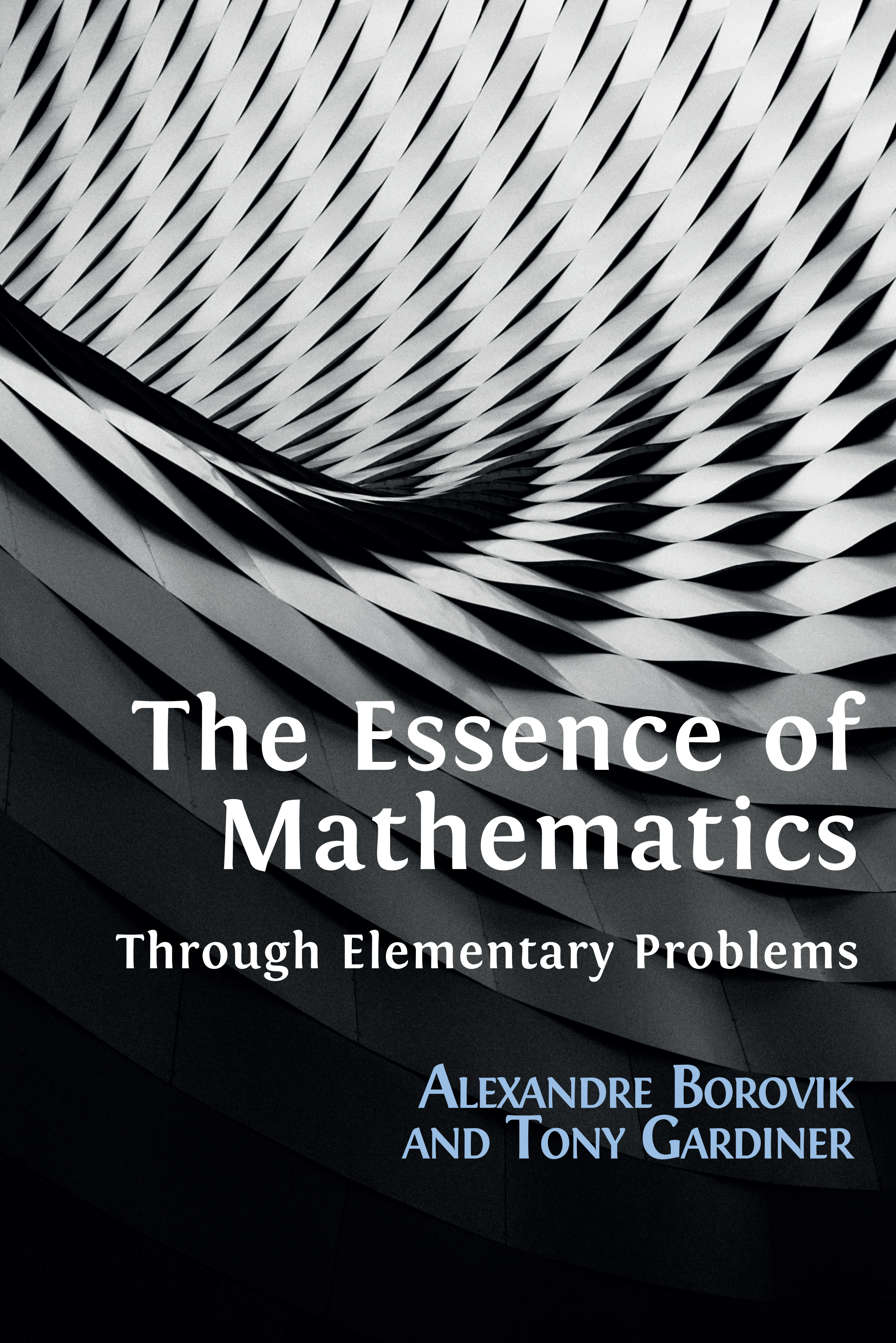


The Essence of Mathematics

Through Elementary Problems

Alexandre Borovik
and Tony Gardiner

THE ESSENCE OF MATHEMATICS

THROUGH ELEMENTARY PROBLEMS

The Essence of Mathematics

Through Elementary Problems

Alexandre Borovik and Tony Gardiner

OpenBook Publishers
Knowledge is for sharing

http://www.openbookpublishers.com

Digital material and resources associated with this volume are available at
http://www.openbookpublishers.com/product/979#resources

This is the third volume of the OBP Series in Mathematics:

ISSN 2397-1126 (Print)
ISSN 2397-1134 (Online)

ISBN Paperback 9781783746996
ISBN Hardback: 9781783747009
ISBN Digital (PDF): 9781783747016
DOI: 10.11647/OBP.0168

Cover photo: *Abstract Spiral Pattern* (2015) by Samuel Zeller,
https://unsplash.com/photos/j0g8taxHZa0

Cover design by Anna Gatti.

Contents

Simon Phillips Norton
1952–2019

In memoriam

Preface

Understanding mathematics cannot be
transmitted by painless entertainment
. . . actual contact with the ***content*** *of*
living mathematics is necessary.
The present book . . . is not a concession
to the dangerous tendency toward
dodging all exertion.
Richard Courant (1888–1972) and Herbert Robbins (1915–2001)
Preface to the first edition of *What is mathematics?*

Interested students of mathematics, who seek insight into the "essence of the discipline", and who read more widely with a view to discovering what the subject is really about, may emerge with the justifiable impression of serious mathematics as an austere, but distant mountain range – accessible only to those who devote their lives to its exploration. And they may conclude that the beginner can only appreciate its rough outline through a haze of unbridgeable distance. The best popularisers sometimes manage to convey more than this – including hints of the human story behind recent developments, and the way different branches and results interact in unexpected ways; but the essence of mathematics still tends to remain elusive, and the picture they paint is inevitably a broad brush substitute for the detail of living mathematics.

This collection takes a different approach. We start out by observing that mathematics is not a fixed entity – as one might unconsciously infer from the metaphor of an "austere mountain range". Mathematics is a *mental universe*, a work-in-progress in our collective imagination, which grows dramatically over time, and whose eventual extent would seem to be unconstrained – without any obvious limits. This boundlessness also works in reverse, when applied to small details: features which we thought we had understood are repeatedly filled in, or reinterpreted, in new ways to reveal finer and finer micro-structures.

Hence whatever the essence of the discipline may be, it is clearly not something which can only be accessed through the complete exploration of some fixed corpus of knowledge. Rather the essential character of mathematics seems to be related to

- the kind of material that counts as mathematical,
- the way this material is addressed,
- the changes in perspective that occur as our understanding grows and deepens, and
- the unexpected connections that regularly emerge between separate strands and layers.

There are a number of books giving excellent *general* advice to prospective students about how university mathematics differs from school mathematics. In contrast, this collection – which we hope will be enjoyed by interested high school students and their teachers, by undergraduates and postgraduates, and by many others – is more like a messy workshop than a polished exposition. Here the reader is asked to tackle a sequence of problems, to reflect on what they discover, and mostly to draw their own conclusions (though some key messages are explicitly discussed in the text, or in the solutions at the end of each chapter). This attempt to engage the reader as an active participant along the way is inevitably untidy – and may sometimes prove frustrating. In particular, whereas a polished exposition would break up the text with eye-catching diagrams, an untidy workshop will usually leave the reader to draw their own figures as an essential part of the struggle. This temporary untidiness and frustration is an integral part of "the essence" that we seek to capture – provided it leads to occasional glimpses of the power, and the elegance of mathematics.

Young children and students of all ages regularly experience the power, the economy, the beauty, and the elegance of mathematics and of mathematical thinking *on a small scale*, through struggling with certain elementary results and problems (or groups of problems). For example, one of the problems we have included in Chapter 3 was mentioned explicitly in an interview[1] with the leading Russian mathematician Vladimir Arnold (1937–2010):

> **Interviewer**: *Please tell us a little bit about your early education. Were you already interested in mathematics as a child?*
>
> **Arnold**: [...] *The first real mathematical experience I had was when our schoolteacher I.V. Morotzkin gave us the following problem* [VA then formulated Problem **89** in Chapter 3].
>
> *I spent a whole day thinking on this oldie, and the solution (based on what are now called scaling arguments, dimensional analysis, or toric variety theory, depending on your taste) came as a revelation.*

[1] *Notices of the AMS,* vol 44, no. 4.

> *The feeling of discovery that I had then (1949) was exactly the same as in all the subsequent much more serious problems – be it the discovery of the relation between algebraic geometry of real plane curves and four-dimensional topology (1970), or between singularities of caustics and of wave fronts and simple Lie algebras and Coxeter groups (1972). It is the greed to experience such a wonderful feeling more and more times that was, and still is, my main motivation in mathematics.*

This suggests that school mathematics need not be seen solely as an extended apprenticeship, which is somehow different from the craft of mathematics itself. Maybe some aspects of elementary mathematics can be experienced *as if they were a part of mathematics proper*, in which case suitably chosen elementary material, addressed in the appropriate spirit, might serve as a microcosm, or mini-universe, in which many features of the larger mathematical cosmos can be directly, and faithfully experienced by a relative novice (at least to some extent).

This collection of problems (and solutions) is an attempt to embody this idea in a form that might offer students, teachers, and interested readers a glimpse of "the essence of mathematics" – where this insight is experienced, not vicariously through the authors' elegant prose, or broad-brush descriptions, but **through the reader's own engagement with carefully chosen, accessible problems from elementary mathematics**.

Our understanding of the human body and how it works owes much to those (such as the ancient Greeks from 500 BC to Galen in the 2nd century AD, and much later Vesalius in the 16th century AD), who went beyond merely writing *about* such things in high-sounding prose, and who got their hands dirty by procuring cadavers, and cutting them up in order to see things from the inside – while asking themselves all the time how the different parts of the body were connected, and what function they served. In a similar way, the European discovery of the New World in the 15th century, and the confirmation that the Earth can be circumnavigated, depended on those who dared to set sail into uncharted waters and to keep a careful record of what they found.

The process of trying to understand things *from the inside* is not a deterministic procedure: it depends on a mixture of experience and inspiration, intelligence and inference, error and self-criticism. At any given time, the prevailing view may be incomplete, or misguided. But the underlying approach (of checking current ideas against the reality they purport to describe) is the only way we human beings know that allows us to gradually overcome errors and to gain fresh insight.

Our goal in this book is universal (namely to illustrate the idea that a suitably selected elementary microcosm can capture something of the essence of mathematics): hence the problems have all been chosen because we believe they convey something universal in a relatively elementary setting. But the particular set of problems chosen to illustrate the central goal is personal. So we encourage the reader to engage with these problems and results in the same way that old anatomists engaged with cadavers, or old explorers set out on voyages of discovery – getting their hands dirty while asking questions, such as:

> How do the things we see relate to what we know?
> What does this tell us about the subject of mathematics that we want to understand better?

In recent years schools and teachers in many countries have been under increasing political pressure to concentrate on measurable, short term "improvements". Such pressures have often been linked to central testing, with negative consequences for low scores. This has encouraged teachers to play safe, and to focus on *backward-looking* methods that allow students to produce answers to predictable one-step problems. The effect has been to downgrade the more important challenges which every student should face: namely

- of developing a robust mastery of new, *forward-looking* techniques (such as fractions, proportion, and algebra), and
- of integrating the single steps students have at their disposal into larger, systematic schemes, so that they can begin to tackle and solve simple multi-step problems.

Focusing on short-term goals is incompatible with good mathematics teaching. Learning mathematics is a long game; and teachers and students need the freedom to digress, to look ahead, and to build slowly over time. Teachers at each stage must be free to recognise that their primary responsibility is not just to improve their students' performance on the next test, but to establish a firm platform on which subsequent stages can build.

The pressures referred to above will be recognised in many countries, where well-intentioned, but ill-considered, centrally imposed accountability mechanisms have given rise to short-sighted "reforms". A didactical and pedagogical framework that is consistent with the essence, and the educational value of elementary mathematics cannot be rooted in false alternatives to mathematics (such as numeracy, or mathematical literacy). Nor can it be based on tests measuring cheap success on questions that require only one-step routines. We need a framework that encourages a rich combination of childlike curiosity, persistence, fruitful frustration, and the solid satisfaction of structural sense-making.

A problem sequence such as ours should ideally be distilled and refined over decades. However, the best is sometimes the enemy of the good:

Striving to better,
Oft we mar what's well.
(William Shakespeare, *King Lear*)

Hence, as a mild contribution to this process of rediscovering the essence of elementary mathematics, we risk this collection in its present form. And we encourage interested readers to take up pencil and paper, and to join us on this voyage of discovery through elementary mathematics.

Those who enjoy watching professional football (i.e. soccer) must sometimes marvel at the way experienced players seem to be instinctively aware of the movements of other players, and manage to feed the ball into gaps and spaces *that we mere spectators never even noticed were there.* What we overlook is that the best players practise the art of constantly looking around them, and updating their mental record – "viewing the field of play, with their heads up" – so that when the ball arrives and their eyes have to focus on the ball, their ever-changing mental record keeps updating itself to tell them (sometimes apparently miraculously) where the best tactical options lie. Implementing those tactical options depends in part on endless practice of skills; but practice is only one part of the story. What we encourage readers to develop here is the mathematical equivalent of this habit of "viewing the field of play, with one's head up", so that what is noticed can continue to guide the choice of tactical options when one is subsequently immersed in the thick of calculation.

Ours is a unique discipline, which is so much richer than the predictable routines that dominate many contemporary classrooms and assessments. We hope that **all** readers will find that the experience of struggling with, and savouring, this little collection reveals the occasional fresh and memorable insight into "the essence of mathematics".

We should not worry if students don't know everything,
but only if they know everything badly.
Peter Kapitsa, (1894–1984)
Nobel Prize for Physics 1978

To ask larger questions is to risk getting things wrong.
George Steiner (1929–)

Acknowledgements

Our thanks for suggestions, corrections comments and other contributions go to: Jean Bacon, Ayşe Berkman, Anna Borovik, Raul Cordovil, Serkan Dogan, Gwyneth Gardiner, Dick Hudson, Martin Hyland, Hovhannes Khuderverdyan, Ali Nesin, Martin Richards, Simon Singh, Gunnar Traustason, Ozge Uklem, Yusuf Ulum, and numerous students from the 2014 UKMT Summer School in Apperley Bridge.

About this text

And as this is done, so all
similar problems are done.
Paolo dell'Abbaco (1282–1374)
Trattato d'aritmetica

It is better to solve one problem in five different ways
than to solve five problems in one way.
George Pólya (1887–1985)

If you go on hammering away at a problem,
it seems to get tired, lies down,
and lets you catch it.
Sir William Lawrence Bragg (1890–1971)
Nobel Prize for Physics 1915

Young man,
in mathematics you don't understand things.
You just get used to them.
John von Neumann (1903–1957)

This is not a random collection of nice problems. Each item or problem, and each group of problems, is included for two reasons:

- they constitute good mathematics – mathematics which repays the effort of engaging with it for the first time, or revisiting it (should it already be familiar);

and

- they embody in a distilled form the quintessential ***spirit*** *of elementary* (initially pre-university) *mathematics* in a style which can be actively enjoyed by committed students and teachers in schools and colleges, and by the interested general reader.

Some items exemplify core general methods, which can be used over and over again (as hinted by the dell'Abbaco quotation). Some items require us to take different views of ostensibly the same material (as illustrated by the contrasting Pólya quote). Many items will at first seem elusive; but persistence may sometimes lead to an unexpected reward (in the spirit of the Bragg quote). In other instances, a correct answer may be obtained – yet leave the solver less than fully satisfied (at least in the short term, as illustrated by the von Neumann quote). And some items are of little importance in themselves – except that they force the solver to engage in *a kind of thinking* which is mathematically important.

Almost all of the included items are likely to involve – in some degree – that frustration which characterises all fruitful problem solving (as represented by the Bragg quote, and the William Golding quotation below), where, if we are lucky, a bewildering initial fog of incomprehension is sometimes magically dissipated by the process of struggling intelligently to make sense of things. And since one cannot always expect to succeed, there are bound to be occasions when the fog *fails* to lift. One may then have no choice but to consult the solutions (either because some essential idea or technique is not yet part of one's stock-in-trade, or because one has overlooked some simple connection). The only advice we can give here is: *the longer you can delay looking at the solutions the better*. But these solutions have been included both to help you improve your own efforts, and to show the way when you get truly stuck.

The "essence of mathematics", which we have tried to capture in these problems is mostly implicit, and so is often left for the reader to extract. Occasionally it has seemed appropriate to underline some aspect of a particular problem or its solution. Some comments of this kind have been included in the text that is interspersed between the problems. But in many instances, the comment or observation that needs to be made can only be appreciated **after** readers have struggled to solve a problem for themselves. In such cases, positioning the observation in the main text might risk spilling the beans prematurely. Hence, many important observations are buried away in the solutions, or in the **Notes** which follow many of the solutions. More often still, we have chosen to make no explicit remark, but have simply tried to shape and to group the problems in such a way that the intended message is conveyed silently by the problems themselves.

Roughly speaking, one can distinguish three types of problems: these may be labelled as *Core*, as *Gems*, or as focusing on more general *Cognition*.

1. *Core* problems or ideas encapsulate important mathematical concepts and mathematical knowledge in a relatively mundane way, yet in a manner that is in some way canonical. These have sometimes been included here to emphasise some important aspect, which contemporary treatments may have forgotten.

2. *Gems* constitute some kind of paradigm that all aspiring students of mathematics should encounter at some stage. These are likely to be encountered as fully fresh, or surprising, only once in a lifetime. But they then continue to serve as beacons, or trig points, that help to delineate the mathematical landscape.

3. The third type of problem plays an *auxiliary* role – namely problems which emphasise the importance of basic *cognitive skills* for doing mathematics (for example: instant mental calculation, visualisation of abstract concepts, short-term memory, attention span, etc.)

The items are grouped into chapters – each with a recognisable theme. Later chapters tend to have a higher level of technical demand than earlier chapters; and the sequence is broadly consistent with a rising level of sophistication. However, this is not a didactically organised text. Each problem is listed where it fits most naturally, even if it involves an idea which is not formally introduced until somewhat later. Detailed solutions, together with any comments which would be out of place in the main text, are grouped together at the end of each chapter.

The first few chapters tend to focus on more elementary material – partly to emphasise the hierarchical structure of mathematics, partly as a reminder that the essence of mathematics can be experienced at *all* levels, and partly to offer a gentle introduction to readers who may appreciate something slightly more structured before they tackle selected parts of later chapters. Hence these early chapters include more discursive commentary than later chapters. **Readers who choose to skip these nursery slopes on a first reading may wish to return to them later, and to consider what this relatively elementary material tells us about the essence of mathematics.**

The collection is offered as a *supplement* to the standard school curriculum. Some items could (and perhaps should) be incorporated into any official curriculum. But the collection as a whole is mainly designed for those who have good reason, and the time and inclination, to go beyond the usual institutional constraints, and to begin to explore the broader landscape of elementary mathematics in order to experience real, "free range" mathematics – as opposed to artificially reconstituted, or processed products.

> *It has come to me in a flash! One's intelligence may march about and about a problem, but the solution does not come gradually into view. One moment it is not. The next and it is there.*
>
> William Golding (1911–1993), *Rites of Passage*

I. Mental Skills

Even a superficial glance at history shows ...
great innovators ... did vast amounts of computation
and gained much of their insight in this way.
I deplore the fact that contemporary mathematical education
tends to give students the idea that computation is
demeaning drudgery to be avoided at all costs.
Harold M. Edwards (1936–)
Fermat's Last Theorem

We start our journey in a way that should be accessible to everyone – with a quick romp through important ideas from secondary school mathematics. The content is at times very elementary; but the problems often hint at something more challenging. The items included in this first chapter also highlight selected facts, techniques and ideas. Some of this early material is included to introduce certain ideas and techniques that later chapters will assume to be "known". A few problems appeal to more advanced ideas (such as *complex numbers*), and are included here to indicate that "mental skills" are not restricted to elementary material.

Pencil and paper will be needed, but the items tend to focus on things which a student of mathematics should know by heart, or should learn to see at a glance, or should be able to calculate inside the head. In later problems (e.g. from Problem **18** onwards) the emphasis on *mental skills* should be interpreted as "ways of thinking", rather than being taken to mean that everything should be done in your head. This is especially true where extended calculations or proofs are required.

Some of the items in this chapter (such as Problems **1** and **2**) should be thoroughly familiar, and are included to underline this fact, rather than because we anticipate that they will need much active attention. Most of the early items in this first chapter are either *core* or *auxiliary*. However, there are also some real *gems*, which may even warrant a place in the the standard *core*.

The chapter is largely devoted to underlining the need for mastery of a repertoire of instantly available techniques, that can be used mentally, quickly, and flexibly to analyse less familiar problems at sight. But it also seeks to emphasise *connections*. Hence readers should be prepared to challenge their previous experience, in case it may have led to methods and results being perceived too narrowly.

We repeat the comment made in the section *About this book*. The "essence of mathematics", which is referred to in the title, is largely implicit in the problems, and is there for the reader to extract. There is some discussion of this essence in the text interspersed between the problems. But, to avoid spilling the beans prematurely, and hence spoiling the problems, many important observations are buried away in the solutions, or in the **Notes** which follow many of the solutions.

1.1. Mental arithmetic and algebra

1.1.1 Times tables.

Problem 1 Using only mental arithmetic:

(a) Compute for yourself, and learn by heart, the times tables up to 9×9.

(b) Calculate instantly:

(i) 0.004×0.02 (ii) 0.0008×0.07 (iii) 0.007×0.12
(iv) $1.08 \div 1.2$ (v) $(0.08)^2$ △

Multiplication tables are important for many reasons. They allow us to appreciate directly, at first hand, the efficiency of our miraculous *place value system* – in which representing any number, and implementing any operation, are reduced to a combined mastery of

(i) the arithmetical behaviour of the ten digits 0–9, and

(ii) the index laws for powers of 10.

Fluency in mental and written arithmetic then leaves the mind free to notice, and to appreciate, the deeper patterns and structures which may be lurking just beneath the surface.

1.1.2 Squares, cubes, and powers of 2.

Algebra begins in earnest when we start to calculate with expressions involving *powers*. As one sees in the language we use for *squares* and *cubes* (i.e. 2^{nd} and 3^{rd} powers), these powers were interpreted *geometrically* for hundreds and thousands of years – so that higher powers, beyond the third power, were seen as being somehow unreal (like the 4^{th} dimension). Our uniform algebraic notation covering all powers emerged in the 17^{th} century (with Descartes (1596–1650)). But before one begins to work with *algebraic* powers, one should first aim to achieve complete fluency in working with *numerical* powers.

Problem 2

(a) Compute by mental arithmetic (using pencil only to record results), then learn by heart:

(i) the squares of positive integers: first up to 12^2; then to 31^2

(ii) the cubes of positive integers up to 11^3

(iii) the powers of 2 up to 2^{10}.

(b) How many squares are there: (i) < 1000? (ii) $< 10\,000$? (iii) $< 100\,000$?

(c) How many cubes are there: (i) < 1000? (ii) $< 10\,000$? (iii) $< 1\,000\,000$?

(d) (i) Which powers of 2 are squares? (ii) Which powers of 2 are cubes?

(e) Find the smallest square greater than 1 that is also a cube. Find the next smallest. △

Evaluating powers, and the associated index laws, constitute an example of a *direct* operation. For each *direct* operation, we need to think carefully about the corresponding *inverse* operation – here "extracting roots". In particular, we need to be clear about the distinction between the fact that the equation $x^2 = 4$ has *two* different solutions, while $\sqrt{4}$ has *just one* value (namely 2).

Problem 3

(a) The operation of "squaring" is a ***function***: it takes a single real number x as *input*, and delivers a definite real number x^2 as *output*.

- Every positive number arises as an output ("is the square of something").
- Since $x^2 = (-x)^2$, each output (other than 0) arises from **at least two** different inputs.
- If $a^2 = b^2$, then $0 = a^2 - b^2 = (a-b)(a+b)$, so either $a = b$, or $a = -b$. Hence no two positive inputs have the same square, so each output (other than 0) arises from **exactly two** inputs (one positive and one negative).
- Hence each positive output y corresponds to **just one** positive input, called $\sqrt{y}$.

Find:

(i) $\sqrt{49}$ (ii) $\sqrt{144}$ (iii) $\sqrt{441}$ (iv) $\sqrt{169}$
(v) $\sqrt{196}$ (vi) $\sqrt{961}$ (vii) $\sqrt{96\,100}$

(b) Let $a > 0$ and $b > 0$. Then $\sqrt{ab} > 0$, and $\sqrt{a} \times \sqrt{b} > 0$, so both expressions are positive.

Moreover, they *have the same square, since*

$$(\sqrt{ab})^2 = ab = (\sqrt{a})^2 \cdot (\sqrt{b})^2 = (\sqrt{a} \times \sqrt{b})^2.$$

$\therefore \sqrt{\boldsymbol{a} \times \boldsymbol{b}} = \sqrt{\boldsymbol{a}} \times \sqrt{\boldsymbol{b}}$.

Use this fact to simplify the following:

(i) $\sqrt{8}$ (ii) $\sqrt{12}$ (iii) $\sqrt{50}$
(iv) $\sqrt{147}$ (v) $\sqrt{288}$ (vi) $\sqrt{882}$

(c) [This part requires some written calculation.] Exact expressions involving square roots occur in many parts of elementary mathematics. We focus here on just one example – namely the regular pentagon.

Suppose that a regular pentagon $ABCDE$ has sides of length 1.

(i) Prove that the diagonal AC is parallel to the side ED.
(ii) If AC and BD meet at X, explain why $AXDE$ is a rhombus.
(iii) Prove that triangles ADX and CBX are similar.
(iv) If AC has length x, set up an equation and find the exact value of x.
(v) Find the exact length of BX.
(vi) Prove that triangles ABD and BXA are similar.
(vii) Find the exact values of $\cos 36°$, $\cos 72°$.
(viii) Find the exact values of $\sin 36°$, $\sin 72°$. △

Every calculation with square roots depends on the fact that "$\sqrt{}$ is a function". That is: given $y > 0$,

$\sqrt{y}$ denotes a **single value** – the **positive** number whose square is y.

The equation $x^2 = y$ has **two** roots, namely $x = \pm\sqrt{y}$; however, $\sqrt{y}$ has **just one** value (which is positive).

The mathematics of the regular pentagon is important – and generally neglected. It is included here to underline the way exact expressions involving square roots arise naturally.

In Problem **3**(c), parts (iii) and (vi) require one to identify similar triangles using angles. The fact that "corresponding sides are then proportional" leads to a quadratic equation – and hence to square roots.

Parts (vii) and (viii) illustrate the fact that basic tools, such as

- the trigonometric identity $\cos^2\theta + \sin^2\theta = 1$,
- the *Cosine Rule*, and
- the *Sine Rule*

should be part of one's stock-in-trade. Notice that the exact values for

$$\cos 36^\circ,\ \cos 72^\circ,\ \sin 36^\circ,\ \text{and}\ \sin 72^\circ$$

also determine the exact values of

$$\sin 54^\circ = \cos 36^\circ,\ \sin 18^\circ = \cos 72^\circ,\ \cos 54^\circ = \sin 36^\circ,\quad \text{and}\quad \cos 18^\circ = \sin 72^\circ.$$

1.1.3 Primes

Problem 4

(a) Factorise 12 345 as a product of primes.

(b) Using only mental arithmetic, make a list of all prime numbers up to 100.

(c)(i) Find a prime number which is one less than a square.

(ii) Find another such prime. △

There are 4 prime numbers less than 10; 25 prime numbers less than 100; and 168 prime numbers less than 1000.

Problem **4**(c) is included to emphasise a frequently neglected message:

> Words and images are part of the way we communicate.
> But most of us cannot *calculate* with words and images.

To make use of mathematics, we must routinely translate *words* into *symbols*. For example, unknown numbers need to be represented by symbols, and points in a geometric diagram need to be properly labelled, before we can begin to calculate, and to reason, effectively.

1.1.4 Common factors and common multiples

To add two fractions we need to find a common multiple, or the LCM, of the two given denominators. To cancel fractions, or to simplify ratios, we need to be able to spot common factors and to find HCFs. Two positive integers a, b which have no (positive) common factors other than 1 (that is, with $HCF(a,b) = 1$) are said to be *relatively prime*, or *coprime*.

Problem 5 [This problem requires a mixture of serious thought and written proof.]

(a) I choose six integers between 10 and 19 (inclusive).

(i) Prove that some pair of integers among my chosen six must be relatively prime.

(ii) Is it also true that some pair must have a common factor?

(b) I choose six integers in the nineties (from 90–99 inclusive).

(i) Prove that some pair among my chosen integers must be relatively prime.

(ii) Is it also true that some pair must have a common factor?

(c) I choose $n+1$ integers from a run of $2n$ consecutive integers.

(i) Prove that some pair among the chosen integers must be relatively prime.

(ii) Is it also true that some pair must have a common factor? △

1.1.5 The Euclidean algorithm

School mathematics gives the impression that to find the HCF of two integers m and n, one must first obtain the *prime power factorisations* of m and of n, and can then extract the HCF from these two expressions. This is fine for beginners. But arithmetic involves unexpected subtleties. It turns out that, as the numbers get larger, factorising integers quickly becomes extremely difficult – a difficulty that is exploited in modern encryption systems. (The limitations of any method that depends on first finding the prime power factorisation of an integer should have become clear in Problem **4**(b), where it is all too easy to imagine that 91 is prime, and in Problem **4**(c)(ii), where students regularly think that 143, or that 323 are prime.)

Hence we would like to have a simple way of finding the HCF of two integers without having to *factorise* each of them first. That is what the Euclidean algorithm provides. We will look at this in more detail later. Meanwhile here is a first taste.

Problem 6

(a)(i) Explain why any integer that is a factor (or a divisor) of both m and n must also be a factor of their difference $m-n$, and of their sum $m+n$.

(ii) Prove that

$$HCF(m,n) = HCF(m-n,n).$$

(iii) Use this to calculate in your head $HCF(1001, 91)$ without factorising either number.

(b)(i) Prove that: $HCF(m, m+1) = 1$.

(ii) Find $HCF(m, 2m+1)$.

(iii) Find $HCF(m^2+1, m-1)$. △

1.1.6 Fractions and ratio

Problem 7 Which is bigger: 17% of nineteen million, or 19% of seventeen million? △

Problem 8

(a) Evaluate

$$\left(1+\frac{1}{2}\right)\left(1+\frac{1}{3}\right)\left(1+\frac{1}{4}\right)\left(1+\frac{1}{5}\right).$$

(b) Evaluate

$$\sqrt{1+\frac{1}{2}} \times \sqrt{1+\frac{1}{3}} \times \sqrt{1+\frac{1}{4}} \times \sqrt{1+\frac{1}{5}} \times \sqrt{1+\frac{1}{6}} \times \sqrt{1+\frac{1}{7}}.$$

(c) We write the product "$4 \times 3 \times 2 \times 1$" as "4!" (and we read this as "4 *factorial*"). Using only pencil and paper, how quickly can you work out the number of weeks in 10! seconds? △

Problem 9 The "DIN A" series of paper sizes is determined by two conditions. The basic requirement is that all the DIN A rectangles are *similar*; the second condition is that when we fold a given size exactly in half, we get the next smaller size. Hence

- a sheet of paper of size A3 folds in half to give a sheet of size A4 – which is *similar* to A3; and
- a sheet of size A4 folds in half to give a sheet of size A5; etc..

(a) Find the constant ratio

$$r = \text{"(longer side length) : (shorter side length)"}$$

for all DIN A paper sizes.

(b)(i) To enlarge A4 size to A3 size (e.g. on a photocopier), each length is enlarged by a factor of r. What is the "enlargement factor" to get from A3 size back to A4 size?

(ii) To "enlarge" A4 size to A5 size (e.g. on a photocopier), each length is "enlarged" by a factor of $\frac{1}{r}$. What is the enlargement factor to get from A5 size back to A4 size? △

Problem 10

(a) In a sale which offers "15% discount on all marked prices" I buy three articles: a pair of trainers priced at £57.74, a T-shirt priced at £17.28, and a yo-yo priced at £4.98. Using only mental arithmetic, work out how much I should expect to pay altogether.

(b) Some retailers display prices without adding VAT – or "sales tax" – at 20% (because their main customers need to know the pre-VAT price). Suppose the prices in part (a) are the prices before adding VAT. Each price then needs to be adjusted in two ways – adding VAT and subtracting the discount. Should I add the VAT first and then work out the discount? Or should I apply the discount first and then add the VAT?

(c) Suppose the discount in part (b) is no longer 15%. What level of discount would exactly cancel out the addition of VAT at 20%? △

Problem 11

(a) Using only mental arithmetic:

(i) Determine which is bigger:

$$\frac{1}{2}+\frac{1}{5} \quad \text{or} \quad \frac{1}{3}+\frac{1}{4}?$$

(ii) How is this question related to the observation that $10 < 12$?

(b) [This part will require some written calculation and analysis.]

(i) For positive real numbers x, compare

$$\frac{1}{x+2}+\frac{1}{x+5} \quad \text{and} \quad \frac{1}{x+3}+\frac{1}{x+4}.$$

(ii) What happens in part (i) if x is negative? △

1.1.7 Surds

Problem 12

(a) Expand and simplify in your head:

(i) $(\sqrt{2}+1)^2$ (ii) $(\sqrt{2}-1)^2$ (iii) $(1+\sqrt{2})^3$

(b) Simplify:

(i) $\sqrt{10+4\sqrt{6}}$ (ii) $\sqrt{5+2\sqrt{6}}$

(iii) $\sqrt{\frac{3+\sqrt{5}}{2}}$ (iv) $\sqrt{10-2\sqrt{5}}$ △

The expressions which occur in exercises to develop fluency in working with surds often appear arbitrary. But they may not be. The arithmetic of surds arises naturally: for example, some of the expressions in the previous problem have already featured in Problem **3**(c). In particular, surds will feature whenever Pythagoras' Theorem is used to calculate lengths in geometry, or when a proportion arising from similar triangles requires us to solve a quadratic equation. So surd arithmetic is important. For example:

- A regular octagon with side length 1 can be surrounded by a square of side $\sqrt{2}+1$ (which is also the diameter of its incircle); so the area of the regular octagon equals $(\sqrt{2}+1)^2-1$ (the square minus the four corners).
- $\sqrt{2}-1$ features repeatedly in the attempt to apply the Euclidean algorithm, or *anthyphairesis*, to express $\sqrt{2}$ as a "continued fraction".
- $\sqrt{10-2\sqrt{5}}$ may look like an arbitrary, uninteresting repeated surd, but is in fact very interesting, and has already featured as $4\sin 36^\circ$ in Problem **3**(c).
- One of the simplest ruler and compasses constructions for a regular pentagon $ABCDE$ (see Problem **185**) starts with a circle of radius 2, centre O, and a point A on the circle, and in three steps constructs the next point B on the circle, where $\underline{AB}$ is an edge of the inscribed regular pentagon, and

$$\underline{AB} = \sqrt{10-2\sqrt{5}}.$$

1.2. Direct and inverse procedures

We all learn to calculate – with numbers, with symbols, with functions, etc. But we may not notice that most calculating procedures come *in pairs*.

- First we learn a *direct*, deterministic, handle-turning technique, where answers are easy to churn out (as with addition, or multiplication, or working out powers, or multiplying brackets in algebra, or differentiating).
- Then we try to work backwards, or to "undo" this direct operation (as with subtraction, or division, or finding roots, or factorising, or integrating). This *inverse* procedure requires one to be completely fluent in the corresponding *direct* procedure; but it is much more demanding, in that one has to juggle possibilities as one goes, in order to home in on the required answer.

To master *inverse* procedures requires a surprising amount of time and effort. And because they are harder to master, they can easily get neglected. Even where they receive a lot of time, there are aspects of *inverse* procedures which tend to go unnoticed.

Problem 13 In how many different ways can the missing digits in this short multiplication be completed?

$$\begin{array}{r} \square\,6 \\ \times \quad \square \\ \hline \square\,2\;8 \\ \hline \end{array}$$

△

One would like students not only to master the *direct* operation of multiplying digits effectively, but also to notice that the *inverse* procedure of

> "identifying the multiples of a given integer that give rise to a specified output"

depends on

> the HCF of the *multiplier* and the *base* (10) of the numeral system.

- Multiplying by 1, 3, 7, or 9 induces a *one-to-one mapping* on the set of ten digits 0–9; so an inverse problem such as "$7 \times \square$ ends in 6" has just one digit-solution.
- Multiplying by 2, 4, 6, or 8 induces a *two-to-one mapping* onto the set of even digits (multiples of 2); so an inverse problem such as "$6 \times \square$ ends in 4" has two digit-solutions, and an inverse problem such as "$6 \times \square$ ends in 3" has no digit-solutions.
- Multiplying by 5 induces a *five-to-one mapping* onto the multiples (0 and 5) of 5, so an inverse problem such as "$5 \times \square$ ends in 0" has five digit-solutions and an inverse problem such as "$5 \times \square$ ends in 3" has no digit-solutions at all.

- Multiplying by 0 induces a *ten-to-one mapping* onto the multiples of 0 (namely 0); so an inverse problem such as "$0 \times \square$ ends in 0" has ten digit-solutions and an inverse problem such as "$0 \times \square$ ends in 3 (or any digit other than 0)" has no digit-solutions at all.

The next problem shows – in a very simple setting – how elusive inverse problems can be. Here, instead of being asked to perform a *direct* calculation, the rules and the answer are given, and we are simply asked to invent a calculation that gives the specified output.

Problem 14

(a) In the "24 game" you are given four numbers. Your job is to use each number once, and to combine the four numbers using any three of the four basic arithmetical operations – using the same operation more than once if you wish, and as many brackets as you like (but never concatenating different numbers, such as "3" and "4" to make "34"). If the given numbers are 3, 3, 4, 4, then one immediately sees $3 \times 4 + 3 \times 4 = 24$. With 3, 3, 5, 5 it may take a little longer, but is still fairly straightforward. However, you may find it more challenging to make 24 in this way:

  (i) using the four numbers 3, 3, 6, 6
  (ii) using the four numbers 3, 3, 7, 7
  (iii) using the four numbers 3, 3, 8, 8.

(b) Suppose we restrict the numbers to be used each time to "four 4s" $(4, 4, 4, 4)$, and change the goal from "make 24", to "make each answer from 0–10 using exactly four 4s".

  (i) Which of the numbers 0–10 cannot be made?
  (ii) What if one is allowed to use squaring and square roots as well as the four basic operations? What is the first inaccessible integer? △

Calculating by turning the handle deterministically (as with addition, or multiplication, or multiplying out brackets, or differentiating) is a valuable skill. But such *direct* procedures are usually only the beginning. Using mathematics and solving problems generally depend on the corresponding *inverse* procedures – where a certain amount of juggling and insight is needed in order to work backwards (as with subtraction, or division, or factorisation, or integration). For example, in applications of calculus, the main challenge is to solve *differential equations* (an *inverse* problem) rather than to differentiate known functions.

Problem **14** captures the spirit of this idea in the simplest possible context of arithmetic: the required answer is given, and we have to find how (or whether) that answer can be generated. We will meet more interesting examples of this kind throughout the rest of the collection.

1.2.1 Factorisation

Problem 15

(a)(i) Expand $(a+b)^2$ and $(a+b)^3$.

(ii) Without doing any more work, write out the expanded forms of $(a-b)^2$ and $(a-b)^3$.

(b) Factorise (i) x^2+2x+1 (ii) x^4-2x^2+1 (iii) $x^6-3x^4+3x^2-1$.

(c)(i) Expand $(a-b)(a+b)$.

(ii) Use (c)(i) and (a)(i) to write down (with no extra work) the expanded form of

$$(a-b-c)(a+b+c)$$

and of

$$(a-b+c)(a+b-c).$$

(d) Factorise $3x^2+2x-1$. △

1.3. Structural arithmetic

Whenever the answer to a question turns out to be unexpectedly nice, one should ask oneself whether this is an accident, or whether there is some explanation which should perhaps have led one to expect such a result. For example:

- Exactly 25 of the integers up to 100 are prime numbers – and 25 is exactly **one quarter** of 100. This is certainly a beautifully memorable fact. But it is a numerical fluke, with no hidden mathematical explanation.
- 11 and 101 are prime numbers. Is this perhaps a way of generating lots of prime numbers:

$$11, 101, 1001, 10\,001, 100\,001, \ldots?$$

  It may at first be tempting to think so – until, that is, you remember what you found in Problem **6**(a)(iii).

Problem 16 Write out the first 12 or so powers of 4:

$$4,\ 16,\ 64,\ 256,\ 1024,\ 4096,\ 16\,384,\ 65\,536,\ \ldots$$

Now create two sequences:

the sequence of **final** digits: 4, 6, 4, 6, 4, 6, ...
the sequence of **leading** digits: 4, 1, 6, 2, 1, 4, 1, 6, ...

Both sequences seem to consist of a single "block", which repeats over and over for ever.

(a) How long is the apparent repeating block for the first sequence? How long is the apparent repeating block for the second sequence?

(b) It may not be immediately clear whether either of these sequences really repeats forever. Nor may it be clear whether the two sequences are alike, or whether one is quite different from the other. Can you give a simple proof that one of these sequences *recurs*, that is, repeats forever?

(c) Can you explain why the other sequence seems to recur, and decide whether it really does recur forever? △

Problem 17 The 4 by 4 "multiplication table" below is completely familiar.

1	2	3	4
2	4	6	8
3	6	9	12
4	8	12	16

What is the total of all the numbers in the 4 by 4 square? How should one write this answer in a way that makes the total obvious? △

1.4. Pythagoras' Theorem

From here on the idea of "mental skills" tends to refer to *ways of thinking* rather than to doing everything in your head.

Pythagoras' Theorem is one of the first truly surprising results in school mathematics: it is hard to see why anyone would think of "adding the squares of the two shorter sides". Despite the apparent attribution to a named person (Pythagoras), the origin of the theorem, and its proof, are unclear. There certainly was someone called Pythagoras (around 500 BC). But the main ancient references to him were written many hundreds of years after he died, and are not very reliable. The truth is that we know very little about him, or his theorem. The proof in Problem **18** below appeared in Book I of Euclid's thirteen books of *Elements* (written around 300 BC – two hundred years after Pythagoras). Much that is said (wrongly) to stem from Pythagoras is attributed in some sources to the *Pythagoreans* – a loose term which refers to any philosopher in what is seen as a tradition going back to Pythagoras.

(This is a bit like interpreting anything called Christian in the last 2000 years as stemming directly from Christ himself.)

Clay tablets from around 1700 BC suggest that some Babylonians must have known "Pythagoras' Theorem"; and it is hard to see how one could know the result without having some kind of justification. But we have no evidence of either a clear statement, or a proof, at that time. There are also Chinese texts that refer to Pythagoras' Theorem (or as they call it, "Gougu"), which are thought to have originated BC – though the earliest surviving edition is from the 13^{th} century AD. There is even an interesting little book by Frank Swetz, with the tongue-in-cheek title *Was Pythagoras Chinese?*.

The history may be confused, but the result – and its Euclidean proof – embodies something of the surprise and elegance of the very best mathematics. The Euclidean proof is included here partly because it is one that can, and should, be remembered (or rather, reconstructed – once one realises that there is really only one possible way to split the "square on the hypotenuse" in the required way). But, as we shall see, the result also links to *exact mental calculation* with surds, to trigonometry, to the familiar mnemonic "CAST", to the idea of a "converse", to sums of two squares, and to Pythagorean triples.

1.4.1 Pythagoras' Theorem, trig for special angles, and CAST

Problem 18 (Pythagoras' Theorem) Let $\triangle ABC$ be a right angled triangle, with a right angle at C. Draw the squares $ACQP$, $CBSR$, and $BAUT$ on the three sides, external to $\triangle ABC$. Use the resulting diagram to prove *in your head* that the square $BAUT$ on BA is equal to the sum of the other two squares by:

- drawing the line through C perpendicular to AB, to meet AB at X and UT at Y
- observing that PA is parallel to QCB, so that $\triangle ACP$ (half of the square $ACQP$, with base AP and perpendicular height AC) is equal in area to $\triangle ABP$ (with base AP and the same perpendicular height)
- noting that $\triangle ABP$ is SAS-congruent to $\triangle AUC$, and that $\triangle AUC$ is equal in area to $\triangle AUX$ (half of rectangle $AUYX$, with base AU and height AX).
- whence $ACQP$ is equal in area to rectangle $AUYX$
- similarly $BCRS$ is equal in area to $BTYX$. △

The proof in Problem **18** is the proof to be found in Euclid's *Elements* Book 1, Proposition 47. Unlike many proofs,

- it is clear what the proof depends on (namely SAS triangle congruence, and the area of a triangle), and
- it reveals exactly **how** the square on the hypotenuse AB divides into two summands – one equal to the square on AC and one equal to the square on BC.

Problem 19

(a) Use Pythagoras' Theorem in a square $ABCD$ of side 1 to show that the diagonal AC has length $\sqrt{2}$. Use this to work out *in your head* the exact values of $\sin 45°$, $\cos 45°$, $\tan 45°$.

(b) In an equilateral triangle $\triangle ABC$ with sides of length 2, join A to the midpoint M of the base BC. Apply Pythagoras' Theorem to find AM. Hence work out *in your head* the exact values of $\sin 30°$, $\cos 30°$, $\tan 30°$, $\sin 60°$, $\cos 60°$, $\tan 60°$.

(c)(i) On the unit circle with centre at the origin $O : (0, 0)$, mark the point P so that P lies in the first quadrant, and so that OP makes an angle θ with the positive x-axis (measured anticlockwise from the positive x-axis). Explain why P has coordinates $(\cos\theta, \sin\theta)$.

(ii) Extend the definitions of $\cos\theta$ and $\sin\theta$ to apply to angles beyond the first quadrant, so that for any point P on the unit circle, where OP makes an angle θ measured anticlockwise from the positive x-axis, the coordinates of P are $(\cos\theta, \sin\theta)$. Check that the resulting functions sin and cos satisfy:

∗ sin and cos are both positive in the first quadrant,
∗ sin is positive and cos is negative in the second quadrant,
∗ sin and cos are both negative in the third quadrant, and
∗ sin is negative and cos is positive in the fourth quadrant.

(iii) Use (a), (b) to calculate the exact values of $\cos 315°$, $\sin 225°$, $\tan 210°$, $\cos 120°$, $\sin 960°$, $\tan(-135°)$.

(d) Given a circle of radius 1, work out the exact area of a regular n-gon **inscribed in** the circle:

(i) when $n = 3$ (ii) when $n = 4$ (iii) when $n = 6$
(iv) when $n = 8$ (v) when $n = 12$.

(e) Given a circle of radius 1, work out the area of a regular n-gon **circumscribed around** the circle:

(i) when $n = 3$ (ii) when $n = 4$ (iii) when $n = 6$
(iv) when $n = 8$ (v) when $n = 12$. △

Knowing the exact values of sin, cos and tan for the special angles $0°$, $30°$, $45°$, $60°$, $90°$ is like knowing one's tables. In particular, it allows one to evaluate trigonometric functions mentally for related angles in all four quadrants (using the CAST mnemonic – C being in the SE of the unit circle, A in the NE quadrant, S in the NW quadrant, and T in the SW quadrant – to remind us which functions are positive in each quadrant). These special angles arise over and over again in connection with equilateral triangles, squares, regular hexagons, regular octagons, regular dodecagons, etc., where one can use what one knows to calculate *exactly* in geometry.

Problem 20

(a) Use Pythagoras' Theorem to calculate the exact length of the diagonal AC in a square $ABCD$ of side length 2.

(b) Let X be the centre of the square $ABCD$ in part (a). Draw lines through X parallel to the sides of $ABCD$ and so divide the large square into four smaller squares, each of side 1. Find the length of the diagonals AX and XC.

(c) Compare your answers to parts (a) and (b) and your answer to Problem **3**(b)(i). △

Pythagoras' Theorem holds the key to calculating exact distances in the plane. To calculate distances on the Earth's surface one needs a version of Pythagoras for "right angled triangles" on the sphere. We address this in Chapter 5.

1.4.2 Converses and Pythagoras' Theorem

Each mathematical statement of the form

> "if ... (Hypothesis H),
> then ... (Consequence C)"

has a *converse* statement – namely

> "if C,
> then H".

If the first statement is true, there is no *a priori* reason to expect its converse to be true. For example, part (c) of Problem **25** below proves that

> "if an integer has the form $4k + 3$,
> then it cannot be written as the sum of two squares".

However, the converse of this statement

> "if an integer cannot be written as a sum of two squares,
> then it has the form $4k + 3$"

is false – since 6 cannot be written as the sum of two squares.

Despite this counterexample, whenever we prove a standard result, it makes sense to ask whether the converse is also true. For example,

> "if $PQRS$ is a parallelogram, then opposite angles are equal: $\angle P = \angle R$, and $\angle Q = \angle S$" (see Problem **157**(ii)).

However you may not have considered the truth (or otherwise) of the converse statement:

> If $ABCD$ is a quadrilateral in which opposite angles are equal ($\angle A = \angle C$ and $\angle B = \angle D$), is it true that $ABCD$ has to be a parallelogram?

The next problem invites you to prove the converse of Pythagoras' Theorem. You should not use the Cosine Rule, since this is a generalisation of both Pythagoras' Theorem and its converse.

Problem 21 Let ABC be a triangle. We use the standard labelling convention, whereby the side BC opposite A has length a, the side CA opposite B has length b, and the side AB opposite C has length c.

Prove that, if $c^2 = a^2 + b^2$, then $\angle BCA$ is a right angle. △

1.4.3 Pythagorean triples

The simplest example of a right angled triangle with integer length sides is given by the familiar triple 3, 4, 5:

$$3^2 + 4^2 = 5^2.$$

Any such integer triple is called a *Pythagorean triple*.

The classification of *all* Pythagorean triples is a delightful piece of elementary number theory, which is included in this chapter both because the result deserves to be memorised, and because (like Pythagoras' Theorem itself) the proof only requires one to juggle a few simple ideas that should be part of one's armoury.

Pythagorean triples arise in many contexts (e.g. see the text after Problem **180**). The classification given here shows that Pythagorean triples form a family depending on *three* parameters p, q, s (in which s is simply a "scaling" parameter, so the most important parameters are p, q). As a warm-up we consider two "one-parameter subfamilies" related to the triple 3, 4, 5.

Problem 22 Suppose $a^2 + b^2 = c^2$ and that b, c are consecutive integers.

(a) Prove that a must be odd – so we can write it as $a = 2m + 1$ for some integer m.

(b) Prove that c must be odd – so we can write it as $c = 2n + 1$ for some integer n. Find an expression for n in terms of m. △

Problem **22** reveals the triple $(3, 4, 5)$ as the first instance ($m = 1$) of a one-parameter infinite family of triples, which continues

$$(5, 12, 13)\ (m = 2),\ (7, 24, 25)\ (m = 3),\ (9, 40, 41)\ (m = 4), \ldots,$$

whose general term is

$$(2m + 1,\ 2m(m + 1),\ 2m(m + 1) + 1).$$

The triple $(3, 4, 5)$ is also the first member of a quite different "one-parameter infinite family" of triples, which continues

$$(6, 8, 10),\ (9, 12, 15),\ \ldots.$$

Here the triples are scaled-up versions of the first triple $(3, 4, 5)$.

In general, common factors simply get in the way:

> If $a^2 + b^2 = c^2$ and $HCF(a, b) = s$, then s^2 divides $a^2 + b^2$, and $a^2 + b^2 = c^2$; so s divides c.
> And if $a^2 + b^2 = c^2$ and $HCF(b, c) = s$, then s^2 divides $c^2 - b^2 = a^2$, so s divides a.

Hence a typical Pythagorean triple has the form (sa, sb, sc) for some scale factor s, where (a, b, c) is a triple of integers, no two of which have a common factor: any such triple is said to be *primitive* (that is, basic – like prime numbers). Every Pythagorean triple is an integer multiple of some *primitive Pythagorean triple*. The next problem invites you to find a simple formula for all primitive Pythagorean triples.

Problem 23 Let (a, b, c) be a primitive Pythagorean triple.

(a) Show that a and b have opposite parity (i.e. one is odd, the other even) – so we may assume that a is odd and b is even.

(b) Show that

$$\left(\frac{b}{2}\right)^2 = \left(\frac{c-a}{2}\right)\left(\frac{c+a}{2}\right),$$

where

$$HCF\left(\frac{c-a}{2}, \frac{c+a}{2}\right) = 1$$

and $\frac{c-a}{2}, \frac{c+a}{2}$ have opposite parity.

(c) Conclude that

$$\frac{c+a}{2} = p^2 \text{ and } \frac{c-a}{2} = q^2,$$

where $HCF(p,q) = 1$ and p and q have opposite parity, so that $c = p^2+q^2$, $a = p^2 - q^2$, $b = 2pq$.

(d) Check that any pair p, q having opposite parity and with $HCF(p,q) = 1$ gives rise to a primitive Pythagorean triple

$$c = p^2 + q^2, \quad a = p^2 - q^2, \quad b = 2pq$$

satisfying $a^2 + b^2 = c^2$. △

Problem 24 The three integers $a = 3$, $b = 4$, $c = 5$ in the Pythagorean triple $(3,4,5)$ form an *arithmetic progression*: that is, $c - b = b - a$. Find all Pythagorean triples (a,b,c) which form an arithmetic progression – that is, for which $c - b = b - a$. △

1.4.4 Sums of two squares

The classification of Pythagorean triples tells us precisely which **squares** can be written as the sum of two squares. We now turn to the wider question: "Which *integers* are equal to the sum of two squares?"

Problem 25

(a) Which of the prime numbers < 100 can be written as the sum of two squares?

(b) Find an easy way to immediately write $(a^2 + b^2)(c^2 + d^2)$ in the form $(x^2 + y^2)$. (This shows that the set of integers which can be written as the sum of two squares is "closed" under multiplication.)

(c) Prove that no integer (and hence no prime number) of the form $4k + 3$ can be written as the sum of two squares.

(d) The only *even* prime number can clearly be written as a sum of two squares: $2 = 1^2 + 1^2$. Euler (1707–1783) proved that every *odd* prime number of the form $4k + 1$ can be written as the sum of two squares in exactly one way. Find all integers < 100 that can be written as a sum of two squares.

(e) For which integers $N < 100$ is it possible to construct a square of area N, with vertices having integer coordinates? △

In Problem **25** parts (a) and (d) you had to decide which integers < 100 can be written as a sum of two squares as an exercise in mental arithmetic. In part (b) the fact that this set of integers is closed under multiplication turned out to be an application of the arithmetic of *norms* for complex numbers. Part (e) then interpreted sums of two squares geometrically by using Pythagoras' Theorem on the square lattice. These exercises are worth engaging in for their own sake. But it may also be of interest to know that writing an integer as a sum of two squares is a serious mathematical question – and in more than one sense.

Gauss (1777–1855), in his book *Disquisitiones arithmeticae* (1801) gave a complete analysis of when an integer can be represented by a 'quadratic form', such as $x^2 + y^2$ (as in Problem **25**) or $x^2 - 2y^2$ (as in Problem **54**(c) in Chapter 2).

A completely separate question (often attributed to Edward Waring (1736–1798)) concerns which integers can be expressed as a k^{th} power, or as a sum of n such powers. If we restrict to the case $k = 2$ (i.e. squares), then:

- When $n = 2$, Euler (1707–1783) proved that the integers that can be written as a sum of *two squares* are precisely those of the form

$$m^2 \times p_0 \times p_1 \times p_2 \times \cdots \times p_s,$$

  where $p_0 = 1$ or 2, and $p_1 < p_2 < \cdots < p_s$ are odd primes of the form $4l + 1$.

- When $n = 3$, Legendre (1752–1833) and Gauss proved between them that the integers which can be written as a sum of *three squares* are precisely those that are **not** of the form $4^m \times (8l + 7)$.

- When $n = 4$, Lagrange (1736–1813) had previously proved that **every** positive integer can be written as a sum of *four squares*.

1.5. Visualisation

Problem 26 (Pages of a newspaper) I found a (double) sheet from an old newspaper, with pages 14 and 27 next to each other. How many pages were there in the original newspaper? △

Problem 27 (Overlapping squares) A square $ABCD$ of side 2 sits on top of a square $PQRS$ of side 1, with vertex A at the centre O of the small square, side AB cutting the side PQ at the point X, and $\angle AXQ = \theta$.

(a) Calculate the area of the overlapping region.

(b) Replace the two squares in part (a) with two equilateral triangles. Can you find the area of overlap in that case? What if we replace the squares (i.e. regular 4-gons) in part (a) with regular $2n$-gons? △

Problem 28 (A folded triangle) The equilateral triangle $\triangle ABC$ has sides of length 1 cm. D and E are points on the sides AB and AC respectively, such that folding $\triangle ADE$ along DE folds the point A onto A' which lies outside $\triangle ABC$.

What is the total perimeter of the region formed by the three single layered parts of the folded triangle (i.e. excluding the quadrilateral with a folded layer on top)? △

Problem 29 ($\boldsymbol{A+B=C}$) The 3 by 1 rectangle $ADEH$ consists of three adjacent unit squares: $ABGH$, $BCFG$, $CDEF$ left to right, with A in the top left corner. Prove that

$$\angle DAE + \angle DBE = \angle DCE.$$

△

Problem 30 (Dissections)

(a) Joining the midpoints of the edges of an equilateral triangle ABC cuts the triangle into four identical smaller equilateral triangles. Removing one of the three outer small triangles (say AMN, with M on AC) leaves three-quarters of the original shape in the form of an isosceles trapezium $MNBC$. Show how to cut this isosceles trapezium into four congruent pieces.

(b) Joining the midpoints of opposite sides of a square cuts the square into four congruent smaller squares. If we remove one of these squares, we are left with three-quarters of the original square in the form of an L-shape. Show how to cut this L-shape into four congruent pieces. △

Problem 31 (Yin and Yang) The shaded region in Figure 1, shaped like a large comma, is bounded by three semicircles – two of radius 1 and one of radius 2.

Cut each region (the shaded region and the unshaded one) into two 'halves', so that all four parts are congruent (i.e. of identical size and shape, but with possibly different orientations). △

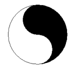


Figure 1: Yin and Yang

In Problem **31** your first thought may have been that this is impossible. However, since the wording indicated that you are expected to succeed, it was clear that you must be missing something – so you tried again. The problem then tests both flexibility of thinking, and powers of visualisation.

1.6. Trigonometry and radians

1.6.1 Sine Rule

School textbooks tend to state the *Sine Rule* for a triangle ABC without worrying why it is true. So they often fail to give the result in its full form:

Theorem If R is the radius of the circumcircle of the triangle ABC, then

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

This full form explains that the three ratios

$$\frac{a}{\sin A}, \quad \frac{b}{\sin B}, \quad \frac{c}{\sin C}$$

are all equal *because they are all equal to the diameter* $2R$ *of the circumcircle* of $\triangle ABC$ – an additional observation which may well suggest how to prove the result (see Problem **32**).

Problem 32 Given any triangle ABC, construct the perpendicular bisectors of the two sides AB and BC. Let these two perpendicular bisectors meet at O.

(a) Explain why $OA = OB = OC$.

(b) Draw the circle with centre O and with radius OA. There are three possibilities:

(i) The centre O lies on one of the sides of triangle ABC.

(ii) The centre O lies inside triangle ABC.

(iii) The centre O lies outside triangle ABC.

Case (i) leads directly to the *Sine Rule* for a right angled triangle ABC (remembering that $\sin 90° = 1$). We address case (ii), and leave case (iii) to the reader.

(ii) Extend the line BO to meet the circle again at the point A'. Explain why $\angle BA'C = \angle BAC = \angle A$, and why $\angle A'CB$ is a right angle. Conclude that

$$\sin A = \frac{BC}{A'B} = \frac{a}{2R},$$

and hence that

$$\frac{a}{\sin A} = 2R \quad \left(= \frac{b}{\sin B} = \frac{c}{\sin C} \right). \qquad \triangle$$

Problem 33 Let $\Delta = \text{area}(\triangle ABC)$.

(a) Prove that

$$\Delta = \frac{1}{2} \cdot ab \cdot \sin C.$$

(b) Prove that $4R\Delta = abc$. $\triangle$

1.6.2 Radians and spherical triangles

There is no God-given unit for measuring distance; different choices of unit give rise to answers that are related by *scaling*. However the situation is different for *angles*. In primary and secondary school we measure *turn* in *degrees* – where a half turn is 180°, a right angle is 90°, and a complete turn is 360°. This angle unit dates from the ancient Babylonians ($\sim$ 2000 BC). We are not sure why they chose 360 units in a full turn, but it seems to be related to the approximate number of days in a year (the time required for the heavens to make a complete rotation in the night sky), and to the fact that they wrote their numbers in "base 60". However the choice is no more objectively mathematical than measuring distance in inches or in centimetres.

After growing up with the idea that angles are measured in degrees, we discover towards the end of secondary school that:

> there is another unit of measure for *angles* – namely **radians**.

It may not at first be clear that this is an entirely natural, God-given unit. The size of, or amount of turn in, an angle at the point A can be captured in an absolute way by drawing a circle of radius r centred at the point A, and measuring the arc length which the angle cuts off on this circle. The angle size (in *radians*) of the angle at A is then defined to be the ratio

$$\frac{\text{arc length}}{\text{radius}}.$$

That is,

size of angle at the point A = arc length cut off on a circle of radius 1, centred at the apex A.

Hence a right angle is of size $\frac{\pi}{2}$ radians; a half turn is equal to π (radians); a full turn is equal to 2π (radians); each angle in an equilateral triangle is equal to $\frac{\pi}{3}$ (radians); the three angles of a triangle have sum π; and the angles of a polygon with n sides have sum $(n-2)\pi$ (see Problem **230** in Chapter 6).

For a while after the introduction of radians we continue to emphasise the word *radians* each time we give the measure of an angle in order to stress that we are no longer using degrees. But this is not really a switch to a new unit: this new way of measuring angles is in some sense objective – so we soon drop all mention of the word "radians" and simply refer to the size of an angle (in radians) as if it were a pure number.

This switch affects the meaning of the familiar *trigonometric functions*. And though we continue to use the same names (sin, cos, tan, etc.), they become slightly different *as functions*, since the inputs are now always assumed to be in radians.

The real payoff for making this change stems from the way it recognizes the connection between *angles* and *circles*. This certainly makes calculating circular arc lengths and areas of sectors easy (an arc with angle θ on a circle of radius r now has length θr; and a circular sector with angle 2θ now has area θr^2). But the main benefit – which one hopes all students appreciate eventually – is that this change of perspective highlights the fundamental link between $\sin x$, $\cos x$, and e^x:

- "$\cos x$" becomes the derivative of $\sin x$
- "$-\sin x$" becomes the derivative of $\cos x$, and
- the three functions are related by the totally unexpected identity

$$e^{i\theta} = \cos\theta + i\sin\theta.$$

The next problem draws attention to a beautiful result which reveals, *in a pre-calculus, pre-complex number setting*, a beautiful consequence of thinking about angles in terms of radians. The goal is to discover a formula for the area of a *spherical triangle* in terms of its angles and π, which links the formula for the circumference of a circle with that for the surface area of a sphere.

Suppose we wish to do geometry on the sphere. There is no problem deciding how to make sense of *points*. But it is less clear what we mean by (*straight*) *lines*, or line segments.

Before making due allowance for the winds and the tides, an airline pilot and a ship's Captain both need to know how to find the *shortest path* joining two given points A, B on a sphere. If the two points both lie on the equator, it is plausible (and correct) that the shortest route is to travel from A to B *along the equator*. If we think of the equator as being in a *horizontal* plane through the centre O of the sphere, then we may notice that we can change the equator into a circle of longitude by rotating the sphere so that the "horizontal" plane (through O) becomes a "vertical" plane (through O). So we may view two points A and B which both lie on the same circle of longitude as lying on a "vertical equator" passing through A, B and the North and South poles: the shortest distance from A to B must therefore lie along that circle of longitude.

If we now rotate the sphere through some other angle, we get a "tilted equator" passing through the images of the (suitably tilted) points A and B: these "tilted equators" are called *great circles*. Each great circle is the intersection of the sphere with a plane through the centre O of the sphere. So

> to find the shortest path from A to B:
>
> - take the plane determined by the points A, B and the centre of the sphere O;
> - find the great circle where this plane cuts the sphere;
> - then follow the arc from A to B along this great circle.

Once we have points and line segments (i.e. arcs of great circles) on the sphere, we can think about *triangles*, and about the *angles* in such a triangle. In a triangle ABC on the sphere, the sides AB and AC are arcs of great circles meeting at A. By rotating the sphere we can imagine A as being at the North pole; so the two sides AB and AC behave just like arcs of two circles of longitude emanating from the North pole. In particular, we can measure the angle between them (this is exactly how we measure *longitude*): the two arcs AB, AC of circles of longitude set off from the North pole A in different horizontal directions before curving southwards, and the angle between them is the angle between these two initial horizontal directions (that is, the angle between the plane determined by O, A, B and the plane determined by O, A, C).

Problem 34 Imagine a triangle ABC on the unit sphere (with radius $r = 1$), with angle α between AB and AC, angle β between BC and BA, and angle γ between CA and CB. You are now in a position to derive the remarkable formula for the area of such a spherical triangle.

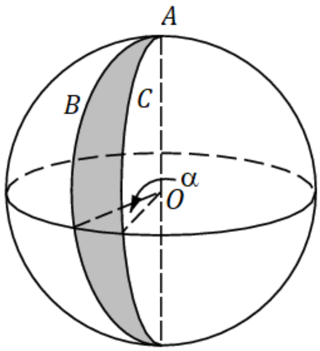


Figure 2: Angles on a sphere

(a) Let the two great circles containing the sides AB and AC meet again at A'. If we imagine A as being at the North pole, then A' will be at the South pole, and the angle between the two great circles at A' will also be α. The slice contained between these two great circles is called a *lune* with angle α.

(i) What fraction of the surface area of the whole sphere is contained in this lune of angle α? Write an expression for the actual area of this lune.

(ii) If the sides AB and AC are extended backwards through A, these backward extensions define another lune with the same angle α, and the same surface area. Write down the total area of these two lunes with angle α.

(b)(i) Repeat part (a) for the two sides BA, BC meeting at the vertex B, to find the total area of the two lunes meeting at B and B' with angle β.

(ii) Do the same for the two sides CA, CB meeting at the vertex C, to find the total area of the two lunes meeting at C and C' with angle γ.

(c)(i) Add up the areas of these six lunes (two with angle α, two with angle β, and two with angle γ). Check that this total includes every part of the sphere at least once.

(ii) Which parts of the sphere have been covered more than once? How many times have you covered the area of the original triangle ABC? And how many times have you covered the area of its sister triangle $A'B'C'$?

(iii) Hence find a formula for the area of the triangle ABC in terms of its angles – α at A, β at B, and γ at C. △

1.6.3 Polar form and sin(A+B)

The next problem is less elementary than most of Chapter 1, but is included here to draw attention to the ease with which the addition formulae in trigonometry can be reconstructed once one knows about the *polar form* representation of a complex number. Those who are as yet unfamiliar with this material may skip the problem – but should perhaps remember the underlying message (namely that, once one *is* familiar with this material, there is no need ever again to get confused about the trig addition formulae).

Problem 35

(a) You may know that any complex number $z = \cos\theta + i\sin\theta$ of modulus 1 (that is, which lies on the unit circle centred at the origin) can be written in the modulus form $z = e^{i\theta}$. Use this fact to reconstruct in your head the trigonometric identities for $\sin(A + B)$ and for $\cos(A + B)$. Use these to derive the identity for $\tan(A + B)$.

(b) By choosing X, Y so that $A = \frac{X+Y}{2}$, and $B = \frac{X-Y}{2}$, use part (a) to reconstruct the standard trigonometric identities for

$$\sin X + \sin Y,\ \sin X - \sin Y,\ \cos X + \cos Y,\ \cos X - \cos Y.$$

(c)(i) Check your answer to (a) for $\sin(A + B)$ by substituting $A = 30°$, and $B = 60°$.

(ii) Check your answer to (b) for $\cos X - \cos Y$ by substituting $X = 60°$, and $Y = 0°$.

(d)(i) If $A + B + C + D = \pi$, prove that

$$\sin A \sin B + \sin C \sin D = \sin(B + C)\sin(B + D).$$

(ii) Given a cyclic quadrilateral $WXYZ$, with $\angle XWY = A$, $\angle WXZ = B$, $\angle YXZ = C$, $\angle WYX = D$, deduce *Ptolemy's Theorem*:

$$WX \times YZ + WZ \times XY = WY \times XZ.$$

△

1.7. Regular polygons and regular polyhedra

Regular polygons have already featured rather often (e.g. in Problems **3**, **12**, **19**, **27**, **28**, **29**). This is a general feature of elementary mathematics; so the neglect of the geometry of regular polygons, and their 3D companions, the regular polyhedra, is all the more unfortunate. We end this first chapter with a first brief look at polygons and polyhedra.

1.7.1 Regular polygons are cyclic

Problem 36 A polygon $ABCDE\cdots$ consists of n vertices A, B, C, D, E, $\ldots$, and n sides AB, BC, CD, DE …which are disjoint except that successive pairs meet at their common endpoint (as when AB, BC meet at B). A polygon is *regular* if any two sides are congruent (or equal), and any two angles are congruent (or equal). Can a regular n-gon $ABCDE\cdots$ always be inscribed in a circle? In other words, does a regular polygon automatically have a "centre", which is equidistant from all n vertices? △

1.7.2 Regular polyhedra

Problem 37 (Wrapping)

(a) You are given a regular tetrahedron with edges of length 2. Is it possible to choose positive real numbers a and b so that an a by b rectangular sheet of paper can be used to "wrap", or cover, the regular tetrahedron without leaving any gaps or overlaps?

(b) Given a cube with edges of length 2, what is the smallest sized rectangle that can be used to wrap the cube in the same way without cutting the paper? (In other words, if we want to completely cover the cube, what is the smallest area of overlap needed? How small a fraction of the paper do we have to waste?) △

Problem 38 (Cross-sections) Can a cross-section of a cube be:

(i) an equilateral triangle?

(ii) a square?

(iii) a polygon with more than six sides?

(iv) a regular hexagon?

(v) a regular pentagon? △

Problem 39 (Shadows) Can one use the Sun's rays to produce a plane shadow of a cube:

(i) in the form of an equilateral triangle?

(ii) in the form of a square?

(iii) in the form of a pentagon?

(iv) in the form of a regular hexagon?

(v) in the form of a polygon with more than six sides? △

The imparting of factual knowledge is for us a secondary consideration. Above all we aim to promote in the reader a correct attitude, a certain discipline of thought, which would appear to be of even more essential importance in mathematics than in other scientific disciplines. ...

General rules which could prescribe in detail the most useful discipline of thought are not known to us. Even if such rules could be formulated, they would not be very useful. Rather than knowing the correct rules of thought theoretically, one must have assimilated them into one's flesh and blood ready for instant and instinctive use. Therefore for the schooling of one's powers of thought only the practice of thinking is really useful.

G. Pólya (1887–1985) and G. Szegő (1895–1985)

1.8. Chapter 1: Comments and solutions

1.

(a) Assuming that the 2×, 3×, 4×, and 5× tables are known, and that one has understood that the order of the factors does not matter, all that remains to be learned is 6×6, 6×7, 6×8, 6×9; 7×7, 7×8, 7×9; 8×8, 8×9; and 9×9.

(b)(i) $(4 \times 10^{-3}) \times (2 \times 10^{-2}) = 8 \times 10^{-5} = 0.00008$

(ii) $(8 \times 10^{-4}) \times (7 \times 10^{-2}) = 56 \times 10^{-6} = 0.000056$

(iii) $(7 \times 10^{-3}) \times (12 \times 10^{-2}) = 84 \times 10^{-5} = 0.00084$

(iv) $1.08 \div 1.2 = 10.8 \div 12 = 108 \div (12 \times 10) = 0.9$

(v) $(8 \times 10^{-2})^2 = 64 \times 10^{-4} = 0.0064$

2.

(a)(i) 1, 4, 9, 16, 25, 36, 49, 64, 81, 100, 121, 144;
169, 196, 225, 256, 289, 324, 361, 400, 441, 484, 529, 576, 625, 676, 729, 784, 841, 900, 961

(ii) 1, 8, 27, 64, 125, 216, 343, 512, 729, 1000, 1331

(iii) $1 = 2^0$, 2, 4, 8, 16, 32, 64, 128, 256, 512, 1024

(b)(i) 31 ($31^2 = 961$, $32^2 = 2^{10} = 1024$)

(ii) 99 ($100^2 = 10^4 = 10\,000$)

(iii) 316 ($310^2 = 96\,100 < 100\,000 < 320^2$; so look more carefully between 310 and 320)

(c) (i) 9 ($9^3 = 729 < 10^3 = 1000$)

(ii) 21 ($20^3 = 8000$, $22^3 = 10\,648$)

(iii) 99 ($100^3 = 10^6 = 1\,000\,000$)

(d) (i) Those powers 2^e of the form 2^{2n} for which the exponent e is a multiple of 2: i.e. $e \equiv 0 \pmod 2$.

(ii) Those powers 2^e of the form 2^{3n} for which the exponent e is a multiple of 3: i.e. $e \equiv 0 \pmod 3$.

(e) $64 = 2^6 = 4^3 = 8^2$. $729 = 3^6 = 9^3 = 27^2$.

3.

(a) (i) 7; (ii) 12; (iii) 21; (iv) 13; (v) 14; (vi) 31; (vii) $10 \times 31 = 310$

(b) (i) $2\sqrt{2}$; (ii) $2\sqrt{3}$; (iii) $5\sqrt{2}$; (iv) $7\sqrt{3}$; (v) $12\sqrt{2}$; (vi) $21\sqrt{2}$

(c) (i) $\angle ABC = 108°$. $\triangle BAC$ is isosceles ($BA = BC$), so $\angle BAC = \angle BCA = 36°$.
$\therefore \angle CAE = \angle BAE - \angle BAC = 72° = 180° - \angle AED$.
So AC is parallel to ED (since corresponding angles add to 180°).

(ii) AX is parallel to ED; similarly DX is parallel to EA. Hence $AXDE$ is a parallelogram, with $EA = ED$.

(iii) The two triangles are both isosceles and $\angle AXD = \angle CXB = 108°$ (vertically opposite angles).
Hence $\angle XAD = \angle XCB = 36°$, and $\angle XDA = \angle XBC = 36°$.

(iv) $AD : CB = DX : BX$, so $x : 1 = 1 : (x-1)$; hence $x^2 - x - 1 = 0$ and $x = \frac{1+\sqrt{5}}{2}$ – the *Golden Ratio*, usually denoted by the Greek letter τ (*tau*), with approximate value 1.6180339887....

(v) $BX = x - 1 = \frac{\sqrt{5}-1}{2}$ $(= \frac{1}{\tau} = \tau - 1)$, with approximate value 0.6180339887....

(vi) We may either check that corresponding angles are equal in pairs $(36°, 72°, 72°)$, or that corresponding sides are in the same ratio $x : 1 = 1 : (x-1)$.

(vii) $\cos 36° = \frac{\sqrt{5}+1}{4}$; $\cos 72° = \frac{\sqrt{5}-1}{4}$ (drop perpendiculars from D to AB and from X to BC; or use the Cosine Rule).

(viii) Use $\sin^2 36° + \cos^2 36° = 1$: $\sin 36° = \frac{\sqrt{10-2\sqrt{5}}}{4}$; $\sin 72° = \frac{\sqrt{10+2\sqrt{5}}}{4}$.

Note: The *Golden Ratio* crops up in many unexpected places (including the regular pentagon, and the *Fibonacci* numbers). Unfortunately much that is written about its ubiquity is pure invention. One of the better popular treatments, that highlights the number's significance, while taking a sober view of spurious claims, is the book *The Golden Ratio* by Mario Livio.

4.

(a) $12\,345 = 5 \times 2469 = 3 \times 5 \times 823$. But is 823 a prime number?
It is easy to check that 823 is not divisible by 2, or 3, or 5, or 7, or 11. The *Square Root Test* (displayed below) tells us that it is only necessary to check four more potential prime factors.

Square Root Test: If $N = a \times b$ with $a \leqslant b$, then $a \times a \leqslant a \times b = N$, so the smaller factor $a \leqslant \sqrt{N}$.

Hence, if N (= 823 say) is not prime, its smallest factor > 1 is at most equal to $\sqrt{N}$ (= $\sqrt{823} < 29$). Checking $a = 13, 17, 19, 23$ shows that the required prime factorisation is $12\,345 = 3 \times 5 \times 823$.

(b) There are 25 (2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61, 67, 71, 73, 79, 83, 89, 97).

Notes:

(i) For small primes, mental arithmetic should suffice. But one should also be aware of the general *Sieve of Eratosthenes* (a Greek polymath from the 3rd century BC). Start with the integers 1–100 arranged in ten columns, and proceed as follows:

1	2	3	4	5	6	7	8	9	10
11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30
31	32	33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48	49	50
61	62	63	64	65	66	67	68	69	70
71	72	73	74	75	76	77	78	79	80
81	82	83	84	85	86	87	88	89	90
91	92	93	94	95	96	97	98	99	100

Delete 1 (which is not a prime: see (ii) below).

Circle the first undeleted integer; remove all other multiples of 2.

Circle the first undeleted integer; remove all other multiples of 3.

Circle the first undeleted integer; remove all other multiples of 5.

Circle the first undeleted integer; remove all other multiples of 7.

All remaining undeleted integers < 100 must be prime (by the *Square Root Test* (see part (a)).

(ii) The multiplicative structure of integers is surprisingly subtle. The first thing to notice is that 1 has a special role, in that it is the *multiplicative identity:* for each integer n, we have $1 \times n = n$. Hence 1 is "multiplicatively neutral" – it has no effect.

The "multiplicative building blocks" for integers are the *prime numbers*: every integer > 1 can be broken down, or *factorised* as a product of prime numbers, in exactly one way. The integer 1 has no proper factors, and has no role to play in breaking down larger integers by *factorisation*. So 1 is **not** a prime.

(If we made the mistake of counting 1 as a prime number, then we would have to make all sorts of silly exceptions – for example, to allow for the fact that $2 = 2 \times 1 = 2 \times 1 \times 1 = \ldots$, so 2 could then be factorised in infinitely many ways.)

(iii) Notice that $91 = 7 \times 13$ is not a prime; so there is exactly one prime in the 90s – namely 97.

How many primes are there in the next run of 10 (from 100–109)?

How many primes are there from 190–199? How many from 200–210?

(c) (i) $3 = 2^2 - 1$.

(ii) Many students struggle with this, and may suggest 143, or 323, or even 63. The problem conceals a (very thinly) disguised message:

One cannot calculate with words.

To make use of mathematics, we must routinely translate *words* into *symbols*. As soon as "one less than a square" is translated into symbols, bells should begin to ring. For you know that $x^2 - 1 = (x-1)(x+1)$, so $x^2 - 1$ can only be prime if the smaller factor $(x-1)$ is equal to 1.

5.

(a) (i) If we try to avoid such a "relatively prime pair", then we must **not** choose any of 11, 13, 17, 19 (since they are prime, and have no multiples in the given range). So we are forced to choose the other six integers: 10, 12, 14, 15, 16, 18 – and there are then exactly two pairs which are relatively prime, namely 14, 15 and 15, 16.

(ii) If we try to avoid such a pair, then we can choose at most one even integer. So we are then forced to choose **all five** available odd integers, and our list will be: "unknown even, 11, 13, 15, 17, 19". If the even integer is chosen to be 14, or 16, then every pair in my list has LCM = 1. So the answer is "No".

(b) (i) If we try to avoid such a pair, then we must **not** choose 97 (the only prime number in the nineties). And we must not choose $95 = 5 \times 19$ (which is relatively prime to all other integers in the given range – except for 90); and we must not choose $91 = 7 \times 13$ (which is relatively prime to all other integers in the given range – except for 98). So we are forced to choose six integers from 90, 92, 93, 94, 96, 98, 99. Whichever integer we then omit leaves a pair which is relatively prime.

(ii) If we try to avoid such a pair, then we can choose at most one even integer. So we are then forced to choose all five available odd integers, and our list will be: "unknown even, 91, 93, 95, 97, 99", and so must include the pair 93, 99 – with common factor 3. So the answer is "Yes".

(c) In parts (a) and (b), the possible integers are limited (in (a) to the "teens", and in (b) to the "nineties"); so it is natural to reach for *ad hoc* arguments as we did above. But in part (c) you know nothing about the numbers chosen.

Note: The question says that "*I* choose", and asks whether "*you*" can be sure. So you have to find **either** a *general* argument that works for any n, **or** a counterexample. And the theme of this chapter indicates that it should not require any extended calculation.

The relevant "general idea" is the *Pigeon Hole Principle* which we may meet in the second part of this collection. So this problem may be viewed as a gentle introduction.

(i) Group the $2n$ consecutive integers

$$a+1, a+2, \ldots, a+2n$$

into n pairs of consecutive integers

$$\{a+1, a+2\}, \{a+3, a+4\}, \ldots, \{a+(2n-1), a+2n\}.$$

- If we choose at most one integer from each pair, then we never get more than n integers.
- So as soon as we choose $n+1$ integers from $2n$ consecutive integers, we are forced to choose **both** integers in some pair k, $k+1$, and this pair of consecutive integers is always relatively prime (see Problem **6**(b)(i)).

(ii) We saw in part (a)(ii) that, if $n = 5$ and the $2n$ integers start at 10, then we can choose six integers (either 11, 13, 14, 15, 17, 19, or 11, 13, 15, 16, 17, 19), and in each case every pair has LCM = 1. So for $n = 5$ the answer is "No" (because there is *at least one case* where one cannot be sure).

However, as soon as n is at least 6, we show that the argument in part (a)(ii) breaks down. As before, if we try to choose a subset in which no pair has a common factor, then we can choose at most one even integer. So we are forced to choose **all** the odd integers. But any run of **at least six** consecutive odd integers includes two multiples of 3. So for $n \geqslant 6$, the answer is "Yes".

6.

(a)(i) Suppose k is a factor of m and n. Then we can write $m = kp$ and $n = kq$ for some integers p, q. Hence $m - n = k(p - q)$, so k is a factor of $m - n$. Also $m + n = k(p + q)$, so k is a factor of $m + n$.

(ii) Any factor of m and n is also a factor of their difference $m - n$; so the set of common factors of m and n is a **subset** of the set of common factors of $m - n$ and n.

And any factor of $m - n$ and n is also a factor of their sum m; so the set of common factors of $m - n$ and n is a **subset** of the set of common factors of m and n.

Hence the two sets of common factors are identical. In particular, the two "highest common factors" are equal.

(iii) Subtract 91 from 1001 ten times to see that

$$HCF(1001, 91) = HCF(1001 - 910, 91) = 91.$$

(b)(i) Subtract m from $m + 1$ once to see that

$$HCF(m+1, m) = HCF(1, m) = 1.$$

(ii) Subtract m from $2m + 1$ twice to see that

$$HCF(2m+1, m) = HCF(m+1, m) = HCF(1, m) = 1.$$

(iii) Subtract $m - 1$ from $m^2 + 1$ "$m + 1$ times" to see that

$$HCF(m^2+1, m-1) = HCF((m^2+1) - (m^2-1), m-1) = HCF(2, m-1).$$

Hence, if m is odd, the HCF = 2; if m is even, the HCF = 1.

7. They are equal. (The first is

$$\frac{17}{100} \times 19\,000\,000,$$

the second is

$$\frac{19}{100} \times 17\,000\,000,$$

which are equal since multiplication is commutative and associative.)

8.

(a)

$$\frac{3}{2} \times \frac{4}{3} \times \frac{5}{4} \times \frac{6}{5} = \frac{6}{2} = 3$$

(b)

$$\sqrt{\frac{3}{2} \times \frac{4}{3} \times \frac{5}{4} \times \frac{6}{5} \times \frac{7}{6} \times \frac{8}{7}} = \sqrt{\frac{8}{2}} = 2$$

(c)

$$\begin{aligned}
10 \times 9 \times 8 \times 7 \times \times 5 \times 4 \times 3 \times 2 \times 1 \;\; &\text{seconds} \\
&= \frac{10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1}{60} \;\; \text{minutes} \\
&= \frac{10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1}{60 \times 60} \;\; \text{hours} \\
&= \frac{10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1}{60 \times 60 \times 24} \;\; \text{days} \\
&= \frac{10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1}{60 \times 60 \times 24 \times 7} \;\; \text{weeks} \\
&= 6 \;\; \text{weeks (after cancelling).}
\end{aligned}$$

Note: These three questions underline what we mean by *structural arithmetic.* Fractions should **never** be handled by *evaluating* numerators and denominators. Instead one should always be on the lookout for structural features which simplify calculation – such as *cancellation.*

9.

(a) Suppose a rectangle in the "DIN A" series has dimensions a by b, with $a < b$. Folding in half produces a rectangle of size $\frac{b}{2}$ by a. Hence $b : a = a : \frac{b}{2}$, so $b^2 = 2a^2$, and $b : a = \sqrt{2} : 1$.

(b) (i) $\frac{1}{r}$. (ii) r.

10.

(a) "15% discount" means the price actually charged is "85% of the marked price". Hence each marked price needs to be **multiplied by 0.85**.

The distributive law says we may add the marked prices first and then multiply the total (exactly £80) by 0.85 to get

$$£\left(\frac{85}{100} \times 80\right) = £(17 \times 4) = £68.$$

Note: The context (shopping, sales tax, and discount) is mathematically uninteresting. What matters here is the underlying *multiplicative structure* of the solution, which arises in many different contexts.

(b) "Add 20% VAT" means multiplying the discounted pre-VAT total (£68) by 1.2, or $\frac{6}{5}$. Hence the final price, with VAT added, is £$(1.2 \times 0.85 \times 80)$.

If the VAT were added first, the price before discount would be £(1.2×80), and the final price after allowing for the discount would be £$(0.85 \times 1.2 \times 80)$.

Since multiplication is commutative, the two calculations have the same result, so the order does not matter (just as the final result in Problem **9** is the same whether one first enlarges A4 to A3 and then reduces A3 to A4, or first reduces A4 to A5 and then enlarges A5 to A4).

Note: Notice that we did not evaluate the two answers to see that they gave the same output £81.60. If we had, then the equality might have been a fluke due to the particular numbers chosen. Instead we left the answer unevaluated, in *structured form*, which showed that the equality would hold for any input.

(c) To cancel out multiplying by $\frac{6}{5}$ we need to multiply by $\frac{5}{6}$ – a discount of $\frac{1}{6}$, or $16\frac{2}{3}\%$.

Note: This question has nothing to do with financial applications. It is included to underline the fact that although percentage change questions use the language of *addition* and *subtraction* ("increase", or "decrease"), the mathematics suggests they should be handled **multiplicatively**.

11.

(a)(i) $2 \times 5 < 3 \times 4$, so

$$\frac{7}{2 \times 5} > \frac{7}{3 \times 4}.$$

Hence

$$\frac{1}{2} + \frac{1}{5} > \frac{1}{3} + \frac{1}{4}.$$

(ii) At first sight, "$10 < 12$" may not seem related to "$\frac{1}{2} + \frac{1}{5} > \frac{1}{3} + \frac{1}{4}$". Yet the crucial fact we started from in part (i) was "$2 \times 5 = 10 < 12 = 3 \times 4$".

(b) $10 < 12$, so

$$(x + 2)(x + 5) = x^2 + 7x + 10 < x^2 + 7x + 12 = (x + 3)(x + 4).$$

(i) If all four brackets are positive (i.e. if $x > -2$), then we also have $2x + 7 > 0$, and it follows that

$$\begin{aligned}\frac{1}{x+2} + \frac{1}{x+5} &= \frac{2x+7}{(x+2)(x+5)} \\ &> \frac{2x+7}{(x+3)(x+4)} \\ &= \frac{1}{x+3} + \frac{1}{x+4}.\end{aligned}$$

(ii) When calculating with the given algebraic expression, the values

$$x = -2, -3, -4, -5$$

are "forbidden values".

If $x > -2$, then (as in part (i)) we have

$$\begin{aligned}\frac{1}{x+2} + \frac{1}{x+5} &= \frac{2x+7}{(x+2)(x+5)} \\ &> \frac{2x+7}{(x+3)(x+4)} \\ &= \frac{1}{x+3} + \frac{1}{x+4}.\end{aligned}$$

For permitted values of $x < -2$, one or more of the brackets $(x+2)$, $(x+5)$, $(x+3)$, $(x+4)$ will be negative. However, one can still carry out the algebra to simplify

$$\frac{1}{x+2} + \frac{1}{x+5} = \frac{2x+7}{(x+2)(x+5)} \quad \text{and} \quad \frac{1}{x+3} + \frac{1}{x+4} = \frac{2x+7}{(x+3)(x+4)}$$

When $x = -\frac{7}{2}$ both expressions are **equal**, and equal to 0. The simplified numerators are both positive if $x > -\frac{7}{2}$, and both negative if $x < -\frac{7}{2}$; and the sign of the denominators changes as one moves through the four intervals $-3 < x < -2$, $-4 < x < -3$, $-5 < x < -4$, $x < -5$, with the inequality switching

from " $>$ " (for $x > -2$) to " $<$ " (for $-3 < x < -2$),
to " $>$ " (for $-3.5 < x < -3$),
to " $<$ " (for $-4 < x < -3.5$),
to " $>$ " (for $-5 < x < -4$),
to " $<$ " (for $x < -5$).

12.

(a) (i) $3 + 2\sqrt{2}$; (ii) $3 - 2\sqrt{2}$; (iii) $7 + 5\sqrt{2}$.

Note: Notice that you can write down the answer to (ii) as soon as you have finished (i), without doing any further calculation.

(b) (i) $2+\sqrt{6}$; (ii) $\sqrt{2}+\sqrt{3}$; (iii) $\frac{1+\sqrt{5}}{2}$ (the *Golden Ratio* $\frac{1+\sqrt{5}}{2} = \tau$ is the larger root of the quadratic equation $x^2 - x - 1 = 0$. Hence $\frac{3+\sqrt{5}}{2} = \tau + 1 = \tau^2$);
(iv) $\sqrt{10-2\sqrt{5}}$: this does not simplify further.

Note: Some readers may think an apology is in order for part (iv). The lesson here is that, while one should always try to simplify, there is no way of knowing in advance whether a simplification is possible. And there is no way out of this dilemma. So one is reduced to thinking: any simplification would involve $\sqrt{5}$, and if one tries to solve $(a+b\sqrt{5})^2 = 10-2\sqrt{5}$, then the solutions for a and b do not lead to anything "simpler". (This repeated surd should perhaps have rung bells, as it was equal to the exact expression for $4\sin 36^\circ$ in Problem **3**(c). It was included here partly because the question of its simplification should already have arisen when it featured in that context.)

13. In reconstructing the missing digits the number of possible solutions is determined by *the highest common factor of the multiplier and* 10. At the first step (in the units column):

> because $HCF(6,10) = 2$, $\square \times 6 = 8 \pmod{10}$ has **two** solutions which differ by 5 – namely 3 and 8.

The first possibility then requires us to solve $(\square \times 3) + 1 = 2 \pmod{10}$: because $HCF(3,10) = \mathbf{1}$, this has just **one** solution – namely 7. This gives rise to the solution **76** × **3** = **228**.

The second possibility requires us to solve $(\square \times 8) + 4 = 2 \pmod{10}$: because $HCF(8,10) = \mathbf{2}$, this has **two** solutions which differ by 5 – namely 1 and 6. This gives rise to two further solutions: **16** × **8** = **128**, and **66** × **8** = **528**.

14.

(a) The solutions are entirely elementary, with no trickery. But they can be surprisingly elusive. And since this elusiveness is the only reason for including the problem, we hesitate to relieve any frustration by giving the solution.

The whole thrust of the "24 game" is to underline the scope for "getting to know" the many faces of a number like 24: for example, as $24 = 12+12$ $(= 3\times4+3\times4$ for 3, 3, 4, 4); as $24 = 25-1$ $(= 5\times5-3\div3$ for 3, 3, 5, 5); and as $24 = 27-3$ $(= 3\times3\times3-3$ for 3, 3, 3, 3). So one should be looking for ways of exploiting other important arithmetical aspects of 24 – in particular, as 4×6 and as 3×8.

(b) (i) $0 = (4-4)+(4-4)$; $1 = (4\div4)\times(4\div4)$; $2 = (4\div4)+(4\div4)$; $3 = (4+4+4)\div4$; $4 = ((4-4)\times4)+4$; $5 = ((4\times4)+4)\div4$; $6 = 4+((4+4)\div4)$; $7 = 4+4-(4\div4)$; $8 = (4+4)\times(4\div4)$; $9 = (4+4)+(4\div4)$. The output 10 seems to be impossible with the given restrictions.

(ii) With squaring and $\sqrt{}$ allowed we can manage $10 = 4+4+4-\sqrt{4}$. Indeed, one can make everything up to 40 except (perhaps) 39.

15.

(a) (i) $a^2 + 2ab + b^2$; $a^3 + 3a^2b + 3ab^2 + b^3$

(ii) Replace b by $(-b)$: $a^2 - 2ab + b^2$; $a^3 - 3a^2b + 3ab^2 - b^3$

(b) (i) $(x+1)^2$; (ii) $(x^2-1)^2$; (iii) $(x^2-1)^3$

(c) (i) $a^2 - b^2$

(ii) Replace "b" by "$b+c$": $a^2 - (b+c)^2 = a^2 - b^2 - c^2 - 2bc$

Replace "b" by "$b-c$": $a^2 - (b-c)^2 = a^2 - b^2 - c^2 + 2bc$

(d) One way is to rewrite this expression as a difference of two squares:

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

Note: As so often, the messages here are largely implicit. In part (a)(ii) we explicitly highlight the intention to use what you already know (by simply substituting "$-b$" in place of "b". In part (b), you are expected to recognise (i), and then to see (ii) and (iii) as mild variations on the expansions of $(a-b)^2$ and $(a-b)^3$ in part (a). Part (c) repeats (in silence) the message of (a)(ii): **think** – don't slog it out. And part (d) encourages you to keep an eye out for thinly disguised instances of "a difference of two squares".

16.

(a) **Final digits**: 'block' 4, 6 of length 2;

leading digits: "block" 4, 1, 6, 2, 1 of length 5.

(b) **Claim** The sequence of "units digits" really does recur.

Proof Given a power of 4 that has units digit 4, the usual multiplication algorithm for multiplying by 4 produces a number with units digit 6.

Given this new power of 4 with units digit 6, the usual multiplication algorithm for multiplying by 4 produces a number with units digit 4.

At this stage the sequence of units digits begins a new cycle.

[Alternatively: The units digit is simply equal to the relevant power of 4 (mod 10). Multiplying by 4 changes 4 to 6 (mod 10); multiplying by 4 changes 6 to 4 (mod 10); – and the cycle repeats.]

(c) The sequence of **leading** digits *seems* to recur every 5 terms, because $4^5 = 2^{10} = 1024$ is almost exactly equal to 1000. Each time we move on 5 steps in the sequence, we multiply by $4^5 = 1024$. As far as the leading digit is concerned, this has the same effect as multiplying the initial term (4) by slightly more than 1.024 (then adding any 'carries'), which is very like multiplying by 1 – and so does not change the leading digit (yet).

However, each time we move on 10 steps in the sequence, we multiply by $4^{10} = 1024^2 = 1\,048\,576$. As far as the leading digit is concerned, this has the same effect as multiplying by slightly more than 1.048576.

When we move on 25 steps, we multiply by $4^{25} = 1\,125\,899\,906\,842\,624$. And as far as the leading digit is concerned, this has the same effect as multiplying by slightly more than 1.12599906842624. And so on.

Eventually, the multiplier becomes large enough to change one of the leading digits.

17. The total is 100.

Having found this by *direct* calculation, we should think *indirectly* and notice that $100 = 10^2$.

And we should then ask: "Why 10? What has **10** got to do with the **4**× multiplication table?"

A quick check of the 1× multiplication table (total = 1), the 2× multiplication table (total = 9), etc. may suggest what we should have seen immediately.

The first row has sum:	$(1+2+3+4)$.
The second row has total	$2\times(1+2+3+4)$.
The third row has total	$3\times(1+2+3+4)$.
The fourth row has total	$4\times(1+2+3+4)$.

$\therefore$ The total is $\quad (1+2+3+4)\times(1+2+3+4) = (1+2+3+4)^2$.

19.

(a) $\sin 45^\circ = \frac{1}{\sqrt{2}} = \frac{\sqrt{2}}{2} = \cos 45^\circ$; $\tan 45^\circ = 1$

(b) $AM = \sqrt{3}$; $\sin 30^\circ = \frac{1}{2}$, $\cos 30^\circ = \frac{\sqrt{3}}{2}$, $\tan 30^\circ = \frac{1}{\sqrt{3}} = \frac{\sqrt{3}}{3}$; $\sin 60^\circ = \frac{\sqrt{3}}{2}$, $\cos 60^\circ = \frac{1}{2}$, $\tan 60^\circ = \sqrt{3}$.

(c) (iii) $\cos 315^\circ = \cos 45^\circ = \frac{\sqrt{2}}{2}$; $\sin 225^\circ = -\sin 45^\circ = -\frac{\sqrt{2}}{2}$; $\tan 210^\circ = \tan 30^\circ = \frac{\sqrt{3}}{3}$; $\cos 120^\circ = -\cos 60^\circ = -\frac{1}{2}$; $\sin 960^\circ = \sin 240^\circ = -\sin 60^\circ = -\frac{\sqrt{3}}{2}$; $\tan(-135^\circ) = \tan 45^\circ = 1$.

(d) Cut the n-gon into n "cake slices", and use the formula "$\frac{1}{2}ab\sin C$" for each slice.
(i) $\frac{3\sqrt{3}}{4}$; (ii) 2; (iii) $\frac{3\sqrt{3}}{2}$; (iv) $2\sqrt{2}$; (v) 3

(e) Work out the side length of the n-gon, then cut the n-gon into n "slices", and use the formula "$\frac{1}{2}$(base × height)" for each slice.
(i) $3\sqrt{3}$; (ii) 4; (iii) $2\sqrt{3}$; (iv) $8(\sqrt{2}-1)$; (v) $12(2-\sqrt{3})$

Note: There is no hidden trig here: all you need is Pythagoras' Theorem. For example, in part (e)(iv) we can extend alternate sides of the regular octagon to form the circumscribed 2 by 2 square. The four corner triangles are isosceles right angled triangles with hypotenuse of length s (the side of the octagon). Hence each side of the square is equal to $s + 2\cdot\frac{s}{\sqrt{2}} = 2$, whence $s = 2(\sqrt{2}-1)$.

20. (a) $\sqrt{8}$; (b) $\sqrt{2}$, $\sqrt{2}$; (c) $\sqrt{8} = \sqrt{4 \times 2} = 2\sqrt{2}$

21. Construct the perpendicular from A to BC (possibly extended); let this meet the line BC at X. There are four possibilities:

(i) either $X = C$, in which case $\angle BCA$ is a right angle as required; or $X = B$, in which case $b^2 = a^2 + c^2$, contradicting $a^2 + b^2 = c^2$;

(ii) $X \neq B, C$, and C lies between B and X;

(iii) $X \neq B, C$, and X lies between B and C;

(iv) $X \neq B, C$, and B lies between X and C.

We analyse case (ii) and leave cases (iii) and (iv) to the reader.

(ii) $\triangle AXC$ and $\triangle AXB$ are both right angled triangles; so by Pythagoras' Theorem we know that

$$\begin{aligned} AC^2 &= AX^2 + XC^2, \quad \text{and} \\ AB^2 &= AX^2 + XB^2 \\ &= AX^2 + (XC + CB)^2 \\ &= AX^2 + XC^2 + CB^2 + 2XC \cdot CB \\ &= AC^2 + CB^2 + 2XC \cdot CB. \end{aligned}$$

Since we are told that $AC^2 + CB^2 = AB^2$, it follows that $2XC \cdot CB = 0$, contrary to $X \neq C$.

Note: Notice that the proof of the converse of Pythagoras' Theorem makes use of Pythagoras' Theorem itself.

22.

(a) $c = b + 1$, so $a^2 = c^2 - b^2 = 2b + 1$. Hence a is odd, and we can write $a = 2m + 1$.

(b) Suppose $b = 2n - 1$ is also odd. Then $c^2 = 4n^2$ is divisible by 4 – which contradicts the fact that $b^2 = 4(n^2 - n) + 1$, and $a^2 = 4(m^2 + m) + 1$, so $a^2 + b^2$ leaves remainder 2 on division by 4.

Hence $b = 2n$ is even and $c = 2n + 1$ is odd. But then

$$(2m + 1)^2 + (2n)^2 = a^2 + b^2 = c^2 = (2n + 1)^2,$$

so $4(m^2 + m) = 4n$, and $n = m(m + 1)$.

23.

(a) If a and b are both even, then $HCF(a, b) \neq 1$, so the triple would not be primitive.

If a and b are both odd, we use the idea from Problem **22**(b). Suppose $a = 2m+1$, $b = 2n+1$; then $a^2 = 4(m^2+m)+1$, and $b^2 = 4(n^2+n)+1$, so $a^2+b^2 = 2 \times (2(m^2+m+n^2+n)+1)$. But this is "twice an odd number", so cannot be equal to c^2 (since c would have to be even, and any even square must be a multiple of 4).

Hence we may assume that a is odd and b is even: so c is is odd.

(b) Then $a^2+b^2 = c^2$ yields $b^2 = c^2 - a^2 = (c-a)(c+a)$, so

$$\left(\frac{b}{2}\right)^2 = \left(\frac{c-a}{2}\right)\left(\frac{c+a}{2}\right).$$

Any common factor of $\frac{c+a}{2}$ and $\frac{c-a}{2}$ divides their sum c and their difference a, so $HCF(\frac{c-a}{2}, \frac{c+a}{2}) = 1$. Since the difference of these two factors is a, which is odd, they have opposite parity.

(c) If two integers are relatively prime, and their product is a square, then each of the factors has to be a square (consider their prime factorisations). Hence $\frac{c+a}{2} = p^2$ and $\frac{c-a}{2} = q^2$, where $HCF(p,q) = 1$ and p and q have opposite parity. Therefore

$$c = p^2 + q^2, \quad a = p^2 - q^2, \quad b = 2pq.$$

(d) It is easy to check that any triple of the given form is (i) primitive, and (ii) satisfies $a^2 + b^2 = c^2$.

24. Claim The only such triples are those of the form $(3s, 4s, 5s)$.

Proof We show that the only *primitive* Pythagorean triple which forms an arithmetic progression is the familiar triple $(3,4,5)$.

By Problem **23**, one of the numbers in any *primitive* Pythagorean triple is *even* (namely $2pq$) and two are *odd* (p and q are of opposite parity, so p^2-q^2 and p^2+q^2 are both odd).

$\therefore$ $2pq$ is the "middle term", and the smallest and largest terms differ by $2q^2$.

$\therefore$ the common difference $d = c - b = b - a$ is equal to q^2.

$\therefore$ $2pq = p^2$, so $p = 2q$.

Finally, since p and q are relatively prime, we must have $q = 1$, $p = 2$. QED

Note: Alternatively, let (a,b,c) be any Pythagorean triple (not necessarily primitive), which forms an arithmetic progression. Then

$$a^2 = c^2 - b^2 = (c-b)(c+b) = (b-a)(c+b).$$

So $b(c+b) = a(a+b+c)$. Hence $a \cdot 3b = a(a+b+c) = b(c+b)$. It then follows that $3b^2 = b(a+b+c) = 4ba$, so $3b = 4a$ and $a : b = 3 : 4$.

25.

(a) $2 = 1^2 + 1^2$, $5 = 2^2 + 1^2$, $13 = 3^2 + 2^2$, $17 = 4^2 + 1^2$, $29 = 5^2 + 2^2$, $37 = 6^2 + 1^2$, $41 = 5^2 + 4^2$, $53 = 7^2 + 2^2$, $61 = 6^2 + 5^2$, $73 = 8^2 + 3^2$, $89 = 8^2 + 5^2$, $97 = 9^2 + 4^2$.

(b) $(a^2 + b^2)(c^2 + d^2) = (ac - bd)^2 + (ad + bc)^2$.

Note: It is easy to check this identity once it is given, but most of us are not so fluent in algebra as to spot this handy identity without help! However, Chapter 1 is about "Mental skills", and one such technique (once you have mastered it) arises from the arithmetic of *complex numbers*. If you have met complex numbers, then this identity can be written down immediately. Let us explain briefly how.

Every complex number $w = a + bi$ (where $i^2 = -1$) can be represented as a point in the complex plane with coordinates (a, b). The "size", or *modulus*, of w is its length $|w|$ (the distance of (a, b) from the origin $(0, 0)$); and the square of this length $a^2 + b^2$ is referred to as the *norm* of the complex number $w = a + bi$. The required identity is an immediate consequence of the two facts:

- the modulus of a product is equal to the product of the two moduli: $|wz| = |w| \cdot |z|$, and
- the norm $a^2 + b^2$ can be expressed algebraically as $a^2 + b^2 = (a + bi)(a - bi)$.

Once we know these facts:

- $a^2 + b^2$ can be interpreted as the norm of $w = a + bi$, and
- $c^2 + d^2$ as the norm of $z = c + di$;

the product of the two norms $(a^2 + b^2)(c^2 + d^2)$ is then equal to the norm of the product $w \cdot z = (ac - bd) + (ad + bc)i$.

Note: If we choose $z = c - di$, then $wz = (ac + bd) + (bc - ad)i$, and we get a second identity: $(a^2 + b^2)(c^2 + d^2) = (ac + bd)^2 + (bc - ad)^2$.

(c) The square $(2n)^2$ of any even number $2n$ is a multiple of $2^2 = 4$.

Any odd number has the form $2n + 1$; its square $(2n + 1)^2 = 4n^2 + 4n + 1$ is 1 more than a multiple of 4. So in the sum of two squares, we have

(i) both squares are even and their sum is a multiple of 4, or

(ii) one square is even and one is odd and their sum is of the form $4k + 1$, or

(iii) both squares are odd and their sum is of the form $4k + 2$.

Hence no number of the form $4k + 3$ can be written as a sum of two squares.

(d) We are told that $2 = 1^2 + 1^2$, and that Euler showed every prime of the form $4k + 1$ can be written as the sum of two squares. Part (b) then shows that any product of such primes can be written as the sum of two squares. And if we multiply a sum of two squares by a square, the result can again be written as the sum of two squares. This allows us to construct the list of forty six integers $N < 100$ which can be so written. These are precisely the integers of the form "(a square) × (a product of distinct primes p, where $p = 2$ or $p = 4k+1$)":

0, 1, 2, 4, 5, 8, 9, 10, 13, 16, 17, 18, 20, 25, 26, 29, 32, 34, 36, 37, 40, 41, 45, 49, 50, 52, 53, 58, 60, 61, 64, 65, 68, 72, 73, 74, 80, 81, 82, 83, 85, 87, 89, 90, 97, 98.

(e) The side of such a square is the hypotenuse of a right angled triangle whose legs run in the x- and y- directions, and have integer lengths (because their vertices are at points with integer coordinates). Hence the answer is exactly the same as for part (d) (provided one does not quibble about the idea of a square with side 0 and area 0).

26. Most sheets in a newspaper are double sheets with four pages. If we assume that all sheets are double sheets, then the 13 pages before page 14 match up with the 13 pages after page 27, so there are $27 + 13 = 40$ pages in all. (If the paper included inserted 'single sheets' – with just two pages, then there is no solution.)

27.

(a) If $\theta = 90^\circ$, then the overlap is clearly one quarter of the small square. In general, the continuations of the sides BA and DA cut the small square into four congruent quadrilaterals, one of which is the area of overlap. So the overlap is always **one quarter** of the lower square.

(b) The area of overlap for "a large equilateral $\triangle ABC$ on top of a small equilateral $\triangle PQR$" is not constant, but depends on the angle of orientation of the large triangle. However, if viewed in the right way, something similar works for a large regular $2n$-gon on top of a small regular $2n$-gon with one corner of the large polygon at the centre of the small one.

The key is to realise how the fraction "one quarter" arises for a regular 4-gon. There $2n = 4$, so $n = 2$, and each vertex angle is equal to $\left(\frac{360^\circ}{2n}\right)(n-1) = 90^\circ$, which is exactly $\frac{n-1}{2n} = \frac{1}{4}$ of 360°. For a regular $2n$-gon, the large polygon always covers a fraction equal to exactly $\frac{n-1}{2n}$ of the small polygon.

28. 3 cm, the same as the perimeter of triangle ABC.[2]

What if A were folded to some point A'' on BC?

29. *In extremis* one may reach for trigonometry: if we denote the three angles by α (at A), β (at B), and γ (at C), then the arrangement of squares implies that $\tan\alpha = \frac{1}{3}$, $\tan\beta = \frac{1}{2}$, and $\tan\gamma = 1$, so we can use the standard identity

$$\tan(\alpha+\beta) = \frac{\tan\alpha + \tan\beta}{1 - \tan\alpha\tan\beta}$$

to see that $\tan(\alpha+\beta) = 1 = \tan\gamma$.

---

[2] From: Y. Wu, The examination system in China: the case of zhongkao mathematics. 12th International Congress on Mathematical Education. 8 July – 15 July, 2012, COEX, Seoul, Korea

However, it is worth looking for a more elementary explanation than 'brute force calculation'. If we embed the horizontal 3 by 1 rectangle $ADEH$ in the top right hand corner of a 4 by 4 square $ZDXY$, (with Z labelling the top left hand corner), then we can complete the square $AEPQ$, which has AE as one side, with P on side XY and Q on side YZ.

Then $\angle AEH = \angle DAE$, and $\angle AEQ = \angle DCE$. So all we need to explain is why $\angle HEQ = \angle DBE$ – and this follows from the fact that EQ passes through the centre of the 4 by 4 square $ZDXY$.

30.

(a) Construct points P and Q inside the trapezium so that $MNPQ$ is similar to $BCMN$. If the line through P parallel to AB meets BC at X, and the line through Q parallel to AC meets BC at Y, then $MNPQ$, $NBXP$, $XYQP$, $MCYQ$ are the required pieces.

(b) Each of the four pieces must be *three-quarters* of one of the small squares. So we have to lose one quarter of each small square. There are various ways to do this, but most create non-congruent parts. Cut each of the smaller squares into quarters as for the original square. If O is the centre of the original square, lump together the three mini-squares which have O as a vertex to form an L-shape. Each of the three remaining small squares has lost a quarter and forms an identical L-shape.

31. To divide the shaded region in 2 congruent parts, rotate the lower small semicircle through the angle $\frac{\pi}{2}$ anticlockwise about the centre of the large circle.

Note: The same idea allows one to divide the shaded region into n congruent parts: rotate the lower small semicircle successively through the angle $\frac{\pi}{n}$ anticlockwise about the centre of the large circle.[3]

32.

(a) The point O lies on the perpendicular bisector of AB, and so is equidistant from the two endpoints A and B, so $OA = OB$. The point O also lies on the perpendicular bisector of BC, and so is equidistant from the two endpoints B and C, so $OB = OC$.

(b) (ii) $\angle BA'C$ and $\angle BAC$ are angles subtended in the same segment of the circle by the same chord BC, so are equal ("angles in the same segment").

$\angle A'CB$ is the angle subtended on the circumference (at C) by the diameter $A'B$, and so must be a right angle. In $\triangle A'BC$ we then see that

$$\sin A = \sin A' = \frac{a}{2R}.$$

If we now switch attention from the angle at A to the angle at B, and then to the angle at C, we can show that $\sin B = \frac{b}{2R}$, and that $\sin C = \frac{c}{2R}$.

[3] From: Introductory Assignment, Gelfand Correspondence Program in Mathematics.

33.

(a) Drop a perpendicular from A to meet the line BC at X. Then $AX = b \cdot \sin C$, so

$$\Delta = \frac{1}{2} \cdot (a \times b \sin C).$$

(b) Substitute "$\sin C = \frac{c}{2R}$" (from the Sine Rule) into the formula in part (a).

34.

(a) (i) $\frac{\alpha}{2\pi}$; $\frac{\alpha}{2\pi} \times$ (surface area of unit sphere $= 4\pi$) $= 2\alpha$.

(ii) 4α

(b) (i) 4β; (ii) 4γ

(c) (i) $4(\alpha + \beta + \gamma)$

(ii) Triangle ABC and its sister triangle $A'B'C'$ are congruent, and each is covered 3 times.

(iii)

$$\begin{aligned} 4(\alpha + \beta + \gamma) &= \text{(total surface area of the unit sphere)} \\ &\quad + (4 \times \text{(area of the spherical triangle } ABC)) \\ \therefore \text{area}(\triangle ABC) &= (\alpha + \beta + \gamma) - \pi. \end{aligned}$$

Note: In particular, the formula for the area of a spherical triangle implies:

- the angle sum $\alpha + \beta + \gamma$ in any spherical triangle is **always greater than** π, and
- the larger the triangle ABC, the more its angle sum must exceed π.

35.

(a)

$$\begin{aligned} \cos(A + B) + i \sin(A + B) &= e^{i(A+B)} \\ &= e^{iA} \cdot e^{iB} \\ &= (\cos A + i \sin A)(\cos B + i \sin B). \end{aligned}$$

Hence

$$\sin(A + B) = \sin A \cos B + \cos A \sin B,$$

and

$$\cos(A + B) = \cos A \cos B - \sin A \sin B.$$

To reconstruct $\tan(A + B)$, divide these two expressions, and then divide numerator and denominator by "$\cos A \cos B$" to get

$$\tan(A + B) = \frac{\sin(A + B)}{\cos(A + B)} = \frac{\tan A + \tan B}{1 - \tan A \tan B}.$$

(b)

$$\begin{aligned}\sin X &= \sin(A+B)\\ &= \sin A\cos B + \cos A\sin B\\ &= \sin\left(\frac{X+Y}{2}\right)\cos\left(\frac{X-Y}{2}\right) + \cos\left(\frac{X+Y}{2}\right)\sin\left(\frac{X-Y}{2}\right),\end{aligned}$$

and

$$\begin{aligned}\sin Y &= \sin(A-B)\\ &= \sin A\cos B - \cos A\sin B\\ &= \sin\left(\frac{X+Y}{2}\right)\cos\left(\frac{X-Y}{2}\right) - \cos\left(\frac{X+Y}{2}\right)\sin\left(\frac{X-Y}{2}\right),\end{aligned}$$

$$\therefore \sin X + \sin Y = 2\sin\left(\frac{X+Y}{2}\right)\cos\left(\frac{X-Y}{2}\right).$$

For $\sin X - \sin Y$, substitute "$-Y$" in place of Y to get:

$$\sin X - \sin Y = 2\sin\left(\frac{X-Y}{2}\right)\cos\left(\frac{X+Y}{2}\right).$$

Similarly

$$\begin{aligned}\cos X + \cos Y &= \cos(A+B) + \cos(A-B)\\ &= (\cos A\cos B - \sin A\sin B) + (\cos A\cos B + \sin A\sin B)\\ &= 2\cos A\cos B\\ &= 2\cos\left(\frac{X+Y}{2}\right)\cos\left(\frac{X-Y}{2}\right)\end{aligned}$$

$$\begin{aligned}\cos X - \cos Y &= \cos(A+B) - \cos(A-B)\\ &= (\cos A\cos B - \sin A\sin B) - (\cos A\cos B + \sin A\sin B)\\ &= -2\sin A\sin B\\ &= -2\sin\left(\frac{X+Y}{2}\right)\sin\left(\frac{X-Y}{2}\right).\end{aligned}$$

(c) (i) $\sin(A+B) = \sin 90^\circ = 1$;

$$\sin A\cos B + \cos A\sin B = \left(\frac{1}{2}\right)^2 + \left(\frac{\sqrt{3}}{2}\right)^2 = 1.$$

(ii) $\cos X - \cos Y = \frac{1}{2} - 1 = -\frac{1}{2}$;

$$-2\sin\left(\frac{X+Y}{2}\right)\sin\left(\frac{X-Y}{2}\right) = -2\sin^2 30^\circ = -\frac{1}{2}.$$

(d) (i) $2\sin A \sin B = \cos(A-B) - \cos(A+B)$.

$$\begin{aligned}\therefore\ 2\sin A\sin B + 2\sin C\sin D &\\ &= [\cos(A-B) - \cos(A+B)]\\ &\qquad + [\cos(C-D) - \cos(C+D)]\\ &= \cos(A-B) + \cos(C-D)\\ &\qquad (\text{since } C+D = \pi - (A+B))\\ &= 2\cos\left(\frac{A+C-(B+D)}{2}\right)\cos\left(\frac{A+D-(B+C)}{2}\right)\\ &= 2\cos\left(\frac{\pi}{2} - (B+D)\right)\cos\left(\frac{\pi}{2} - (B+C)\right)\\ &= 2\sin(B+D)\sin(B+C).\end{aligned}$$

Note: We can swap A and B without changing the expression "$\sin A \sin B + \sin C \sin D$". Hence the same should be true of the RHS "$\sin(B+C)\sin(B+D)$". Fortunately, since $A+B+C+D = \pi$, we know that $\sin(A+C) = \sin(B+D)$, and $\sin(A+D) = \sin(B+C)$.

(ii) In triangle WXY we see that $A+B+C+D = \pi$. Hence

$$\sin A \sin B + \sin C \sin D = \sin(A+D)\sin(B+D).$$

Now let R be the radius of the circle. Use "equality of angles in the same segment" and the Sine Rule (in its full form: see Problem **32**) to write:

$2R\sin A = XY$, $2R\sin B = WZ$, $2R\sin C = YZ$, $2R\sin D = WX$,

$2R\sin(A+D) = WY$, $2R\sin(B+D) = XZ$.

$\therefore WX \times YZ + WZ \times XY = WY \times XZ$.

36. Yes.

Let the perpendicular bisectors of AB and BC meet at the point O.

Then $OA = OB$ and $OB = OC$, so the circle with centre O passing through A also goes through B and C.

We have to prove that this circle also passes through D, E, etc..

To do this we prove that $\triangle OBC \equiv \triangle OCD$.

We know that $\triangle OAB \equiv \triangle OBC$ (by SSS-congruence: $OA = OB$ and $OB = OC$ by the construction of O; and $AB = BC$ since both are sides of a regular n-gon). Moreover

$$\begin{aligned}\angle OAB &= \angle OBA \text{ (base angles of the isosceles triangle } \triangle OAB)\\ &= \angle OCB \text{ (since } \triangle OAB \equiv \triangle OBC)\\ &= \angle OBC \text{ (base angles of the isosceles triangle } \triangle OBC).\end{aligned}$$

And $\angle ABC = \angle BCD$ (angles of the same regular n-gon).

$\therefore \angle OCD = \angle BCD - \angle OCB = \angle ABC - \angle OBA = \angle OBC$.
$\therefore \triangle OBC \equiv \triangle OCD$ (by the SAS-congruence criterion).
Hence $OC = OD$.
Continuing in this way we can prove that $OA = OD = OE$, etc..

37.

(a) Yes. There are two nets for a regular tetrahedron. One of these consists of four equilateral triangles in a row (alternately right side up and right side down). In making the tetrahedron, the two sloping ends are glued together. So if we cut half a triangle from one end and stick it on the other end, we get a 4 by $\sqrt{3}$ rectangle which folds round the tetrahedron exactly without any gaps or overlaps.

(b) The usual way to wrap a cube with edges of length 2 is to take a 4 by 8 rectangular piece of wrapping paper, to position the cube centrally on an edge of length 4 (1 unit from each edge), and to wrap the paper to cover a circuit of four faces. The overlapping residue can then be folded down to cover each side face, with overlaps. Hence the ratio

$$\text{“area of paper”} : \text{“surface area of cube”} = 8 : 6.$$

The same ‘wastage rate’ can be achieved with a square $4\sqrt{2}$ by $4\sqrt{2}$ piece of paper. Position the cube centrally on the paper, but turned through an angle of 45°. Then fold the four corners of the paper up each of the four side faces (with folds to tuck in four ‘wasted’ isosceles right angled triangles – one in the middle of each edge of the paper, with total wasted area 8). Finally, the four isosceles right angled triangles at the four corners of the paper can be folded in to exactly cover the top face without further overlaps.

However, one can do significantly better if the paper can be folded back on itself. Take a 2 by 14 rectangle, and think of this as being marked into seven 2 by 2 squares. Place the cube to cover the central 2 by 2 square – leaving three 2 by 2 squares sticking out each side. Fold one 2 by 6 strip up to cover the top square, before folding back along a diagonal of the top square to reveal the inside of the paper and to cover half of the top square *twice* before folding down to cover one side square. Do the same with the other 2 by 6 strip, with the reverse fold along the diagonal of the top square resulting in the other half of the top square being covered twice, with the tail folding down to cover the other side square. Hence the ratio

$$\text{“area of paper”} : \text{“surface area of cube”} = 7 : 6.$$

(This lovely solution was provided by Julia Gog. We do not know whether one can do better.)

38.

(i) Yes. (Cut off a corner A say with a plane passing through the three neighbours of A.)

(ii) Yes. (Cut the cube parallel to a face.)

(iii) No. (Any cross-section of the cube is a polygon. Each edge of this polygonal cross-section is the line segment formed by the intersection of the cutting plane with one of the faces of the cube. Since the cube has just six faces, the cross-section can have at most six sides.)

(iv) Yes. (Let A and G be two opposite corners – so that AG passes through the centre O of the cube. Then the plane through O which is perpendicular to AG cuts the surface of the cube in a regular hexagon.)

(v) No. (It may not be clear how to prove this easily. It is perfectly possible to obtain a pentagonal cross-section by cutting with a plane that misses exactly one face. But if the cutting plane misses exactly one face, we can be sure that it must cut both faces belonging to some "opposite pair"; and these two faces are **parallel**, so the resulting edges of the cross-sectional pentagon are **parallel**. Hence the pentagon can never be regular.)

39.

(i) No; (ii) Yes; (iii) No; (iv) Yes; (v) No.

The 12 edges of the cube come in three groups of four – namely the four parallel edges in each of three directions.

Consider the four edges in one of these parallel groups. If the Sun's rays are parallel to these four edges, then each of these edges projects to a single vertex of the outline of the shadow – which is a projection of a square.

In all other cases two of the four parallel edges in the group give rise to shadows that form part of the *boundary* of the shadow polygon, while the other two edges project to the inside of the shadow (and so do not feature in the boundary of the shadow). Hence each of the three groups provides two edges to the boundary of the shadow polygon, and we obtain a hexagon.

To obtain a *regular* hexagon, align the Sun's rays parallel to the line AG joining two opposite corners A and G of the cube, and position the shadow plane perpendicular to this direction. The three edges at these two corners A and G then project to the inside of the shadow, while the six remaining edges project to a regular hexagon. (Since there are four body diagonals like AG, there are four ways to make such a projection. In each case, the six edges of the cube that project to the regular hexagon form a non-planar hexagon on the surface of the cube, that zig-zags its way round the polyhedron like a 'wobbly equator', turning alternately left and right each time it reaches a vertex. Such a closed circuit on a regular polyhedron is called a *Petrie polygon* – named after John Flinders Petrie (1907–1972), son of the famous Egyptologist Flinders Petrie).

II. Arithmetic

A child of the new generation
Refused to learn multiplication
He said, "Don't conclude
That I'm stupid, or rude.
I am simply without motivation."
Joel Henry Hildebrand (1881–1983)

Many important aspects of serious mathematics have their roots in the world of arithmetic. This is a world everyone can enjoy and master. In this chapter we re-visit, or maybe meet for the first time, key aspects of arithmetic that are often overlooked – ending with an introduction to the basic result on the distribution of primes.

The place of arithmetic in elementary mathematics can only be understood if one realises that, from upper primary school onwards, mathematics should no longer focus on more and more complicated calculations. Rather it moves beyond a set of procedures for grinding out answers, and should become a *structural laboratory*, where we gain insight into simple phenomena, and where we begin to appreciate how calculation can be managed, or tamed. The focus on structure leads in the main to matters which can be best expressed *algebraically*. This chapter concentrates mainly on structural aspects of number that are strictly arithmetical (e.g. related to numerals and place value), or where the relevant structural approach is "pre-algebraic" – with occasional forays into the world of algebra.

We repeat the observation that the "essence of mathematics" in the title is mostly left implicit in the problems. And while there is some discussion of this "essence" in the text between the problems, most of the relevant observations that we make are to be found in the solutions, or in the **Notes** which follow many of the solutions.

2.1. Place value and decimals: basic structure

Problem 40 Without using a calculator:

(a) Work out

(i) $12\,345\,679 \times 9$

(ii) $7 \times 9 \times 11 \times 13$.

(b) Divide

(i) $123\,123\,123$ by 123

(ii) $111\,111\,111$ by 111

(iii) $111\,111\,111$ by 37. △

Problem 41 Work out in your head (i) 11^2 (ii) 11^3 (iii) 101^2. △

Problem 42 Try to answer the following questions using only mental arithmetic:

(a)(i) What is the largest and the smallest possible number of digits in the answer when you multiply a 3-digit integer by a 5-digit integer?

(ii) What if we multiply an m-digit integer by an n-digit integer?

(b)(i) How many (base 10) digits are there in the evaluated form of 2^{20}?

(ii) Estimate $\left(\frac{1}{2}\right)^{20}$ to 6 decimal places.

(c) Can a natural number (i.e. a positive integer) be smaller than the product of its (base 10) digits?

(d) Work out how many zeros there are on the end, and work out the last non-zero digit of (i) $2^{15} \times 5^3$ (ii) $20!$. △

Problem 43 Imagine the sequence of positive integers from 1 to 60 written in a single row as the digits of a very large integer:

$$1234567891011121314151617181920212223\cdots5960.$$

You have to cross out 100 of these digits.

(a) Suppose you want to make the remaining number as *small* as possible. What number is left?

(b) Now suppose that you want to make the remaining number as *large* as possible. What number is left? △

2.2. Order and factors

Problem 44 Find the remainder when we divide

$$1111\cdots1111 \quad \text{(with 1111 digits 1)}$$

by 1111. △

Problem 45 Which of the numbers

$$\frac{100\,001}{100\,002} \text{ and } \frac{10\,000\,001}{10\,000\,002}$$

is bigger? △

Problem 46 Show that the integer

$$100\,000\,000\,003\,000\,000\,000\,000\,700\,000\,000\,021$$

is not prime. △

Problem 47 How many prime numbers are there in each of these sequences? (Can you identify infinitely many primes in either sequence? Can you identify infinitely many non-primes?)

(a) 1, 11, 111, 1111, 11 111, 111 111, 1 111 111, ...

(b) 11, 1001, 100 001, 10 000 001, ... △

2.3. Standard written algorithms

Problem 48 Use standard column arithmetic (i.e. long multiplication) to evaluate 9009×37. Why should you have foreseen the outcome? △

Problem 49 In the long division shown here, all the digits are missing.

```
           * * * * *
       ---------------
** |   * * * * * * *
         * *
         ---
           * * *
             * *
             ---
               * *
               * *
               ---
               * * *
               * * *
               -----
```

But the “shape” of the constituent numbers is clear.

Can you work out all possibilities for the two-digit divisor? △

Problem 50 (*For those readers who can write simple computer code.*) In these problems you may choose your favourite programming language, and a device of your choice.

(a) Two non-negative integers m and n are to be entered in base 10, digit by digit, via a keyboard. Write computer code to implement the standard algorithms of column arithmetic in order to output to the screen (in the same format):

(i) $m + n$

(ii) $m - n$

(iii) $m \times n$

(iv) (if n is a divisor of m) $m \div n$

(v) (if n is not a divisor of m) the integer part q of the quotient $m \div n$ and the remainder r.

(b) Repeat the challenge of part (a), but this time try to write shorter code by using recursion (or other programming tricks).

(c) Repeat the challenge of parts (a) and (b), but this time with inputs and outputs in the binary numeral system (see Section 2.8). △

2.4. Divisibility tests

An integer written in base 10:

is divisible by 10 precisely when the units digit is 0.

Because $10 = 2 \times 5$, it follows that an integer (in base 10):

> is divisible by 5 precisely when the units digits is 0 or 5 (i.e. a multiple of 5); and
>
> is divisible by 2 precisely when the units digit is 0, 2, 4, 6, or 8 (i.e. a multiple of 2).

Because $100 = 4 \times 25$, it follows that an integer:

> is divisible by 4 precisely when the integer formed by its last two digits is a multiple of 4; and
>
> is divisible by 25 precisely when its last two digits are 00, 25, 50, or 75 (that is, a multiple of 25).

Because $1000 = 8 \times 125$, it follows that an integer:

> is divisible by 8 precisely when the integer formed by its last three digits is a multiple of 8.

Hence simple tests for divisibility by 2, by 4, by 5, by 8, and by 10 all follow easily from the way we write numbers in base 10.

Problem 51

(a) Prove that, when an integer is written in base 10, the *remainder* when it is divided by 9 is equal to the *remainder* when its "digit-sum" is divided by 9. Conclude that the remainder when an integer is divided by 3 is equal to the remainder when its "digit-sum" is divided by 3.

(b) Explain why an integer is divisible by 6 precisely when it is divisible both by 2 and by 3. △

Problem 52

(a) What can you say about an integer N which is divisible by three times the sum of its base 10 digits?

(b) Find all integers which are *equal* to three times the sum of their base 10 digits.

(c) Find the smallest positive multiple of 9 with no odd digits. △

Problem 53 Prove than an integer written in base 11 is divisible by ten precisely when its digit-sum is divisible by ten. △

2.5. Sequences

We have already met

- the sequence of natural numbers (1, 2, 3, 4, 5, ...),
- the sequence of squares (1, 4, 9, 16, 25, ...),
- the sequence of cubes (1, 8, 27, 64, 125, ...),
- the sequence of prime numbers (2, 3, 5, 7, 11, 13, 17, ...),
- the sequence of powers of 2 (1, 2, 4, 8, 16, 32, ...), and the sequence of powers of 4 (1, 4, 16, 64, 256, ...).

We have also considered

- the sequence of *units* digits of the powers of 4 (1, 4, 6, 4, 6, 4, 6, ...),
- the sequence of *leading* digits of the powers of 4 (1, 4, 1, 6, 2, 1, 4, ...).

2.5.1 Triangular numbers

Problem 54

(a) Evaluate the first twelve terms of the sequence of *triangular* numbers:

$$1, 1+2, 1+2+3, 1+2+3+4, \ldots, 1+2+3+\cdots+10+11+12.$$

(b) Find and prove a formula for the n^{th} triangular number

$$T_n = 1+2+3+\cdots+n.$$

(c) Which triangular numbers are also (i) powers of 2? (ii) prime? (iii) squares? (iv) cubes? △

2.5.2 Fibonacci numbers

The Hindu-Arabic numeral system emerged in the Middle East in the 10^{th} and 11^{th} centuries. Fibonacci, also known as *Leonardo of Pisa*, is generally credited with introducing this system to Europe around 1200 – especially through his book *Liber Abaci* (1202). One of the problems in that book introduced the sequence that now bears his name.

The sequence of *Fibonacci numbers* begins with the terms $F_0 = 0$, $F_1 = 1$, and continues via the Fibonacci recurrence relation:

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

The sequence was introduced through a curious problem about breeding rabbits; but to this day it continues to feature in many unexpected corners of mathematics and its applications.

Problem 55

(a)(i) Generate the first twelve terms of the Fibonacci sequence:

$$F_0,\ F_1,\ \ldots,\ F_{11}.$$

(ii) Use this to generate the first eleven terms of the sequence of "differences" between successive Fibonacci numbers. Then generate the first ten terms of the sequence of "differences between successive differences".

(iii) Find an expression for the m^{th} term of the k^{th} sequence of differences.

(b)(i) Generate the first twelve terms of the sequence of powers of 2:

$$2^0,\ 2^1,\ 2^2,\ \ldots,\ 2^{11}.$$

(ii) Use this to generate the first eleven terms of the sequence of "differences" between successive powers of 2. Then generate the first ten terms of the sequence of "differences between successive differences".

(iii) Find an expression for the m^{th} term of the k^{th} sequence of differences.

△

The sequence of differences between successive terms in the sequence of triangular numbers is just the sequence of natural numbers (starting with 2):

$$2,\ 3,\ 4,\ 5,\ 6,\ \ldots;$$

and the sequence of "second differences" is then *constant*:

$$1,\ 1,\ 1,\ 1,\ 1,\ \ldots.$$

The sequences of powers of 2 and the Fibonacci numbers behave very differently from this, in that taking differences reproduces something very like the initial sequence. In particular, taking differences can never lead to a *constant* sequence.

Logically the next four problems should wait until Chapter 6, where we address the delicate matter of "proof by mathematical induction". However, that would deprive us of the chance to sample the kind of surprises that lie just beneath the surface of the Fibonacci sequence, and to experience the process of fumbling our way towards a structural understanding of the apparent patterns that emerge. Of course, each time we think we have managed to *guess* what seems to be true, we face the challenge of *proof*. Those who have not yet mastered "proof by induction" are encouraged to get what they can from the solutions, and to view this as an informal introduction to ideas that will be squarely addressed in Chapter 6.

Problem 56

(a)(i) Generate the sequence of *partial sums* of the sequence of powers of 2:

$$2^0,\ 2^0+2^1,\ 2^0+2^1+2^2,\ 2^0+2^1+2^2+2^3,\ldots$$

(ii) Prove that each partial sum is 1 less than the *next* power of 2.

(b)(i) Generate the sequence of partial sums of the Fibonacci sequence:

$$F_0,\ F_0+F_1,\ F_0+F_1+F_2,\ F_0+F_1+F_2+F_3,\ldots$$

(ii) Prove that each partial sum is 1 less than the *next but one* Fibonacci number. △

Problem **56**(b) starts out with the observation that

$$F_0+F_1=F_3-1$$

which is a consequence of the first two instances of the fundamental recurrence relation

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

and derives a surprising value for the n^{th} partial sum:

$$F_0+F_1+F_2+\cdots+F_{n-1}.$$

Fibonacci numbers make their mathematical presence felt in a quiet way – partly through the almost spooky range of unexpected internal relations which they satisfy, as illustrated in Problem **56**(b) and in the next few problems.

Problem 57

(a) Note that

$$F_n^2=F_{n-0}F_{n+0}=F_n^2+(-1)^{n-1}F_0.$$

(i) Evaluate the succession of terms:

$$F_{1-1}F_{1+1},\ F_{2-1}F_{2+1},\ F_{3-1}F_{3+1},\ F_{4-1}F_{4+1},\ \ldots.$$

(ii) Guess a simpler expression for the product $F_{n-1}F_{n+1}$. Prove your guess is correct.

(b) Let $a, b, c, d \geqslant 0$.

(i) Show that the parallelogram $OABC$ spanned by the origin O, and the points $A = (a, b)$, $C = (c, d)$ and their sum $B = (a + c, b + d)$ has area $|ad - bc|$.

(ii) Find the area of the first parallelogram in the sequence of "Fibonacci parallelograms", spanned by the origin O, and the points $A = (F_0, F_1) = (0, 1)$, $C = (F_1, F_2) = (1, 1)$.

(iii) Show that the n^{th} parallelogram $OACB$ in this sequence, spanned by the origin O, and the points $A = (F_{n-1}, F_n)$ and $B = (F_n, F_{n+1})$, and the $(n + 1)^{\text{th}}$ parallelogram $OBDC$ spanned by the origin O, and the points $B = (F_n, F_{n+1})$ and $C = (F_{n+1}, F_{n+2})$ overlap in the triangle OBC, which is exactly half of each parallelogram.

Conclude that every such parallelogram has area 1. Relate this to the conclusion of (a)(ii). △

The basic recurrence relation for Fibonacci numbers specifies the next term as the sum of two successive terms. We now consider what this implies about the sum of the *squares* of two successive terms.

Problem 58

(a) Evaluate the first few terms of the sequence

$$F_0^2 + F_1^2,\ F_1^2 + F_2^2,\ F_2^2 + F_3^2,\ \ldots.$$

(b) Guess a simpler expression for the sum $F_{n-1}^2 + F_n^2$. Prove your guess is correct. △

Problem 59

(a) Note that

$$F_0F_4 = 0 = F_2^2 - 1, \quad F_1F_5 = 5 = F_3^2 + 1.$$

(i) Evaluate the succession of terms:

$$F_{2-2}F_{2+2},\ F_{3-2}F_{3+2},\ F_{4-2}F_{4+2},\ F_{5-2}F_{5+2},\ F_{6-2}F_{6+2},\ \ldots.$$

(ii) Guess a simpler expression for the product $F_{n-2}F_{n+2}$. Prove your guess is correct.

(b)(i) Evaluate the succession of terms:

$$F_{3-3}F_{3+3},\ F_{4-3}F_{4+3},\ F_{5-3}F_{5+3},\ F_{6-3}F_{6+3},\ldots.$$

(ii) Guess a simpler expression for the product $F_{n-3}F_{n+3}$. Prove your guess is correct. △

2.6. Commutative, associative and distributive laws

In this short section we re-emphasise the shift away from blind calculation, and towards consideration of the *structure* of arithmetic, which was already implicit in Problems **7–10**, and Problems **16–17** in Chapter 1.

Problem 60 Each of two positive numbers a and b is increased by 10%.

(i) What is the percentage change of their sum $a + b$?

(ii) What is the percentage change of their product $a \times b$?

(iii) What is the percentage change in their quotient $\frac{a}{b}$? △

Problem 61 The numbers a, b, c, d, e, f are positive. How will the value of the expression

$$a \div (b \div (c \div (d \div (e \div f))))$$

change if the value of f is doubled? △

Problem 62 In Problem **17** we saw that it is no accident that the sum of entries in the 4 by 4 'multiplication table' is equal to 100.

1	2	3	**4**
2	4	6	**8**
3	6	9	**12**
4	**8**	**12**	**16**

(a) Go back to the proof that the total is equal to $(1 + 2 + 3 + 4)^2$ and see how this depends on the distributive law.

(b) The total of all entries in the multiplication square can be broken down into a succession of "reverse L-shapes", such as the one formed by the bottom row and right hand column (shown above in **bold**).

(i) Work out the subtotal in each of the four reverse L-shapes in the 4 by 4 multiplication table. What do you notice about these four subtotals?

(ii) Use the formulae for the k^{th} and $(k-1)^{\text{th}}$ triangular numbers T_k and T_{k-1} to prove that, in the n by n multiplication table, the k^{th} reverse L-shape always gives rise to a subtotal k^3.

Conclude that

$$T_n^2 = 1^3 + 2^3 + 3^3 + \cdots + n^3.$$

Hence find a simple formula for the sum C_n of the first n cubes. △

Now that we have a compact formula

- for the sum T_n of the first n positive integers, and
- for the sum C_n of the first n positive cubes,

we would naturally like to find a similar formula

- for the sum S_n of the first n squares:

$$S_n = 1^2 + 2^2 + 3^2 + \cdots + n^2$$

(that is, the sum of the entries on the leading diagonal of the n by n multiplication square).

This can be surprisingly elusive. But one way of obtaining it is to look instead for the sum of the entries in the sloping diagonal 2, 6, 12, 20, . . . *just above* the main diagonal in the n by n multiplication square.

Problem 63 Consider the n by n multiplication square.

(a) Express the r^{th} term in the sloping diagonal just above the main diagonal in terms of r. Hence show that the sum of entries in this sloping diagonal is equal to $S_{n-1} + T_{n-1}$.

(b) **Multiply by 3** each of the terms in the sloping diagonal just above the main diagonal.

(i) Guess a formula for the successive sums of these terms (6, 6 + 18, 6 + 18 + 36, . . .), and prove that your formula is correct.

(ii) Hence derive a formula for the sum S_n of the first n squares. △

2.7. Infinite decimal expansions

The standard written algorithms for calculating with integers extend naturally to *terminating* decimals. But how is one supposed to calculate *exactly* with decimals that go on for ever?

Problem 64 The decimals listed here all continue forever, recurring in the expected way. Calculate:

(a) $0.55555\cdots + 0.66666\cdots =$

(b) $0.99999\cdots + 0.11111\cdots =$

(c) $1.11111\cdots - 0.22222\cdots =$

(d) $0.33333\cdots \times 0.66666\cdots =$

(e) $1.22222\cdots \times 0.818181\cdots =$ △

Problem 65

(a) Show that any decimal $b_n b_{n-1} \cdots b_0 . b_{-1} b_{-2} \cdots b_{-k}$ that terminates can be written as a fraction with denominator a power of 10.

(b) Show that any fraction that is equivalent to a fraction with denominator a power of 10 has a decimal that terminates.

(c) Conclude that a fraction $\frac{p}{q}$, for which $HCF(p,q) = 1$, has a decimal that terminates precisely when q divides some power of 10 (that is, when $q = 2^a \times 5^b$ for some non-negative integers a, b).

(d) Prove that any fraction $\frac{p}{q}$, for which $HCF(p,q) = 1$, and where q is not of the form $q = 2^a \times 5^b$, has a decimal which recurs, with a recurring block of length at most $q - 1$.

(e) Prove that any decimal which recurs is the decimal of some fraction. △

Problem 66

(a) Find the fraction equivalent to each of these recurring decimals:

(i) $0.037037037\cdots$

(ii) $0.370370370\cdots$

(iii) $0.703703703\cdots$

(b) Let a, b, c be digits ($0 \leqslant a, b, c \leqslant 9$).

(i) Write the recurring decimal "$0.aaaaa\cdots$" as a fraction.

(ii) Write the recurring decimal "$0.ababababab\cdots$" as a fraction.

(iii) Write the recurring decimal "$0.abcabcabcabcabc\cdots$" as a fraction. △

Problem 67 Find the lengths of the recurring blocks for:

(a) $\frac{1}{6}, \frac{5}{6}$

(b) $\frac{1}{7}, \frac{2}{7}, \frac{3}{7}, \frac{4}{7}, \frac{5}{7}, \frac{6}{7}$

(c) $\frac{1}{11}, \frac{2}{11}, \frac{3}{11}, \frac{4}{11}, \frac{5}{11}, \frac{6}{11}, \frac{7}{11}, \frac{8}{11}, \frac{9}{11}, \frac{10}{11}$

(d) $\frac{1}{13}, \frac{2}{13}, \frac{3}{13}, \frac{4}{13}, \frac{5}{13}, \frac{6}{13}, \frac{7}{13}, \frac{8}{13}, \frac{9}{13}, \frac{10}{13}, \frac{11}{13}, \frac{12}{13}$ △

Problem 68 Decide whether each of these numbers has a decimal that recurs. Prove each claim.

(a) $0.12345678910111213141516171819202122232425262728293031\cdots$

(b) $0.10010001000010000010000001000000010000000010000000001 0\cdots$

(c) $\sqrt{2}$ △

Problem 69 For which real numbers x is the decimal representation of x unique? △

Problem **68** raises the question as to whether one person, who has total control, can specify the digits of a decimal so as to be sure that it neither terminates nor recurs: that is, so that it represents an *irrational* number. The next problem asks whether one person can achieve the same outcome with less control over the choice of digits.

Problem 70 Players A and B specify a real number between 0 and 1. The first player A tries to make sure that the resulting number is *rational*; the second player B tries to make sure that the resulting number is *irrational*. In each of the following scenarios, decide whether either player has a strategy that guarantees success.

(a) Can either player guarantee a "win" if the two players take turns to specify successive digits: first A chooses the entry in the first decimal place, then B chooses the entry in the second decimal place, then A chooses the entry in the third decimal place, and so on?

(b) Can either player guarantee a win if A chooses the digits to go in the odd-numbered places, and (entirely separately) B chooses the digits to go in the even-numbered places?

(c) What if A chooses the digits that go in almost all the places, but allows B to choose the digits that are to go in a sparse infinite collection of decimal places (e.g. the prime-numbered positions; or the positions numbered by the powers of 2; or ...)?

(d) What if A controls the choice of all but a finite number of decimal digits? △

2.8. The binary numeral system

There are all sorts of reasons why one should give thought to numeral systems using bases different from the familiar base 10. This is especially true of base 2, which is the simplest system of all, and is also (in some sense) the most widely used. What follows is only intended to offer a restricted glimpse into this alternative universe.

Problem 71 The numbers in this item are all written in base 2.

(a) Carry out the addition

$$\begin{array}{r} 11100 \\ +\ \ 1110 \\ \hline \end{array}$$

without changing the numbers into their base 10 equivalents – simply by applying the rules for base 2 column addition and "carrying".

(b) Carry out these long multiplications without changing the numbers into their base 10 equivalents – simply by applying the rules for base 2 column multiplication.

$$\text{(i)}\ \begin{array}{r} 10110 \\ \times \quad\ 10 \\ \hline \end{array} \qquad \text{(ii)}\ \begin{array}{r} 1110 \\ \times \quad 11 \\ \hline \end{array} \qquad \text{(iii)}\ \begin{array}{r} 110 \\ \times 111 \\ \hline \end{array}$$

(c) Try to add these fractions (where the numerators and denominators are numerals written in base 2) without changing the fractions into more familiar base 10 form.

$$\frac{110}{1111} + \frac{1}{10} + \frac{1001}{1110}$$ △

The next problem invites you to devise divisibility tests for integers written in base 2 like those for base 10 (that is, tests which implement some check involving the base 2 digits in place of carrying out the actual division).

Problem 72 Let N be a positive integer written in base 2. Describe and justify a simple test, based on the digits of $N_{\text{base}2}$:

(i) for N to be divisible by 2

(ii) for N to be divisible by 3

(iii) for N to be divisible by 4

(iv) for N to be divisible by 5. △

Problem 73 A mathematical merchant has a pair of scales and an infinite set of calibrated integral weights with values $w_0, w_1, w_2, \ldots$ (where $w_0 < w_1 < w_2 < \ldots$), but with only one weight of each value.

(a) Suppose that, for each object of positive integer weight w whose weight is to be determined, when the object is placed in one scale pan, the merchant is able to select some combination of his weights $w_0, w_1, w_2, \ldots$ to put in the other scale pan to balance, and hence to determine the weight of, the object to be weighed.

  (i) If for each weight w there is a *unique* choice of weights w_i that balance w, prove that the collection of weights must consist of all the powers of 2.

  (ii) If every object of unknown integral weight w can be balanced by some collection of the weights w_i, but some weights w can be balanced, or "represented", in more than one way, is it true that the merchant's collection of weights has to *include* all the powers of 2?

(b) What can you prove if the merchant's set of weights allow him to balance every unknown integer weight w in exactly one way by varying his weighing procedure, so that he can place his "known weights" *in either scale pan* (either in the same scale pan as the unknown weight to add to its weight, or in the opposite scale pan to balance it)? △

Problem 74 Explain how to express any fraction

$$\frac{m}{2^n}$$

where $0 < m < 2^n$ as a sum of distinct *unit* fractions with denominator a power of 2. △

You may have heard of an algorithm (a bit like long division) which allows one to compute by hand the *square root* of any number N given in base 10. The algorithm starts by grouping the digits of N in pairs, starting from the decimal point. It then extracts the square root, digit by digit, with the square root having one digit for each successive pair of digits of N, starting with the left-most pair (which may be a single digit).

We all know how to start the process. For example, if the left-most pair of digits in N is "12", then we know that the square root starts with a "3". Successive digits are then identified using the algebraic identity

$$N = (x + y)^2 = x^2 + 2xy + y^2,$$

where x is the sequence of leading digits in the "partial square root" extracted so far (followed by an appropriate string of 0s), and y stands for the residual part of the required square root.

The key is to concentrate each time on the leading digit Y of the residue "$N - x^2$", and at each stage to choose the leading digit Y of y so that $2xy + y^2$ does not exceed $N - x^2$. This sequence of steps is traditionally (and helpfully) laid out in much the same way as long division, where at each stage we subtract the square of the current approximate square root x, from the original number N, and "bring down" the next pair of digits, and then choose the next digit Y in the square root (the leading digit of y) so that "$2xy + y^2$" does not exceed the residue $N - x^2$.

In base 10 each stage requires one to juggle possibilities to decide on the next digit in the partial square root. However, in base 2 the process should be simpler, since at each stage we only have to decide whether the next digit is a 1 or a 0.

Problem 75 Work out how to calculate the square root of any square given in *base* 2. △

2.9. The Prime Number Theorem

We have already observed that there are 4 primes less than 10, 25 primes less than 100, and 168 primes less than 1000. And there are 78 498 primes less than 10^6. So

> 40% of integers < 10 are prime;
> 25% of integers < 100 are prime;
> 16.8% of integers < 1000 are prime; and
> 7.8498% of integers $< 10^6$ are prime.

In other words, the fraction of integers which are prime numbers diminishes as we go up.

The first question to ask is whether prime numbers themselves "run out" at some stage, or whether they go on for ever. The answer is very like that for the counting numbers, or positive integers $1, 2, 3, 4, 5, \ldots$:

> the counting process certainly gets started (with 1); and
> no matter how far we go, we can always "add 1" to get a larger counting number.

Hence we conclude that the counting numbers "go on for ever".

Problem 76

(a)(i) Start the process of generating prime numbers by choosing your favourite small prime number and call it p_1.

(ii) Then define $n_1 = p_1 + 1$.

(b)(i) Since $n_1 > 1$, n_1 must be divisible by some prime. Explain why p_1 is not a factor of n_1. (What is the *remainder* when we divide n_1 by p_1?)

(ii) Let p_2 be the *smallest* prime factor of n_1.

(iii) Define $n_2 = p_1 \times p_2 + 1$

(c)(i) Since $n_2 > 1$, n_2 must be divisible by some prime. Explain why p_1 and p_2 are not factors of n_2. (What is the remainder when we divide n_1 by p_1, or by p_2?)

(ii) Let p_3 be the *smallest* prime factor of n_2.

(iii) Define $n_3 = p_1 \times p_2 \times p_3 + 1$

(d) Suppose we have constructed k distinct prime numbers $p_1, p_2, p_3, \ldots, p_k$. Explain how we can always construct a prime number p_{k+1} different from $p_1, p_2, \ldots, p_k$.

(e) Apply the above process with $p_1 = 2$ to find p_2, p_3, p_4, p_5. △

Once we know that the prime numbers go on for ever, we would like to have a clearer idea as to the *frequency* with which prime numbers occur among the positive integers. We have already noted that

- there are 4 primes between 1 and 10,
- and again 4 primes between 10 and 20;
- but there is only 1 prime in the 90s;
- and then 4 primes between 100 and 110.
- And there are *no primes at all* between 200 and 210.

In other words, the distribution of prime numbers seems to be fairly chaotic. Our understanding of the full picture remains fragmentary, but we are about to see that the apparent chaos in the distribution of prime numbers conceals a remarkable pattern just below the surface.

The next item is only an experiment; but it is a very suggestive experiment. It is artificial, in that what you are invited to count has been carefully chosen to point you in the right direction. The resulting observation is generally referred to as the *Prime Number Theorem*. The result was conjectured by Legendre (1752–1833) and by Gauss (1777–1855) in the late 1790s – and was proved 100 years later (independently and almost simultaneously) in 1896 by the French mathematician Hadamard (1865–1963) and by the Belgian mathematician de la Vallée Poussin (1866–1962). You will need to access a list of prime numbers up to 5000 say.

Problem 77 Let $\pi(x)$ denote the number of prime numbers $\leqslant x$: so $\pi(1) = 0$, $\pi(2) = 1$, $\pi(3) = \pi(4) = 2$, $\pi(100) = 25$. You are invited to count the number of primes up to certain carefully chosen numbers, and then to study the results. The pattern you should notice works just as well for other numbers – but is considerably harder to spot.

The special values we choose for "x" are

the next integer above successive powers of the special number e,

where e is an important constant in mathematics – an irrational number whose decimal begins $2.7182818\cdots$, and which has its own button on most calculators (see Problem **248**).

(a) Complete the following table.

n	e^n	next integer N	$\pi(N)$
1	$2.718\cdots$	3	2
2	$7.389\cdots$		
3	$20.08\cdots$		
4	$54.59\cdots$		
5	$148.41\cdots$		
6	$403.42\cdots$		
7	$1096.63\cdots$		
8	$2980.95\cdots$		
9	$8103.08\cdots$		1019

(b) Find an expression that seems to specify $\pi(N)$ as a function of n. Hence conjecture an expression for $\pi(x)$ in terms of x. △

Durch planmässiges Tattonieren.
[Through systematic fumbling.]
Carl Friedrich Gauss (1777–1855),
when asked how he came to make so many
profound discoveries in mathematics.

2.10. Chapter 2: Comments and solutions

40.

(a) (i) 111 111 111

(ii) 9009 ($1001 = 7 \times 11 \times 13$ is a factorisation that is worth remembering for all sorts of reasons: for example, it incorporates $91 = 7 \times 13$; and it lies behind certain tests for divisibility by 7).

(b) (i) 1 001 001; (ii) 1 001 001; (iii) 3 003 003 (since $111 = 3 \times 37$)

41.

(i) $(10+1)^2 = 10^2 + 2 \times 10 + 1^2 = 121$;

(ii) $(10+1)^3 = 10^3 + 3 \times 10^2 + 3 \times 10 + 1^3 = 1331$;

(iii) $(100+1)^2 = 100^2 + 2 \times 100 + 1^2 = 10\,000 + 200 + 1 = 10\,201$

42.

(a) (i) Largest 8, smallest 7. (The smallest 3-digit number is 100 and the smallest 5-digit number is 10 000, so the smallest possible product is $10^2 \times 10^4 = 10^6$ – and so has 7 digits. The largest 3-digit number is just less than 1000 and the largest 5-digit number is just less than 100 000, so the largest possible product is just less than $10^3 \times 10^5 = 10^8$ – and so has 8 digits.)

(ii) Largest $m+n$, smallest $m+n-1$. (The smallest m-digit number is 10^{m-1} and the smallest n-digit number is 10^{n-1}, so the smallest possible product is 10^{m+n-2} – and so has $m+n-1$ digits. The largest m-digit number is just less than 10^m and the largest n-digit number is just less than 10^n, so the largest possible product is just less than $10^m \times 10^n = 10^{m+n}$ – and so has $m+n$ digits.)

(b) (i) $2^{10} = 1024$ is very slightly larger than 10^3. Hence $2^{20} = 1024^2$ is very slightly larger than 10^6, so has 7 digits.

(ii) 2^{20} is very slightly larger than 10^6. In fact

$$(10^3 + 24)^2 = 10^6 + 2 \times 10^3 + 24^2 = 10^6 + 2 \times 10^3 + 576 = 1\,002\,576.$$

$\left(\frac{1}{2}\right)^{20}$ is its reciprocal, so is slightly smaller than $10^{-6} = 0.000001$, so it starts with six 0s after the decimal point and rounds up to 0.000001 (to 6 d.p.).

(c) No. (It can be **equal** to the product of its digits if it has just 1 digit. If a number N has k digits, with leading digit $= m$, then $N \geqslant m \times 10^{k-1}$, but the product of its digits is at most $m \times 9^{k-1}$.)

(d)(i) 3, 6. ($2^{15} \times 5^3 = 2^{12} \times 10^3 = 4096 \times 10^3$)

(ii) 4, 4. (Most of us will need some rough work to supplement mental arithmetic here.

$$\begin{aligned} 20! &= 20 \times 19 \times 18 \times \cdots \times 2 \times 1 \\ &= 2^{18} \times 3^8 \times 5^4 \times 7^2 \times 11 \times 13 \times 17 \times 19 \\ &= 10^4 \times 2^{14} \times 3^8 \times 7^2 \times 11 \times 13 \times 17 \times 19. \end{aligned}$$

So 20! ends in 4 zeros, and its last non-zero digit is equal to the units digit of $2^{14} \times 3^8 \times 7^2 \times 11 \times 13 \times 17 \times 19$. If we work "mod 10" this is equal to the units digit of $4 \times 1 \times 9 \times 1 \times 3 \times 7 \times 9$.)

Note: The reader may notice that we have used "congruences", or "modular arithmetic" (mod 10) here and at several points in Chapter 1 (e.g. in the solutions to Problem **2**(d), Problem **13**, Problem **16**(b)).

In all these contexts one only needs to know that, if we fix the divisor n, then the *remainders* on division by n can be added and multiplied like ordinary numbers, since

$$(an + r) + (bn + s) = (a + b)n + (r + s),$$

and

$$(an + r)(bn + s) = (abn + as + br)n + rs.$$

Division is more delicate. We leave the reader to look up the details in any book on elementary number theory.

43. (a) 00 000 123 450 (b) 99 999 785 960

The initial number (12 $\cdots$ 9 10 11 $\cdots$ 59 60) has $9 + 50 \times 2 + 2 = 111$ digits. Hence we are left with a number having exactly 11 digits.

For the smallest integer, we delete digits to leave the smallest initial digits (preferably 0s).

For the largest integer, we delete digits to leave as many 9s at the front as possible (and then sort out the tail).

44.

$$11\,111\,111 = 11\,110\,000 + 1111 = 1111 \times 10\,001.$$

In much the same way

$$1111 \cdots 1111000$$

(with 1108 1s and three 0s) is exactly divisible by 1111. So the remainder is **111**.

45. Compare $(10^5 + 1)(10^7 + 2)$ and $(10^5 + 2)(10^7 + 1)$.

The second is $10^7 - 10^5$ bigger than the first, so the second fraction is bigger than the first.

46. The fact that $3 \times 7 = 21$, and the position of the zeros, suggests that we express the integer as:

$$10^{35} + 3 \times 10^{24} + 7 \times 10^{11} + 3 \times 7 = (10^{11} + 3)(10^{24} + 7).$$

Note: If you feel you should have been "given a hint", then pause for a moment. There is nothing misleading here. We have no standard techniques for analysing such large numbers. The very size of the number forces you to think whether there is anything familiar about it that you might use. And the number is so simple that the only thing that can possibly stand out is the 3, 7, and 21. The rest is up to you.

47.

(a) 11 is prime. And 111 is a multiple of 3: $111 = 3 \times 37$. You should also be able to see that 1111 is a multiple of 11: $1111 = 11 \times 101$.

It is unclear whether 11 111 is prime or not. The *Square Root Test* says that to decide, we only need to check possible prime factors up to $\sqrt{11\,111} < 107$. We can eliminate 2, 3, 5, 7, 11 mentally, with very little effort. And with a calculator, it is easy to check 13, 17, 19, 23, 29, 31, 37, 41, ... and to discover that $11\,111 = 41 \times 271$.

Clearly $111\,111 = 11 \times 10\,101 = 111 \times 1001$.

So the sequence does not look too promising. All the even-numbered terms are divisible by 11; every third term is divisible by 111 (and of course, by 3); every fourth term is divisible by 1111 (and hence by 101); and so on. So the only possible candidates for primes are the second, third, fifth, seventh, eleventh, ... terms: that is the terms in **prime** positions.

Each of these terms is equal to the second bracket in the factorisation:

$$10^p - 1 = (10 - 1)(10^{p-1} + 10^{p-2} + \cdots + 10 + 1),$$

where p is a prime number.

We have seen that $111 = 3 \times 37$, and that $11\,111 = 41 \times 271$, which is not very encouraging. The 7^{th}, 11^{th}, 13^{th}, and 17^{th} terms are also not prime. But the 19^{th} term and the 23^{rd} terms are prime.

So primes seem scarce, but 11 is **not** the only prime in the sequence.

Note: Again, if you feel the problem was misleading, then pause for a moment. Part of "the essence of mathematics" is learning that some problems have a tidy solution, while others open up a rather different agenda. The only obvious way to begin to recognise this distinction is occasionally to be left to struggle to solve something that is presented as if it were a *closed* problem (with a tidy solution), only to discover that it is messier than one expected.

(b) We have already seen that $1001 = 7 \times 11 \times 13$.

Another reason for remembering this is that it is a simple instance of the standard factorisation:

$$10^3 + 1 = (10 + 1)(10^2 - 10 + 1)$$

Because the signs in the second bracket are alternately "+" and "–', this factorisation extends to all **odd** powers of 10: for example,

$$100\,001 = 10^5 + 1 = (10 + 1)(10^4 - 10^3 + 10^2 - 10 + 1)$$

So this time, 11 **is** the only prime in the list.

Note: The missing "odd" terms

$$101, 10\,001, 1\,000\,001, 100\,000\,001, \ldots\ldots$$

are slightly different – each being of the form $x^2 + 1$.

The fact that there is an algebraic factorisation of

$$x^3 + 1 = (x + 1)(x^2 - x + 1)$$

implies that $1001 = 10^3 + 1$ has to factorise. But the lack of an algebraic factorisation of $x^2 + 1$ does not *prevent* any particular number of the form $x^2 + 1$ from factorising: for example, $3^2 + 1 = 2 \times 5$, and $5^2 + 1 = 2 \times 13$ both factorise; but $4^2 + 1$, $6^2 + 1$, and $10^2 + 1$ do not.

One may be forgiven for not knowing that $10^4 + 1 = 10\,001 = 73 \times 137$. But one should realize that

$$10^6 + 1 = 100^3 + 1 = (100 + 1)(100^2 - 100 + 1).$$

48. The prime factorisation $111 = 3 \times 37$ is worth remembering. If this is second nature, then one can do better in this problem than merely grind out the answer using long multiplication, by seeing how the output to the calculation 1001×333 simply positions "333 thousands" and "333 units" next to each other:

$3 \times 37 = 111$, so $9 \times 37 = 333$.

Hence $9009 \times 37 = 1001 \times 333 = 333\,333$.

Note: The prime factorisation of 1001 is not needed here. But it is important elsewhere.

49. The very first step shows that the leading digit of the dividend must be 1; and since "three-digit minus two-digit leaves one-digit (d say)" the divisor has a multiple in the 90s.

The very next stage again gives "three-digit minus two-digit leaves one-digit", and the remainder from the first division is now the hundreds digit, so $d = 1$. Hence the two-digit divisor has 99 as a multiple (at the first step of the long division) – so the divisor must be 11, 33, or 99.

The next division shows that the divisor has a two-digit multiple, which when subtracted from a two-digit number leaves a two-digit remainder, so the divisor cannot be 99.

The final stage shows that the divisor has a three-digit multiple, so it cannot be 11.

Hence the divisor must be 33.

50. Your solution will depend on the programming language used. We use this problem to attract the reader's attention to some not so frequently discussed issues:

- The "formal written algorithms" of arithmetic are not entirely obvious.
- Their practical use is not really "formal", it uses a number of unstated conventions. For example, it requires from the user an intuitive feel for the "data

structures" involved and starts by writing one base 10 integer under another keeping digits in the same decimal position aligned in a column (a computer scientist would call it "parsing the input").

- Base 10 integers contain different numbers of digits and shorter ones may need to be padded with zeroes (mentally, in calculations on paper, or explicitly, as may be necessary when writing code), that is, $1234 + 56$ has to be treated as $1234 + 0056$.
- Digits in the number are read and used from right to left, the opposite way to reading text. (This may be a piece of fossilised history: our decimals are Arabic, and Arabs write from right to left.)

51.

(a) This exploits the fact that

$$(10^k - 1) = (10 - 1)(10^{k-1} + 10^{k-2} + \cdots + 10 + 1),$$

and so is divisible by $(10 - 1)$ (a fact which is obvious when we write $10 - 1 = 9$, $10^2 - 1 = 99$, $10^3 - 1 = 999$, etc.). For example:

$$\begin{aligned}\mathbf{12\,345} &= \mathbf{1} \times 10^4 + \mathbf{2} \times 10^3 + \mathbf{3} \times 10^2 + \mathbf{4} \times 10 + \mathbf{5} \\ &= [\mathbf{1} \times (10^4 - 1) + \mathbf{2} \times (10^3 - 1) + \mathbf{3} \times (10^2 - 1) + \mathbf{4} \times (10 - 1)] \\ &\qquad + [\mathbf{1} + \mathbf{2} + \mathbf{3} + \mathbf{4} + \mathbf{5}] \\ &= [\text{a sum of terms, each of which is a multiple of 9}] \\ &\qquad + [\text{the sum of the digits of "}\mathbf{12\,345}\text{"}]\end{aligned}$$

If the LHS is divided by 9, the remainder from the first bracket on the RHS is zero, so the overall remainder is the same as the remainder from dividing the digit sum by 9.

Since 9 is a multiple of 3, the first bracket is exactly divisible by 3. Hence if the LHS is divided by 3, the remainder from the first bracket on the RHS is zero, and the overall remainder is the same as the remainder from dividing the digit sum by 3.

Note: If we were only interested in "divisibility by 9", then we could have managed without appealing to the algebraic factorisation

$$(10^k - 1) = (10 - 1)(10^{k-1} + 10^{k-2} + \cdots + 10 + 1),$$

since

$$10 - 1 = 9, \quad 10^2 - 1 = 99, \quad 10^3 - 1 = 999, \ldots$$

are all visibly "multiples of 9". However, the structure of the above solution extends naturally to prove that, when an integer is written in base b, the remainder on division by $b - 1$ is the same as the remainder on dividing the base b "digit sum" by $b - 1$.

(b) If an integer N is divisible by 6, then we can write $N = 6m$ for some integer m.

Hence $N = (2 \times 3)m = 2 \times (3m)$, so N is a multiple of 2; and $N = 3 \times (2m)$, so N is a multiple of 3.

If an integer N is divisible by 2, then we can write $N = 2k$ for some integer k. If N is also divisible by 3, then 3 divides exactly into $2k$. But $HCF(2,3) = 1$, so the 3 must go exactly into the second factor k, so $k = 3m$ for some integer m, and $N = 6m$ is divisible by 6.

Note: It is crucial that $HCF(2,3) = 1$. (E.g. 12 is divisible by 6 and by 4; but 12 is **not** divisible by 6×4.)

52.

(a) N is divisible by 3. Hence its digit-sum is divisible by 3.
But then "three times the sum of its digits" is a multiple of 9: hence the integer is divisible by 9, and so the sum of its digits is divisible by 9.
But then it is divisible by "three time a multiple of 9" – that is divisible by 27. So $N = 27$, or 54, or 81, or 108, or .... (However, you soon come to the first multiple of 27 that is **not** "divisible by 3 times the some of its digits".)

(b) 27. (Suppose the integer N has k digits. Then $N \geqslant 10^{k-1}$, and its digit-sum is at most $9k$. If N is equal to "three times the sum of its digits", then $10^{k-1} \leqslant N \leqslant 27k$ which means $k \leqslant 2$. And from part (a) we know that N is a multiple of 27.)

(c) 288. (If the digit sum is equal to 9 (or any odd multiple of 9), then at least one digit must be odd; so we only need to worry about integers with digit-sum equal to 18, or 36, or .... The only such multiple of 9 up to 100 is 99. All multiples of 9 between 100 and 200 have an odd hundreds digit. In the 200s, the first integer with digit-sum 18 is 279 – with two odd digits. The next is 288.)

53. (a) This exploits the fact that

$$(11^k - 1) = (11 - 1)(11^{k-1} + 11^{k-2} + \cdots + 11 + 1),$$

and so is divisible by $(11 - 1)$ – a fact which is obvious if we introduce a new digit X in base 11 to stand for "ten", and then notice that

$$11 - 1 = X_{\text{base 11}},\ 11^2 - 1 = XX_{\text{base 11}},\ 11^3 - 1 = XXX_{\text{base 11}},\ \text{etc.}$$

For example:

$$\begin{aligned}
\mathbf{12\,345}_{\text{base 11}} &= \mathbf{1} \times 11^4 + \mathbf{2} \times 11^3 + \mathbf{3} \times 11^2 + \mathbf{4} \times 11 + \mathbf{5} \\
&= [\mathbf{1} \times (11^4 - 1) + \mathbf{2} \times (11^3 - 1) + \mathbf{3} \times (11^2 - 1) + \mathbf{4} \times (11 - 1)] \\
&\qquad + [\mathbf{1} + \mathbf{2} + \mathbf{3} + \mathbf{4} + \mathbf{5}] \\
&= [\text{a sum of terms, each of which is a multiple of ten}] \\
&\qquad + [\text{the sum of the digits of "}\mathbf{12\,345}\text{"}]
\end{aligned}$$

If the LHS is divided by ten, the remainder from the first bracket on the RHS is zero, so the overall remainder is the same as the remainder from dividing the digit sum by ten.

54.

(a) 1, 3, 6, 10, 15, 21, 28, 36, 45, 55, 66, 78.

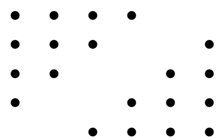
(b) Combine two copies of the required sum. If we do this algebraically, we get

$$\begin{array}{ccccccccc} 1 & + & 2 & + & 3 & + & \cdots & + & n \\ n & + & n-1 & + & n-2 & + & \cdots & + & 1 \end{array}$$

and observe that each of the n vertically aligned columns adds to $n+1$.
Hence

$$T_n = 1 + 2 + 3 + \cdots + n = \frac{n(n+1)}{2}.$$

If we do the same geometrically, then we can combine two "staircases"



of dots (one of which is inverted) into an n by $n+1$ array of dots (either with n columns and $n+1$ dots in each column, or with $n+1$ columns and n dots in each column).

Note: The n^{th} triangular number is defined by the "formula"

$$T_n = 1 + 2 + 3 + \cdots + n.$$

But this "formula" has serious limitations: in particular, there is no way to calculate T_{100} without first calculating T_1, then T_2, then T_3, ... all the way up to T_{99}. Hence it is just a "recurrence relation", which tells us how to find T_n once we know T_{n-1} (just "add n").
The formula

$$T_n = \frac{n(n+1)}{2}$$

derived in part (b) is much more useful, in that it allows us to work out the value of T_n as soon as we know the value of n. This is what we call a "**closed** formula". (The language may seem strange, but it refers to the fact that the calculation is direct, and that the formula involves a small, fixed number of operations – whereas using the recurrence requires more and more work as n gets larger.)

(c) **Note:** There are two reasons why these questions are worth asking. The first is that whenever we focus attention on certain special classes of objects, it is always good practice to consider whether the notions we have defined are completely separate, and to try to identify any overlaps. The second reason is less obvious, but can be surprisingly fruitful: sometimes two ideas may be interesting, yet have nothing to do with each other; but at other times, the two ideas may not only be interesting in their own right, but may "combine" in a way that gives rise to surprising subtleties. Here two of the combinations are routine and uninteresting; but two combinations generate more interesting mathematics than we have a right to expect.

(i) We know that one of the two factors n and $n+1$ in the numerator is odd, and the other is even. If the triangular number

$$T_n = \frac{n(n+1)}{2}$$

is to be a **power of 2**, then any odd factor of T_n must be equal to 1, so $n < 3$: $n = 2$ does not give a power of 2. Hence $n = 1$, and $T_n = 1$ is the only triangular number which is also a power of 2.

(ii) If the triangular number T_n is to be **prime**, then either

- ∗ n is odd and one of n, $\frac{n+1}{2}$ is equal to 1 (so $n = 1$ and $T_1 = 1$ is not prime), or
- ∗ n is even and one of $\frac{n}{2}$, $n+1$ is equal to 1, so $n = 2$, and $T_2 = 3$ is the only triangular number which is also prime.

(iii) The only immediately obvious "**square** triangular numbers T_n" are the first and the eighth – namely $T_1 = 1$ and $T_8 = 36$. But what seems obvious is rarely the whole truth. There are in fact *infinitely many* such "square triangular numbers" (e.g. $T_{49} = 1225$, $T_{288} = 41\,616$, $T_{1681} = 1\,413\,721$, ...). This is a consequence of the formula in part (b). For example:

When n is even, we notice that $a = \frac{n}{2}$ and $n+1 = 2a+1$ are integers with no common factors. We want their product to be a square. Because $HCF(a, 2a+1) = 1$, this occurs precisely when both a $(= b^2)$ and $2a+1$ $(= c^2)$ are both squares. So we see that solutions correspond to pairs of integers b, c which satisfy the *Pell* equation $c^2 = 2b^2 + 1$. Notice that $b = 2$, $c = 3$ is one solution, and that they satisfy the equation $c^2 - 2b^2 = 1$.

We have already met

$$a^2 + b^2 = (a+bi)(a-bi)$$

as the *norm* (or square of the length) of the complex number $a + bi$ (Problem **25**). In a similar way, we can "factorise"

$$c^2 - 2b^2 = (c + b\sqrt{2})(c - b\sqrt{2}).$$

So once we have one solution of the equation $c^2 - 2b^2 = 1$, we can take powers to get more solutions:

$$[(c + b\sqrt{2})^2][(c - b\sqrt{2})^2] = 1^2 = 1, \qquad \text{etc..}$$

Hence, for example,

$$(3 + 2\sqrt{2})^2 = 17 + 12\sqrt{2}$$

gives rise to the solution $b = 12$, $c = 17$ – corresponding to $a = 144$, $n = 288$. Similarly

$$(3 + 2\sqrt{2})^3 = \ldots + \ldots\sqrt{2}$$

gives rise to the solution $b = \ldots$, $c = \ldots$, corresponding to $(a = \ldots)$, $n = \ldots$.

Note: If you are not yet familiar with complex numbers, or with the idea of a *norm*, don't worry. Make a note of it as something that seems to be powerful and is worth learning. It will reappear later.

(iv) The only obvious **cube** triangular number is the first one – namely $T_1 = 1$. Basic algebra leads quickly to an equation as in part (i):

$$\frac{n(n+1)}{2} = m^3,$$

which is equivalent to

$$(2n+1)^2 - 1 = (2m)^3.$$

So $(2m)^3$ and $(2m)^3 + 1$ are consecutive integers that are both proper **powers**. Catalan (1814–1894) conjectured in 1844 that $8 = 2^3$ and $9 = 3^2$ are the only consecutive powers (other than 0 and 1). This simple-sounding conjecture was finally proved only in 2004. It follows that $T_1 = 1$ is the only triangular number that is also a cube.

55.

(a)(i) 0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89

(ii) 1, 0, 1, 1, 2, 3, 5, 8, 13, 21, 34; 1, 1, 0, 1, 1, 2, 3, 5, 8, 13

(iii) m^{th} term of k^{th} sequence of differences $= F_{m-k}$

(b)(i) 1, 2, 4, 8, 16, 32, 64, 128, 256, 512, 1024, 2048

(ii) 1, 2, 4, 8, 16, 32, 64, 128, 256, 512, 1024; 1, 2, 4, 8, 16, 32, 64, 128, 256, 512

(iii) m^{th} term of k^{th} sequence of differences $= 2^m$

56.

(a)(i) 1, 3, 7, 15, 31, 63, 127, 255, 511, 1023, ...

(ii) $x^{n+1} - 1 = (x-1)(x^n + x^{n-1} + \cdots + x + 1)$.
When $x = 2$, the first factor on the RHS $(x-1) = 1$, so

$$2^0 + 2^1 + 2^2 + \cdots + 2^n = 2^{n+1} - 1.$$

[Alternatively:

$$\begin{aligned} 2^0 + (2^0 + 2^1 + 2^2 + \cdots + 2^n) &= (2^0 + 2^0 \ [= 2^1]) + (2^1 + 2^2 + \cdots + 2^n) \\ &= (2^1 + 2^1 \ [= 2^2]) + (2^2 + 2^3 + \cdots + 2^n) \\ &= (2^2 + 2^2 \ [= 2^3]) + (2^3 + 2^4 + \cdots + 2^n) \\ &= \ldots \\ &= (2^n + 2^n) = 2^{n+1}. \end{aligned}$$

(b)(i) 0, 1, 2, 4, 7, 12, 20, 33, 54, 88, ...

(ii) $F_0 + F_1 = F_2 = F_3 - F_1$
$F_0 + F_1 + F_2 = (F_3 - F_1) + F_2 = (F_3 + F_2) - F_1 = F_4 - F_1$
$F_0 + F_1 + F_2 + F_3 = (F_4 - F_1) + F_3 = (F_4 + F_3) - F_1 = F_5 - F_1.$

Claim:

$$F_0 + F_1 + F_2 + \cdots + F_{n-1} = F_{n+1} - F_1$$

holds for all $n \geqslant 1$.

Proof: When $n = 1$, the LHS $= F_0 = 0 = 1 - 1 = F_2 - F_1 =$ RHS.
We proved the next few case $n = 2$, $n = 3$, $n = 4$ above.
Suppose we have already proved the required relation holds all the way up to the $(k-1)^{\text{th}}$ equation:

$$F_0 + F_1 + F_2 + \cdots + F_{k-1} = F_{k+1} - F_1.$$

Then the k^{th} equation follows like this:

$$\begin{aligned} (F_0 + F_1 + F_2 + \cdots + F_{k-1}) + F_k &= (F_{k+1} - F_1) + F_k \\ &= (F_{k+1} + F_k) - F_1 \\ &= F_{k+2} - F_1. \end{aligned}$$

So we have shown

* that the identity holds for the first few values, and
* that whenever we know it is true up to the $(k-1)^{\text{th}}$ identity, it also holds for the k^{th} identity.

Hence the identity holds for all $n \geqslant 1$. QED

Alternatively:

$$\begin{aligned} F_1 + (F_0 + F_1 + \cdots + F_k) &= (F_1 + F_0\ \ [= F_2]) + (F_1 + F_2 + \cdots + F_k) \\ &= (F_2 + F_1\ \ [= F_3]) + (F_2 + F_3 + \cdots + F_k) \\ &= (F_3 + F_2\ \ [= F_4]) + (F_3 + F_4 + \cdots + F_k) \\ &= \ldots \\ &= F_{k+1} + F_k = F_{k+2}. \end{aligned}$$

Note: In **56**(a)(ii) we appealed directly to the factorisation of $x^{n+1} - 1$ as though this were a "known fact" which is easy to prove. And in the "alternative" proof, we repeatedly combined "$2^k + 2^k$" to make 2^{k+1}, inserting dots "…" to indicate that this replacement operation is repeated $n + 1$ times. Both of these involved thinly veiled applications of the principle of *Mathematical Induction*, which is addressed in detail in Chapter 6. In **56**(b)(ii) we had no way of concealing the use of "proof by *Mathematical Induction*", which is likely to be lurking whenever we have

a proposition, or statement, $\mathbf{P}(n)$ involving the parameter n

and

we wish to prove the **infinite** collection of assertions:

$$\text{“}\mathbf{P}(\boldsymbol{n}) \text{ is true } \textbf{for every } \boldsymbol{n} = 1, 2, 3, \ldots \text{”}.$$

The standard way of achieving this apparent miracle of proving infinitely many things at once is:

> to check that $\mathbf{P}(1)$ holds (that is, to check that $\mathbf{P}(n)$ is true when $n = 1$);

then

> to suppose that we have checked all of the instances $\mathbf{P}(1), \mathbf{P}(2), \ldots,$ up to $\mathbf{P}(k)$ for some $k \geqslant 1$,

and

> to show that the next instance $\mathbf{P}(k+1)$ must then also be true.

We then conclude that $\mathbf{P}(\boldsymbol{n})$ **is true for all** $\boldsymbol{n} \geqslant \mathbf{1}$.

57.

(a)(i) 0, 2, 3, 10, 24, 65, 168, ...

(ii) **Guess**:

$$F_{n-1}F_{n+1} = F_n^2 + (-1)^n F_1, \text{ for all } n \geqslant 1.$$

Proof: By part (i), this identity holds for $n = 1, 2, 3, 4, 5, 6, 7$.
Suppose we have checked it as far as the k^{th} instance:

$$F_{k-1}F_{k+1} = F_k^2 + (-1)^k F_1.$$

Then the next instance follows, since

$$\begin{aligned}
F_{(k+1)-1}F_{(k+1)+1} &= F_k F_{k+2} \\
&= (F_{k+1} - F_{k-1})(F_k + F_{k+1}) \\
&= F_{k+1}^2 + (F_{k+1}F_k - F_{k-1}F_k) - F_{k-1}F_{k+1} \\
&= F_{k+1}^2 + (F_k^2 - F_{k-1}F_{k+1}) \\
&= F_{k+1}^2 + (-1)^{k+1} F_1.
\end{aligned}$$

So we have shown that the identity holds for the first few values of n, and whenever we know it is true up to the k^{th} identity, it also holds for the $(k+1)^{\text{th}}$ identity. Hence the identity holds for all $n \geqslant 1$. QED

(b)(i) We suppose that

$$\frac{b}{a} < \frac{d}{c}$$

(if the inequality is reversed, the expression for the area is multiplied by “-1”).

The lines $x = 0$, $y = 0$, $x = a + c$, $y = b + d$ form a rectangle of area $(a + c)(b + d)$, which surrounds the parallelogram.

To get from this to the area of the parallelogram, we must subtract

* the two external corner rectangles (top left, and bottom right) – each of area bc; and
* the four external right angled triangles–which fit together in pairs to make rectangles of areas ab and cd. Hence

$$\text{area}(OABC) = (a + c)(b + d) - 2bc - ab - cd = ad - bc.$$

(ii) 1

(iii) Half of the 2^{nd} parallelogram is equal to half of the 1^{st} – so both have the same area, namely 1.

Half of the 3^{rd} parallelogram is equal to half of the 2^{nd} – so they both have the same area, namely 1.

And so on. Hence the area of the n^{th} parallelogram is equal to

$$|ad - bc| = |F_{n-1}F_{n+1} - F_n^2| = 1.$$

Part (a)(ii) is more precise in that it says that $F_{n-1}F_{n+1} - F_n^2 = (-1)^n$: this says that the relative positions of the generators (a, b), (c, d) for successive Fibonacci parallelograms alternate, with first $\frac{b}{a} > \frac{d}{c}$, and then $\frac{b}{a} < \frac{d}{c}$. (In fact the gradient of successive versions of the line OA, or the ratio of successive Fibonacci numbers, converges to the *Golden Ratio* τ, with successive Fibonacci points $A = (F_{n-1}, F_n)$ alternately above and below the line with equation $y = \tau x$.)

58.

(a) 1, 2, 5, 13, 34, ...

(b) **Guess**:

$$F_{n-1}^2 + F_n^2 = F_{2n-1}.$$

Note: When part (a) gives rise unexpectedly to "the odd-numbered terms of the Fibonacci sequence", it is almost impossible to believe that this is an accident. Yet the attempt to prove that this "Guess" is correct may well prove elusive – for it is hard to see how to relate the $(n-1)^{\text{th}}$ and n^{th} terms to the $(2n-1)^{\text{th}}$ term.

One approach is to

"try to prove something stronger than what seems to be required".

Claim: For each $n \geqslant 1$, **both** of the following are true:

$$F_{n-1}^2 + F_n^2 = F_{2n-1} \text{ and } F_{n+1}^2 - F_{n-1}^2 = F_{2n}.$$

Proof: We have already checked that the first relation holds for $n = 1, 2, 3, 4, 5$. And it is easy to check that

$$\begin{aligned} F_{1+1}^2 - F_{1-1}^2 &= 1 - 0 = 1 = F_2, \\ F_{2+1}^2 - F_{2-1}^2 &= 4 - 1 = 3 = F_4, \\ F_{3+1}^2 - F_{3-1}^2 &= 9 - 1 = 8 = F_6. \end{aligned}$$

So both identities hold for the first few values of n.

Now suppose we have checked that **both** relations hold all the way up to the k^{th} pair of relations.

Then simply adding the two relations in the k^{th} pair gives the first relation of the next pair:

$$\begin{aligned} F_k^2 + F_{k+1}^2 &= (F_{k-1}^2 + F_k^2) + (F_{k+1}^2 - F_{k-1}^2) \\ &= F_{2k-1} + F_{2k} \\ &= F_{2k+1} \end{aligned}$$

To see that the second relation of the next pair also follows, consider

$$\begin{aligned} F_{k+2}^2 - F_k^2 &= (F_k + F_{k+1})^2 - F_k^2 \\ &= F_{k+1}^2 + 2F_kF_{k+1} \\ &= (F_{k+1}^2 - F_{k-1}^2) + F_{k-1}^2 + 2F_kF_{k+1} \\ &= F_{2k} + (F_{k+1} - F_k)^2 + 2F_kF_{k+1} \\ &= F_{2k} + (F_{k+1}^2 + F_k^2) \\ &= F_{2k} + F_{2k+1} \\ &= F_{2k+2}. \end{aligned}$$

So we have shown

- that the identities hold for the first few values of n, and
- that whenever we know the k^{th} pair of identities hold, the $(k+1)^{\text{th}}$ pair also hold.

Hence the two identities hold for all $n \geqslant 1$. QED

59.

(a) (i) $0, 5, 8, 26, 63, \ldots$

(ii) **Guess**: $F_{n-2}F_{n+2} = F_n^2 + (-1)^{n+1}$.

Proof: By part (i), this identity holds for $n = 2, 3, 4, 5, 6$.

Suppose we have checked it as far as the k^{th} instance:

$$F_{k-2}F_{k+2} = F_k^2 + (-1)^{k+1}.$$

Then the next instance follows using **57**, since

$$\begin{aligned} F_{(k+1)-2}F_{(k+1)+2} &= F_{k-1}F_{k+3} \\ &= F_{k-1}(F_{k+1} + F_{k+2}) \\ &= F_{k-1}F_{k+1} + F_{k-1}F_{k+2} \\ &= F_k^2 + (-1)^k + (F_{k+1} - F_k)(F_k + F_{k+1}) \\ &= (-1)^k + F_{k+1}^2. \end{aligned}$$

(b) (i) $0, 13, 21, 68, \ldots$

(ii) **Guess**:

$$F_{n-3}F_{n+3} = F_n^2 + (-1)^{n+3-1}F_3^2.$$

This suggests that we should reinterpret our previous guesses, and that the "correction terms" on the RHS:

* in Problem **57**(a) should have been written as "$(-1)^{n+0-1}F_0^2$",
* in Problem **57**(a)(ii) should have been written as "$(-1)^{n+1-1}F_1^2$", and
* in Problem **59**(a)(ii) should have been written as "$(-1)^{n+2-1}F_2^2$".

We leave the proof (or otherwise) of this conjecture as an exercise for the reader.

60.

(i) 10%

(ii) 21% – notice that

$$(1.1a)(1.1b) = (1+0.1)^2 ab = (1+0.2+0.01)ab = 1.21ab.$$

(iii) 0% – notice that

$$\frac{1.1a}{1.1b} = \frac{a}{b}.$$

61. If x is doubled in the expression "x", then the value of the expression doubles.

If y is doubled in the expression $x \div y$, then the value of the expression is halved.

If z is doubled in the expression $x \div (y \div z)$, then the bracket is halved, and the expression is doubled.

Replacing "x, y, z" by "d, e, f" we see that, if the value of f is doubled, the value of the bracket $(d \div (e \div f))$ is also doubled.

If we now take $x = b$, $y = c$, $z = (d \div (e \div f))$, then, when f is doubled, z is doubled, and the value of $(b \div (c \div (d \div (e \div f))))$ is doubled.

Hence the value of the whole expression

$$a \div (b \div (c \div (d \div (e \div f))))$$

is halved.

62.

(a) The fact that one can add the entries in any order depends on the commutative and associative laws of addition. Expressing the subtotal in the second row as $2(1+2+3+4)$ uses the distributive law. And expressing the overall sum

$$(1+2+3+4) + 2(1+2+3+4) + 3(1+2+3+4) + 4(1+2+3+4)$$

as $(1+2+3+4)^2$ uses the distributive law again.

(b)(i) $1 = 1^3$, $8 = 2^3$, $27 = 3^3$, $64 = 4^3$.

(ii) $(4+8+12+16)+(12+8+4) = 4T_4 + 4T_3$. Similarly, the k^{th} reverse L-shape has sum

$$k \cdot T_k + k \cdot T_{k-1} = \frac{1}{2}k^2(k+1) + \frac{1}{2}k^2(k-1) = k^3.$$

Hence

$$C_n = 1^3 + 2^3 + 3^3 + \cdots + n^3 = (1+2+3+\cdots+n)^2 = \frac{1}{4}\cdot n^2(n+1)^{2.}.$$

63.

(a) The terms are 1×2, 2×3, 3×4, etc,; so the r^{th} term is $r(r{+}1)$, and the last term is $(n-1)((n-1)+1)$.
The r^{th} term can be expressed as "$r^2 + r$", so the sum

$$1 \times 2 + 2 \times 3 + 3 \times 4 + \cdots + r(r+1) + \cdots + (n-1)n$$

can be expressed as

$$(1^2 + 2^2 + 3^2 + \cdots + (n-1)^2) + (1+2+3+\cdots+(n-1)) = S_{n-1} + T_{n-1}.$$

(b)(i) * $n = 2$: $6 = 1 \times 2 \times 3$.
* $n = 3$: $6 + 18 = 24 = 2 \times 3 \times 4$.
* $n = 4$: $6 + 18 + 36 = 60 = 3 \times 4 \times 5$.

Guess: $3(S_{n-1} + T_{n-1}) = (n-1)n(n+1)$.
Proof: This is true for $n = 1, 2, 3, 4$.
Suppose we have checked the claim for all values up to

$$3(S_{k-1} + T_{k-1}) = (k-1)k(k+1).$$

Then

$$\begin{aligned} 3(S_k + T_k) &= 3([S_{k-1} + k^2] + [T_{k-1} + k]) \\ &= (k-1)k(k+1) + 3k(k+1) \\ &= k(k+1)(k+2). \end{aligned}$$

Hence our guess is true for all $n \geqslant 1$.

(ii)

$$S_n + T_n = \frac{n(n+1)(n+2)}{3},$$

so

$$S_n = \frac{n(n+1)(n+2)}{3} - T_n = \frac{n(n+1)(2n+1)}{6}.$$

64. If one tries to apply the usual algorithms for decimals, then one is likely to get in something of a mess. But if we re-interpret each decimal as a fraction, then things are much easier.

(a) $\frac{5}{9} + \frac{6}{9} = \frac{11}{9} = 1.22222\cdots$.

(b) $0.99999\cdots = \frac{9}{9} = 1$; $1 + \frac{1}{9} = 1.11111\cdots$.

(c) $\frac{10}{9} - \frac{2}{9} = \frac{8}{9} = 0.88888\cdots$

(d) $\frac{1}{3} \times \frac{2}{3} = \frac{2}{9} = 0.22222\cdots$.

(e) $\frac{11}{9} \times \frac{9}{11} = 1$.

65.

(a) Such a decimal is by definition equal to the fraction with numerator

$$b_n b_{n-1} \cdots b_1 b_0 b_{-1} b_{-2} \cdots b_{-k}$$

(an integer with $n + k + 1$ decimal digits) and with and denominator 10^k.

(b) If $\frac{p}{q}$ is equivalent to a fraction with numerator

$$m = b_n b_{n-1} \cdots b_1 b_0 \text{ base 10}$$

and denominator 10^k, then m has decimal representation

$$b_n b_{n-1} \cdots b_k . b_{k-1} \cdots b_1 b_0 .$$

(c) Parts (a) and (b) show that fractions $\frac{p}{q}$ with $HCF(p, q) = 1$, whose decimals terminate are precisely those which are equivalent to fractions having denominator a power of 10: that is, those for which the denominator q is a factor of some integer of the form $10^k = 2^k \times 5^k$.

(d) If q does not divide some power of 10, then its decimal does not terminate. Hence, when carrying out the division of p by q we never get remainder 0. So the only possible remainders are $1, 2, \ldots, q - 1$. The first remainder after the decimal point is equal to $p \pmod{q}$). In the ensuing q decimal places, there are just $q - 1$ distinct possible remainders, so some remainder (say r) must occur for the second time by the q^{th} step, and the outputs (and remainders) thereafter will then be the same as they were the first time that the remainder r occurred.

(e) Suppose d has a decimal with a repeating block of length b starting in the $(k+1)^{\text{th}}$ decimal place. (e.g. $d = 1234.567890909090909090 \cdots$ has $b = 2$, $k = 4$). Then the infinite decimal tails cancel when we subtract $M = 10^b d - d$, and the difference M becomes an integer N if we multiply by 10^k: $N = M \times 10^k$. Hence $d(10^b - 1)10^k = N$, and d is equal to a fraction with denominator $(10^b - 1)10^k$.

66.

(a) (i) $\frac{1}{27}$; (ii) $\frac{10}{27}$; (iii) $\frac{19}{27}$

(b) (i) $\frac{a}{9}$; (ii) $\frac{ab}{99}$; (iii) $\frac{abc}{999}$

67.

(a) $0.166666 \cdots$ (block length 1); $0.833333 \cdots$ (block length 1)

(b) All have block length 6:

$$0.142857142857142857 \cdots;$$
$$0.285714285714285714 \cdots;$$
$$0.428571428571428571 \cdots;$$
$$0.571428571428571428 \cdots;$$
$$0.714285714285714285 \cdots;$$
$$0.857142857142857142 \cdots.$$

Note: The repeating blocks are all cyclically related: e.g. the block for $\frac{2}{7}$ is the same as for $\frac{1}{7}$, but starting at "2" instead of at "1".

(c) All have block length 2:

$$0.090909\cdots;\ 0.181818\cdots;\ 0.272727\cdots;\ 0.363636\cdots;\ 0.454545\cdots;$$
$$0.545454\cdots;\ 0.636363\cdots;\ 0.727272\cdots;\ 0.818181\cdots;\ 0.909090\cdots.$$

Note: The repeating blocks are not all cyclically the same, but fall into five pairs:

- $\frac{1}{11}$ and $\frac{10}{11}$ are cyclically related;
- as are those for $\frac{2}{11}$ and $\frac{9}{11}$;
- and those for $\frac{3}{11}$ and $\frac{8}{11}$;
- and those for $\frac{4}{11}$ and $\frac{7}{11}$;
- and those for $\frac{5}{11}$ and $\frac{6}{11}$.

(d) All have block length 6.

Note: They fall into two families of six, where each family is cyclically related :

$$\begin{aligned}
\tfrac{1}{13} &= 0.076923076923076923\cdots,\\
\tfrac{3}{13} &= 0.230769230769230769\cdots,\\
\tfrac{4}{13} &= 0.307692307692307692\cdots,\\
\tfrac{9}{13} &= 0.692307692307692307\cdots,\\
\tfrac{10}{13} &= 0.769230769230769230\cdots,\\
\tfrac{12}{13} &= 0.923076923076923076\cdots;
\end{aligned}$$

and

$$\begin{aligned}
\tfrac{2}{13} &= 0.153846153846153846\cdots;\\
\tfrac{5}{13} &= 0.384615384615384615\cdots,\\
\tfrac{6}{13} &= 0.461538461538461538\cdots,\\
\tfrac{7}{13} &= 0.538461538461538461\cdots,\\
\tfrac{8}{13} &= 0.615384615384615384\cdots,\\
\tfrac{11}{13} &= 0.846153846153846153\cdots.
\end{aligned}$$

68.

(a) Does not recur. (If it did, it would have a recurring block of length b say. But by the time the counting sequence $1, 2, 3, \ldots$ reaches 10^{2b} the decimal will contain a periodic block of $2b$ zeros, so the recurring block must consist of 0s, in which case the decimal terminates.)

(b) Does not recur. (Similar to part (a).)

(c) Does not recur. (If it did recur, then $\sqrt{2}$ would be a rational number: see Problems **267**, **268**, **270**.)

69. Claim Decimal fractions have two decimal representations. All other numbers have exactly one decimal representation.

Proof: Every "decimal fraction" (that is, any fraction which can be written with denominator a power of 10) has two representations – one that terminates and one that ends with an endless string of 9s: if the last non-zero digit of the terminating decimal is k, then the second representation of the same number is obtained by changing the "k" to "$k-1$" and following it with an endless string of 9s.

Consider an unknown number with two different decimal representations α and β. Since they are "different", α and β must differ in at least one position. Suppose the first, or left-most, position in which they differ is that in the k^{th} decimal place (corresponding to 10^{-k}), and that the two digits in that position are a_k (for α) and b_k (for β).

We may suppose that $a_k < b_k$. Then $b_k = a_k + 1$ (otherwise $b_k - a_k > 1$, and $\beta - \alpha > 10^{-k}$, so $\alpha \neq \beta$).

Moreover, since β is not larger than α, the digits following b_k must all be equal to 0, and the digits following a_k must all be equal to 9. QED

70. In case (d), A only has to choose a recurring block (such as "$55555\cdots$", or "$090909\cdots$", or "$123123123\cdots$") for his/her positions – no matter where they are. B's control terminates at some stage, after which A's recurring block guarantees that the resulting number is rational.

The other parts all offer a guaranteed strategy for B. Let the positions chosen by B be numbered

$$n_1, n_2, n_3, n_4, \ldots, n_k, \ldots.$$

Now exploit the fact that the positive rationals are *countable* – that is, can be included in a **single list**. To see this we can use Cantor's (1845–1918) diagonal enumeration

$$\frac{0}{1};\quad \frac{1}{1};\quad \frac{1}{2},\frac{2}{1};\quad \frac{1}{3},\frac{3}{1};\quad \frac{1}{4},\frac{2}{3},\frac{3}{2},\frac{4}{1};\quad \frac{1}{5},\frac{5}{1};\quad \frac{1}{6},\frac{2}{5},\frac{3}{4},\frac{4}{3},\frac{5}{2},\frac{6}{1};\quad \frac{1}{7},\ldots,$$

which lists all rationals $\frac{p}{q}$ with $HCF(p,q) = 1$

- first those with $p + q = 1$,
- then those with $p + q = 2$,
- then those with $p + q = 3$,

and so on.

All B needs to do is to make sure that the resulting decimal is not the decimal of any number in this list, and s/he can do this by choosing a digit in the n_k^{th} position which is different from the digit which the k^{th} rational in the above list has in that position. The resulting real number is then different from every number in the list – and hence must be irrational.

71.

(a) 101 010 (in each column (i) $0 + 0 = 0$, (ii) $1 + 0 = 1$, (iii) $1 + 1 =$ "0 and carry 1").

(b) (i) 1 010 100 (ii) 101 010 (iii) 101 010

(c) 2

Note: Trying to do this should make it clear how easily we confound "the fourteenth positive integer" with its familiar base 10 representation. It takes time and effort to learn to see "$14_{\text{base }10}$" as "2×7", and "$21_{\text{base }10}$" as 3×7, and hence to spot the common multiple "$42_{\text{base }10}$". In *base* 2 the same numbers evoke no such familiar echoes.

72. Let $N = (a_k a_{k-1} \cdots a_1 a_0)_{\text{base }2}$.

(i) N is divisible by 2 precisely when the units digit a_0 is equal to 0.

(ii) N is divisible by 3 precisely when the alternating sum

$$\text{“}a_0 - a_1 + a_2 - a_3 + \cdots \pm a_k\text{”}$$

is divisible by 3.

Proof

$$\begin{aligned} N &= (a_k a_{k-1} \cdots a_1 a_0)_{\text{base }2} \\ &= 2^k a_k + 2^{k-1} a_{k-1} + \cdots + 2a_1 + a_0. \end{aligned}$$

For each odd suffix m, *increase* the coefficient 2^m by 1: then

$$2^m + 1 = (2+1)(2^{m-1} - 2^{m-2} + \cdots - 2 + 1)$$

has 3 as a factor.

For each even suffix $m = 2n$, *decrease* the coefficient by 1: then

$$2^{2n} - 1 = (2^2 - 1)(2^{2n-2} + 2^{2n-4} + \cdots + 2^2 + 1)$$

has 3 as a factor.

Hence

$$\begin{aligned} N &= 2^k a_k + 2^{k-1} a_{k-1} + \cdots + 2a_1 + a_0 \\ &= (\text{multiple of } 3) + (a_0 - a_1 + a_2 - \cdots \pm a_k). \end{aligned}$$

(iii) N is divisible by 4 precisely when the last two digits a_1 and a_0 are both equal to 0.

(iv) N is divisible by 5 precisely when the alternating sum

$$\text{“}a_1 a_0\text{”} - \text{“}a_3 a_2\text{”} + \text{“}a_5 a_4\text{”} - \cdots$$

is divisible by 5.

Proof:

$$\begin{aligned} N &= (a_k a_{k-1} \cdots a_1 a_0)_{\text{base }2} \\ &= 2^k a_k + 2^{k-1} a_{k-1} + \cdots + 2a_1 + a_0 \\ &= (2a_1 + a_0) + 2^2(2a_3 + a_2) + 2^4(2a_5 + a_4) + \cdots \\ &= (2^2 + 1)(2a_3 + a_2) + (2^4 - 1)(2a_5 + a_4) + \cdots \\ &\qquad + [(2a_1 + a_0) - (2a_3 + a_2) + (2a_5 + a_4) - \cdots] \\ &= (\text{a multiple of } 5) + [\text{“}a_1 a_0\text{”} - \text{“}a_3 a_2\text{”} + \text{“}a_5 a_4\text{”} - \cdots]. \end{aligned}$$

73.

(a) To weigh an object with weight 1, we must have $w_0 = 1$.

To weigh an object with weight 2, we must have $w_1 = 2$. We can then weigh any object of weight 3, but not one of weight 4.

(i) Now assume each positive weight w can be balanced in exactly one way. Then we cannot have $w_2 = 3$, so $w_2 = 4$.

Suppose that, continuing in this way, we have deduced that $w_i = 2^i$ for each $i = 0, 1, 2, \ldots, k$.

Then the binary numeral system reveals precisely that every weight w from 0 up to

$$2^{k+1} - 1 = 1 + 2 + 2^2 + \cdots + 2^k$$

can be uniquely represented, but 2^{k+1} cannot. Hence

$$w_{k+1} = 2^{k+1}.$$

The result follows by induction.

(ii) If the representation of each integer is not unique, then the sequence

$$w_0,\ w_1,\ w_2,\ \ldots$$

need not include the powers of 2. For example, it could begin

$$1,\ 2,\ 3,\ 5,\ \ldots$$

(b) If each integer w is to be weighed in this way, then w has to be represented in the form

$$w = a_1w_1 + a_2w_2 + a_3w_3 + \cdots$$

where each coefficient $a_i = 0$ (if the weight w_i is not used to weigh w), or $= 1$ (if the weight w_i is used to balance w), or $= -1$ (if the weight w_i is used to supplement w).

If each representation is to be unique, then one can prove as in (a)(i) that the sequence of weights must be the successive powers of 3.

74. Write m in "base 2":

$$m = (a_{n-1} \cdots a_1a_0)_{\text{base } 2},$$

where each $a_k = 0$ or 1. Then

$$\frac{m}{2^n} = \frac{a_0}{2^n} + \frac{a_1}{2^{n-1}} + \cdots + \frac{a_{n-1}}{2}.$$

That is,

$$\frac{m}{2^n} = (0.a_{n-1} \cdots a_1a_0)_{\text{base } 2}.$$

75. We give an example, starting with $N = 110\,111\,001_{\text{base } 2}$.

Write N, and pair off the digits, starting at the units digit.

$$\mathbf{1} \parallel 10 \parallel 11 \parallel 10 \parallel 01$$

The left-most digit stands for 2^8, so the square root is at least 2^4 (and less than 2^5). Hence the required square root has five digits (one for each "pair" of digits of N), and starts with a 1.

Root $\quad \mathbf{1} \parallel ? \parallel ? \parallel ? \parallel ?$

[We can also see that the final units digit will have to be a "1". But this is not the time to add such information.]

Let $x = \mathbf{1}0\,000$, and subtract $x^2 = 100\,000\,000$ from N:

$$\begin{array}{l} 1\quad 00 \quad 00 \quad 00 \quad 00 \\ \;\parallel \mathbf{10} \parallel 11 \parallel 10 \parallel 01 \end{array}$$

This residue has to be equal to "$2xy + y^2$". However, as with long division, our immediate interest is in determining the **next digit** of our "partial square root".

If the next digit is a 1 (contributing 2^3), then $2xy \geqslant 2^8$, which would spill over and change the digit we have already determined. Hence the next digit is a 0.

Root $\quad \mathbf{1} \parallel \mathbf{0} \parallel ? \parallel ? \parallel ?$

So we can again let $x = \mathbf{10}\,000$ giving the same remainder, which has to be equal to "$2xy + y^2$", but this time $y < 2^3$ has at most three digits.

The remainder

$$\parallel \mathbf{10} \parallel 11 \parallel 10 \parallel 01$$

is greater than 2^7, so $y \geqslant 2^2$ and the next digit must be a "1".

Root $\quad \mathbf{1} \parallel \mathbf{0} \parallel \mathbf{1} \parallel ? \parallel ?$

Now let $x = \mathbf{10\,1}00$, and subtract $x^2 = 110\;010\;000$ from N, leaving

$$\parallel \mathbf{10} \parallel 10 \parallel 01$$

This residue has to equal $2xy + y^2$, with $x = 10\,100$.

If the next digit in the square root is 1, then $2xy \geqslant 2^6 > 101\,001 = 2xy + y^2$.

Hence the next digit is 0, and the last digit is then 1.

Hence the required square root is equal to:

Root $\quad \mathbf{1} \parallel \mathbf{0} \parallel \mathbf{1} \parallel \mathbf{0} \parallel \mathbf{1}$

76.

(b)(i) The fact that

$$n_1 = p_1 + 1$$

says that

"n_1 is equal to a multiple of p_1 with remainder = 1".

(c)(i) The fact that

$$n_2 = p_1 \times p_2 + 1$$

says that

"n_2 is equal to a multiple of p_1 with remainder = 1",

and that

"n_2 is equal to a multiple of p_2 with remainder = 1".

Hence neither p_1 nor p_2 are factors of n_2.

(d) The fact that

$$n_k = p_1 \times p_2 \times \cdots \times p_k + 1$$

says that

"n_k is equal to a multiple of p_i with remainder = 1"

for each suffix i, $1 \leqslant i \leqslant k$. Hence none of the primes $p_1, p_2, p_3, \ldots, p_k$ is a factor of n_k.

So the smallest prime factor of n_k always gives us a new prime p_{k+1}.

(e) If we start with $p_1 = 2$, then $n_1 = p_1 + 1 = 3$, so $p_2 = 3$.
Then $n_2 = p_1 \times p_2 + 1 = 7$, so $p_3 = 7$.
Then $n_3 = p_1 \times p_2 \times p_3 + 1 = 43$, so $p_4 = 43$.
Then $n_4 = p_1 \times p_2 \times p_3 \times p_4 + 1 = 1807 = 13 \times 139$, so $p_5 = 13$.

77.

(a) We write $\lceil x \rceil$ for the "first integer $\geqslant x$". Then

$$\begin{aligned}
\pi(\lceil e^1 \rceil) &= \pi(3) = 2;\\
\pi(\lceil e^2 \rceil) &= \pi(8) = 4;\\
\pi(\lceil e^3 \rceil) &= \pi(21) = 8;\\
\pi(\lceil e^4 \rceil) &= \pi(55) = 16;\\
\pi(\lceil e^5 \rceil) &= \pi(149) = 35;\\
\pi(\lceil e^6 \rceil) &= \pi(404) = 79;\\
\pi(\lceil e^7 \rceil) &= \pi(1097) = 184;\\
\pi(\lceil e^8 \rceil) &= \pi(2981) = 429;\\
\pi(\lceil e^9 \rceil) &= \pi(8104) = 1019.
\end{aligned}$$

(b) The initial "doubling" is an accident of small numbers, which soon turns into "slightly more than doubling".

The observation that should (eventually) jump out at you concerns the ratio $e^N : \pi(N)$, which seems to be surprisingly close to $N - 1$. This suggests the possible

Conjecture: $\pi(x) \sim \frac{x}{\ln(x)-1}$ (where $\ln(x) = \log_e(x)$).

III. Word Problems

All the evidence suggests that
the shapes of reality
are mathematical.
George Steiner (1929–)

The previous chapter focused on aspects of the arithmetic of *pure numbers* – mostly without any surrounding context. However, our mathematical experience does not begin with pure numbers. At school level, mathematical concepts, and the reasoning we bring to understanding and using them, have their roots in *language*. And in real life, every application of mathematics starts out with a situation which is described **in words**, and which has to be reformulated mathematically before we can begin to *calculate*, and to draw meaningful mathematical conclusions. *Word problems* play an important, if limited, role in helping students to appreciate, and to handle the subtleties involved in

the art of *using the mathematics we know*
to solve problems *given in words.*

This art of using mathematics involves two distinct – but interacting – processes, which we refer to here as "simplifying" and "recognising structure".

- To identify the mathematical heart of a problem arising in the real world, one may first have to *simplify* – that is, to side-line details that seem unimportant or irrelevant, and then simplify as much as possible *without changing the underlying problem* (e.g. by replacing some awkward feature by a different quantity which is easier to measure, or by an approximation which is easier to work with).

This "simplifying" stage is well-illustrated by the tongue-in-cheek title of the classic textbook *Consider a spherical cow ...* by John Harte (1985):

> Milk production at a dairy farm was low, so [...] a multidisciplinary team of professors was assembled. [...After] two weeks of intensive on-site investigation [...] the farmer received the write up, and opened it to read [...] "Consider a spherical cow ...".

The point to emphasise here is that the judgements needed when "simplifying" are subtle, depend on an understanding of the particular situation being modelled, and may lead to a model which at first sight seems to be counterintuitive, but which may not be as silly as it seems – and which therefore needs to be explained sensitively to non-mathematicians.

In contrast *word problems* by-pass the "simplifying" stage, and focus instead on "recognising structure": they present the solver with a problem which is already essentially mathematical, but where the inner structure is contextualised, and is described in words. All the solver has to do is to interpret the verbal description in a way that extracts the structure just beneath the surface, and to translate it into a familiar mathematical form. That is, *word problems* are designed to develop facility with the process of "recognising structure", while avoiding the complication of expecting students to make modelling judgements of the kind required by the subtler "simplifying" process.

Because *word problems* focus on the second process of "recognising structure", they tend to incorporate the relevant mathematical structure *isomorphically*. The underlying structure still needs to be identified and interpreted, but the interpretations are likely to be standard, with no need for imaginative assumptions and simplifications before the structure can be discerned. For example, if a problem in primary school refers to an unknown number of "sweets" to be "shared" between six children, then the collection of "sweets" is isomorphic to a pure number (the number of sweets); and the act of "sharing" is a thinly veiled reference to numerical division.

The story in a word problem may be a purely mathematical problem in disguise. But the art of identifying the *correspondence* between

> the *data* given in the story line, and
>
> the *mathematical entities* to which they correspond

and between

> the *actions* in the story line, and
>
> the corresponding *mathematical operations* on those mathematical entities

is non-trivial, and has to be learned the hard way. The first problem below illustrates the remarkable variety of instances of even the simplest subtraction, or difference.

As in Chapters 1 and 2 the "essence of mathematics" is to be found in the problems themselves. Some discussion of this "essence" is presented in the text between the problems; but most of the relevant observations are either to be found in the solutions (or in the **Notes** which follow many of the solutions), or are left for readers to extract for themselves.

3.1. Twenty problems which embody "3 − 1 = 2"

The answer to every one of the questions in Problem **78** is the same – at least, as a 'pure number'. The goal is therefore not to "solve" each problem, but to distinguish between, and to reflect upon, the different ways in which the very simple mathematical structure "$3 - 1 = 2$" turns out to be the relevant "model" in each case.

Problem 78

(a) I was given three apples, and then ate two of them. How many were left?

(b) A barge-pole three metres long stands upright on the bottom of the canal, with one metre protruding above the surface. How deep is the water in the canal?

(c) Tanya said: "I have three more brothers than sisters". How many more boys are there in Tanya's family than girls?

(d) How many cuts do you have to make to saw a log into three pieces?

(e) A train was due to arrive one hour ago. We are told that it is three hours late. When can we expect it to arrive?

(f) A brick and a spade weigh the same as three bricks. What is the weight of the spade?

(g) The distance between each successive pair of milestones is 1 mile. I walk from the first milestone to the third one. How far do I walk?

(h) The *arithmetic mean* (or average) of two numbers is 3. If half their difference is 1, what is the smaller number?

(i) The distance from our house to the train station is 3 km. The distance from our house to Mihnukhin's house along the same road is 1 km. What is the distance from the station to Mihnukhin's house?

(j) In one hundred years' time we will celebrate the tercentenary of our university. How many centuries ago was it founded?

(k) In still water I can swim 3 km in three hours. In the same time a log drifts 1 km downstream in the river. How many kilometres would I be able to swim in the same time travelling upstream in the same river?

(l) December 2^{nd} fell on a Sunday. How many working days preceded the first Tuesday of that month?[4]

[4] This question is historically correct. In 1946, in the Soviet Union, when these problems were formulated, Saturday was a working day.

(m) I walk with a speed of 3 km per hour. My friend is some distance ahead of me, and is walking in the same direction pushing his broken down motorbike at 1 km per hour. At what rate is the distance between us diminishing?

(n) A trench 3 km long was dug in a week by three crews of diggers, all working at the same rate as each other. How many such crews would be needed to dig a trench 1 km shorter in the same time?

(o) Moscow and Gorky are cities in adjacent time zones. What is the time in Moscow when it is 3 pm in Gorky?[5]

(p) An old 'rule-of-thumb' for anti-aircraft gunners stated that: To hit a plane from a stationary anti-aircraft gun, one should aim at a point exactly three plane's lengths ahead of the moving plane. Now suppose that the gun was actually moving in the same direction as the plane with one third of the plane's speed. At what point should the gunner aim his fire?

(q) My brother is three times as old as I am. How many times my present age was his age when I was born?

(r) I add 1 to a number and the result is a multiple of 3. What would the remainder be if I were to divide the original number by 3?

(s) It takes 1 minute for a train 1 km long to completely pass a telegraph pole by the track side. At the same speed the train passes right through a tunnel in 3 minutes. What is the length of the tunnel?

(t) Three trams operate on a two-track route, with trams travelling in one direction on one track and returning on the other track. Each tram remains a fixed distance of 3 km behind the tram in front. At a particular moment one tram is exactly 1 km away from the tram on the opposite track. How far is the third tram from its nearest neighbour? △

3.2. Some classical examples

Problem 79 Katya and her friends stand in a circle in such a way that the two neighbours of each child are of the same gender. If there are five boys in the circle, how many girls are there? △

Problem 80 How much pure water must be added to a vat containing 10 litres of 60% solution of acid to dilute it into a 20% solution of acid? △

---

[5] Gorky (now the city of Nizhny Novgorod) lies to the east of Moscow.

Problem 81 A mother is $2\frac{1}{2}$ times as old as her daughter. Six years ago the mother was 4 times as old as her daughter. How old are mother and daughter? △

Problem 82

(a) Tom takes 2 hours to complete a job. Dick takes 3 hours to complete the same job. Harry takes 4 hours to complete the same job. How long would they take to complete the job, all working together (at their own rates)?

(b) Tom and Dick take 2 hours to complete a job working together. Dick and Harry take 3 hours to complete the same job. Harry and Tom take 4 hours to complete the same job. How long would they take to complete the same job, all working together? △

Problem 83 A team of mowers had to mow two fields, one twice as large as the other. The team spent half-a-day mowing the larger field. After that the team split: one half continued working on the big field and finished it by evening; the other half worked on the smaller field, and did not finish it that day – but the remaining part was mowed by one mower in one day. How many mowers were there? △

3.3. Speed and acceleration

Problem 84 Jack and Jill went up the hill, and averaged 2 mph on the way up. They then turned round and went straight back down by the same route, this time averaging 4 mph. What was their average speed for the round trip (up and down)? △

Problem 85

(a)(i) A cycling road race requires one to complete 3 laps of a long road circuit. On the first lap I average 40 km/h; on the second lap I average 30 km/h; and on the third lap I only average 20 km/h. What is my average speed for the whole race?

(ii) I cycle for 3 hours round the track of a velodrome, averaging 40 km/h for the first hour, 30 km/h for the second hour, and 20 km/h for the final hour. What is my average speed over the whole 3 hours?

(b) Two cyclists compete in an endurance event.

(i) The first cyclist pedals at 60 km/h for half the time and then at 40 km/h for the other half. The second cyclist pedals at 60 km/h for half of the total distance and then at 40 km/h for the remaining half. Who wins?

(ii) In a two hour event, the first cyclist pedals at u km/h for the first hour and then at v km/h for the second hour. The second cyclist pedals at u km/h for half of the total distance and then at v km/h for the remaining half. Who wins?

(c)(i) Apply your argument in (b)(ii) to prove an inequality between

* the **arithmetic** mean
$$\frac{u+v}{2}$$
of two positive quantities u, v, and
* the **harmonic** mean
$$\frac{2}{\frac{1}{u}+\frac{1}{v}}.$$

(ii) Give a purely algebraic proof of your inequality in (i). △

Problem 86 A train started from a station and, moving with a constant acceleration, covered a distance of 4 km, finally reaching a speed of 72 km/hour. Find the acceleration of the train, and the time taken for the 4 km. △

Problem 87 (**Average speed of an accelerating car**) A typical car (and maybe also a typical train!) does not move with constant acceleration. Starting from a standstill, a car moves through the gears and "accelerates more quickly" in lower gears, when travelling at lower speeds, than it does in higher gears, when travelling at higher speeds. Use this empirical fact to prove that the *average speed* of a car accelerating from rest is *more than half* of its final measured speed after the acceleration. △

3.4. Hidden connections

Problem 88 Two old women set out at sunrise and each walked with a constant speed. One went from A to B, and the other went from B to A. They met at noon, and continuing without a stop, they arrived respectively at B at 4 pm and at A at 9 pm. At what time was sunrise on that day? △

Problem 89 A paddle-steamer takes five days to travel from St Louis to New Orleans, and takes seven days for the return journey. Assuming that the rate of flow of the current is constant, calculate how long it takes for a raft to drift from St Louis to New Orleans. △

Problem 90 [From Paolo dell'Abbaco's *Trattato d'aritmetica*] "From here to Florence is 60 miles, and there is one who walks it in 8 days [in one direction], another in five days [in the opposite direction]. It is asked: Departing at the same time, in how many days will they meet?" △

Problem 91 Notice that in Problem **88** sunrise occurs $t = \sqrt{4 \times 9}$ hours before noon, and that $\sqrt{4 \times 9}$ is the **geometric mean** of 4 and 9. Once this is pointed out, can you reformulate your solution to Problem **88** to solve a more general problem? △

3.5. Chapter 3: Comments and solutions

78.

(a) This is the simplest form of all: 3 are given; 2 are removed; so what remains is "3 – 2".

(b) Length is a *continuous* quantity (rather than *discrete* quantity – like apples, or sweets). So we have to perceive a line segment (partially hidden beneath the water) rather than a quantity. We know the total length of the pole, and the length of the protruding portion. We can then infer the hidden length by subtraction.

Note: This kind of "geometrical subtraction" is needed in many contexts (such as: proving the general formula

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

for the area of a triangle, or showing that the area of the parallelogram spanned by the origin and vectors (a, b), (c, d) is $|ad - bc|$, or in Euclid's *Elements*, Book I, Proposition 2). The idea can be strangely elusive.

(c) The situation here is significantly different. We start with Tanya's brothers and sisters, and finish with the related, but different, notion of "boys and girls in Tanya's family". The "3" does not represent anything specific: it is a numerical *excess* (of Tanya's brothers over her sisters). In contrast, the "1" seems to represent Tanya herself, who needs to be taken into account when we switch from the initial scenario (Tanya's brothers and sisters) to the final question about "boys and girls in Tanya's family".

(d) No doubt this can be solved by drawing a picture in which the underlying structure is only appreciated superficially. But beneath the surface, it seems to be a much more abstract representation of "$3-1=2$". The "3" certainly stands for the "three pieces". But the operation "-1" is not obviously subtracting anything.

The relevant observation is simply that, starting from one end, pieces and cuts *alternate*. So if we ignore the starting end, there must be the same number of pieces and cuts – except that if we start with a log (rather than a long tape from which we are cutting off pieces), the "last cut" is "the other end of the log", which has already been cut – so does not need to be cut again, and this obliges us to subtract 1 from the number of pieces to get the number of additional cuts.

Note: This idea arises in many settings, and is sometimes referred to as "Posts and gaps". Sometimes one has to "subtract 1" as here; at other times one has to "add 1" (e.g. when counting the number of "posts", if we are given the number of "gaps", or "fence panels").

(e) Once again we are dealing with a continuous quantity – *time*. On this occasion the problem invites us to construct a (horizontal?) diagram very like the pole and the water in (b). But this time, the origin is likely to be perceived as "now", with a time-line stretching back 1 hour (to the left?) to mark the time when the train was due, and then moving on 3 hours (to the right), passing through the origin to a point 2 hours from now.

(f) It is unclear how young children might tackle this with "bare hands". However, at some stage one would like them to see the words as evoking the powerful (and rather different) underlying image of "scales", or an imagined "equation". Once one 'sees' the two pans of a balance, with a "brick and a spade" on one side being balanced by "three bricks" on the other, one can imagine removing "1 brick" from each pan to be left with the spade on its own balanced by "$3-1=2$" bricks.

(g) This is in some ways a simpler version of "Posts and gaps". However, there is an additional step – since we are no longer merely counting the gaps, but translating this counting number into a *distance*. In this instance, if one does not pay too much attention to the extra step, both give an answer "2".

Note: The impact of the extra step (switching from discrete counting number to continuous distance) can be seen more clearly in the number of errors made when students are faced with such variations as:

> "There are ten lamp posts in my street, and they are 70 metres apart. How far is it from the first to the last?"

(h) One suspects that this superficially simple problem would prove inaccessible unless pupils have learned to represent word problems diagrammatically, or have already mastered simple algebra. The "3" and the "1" do not represent real-world entities; so one has to be prepared to mark a "3" on a number line, and to interpret "average" as indicating that the two unknown quantities lie equally spaced either side of it. "Half their difference" is then staring one in the face, and the smaller number (to the left) is clearly "$3-1$".

Note: This may look rather like the overdue train in part (e). We suggest that it is significantly different.

(i) The story line clearly adds layers of difficulty which we tend to overlook. Learning to "recognise structure" and to translate words into a form that allows one to calculate is clearly a non-trivial (and neglected) art. Distances in kilometres may convey something more active than the given "length of a barge pole" in part (b), or the reported times in part (e), even if the diagram – once constructed – is very similar (provided of course that "along the same road" is interpreted as meaning "in the same direction").

Note: Consider the following item from an authoritative international study TIMSS 2011[6] for pupils aged around 14:

> "Points A, B, and C lie in a line and B is between A and C. If $AB = 10$ cm and $BC = 5.2$ cm, what is the distance between the midpoints of AB and BC?
>
> A 2.4 cm B 2.6 cm C 5.0 cm D 7.6 cm"

The question is a multiple choice question, and the options represent different ways of failing to translate the words into a suitable diagram, or to interpret them correctly. The sampling (in around 50 countries) was done very carefully. So the different success rates in different countries (of which 5 are given below) suggest that some systems give far too little attention to helping pupils to learn the relevant underlying art:

> Russia 60%, Hungary 41%, Australia 40%, England 38%, USA 29%

(j) The story line here has a different flavour. The time-line is the reverse of the overdue train in part (e), yet the measuring in centuries may make the question less immediately accessible. It may be harder to "feel" a natural interpretation, and so success may be more dependent on a willingness to represent the given information abstractly.

(k) Up to now, all problems were either static, or involved motion in a directly accessible form. Here we meet for the first time the need to interpret the words in terms of "relative motion". I may get as far as picturing myself swimming upstream in the river (against the current); but neither the "3" nor the "1" have any direct relevance to me at that time: they have to be **imagined** (as "me swimming in still water", and "the effect of the river in slowing me down"), and then interpreted in a way that allows a simple calculation.

(l) The words need to be interpreted from a very different kind of story line: if the 2^{nd} is on a Sunday, then the "first Tuesday" must be the 4^{th}. There are therefore "3" days preceding the first Tuesday – of which just "1" (Sunday) is not a working day. All that is needed is "counting"; but the wording requires a different kind of interpretation.

---

[6] Trends in International Mathematics and Science Study, https://timssandpirls.bc.edu/timss2011/index.html

(m) This is another example of "relative velocities" – but the need for subtraction no longer arises because of travel in opposite directions. In some ways it is simpler than (k); yet the final question relates to something less tangible – namely the "rate at which the distance between us is diminishing". Before one understands relative velocities, one has to choose to focus on "what happens during each hour", where I cover 3 km and my friend covers only 1 km, with the difference "$3 - 1$" measuring nothing tangible, but being *the amount by which our separation decreases* during that hour.

(n) Here it is even more important to translate the given information about "rates" into concrete form. "In the same time" should trigger the questions: "How many crews would be needed for $(3 - 1)$ km?", which may then trigger the question: "How long a ditch could 1 crew dig *in the same time*?". Whatever approach is taken, it is worth asking "If the answer is "$3-1$", what exactly is the "3"? And what is the "1"?"

(o) This does presume a degree of fluency in "modelling" the given information (e.g. knowing that "adjacent time zones" almost always differ by 1 hour, and that the Earth's rotation is from West to East, so that the Sun "rises" first in the East). On the surface, if the "3" is interpreted as the "3" in 3 pm, then the calculation "$3 - 1$" is an *adjustment*, rather than a strict subtraction (the 3 pm and the "1 hour time difference" are not really comparable quantities with which one can do arithmetic). At a deeper level one can turn both the "3" and the "1" into comparable quantities, and so justify the arithmetic.

(p) Here we face full-on what has been lurking just below the surface of certain earlier problems (such as (n)) – namely that we are dealing with (approximate) *proportion*. We ignore marginal differences in the distance to a distant object at slightly different angles, and compare on the one hand

distances along the plane's path (measured in "plane's lengths"),

and on the other hand

the time taken by the anti-aircraft fire to reach the plane.

This comparison has to be made because of the added complication of the change in the relative velocity of the gun and the plane.

The given rule of thumb specifies the direction in which a stationary gunner should aim; and the reported (unrealistically fast, yet presumed to be steady) motion of the gun introduces a 2-dimensional (vector) version of "swimming upstream" – which suggests the expected answer "two thirds of 3 plane lengths", so that "1" of the "3 plane's lengths" is compensated by the gun's motion.

(q) A solution is again dependent on representing the given information in some form. Whether or not one uses symbols, the wording invites the solver to use "my present age (in years)" as a preferred unit, and to represent "my brother's present age" as "3" of these basic units. The "$3-1$" then represents how much older he is than I am – and hence how old he was when I was born, or "how many times my present age he was when I was born".

Note: The choice of unit may conceal the fact that the question and solution are rooted in ratio and proportion.

(r) The subtraction "3 − 1" here only makes sense in arithmetic (mod 3), where "remainder 0 (on division by 3)" and "remainder 3" are in some sense equivalent. Although the "1" in "3 − 1" may be taken to be the "1" that is added to the original number in the question, the "3" is an invented remainder – which is interchangeable with "0" when working (mod 3).

(s) In the usual answer "3 − 1", one could argue that the "1" **does** appear in the question, but that the "3" **does not**. Again we are dealing with *proportion*, where the times taken (at constant speed) are proportional to lengths, or to distances travelled. First the given length (1 km) of the train and the given time (1 minute) in relation to the "pole by the track side" gives a simple *constant of proportion* (= 1), which allows us to translate the time taken into the distance travelled (and hence to calculate speed). If we re-interpret the "endpoint of the tunnel" as being just like another "pole by the track side", then it takes 1 minute for the train to emerge from the tunnel, and hence "3 − 1 minutes" for the front buffers of the train to cover the full length of the tunnel, which is therefore "(3 − 1) km" long (given that the constant of proportionality = 1).

(t) It is not clear how to interpret the "3" and the "1" in "3 − 1" without getting one's hands dirty with the configuration described. In particular, somewhere along the line one has to interpret the "3 km" separation between trams as revealing that the total length of the track is 9 km, and hence that each of the two parallel stretches of track is 4.5 km.

The "tram on the opposite track" is travelling in the opposite direction, is 1 km away, and is "3 km ahead" (or "3 km behind"); so one of these trams is 1 km from the end of the track, and the other is on the other track and 2 km from one end (travelling in the opposite direction). There are exactly two possible configurations – each arising from the other if we reverse the direction of travel. By choosing the direction of travel (or by allowing "negative speed") we may assume that tram A is 2 km from the same end of the track and that tram B in front of it is 1 km beyond the end of the track on the opposite side. Tram C is 3 km ahead of B, and hence 4 km down that 4.5 km stretch of track (so has not yet "turned the corner"). Hence it is 1 km closer to its nearest neighbour (A) than it is to B.

79. If we ignore the first sentence, then there could be zero girls (and five boys). But the first sentence guarantees that there is at least one girl ("Katya and **her** friends"). So boys and girls must alternate, giving rise to 5 girls.

80. The problem requires a degree of "modelling" in that "60% solution of acid" suggests that the initial ratio

$$\text{“acid : water”} = 60 : 40.$$

Hence the initial 10 litres is made up of 4 litres of water and 6 litres of acid. Adding water does not change the amount of acid; so we want 6 litres to be 20% of the final mix – which must therefore be 30 litres. Hence we should add 20 litres.

81. The difference in ages is $\frac{3}{2} \times d$, where d is the daughter's age in years. Six years ago the difference was three times the daughter's age, which was then $d - 6$

years. Hence

$$3(d-6) = \frac{3}{2} \times d,$$

so $d = 12$.

82.

Note: Underpinning all such problems is the "**unitary method**", which here comes into its own. It is an essential tool, which is scarcely taught, and not sufficiently practised. (As a result many students mindlessly translate "Tom takes 2 hours" as "$T = 2$", etc..)

(a) When they all work together we need to know **not** how long each takes to do the job, but **at what rate** each contributor works.

Tom does the job in 2 hours, so works at the rate of "$\frac{1}{2}$ of a job in **1 hour**". Dick works at a rate of "$\frac{1}{3}$ of a job in **1 hour**", and Harry works at the rate of "$\frac{1}{4}$ of a job in **1 hour**".

So working together, they can manage

$$\frac{1}{2} + \frac{1}{3} + \frac{1}{4} = \frac{13}{12}$$

of a job in 1 hour.

Hence, to complete 1 job they require $\frac{12}{13}$ of an hour.

(b) As in part (a), we need to know the rate at which each man works.

Suppose that Tom completes the fraction t of a job in **1 hour**, that Dick completes the fraction d of a job in **1 hour**, and that Harry completes the fraction h of a job in **1 hour**.

Then in 1 hour, working together, they complete $(t+d+h)$ jobs; so to complete 1 job takes them

$$\frac{1}{t+d+h} \quad \text{hours.}$$

We therefore need to find "$t + d + h$".

In 1 hour, Tom and Dick together complete $t + d$ jobs. And we are told that in 2 hours they complete 1 job, so $t + d = \frac{1}{2}$. Similarly $d + h = \frac{1}{3}$, and $h + t = \frac{1}{4}$. Adding yields

$$2(t+d+h) = \frac{1}{2} + \frac{1}{3} + \frac{1}{4},$$

so

$$t + d + h = \frac{13}{24}.$$

Hence the time required for Tom, Dick and Harry to finish 1 job working together is

$$\frac{1}{t+d+h} = \frac{24}{13}$$

hours.

Note: Alternatively, one might let Tom take T hours to complete **1 job**, Dick take D hours to complete **1 job**, and Harry take H hours to complete **1 job**. Then

$$t = \frac{1}{T}, \quad d = \frac{1}{D}, \quad h = \frac{1}{H}.$$

83. Imagine the two fields as strips of equal width – with the larger field *twice as long* as the smaller one.

The large strip was completely mowed in two parts:

(i) by the whole team working for the first half day, and

(ii) by half the team working for the second half of the day.

Hence the whole team mowed *two thirds* of the large field and the half team mowed the remaining *one third*.

So the half team, who worked on the smaller field, mowed *the equivalent of one third of the larger field* – that is, *two thirds of the (half-size) smaller field*. Therefore the remaining one third of the smaller field was mowed by a single man on the second day.

The previous two thirds of the smaller field (twice as much) was mowed in half a day (half the time), so must have required 4 (= *four times as many*) men. So the whole team contained 8 mowers.

Alternatively, we may suppose that there are $2n$ mowers (since the team is said to split into two halves), and that each mower mows at the rate of "r large fields per day".

The total work done in completing the larger field is then

(i) $(2n \times r) \times \frac{1}{2}$ in the morning and

(ii) $(n \times r) \times \frac{1}{2}$ in the afternoon

where each part is equal to

$$(\text{number of men} \times \text{rate of working}) \times (\text{length of time worked}).$$

That is $\frac{3}{2}nr$. So $\frac{3}{2}nr = 1$.

The total work done on the smaller field is

(i) $(n \times r) \times \frac{1}{2}$ in the afternoon of the first day, and

(ii) $(1 \times r)$ on the second day.

That is $\frac{n+2}{2} \times r$. So $\frac{n+2}{2} \times r = \frac{1}{2}$ (since the smaller field is half the larger field). Hence $\frac{3}{2}n = n + 2$.

84. The words "average speed" often provoke an unthinking assumption that one is simply being asked to find the average of the "speed ***numbers***" given in the problem. A moment's thought should remind us that the "average speed" for a journey is **not** equal to the "average of the various speeds taken as pure numbers"; it is equal to

$$(\text{the total distance travelled}) \div (\text{the total time taken}).$$

If the distance up the hill is m miles, then the climb takes $\frac{m}{2}$ hours, and the descent takes $\frac{m}{4}$ hours. The total distance for the round trip is $2m$ miles, so Jack and Jill's average speed is

$$\frac{2m}{\frac{3m}{4}} = \frac{8}{3} \text{ mph.}$$

Note: We first meet averages for *discrete* quantities, or whole numbers, where the goal is to replace a collection of quantities, or numbers, by a single representative statistic. If n quantities contribute equally, then each contributes exactly $(\frac{1}{n})^{\text{th}}$ to the average.

One way of looking at this is to represent each of the quantities being averaged in a bar chart – as rectangles of width 1, and with height corresponding to the quantity represented. "Adding all the quantities and dividing by n" is then the same as "calculating the total area under the graph and then dividing by the total length of the interval". In other words, we have replaced the complicated bar chart by a *constant function* (or a single rectangle), which has the same domain as the bar chart, and which has the same area under it (or integral) as the more complicated bar chart.

More generally, given a function $y = f(x)$ defined for values of x in the interval $[a, b]$, its *average* $f_{[a,b]}$ (over the interval $[a, b]$) is defined to be

$$f_{[a,b]} = \frac{\int_a^b f(x)dx}{|b-a|}.$$

When we talk about "average speed", we are thinking of speed changing *as a function of time*; and the total distance covered in any given time interval $[a, b]$ is equal to the area under the graph. We want a single "average speed" $v_{[a,b]}$ (a *constant function*) that would cover the same distance in the same time as the more complicated reality of varying speed. That is,

- we consider the speed $v(t)$ as a function of time t,
- then we integrate with respect to t over the specified time interval $[a, b]$, and
- finally we divide the result by the total length $|b - a|$ of the time interval:

$$v_{[a,b]} = \frac{\int_a^b v(t)dt}{|b-a|}.$$

In Problem **84** the walking speed is misleadingly given in terms of "up" and "down" – which represent the first *half* ***distance*** travelled, and the second *half* ***distance*** travelled. The careful solver knows that s/he has to find "total distance travelled" and divide by "total time taken"; but s/he may not notice that s/he has in fact reinterpreted the given information so that speed is seen as a function of *time* (rather than of distance).

85.

(a) (i) Let the distance covered on each lap be m km. Then the first lap takes me $\frac{m}{40}$ hours; the second lap takes me $\frac{m}{30}$ hours; the third lap takes me $\frac{m}{20}$ hours. So the total time taken for the three laps is

$$\frac{m}{40} + \frac{m}{30} + \frac{m}{20} = \frac{13m}{120} \text{ hours.}$$

Hence my average speed for the race covering $3m$ km is

$$\frac{3m}{\left(\frac{13m}{120}\right)} = \frac{360}{13} \text{ km/h.}$$

Note: Alternatively, because the two factors of m in the numerator and the denominator cancel each other, this answer may be formulated as the *harmonic mean* of the given speeds:

$$\frac{3}{\left[\frac{1}{40}+\frac{1}{30}+\frac{1}{20}\right]}.$$

(ii) In the first hour I cycle 40 km; in the second hour I cycle 30 km; in the third hour I cycle 20 km. So in the three hours I cycle $40 + 30 + 20 = 90$ km. So my average speed is 30 km/h.

Note: Alternatively, as long as the three time intervals t are equal, we land up with t as a factor in both the numerator and the denominator, so these common factors cancel out, and the answer is simply the *arithmetic mean* of the given speeds:

$$\frac{20+30+40}{3}.$$

(b)(i) The second cyclist spends more time cycling at 40 km/h than at 60 km/h, so the **first** cyclist spends **more time** cycling **at the higher speed**. Hence the first cyclist wins.

(ii) Again (unless $u = v$), the first cyclist spends more time cycling at the higher speed. Hence the first cyclist wins.

(c)(i) As in part (a)(ii), the first cyclist finishes with average speed $\frac{u+v}{2}$ km/h; and as in part (a)(i) the second cyclist finishes with average speed

$$\frac{2}{\left[\frac{1}{u}+\frac{1}{v}\right]}\ \text{km/h}.$$

Hence, part (b)(ii) shows that

$$\frac{u+v}{2} \geqslant \frac{2}{\left[\frac{1}{u}+\frac{1}{v}\right]} = \frac{2uv}{u+v}.$$

(ii) If we rearrange the required inequality

$$\frac{u+v}{2} \geqslant \frac{2uv}{u+v},$$

then we see that it is equivalent to proving that $(u+v)^2 \geqslant 4uv$. This suggests that we should **start** with the universally true statement:

$$(u-v)^2 \geqslant 0 \text{ for all } u, v \geqslant 0.$$

Adding $4uv$ to both sides yields $(u+v)^2 \geqslant 4uv$.
Multiplying both sides by the non-negative quantity $\frac{1}{2(u+v)}$ then gives the required inequality.

86. The only "modelling" required here is to translate the problem using the standard equations of kinematics. For motion from rest we have

(i) $v = at$, where t is the time, a is the uniform acceleration, and v the final speed, and

(ii) $s = \frac{1}{2}at^2$, where s is the distance travelled.

There is a question as to what units we should use. For the moment we stick to measuring v in km/h as given, s in km, t in hours, and a in the (unfamiliar) units of km/h^2: so $72 = at$ and $4 = \frac{1}{2}at^2$.

Dividing the second equation by the first gives $\frac{1}{18} = \frac{1}{2}t$, so $t = \frac{1}{9}$ hours (= 400 seconds).

Substituting in the first equation gives $a = 72 \times 9$ km/h^2 (= $\frac{1}{20}$ m/sec^2).

Note: Equations (i) and (ii) can be summarised as saying that, under uniform acceleration a, the distance travelled is $s = (\frac{1}{2}at) \times t$. Hence the *average* speed for the complete journey is equal to exactly *half of the final speed* $v = at$.

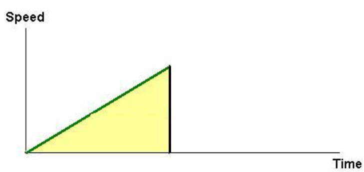
In general, those tackling the problem may agree that the familiar units of speed and distance do not give us a very good gut feeling for the scale of acceleration. If we measure acceleration in km/h^2, then we get huge numbers for acceleration which one cannot easily relate to. And if we switch to m (metres), m/sec, and m/sec^2, then we get rather small numbers for the acceleration, which again convey relatively little.

[The original (Russian) version of this problem had the train travelling 2.1 km and reaching a speed of 54 km/h. This produces a nice answer for the time taken, but a relatively inscrutable answer for the acceleration. So we have changed the parameters.]

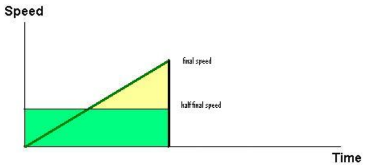
87.

‘‘*We explain why, when a vehicle accelerates from* 0 *to* 20 *mph, its average speed is* **more than** 10 *mph. In general, the average speed of an accelerating vehicle is more than half the final speed after the acceleration.*

Consider first the case when the acceleration is constant: this means that the graph which represents the speed of the vehicle as a function of time is a straight line:

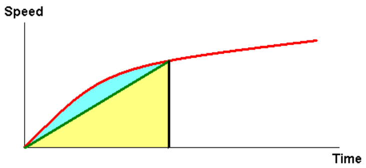


In that case, the distance travelled is equal to the area under the graph. But from the formula for the area of a triangle we know that this area equals the area of the rectangle with the same base and half the height of the triangle:



This means that the ***average speed*** *in that case is* ***exactly half*** *of the final (maximum) speed.*

But ***a car has higher acceleration in lower gears, that is, at smaller speeds.*** *Therefore the graph of speed as a function of time is concave, and the area under the graph is greater than in the case of constant acceleration. Hence, while reaching the same speed, the car travels further and its average speed is higher:*



We come to the conclusion that the average speed of an accelerating car is greater than half its speed at the end of acceleration."

Note: The text of this solution is reproduced from the appendix to a document prepared for, and submitted to, the *Crown Prosecution Service* in England. This may partly explain why it contains not a single formula. It was written by a student studying economics, and the mixture of language and graphs used illustrates the typical economist's way of thinking. Economists rarely have complete data, so they tend to rely on a combination of common sense and the basic patterns of economic variables – such as the "convexity" or "concavity" of functions. Indeed some chapters of mathematical economics could be described as outlining "the kinematics of money", and have surprising similarities to mechanics.

88. Suppose sunrise was t hours before noon – so that the first woman covers the total distance in $t + 4$ hours, while the second covers the same distance in $t + 9$ hours.

We know nothing about the distance from A to B, so it makes sense to choose this distance as our unit.

Then the first woman's speed is $\frac{1}{t+4}$, while the second woman's speed is $\frac{1}{t+9}$ units per hour.

The relative speed of A and B (the speed with which the distance between them changes) is $\frac{1}{t+4} + \frac{1}{t+9}$.

They meet at noon, so in t hours, the distance between them reduces from 1 unit to 0.

Hence

$$1 = t \times \left(\frac{1}{t+4} + \frac{1}{t+9} \right);$$

that is, $t^2 = 36$, so $t = 6$, and sunrise was at $(12 - 6) = 6$ am.

89. Let us introduce a new measure of distance – which we call a *league*. (Readers may know from old documents or from poetry that this was an old measure of distance for journeys, without knowing exactly how far it was; so we feel free to use it as an *abstract* unit *of unknown size*.)

To mesh distance and time, the journey from St Louis to New Orleans needs to be some multiple of 7, and the journey from New Orleans to St Louis needs to be some multiple of 5. Hence we choose the distance to be equal to $5 \times 7 = 35$ "*leagues*".

Then the speed of the paddle-steamer upstream is:

$$\frac{35}{7} = 5 \text{ "leagues per day"}$$

and the speed downstream is:

$$\frac{35}{5} = 7 \text{ "leagues per day"}.$$

The speed of the current gets subtracted from the speed of the paddle-steamer going upstream, and gets added to the speed of the paddle-steamer going downstream; so the speed of the current is:

$$\frac{7-5}{2} = 1 \text{ "league per day"}.$$

Hence a raft will drift from St Louis to New Orleans in $\frac{35}{1} = 35$ days.

Note: This elegant solution involves the introduction of a *hidden intermediate parameter*, an unknown quantity which helps us reason about the problem. The parameter is apparently the *distance* (from St Louis to New Orleans); but it is in fact a measure of distance chosen so as to be *compatible with the time taken.*

The art of identifying, and choosing, relevant "hidden parameters", and the analysis of their relation to the data, and their mutual relations, constitute an important and challenging part of the mathematical modelling process.

Notice that if we reformulate the problem in more general terms, with the paddle-steamer taking "a days" downstream and "b days" upstream, then the answer "d days" (for the time to drift *downstream*) happens to be the *harmonic mean* of the quantities "a" and "$-b$":

$$d = \frac{2}{\frac{1}{a} + \frac{1}{-b}}.$$

90. [This is “Problem 108” in Paolo dell’Abbaco’s *Trattato d’aritmetica* (c.1370), with a rough translation of the solution procedure given there courtesy of Roy Wagner.]

> *“Do the following: multiply* 5 *by* 8*, which makes* 40*. Then say thus: in* 40 *days one will make the trip* 8 *times, and the other* 5 *times, so both together will make the trip* 13 *times.*
>
> *Now say: if* 40 *days equals* 13 *trips, how many days are needed [on average] for one trip? And so multiply* 1 *times* 40*, which makes* 40*; then divide this by* 13*, which makes* 3 *days and* $\frac{1}{13}$ *of a day.*
>
> *And so I say that in* 3 *days and* $\frac{1}{13}$ *of a day the two will come together.*
>
> *And as this is done, so all similar problems are done.”*

Note: The problem as stated conveys an air of reality by giving the distance “from here to Florence” in miles; but this fact is not mentioned in the solution! Instead, the solution starts by introducing a hidden parameter, measured by a *dimensionless* unit: a **trip**.

This move (to invent a natural unit of measurement) also featured in Problem **89** above and has deep mathematical reasons. Problem **89** was borrowed from an interview with Vladimir Arnold (*Notices of the AMS*, vol. 44, no. 4), where we read:

> **Interviewer**: *Please tell us a little bit about your early education. Were you already interested in mathematics as a child?*
>
> **Arnold**: [...] *The first real mathematical experience I had was when our schoolteacher I. V. Morotzkin gave us the following problem* [VA then formulated Problem **89**].
>
> *I spent a whole day thinking on this oldie, and the solution (based on what are now called scaling arguments, dimensional analysis, or toric variety theory, depending on your taste) came as a revelation.*
>
> *The feeling of discovery that I had then (1949) was exactly the same as in all the subsequent much more serious problems – be it the discovery of the relation between algebraic geometry of real plane curves and four-dimensional topology (1970), or between singularities of caustics and of wave fronts and simple Lie algebras and Coxeter groups (1972). It is the greed to experience such a wonderful feeling more and more times that was, and still is, my main motivation in mathematics.*

Arnold refers here to *scaling arguments* or *dimensional analysis*: that is, the mathematical art of choosing and analysing the use of units of measurement. This has its origins in, and includes as an integral part, Euclid’s classical theory of proportion.

91. Suppose as before that the sun rises t hours before noon; but replace 4 pm (the time the woman starting at A arrived at B) by a pm, and replace 9 pm (the time the woman starting at B arrived at A) by b pm. Let C be the point where they meet (at noon).

Then, since each woman walks at a constant speed, we have

$$\frac{t}{a} = \frac{|AC|}{|CB|} \text{ (for the woman starting from } A\text{),}$$

and

$$\frac{t}{b} = \frac{|BC|}{|CA|} \text{ (for the woman starting from } B\text{).}$$

Hence

$$\frac{t}{a} = \frac{|AC|}{|CB|} = \frac{b}{t},$$

so $t^2 = ab$.

Note: This totally unexpected result validates the choice of the unknown t as the time in hours from sunrise to noon. Not knowing its significance in advance, this choice was motivated by the observation that “noon” occurs in the problem as the only common “origin”, or reference point for time data.

IV. Algebra

The first rule of intelligent tinkering
is to save all the parts.
Paul R. Ehrlich (1932–)

Many important aspects of serious mathematics have their roots in the world of arithmetic. However, when we implement an arithmetical procedure by combining numbers with very different meanings to produce *a single numerical output*, it becomes almost impossible to see how the separate ingredients contribute to the final answer. In other words, calculating exclusively with numbers contravenes Paul Ehrlich's "first rule of intelligent tinkering". This is why in Chapters 1 and 2 we stressed the need to move beyond blind calculation, and to begin to think *structurally* – even when calculating purely with numbers. *Algebra* can be seen as a remarkable way of "tinkering with numbers", so that we not only "keep all the parts", but manage to *keep them separate* (by giving them different names), and hence can see clearly what contribution each ingredient variable makes to the final output. To benefit from this feature of algebra, we need to learn to "read" algebraic expressions, and to interpret what they are telling us – in much the same way that we learn to read numbers (so that, where appropriate, 100 is seen as 10^2, and 10 is seen as $1 + 2 + 3 + 4$).

Before algebra proper was invented (around 1600), the ability to extract the general picture lying hidden inside each calculation was restricted to specialists. The ancient Babylonians (1700–1500 BC) described their general procedures as *recipes*, presented in the context of problems involving particular numbers. But they did this in such a way as to demonstrate convincingly that whoever formulated the procedure had managed to see "the general in the particular". The ancient Greeks used a geometrical setting to reveal generality, and encoded what we would see as "algebraic" methods in geometrical language. In the 9$^{\text{th}}$ century AD, Arabs such as Al-Khwarizmi (c.780–c.850), managed to encapsulate generality using a very limited kind of algebra, without the full symbolical language that would emerge later. The *abacists*, such as Paolo dell'Abbaco (1282–1374) who featured in Chapter 3, clearly saw that the power and spirit of mathematics was rooted in this generality. But modern algebraic symbolism – in particular, the idea that to express generality we need to use letters to represent not only variables, but also important *parameters* (such as the coefficients a, b, c in a general quadratic $ax^2 + bx + c$) – had to wait for the inscrutable writings of Viète (1540–1603), and especially for Fermat (1601–1665) and Descartes (1596–1650) who simplified and extended Viète's ideas in the 1630s.

Within a generation, the huge potential of this systematic use of symbols was revealed by the triumphs of Newton (1642–1727), Leibniz (1646–1716), and others in the years before 1700. Later, the refinements proposed by Euler (1707–1783) in his many writings throughout the 18^{th} century, made this new language and its discoveries accessible to us all – much as Stevin's (1548–1620) version of place value for numbers made calculation accessible to Everyman.

Our coverage of algebra is necessarily selective. We focus on a few ideas that are needed in what follows, and which should ideally be familiar – but with an emphasis that may be less familiar. When working algebraically, the key mathematical messages are mostly implicit in the manipulations themselves. Hence many of the additional comments in this chapter are to be found as part of the solutions, rather than within the main text.

4.1. Simultaneous linear equations and symmetry

Problem 92 Dad took our new baby to the clinic to be weighed. But the baby would not stay still and caused the needle on the scales to wobble. So Dad held the baby still and stood on the scales, while nurse read off their combined weight: 78kg. Then nurse held the baby, while Dad read off their combined weight: 69kg. Finally Dad held the nurse, while the baby read off their combined weight: 137kg. How heavy was the baby? △

The situation described in Problem **92** is representative of a whole class of problems, where the given information incorporates a certain *symmetry*, which the solver would be wise to respect. Hence one should hesitate before applying systematic brute force (as when using the information from one weighing to substitute for one of the three unknown weights – a move which effectively reduces the number of unknowns, but which fails to respect the symmetry in the data).

A similar situation arises in certain puzzles like the following.

Problem 93 Numbers are assigned (secretly) to the vertices of a polygon. Each edge of the polygon is then labelled with the sum of the numbers at its two end vertices.

(a) If the polygon is a triangle ABC, and the labels on the three sides are c (on AB), b (on AC), and a (on BC), what were the numbers written at each of the three vertices?

(b) If the polygon is a quadrilateral $ABCD$, and the labels on the four sides are w (on AB), x (on BC), y (on CD), and z (on DA), what numbers were written at each of the four vertices?

(c) If the polygon is a pentagon $ABCDE$, and the labels on the five sides are d (on AB), e (on BC), a (on CD), b (on DE), and c (on EA), what numbers were written at each of the five vertices? △

In case any reader is inclined to dismiss such problems as "artificial puzzles", it may help to recall two familiar instances (Problems **94** and **96**) which give rise to precisely the above situation.

Problem 94 In the triangle ABC with sides of lengths a (opposite A), b (opposite B), and c (opposite C), we want to locate the three points where the *incircle* touches the three sides – at point P (on BC), Q (on CA), and R (on AB). To this end, let the two tangents to the incircle from A (namely AQ and AR) have length x, the two tangents from B (namely BP and BR) have length y, and the two tangents from C (namely CP and CQ) have length z. Find the values of x, y, z in terms of a, b, c. △

The second instance requires us first to review the basic properties of midpoints in terms of vectors.

Problem 95

(a) Write down the coordinates of the midpoint M of the line segment joining $Y = (a, b)$ and $Z = (c, d)$. Justify your answer.

(b) Position a general triangle XYZ so that the vertex X lies at the origin $(0, 0)$. Suppose that Y then has coordinates (a, b) and Z has coordinates (c, d). Let M be the midpoint of XY, and N be the midpoint of XZ. Prove the *Midpoint Theorem*, namely that

"MN is parallel to YZ and half its length".

(c) Given any quadrilateral $ABCD$, let P be the midpoint of AB, let Q be the midpoint of BC, let R be the midpoint of CD, and let S be the midpoint of DA. Prove that $PQRS$ is always a parallelogram. △

Problem 96

(a) Suppose you know the position vectors $\mathbf{p}$, $\mathbf{q}$, $\mathbf{r}$ corresponding to the midpoints of the three sides of a triangle. Can you reconstruct the vectors $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$ corresponding to the three vertices?

(b) Suppose you know the vectors $\mathbf{p}$, $\mathbf{q}$, $\mathbf{r}$, $\mathbf{s}$ corresponding to the midpoints of the four sides of a quadrilateral. Can you reconstruct the vectors $\mathbf{w}$, $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$ corresponding to the four vertices?

(c) Suppose you know the vectors $\mathbf{p}$, $\mathbf{q}$, $\mathbf{r}$, $\mathbf{s}$, $\mathbf{t}$ corresponding to the midpoints of the five sides of a pentagon. Can you reconstruct the vectors $\mathbf{v}$, $\mathbf{w}$, $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$ corresponding to the five vertices? △

The previous five problems explore a common structural theme – namely the link between certain sums (or averages) and the original, possibly unknown, data. However this algebraic link was in every case embedded in some practical, or geometrical, context. The next few problems have been stripped of any context, leaving us free to focus on the underlying structure in a purely algebraic, or arithmetical, spirit.

Problem 97 Solve the following systems of simultaneous equations.

(a)(i) $x + y = 1, \quad y + z = 2, \quad x + z = 3$

(ii) $uv = 2, \quad vw = 4, \quad uw = 8$

(b)(i) $x + y = 2, \quad y + z = 3, \quad x + z = 4$

(ii) $uv = 6, \quad vw = 10, \quad uw = 15$

(iii) $uv = 6, \quad vw = 10, \quad uw = 30$

(iv) $uv = 4, \quad vw = 8, \quad uw = 16$ △

Problem 98 Use what you know about solving two simultaneous linear equations in two unknowns to construct the general positive solution to the system of equations:

$$u^a v^b = m, \quad u^c v^d = n.$$

Interpret your result in the language of *Cramer's Rule*. (Gabriel Cramer (1704–1752)). △

Problem 99

(a) For which values b, c does the following system of equations have a unique solution?

$$x + y + z = 3, \quad xy + yz + zx = b, \quad x^2 + y^2 + z^2 = c$$

(b) For which values a, b, c does the following system of equations have a unique solution?

$$x + y + z = a, \quad xy + yz + zx = b, \quad x^2 + y^2 + z^2 = c$$ △

4.2. Inequalities and modulus

The transition from school to university mathematics is in many ways marked by a shift from simple variables, equations and functions, to conditions and analysis involving inequalities and modulus.

Problem 100 What is $|-x|$ equal to: x or $-x$? (What if x is negative?) △

4.2.1 Geometrical interpretation of modulus, of inequalities, and of modulus inequalities

Problem 101

(a) Mark on the coordinate line all those points x in the interval $[0, 1)$ which have the digit "1" immediately after the decimal point in their decimal expansion. What fraction of the interval $[0, 1)$ have you marked?

Note: "$[0, 1)$" denotes the set of all points *between* 0 and 1, together with 0, but not including 1. $[0, 1]$ denotes the interval including *both* endpoints; and $(0, 1)$ denotes the interval *excluding* both endpoints.

(b) Mark on the interval $[0, 1)$ all those points x which have the digit "1" in *at least one* decimal place. What fraction of the interval $[0, 1)$ have you marked?

(c) Mark on the interval $[0, 1)$ all those points x which have a digit "1" in at least one position of their *base* 2 expansion. What fraction of the interval $[0, 1)$ have you marked?

(d) Mark on the interval $[0, 1)$ all those points x which have a digit "1" in at least one position of their *base* 3 expansion. What fraction of the interval $[0, 1)$ have you marked? △

Problem 102 Mark on the coordinate line all those points x for which *two* of the following inequalities are true, and *five* are false:

$$x > 1,\ x > 2,\ x > 3,\ x > 4,\ x > 5,\ x > 6,\ x > 7.$$

△

Problem 103 Mark on the coordinate line all those points x for which

$$|x - 5| = 3.$$

△

Problem 104

(a) Mark on the coordinate line all those points x for which *two* of the following inequalities are true, and *five* are false:

$$|x| > 1,\ |x| > 2,\ |x| > 3,\ |x| > 4,\ |x| > 5,\ |x| > 6,\ |x| > 7.$$

(b) Mark on the coordinate line all those points x for which *two* of the following inequalities are true, and *five* are false:

$$|x-1| > 1,\ |x-2| > 2,\ |x-3| > 3,\ |x-4| > 4,\ |x-5| > 5,\ |x-6| > 6,\ |x-7| > 7.$$

△

Problem 105 Mark on the coordinate line all those points x for which

$$|x + 1| + |x + 2| = 2.$$

△

Problem 106 Find numbers a and b with the property that the set of solutions of the inequality

$$|x - a| < b$$

consists of the interval $(-1, 2)$. △

Problem 107

(a) Mark on the coordinate plane all points (x, y) satisfying the inequality

$$|x - y| < 3.$$

(b) Mark on the coordinate plane all points (x, y) satisfying the inequality

$$|x - y + 5| < 3.$$

(c) Mark on the coordinate plane all points (x, y) satisfying the inequality

$$|x - y| < |x + y|.$$

△

4.2.2 Inequalities

Problem 108 Suppose real numbers a, b, c, d satisfy $\frac{a}{b} < \frac{c}{d}$.

(i) Prove that

$$\frac{a}{b} < \frac{\left(\frac{a}{b} + \frac{c}{d}\right)}{2} < \frac{c}{d}.$$

(ii) If $b, d > 0$, prove that

$$\frac{a}{b} < \frac{a+c}{b+d} < \frac{c}{d}.$$ △

Problem 109 (**Farey series**) When the fully cancelled fractions in $[0, 1]$ with denominator $\leqslant n$ are arranged in increasing order, the result is called the *Farey series* (or *Farey sequence*) *of order n.*

Order 1: $\frac{0}{1} < \frac{1}{1}$

Order 2: $\frac{0}{1} < \frac{1}{2} < \frac{1}{1}$

Order 3: $\frac{0}{1} < \frac{1}{3} < \frac{1}{2} < \frac{2}{3} < \frac{1}{1}$

Order 4: $\frac{0}{1} < \frac{1}{4} < \frac{1}{3} < \frac{1}{2} < \frac{2}{3} < \frac{3}{4} < \frac{1}{1}$

(a) Write down the full Farey series (or sequence) of order 7.

(b)(i) Imagine the points $0.1, 0.2, 0.3, \ldots, 0.9$ dividing the interval $[0, 1]$ into ten subintervals of length $\frac{1}{10}$. Now insert the eight points corresponding to

$$\frac{1}{9}, \frac{2}{9}, \frac{3}{9}, \ldots, \frac{8}{9}.$$

Into which of the ten subintervals do they fall?

(ii) Imagine the n points

$$\frac{1}{n+1}, \frac{2}{n+1}, \frac{3}{n+1}, \ldots, \frac{n}{n+1}$$

dividing the interval $[0, 1]$ into $n + 1$ subintervals of length $\frac{1}{n+1}$. Now insert the $n - 1$ points

$$\frac{1}{n}, \frac{2}{n}, \frac{3}{n}, \ldots, \frac{n-1}{n}.$$

Into which of the $n + 1$ subintervals do they fall?

(iii) In passing from the Farey series of order n to the Farey series of order $n+1$, we insert fractions of the form $\frac{k}{n+1}$ between certain pairs of adjacent fractions in the Farey series of order n. If $\frac{a}{b} < \frac{c}{d}$ are adjacent fractions in the Farey series of order n, prove that, when adding fractions for the Farey series of order $n+1$, **at most one** fraction is inserted between $\frac{a}{b}$ and $\frac{c}{d}$.

(c) **Note:** It is worth struggling to prove the two results in part (c). But do not be surprised if they prove to be elusive – in which case, be prepared to simply use the result in part (c)(ii) to solve part (d).

(i) In the Farey series of order n the first two fractions are $\frac{0}{1} < \frac{1}{n}$, and the last two fractions are $\frac{n-1}{n} < \frac{1}{1}$. Prove that every other adjacent pair of fractions $\frac{a}{b} < \frac{c}{d}$ in the Farey series of order n satisfies $bd > n$.

(ii) Let $\frac{a}{b} < \frac{c}{d}$ be adjacent fractions in the Farey series of order n. Prove (by induction on n) that $bc - ad = 1$.

(d) Prove that if

$$\frac{a}{b} < \frac{c}{d} < \frac{e}{f}$$

are three successive terms in any Farey series, then

$$\frac{c}{d} = \frac{a+e}{b+f}.$$

△

Problem 110 Solve the following inequalities.

(a) $x + \dfrac{1}{x} < 2$

(b) $x \leqslant 1 + \dfrac{2}{x}$

(c) $\sqrt{x} < x + \dfrac{1}{4}$

△

Problem 111

(a) The sum of two positive numbers equals 5. Can their product be equal to 7?

(b) (**Arithmetic mean, Geometric mean, Harmonic mean, Quadratic mean**) Prove that, if $a, b > 0$, then

$$\frac{2}{\left[\frac{1}{a} + \frac{1}{b}\right]} = \frac{2ab}{a+b} \leqslant \sqrt{ab} \leqslant \frac{a+b}{2} \leqslant \sqrt{\frac{a^2+b^2}{2}}$$

$$(\text{HM} \leqslant \text{GM} \leqslant \text{AM} \leqslant \text{QM})$$

△

Problem 112 The two hundred numbers

$$1,\ 2,\ 3,\ 4,\ 5,\ \ldots,\ 200$$

are written on the board. Students take turns to replace two numbers a, b from the current list by their sum divided by $\sqrt{2}$. Eventually one number is left on the board. Prove that the final number must be less than 2000. △

4.3. Factors, roots, polynomials and surds

Problem 113

(a)(i) Find a prime number which is one less than a square.
(ii) Find another such prime.

(b)(i) Find a prime number which is one more than a square.
(ii) Find another such prime.

(c)(i) Find a prime number which is one less than a cube.
(ii) Find another such prime.

(d)(i) Find a prime number which is one more than a cube.
(ii) Find another such prime. △

Problem 114 Factorise $x^4 + 1$ as a product of two quadratic polynomials with real coefficients. △

4.3.1 Standard factorisations

The challenge to factorise unfamiliar expressions, may at first leave us floundering. But if we assume that each such problem is solvable with the tools at our disposal, we then have no choice but to fall back on the standard tools we have available (in particular, the standard factorisation of a difference of two squares, in which "cross terms" cancel out). The next problem extends this basic repertoire of standard factorisations.

Problem 115

(a)(i) Factorise $a^3 - b^3$.
(ii) Factorise $a^4 - b^4$ as a product of one linear factor and one factor of degree 3, and as a product of two linear factors and one quadratic factor.

(iii) Factorise $a^n - b^n$ as a product of one linear factor and one factor of degree $n-1$.

(b)(i) Factorise $a^3 + b^3$.

(ii) Factorise $a^5 + b^5$ as a product of one linear factor and one factor of degree 4.

(iii) Factorise $a^{2n+1} + b^{2n+1}$ as a product of one linear factor and one factor of degree $2n$. △

Problem **115** develops the ideas that were implicit in Problem **113**. The clue lies in Problem **113**(a), and in the comment made in the main text in Chapter 1 (after Problem **4** in Chapter 1), which we repeat here:

> "The last part [of Problem **113**(a)] is included to emphasise a frequently neglected message:
>
> Words and images are part of the way we communicate.
> But most of us cannot *calculate* with words and images.
>
> To make use of mathematics, we must routinely translate *words* into *symbols*. So "numbers" need to be represented by symbols, and points in a geometric diagram need to be properly labelled before we can begin to calculate, and to reason, effectively."

As soon as one reads the words "one less than a square", one should instinctively translate this into the form "$x^2 - 1$". Bells will then begin to ring; for it is impossible to forget the factorisation

$$x^2 - 1 = (x-1)(x+1).$$

And it follows that:

> for a number that factorises in this way to be prime, the smaller factor $x - 1$ must be equal to 1;
> $\therefore x = 2$, so there is only one such prime.

The integer factorisations in Problem **113**(c) – namely

$$3^3 - 1 = 2 \times 13,\ 4^3 - 1 = 3 \times 21,\ 5^3 - 1 = 4 \times 31,\ 6^3 - 1 = 5 \times 43,\ \ldots$$

may help one to remember (or to discover) the related factorisation

> $x^3 - 1 = (x-1)(x^2 + x + 1)$.
> $\therefore$ For a number that factorises in this way to be prime, the smaller factor "$x - 1$" must be equal to 1;
> $\therefore x = 2$, so there is only one such prime.

Problem **113** parts (a) and (c) highlight the completely general factorisation (Problem **115**(a)(iii)):

$$x^n - 1 = (x-1)(x^{n-1} + x^{n-2} + \cdots + x^2 + x + 1).$$

This family of factorisations also shows that we should think about the factorisation of $x^2 - 1$ as $(\boldsymbol{x} - \mathbf{1})(x+1)$, with the uniform factor $(\boldsymbol{x} - \mathbf{1})$ **first** (rather than as $(x+1)(x-1)$). Similarly, the results of Problem **115** show that we should think of the familiar factorisation of $a^2 - b^2$ as $(\boldsymbol{a} - \boldsymbol{b})(a+b)$, (not as $(a+b)(a-b)$, but always with the factor $(\boldsymbol{a} - \boldsymbol{b})$ **first**).

The integer factorisations in Problem **113**(d) – namely

$$3^3 + 1 = 4 \times 7, 4^3 + 1 = 5 \times 13, 5^3 + 1 = 6 \times 21, 6^3 + 1 = 7 \times 31, 7^3 + 1 = 8 \times 43, \ldots$$

may help one to remember (or to discover) the related factorisation

$$x^3 + 1 = (x+1)(x^2 - x + 1).$$

> $\therefore$ For such a number to be prime, one of the factors must be equal to 1.
> This time one has to be more careful, because the first bracket may not be the "smaller factor" – so there are two cases to consider:
>
> (i) if $x + 1 = 1$, then $x = 0$, and $x^3 + 1 = 1$ is not prime;
>
> (ii) if $x^2 - x + 1 = 1$, then $x = 0$ or $x = 1$, so $x = 1$ and we obtain the prime 2 as the only solution.

The factorisation for $x^3 + 1$ works because "3 is **odd**", which allows the alternating $+/-$ signs to end in a "+" as required. Hence Problem **113**(d)(iii) highlights the completely general factorisation **for odd powers**:

$$x^{2n+1} + 1 = (x+1)(x^{2n} - x^{2n-1} + x^{2n-2} - \cdots + x^2 - x + 1).$$

You probably know that there is no standard factorisation of $x^2 + 1$, or of $x^4 + 1$ (but see Problem **114** above).

Problem 116

(a) Derive a *closed formula* for the sum of the geometric series

$$1 + r + r^2 + r^3 + \cdots + r^n.$$

(The meaning of *closed formula* was discussed in the **Note** to the solution to Problem **54**(b) in Chapter 2.)

(b) Derive a closed formula for the sum of the geometric series

$$a + ar + ar^2 + ar^3 + \cdots + ar^n.$$ △

We started this subsection by looking for prime numbers of the form $x^2 - 1$. A simple-minded approach to the distribution of prime numbers might look for formulae that generate primes – all the time, or infinitely often, or at least much of the time. In Chapter 1 (Problem **25**) you showed that no prime of the form $4k + 3$ can be "represented" as a sum of two squares (i.e. in the form "$x^2 + y^2$"), and we remarked that every other prime can be so represented in exactly one way. It is true (but not obvious) that roughly half the primes fall into the second category; so it follows that substituting integers for the two variables in the polynomial $x^2 + y^2$ produces a prime number infinitely often.

Problem 117 Experiment suggests, and Goldbach (1690–1764) showed in 1752 that no polynomial in one variable, and with integer coefficients, can give prime values for all integer values of the variable. But Euler (1707–1783) was delighted when he discovered the quadratic

$$f(x) = x^2 + x + 41.$$

Clearly $f(0) = 41$ is prime. And $f(1) = 43$ is also prime. What is the first positive integer n for which $f(n)$ is **not** prime? △

Problem **117** should be seen as a particular instance of the question as to whether prime numbers can be captured by a *polynomial* with integer coefficients, and in particular by a *quadratic*. The next two problems consider the simplest instances of representing prime numbers by expressions involving *exponentials* (that is, where the variable is in the *exponent*).

Problem 118

(a)(i) Suppose $a^n - 1 = p$ is a prime. Prove that $a = 2$ and that n must itself be prime.

(ii) How many primes are there among the first five such numbers

$$2^2 - 1,\ 2^3 - 1,\ 2^5 - 1,\ 2^7 - 1,\ 2^{11} - 1?$$

(b)(i) Suppose $a^n + 1 = p$ is a prime. Prove that either $a = 1$, or a must be even and that n must then be a power of 2.

(ii) In the simplest case, where $a = 2$, how many primes are there among the first five such numbers

$$2^1 + 1,\ 2^2 + 1,\ 2^4 + 1,\ 2^8 + 1,\ 2^{16} + 1?$$ △

Primes of the form $2^p - 1$ are called *Mersenne primes* (after Marin Mersenne (1588–1648)). We now know at least fifty such primes (with the exponent p ranging up to around 80 million). Finding new primes is not in itself important, but the search for Mersenne primes has been used as a focus for many new developments in programming, and in number theory.

Primes of the form $2^n + 1$ are called *Fermat primes* (after Pierre de Fermat (1601–1665)). The story here is very different. We now refer to the number $2^n + 1$ with $n = 2^k$ as the k^{th} *Fermat number* f_k. You showed in Problem **118** (as Fermat did himself) that f_0, f_1, f_2, f_3, f_4 are all prime. Fermat then rather rashly claimed that f_n is always prime. However, Euler showed (100 years later) that the very next Fermat number f_5 fails to be prime. And despite all the power of modern computers, we have still not found another Fermat number that is prime!

4.3.2 Quadratic equations

The general solution of quadratic equations dates back to the ancient Babylonians ($\approx$ 1700 BC). Our modern understanding depends on two facts:

- an equation of the form $x^2 = a$ where $a > 0$, has exactly two solutions: $x = \pm\sqrt{a}$;
- any product $X \cdot Y$ is equal to 0 precisely when one of the two factors X, Y is equal to 0.

Problem 119 Solve the following quadratic equations:

(a) $x^2 - 3x + 2 = 0$

(b) $x^2 - 1 = 0$

(c) $x^2 - 2x + 1 = 0$

(d) $x^2 + \sqrt{2}x - 1 = 0$

(e) $x^2 + x - \sqrt{2} = 0$

(f) $x^2 + 1 = 0$

(g) $x^2 + \sqrt{2}x + 1 = 0$ △

Problem 120 Let

$$p(x) = x^2 + \sqrt{2}x + 1.$$

Find a polynomial $q(x)$ such that the product $p(x)q(x)$ has integer coefficients. △

Problem 121

(a) I am thinking of two numbers, and am willing to tell you their sum s and their product p. Express the following procedure algebraically and explain why it will always determine my two unknown numbers.

> Halve the sum s, and square the answer.
> Then subtract the product p and take the square root of the result, to get the answer.
> Add "the answer" to half the sum and you have one unknown number; subtract "the answer" from half the sum and you have the other unknown number.

(b) I am thinking of the length of one side of a square. All I am willing to tell you are two numbers b and c, where when I add b times the side length to the area I get the answer c. Express the following procedure algebraically and explain why it will always determine the side length of my square.

> Take one half of b, square it and add the result to c.
> Then take the square root.
> Finally subtract half of b from the result.

(c) A regular pentagon $ABCDE$ has sides of length 1.

(i) Prove that the diagonal AC is parallel to the side ED.

(ii) If AC and BD meet at X, explain why $AXDE$ is a rhombus.

(iii) Prove that triangles ADX and CBX are similar.

(iv) If AC has length x, set up an equation and find the exact value of x.

△

Problem **121**(a), (b) link to Problem **111**(a) (and to Problem **129** below), in relating the roots and the coefficients of a quadratic. If we forget for the moment that the coefficients are usually known, while the roots are unknown, then we see that if α and β are the roots of the quadratic

$$x^2 + bx + c,$$

then

$$(x - \alpha)(x - \beta) = x^2 + bx + c,$$

so

$$\alpha + \beta = -b \text{ and } \alpha\beta = c.$$

In other words, the two coefficients b, c are equal to the two simplest *symmetric* expressions in the two roots α and β. Part (a) of the next problem is meant to suggest that all other symmetric expressions in α and

β (that is, any expression that is unchanged if we swap α and β) can then be written in terms of b and c. The full result proving this fact is generally attributed to Isaac Newton (1642–1727). Part (b) suggests that, provided one is willing to allow case distinctions, something similar may be true of *anti-symmetric* expressions (where the effect of swapping α and β is to multiply the expression by "-1").

Problem 122 Let α and β be the roots of the quadratic equation

$$x^2 + bx + c = 0.$$

(a)(i) Write $\alpha^2 + \beta^2$ in terms of b and c only.

(ii) Write $\alpha^2\beta + \beta^2\alpha$ in terms of b and c only.

(iii) Write $\alpha^3 + \beta^3 - 3\alpha\beta$ in terms of b and c only.

(b)(i) Write $\alpha - \beta$ in terms of b and c only.

(ii) Write $\alpha^2\beta - \beta^2\alpha$ in terms of b and c only.

(iii) Write $\alpha^3 - \beta^3$ in terms of b and c only. △

Problem 123 (**Nested surds, simplification of surds**)

(a)(i) For any positive real numbers a, b, prove that

$$\sqrt{a} + \sqrt{b} = \sqrt{a + b + \sqrt{4ab}}$$

(ii) Simplify $\sqrt{5 + \sqrt{24}}$.

(b)(i) Find a similar formula for $\sqrt{a} - \sqrt{b}$.

(ii) Simplify $\sqrt{5 - \sqrt{16}}$ and $\sqrt{6 - \sqrt{20}}$. △

Problem 124 (**Integer polynomials with a given root**) We know that $\alpha = 1$ is a root of the polynomial equation $x^2 - 1 = 0$; that $\alpha = \sqrt{2}$ is a root of $x^2 - 2 = 0$; and that $\alpha = \sqrt{3}$ is a root of $x^2 - 3 = 0$.

(a) Find a quadratic polynomial with integer coefficients which has

$$\alpha = 1 + \sqrt{2}$$

as a root.

(b) Find a quadratic polynomial with integer coefficients which has

$$\alpha = 1 + \sqrt{3}$$

as a root.

(c) Find a polynomial with integer coefficients which has

$$\alpha = \sqrt{2} + \sqrt{3}$$

as a root. What are the other roots of this polynomial?

(d) Find a polynomial with integer coefficients which has

$$\alpha = \sqrt{2} + \frac{1}{\sqrt{3}}$$

as a root. What are the other roots of this polynomial? △

Problem 125

(a) Prove that the number $\sqrt{2} + \sqrt{3}$ is irrational.

(b) Prove that the number $\sqrt{2} + \sqrt{3} + \sqrt{5}$ is irrational. △

Problem 126 (**Polynomial long division**) Find

(i) the quotient and the remainder when we divide $x^{10} + 1$ by $x^3 - 1$

(ii) the remainder when we divide $x^{2013} + 1$ by $x^2 - 1$

(iii) the quotient and the remainder when we divide $x^m + 1$ by $x^n - 1$, for $m > n \geqslant 1$. △

Problem 127 Find the remainder when we divide $x^{2013} + 1$ by $x^2 + x + 1$. △

4.4. Complex numbers

Up to this point, the chapter and solutions have largely avoided mentioning *complex numbers*. However, the present chapter would be incomplete were we not to interpret some of the earlier material in terms of complex numbers. Readers who have already met complex numbers will probably still find much in the next two sections that is new. Those for whom complex numbers are as yet unfamiliar should muddle through as best they can, and may then be motivated to learn more in due course.

We already know that the square x^2 of any real number x is $\geqslant 0$.

- If $a = 0$, then the equation $x^2 = a$ has exactly one root, namely $x = 0$;
- if $a > 0$, then the equation $x^2 = a$ has exactly two roots – namely $\pm\sqrt{a}$, where $\sqrt{a}$ denotes the root that is *positive*;
- if $a < 0$, then the equation $x^2 = a$ has no real roots.

And that is where the matter would have rested.

From a modern perspective, we can see that *complex numbers* are implicit in the formula for the roots of a quadratic equation: complex numbers become explicit as soon as the coefficients of a quadratic $ax^2 + bx + c$ give rise to a negative *discriminant* $b^2 - 4ac < 0$.

But this may not have been quite how complex numbers were discovered. Contrary to oft-repeated myths, complex numbers may not have forced themselves on our attention by someone asking about "solutions to the *quadratic* equation $x^2 = -1$". As long as we inhabit the domain of *real* numbers, we can be sure that no known number x could possibly have such a square, so we are unlikely to go in search of it.

New ideas in the history of mathematics tend to emerge when a fresh analysis of *familiar* entities forces us to consider the possible existence of some previously unsuspected universe. In the time from the ancient world up to the fifteenth century, the idea of "number", and of calculation, was restricted to the world of *real* (usually positive) numbers. In such a world, quadratic equations with non-real solutions simply could not arise.

However, in the Brave New World of the Renaissance, where novelty, exploration, and discovery were part of the *Zeitgeist*, a general method for solving *cubic* equations was part of the as-yet-undiscovered "wild west" of mathematics, part of the mathematical New World which invited exploration. Notice that this was not a wildly speculative venture (like trying to solve the meaningless equation "$x^2 = -1$"), since a cubic polynomial **always** has at least one real root. After three thousand years in which little progress had been made, the first half of the sixteenth century witnessed an astonishing burst of progress, resulting in the solution not only of cubic equations, but also of quartic equations. We postpone the details until Section 4.5. All we note here is that,

> the general method for solving cubic equations published in 1545, was given as a procedure, illustrated by examples, that showed how to find *genuinely* **real** *solutions to equations of the third degree having genuinely real* (*and positive*) *coefficients.*

The procedure clearly worked. And it proceeded as follows:

> Construct the real solution x as the sum $x = u + v$ of two intermediate answers u and v – where the two summands u and v sometimes turned out to be what we would call "conjugate complex numbers", *whose imaginary parts cancelled out*, leaving a real result for the required root x.

Those who devised the procedure had no desire to leave the *real* domain: they were focused on a problem in the *real* domain (a cubic equation with *real* coefficients, having a *real* root), and devised a general procedure to find that genuinely *real* root. But the procedure they discovered led the solver on a journey that sometimes "passed into the complex domain", before returning to the real domain! (See Problem **129**.)

Working with complex numbers depends on two skills – one very familiar, and one less so.

- The familiar skill is a willingness to work with a "number" *in terms of its properties only*, without wishing to evaluate it.

We are thoroughly familiar with this when we work with $\frac{2}{3}$ and other fractions: we know that $\frac{2}{3} = 2 \times \frac{1}{3}$; and all we know about $\frac{1}{3}$ is that "whenever we have 3 copies of $\frac{1}{3}$, we can simplify this to 1". Much the same happens when we first learn to work with $\sqrt{2}$, where we carry out such calculations as $(1+\sqrt{2})^2 = 3 + 2\sqrt{2}$, based only on collecting up like terms and the fact that $\sqrt{2} \times \sqrt{2}$ can always be replaced by 2.

- The less familiar skill is easily overlooked. When, for whatever reason, we decide to allow solutions to the equation $x^2 = -1$, three things need to be understood.
  - First, these new solutions come **in pairs**: if i *is one solution of* $x^2 = -1$ then $-i$ is another (because $(-1) \times (-1) = 1$ means that $(-x)^2 = x^2$ for all "numbers" x).
  - Second, the equation $x^2 = -1$ has **exactly two solutions** – one the negative of the other (if x and y are both solutions, then $x^2 = y^2$, so $x^2 - y^2 = (x-y)(x+y) = 0$, so either $x = y$, or $x = -y$).
  - Third, *we have no way of telling these two solutions apart*: we know that each is the negative of the other, but there is no way of singling out one of them as "the main one" (as we could when defining the square root of a positive real such as 2). We can call them $\pm i$, but they are *each as good as the other*. This important fact is often undermined by referring to one of these roots as $\sqrt{-1}$ (as if it were the dominant partner), and to the other as $-\sqrt{-1}$ (as if it were somehow just the "negative" of the main root).

The truth is that "$\sqrt{-1}$" is a serious abuse of notation, because there is no way to extend the definition of the *function* "$\sqrt{}$" in the way that this implies: when we try to "take square roots" of negative (or complex) numbers, the output is inescapably "two-valued", so "$\sqrt{}$" is no longer a function. The two roots of $x^2 = -1$ are like Tweedledum and Tweedledee: we know there are two of them, and we know how they are related; but we have no way of distinguishing them, or of singling one of them out.

Once we accept this, we can write complex numbers in the form $a + bi$, where a and b are real numbers (just as we used to write numbers in the form $a + b\sqrt{2}$, where a and b are rational numbers). And we can proceed to add, subtract, multiply, and divide such expressions, and then collect up the "real" and "imaginary" parts to tidy up the answer.

Problem 128

(a) Write the inverse $(a + bi)^{-1}$ in the form $c + di$.

(b) Write down a quadratic equation with real coefficients, which has $a + bi$ as one root (where a and b are real numbers). △

Problem 129 Divide 10 into two parts, whose product is 40. △

Problem **129** appears in Chapter XXXVII of Girolamo Cardano's (1501–1576) book *Ars Magna* (1545). Having previously presented the general methods for solving quadratic, cubic, and quartic equations, he honestly confronts the phenomenon that his method for solving cubic equations (see Problem **135**) produces the required real (and positive) solution x as a sum of *complex* conjugates u and v – involving not only *negative* numbers, but ***square roots*** *of negative numbers.* After presenting the formal solution of Problem **129**, and having shown that the calculation works exactly as it should, he adds the bemused remark:

> "*So progresses arithmetic subtlety,*
> *the end of which ... is as refined as it is useless.*"

Arithmetic with complex numbers in the form $a + bi$ is done by carrying out the required operations, and then collecting up the "real" and "imaginary" parts as separate components – just as with adding vectors (a, b). We treat the two parts as Cartesian coordinates, and so identify the complex number $a + bi$ with the point (a, b) in the complex plane.

The "Cartesian" representation $a + bi$ is very convenient for *addition.* But the essential definition (and significance) of complex numbers is rooted in *multiplication.* And for multiplication it is often much better to work with complex numbers written in *polar form.* Suppose we mark the complex number $w = a + bi$ in the complex plane.

> The *modulus* $|w|$ of w (often denoted by r) is the distance $r = \sqrt{a^2 + b^2}$ of the complex number $a + bi$ from the origin in the complex plane.
> The angle θ, measured anticlockwise from the positive real axis to the line joining the complex number w to the origin, is called the *argument*, $\mathrm{Arg}(w) = \theta$, of w.

It is then easy to check that $a = r\cos\theta$, $b = r\sin\theta$, and that

$$w = r(\cos\theta + i\sin\theta).$$

This is the *polar form* for w. Instead of focusing on the Cartesian coordinates a, b, the polar form pinpoints w in terms of

- its *length*, or *modulus*, r (which specifies the circle, with centre at the origin, on which the complex number w lies), and
- the *argument* θ (which tells us where on this circle w is to be found).

Problem 130

(a) Given two complex numbers in polar form:

$$w = r(\cos\theta + i\sin\theta),\ z = s(\cos\phi + i\sin\phi),$$

show that their product is precisely

$$wz = rs(cos(\theta + \phi) + i\sin(\theta + \phi)).$$

(b) (**de Moivre's Theorem**: Abraham de Moivre (1667–1754)) Prove that

$$(\cos\theta + i\sin\theta)^n = \cos(n\theta) + i\sin(n\theta).$$

(c) Prove that, if

$$z = r(\cos\theta + i\sin\theta)$$

satisfies $z^n = 1$ for some integer n, then $r = 1$. △

The last three problems in this subsection look more closely at "roots of unity" – that is, roots of the polynomial equation $x^n = 1$. In the *real* domain, we know that:

(i) when n is odd, the equation $x^n = 1$ has exactly one root, namely $x = 1$; and

(ii) when n is even, the equation $x^n = 1$ has just two solutions, namely $x = \pm 1$.

In contrast, in the *complex* domain, there are n "n^{th} roots of unity". Problem **130**(c) shows that these "roots of unity" all lie on the unit circle, centered at the origin. And if we put $n\theta = 2k\pi$ in Problem **130**(b) we see that the n n^{th} roots of unity include the point "$1 = \cos 0 + i \sin 0$", and are then equally spaced around that circle with $\theta = \frac{2k\pi}{n}$ $(1 \leqslant k \leqslant n-1)$, and form the vertices of a regular n-gon.

Problem 131

(a) Find all the complex roots of unity of degree 3 (that is, the roots of $x^3 = 1$) in surds form.

(b) Find all the complex roots of unity of degree 4 in surd form.

(c) Find all the complex roots of unity of degree 6 in surd form.

(d) Find all the complex roots of unity of degree 8 in surd form. △

Problem 132 Use Problem **131**(d) to factorise $x^4 + 1$ as a product of four linear factors, and hence as a product of two quadratic polynomials with real coefficients. △

Problem 133

(a) Find all the complex roots of unity of degree 5 in surd form.

(b) Factorise $x^5 - 1$ as a product of one linear and two quadratic polynomials with real coefficients. △

4.5. Cubic equations

The first recorded procedure for finding the positive roots of any given quadratic equation dates from around 1700 BC (ancient Babylonian). A corresponding procedure for cubic equations had to wait until the early sixteenth century AD. The story is a slightly complicated one – involving public contests, betrayal, and much else besides.

In Section 4.4 we saw that the cubic equation $x^3 = 1$ has three solutions – two of which are complex numbers. But in the sixteenth century, even negative numbers were viewed with suspicion, and complex numbers were still unknown. Moreover, symbolical algebra had not yet been invented, so everything was carried out in words: constants were "numbers"; a given multiple of the unknown was referred to as so many "things"; a given multiple of the square of the unknown was simply referred to as "squares"; and so on.

In short, we know that an improved method for sometimes finding a (positive) unknown which satisfied a cubic equation was devised by Scipione del Ferro (1465–1526) around 1515. He kept his method secret until just before his death, when he told his student Antonio del Fiore (1506–??). Niccolò Tartaglia (1499–1557) then made some independent progress in solving cubic equations. At some stage (around 1535) Fiore challenged Tartaglia to a public "cubic solving contest". In preparing for this event, Tartaglia managed to improve on his method, and he seems to have triumphed in the contest. Tartaglia naturally hesitated to divulge his method in order to preserve his superiority, but was later persuaded to communicate what he knew to Girolamo Cardano (1501–1576) after Cardano promised not to publish it (either never, or not before Tartaglia himself had done so). Cardano improved the method, and his student Ferrari (1522–1565) extended the idea to give a method for solving quartic equations – all of which Cardano then published, contrary to his promise, but with full attribution to the rightful discoverers, in his groundbreaking book *Ars Magna* (1545 – just two years after Copernicus (1473–1543) published his *De revolutionibus* ...). Problem **134** illustrates the necessary first move in solving any cubic equation. Problem **135** then illustrates the general method in a relatively simple case.

Problem 134

(a) Given the equation $x^3 + 3x^2 - 4 = 0$, choose a constant a, and then change variable by substituting $y = x + a$ to produce an equation of the form $y^3 + ky = \text{constant}$.

(b) In general, given any cubic equation $ax^3 + bx^2 + c^x + d = 0$ with $a \neq 0$, show how to change variable so as to reduce this to a cubic equation with no quadratic term. △

Problem 135 The equation $x^3 + 3x^2 - 4 = 0$ clearly has "$x = 1$" as a positive solution. (The other two solutions are $x = -2$, and $x = -2$ – a repeated root; however negatives were viewed with suspicion in the sixteenth century, so this root might well have been ignored.) Try to understand how the following sequence of moves "finds the root $x = 1$":

(i) substitute $y = x + 1$ to get a cubic equation in y with no term in y^2;

(ii) imagine $y = u + v$ and interpret the identity for

$$(u + v)^3 = u^3 + 3uv(u + v) + v^3$$

as your cubic equation in y;

(iii) solve the simultaneous equations "$3uv = 3$", "$u^3 + v^3 = 2$" (not by guessing, but by substituting $v = \frac{1}{u}$ from the first equation into the second to get a quadratic equation in "u^3", which you can then solve for u^3 before taking cube roots);

(iv) then find the corresponding value of v, hence the value of $y = u + v$, and hence the value of x. △

The simple method underlying Problem **135** is in fact completely general. Given any cubic equation

$$ax^3 + bx^2 + cx + d = 0 \quad (\text{with } a \neq 0)$$

we can divide through by a to reduce this to

$$x^3 + px^2 + qx + r = 0$$

with leading coefficient $= 1$. Then we can substitute $y = x + \frac{p}{3}$ and reduce this to a cubic equation in y

$$y^3 - 3\left(\frac{p}{3}\right)^2 y + qy + \left[r + 2\left(\frac{p}{3}\right)^3 - q\left(\frac{p}{3}\right)\right] = 0$$

which we can treat as having the form

$$y^3 - my - n = 0.$$

So we can set $y = u + v$ (for some unknown u and v yet to be chosen), and treat the last equation as an instance of the identity

$$(u + v)^3 - 3uv(u + v) - (u^3 + v^3) = 0$$

which it will become if we simply choose u and v to solve the simultaneous equations

$$3uv = m, \quad u^3 + v^3 = n.$$

We can then solve these equations to find u, then v – and hence find $y = u+v$ and $x = y - \frac{p}{3}$.

4.6. An extra

Back in Chapter 1, Problem **6** we introduced the Euclidean algorithm for integers. The same idea was extended to polynomials with integer coefficients in Problem **126**. In both these settings one starts with a domain (whether the set of integers, or the set of all polynomials with integer coefficients) where there is a notion of divisibility: given two elements m, n in the relevant domain, we say

> "n **divides** m" if there exists an element q in the domain such that $m = qn$.

The next problem invites you to think how one might extend the Euclidean algorithm to a new domain, namely the *Gaussian integers* $\mathbb{Z}[i]$ – the set of all complex numbers $a + bi$ in which the real and imaginary "coordinates" a and b are integers.

Problem 136 Complex numbers $a + bi$, where both a and b are integers, are called Gaussian integers. Try to formulate a version of the "division algorithm" for "division with remainder" (where the remainder is always "less than" the divisor in some sense) for pairs of Gaussian integers. Extend this to construct a version of the Euclidean algorithm to find the HCF of two given Gaussian integers. △

It is a profoundly erroneous truism ...
that we should cultivate the habit of thinking what we are doing.
The precise opposite is the case.
Civilisation advances by extending the number of important
operations which we can perform without thinking about them.

Alfred North Whitehead (1861–1947)

4.7. Chapter 4: Comments and solutions

92. Answer: Humour aside, this is a common situation.

We know $d + b$, $n + b$, $d + n$ rather than the values of d, b, n.

The key is to exploit the symmetry in the given data, rather than solving blindly. Adding all three two-way totals gives $2(d + b + n) = 284$, whence $d + b + n = 142$. We can then subtract the given value of $d + n = 137$ to get the value of $b = 5$.

93.

(a) Let the numbers at the three vertices be A, B, C. Adding shows that

$$a + b + c = 2(A + B + C)$$

so

$$A = \frac{a + b + c}{2} - (B + C) = \frac{b + c - a}{2}$$

etc.

(c) **Note:** We postpone the "solution" of part (b), and address part (c) first. Let the numbers at the five vertices be A, B, C, D, E. Adding shows that

$$d+e+a+b+c=2(A+B+C+D+E)$$

so

$$\begin{aligned} A &= \frac{d+e+a+b+c}{2}-(B+C)-(D+E) \\ &= \frac{d-e+a-b+c}{2} \end{aligned}$$

etc.

(b) The second part is different. The four given edge-values do not determine the four unknown vertex-values. It may look as though four pieces of information should suffice to find four unknowns; but there is a catch: the sum of the numbers on the two opposite edges AB and CD is just the sum of the numbers at the four vertices, and so is equal to the sum of the numbers on the edges BC and DA. Hence one of the given edge-values is determined by the other three.

Note: This distinction between polygons with an odd and an even number of vertices would arise in exactly the same way if each edge was labelled with the average ("half the sum") of the numbers at its two end vertices.

94. $a=BC=BP+PC=y+z$; $b=x+z$; $c=x+y$. Hence

$$a+b+c=2(x+y+z)$$

so

$$x+y+z=\frac{a+b+c}{2}.$$

So

$$\begin{aligned} x &= \frac{a+b+c}{2}-(y+z) \\ &= \frac{b+c-a}{2} \end{aligned}$$

etc.

95.

(a) Let

$$M=\left(\frac{a+c}{2},\frac{b+d}{2}\right).$$

The shift, or vector, from (a,b) to (c,d) goes

"along $c-a$ in the x-direction" and "up $d-b$ in the y-direction".

Draw the ordinate through Y and the abscissa through Z, to meet at P, so creating a right angled triangle with legs YP of length $|c-a|$ and PZ of length

$|d - b|$. The midpoint of YP clearly lies halfway along YP at

$$S = \left(a + \frac{c-a}{2}, b\right)$$

and the midpoint of PZ clearly lies halfway up PZ at

$$T = \left(c, d - \frac{d-b}{2}\right).$$

Then $\triangle YSM$ and $\triangle MTZ$ are both right-angled triangles and are congruent (by RHS congruence). Hence $YM = MZ$, so M is the midpoint of YZ.

(b)

$$M = \left(\frac{a}{2}, \frac{b}{2}\right), \quad N = \left(\frac{c}{2}, \frac{d}{2}\right)$$

so vector

$$\mathbf{MN} = \left(\frac{c-a}{2}, \frac{d-b}{2}\right) = \frac{1}{2}\mathbf{BC}.$$

(c) **Note:** We use the result from part (b), but not the method from part (b).
By part (b) applied to $\triangle BAC$, PQ is half the length of AC and parallel to AC.
By part (b) applied to $\triangle DAC$, SR is half the length of AC and parallel to AC.
Hence PQ is parallel to SR.
Similarly one can prove (applying part (b) twice – first to $\triangle ABD$, and then to $\triangle CBD$) that PS is parallel to QR.
Hence $PQRS$ is a parallelogram.

96.

(a) $\mathbf{p} = \frac{1}{2}(\mathbf{x} + \mathbf{y})$, $\mathbf{q} = \frac{1}{2}(\mathbf{y} + \mathbf{z})$, $\mathbf{r} = \frac{1}{2}(\mathbf{z} + \mathbf{x})$, so

$$\mathbf{p} + \mathbf{q} + \mathbf{r} = \mathbf{x} + \mathbf{y} + \mathbf{z}.$$

Hence

$$\begin{aligned} \mathbf{x} &= (\mathbf{p} + \mathbf{q} + \mathbf{r}) - (\mathbf{y} + \mathbf{z}) = \mathbf{p} - \mathbf{q} + \mathbf{r} \\ \mathbf{y} &= (\mathbf{p} + \mathbf{q} + \mathbf{r}) - (\mathbf{x} + \mathbf{z}) = \mathbf{p} + \mathbf{q} - \mathbf{r} \\ \mathbf{z} &= (\mathbf{p} + \mathbf{q} + \mathbf{r}) - (\mathbf{x} + \mathbf{y}) = \mathbf{q} + \mathbf{r} - \mathbf{p}. \end{aligned}$$

(b)

$$\mathbf{p} = \frac{1}{2}(\mathbf{w} + \mathbf{x}), \ \mathbf{q} = \frac{1}{2}(\mathbf{x} + \mathbf{y}), \ \mathbf{r} = \frac{1}{2}(\mathbf{y} + \mathbf{z}), \ \mathbf{s} = \frac{1}{2}(\mathbf{z} + \mathbf{w})$$

so

$$\mathbf{p} + \mathbf{q} + \mathbf{r} + \mathbf{s} = \mathbf{w} + \mathbf{x} + \mathbf{y} + \mathbf{z}.$$

Hence

$$\mathbf{w} = (\mathbf{p} + \mathbf{q} + \mathbf{r} + \mathbf{s}) - (\mathbf{x} + \mathbf{y} + \mathbf{z});$$

but there is no obvious way to pin down $(\mathbf{x} + \mathbf{y} + \mathbf{z})$.

In fact different quadrilaterals may give rise to the same four "midpoints". (It is an interesting exercise to identify the family of quadrilaterals corresponding to a given set of four midpoints.)

(c) As in parts (a) and (b),

$$\mathbf{p} = \frac{1}{2}(\mathbf{v}+\mathbf{w}), \mathbf{q} = \frac{1}{2}(\mathbf{w}+\mathbf{x}), \mathbf{r} = \frac{1}{2}(\mathbf{x}+\mathbf{y}), \mathbf{s} = \frac{1}{2}(\mathbf{y}+\mathbf{z}), \mathbf{t} = \frac{1}{2}(\mathbf{z}+\mathbf{v}).$$

Hence

$$\mathbf{p}+\mathbf{q}+\mathbf{r}+\mathbf{s}+\mathbf{t} = \mathbf{v}+\mathbf{w}+\mathbf{x}+\mathbf{y}+\mathbf{z}$$

so

$$\begin{aligned}
\mathbf{v} &= (\mathbf{p}+\mathbf{q}+\mathbf{r}+\mathbf{s}+\mathbf{t}) - (\mathbf{w}+\mathbf{x}) - (\mathbf{y}+\mathbf{z}) = \mathbf{p}-\mathbf{q}+\mathbf{r}-\mathbf{s}+\mathbf{t}\\
\mathbf{w} &= (\mathbf{p}+\mathbf{q}+\mathbf{r}+\mathbf{s}+\mathbf{t}) - (\mathbf{x}+\mathbf{y}) - (\mathbf{z}+\mathbf{v}) = \mathbf{p}+\mathbf{q}-\mathbf{r}+\mathbf{s}-\mathbf{t}\\
\mathbf{x} &= (\mathbf{p}+\mathbf{q}+\mathbf{r}+\mathbf{s}+\mathbf{t}) - (\mathbf{v}+\mathbf{w}) - (\mathbf{y}+\mathbf{z}) = -\mathbf{p}+\mathbf{q}+\mathbf{r}-\mathbf{s}+\mathbf{t}\\
\mathbf{y} &= (\mathbf{p}+\mathbf{q}+\mathbf{r}+\mathbf{s}+\mathbf{t}) - (\mathbf{w}+\mathbf{x}) - (\mathbf{z}+\mathbf{v}) = \mathbf{p}-\mathbf{q}+\mathbf{r}+\mathbf{s}-\mathbf{t}\\
\mathbf{z} &= (\mathbf{p}+\mathbf{q}+\mathbf{r}+\mathbf{s}+\mathbf{t}) - (\mathbf{v}+\mathbf{w}) - (\mathbf{x}+\mathbf{y}) = -\mathbf{p}+\mathbf{q}-\mathbf{r}+\mathbf{s}+\mathbf{t}.
\end{aligned}$$

97.

(a)(i) As in Problems **93-95** we instinctively add to get

$$2(x+y+z) = 6$$

so

$$x+y+z = 3.$$

Hence

$$\begin{aligned}
x &= 3-(y+z) = 1\\
y &= 3-(x+z) = 0\\
z &= 3-(x+y) = 2.
\end{aligned}$$

(ii) The same idea (replacing addition by multiplication) leads to

$$2 \times 4 \times 8 = 64 = uv \cdot vw \cdot wu = (uvw)^2$$

so $uvw = \pm 8$. Hence

$$\begin{aligned}
u &= \frac{uvw}{vw} = \frac{\pm 8}{4} = \pm 2\\
v &= \frac{uvw}{uw} = \frac{\pm 8}{8} = \pm 1\\
w &= \frac{uvw}{uv} = \frac{\pm 8}{2} = \pm 4.
\end{aligned}$$

$\therefore (u, v, w) = (2, 1, 4)$ or $(-2, -1, -4)$.

Note: Alternatively, we may notice that u, v, w are either all positive, or all negative. If we restrict in the first instance to purely positive solutions, then we may set $u = 2^x$, $v = 2^y$, $w = 2^z$, translate (ii) into (i), and conclude that $(x, y, z) = (1, 0, 2)$, so that $(u, v, w) = (2, 1, 4)$. We must then remember the negative solution $(u, v, w) = (-2, -1, -4)$.

(b)(i) As in (a)(i) we add to get $2(x + y + z) = 9$, so $x + y + z = \frac{9}{2}$. Hence

$$\begin{aligned} x &= \frac{9}{2} - (y + z) = \frac{3}{2} \\ y &= \frac{9}{2} - (x + z) = \frac{1}{2} \\ z &= \frac{9}{2} - (x + y) = \frac{5}{2}. \end{aligned}$$

(ii) The same idea leads to

$$6 \times 10 \times 15 = 900 = uv \cdot vw \cdot wu = (uvw)^2,$$

so $uvw = \pm 30$.
Hence

$$\begin{aligned} u &= \frac{uvw}{vw} = \frac{\pm 30}{10} = \pm 3 \\ v &= \frac{uvw}{uw} = \frac{\pm 30}{15} = \pm 2 \\ w &= \frac{uvw}{uv} = \frac{\pm 30}{6} = \pm 5. \end{aligned}$$

Either u, v, w are all positive, or all negative.
$\therefore (u, v, w) = (3, 2, 5)$ or $(-3, -2, -5)$.

(iii) The same idea leads to

$$6 \times 10 \times 30 = 2 \times 900 = vw \cdot wu = (uvw)^2,$$

so $uvw = \pm 30\sqrt{2}$. Hence

$$\begin{aligned} u &= \frac{uvw}{vw} = \frac{\pm 30\sqrt{2}}{10} = \pm 3\sqrt{2} \\ v &= \frac{uvw}{uw} = \frac{\pm 30\sqrt{2}}{15} = \pm 2\sqrt{2} \\ w &= \frac{uvw}{uv} = \frac{\pm 30\sqrt{2}}{6} = \pm 5\sqrt{2}. \end{aligned}$$

Either u, v, w are all positive, or all negative.
$\therefore (u, v, w) = (3\sqrt{2}, 2\sqrt{2}, 5\sqrt{2})$ or $(-3\sqrt{2}, -2\sqrt{2}, -5\sqrt{2})$.

(iv) We could of course repeat the same method.

Or we could again look in the first instance for positive solutions, notice that $4 = 2^2$, $8 = 2^3$, $16 = 2^4$, and take logs (to base 2). Then

$$\begin{aligned} \log_2 u + \log_2 v &= 2 \\ \log_2 v + \log_2 w &= 3 \\ \log_2 u + \log_2 w &= 4, \end{aligned}$$

so (from part (i)) any positive solution satisfies

$$\log_2 u = \frac{3}{2},\ \log_2 v = \frac{1}{2},\ \log_2 w = \frac{5}{2},$$

so

$$(u, v, w) = (2\sqrt{2}, \sqrt{2}, 4\sqrt{2}).$$

We must then remember to include

$$(u, v, w) = (-2\sqrt{2}, -\sqrt{2}, -4\sqrt{2}).$$

98. The simplest idea is to take logs, and reduce the system to a familiar linear system:

$$\begin{aligned} a \cdot \log u + b \cdot \log v &= \log m \\ c \cdot \log u + d \cdot \log v &= \log n. \end{aligned}$$

Multiplying the first equation by c and subtracting it from the second equation multiplied by a gives:

$$\log v = \frac{a \cdot \log n - c \cdot \log m}{ad - bc}.$$

Multiplying the first equation by d and subtracting b times the second equation gives:

$$\log u = \frac{d \cdot \log m - b \cdot \log n}{ad - bc}.$$

If the numerators and denominators are expressed in determinant form, we get the 2×2 version of *Cramer's Rule*. The original unknowns u, v can then be obtained by taking suitable powers.

What emerges looks interesting:

$$\begin{aligned} u &= m^{\frac{d}{ad-bc}} \cdot n^{-\frac{b}{ad-bc}} \\ v &= m^{-\frac{c}{ad-bc}} \cdot n^{\frac{a}{ad-bc}} \end{aligned}$$

but it is not clear how it generalises.

99.

(a) $x+y+z = 3$ is the equation of a plane through the three points $(3, 0, 0)$, $(0, 3, 0)$, $(0, 0, 3)$.

$x^2 + y^2 + z^2 = c$ is the equation of a sphere, centered at the origin, with radius $\sqrt{c}$. The sphere misses the plane completely when $c < 3$, meets the plane in a single point when $c = 3$, and cuts the plane in a circle C when $c > 3$ (the circle lying in the positive octant provided $c < 9$).

If $xy + yz + zx = b$ meets this intersection at all, then any permutation of the three coordinates x, y, z produces another point which also satisfies the other two equations (since they are both symmetrical). Hence for the system to have a unique solution, the circle C must contain a point with $x = y = z$. Hence $c = 3$, and $b = 3$, and the unique solution is

$$(x, y, z) = (1, 1, 1).$$

(b) We must have $c \geqslant 0$ for any solution. If $c = 0$, then for a unique solution, we must have $x = y = z = 0$, so $a = b = 0$. If we exclude this case, then we may assume that $c > 0$.

$$x + y + z = a$$

is the equation of a plane through the three points $(a, 0, 0)$, $(0, a, 0)$, $(0, 0, a)$.

$$x^2 + y^2 + z^2 = c$$

is the equation of a sphere, centre the origin, with radius $\sqrt{c}$, which misses the plane completely when $c < \frac{a^2}{3}$, meets the plane in a single point when $c = \frac{a^2}{3}$, and cuts the plane in a circle C when $c > \frac{a^2}{3}$ (the circle lying in the positive octant provided $c < a^2$).

If

$$xy + yz + zx = b$$

meets this intersection at all, then any permutation of the three coordinates x, y, z produces another point which also satisfies the other two equations (since they are both symmetrical). Hence for the system to have a unique solution, the circle C must contain a point with $x = y = z$. Hence that point is $x = y = z = \frac{a}{3}$, so $c = \frac{a^2}{3} = b$, and the unique solution is

$$(x, y, z) = \left(\frac{a}{3}, \frac{a}{3}, \frac{a}{3}\right).$$

100. $|-x|$ is never negative. If $x \geqslant 0$, then $|-x| = x$; if x is negative, then $-x$ is positive, so $|-x| = -x$.

Note: We need to learn to see both x and $-x$ as *algebraic* entities, with x as a placeholder (which may well be negative, in which case $-x$ would be positive).

101.

(a) The interval $[0.1, 0.2)$. We have marked exactly $\frac{1}{10}$ of the interval $[0, 1)$.

(b) This needs a little thought. First we mark the interval $[0.1, 0.2)$, of length $\frac{1}{10}$. Then we mark 9 smaller intervals

$$[0.01, 0.02),\ [0.21, 0.22),\ \ldots,\ [0.91, 0.92)$$

of total length $9 \cdot \left(\frac{1}{10}\right)^2$. Then 9^2 smaller intervals

$$[0.001, 0.002),\ [0.021, 0.022),\ \ldots,\ [0.991, 0.992)$$

of total length $9^2 \cdot \left(\frac{1}{10}\right)^3$. And so on.

$$\begin{aligned}
&[0.1, 0.2) \\
&\quad\cup\quad [0.01, 0.02) \cup [0.21, 0.22) \cup [0.31, 0.32) \cup [0.41, 0.42) \\
&\qquad\qquad \cup [0.51, 0.52) \cup [0.61, 0.62) \cup [0.71, 0.72) \\
&\qquad\qquad\qquad \cup [0.81, 0.82) \cup [0.91, 0.92) \\
&\quad\cup\quad [0.001, 0.002) \cup [0.021, 0.022) \cup [0.031, 0.032) \cup \cdots \\
&\quad\cup\quad \cdots
\end{aligned}$$

It would seem that a vast number of points are left *unmarked* – namely, every point whose decimal representation uses only 0s, 2s, 3s, 4s, 5s, 6s, 7s, 8s, and 9s. However, **the total length** of the *marked* intervals is given by adding:

$$\frac{1}{10} + 9 \cdot \left(\frac{1}{10}\right)^2 + 9^2 \cdot \left(\frac{1}{10}\right)^3 + 9^3 \cdot \left(\frac{1}{10}\right)^4 + 9^4 \cdot \left(\frac{1}{10}\right)^5 + 9^5 \cdot \left(\frac{1}{10}\right)^6 + \cdots$$

That is an infinite geometric series with first term $a = \frac{1}{10}$ and common ratio $r = \frac{9}{10}$, and hence with sum $= 1$. In other words, the **total length of what remains unmarked is zero**.

(c) $(0, 1)$: every real number except 0 has an expansion in base 2 with a "1" in some position. So this time *nothing is left unmarked* (except 0). Hence the complement of the set of marked points consists simply of one point, namely $\{0\}$. So it is not surprising that the total of all the marked intervals has length 1.

(d) First we mark the interval $[0.1, 0.2)$, of length $\frac{1}{3}$. Then we mark 2 smaller intervals

$$[0.01, 0.02),\ [0.21, 0.22)$$

of total length $2 \cdot \left(\frac{1}{3}\right)^2$. Then 2^2 smaller intervals

$$[0.001, 0.002),\ [0.021, 0.022),\ [0.201, 0.202),\ [0.221, 0.222)$$

of total length $2^2 \cdot \left(\frac{1}{3}\right)^3$. And so on.

$$\begin{aligned}
&[0.1, 0.2) \\
&\quad\cup\quad [0.01, 0.02) \cup [0.21, 0.22) \\
&\quad\cup\quad [0.001, 0.002) \cup [0.021, 0.022)) \cup [0.201, 0.202) \cup [0.2201, 0.02202) \\
&\quad\cup\quad \cdots
\end{aligned}$$

The set of marked points is the complement of the famous *Cantor set* (Georg Cantor (1845–1918)) and has total length

$$\frac{1}{3}+2\cdot\left(\frac{1}{3}\right)^2+2^2\cdot\left(\frac{1}{3}\right)^3+2^3\cdot\left(\frac{1}{3}\right)^4+2^4\cdot\left(\frac{1}{3}\right)^5+2^5\cdot\left(\frac{1}{3}\right)^6+\cdots$$

This is an infinite geometric series with first term $a=\frac{1}{3}$ and common ratio $r=\frac{2}{3}$, and so has sum $=1$.
Hence, the **total length of what remains unmarked is zero**.

Note: The set described in (d) leaves as its complement a collection of points – the *Cantor set* – which consists of the "endpoints" of the intervals that have been removed; these are points whose base 3 expansion involves only 0s and 2s. This complement:

(i) is "the same size" as the whole interval $[0,1]$ (since if we interpret the 2s in the base 3 expansion as 1s, we get a correspondence between the set of "endpoints" and the set of all possible *base* 2 expansions for real numbers in $[0,1]$);

(ii) is "nowhere dense" (since every pair of points in the complement is separated by some interval)

(ii) has total length $=0$.

102. (2, 3]. Each inequality implies all the ones before it. Hence the two which are true must be the first two. Hence $x\leqslant 3$, and $x>2$.

103. If $x-5\geqslant 0$, then we must solve $x-5=3$; so $\boldsymbol{x}=\mathbf{8}$; if $x-5<0$, we must solve $x-5=-3$, so $\boldsymbol{x}=\mathbf{2}$.

Note: The fact that $|x|$ denotes the positive value of the pair $\{x,-x\}$ can be rephrased as: $|x|$ is equal to the **distance** from x to 0.

In the same way, $|x-5|$ denotes the positive member of the pair

$$\{x-5,-(x-5)\}$$

so $|x-5|$ is equal to the distance from $x-5$ to 0 (i.e. the distance from x to 5).

This is a very important way to think about expressions like $|x-5|$.

104.

(a) $[-3,-2)\cup(2,3]$. (Each inequality implies all those that go before it. So we need solutions to $|x|>1$ **and** $|x|>2$, which satisfy $|x|\leqslant 3$.)

(b) $(4,6]$. (Each inequality implies the one before it. To see this, think in terms of distances: we want points x whose distance from 1 is >1, whose distance from 2 is >2, etc.. So we need to find points x which solve the first two inequalities, but not the third. Points in the half line $(-\infty,0)$ satisfy all seven inequalities, so we are left with $(4,6]$.)

105. $\left\{-\frac{5}{2}, -\frac{1}{2}\right\}$. (We need all points x for which

"the distance from x to -1" plus "the distance from x to -2"

equals 2. This excludes all points between -2 and -1, for which the sum is equal to 1; for points between $-\frac{5}{2}$ and $-\frac{1}{2}$ the sum is < 2; for points in $\left(-\infty, -\frac{5}{2}\right)$ or $\left(-\frac{1}{2}, \infty\right)$ the sum is > 2.)

106. $a = \frac{1}{2}$, $b = \frac{3}{2}$. (For solutions to exist, we must have $b > 0$. The solutions of the given inequality then consist of all x such that

"the distance from x to a is less than b"

that is, all x in the interval $(a - b, a + b)$. Hence $a - b = -1$, $a + b = 2$.)

107.

(a) "The difference between the x- and y-coordinates is < 3", means that the point (x, y) lies in the infinite strip between the lines $x - y = -3$ and $x - y = 3$.

(b) Shifting the origin of coordinates to $(-5, 0)$ changes the x-coordinate to "$X = x + 5$" and leaves the y-coordinate unaffected (so $Y = y$). In this new frame we want "$|X - Y| < 3$", so the required points lie in the infinite strip between the lines $X - Y = -3$ and $X - Y = 3$; that is, between the lines $x - y + 5 = -3$ and $x - y + 5 = 3$.

(c) $x > 0$ and $y > 0$, or $x < 0$ and $y < 0$. (For any solution at all, we must have $|x + y| > 0$, which excludes points on the line $x + y = 0$. Divide both sides by $|x + y|$ and simplify to get

$$\left|1 - \frac{2y}{x + y}\right| < 1.$$

In other words:

$$0 < \frac{2y}{x + y} < 2.$$

If $y > 0$, then $x + y > 0$, so $2x + 2y > 2y$, whence $x > 0$ (so "$x > 0$ and $y > 0$"). If $y < 0$, then $x + y < 0$, so $2x + 2y < 2y$, whence $x < 0$ (so "$x < 0$ and $y < 0$"). If $x > 0$ and $y > 0$, or $x < 0$ and $y < 0$, then clearly $|x - y| < |x + y|$.)

108. Let

$$x = \frac{a}{b} < \frac{c}{d} = y.$$

(i) Since $x < y$, it follows that

$$x - \frac{x}{2} = \frac{x}{2} < \frac{y}{2},$$

so $x < \frac{x+y}{2}$; moreover $\frac{x}{2} < y - \frac{y}{2}$, so $\frac{x+y}{2} < y$.

(ii) Since $\frac{a}{b} < \frac{c}{d}$ and $b, d > 0$, we can multiply both sides by bd to get $ad < bc$. Therefore

$$a(b+d) = ab + ad < ba + bc = b(a+c),$$

and

$$(a+c)d = ad + cd < bc + dc = (b+d)c.$$

$\therefore \frac{a}{b} < \frac{a+c}{b+d}$, and $\frac{a+c}{b+d} < \frac{c}{d}$ (since b, d, and $b+d$ are all > 0, so we can divide the first inequality by $b(b+d)$ and the second by $d(b+d)$).

109.

(a)

$$\frac{0}{1} < \frac{1}{7} < \frac{1}{6} < \frac{1}{5} < \frac{1}{4} < \frac{2}{7} < \frac{1}{3} < \frac{2}{5} < \frac{3}{7} < \frac{1}{2} < \frac{4}{7} < \frac{3}{5} < \frac{2}{3} < \frac{5}{7} < \frac{3}{4} < \frac{4}{5} < \frac{5}{6} < \frac{6}{7} < \frac{1}{1}$$

(b) (i) It is tempting simply to consider the decimals

$$\frac{1}{9} = 0.111\ldots, \ \frac{2}{9} = 0.222\ldots, \ \frac{3}{9} = 0.333\ldots, \ \ldots, \ \frac{8}{9} = 0.888\ldots$$

in order to conclude that these fractions miss the first and last subinterval, and then fall one in each of the remaining subintervals. In preparation for part (ii), it is better to observe that

* $\frac{1}{10} < \frac{1}{9}$ and $\frac{8}{9} < \frac{9}{10}$, so none of the 9^{th}s land up in the first or last subintervals;
* then rewrite

$$\frac{1}{9} = \frac{1}{10} + \frac{1}{90}, \ \frac{2}{9} = \frac{2}{10} + \frac{2}{90}, \ \frac{3}{9} = \frac{3}{10} + \frac{3}{90}, \ \ldots, \ \frac{8}{9} = \frac{8}{10} + \frac{8}{90}$$

and notice that

$$\frac{0}{10} < \frac{1}{90} < \cdots < \frac{8}{90} < \frac{1}{10},$$

so that, for $1 \leqslant m \leqslant 9$,

$$\frac{m}{10} < \frac{m}{9} < \frac{m+1}{10};$$

hence **exactly one** 9^{th} goes in each of the other eight subintervals.

(ii) Notice that

* $\frac{1}{n+1} < \frac{1}{n}$ and $\frac{n-1}{n} < \frac{n}{n+1}$, so none of the n^{th}s land up in the first or last subintervals;
* then rewrite

$$\frac{1}{n} = \frac{1}{n+1} + \frac{1}{n(n+1)}, \ \frac{2}{n} = \frac{2}{n+1} + \frac{2}{n(n+1)}, \ldots, \ \frac{n-1}{n} = \frac{n-1}{n+1} + \frac{n-1}{n(n+1)}$$

and notice that

$$\frac{0}{n+1} < \frac{1}{n(n+1)} < \cdots < \frac{n-1}{n(n+1)} < \frac{1}{n+1},$$

so, for $1 \leqslant m \leqslant n$,

$$\frac{m}{n+1} < \frac{m}{n} < \frac{m+1}{n+1};$$

hence **exactly one** n^{th} goes in each of the other $n-1$ subintervals.

(iii) Suppose two (or more) fractions are inserted between $\frac{a}{b}$ and $\frac{c}{d}$. Then these two fractions would have to be successive multiples of $\frac{1}{n+1}$; but then they would have a multiple of $\frac{1}{n}$ between them (by part (ii)), and this would be a term of the Farey series of order n sitting between $\frac{a}{b}$ and $\frac{c}{d}$. Since there is no such term, at most one fraction can be inserted between $\frac{a}{b}$ and $\frac{c}{d}$.

(c) **Note 1:** This problem was included because the idea of Farey series seems so simple, and their curious properties are so intriguing. While this remains true, it turns out that Farey series also have something different, and slightly unexpected, to teach us about "the essence of mathematics". Part of us expects that simple-looking results should have short and accessible proofs – even though we know that *Fermat's Last Theorem* shows otherwise. In the case of Farey series, the relevant properties can be proved in ways that should be accessible (in principle); but the proofs are not easy. So do not be upset if, after all your efforts, you land up trying to absorb the solution given here – and the underlying idea that

> simple objects and "elementary" proofs can sometimes be more intricate than one anticipates.

Note 2: If

$$\frac{a}{b} < \frac{c}{d}$$

are consecutive terms in a Farey series, then "$bc - ad$" must be an integer > 0. The fact that this difference is always equal to 1 is easily checked in any *particular* case, but it is unclear exactly why this is *necessarily* true (rather than an accident) – or even how one would go about proving it. Every treatment of Farey series has to find its own way round this difficulty. We give the simplest proof we can (in the sense that it assumes no more than we have already used: a little about numbers and some algebra). But it is not at all "easy". We indicate a different approach in the "**Notes**" at the end of part (d).

(i) [The fact that, except for the two end intervals, we have $bd > n$ will be needed in the proof of part (c)(ii).]

We proceed by mathematical induction on n (the "order" of the Farey series) – a technique which we have already met in Chapter 2 (Problems **54**–**59**, **76**) and which will be addressed more fully in Chapter 6.

∗ When $n = 1$, the Farey series of order n is just:

$$\frac{0}{1} < \frac{1}{1}$$

and this subinterval is both the first and the last, so the claim is "vacuously true" (because there is "nothing to check"). When $n = 2$, the Farey series of order n is:

$$\frac{0}{1} < \frac{1}{2} < \frac{1}{1},$$

and again the only subintervals are the first and the last, so again there is nothing to check.

* We now *suppose that we know the claim is true* for the Farey series of order k, for some $k > 1$, and show that it must then also be true for the Farey series of order $k+1$. (Since we know it is true for $n = 2$, it will then be true for $n = 3$; and once we know it is true for $n = 3$, it must then be true for $n = 4$; and so on.)

  To show that the claim is true for the Farey series of order $k+1$, we consider any adjacent pair of fractions

$$\frac{a}{b} < \frac{c}{d}$$

  (*other than the first pair and the last pair*) in the Farey series of order $k+1$.

Claim $bd > k + 1$.

Proof Note first that, since we are avoiding the two end subintervals, both b and d are > 1.

Suppose first that the pair $\frac{a}{b} < \frac{c}{d}$ are not adjacent in the previous Farey series of order k. Then at least one of the two fractions has been inserted in creating the Farey series of order $k+1$, and so has denominator $= k+1$. (The fractions inserted are precisely those with denominator "$k+1$" which cannot be reduced by cancelling.) Hence the product

$$bd \geqslant 2(k+1) > k+1.$$

Thus we may assume that the pair $\frac{a}{b} < \frac{c}{d}$ were already adjacent in the Farey series of order k. But then by our "induction hypothesis" (namely that the desired result is already known to be true for the Farey series of order k), we know that $bd > k$. If $bd > k+1$, then the pair $\frac{a}{b} < \frac{c}{d}$ satisfies the required condition. Hence we only have to worry about the possibility that $bd = k+1$.

Suppose that $bd = k+1$. Then the interval $\frac{a}{b} < \frac{c}{d}$ has length

$$\frac{bc - ad}{bd} = \frac{r}{k+1}$$

for some positive integer $r = bc - ad$.

If $r > 1$, then the interval would have length $> \frac{1}{k+1}$, so $\frac{a}{b} < \frac{c}{d}$ would **not** be successive terms in the series (for we would have inserted some additional term when moving from the Farey series of order k to the Farey series of order $k+1$).

Hence we can be sure that $r = 1$, that the subinterval $\frac{a}{b} < \frac{c}{d}$ has length exactly $\frac{1}{k+1}$. Now successive fractions with denominator $k+1$ differ by exactly $\frac{1}{k+1}$, so some fraction with denominator $k+1$ must lie in this subinterval. Since no additional fraction is inserted between them in passing from the series of order k to the series of order $k+1$, $\frac{a}{b}$ and $\frac{c}{d}$ must both be "cancelled versions" of two successive fractions with denominator $k+1$. But then, by part (b)(ii), there would have to be a fraction with denominator k in the interval $\frac{a}{b} < \frac{c}{d}$ – which is not the case.

Therefore the possibility $bd = k + 1$ does not in fact occur.

So we can be sure that in every case, $bd > k + 1$.

Hence whenever the result is true for the Farey series of order k, it must then also be true for the Farey series of order $k + 1$.

It follows that the result is true for the Farey series of order n, for all $n \geqslant 1$.

QED

(ii) We proceed by induction on n.

* If $\frac{a}{b} < \frac{c}{d}$ are adjacent fractions in the Farey series of order 1, then $\frac{a}{b} = \frac{0}{1}$ and $\frac{c}{d} = \frac{1}{1}$, so $bc - ad = 1$.
* Now suppose that, for some $k \geqslant 1$, we already know that the result holds for any adjacent pair in the Farey series of order k.

Let $\frac{a}{b} < \frac{c}{d}$ be adjacent fractions in the Farey series of order $k + 1$.

If $\frac{a}{b} < \frac{c}{d}$ were already adjacent fractions in the Farey series of order k (i.e. if no fraction has been inserted between $\frac{a}{b}$ and $\frac{c}{d}$ in passing from the series of order k to the series of order $k + 1$), then we already know (by the induction hypothesis) that $bc - ad = 1$.

Thus we may concentrate on the case where $\frac{a}{b} < \frac{c}{d}$ are not adjacent fractions in the Farey series of order k. By (b)(iii), at most one fraction with denominator $k + 1$ is inserted between any two adjacent fractions in the Farey series of order k, so we have either

$$\frac{a}{b} < \frac{c}{d} < \frac{e}{f},$$

with $\frac{a}{b} < \frac{e}{f}$ being adjacent fractions in the Farey series of order k (so $be - af = 1$), or

$$\frac{e}{f} < \frac{a}{b} < \frac{c}{d},$$

with $\frac{e}{f} < \frac{c}{d}$ being adjacent fractions in the Farey series of order k (so $fc - ed = 1$). We consider the first of these possibilities (the second is entirely similar).

Suppose

$$\frac{a}{b} < \frac{c}{d} < \frac{e}{f},$$

with $\frac{a}{b} < \frac{e}{f}$ being adjacent fractions in the Farey series of order k. By part (i) we know that $bf \geqslant k + 1$; and by induction we know that $be - af = 1$. Hence the interval $\frac{a}{b} < \frac{e}{f}$ has length at most $\frac{1}{k+1}$. We have to prove that $bc - ad = 1$.

Let $bc - ad = r > 0$, and $ed - fc = s > 0$.

Then $sa + re = c$, and $sb + rf = d$.

In particular, $HCF(r, s) = 1$ (since $HCF(c, d) = 1$).

Hence $\frac{c}{d}$ belongs to the family

$$S = \left\{ \frac{xa + ye}{xb + yf} : \text{ where } x, y \text{ are any positive integers with } HCF(x, y) = 1 \right\}.$$

Since everything is positive, easy algebra shows that

$$\frac{a}{b} < \frac{xa + ye}{xb + yf} < \frac{e}{f},$$

so every element of S lies between $\frac{a}{b}$ and $\frac{e}{f}$.
As long as we choose x, y such that $HCF(x, y) = 1$, any common factor of $xa + ye$ and $xb + yf$ would also divide both

$$b(xa + ye) - a(xb + yf) = (be - af)y = y,$$

and

$$e(xb + yf) - f(xa + ye) = (be - af)x = x.$$

Hence

$$HCF(xa + ye, xb + yf) = 1$$

so each element of S is in lowest terms (i.e. no further cancelling is possible).
We have shown that "$\frac{c}{d}$ belongs to the family S", and that *all* elements of S fit *between* $\frac{a}{b}$ and $\frac{e}{f}$; which are *adjacent* fractions in the Farey series of order k. So none of the elements of S can have arisen before the series of order $k + 1$. But each fraction in S arises at some stage in a Farey series.
And the first to *arise* (because it has the smallest denominator) is "$\frac{a+e}{b+f}$". Hence

$$\frac{c}{d} = \frac{a + e}{b + f},$$

so $r = s = 1$, and $bc - ad = 1$ as required. QED

(d) Let

$$\frac{a}{b} < \frac{c}{d} < \frac{e}{f}$$

be three successive terms in any Farey sequence. By (c) we know that $bc - ad = 1$, and that $de - cf = 1$. In particular, $bc - ad = de - cf$, so

$$\frac{c}{d} = \frac{a + e}{b + f}.$$

Note 1: It may not be clear why we are proving this result "again" – since it appeared in the final line of the solution to part (c). However, in part (c) the statement that

$$\frac{c}{d} = \frac{a + e}{b + f}$$

was arrived at *within the induction step*, and so was *subject to other assumptions*. In contrast, now that the result in part (c) has been clearly established, we can use it to prove part (d) without any hidden assumptions.

Note 2: If we represent each fraction $\frac{a}{b}$ in the Farey series of order n by the point (b, a), then each point lies in the right angled triangle joining $(0, 0)$, $(n, 0)$, and (n, n), and each fraction in the Farey series is equal to the gradient of the line, or vector, joining the origin to the integer lattice point (b, a). The ordering of the fractions in the Farey series corresponds to the sequence of increasing gradients, from $\frac{0}{1}$ up to $\frac{1}{1}$. If $\frac{a}{b}$ and $\frac{e}{f}$ are adjacent fractions in some Farey series, then the result in (d) says that the next fraction to be inserted between them is $\frac{a+e}{b+f}$ corresponding to the vector sum of (b, a) and (f, e). And the result in (c) says that the area of the parallelogram with vertices $(0,0)$, (b,a), (f,e), $(b+f, a+e)$ is equal to 1 (see Problem **57**(b)). Hence the result in (c) reduces to the fact that

> **Theorem** Any parallelogram, whose vertices are integer lattice points (i.e. points (b, a) where both coordinates are integers), and with no additional lattice points inside the parallelogram or on the four sides, has area 1.

110.

(a) Suppose that x satisfies $x + \frac{1}{x} < 2$. Then $x \neq 0$ (or $\frac{1}{x}$ is not defined).
$\therefore \frac{x^2+1}{x} < 2$.
If $x > 0$, then $x^2 - 2x + 1 = (x-1)^2 < 0$, which has no solutions.
$\therefore x < 0$, in which case x satisfies $x + \frac{1}{x} < 0 < 2$, so every $x < 0$ is a solution of the original inequality.

(b) Suppose x satisfies $x \leqslant 1 + \frac{2}{x}$. Again $x \neq 0$ (or $\frac{1}{x}$ is not defined).

(i) If $x > 0$, then $x^2 - x - 2 = (x-2)(x+1) \leqslant 0$.
$\therefore -1 \leqslant x \leqslant 2$ (and $x > 0$); hence $0 < x \leqslant 2$, and all such x satisfy the original inequality.

(ii) If $x < 0$, then $x^2 - x - 2 \geqslant 0$, so $(x-2)(x+1) \geqslant 0$.
$\therefore$ either $x \leqslant -1$, or $x \geqslant 2$ (and $x < 0$); hence $x \leqslant -1$, and all such x satisfy the original inequality.

(c) Suppose x satisfies $\sqrt{x} < x + \frac{1}{4}$.
$\therefore 4\left(\sqrt{x}\right)^2 - 4\sqrt{x} + 1 > 0$
$\therefore (2\sqrt{x} - 1)^2 > 0$, so $x \neq \frac{1}{4}$, and all such x satisfy the original inequality.

111.

(a) If $a + b = 5$ and $ab = 7$, then a, b are solutions of

$$(x-a)(x-b) = x^2 - 5x + 7 = 0.$$

But the roots of this quadratic equation are

$$\frac{5 \pm \sqrt{25-28}}{2} = \frac{5 \pm \sqrt{-3}}{2},$$

so a and b cannot be "positive reals".

(b) We abbreviate the "arithmetic mean" by AM, the "geometric mean" by GM, the "harmonic mean" by HM, and the "quadratic mean" by QM.

$$\left(\sqrt{a} - \sqrt{b}\right)^2 \geqslant 0$$

so

$$a + b \geqslant 2\sqrt{ab}$$

therefore

$$\sqrt{ab} \geqslant \frac{2ab}{a+b} \qquad \text{(GM} \geqslant \text{HM)}$$

and

$$\sqrt{ab} \leqslant \frac{a+b}{2} \qquad \text{(GM} \leqslant \text{AM).}$$

Also

$$\left(\frac{a-b}{2}\right)^2 \geqslant 0,$$

so

$$\frac{a^2+b^2}{4} \geqslant \frac{2ab}{4}$$

whence

$$\frac{a^2+b^2}{2} \geqslant \frac{a^2+b^2+2ab}{4} = \left(\frac{a+b}{2}\right)^2.$$

Therefore

$$\sqrt{\frac{a^2+b^2}{2}} \geqslant \frac{a+b}{2} \qquad \text{(QM} \geqslant \text{AM).}$$

QED

112. [This delightful problem was devised by Oleksiy Yevdokimov.] We need to find something which remains constant, or which does not increase, when we replace two terms a, b by $\frac{a+b}{\sqrt{2}}$.

Idea: If the two terms a, b were replaced each time by their sum $a + b$, then the sum of all the numbers in the list would be unchanged, so we could be sure that the final number after 199 such moves would have to be

$$1 + 2 + 3 + \cdots + 200 = \frac{200 \times 201}{2}.$$

This doesn't work here. However, in the spirit of this section on inequalities, one may ask:

> *What happens to the sum of the squares of the terms in the list after each move?*

When we move from one list to the next, only two terms are affected, and for these two terms, the previous sum of squares is replaced by $\left(\frac{a+b}{\sqrt{2}}\right)^2$. How does this affect the sum of all squares on the list?

We know that $a^2 + b^2 \geqslant 2ab$ for all a, b. And it is easy to see that this is equivalent to:

$$a^2 + b^2 \geqslant \left(\frac{a+b}{\sqrt{2}}\right)^2.$$

So when we replace two terms a, b by $\frac{a+b}{\sqrt{2}}$, the sum of the squares of all the terms in the list *never increases*. Hence the final term is less than or equal to the square root of the **initial** sum of squares

$$\begin{aligned} 1^2 + 2^2 + 3^2 + \cdots + 200^2 &= \frac{200 \times 201 \times 401}{6} \qquad \text{(by Problem } \mathbf{62}\text{)} \\ &< \frac{200 \times 300 \times 400}{6} \\ &= 4 \times 10^6. \end{aligned}$$

$\therefore$ the final term is $< \sqrt{4 \times 10^6} = 2000$.

113.

(a) (i) $3 = 2^2 - 1$.

(ii) It seems hard to find another.

(b) (i) $2 = 1^2 + 1$.

(ii) $5 = 2^2 + 1$ (or $17 = 4^2 + 1$; or $37 = 6^2 + 1$; or $101 = 10^2 + 1$; or ...). In other words, there seem to be lots.

Note: At first sight primes of this form "keep on coming". Given that we now know (see Problem **76**) that the list of all prime numbers "goes on for ever", it is natural to ask: Are there infinitely many prime numbers "one more than a square"? Or does the list run out?

This is one of the simplest questions one can ask to which the answer is **not yet known!**

(c) (i) $7 = 2^3 - 1$.

(ii) It seems hard to find another.

(d) (i) $2 = 1^3 + 1$.

(ii) It seems hard to find another.

Note: Parts (a), (c) and (d) should make one suspicious – provided one notices that:

(a) $63 = 7 \times 9$, $143 = 11 \times 13$, $323 = 17 \times 19$;

(c) $511 = 7 \times 73$, $1727 = 11 \times 157$;

(d) $217 = 7 \times 31$, $513 = 9 \times 57$, $1001 = 7 \times 143$.

This problem is so instructive that its solution is discussed in the main text following Problem **115**.

114.

$$x^4 + 1 = \left(x^2 + \sqrt{2} \cdot x + 1\right)\left(x^2 - \sqrt{2} \cdot x + 1\right).$$

(Suppose

$$x^4 + 1 = \left(x^2 + ax + b\right)\left(x^2 + cx + d\right).$$

It is natural to try $b = d = 1$ in order to make the constant term $bd = 1$, and then to try $c = -a$ (so that the coefficients of x^3 and of x are both 0). It then remains to choose the value of a so that the total coefficient "$2 - a^2$" of all terms in x^2 is equal to 0: that is, $a = \sqrt{2}$.)

115.

(a)(i)

$$a^3 - b^3 = (a-b)(a^2 + ab + b^2).$$

(ii)

$$\begin{aligned} a^4 - b^4 &= (a-b)\left(a^3 + a^2b + ab^2 + b^3\right) \\ &= \left(a^2 - b^2\right)\left(a^2 + b^2\right) \\ &= (a-b)(a+b)\left(a^2 + b^2\right). \end{aligned}$$

(iii)

$$a^n - b^n = (a-b)\left(a^{n-1} + a^{n-2}b + a^{n-3}b^2 + \cdots + ab^{n-2} + b^{n-1}\right).$$

Note: The general factorisation

$$x^n - 1 = (x-1)\left(x^{n-1} + x^{n-2} + \cdots + x^2 + x + 1\right)$$

provides a fresh slant on the test for divisibility by 9 in base 10, or in general for divisibility by $b - 1$ in base b (see Problem **51**):

> "an integer written in base b is divisible by $b-1$ precisely when its digit sum is divisible by $b - 1$".

(b) (i)

$$a^3 + b^3 = (a+b)\left(a^2 - ab + b^2\right).$$

(ii)

$$a^5 + b^5 = (a+b)(a^4 - a^3b + a^2b^2 - ab^3 + b^4).$$

(iii)

$$a^{2n+1} + b^{2n+1} = (a+b)\left(a^{2n} - a^{2n-1}b + a^{2n-2}b^2 - a^{2n-3}b^3 + \cdots - ab^{2n-1} + b^{2n}\right).$$

116.

(a) Replace a by 1, b by r, and n by $n + 1$ in the answer to **115**(a)(iii), to see that:

$$1 + r + r^2 + r^3 + \cdots + r^n = \frac{1 - r^{n+1}}{1 - r}.$$

(b) Multiply the closed formula in (a) by "a" to see that:

$$a + ar + ar^2 + ar^3 + \cdots + ar^n = a \cdot \frac{1 - r^{n+1}}{1 - r}.$$

117. When $x = 40$,

$$f(x) = x^2 + (x + 40) + 1 = 40^2 + 2 \times 40 + 1 = 41^2$$

is not prime. So the sequence of prime outputs must stop some time before $f(40)$. But it in fact keeps going as long as it possibly could, so that

$$f(0),\ f(1),\ f(2),\ \ldots,\ f(39)$$

are all prime. (This may explain Euler's delight.)

Note: The links between polynomials with integer coefficients (even lowly quadratics) and prime numbers are still not fully understood. For example, you might like to look up *Ulam's spiral.* (Ulam (1909–1984) plotted the positive integers in a square spiral and found the primes arranging themselves in curious patterns that we still do not fully understand.)

Interest in the connections between polynomials and primes was revived in the second half of the 20th century. It was eventually proved that there exists a polynomial in 10 variables, with *integer* coefficients, which takes both positive and negative values when the variables run through all possible non-negative *integer* values, but which does so in such a way that it generates **all** the primes as the set of positive outputs.

118.

(a) (i) For

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

to be prime, the smaller factor must be $= 1$, so $a = 2$.

If n is not prime, we can factorise $n = rs$, with $r, s > 1$. Then

$$2^n - 1 = 2^{rs} - 1 = (2^r)^s - 1 = (2^r - 1)\left(2^{r(s1)} + 2^{r(s2)} + \cdots + 2 + 1\right);$$

Hence $2^n 1$ also factorises, so could not be prime. Hence n must be prime.

(ii) $2^2 - 1 =$ **3**, $2^3 - 1 =$ **7**, $2^5 - 1 =$ **31**, $2^7 - 1 =$ **127** are all prime; $2^{11} - 1 = 2047 = 23 \times 89$ is not.

Note: This is a simple example of the need to distinguish carefully between the statement

"if $2^n - 1$ is prime, then n is prime" (which is true),

and its converse

"if n is prime, then $2^n - 1$ is prime" (which is false).

(b) (i) Suppose that $a > 1$. Then $a^n + 1 > 2$; so for $a^n + 1$ to be prime, it must be odd, so a must be even.

If n has an odd factor $m > 1$, we can write $n = km$. Then

$$\begin{aligned} a^n + 1 &= a^{km} + 1 \\ &= (a^k)^m + 1 \\ &= (a^k + 1)\left(a^{k(m-1)} - a^{k(m-2)} + \cdots - a^k + 1\right). \end{aligned}$$

Since m is odd and > 1, we have $m \geqslant 3$. It is then easy to show that

$$a^k + 1 \leqslant a^{k(m-1)} - a^{k(m-2)} + \cdots - a^k + 1.$$

And since $a > 1$, neither factor $= 1$, so $a^n + 1$ can never be prime. Hence n can have no odd factor > 1, which is the same as saying that $n = 2^r$ must be a power of 2.

(ii) $2^1 + 1 = \mathbf{3}$, $2^2 + 1 = \mathbf{5}$, $2^4 + 1 = \mathbf{17}$, $2^8 + 1 = \mathbf{257}$, $2^{16} + 1 = \mathbf{65\,537}$ are all prime. (The very next such expression

$$2^{32} + 1 = 4\,294\,967\,297 = 641 \times 6\,700\,417$$

is not prime.)

Note: The sad tale of Fermat's claim that "all Fermat numbers are prime" shows that mathematicians are not exempt from the obligation to distinguish carefully between a statement and its converse!

119.

(a) $x^2 - 3x + 2 = (x-2)(x-1) = 0$ precisely when one of the brackets $= 0$; that is, $x = 2$, or $x = 1$.

(b) $x^2 - 1 = (x-1)(x+1) = 0$ precisely when $x = 1$ or $x = -1$.

(c) $x^2 - 2x + 1 = (x-1)^2 = 0$ precisely when $x = 1$ (a repeated root).

(d) $x^2 + \sqrt{2}x - 1 = 0$ requires us

– to complete the square

$$x^2 + \sqrt{2}x - 1 = \left(x + \frac{\sqrt{2}}{2}\right)^2 - 1 - \frac{1}{2},$$

so

$$x + \frac{\sqrt{2}}{2} = \pm\sqrt{\frac{3}{2}},$$

– or to use the quadratic formula:

$$x = \frac{-\sqrt{2} \pm \sqrt{2+4}}{2}.$$

(e) $x^2 + x - \sqrt{2} = 0$ requires us

– to complete the square

$$x^2 + x - \sqrt{2} = \left(x + \frac{1}{2}\right)^2 - \sqrt{2} - \frac{1}{4},$$

so

$$x + \frac{1}{2} = \pm\sqrt{\sqrt{2} + \frac{1}{4}}$$

– or to use the quadratic formula:

$$x = \frac{-1 \pm \sqrt{1 + 4\sqrt{2}}}{2}.$$

(f) $x^2 + 1 = 0$ yields $x = \pm\sqrt{-1}$.

(g) $x^2 + \sqrt{2}x + 1 = 0$ yields

$$x = \frac{-\sqrt{2} \pm \sqrt{2 - 4}}{2} = \frac{-\sqrt{2} \pm \sqrt{-2}}{2}.$$

120. $q(x) = x^2 - \sqrt{2} \cdot x + 1$. (There is no obvious magic method here. However, it should be natural to try to insert a term $\sqrt{2} \cdot x$ in $q(x)$ to "resolve" the term $\sqrt{2} \cdot x$ in $p(x)$; and the familiar cancelling of cross terms in $(a+b)(a-b)$ should then suggest the possible benefit of trying $q(x) = x^2 - \sqrt{2} \cdot x + 1$.)

Note: $p(x)q(x) = x^4 + 1$ (see Problem **114**).

121.

(a) Let the two unknown numbers be α and β. Then $s = \alpha + \beta$, and $p = \alpha\beta$. "The square of half the sum" $\left(\frac{s}{2}\right)^2 = \left(\frac{\alpha+\beta}{2}\right)^2$.

Subtracting $p = \alpha\beta$ produces $\left(\frac{\alpha-\beta}{2}\right)^2$ whose "square root" will be either $\frac{\alpha-\beta}{2}$, or $-\left(\frac{\alpha-\beta}{2}\right)$ – whichever is positive.

Adding this to "half the sum" gives one root; subtracting gives the other root.

(b) Let the length of one side be x. We are told that $x^2 + bx = c$.

"Take half of b, square it, and add the result to c"

translates as:

"Rewrite the equation as: $\left(x + \frac{b}{2}\right)^2 = c + \left(\frac{b}{2}\right)^2$."

That is, we have "completed the square" $\left(x + \frac{b}{2}\right)^2$. If we now take the (positive) square root and subtract $\frac{b}{2}$, we get a single value for x, which determines the side length of my square as required.

If the same method is applied to the general quadratic equation

$$ax^2 + bx + c = 0,$$

with the extra initial step of "multiply through by $\frac{1}{a}$", we produce first

$$x^2 + \frac{b}{a}x + \frac{c}{a} = 0,$$

then

$$\left(x + \frac{b}{2a}\right)^2 + \left(\frac{c}{a} - \left(\frac{b}{2a}\right)^2\right) = 0,$$

then

$$x + \frac{b}{2a} = \pm\sqrt{\left(\frac{b}{2a}\right)^2 - \frac{c}{a}} = \frac{\pm\sqrt{b^2 - 4ac}}{2a},$$

which leads to the familiar quadratic formula.

(c) See Problem **3**(c)(iv). $AD : CB = DX : BX$, so $x : 1 = 1 : (x - 1)$. Hence $x^2 - x - 1 = 0$. If we use the quadratic formula derived in the answer to part (b) above, and realise that $x > 1$, then we obtain $x = \frac{1+\sqrt{5}}{2}$.

Note: The procedure given in (a) dates back to the ancient Babylonians ($\sim$ 1700 BC) and later to the ancient Greeks ($\sim$ 300 BC). Both cultures worked *without* algebra. The Babylonians gave their verbal procedures as recipes in words, in the context of particular examples. The Greeks expressed everything geometrically. In modern language, if we denote the unknown numbers by α and β, then

$$(x - \alpha)(x - \beta) = x^2 - (\alpha + \beta)x + \alpha\beta.$$

Being told the sum and product is therefore the same as being given the coefficients of a quadratic equation, and being asked to find the two roots.

Our method for factorizing a quadratic involves a mental process of 'inverse arithmetic', where we juggle possibilities in search of α and β, when all we know are the coefficients (that is, the sum $\alpha + \beta$, and the product $\alpha\beta$).

The procedure in (b) also dates back to the ancient Babylonians, and is essentially our process of completing the square. It was given as a procedure, without our algebraic notation. The Babylonians seem not to have been hampered (as the Greeks were) by the fact that it makes no sense to add a *length* and an *area*! They worded things geometrically, but seem to have understood that they were really playing *numerical* games (an idea which European mathematicians found elusive right up to the time of Descartes (1590–1656)).

Similarly, the modern use of symbols – allowing one to represent either positive or negative quantities – was widely resisted right into the nineteenth century. What we would write as a single family of quadratic equations, $ax^2+bx+c = 0$, had to be split into separate cases where two *positive* quantities were equated. For example, the groundbreaking book *Ars Magna* in which Cardano (1501–1576) explained how to solve cubic and quartic equations begins with quadratics – where his procedure distinguishes four different cases: "squares equal to numbers", "squares equal to things", "squares and things equal to numbers", "squares and numbers equal to things".

122.

(a) (i) $\alpha^2 + \beta^2 = (\alpha + \beta)^2 - 2\alpha\beta = b^2 - 2c$.

(ii) $\alpha^2\beta + \beta^2\alpha = \alpha\beta(\alpha + \beta) = c \cdot (-b) = -bc$.

(iii) We rearrange

$$\begin{aligned} \alpha^3 + \beta^3 - 3\alpha\beta &= (\alpha + \beta)\left(\alpha^2 - \alpha\beta + \beta^2\right) - 3\alpha\beta \\ &= (-b) \cdot \left(b^2 - 3c\right) - 3c \\ &= -b^3 + 3bc - 3c. \end{aligned}$$

[Alternatively: $\alpha^3 + \beta^3 = (\alpha + \beta)^3 - 3\alpha\beta(\alpha + \beta)$, etc.]

(b) (i) [Cf **121**(a).] $(\alpha - \beta)^2 = (\alpha + \beta)^2 - 4\alpha\beta$.
Therefore

$$\alpha - \beta = \sqrt{b^2 - 4c} \text{ if } \alpha \geqslant \beta,$$

and

$$\alpha - \beta = -\sqrt{b^2 - 4c} \text{ if } \alpha < \beta.$$

(ii)

$$\alpha^2\beta - \beta^2\alpha = -\alpha\beta(\alpha - \beta) = -c\sqrt{b^2 - 4c} \text{ if } \alpha \geqslant \beta,$$

and

$$\alpha^2\beta - \beta^2\alpha = -\alpha\beta(\alpha - \beta) = c\sqrt{b^2 - 4c} \text{ if } \alpha < \beta.$$

(iii) $\alpha^3 - \beta^3 = (\alpha - \beta)(\alpha^2 + \alpha\beta + \beta^2)$.
Therefore

$$\alpha^3 - \beta^3 = \left[\sqrt{b^2 - 4c}\right]\left(b^2 - c\right) \text{ if } \alpha \geqslant \beta,$$

and

$$\alpha^3 - \beta^3 = \left[-\sqrt{b^2 - 4c}\right]\left(b^2 - c\right) \text{ if } \alpha < \beta.$$

123.

(a) (i) $\sqrt{a} + \sqrt{b}$ and $\sqrt{a + b + \sqrt{4ab}}$ are both positive. And it is easy to check that they have the same square:

$$\left(\sqrt{a} + \sqrt{b}\right)^2 = a + b + 2\sqrt{ab},$$

and

$$\left(\sqrt{a + b + \sqrt{4ab}}\right)^2 = a + b + \sqrt{4ab}.$$

Hence

$$\sqrt{a} + \sqrt{b} = \sqrt{a + b + \sqrt{4ab}}.$$

(ii) $5 = 2 + 3$, and $24 = 4 \times 2 \times 3$;
Therefore

$$\sqrt{2 + 3 + \sqrt{4 \times 2 \times 3}} = \sqrt{2} + \sqrt{3}$$

(which is easy to check).

(b) (i) **Claim** If $a \geqslant b$ ($\neq 0$), then

$$\sqrt{a} - \sqrt{b} = \sqrt{a + b - \sqrt{4ab}}.$$

Proof $\sqrt{a} - \sqrt{b}$ and $\sqrt{a + b - \sqrt{4ab}}$ are both $\geqslant 0$ (Why?). And it is easy to check that

$$\left(\sqrt{a} - \sqrt{b}\right)^2 = a + b - 2\sqrt{ab},$$

and

$$\left(\sqrt{a + b - \sqrt{4ab}}\right)^2 = a + b - \sqrt{4ab}.$$ QED

(ii) Simplify $\sqrt{5 - \sqrt{16}}$ and $\sqrt{6 - \sqrt{20}}$.
$5 = 4 + 1$ and $16 = 4 \times 4 \times 1$, so $\sqrt{5 - \sqrt{16}} = \sqrt{4} - \sqrt{1} = 1$.
Actually, there is a simpler solution:

$$\sqrt{5 - \sqrt{16}} = \sqrt{5 - 4} = \sqrt{1} = 1.$$

$6 = 5 + 1$ and $20 = 4 \times 5 \times 1$, so $\sqrt{6 - \sqrt{20}} = \sqrt{5} - \sqrt{1} = \sqrt{5} - 1$.

124.

(a) Let $\alpha = 1 + \sqrt{2}$. Then $\alpha^2 = 3 + 2\sqrt{2}$. Hence $\alpha^2 - 2\alpha = 1$, so α satisfies the quadratic polynomial equation $x^2 - 2x - 1 = 0$.

Note: Observe that the resulting polynomial is equal to

$$\left(x - \left(1 + \sqrt{2}\right)\right)\left(x - \left(1 - \sqrt{2}\right)\right).$$

In other words, to rationalize the coefficients, we need a polynomial which has both $\alpha = 1 + \sqrt{2}$ and its "conjugate" $1 - \sqrt{2}$ as roots.

(b) Let $\alpha = 1 + \sqrt{3}$. Then $\alpha^2 = 4 + 2\sqrt{3}$. Hence $\alpha^2 - 2\alpha = 2$, so α satisfies the quadratic polynomial equation $x^2 - 2x - 2 = 0$.

Note: Observe that the resulting polynomial is equal to

$$\left(x - \left(1 + \sqrt{3}\right)\right)\left(x - \left(1 - \sqrt{3}\right)\right).$$

In other words, to rationalize the coefficients, we need a polynomial which has both $\alpha = 1 + \sqrt{3}$ and its "conjugate" $1 - \sqrt{3}$ as roots.

(c) Let $\alpha = \sqrt{2} + \sqrt{3}$. Then $\alpha^2 = 5 + 2\sqrt{6}$, so $\alpha^2 - 5 = 2\sqrt{6}$, and $\left(\alpha^2 - 5\right)^2 = 24$. Hence α satisfies the quartic polynomial equation $x^4 - 10x^2 + 1 = 0$.

Note: Observe that the resulting polynomial is equal to

$$\left(x - \left(\sqrt{2} + \sqrt{3}\right)\right)\left(x - \left(\sqrt{2} - \sqrt{3}\right)\right)\left(x - \left(-\sqrt{2} + \sqrt{3}\right)\right)\left(x - \left(-\sqrt{2} - \sqrt{3}\right)\right).$$

In other words, the roots are: $\sqrt{2}+\sqrt{3}$ (as required), and also $\sqrt{2}-\sqrt{3}$, $-\sqrt{2}-\sqrt{3}$, and $-\sqrt{2} + \sqrt{3}$.

(d) Let $\alpha = \sqrt{2} + \frac{1}{\sqrt{3}}$. Then

$$\alpha^2 = \frac{7}{3} + 2\sqrt{\frac{2}{3}},$$

so

$$\left(\alpha^2 - \frac{7}{3}\right)^2 = \frac{8}{3},$$

and α satisfies the quartic polynomial equation

$$x^4 - \frac{14}{3} \cdot x^2 + \frac{25}{9} = 0.$$

Note:

$$x^4 - \frac{14}{3} \cdot x^2 + \frac{25}{9} = \left(x - \left[\sqrt{2} + \frac{1}{\sqrt{3}}\right]\right)\left(x - \left[\sqrt{2} - \frac{1}{\sqrt{3}}\right]\right) \cdot \left(x + \left[\sqrt{2} + \frac{1}{\sqrt{3}}\right]\right)\left(x + \left[\sqrt{2} - \frac{1}{\sqrt{3}}\right]\right),$$

so the roots are:

$$x = \sqrt{2} + \frac{1}{\sqrt{3}},\ \sqrt{2} - \frac{1}{\sqrt{3}},\ -\sqrt{2} - \frac{1}{\sqrt{3}},\ -\sqrt{2} + \frac{1}{\sqrt{3}}.$$

125. A direct approach can be made to work in both cases (but see the **Notes**).

(a) Suppose to the contrary that $\sqrt{2} + \sqrt{3} = \frac{p}{q}$, for some integers p, q with $HCF(p,q) = 1$. Then $\left(5 + 2\sqrt{6}\right)q^2 = p^2$, so $\sqrt{6}$ is rational, and we can write $\sqrt{6} = \frac{r}{s}$ with $HCF(r,s) = 1$. But then $6s^2 = r^2$; hence $r = 2t$ must be even; so $3s^2 = 2t^2$, but then s must be even – contradicting $HCF(r,s) = 1$. Hence $\sqrt{2} + \sqrt{3}$ cannot be rational.

Note: It is slightly easier to rewrite the initial equation in the form

$$\sqrt{3} = \frac{p}{q} - \sqrt{2}$$

before squaring to get

$$\left(\frac{p}{q}\right)^2 - 1 = \frac{2p}{q}\sqrt{2},$$

which would imply that $\sqrt{2}$ is rational.

(b) Suppose to the contrary that $\sqrt{2} + \sqrt{3} + \sqrt{5} = \frac{p}{q}$, for some integers p, q with $HCF(p,q) = 1$. Then

$$10 + 2\left(\sqrt{6} + \sqrt{10} + \sqrt{15}\right) = \left(\frac{p}{q}\right)^2,$$

so $\sqrt{6} + \sqrt{10} + \sqrt{15}$ is rational. Squaring $\sqrt{6} + \sqrt{10} + \sqrt{15}$ then gives that

$$\sqrt{60} + \sqrt{90} + \sqrt{150} = 5\sqrt{6} + 3\sqrt{10} + 2\sqrt{15}$$

is rational. Subtracting $2(\sqrt{6}+\sqrt{10}+\sqrt{15})$ then shows that $3\sqrt{6}+\sqrt{10}$ is rational, and we can proceed as in part (a) to obtain a contradiction. Hence $\sqrt{2}+\sqrt{3}+\sqrt{5}$ cannot be rational.

Note: It is simpler to rewrite the original equation in the form

$$\sqrt{2}+\sqrt{3}=\frac{p}{q}-\sqrt{5}$$

before squaring to obtain

$$5+2\sqrt{6}=\left(5+\left(\frac{p}{q}\right)^2\right)-\frac{2p}{q}\sqrt{5},$$

whence $2\sqrt{6}+\frac{2p}{q}\sqrt{5}$ is rational, and we may proceed as in part (a).

126.

(i) We just have to fill in the missing bits of the partial factorisation

$$x^{10}+1=\left(x^3-1\right)\left(x^7+x^4+\cdots\right)+\text{ remainder.}$$

To produce the required term x^{10} we first insert x^7. This then creates an unwanted term "$-x^7$", so we add $+x^4$ to cancel this out. This in turn creates an unwanted term "$-x^4$", so we add $+x$ to cancel this out. Hence the quotient is x^7+x^4+x, and the remainder is "$x+1$":

$$x^{10}+1=\left(x^3-1\right)\left(x^7+x^4+x\right)+(x+1).$$

Note: It is worth noting a short cut. The factorised term of the form $\left(x^3-1\right)\left(x^7+\cdots\right)$ is equal to zero when $x^3=1$.
So one way to get the remainder is to "treat x^3 as if it were equal to 1". Then

$$x^{10}=\left(x^3\right)^3\cdot x$$

is just like $1\cdot x$, and $x^{10}+1$ behaves as if it were equal to $x+1$, which is the remainder.

(ii)

$$x^{2013+1}=\left(x^2-1\right)\left(x^{2011}+x^{2009}+x^{2007}+\cdots+x\right)+(x+1),$$

so the remainder $=x+1$.

Note: If we treat x^2 "as if it were equal to 1", then

$$x^{2013}+1=\left(x^2\right)^{1006}\cdot x+1$$

behaves as if it were equal to $1\cdot x+1$.

(iii) Apply the Euclidean algorithm to m and n in order to write $m=qn+r$, where $0\leqslant r<n$:

$$x^m=x^{qn+r}=\left(x^n\right)^q\cdot x^r.$$

Then

$$\begin{aligned} x^m + 1 &= x^{qn+r} + 1 \\ &= (x^n - 1)\left(x^{n(q-1)+r} + x^{n(q-2)+r} + x^{n(q-3)+r} + \cdots + x^r\right) + x^r + 1. \end{aligned}$$

So the remainder is $x^r + 1$.

Note: If we treat $x^n - 1$ as if were 0 – that is, if we treat x^n as if it were equal to 1 – then

$$x^m + 1 = x^{qn+r} + 1 = (x^n)^q \cdot x^r + 1$$

which behaves like $1^q \cdot x^r + 1$.

127. Suppose $x^{2013} + 1 = \left(x^2 + x + 1\right) q(x) + r(x)$, where $\deg(r(x)) < 2$. Then

$$\begin{aligned} \left(x^{2013} + 1\right)(x - 1) &= x^{2014} - x^{2013} + x - 1 \\ &= \left(x^3 - 1\right) q(x) + (x - 1)r(x). \end{aligned}$$

Now

$$\begin{aligned} x^{2014} - x^{2013} + x - 1 &= \left(x^3 - 1\right)\left(x^{2011} - x^{2010} + x^{2008} - x^{2007} + x^{2004} - x^{2003} + \cdots + x - \right. \\ &\quad + 2x - 2 \end{aligned}$$

so the remainder $r(x) = 2$.

Note: If x satisfies $x^2 + x + 1 = 0$, then $x^3 - 1 = 0$ and $x \neq 1$.
$\therefore\ x^{2013} + 1 = \left(x^3\right)^{671} + 1$ behaves just like $1^{671} + 1 = 2$, so $r(x) = 2$.

128.

(a)

$$(a + bi)^{-1} = \frac{a}{a^2 + b^2} - \left[\frac{b}{a^2 + b^2}\right] i.$$

(b)

$$p(x) = x^2 - 2ax + \left(a^2 + b^2\right).$$

(Suppose that the quadratic equation

$$p(x) = x^2 + cx + d = 0,$$

with real coefficients c, d, has $x = a + ib$ as a root. Then take the complex conjugate of the equation $p(x) = 0$ to see that $x = a - ib$ is also a rooti of

$$p(x) = x^2 + cx + d = 0.$$

Therefore

$$\begin{aligned} p(x) &= x^2 + cx + d \\ &= (x - (a + ib))(x - (a - ib)), \end{aligned}$$

so $c = -2a$, and $d = a^2 + b^2$.)

129. Let the two unknown numbers be α and β. Then $10 = \alpha + \beta$, and $40 = \alpha\beta$, so α and β are roots of the quadratic equation $x^2 - 10x + 40 = 0$. Hence

$$\alpha, \beta = \frac{10 \pm \sqrt{100 - 160}}{2} = 5 \pm \sqrt{-15}.$$

130.

(a) Applying a simple rearrangement:

$$\begin{aligned} wz &= r(\cos\theta + i\sin\theta) \cdot s(\cos\phi + i\sin\phi) \\ &= rs[(\cos\theta \cdot \cos\phi - \sin\theta\sin\phi) + i(\cos\theta \cdot \sin\phi + sin\theta \cdot \cos\phi)] \\ &= rs[\cos(\theta + \phi) + i\sin(\theta + \phi)] \end{aligned}$$

(by the usual addition formula: Problem **35**)

(b) By part (a),

$$(\cos\theta + i\sin\theta)^2 = \cos(2\theta) + i\sin(2\theta).$$

Hence

$$\begin{aligned} (\cos\theta + i\sin\theta)^3 &= (\cos\theta + i\sin\theta)^2(\cos\theta + i\sin\theta) \\ &= [\cos(2\theta) + i\sin(2\theta)] \cdot (\cos\theta + i\sin\theta) \\ &= \cos(3\theta) + i\sin(3\theta). \end{aligned}$$

Etc.

Note: This should really be presented as a "proof by mathematical induction", where (having established the initial cases) we "suppose the result holds for powers $n = 1, 2, 3, \ldots, k$", and then conclude that

$$\begin{aligned} (\cos\theta + i\sin\theta)^{k+1} &= (\cos\theta + i\sin\theta)^k(\cos\theta + i\sin\theta) \\ &= [\cos(k\theta) + i\sin(k\theta)](\cos\theta + i\sin\theta) \\ &= \cos((k+1)\theta) + i\sin((k+1)\theta). \end{aligned}$$

(c) $z^n = r^n(\cos(n\theta) + i\sin(n\theta))$. Hence if $z^n = 1$, then $|z^n| = r^n = 1$, so $r = 1$ (since $r \geqslant 0$).

131.

(a) We factorise: $x^3 - 1 = (x - 1)\left(x^2 + x + 1\right)$, so the roots are $x = 1$; and

$$x = \frac{-1 \pm \sqrt{1 - 4}}{2} = \frac{-1 \pm \sqrt{-3}}{2} = -\frac{1}{2} \pm \frac{\sqrt{3}}{2}i$$

that is, the other two roots are

$$x = \cos\left(\frac{2\pi}{3}\right) + i\sin\left(\frac{2\pi}{3}\right)$$

and

$$x = \cos\left(-\frac{2\pi}{3}\right) + i\sin\left(-\frac{2\pi}{3}\right).$$

(b) We factorise:

$$x^4 - 1 = \left(x^2 - 1\right)\left(x^2 + 1\right) = (x-1)(x+1)\left(x^2+1\right),$$

so the roots are $x = 1$, $x = -1$, $x = i$, $x = -i$.

(c) We factorise:

$$\begin{aligned} x^6 - 1 &= \left[\left(x^2\right)^3 - 1\right] \\ &= \left(x^2 - 1\right)\left(x^4 + x^2 + 1\right) \\ &= (x-1)(x+1)\left[\left(x^2\right)^2 + x^2 + 1\right], \end{aligned}$$

so the roots are

- $x = 1$, $x = -1$, and
- four further values of x satisfying $x^2 = -\frac{1}{2} \pm \frac{\sqrt{3}}{2}i$: that is,

$$x = \cos\left(\frac{\pi}{3}\right) + i\sin\left(\frac{\pi}{3}\right) = \frac{1}{2} + \frac{\sqrt{3}}{2}i$$

and

$$x = \cos\left(\frac{2\pi}{3}\right) + i\sin\left(\frac{2\pi}{3}\right) = -\frac{1}{2} + \frac{\sqrt{3}}{2}i$$

and

$$x = \cos\left(\frac{-\pi}{3}\right) + i\sin\left(\frac{-\pi}{3}\right) = \frac{1}{2} - \frac{\sqrt{3}}{2}i$$

and

$$x = \cos\left(\frac{-2\pi}{3}\right) + i\sin\left(\frac{-2\pi}{3}\right) = -\frac{1}{2} - \frac{\sqrt{3}}{2}i.$$

(d) We factorise:

$$\begin{aligned} x^8 - 1 &= \left(x^4 - 1\right)\left(x^4 + 1\right) \\ &= \left(x^2 - 1\right)\left(x^2 + 1\right)\left(x^2 + \sqrt{2}\cdot x + 1\right)\left(x^2 - \sqrt{2}\cdot x + 1\right) \end{aligned}$$

so the roots are

- $x = 1$, $x = -1$;
- $x = i$, $x = -i$, and
- the roots of $x^2 + \sqrt{2}\cdot x + 1 = 0$ and $x^2 - \sqrt{2}\cdot x + 1 = 0$, which happen to be

$$x = \cos\left(\frac{\pi}{4}\right) + i\sin\left(\frac{\pi}{4}\right) = \frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2}i$$

and

$$x = \cos\left(-\frac{\pi}{4}\right) + i\sin\left(-\frac{\pi}{4}\right) = \frac{\sqrt{2}}{2} - \frac{\sqrt{2}}{2}i$$

and

$$x = \cos\left(\frac{3\pi}{4}\right) + i\sin\left(\frac{3\pi}{4}\right) = -\frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2}i$$

and

$$x = \cos\left(-\frac{3\pi}{4}\right) + i\sin\left(-\frac{3\pi}{4}\right) = -\frac{\sqrt{2}}{2} - \frac{\sqrt{2}}{2}.$$

132. [In Problem **114** you were left to work out the required factorisation with your bare hands – and a bit of inspired guesswork. The suggested approach here is more systematic.]

The roots of $x^4 + 1 = 0$ are complex numbers whose fourth power is equal to -1: that is,

$$x = \cos\left(\frac{\pi}{4}\right) + i\sin\left(\frac{\pi}{4}\right) = \frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2}i$$

and

$$x = \cos\left(-\frac{\pi}{4}\right) + i\sin\left(-\frac{\pi}{4}\right) = \frac{\sqrt{2}}{2} - \frac{\sqrt{2}}{2}i$$

and

$$x = \cos\left(\frac{3\pi}{4}\right) + i\sin\left(\frac{3\pi}{4}\right) = -\frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2}i$$

and

$$x = \cos\left(-\frac{3\pi}{4}\right)) + i\sin\left(-\frac{3\pi}{4}\right) = -\frac{\sqrt{2}}{2} - \frac{\sqrt{2}}{2}i.$$

The first two are complex conjugates and give rise to two linear factors whose product is $x^2 + \sqrt{2} \cdot x + 1$; the other two are complex conjugates and give rise to two linear factors whose product is $x^2 - \sqrt{2} \cdot x + 1$. Hence

$$x^4 + 1 = \left(x^2 + \sqrt{2} \cdot x + 1\right)\left(x^2 - \sqrt{2} \cdot x + 1\right).$$

133.

(a) The roots of $x^5 - 1 = 0$ are precisely the five complex numbers of the form

$$\cos\left(\frac{2k\pi}{5}\right) + i\sin\left(\frac{2k\pi}{5}\right), \text{ for } k = 0, 1, 2, 3, 4:$$

that is,

$$\begin{aligned} x &= 1 \\ x &= \cos\left(\frac{2\pi}{5}\right) + i\sin\left(\frac{2\pi}{5}\right) \\ x &= \cos\left(\frac{4\pi}{5}\right) + i\sin\left(\frac{4\pi}{5}\right) \\ x &= \cos\left(\frac{6\pi}{5}\right) + i\sin\left(\frac{6\pi}{5}\right) \\ x &= \cos\left(\frac{8\pi}{5}\right) + i\sin\left(\frac{8\pi}{5}\right). \end{aligned}$$

From Problem **3**(c) we know that

$$\begin{aligned} \cos\left(\frac{2\pi}{5}\right) &= \frac{\sqrt{5}-1}{4} = \cos\left(\frac{8\pi}{5}\right) \\ \sin\left(\frac{2\pi}{5}\right) &= \frac{\sqrt{10+2\sqrt{5}}}{4} = -\sin\left(\frac{8\pi}{5}\right) \\ \cos\left(\frac{4\pi}{5}\right) &= -\cos\left(\frac{\pi}{5}\right) = -\frac{\sqrt{5}+1}{4} = \cos\left(\frac{6\pi}{5}\right) \\ \sin\left(\frac{4\pi}{5}\right) &= \frac{\sqrt{10-2\sqrt{5}}}{4} = -\sin\left(\frac{6\pi}{5}\right). \end{aligned}$$

(b) The linear factor is clearly $(x-1)$. Each quadratic factor arises as the product of two conjugate linear factors. We saw in Problem **128**(b) that two linear factors corresponding to roots $a+bi$ and $a-bi$ produce the quadratic factor $x^2 - 2ax + \left(a^2+b^2\right)$. Hence the two quadratic factors are:

$$x^2 - \frac{\sqrt{5}-1}{2}\cdot x + 1, \quad \text{and} \quad x^2 + \frac{\sqrt{5}+1}{2}\cdot x + 1$$

(whose product is equal to $x^4 + x^3 + x^2 + x + 1$).

134.

(a) Put $a = 1$, $y = x+1$: then $x^3 + 3x^2 - 4 = 0$ becomes $y^3 - 3y = 2$.

(b) Divide through by a (which we may assume is non zero, since otherwise it would not be a *cubic* equation), to obtain a cubic equation

$$x^3 + px^2 + qx + r = 0.$$

If we now put $y = x + \frac{p}{3}$, then y^3 incorporates both the x^3 and the x^2 terms, and the equation reduces to:

$$y^3 + \left[q - 3\left(\frac{p}{3}\right)^2\right] y + \left[r + 2\left(\frac{p}{3}\right)^3 - q\left(\frac{p}{3}\right)\right] = 0.$$

135. Given the equation $x^3 + 3x^2 - 4 = 0$. Let $y = x + 1$.

(i) Then $y^3 = x^3 + 3x^2 + 3x + 1$, so $0 = x^3 + 3x^2 - 4 = y^3 - 3y - 2$.

(ii) Set $y = u + v$ and use the fact that

$$(u + v)^3 = u^3 + 3uv(u + v) + v^3$$

is an identity, and so holds for all u and v.

(iii) Solve "$3uv = 3$", "$u^3 + v^3 = 2$". Substitute $v = \frac{1}{u}$ from the first equation into the second to get the quadratic equation in $\left(u^3\right)^2 - 2u^3 + 1 = 0$: that is, $\left(u^3 - 1\right)^2 = 0$, so $u^3 = 1$.

(iv) **Hence $u = 1$ is certainly a solution**. (We know there are also complex cube roots of 1; these lead to the other two solutions of the original cubic, but to "solve the equation" it is enough to find one solution.) **Hence $v = 1$, so $y = u + v = 2$, and $x = 1$.**

136. The Euclidean algorithm for ordinary integers arises by repeating the division algorithm:

> given integers m, n ($\neq 0$), there exists unique integers q, r such that $m = qn + r$ where $0 \leqslant r < n$.

Here q is the *quotient* (the integer part of the division $m \div n$), and r is the *remainder*. If we then replace the initial pair (m, n) by the new pair (n, r) and repeat until we obtain the remainder 0, then the last non-zero remainder is equal to $HCF(m, n)$ (see Problem **6**). The same idea also works for polynomials with integer coefficients (see Problem **126**).

We start by clarifying what we mean by *divisibility* for Gaussian integers. Given two Gaussian integers, $m = a + bi$ and $n = c + di$, we say that $n = c + di$ **divides** $m = a + bi$ (exactly) precisely

> when there exists some other Gaussian integer $q = e + fi$ such that $m = qn$: that is, $a + bi = (e + fi)(c + di)$.

For example, $2 + 3i$ divides $-4 + 7i$ because $(1 + 2i)(2 + 3i) = -4 + 7i$.

If $m = a + bi$ and $n = c + di$ are any old Gaussian integers, then it will not in general be true that "n divides m", but we can imitate the division algorithm. The important idea here when carrying out particular calculations is to realize that "divide by $c + di$" is the same as "multiply by $\frac{c-di}{c^2+d^2}$"

- first carry out the division

$$m \div n = \frac{(a + bi)(c - di)}{c^2 + d^2};$$

- then take the "nearest" Gaussian integer $q = e + fi$, and let the difference $m - qn = r$ be the *remainder*.

As for ordinary integers, any Gaussian integer that is a "common factor of m and n" is then automatically a common factor of n and of $r = m - qn$, and conversely. That is, the common factors of m and n are precisely the same as the common factors of n and r. So we can repeat the process replacing m, n by n, r. Provided the "remainder" r is in some sense "smaller" than n, we can continue until we reach a stage where the remainder $r = 0$ – at which point, the last non-zero remainder is equal to the $HCF(m, n)$ (that is, the Gaussian integer which is the HCF of the two initial Gaussian integers m, n).

The feature of the remainders that gets progressively smaller is their *norm* (see Problem **25**, and Problem **54**). As so often, this becomes clearer when we look at an example.

Let us try to find the HCF of the two Gaussian integers $m = 14-42i$ and $n = 4-7i$.

- First do the division

$$m \div n = \frac{(14 - 42i)(4 + 7i)}{4^2 + 7^2} = \frac{350}{65} - \frac{70}{65}i.$$

- What is meant by the *nearest* Gaussian integer may require an element of judgment; but it is clear that the answer is fairly close to $5 - i = q$, where $qn = 13 - 39i$, with remainder $r = m - qn = 1 - 3i$.
- Now repeat the process with n, r:

$$n \div r = \frac{(4 - 7i)(1 + 3i)}{1^2 + 3^2} = \frac{5}{2} + \frac{1}{2}i.$$

- The *nearest* Gaussian integer is not well-defined, but the answer is fairly close to $3 = q'$. So $q'r = 3 - 9i$, with remainder $r' = n - q'r = 1 + 2i$.
- Now repeat the step with the pair $r = 1 - 3i$ and $r' = 1 + 2i$, to discover that

$$1 - 3i = -(1 + i)(1 + 2i)$$

  with remainder 0. Hence

$$1 + 2i = HCF(14 - 42i, 4 - 7i).$$

Note: One way to picture the process is to learn to "see" the Gaussian integers *geometrically*. Every Gaussian integer (such as $a + bi$) can be written as an integer combination of the two basic Gaussian integers "1" and "i" – namely

$$a + bi = a \times 1 + b \times i.$$

Since 1 and i are both of length 1 and perpendicular to each other, this represents the set of all Gaussian integers as the dots in a "square dot lattice" generated by translations in the x- and y- directions of the basic unit square spanned by 0, 1, i, and $1 + i$.

Any other given Gaussian integer, such as $n = c + di$, then generates a "stretched and rotated" square lattice, which consists of all "Gaussian multiples" of $c + di$ – generated by the basic square which is spanned by

$$0,\ \ (c + di) \times 1,\ \ (c + di) \times i,\ \ \text{and}\ \ (c + di) \times (1 + i).$$

Every Gaussian integer (or rather the point, or dot, which corresponds to it) lies either on the boundary, or inside, one of these larger "stretched and rotated" squares: if the diagonal of one of these larger squares has length $2k$, then any other Gaussian integer $m = a + bi$ lies inside one of these larger squares, and so lies *within distance* k (that is, half a diagonal) of some (Gaussian) multiple qn of $n = c + di$. And the difference $m - qn$ is precisely the required *remainder* r.

Extra: We interpret $\sqrt[3]{8} = 8^{\frac{1}{3}} = 2$. Prove that

$$\sqrt[\sqrt{-1}]{-1} \approx 23\frac{1}{7}$$

(where $\approx$ denotes "approximately equal to").

V. Geometry

Those who fear to experiment with their hands
will never know anything.
George Sarton (1884–1956)

Mathematical truth is not determined arbitrarily
by the rules of some 'man-made' formal system,
but has an absolute nature and lies
beyond any such system of specifiable rules.
Roger Penrose (1930–)

Geometry is in many ways the most natural branch of elementary mathematics through which to convey "the essence" of the discipline.

- The underlying subject matter is rooted in seeing, moving, doing, drawing, making, etc., and so is accessible to everyone.
- At secondary level this practical experience leads fairly naturally to a semi-formal treatment of "geometry as a mental universe"
  - a universe that is bursting with surprising facts, whose statements can be easily understood; and
  - which has a clear logical structure, in terms of which the proofs of these facts are accessible, if sometimes tantalisingly elusive.

This combination of elusive problems to be solved and the steady accumulation of proven results has provided generations of students with their first glimpse of serious mathematics. All readers can imagine the kind of experiences which lie behind the first bullet point above: many of the problems we have already met (such as Problems **4**, **19**, **20**, **26**, **27**, **28**, **29**, **30**, **31**, **37**, **38**, **39**) do not depend on the "semi-formal treatment" referred to in the second bullet point, so can be tackled by anyone who is interested – *provided they accept the importance of learning to construct their own diagrams* (in the spirit of the George Sarton quotation).

The hand is the cutting edge of the mind.
Jacob Bronowski (1908–1974)

But there is a catch – which explains why the present chapter appears so late in the collection. For many problems to successfully convey "the essence of mathematics" there has to be some shared understanding of what constitutes a solution. And in geometry, many solutions require the construction of a **proof**. Yet many readers will never have experienced a coherent "semi-formal treatment" of elementary geometry in the spirit of the second bullet point. Hence in Problems **3**(c), **18**, **21**, **32**, **34**, **36** we committed the cardinal sin of leading the reader by the nose – breaking each problem into steps in order to impose a logical structure. This may have been excusable in Chapter 1; but in a chapter explicitly devoted to geometry, the underlying challenge has to be faced head on: that is, the raw experience of the *hand* has to be refined to provide a deductive structure for the *mind.*

As in Chapter 1, some of the problems listed from Section 5.3 onwards can be tackled without worrying too much about the logical structure of elementary geometry. But in many instances, the "essence" that is captured by a problem requires that the problem be seen within an agreed logical hierarchy – a sequencing of properties, results, and methods, which establishes *what* is a consequence of *what* – and hence, what can be used as part of a solution. In particular, we need to construct proofs that avoid circular reasoning.

> If B is a consequence of A, or if B is equivalent to A, then a 'proof' of A which makes use of B is at best dubious, and may well be a delusion.

The need to avoid such circular reasoning arose already in Problem **21** (the converse of Pythagoras' Theorem), where we felt the need to state explicitly that it would be inappropriate to use the Cosine Rule: (see Problem **192** below).

Such concerns may explain why this chapter on geometry is the last of the chapters relating to elementary 'school mathematics', and why we begin the chapter with

- an apparent digression (Section 5.1), and
- an outline of elementary Euclidean geometry (Section 5.2).

Those with a strong background in geometry may choose to skip these sections on a first reading, and move straight on to the problems which start in Section 5.3. But they may then fail to see how the cumulative architecture of Section 5.2 conveys a rather different aspect of the "essence of mathematics", deriving not just from the individual problems, but from the way a carefully crafted, systematic arrangement of simple "bricks" can create a much more significant mathematical structure.

5.1. Comparing geometry and arithmetic

The opening quotations remind us that the mental universe of formal mathematics draws much of its initial inspiration from human perception and activity – activity which starts with infants observing, moving around, and operating with objects in time and in space. Many of our earliest pre-mathematical experiences are quintessentially proto-geometrical. We make sense of visual inputs; we learn to recognise faces and objects; we crawl around; we learn to look 'behind' and 'underneath' obstructions in search of hidden toys; we sort and we build; we draw and we make; etc.. However, for this experience to develop into *mathematics*, we then need to

- identify certain semi-formal "objects" (points, lines, angles, triangles),
- pinpoint the key relations between them (bisectors, congruence, parallels, similarity), and then
- develop the associated language that allows us to encapsulate insights from prior experience into a coherent framework for calculation and deduction.

Too little attention has been given to achieving a consensus as to how this transition (from *informal experience*, to *formal reasoning*) can best be established for beginners in elementary geometry. In contrast, number and arithmetic move much more naturally

- from our early experience of time and quantity
- to the notation, the operations, the calculational procedures, and the rules of formal arithmetic and algebra.

Counting is rooted in the idea of a *repeated unit* – a notion that may stem from the ever-present, regular heartbeat that envelops every embryo (where the beat is presumably *felt* long before it is *heard*). Later we encounter repeated units with longer time scales (such as the cycles of day and night, and the routines of feeding and sleeping). The first months and years of life are peppered with instances of numerosity, of continuous quantity, of systematic ordering, of sequences, of combinations and partitions, of grouping and replicating, and of relations between quantities and operations – experiences which provide the raw material for the mathematics of number, of place value, of arithmetic, and later of 'internal structure' (or algebra).

The need for political communities to construct a formal school curriculum linking early infant experience and elementary formal mathematics is a recent development. Nevertheless, in the domain of number, quantity, and arithmetic (and later algebra), there is a surprising level of agreement about the steps that need to be incorporated – even though the details may differ in different educational systems and in different classrooms. For example:

- One must somehow establish the idea of a *unit*, which can be replicated to produce larger numbers, or *multiples*.
- One must then group units relative to a chosen base (e.g. 10), iterate this grouping procedure (by taking "ten tens", and then "ten hundreds"), and use *position* to create *place value* notation.
- One must introduce "0" – both as a number in its own right, and as a placeholder for expressing numbers using place value.
- One can then use combinations and differences, multiples and sharing (and partitions), to develop *arithmetic*.
- At some stage one introduces subunits (i.e. *unit fractions*) and submultiples (i.e. multiples of these subunits) to produce *general fractions*; one can then use *equivalence* and common submultiples to extend arithmetic to fractions.
- If we restrict to *decimal fractions*, then our ideas of place value for integers can be extended to the right of the decimal point to produce *decimals*.
- At every stage we need to
  - relate these ideas to *quantities*,
  - require pupils to interpret and solve *word problems*, and
  - cultivate both mental arithmetic and standard written algorithms.
- Towards the end of primary school, attention begins to move beyond bare hands computation, to consciously exploit internal *structure* in preparation for algebra.

Our early *geometrical* experience is just as natural as that relating to number; but it is more subtle. And there is as yet no comparable consensus about the path that needs to be followed if our primitive geometrical experience is to be formalised in a useable way.

The 1960s saw a drive to modernise school mathematics, and at the same time to make it accessible to all. Elementary geometry certainly needed a re-think. But the reformers in most countries simply dismissed the traditional mix (e.g. in England, where one found a blend of technical drawing, Euclidean, and coordinate geometry in different proportions for different groups of students) in favour of more modern-sounding alternatives. Some countries favoured a more abstract, deductive framework; some tried to exploit motion and transformations; some used matrices and groups; some used vectors and linear algebra; some even toyed with topology. More recently we have heard similarly ambitious claims on behalf of dynamic geometry software. And although each approach has its attractions,

> **none** *of the alternatives has succeeded in helping* **more** *students to visualise, to reason, and to calculate effectively in geometrical settings.*

At a much more advanced level, geometry combines

- with abstract algebra (where the approach proposed by Felix Klein (1849–1925) shows how to identify each geometry with a group of transformations), and
- with analysis and linear algebra (where, following Gauss (1777–1855), Riemann (1826–1866) and Grassmann (1809–1877), calculus, vector spaces, and later topology can be used to analyse the geometry of surfaces and other spaces).

However, these subtle formalisms are totally irrelevant for beginners, who need an approach

- based on concepts which are relatively familiar (*points, lines, triangles* etc.), and
- whose basic properties can be formulated relatively simply.

The subtlety and flexibility of dynamic geometry software may be hugely impressive; but if students are to harness this power, they need *prior* mastery of some simple, semi-formal framework, together with the associated language and modes of reasoning. Despite the lack of an accepted consensus, the experience of the last 50 years would seem to suggest that the most relevant framework for beginners at secondary level involves some combination of:

- *static*, relatively traditional Euclidean geometry, and
- Cartesian, or coordinate (analytic) geometry.

5.2. Euclidean geometry: a brief summary

Philosophy is written in this grand book – I mean the
universe – which stands continually open to our gaze,
but it cannot be understood
unless one first learns to comprehend the language
and to interpret the characters in which it is written.
It is written in the language of mathematics,
and its characters are triangles, circles,
and other geometrical figures, without which
it is humanly impossible to understand a single word of it;

without these, one is wandering about in a dark labyrinth.
Galileo Galilei (1564–1642)

This section provides a detailed, but compressed, outline of an initial formalisation of school geometry – of a kind that one would like good students and all teachers to appreciate. It is unashamedly a *semi*-formal approach for beginners, **not** a strictly formal treatment (such as that provided by David Hilbert (1862–1943) in his 1899 book *Foundations of Geometry*, or in the more detailed exposition by Edwin Moise (1918–1998) *Elementary Geometry from an Advanced Standpoint*, published in 1963). In particular:

- we work with relatively informal notions of *points*, *lines*, and *angles* in the plane;
- we focus attention on certain simple issues which really matter at school level (such as how points, lines, line segments, and angles are referred to; the notion of a triangle as an *ordered* triple of vertices; the fact that the vertices of a quadrilateral must be labelled cyclically; etc.);
- we limit the formal deductive structure to just three central criteria, namely the criteria for *congruence*, for *parallels*, and for *similarity*, and show how they allow one to develop results and methods in a logical sequence.

We begin with the intuitive idea of *points* and *lines* in the plane. Two points A, B determine

- the *line segment* $\underline{AB}$ (with endpoints A and B), and
- the *line* AB (which extends the line segment $\underline{AB}$ in both directions – beyond A, and beyond B).

Three points A, B, C determine an angle $\angle ABC$ (between the two line segments $\underline{BA}$ and $\underline{BC}$).

We can then begin to build more complicated figures, such as

- a triangle ABC (with three vertices A, B, C; three sides $\underline{AB}$, $\underline{BC}$, $\underline{CA}$; and three angles $\angle ABC$ at the vertex B, $\angle BCA$ at C, and $\angle CAB$ at A),
- a quadrilateral $ABCD$ (with four vertices A, B, C, D; and four sides $\underline{AB}$, $\underline{BC}$, $\underline{CD}$, $\underline{DA}$ which meet only at their endpoints).

And so on. Two given points A, B also allow us to construct the *circle* with centre A, and passing through B (that is, with *radius* $\underline{AB}$).

This very limited beginning already opens up the world of *ruler and compasses constructions*. In particular, given a line segment $\underline{AB}$, one can draw:

- the circle with centre A, and passing through B, and
- the circle with centre B, and passing through A.

If the two circles meet at C,

- then $\underline{AB} = \underline{AC}$ (radii of the first circle), and $\underline{BA} = \underline{BC}$ (radii of the second circle).

Hence we have constructed the *equilateral triangle* $\triangle ABC$ on the given segment $\underline{AB}$. This construction is the very first proposition in Book 1 of the *Elements* of Euclid (flourished c. 300 BC). Euclid's second proposition is presented next as a problem.

Problem 137 Given three points A, B, C, show how to construct – without measuring – a point D such that the segments $\underline{AB}$ and $\underline{CD}$ are equal (in length). △

Problem **137** looks like a simple starter (where the only available construction is to produce the third vertex of an equilateral triangle on a given line segment). However, to produce a valid solution requires a clear head and a degree of ingenuity.

Given two points A, B, the process of constructing an equilateral triangle $\triangle ABC$ illustrates how we are allowed to construct new points from old.

- Whenever we construct two lines or circles that cross, the points where they cross (such as the point C in the above construction of the equilateral triangle $\triangle ABC$) become available for further constructions. So, if points A and B are given, then once C has been constructed, we may proceed to draw the lines AC and BC.

However, the fact that we can construct a line segment $\underline{AB}$ does not allow us to 'measure' the segment with a ruler, and then to use the resulting measurement to 'copy' the segment $\underline{AB}$ to the point C in order to construct the required point D such that $\underline{AB} = \underline{CD}$. The "ruler" in *ruler and compasses constructions* is used only to draw the line through two known points – not to measure. (Measuring is an *approximate* physical action, rather than an *exact* "mental construction", and so is not really part of mathematics.) Hence in Problem **137** we have to find another way to produce a copy $\underline{CD}$ of the segment $\underline{AB}$ starting at the point C. Similarly, we can construct the circle with centre A and passing through B, but this does not allow us to use the pair of compasses to transfer distances physically (e.g. by picking up the compasses from $\underline{AB}$ and placing the compass point at C, like using the old geometrical drawing instrument that was called *a pair of dividers*). In seeking the construction required in Problem **137**, we are restricted to "exact mental constructions" which may be described in terms of:

- drawing (or constructing) the line joining any two known points,
- constructing the circle with centre at a known point and passing through a known point, and
- obtaining a new point D as the intersection of two constructed lines or circles (or of a line and a circle).

If on the line AB, the point X lies between A and B, then we obtain a *straight angle* $\angle AXB$ at X (or rather *two* straight angles at X – one on each side of the line AB). If we assume that all straight angles are equal, then it follows easily that "vertically opposite angles are always equal".

Problem 138 Two lines AB and CD cross at X, where X lies between A and B and between C and D. Prove that $\angle AXC = \angle BXD$. $\triangle$

Define a *right angle* to be 'half a straight angle'. Then we say that two lines which cross at a point X are *perpendicular* if an angle at X is a right angle (or equivalently, if all four angles at X are equal). The next step requires us to notice two things – partly motivated by experience when coordinating hand, eye and brain to construct, and to think about, physical structures.

- First we need to recognise that *triangles* hold the key to the analysis of more complicated shapes.
- Then we need to realise that triangles in different positions can still be "equal", or *congruent* – which then focuses attention on the **minimal** conditions under which two triangles can be guaranteed to be congruent.

The first of these two bullet points has an important consequence – namely that solving any problem in 2- or in 3-dimensions generally reduces to working with **triangles**. In particular, solving problems in 3-dimensions reduces to working in some 2-dimensional *cross-section* of the given figure (since three points not only determine a triangle, but also determine the plane in which that triangle lies). It follows that 2-dimensional geometry holds the key to solving problems in 3-dimensions, and that **working with triangles is central in all geometry**.

The second bullet point forces us to think carefully about:

- what we mean by a *triangle* (and in particular, to understand why $\triangle ABC$ and $\triangle BCA$ are in some sense *different* triangles, even though they use the same three vertices and sides), and
- what it means for two triangles to be "the same".

A triangle $\triangle ABC$ incorporates six pieces of data, or information: the three sides $\underline{AB}$, $\underline{BC}$, $\underline{CA}$ and the three angles $\angle ABC$, $\angle BCA$, $\angle CAB$. We say that two (ordered) triangles $\triangle ABC$ and $\triangle A'B'C'$ are *congruent* (which we write as

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

where the order in which the vertices are listed matters) if their sides and angles "match up" in pairs, so that

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

As a result of drawing and experimenting with our hands, our minds may realise that certain subsets of these six conditions suffice to imply the others. In particular:

SAS-congruence criterion: if

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

then

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

(where the name "SAS" indicates that the three listed match-ups occur in the specified order S (side), A (angle), S (side) as one goes round each triangle).

SSS-congruence criterion: if

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

then

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

ASA-congruence criterion: if

$$\angle ABC = \angle A'B'C', \quad \underline{BC} = \underline{B'C'}, \quad \angle BCA = \angle B'C'A',$$

then

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

If in a given triangle $\triangle ABC$ we have $\underline{AB} = \underline{AC}$, then we say that $\triangle ABC$ is *isosceles* with **apex** A, and **base** $\underline{BC}$ (*iso* = same, or equal; *sceles* = legs).

Problem 139 Let $\triangle ABC$ be an isosceles triangle with apex A. Let M be the midpoint of the base $\underline{BC}$. Prove that $\triangle AMB \equiv \triangle AMC$ and conclude that AM is perpendicular to the base $\underline{BC}$. $\triangle$

Problem 140 Construct two non-congruent triangles, $\triangle ABC$ and $\triangle A'B'C'$, where $\angle BCA = \angle B'C'A' = 30°$, $|\underline{CA}| = |\underline{C'A'}| = \sqrt{3}$, $|\underline{AB}| = |\underline{A'B'}| = 1$.

Conclude that there is in general no "ASS-congruence criterion". $\triangle$

The congruence criteria allow one to prove basic results such as:

Claim In any isosceles triangle $\triangle ABC$ with apex A (i.e. with $\underline{AB} = \underline{AC}$), the two base angles $\angle B$ and $\angle C$ are equal.

Proof 1 Let M be the midpoint of $\underline{BC}$.
Then $\triangle AMB \equiv \triangle AMC$ (by the SSS-congruence criterion, since
$\underline{AM} = \underline{AM}$,
$\underline{MB} = \underline{MC}$ (by construction of M as the midpoint)
$\underline{BA} = \underline{CA}$ (given)).
$\therefore \angle B = \angle ABM = \angle ACM = \angle C$. QED

Proof 2 $\triangle BAC \equiv \triangle CAB$ (by the SAS-congruence criterion, since
$\underline{BA} = \underline{CA}$ (given),
$\angle BAC = \angle CAB$ (same angle),
$\underline{AC} = \underline{AB}$ (given)).
$\therefore \angle B = \angle ABC = \angle ACB = \angle C$. QED

We also have the converse result:

Claim In any triangle $\triangle ABC$, if the base angles $\angle B$ and $\angle C$ are equal, then the triangle is isosceles with apex A (i.e. $\underline{AB} = \underline{AC}$).

Proof $\triangle ABC \equiv \triangle ACB$ (by the ASA-congruence criterion, since
$\angle ABC = \angle ACB$ (given),
$\underline{BC} = \underline{CB}$, and
$\angle BCA = \angle CBA$ (given)).
$\therefore \underline{AB} = \underline{AC}$. QED

Problem 141

(i) A circle with centre O passes through the point A. The line AO meets the circle again at B. If C is a third point on the circle, prove that $\angle ACB$ is equal to $\angle CAB + \angle CBA$.