime common differences.

Essentially, what we are saying is that given some sequences with common differences that are *pairwise* coprime, the intersections of all these sequences forms a new, *unique* sequence with common difference equal to the product of all the common differences of the original sequences.

Now instead of sequences, we can formalise this idea to *linear congruences* [3], yielding the Chinese remainder theorem:

Given pairwise coprime integers $n_1, n_2, \ldots, n_k$ and arbitrary integers $a_1, a_2, \ldots, a_k$, the system of linear congruences

$$\begin{aligned} x &\equiv a_1 \pmod{n_1} \\ x &\equiv a_2 \pmod{n_2} \\ &\vdots \\ x &\equiv a_k \pmod{n_k} \end{aligned}$$

has a solution x, and this solution is unique (*i.e.* there is only one possible solution) modulo $N = n_1 n_2 \ldots n_k$.

Example Find the smallest positive integer x which leave a remainder of 1, 2 and 3 when divided by 7,4 and 5 respectively.

Solution We can formalise the problem into the following congruences:

$$\begin{aligned} x &\equiv 2 \pmod{4} \\ x &\equiv 3 \pmod{5} \\ x &\equiv 1 \pmod{7} \end{aligned}$$

Notice that the moduli are all pairwise coprime, so by the Chinese remainder theorem, there exists a unique solution $x \pmod{4 \cdot 5 \cdot 7}$.

[3] If you observe carefully, an arithmetic sequence with common difference n is in fact all the possible numbers which give a certain remainder upon division by n. For example, in (4.2.1) each term of the sequence is 3 (mod 5).

To solve and find this x, we will write x as follows

$$x = x_1(5 \cdot 7) + x_2(4 \cdot 7) + x_3(4 \cdot 5)$$

The beauty of this now is that we can take modulo on both sides and things will vanish to 0. Taking $\pmod 4$ gives $x \equiv 35 \cdot x_1 \equiv 3 \cdot x_1 \pmod 4$. But, we also want $x \equiv 2 \pmod 5$. So, if $x_1 = 2$, then we are good.

Similarly, with some strong arithmetic, we can obtain $x_2 = 1$ and $x_3 = 6$.

$$\begin{aligned} \therefore \; x &= 2 \cdot 5 \cdot 7 + 1 \cdot 4 \cdot 7 + 6 \cdot 4 \cdot 5 \\ &= 70 + 28 + 120 \\ &= 218 \end{aligned}$$

But this x is unique modulo $4 \cdot 5 \cdot 7 = 140$, so $x \equiv 218 \equiv 78 \pmod{140}$. Hence, the smallest positive integer which satisifies the congruences is $\boxed{78}$.

4.3 Perfect Squares

4.3.1 Quadratic Residues

Sometimes, it is worth checking how perfect squares behave upon division by some certain integer. For instance, square numbers are never $2 \pmod 3$ and square numbers are also always $0 \pmod 4$ if they are even and $1 \pmod 4$ if they are odd.

Example Prove the very important fact that square numbers are only 0 or $1 \pmod 3$ and never $2 \pmod 3$.

Solution Let a be an integer. We will use the result from modular arithmetic that is $a \equiv b \pmod 3$, then $a^k \equiv b^k \pmod 3$. There are three cases to consider,

$$\begin{aligned} &(1) : a \equiv 0 \pmod 3, \text{ then } a^2 \equiv 0^2 \equiv 0 \pmod 3 \\ &(2) : a \equiv 1 \pmod 3, \text{ then } a^2 \equiv 1^2 \equiv 1 \pmod 3 \\ &(3) : a \equiv 3 \pmod 3, \text{ then } a^2 \equiv 2^2 \equiv 4 \equiv 1 \pmod 3 \quad \square \end{aligned}$$

I will leave it as an exercise for the reader to prove the result for modulo 4.

Example Let k be an integer such that $k = 3m + 1$ for some integer m. Find all integer solutions (x, y) to the equation

$$(x + y - k)(x + y + k) = 1 + xy$$

Solution Since $k = 3m + 1$, we can immediately substitute this into our equation. So we are looking for pairs of integer solutions (x, y) to the equation

$$(x + y - 3m - 1)(x + y + 3m + 1) = 1 + xy$$

Expanding everything out yields

$$\begin{aligned} x^2 + y^2 + 2xy - 9m^2 - 6m - 1 &= 1 + xy \\ \implies x^2 + y^2 + xy - 2 &= 9m^2 + 6m = 3m(3m + 2) \end{aligned}$$

This therefore tells us that if we are to have an integer solution (x, y) to the equation then $x^2 + y^2 + xy - 2$ must be divisible by 3. I suggest we make a substitution now letting $N = m(3m+2)$ so $N \in \mathbb{Z}$. This gives

$$\begin{aligned} x^2 + y^2 + xy - 2 &= 3N \\ x^2 + y^2 + xy &= 3N + 2 \\ (x - y)^2 + 3xy &= 3N + 2 \end{aligned}$$

Taking $\pmod 3$ on both sides, we get $(x - y)^2 \equiv 2 \pmod 3$. This implies that if exists a solution with $x, y \in \mathbb{Z}$ then $(x - y)^2$ must give remainder 2 upon division by 3. Of course, $x - y$ is also an integer. But this is a contradiction - no square number can ever be 2 $(\mathrm{mod}\ 3)$ (see previous example). Hence, there exists no integer solutions to the equation. □

4.3.2 Roots and Rational Numbers

Suppose that p and q are positive integers such that $\sqrt[q]{p}$ is a rational number. Then, it follows that $\sqrt[q]{p}$ must be an integer.

Factors of $4n^2 + 1$

If you take any number of the form $4n^2 + 1$, where n is an integer, then of course all the factors of it are odd, since clearly $4n^2 + 1$ is odd. But it turns out further that all these factors are odds of the form $4m + 1$ for some integer m!

We can show this amazing fact using a proof by contradiction and Fermat's little theorem. Since $4n^2+1$ is odd, all of its *prime factors* will be odd numbers either of the form $4m+1$ or $4m+3$. It is enough to show that $4n^2+1$ has no prime factor of the form $4m+3$.

Let's suppose for contradiction that a prime number $p = 4m+3$ divides $4n^2+1$. As we are trying to engage Fermat's little theorem into play, it is sensible to take $(\text{mod } p)$ of both sides, giving

$$4n^2+1 \equiv 0 \pmod{p} \implies 4n^2 \equiv -1 \pmod{p}$$

Notice though $4k^2 = (2k)^2$. By Fermat's little theorem, we know $x^{p-1} \equiv 1 \pmod{p}$ provided that $x \not\equiv 0 \pmod{p}$. In this case, our 'x' is $2k$ and clearly $2k$ does not divide p as p is an odd prime, so we are good to go. We have

$$(2k)^2 \equiv -1 \pmod{p}$$

As we want to introduce a power of $p-1$, it would make sense to raise both sides to the power of $\frac{p-1}{2}$.

$$\therefore \left((2k)^2\right)^{\frac{p-1}{2}} \equiv (-1)^{\frac{p-1}{2}} \pmod{p} \implies (2k)^{p-1} \equiv (-1)^{\frac{p-1}{2}} \pmod{p}$$

Now, recall that $p = 4m+3$.

$$\therefore (2k)^{p-1} \equiv (-1)^{\frac{4m+3-1}{2}} \equiv (-1)^{2m+1} \equiv -1 \pmod{p}$$

But this is a contradiction! Fermat's little theorem states that we should have $(2k)^{p-1} \equiv 1 \pmod{p}$. Hence, there exist no factors of $4n^2+1$ which have the form $4m+3$. We could follow a similar logic with primes $p = 4m+1$, and we would not arrive at a contradiction. $\square$

Primes of the form $a^n - 1$

There are some nice things we can deduce about numbers of the form a^n-1. If we recall (1.1.4), then a useful idea is that $a^n - 1 = (a-1)(a^{n-1} + a^{n-2} + \cdots + a + 1)$.

If you take any prime number of the form $a^n - 1$, then there are some constraints on a and n. In fact, it must be that $a = 2$ and n is prime. We can go about showing this.

Suppose p is a prime, such that $p = a^n - 1$.

$$p = (a-1)(a^{n-1} + a^{n-2} + \cdots + a + 1)$$

This shows that p is composite (*i.e.* not prime) unless $a-1=1 \implies a=2$. So, $p=2^n-1$. For n, let's now proceed with contradiction. Suppose that n is not prime so it can be written as $n=ab$, for some integers $a,b \geq 2$.

$$\therefore\ p = 2^{ab}-1 = (2^a)^b - 1 = (2^a-1)(2^{a(b-1)} + 2^{a(b-2)} + \cdots + 2^a + 1)$$

As we supposed n is not prime, we had $a \geq 2$. This means that $2^a - 1 \geq 3$, so p would be composite and not prime. This is a contradiction! Hence, n must be prime for p to be prime.

So, primes of the form a^n-1 have $a=2$ and n prime. In fact, these primes are known as *Mersenne primes*.

4.3.3 Fermat's Last Theorem

This idea is pretty useless and pointless as far as an olympiad problem is concerned, but it's cool! Fermat's Last Theorem states that there are no integer solutions (x,y,z) to the equation

$$x^n + y^n = z^n \qquad \text{for } n \geq 3 \tag{4.3.1}$$

4.4 Integer Equations

Many times, we would have to concern ourselves with solving an equation (typically with two variables x,y), where we must restrict our variables to integer values only. As a result, no integer solutions could exist, or some integer solutions could exist, or even infinite integer solutions could exist.

In general, finding integer solutions to an equation requires a very systematic approach, and sometimes brute-force/trial and error can help break into an equation as it is always a good idea to try and find some small cases which work. Below are three common forms of equations we could come across.

4.4.1 Difference of two squares

Here, we will consider integer solutions (x,y) to the equation

$$x^2 - y^2 = k \tag{4.4.1}$$

For any equation of that form, we should turn it into $(x+y)(x-y)=k$, and then consider the possible factorisations of k.

For example, if k is prime then it can only be factored into $1 \times k$ or $k \times 1$, so $x + y = 1$ and $x - y = k$, or $x + y = k$ and $x - y = 1$. This gives us simultaneous equations which we can easily use to compute a possible pair (x, y). Also, note that if x is a solution, then so is $-x$ due to the squares.

However, if a number factors into an odd factor multiplied by an even factor, then we cannot get an integer solution. Take the example $k = 12 = 4 \times 3$. If we take $x + y = 4$, $x - y = 3$, this gives $2x = 7$ but we need x as an integer.

On the other hand, if we use the factoring of $k = 6 \times 2$, then $x + y = 6$, $x - y = 2$ and $2x = 8$, so $(x, y) = (4, 2)$. This way, in the case $k = 12$ the only integer solutions are $(4, 2)$, $(-4, -2)$, $(4, -2)$ and $(-4, 2)$.

By the same logic, there exist no integer solutions in the case $k = 2$ or $k = 6$.

4.4.2 Bezout's Identity

Recall the Euclidean algorithm for computing the gcd of two numbers. We can also work backwards from this, which gives us *Bezout's identity*.

If $\gcd(a, b) = d$, then there exists integer solutions (x, y) to the equation

$$ax + by = d \tag{4.4.2}$$

Furthermore, there exist integer solutions (x, y) to the equation

$$ax + by = n \tag{4.4.3}$$

if and only if $d|n$.

4.4.3 Pell's Equation

Pell's equation concerns integer solutions (x, y) to equations of the form

$$x^2 - ny^2 = 1 \tag{4.4.4}$$

where n is a *nonsquare* positive integer. It can be shown that there are infinitely many solutions and they generate *recursively* from a simple, fundamental solution.

Instead, if n was a square, then the only possible solutions are $(x, y) = (-1, 0)$ and $(x, y) = (1, 0)$. This can be shown by considering a difference of two squares factorisation and considering that $1 = 1 \times 1$ or $1 = -1 \times -1$.

Example Let $\mathbb{Z}[\sqrt{2}]$ denote the set of all real numbers r of form $r = a + b\sqrt{2}$, $a, b \in \mathbb{Z}$. For such an r define $N(r) = a^2 - 2b^2$. Show that $N(rs) = N(r)N(s)$, and hence show that there exist infinitely many pairs of integers (a, b) with $a^2 - 2b^2 = \pm 1$.

Solution Let $r = a + b\sqrt{2}$ and $s = c + d\sqrt{2}$ $(a, b, c, d) \in \mathbb{Z}$. Therefore, $rs = (ac + 2bd) + (ad + bc)\sqrt{2}$.

$$\begin{aligned} \implies N(rs) &= (ac + 2bd)^2 - 2(ad + bc)^2 \\ &= a^2c^2 + 4abcd + 4b^2d^2 - 2a^2d^2 - 4abcd - 2b^2c^2 \\ &= a^2(c^2 - 2d^2) - 2b^2(c^2 - 2d^2) \\ &= \left(a^2 - 2b^2\right)\left(c^2 - 2d^2\right) = \boxed{N(r)N(s)} \end{aligned}$$

The equation $a^2 - 2b^2 = 1$ is effectively $N(r) = 1$, where $r = a + b\sqrt{2} (a, b \in \mathbb{Z})$. We can use the property that $N(rs) = N(r)N(s)$ to prove that there are infinitely many integers satisfying $N(r) = 1$.

Notice that $1^n = 1$ for any $n \in \mathbb{Z}$. Therefore if we know one solution (a, b) for which $N(r) = 1$ then $N(r^2) = N(r)N(r) = 1^2 = 1$. Similarly $N(r^3) = 1$, $N(r^4) = 1$ and so on.

It is not difficult to spot that $(a, b) = (3, 2)$ is a solution since $3^2 - 2(2)^2 = 1$. Hence, it is also true that $N\left((3 + 2\sqrt{2})^n\right) = 1$ for all $n \in \mathbb{N}$. Since the set $\mathbb{N}$ is an infinite set, there exist infinitely many pairs of integers satisfying $a^2 - 2b^2 = 1$.

A similar thing can be done for $a^2 - 2b^2 = N(r) = -1$. We use the fact that $(-1)^n = -1$ for all odd n. A simple solution is $(a, b) = (1, 1)$ since $1^2 - 2(1)^2 = -1$. So, it is also true that $N\left((1 + \sqrt{2})^n\right) = -1$ for all odd numbers n. Since the set of odd numbers is an infinite set, there exist infinitely many pairs of integers satisfying $a^2 - 2b^2 = -1$. $\square$

Bibliography

[1] Geoff Smith, *A Mathematical Olympiad Primer*.

[2] Geoff Smith, *A Mathematical Olympiad Companion*.

[3] Michael Ng, *10 Useful Techniques for Solving Olympiad Problems*.