# Proof patterns

Mark S. Joshi

# Contents

Preface ix

Chapter 1. Induction and complete induction 1
- 1.1. Introduction 1
- 1.2. Examples of induction 1
- 1.3. Why does induction hold? 3
- 1.4. Induction and binomials 3
- 1.5. Triangulating polygons 6
- 1.6. Problems 8

Chapter 2. Double Counting 11
- 2.1. Introduction 11
- 2.2. Summing numbers 11
- 2.3. Vandermonde's identity 13
- 2.4. Fermat's little theorem 13
- 2.5. Icosahedra 15
- 2.6. Pythagoras's theorem 15
- 2.7. Problems 16

Chapter 3. The pigeonhole principle 19
- 3.1. Introduction 19
- 3.2. Rationals and decimals 19
- 3.3. Lossless compression 21
- 3.4. More irrationality 22
- 3.5. Problems 22

Chapter 4. Divisions 25
- 4.1. Introduction 25
- 4.2. Division and well-ordering 25
- 4.3. Algorithms and highest common factors 26
- 4.4. Euclid's Lemma 28

4.5. The uniqueness of prime decompositions 29
4.6. Problems 30

Chapter 5. Contrapositive and contradiction 31
5.1. Introduction 31
5.2. An irrational example 31
5.3. The infinitude of primes 34
5.4. More irrationalities 35
5.5. The irrationality of $e$. 36
5.6. Which to prefer 38
5.7. Contrapositives and converses 38
5.8. The law of the excluded middle 38
5.9. Problems 39

Chapter 6. Intersection-enclosure and Generation 41
6.1. Introduction 41
6.2. Examples of problems 41
6.3. Advanced example 42
6.4. The pattern 43
6.5. Generation 44
6.6. Fields and square roots 46
6.7. Problems 48

Chapter 7. Invariance Difference 51
7.1. Introduction 51
7.2. Dominoes and triminoes 51
7.3. Dimension 52
7.4. Cardinality 56
7.5. Order 59
7.6. Divisibility 61
7.7. Problems 62

Chapter 8. Linear dependence, fields and transcendence 63
8.1. Introduction 63
8.2. Linear dependence 64
8.3. Linear dependence and algebraic numbers 67
8.4. Square roots and algebraic numbers 68
8.5. Transcendental numbers 69
8.6. Problems 69

Chapter 9. Formal equivalence 71

9.1. Introduction 71
9.2. Ruler and compass constructions 71
9.3. Further reading 78
9.4. Problems 78

Chapter 10. Extension-uniqueness 79
10.1. Introduction 79
10.2. Problem examples 79
10.3. The pattern 79
10.4. Problems 80

Chapter 11. Equivalence extension 81
11.1. Introduction 81
11.2. Constructing the integers 81
11.3. Constructing the rationals 84
11.4. The inadequacy of the rationals 87
11.5. Constructing the reals 89
11.6. Convergence of monotone sequences 93
11.7. Existence of square roots 93
11.8. Problems 94

Chapter 12. Proof by classification 95
12.1. Introduction 95
12.2. Co-prime square 95
12.3. Classifying Pythagorean triples 96
12.4. The non-existence of Pythagorean fourth powers 99
12.5. Problems 101

Chapter 13. Specific-generality 103
13.1. Introduction 103
13.2. Reducing the Fermat Theorem 103
13.3. The Four-Colour theorem 104
13.4. Problems 106

Chapter 14. Diagonal tricks and cardinality 107
14.1. Introduction 107
14.2. Definitions 107
14.3. Infinite sets of the same size 108
14.4. Diagonals 110
14.5. Transcendentals 113
14.6. Proving the Schröder–Bernstein theorem 113

14.7. Problems 116

Chapter 15. Connectedness and the Jordan curve theorem 117
15.1. Definitions 117
15.2. Components 118
15.3. The Jordan closed-curve theorem 120
15.4. Problems 123

Chapter 16. The Euler characteristic and the classification of regular polyhedra 125
16.1. Introduction 125
16.2. The Euler characteristic and surgery 126
16.3. Transforming the problem 127
16.4. The result for networks in the plane 129
16.5. Counter-examples 133
16.6. Classifying regular polyhedra 134
16.7. Problems 135

Chapter 17. Discharging 137
17.1. Introduction 137
17.2. The Euler characteristic via discharging 137
17.3. Maps and double counting 138
17.4. Inevitable configurations 141
17.5. Problems 141

Chapter 18. The matching problem 143
18.1. Introduction 143
18.2. Formulating the problem 143
18.3. The algorithm 144
18.4. Uniqueness 144
18.5. Further reading 145
18.6. Problems 145

Chapter 19. Games 147
19.1. Introduction 147
19.2. Defining a game 147
19.3. Termination 148
19.4. Optimal strategy 149
19.5. Second player never wins 150
19.6. Problems 150

Chapter 20. Analytical patterns 153
- 20.1. Introduction 153
- 20.2. The triangle inequality 153
- 20.3. The definition 154
- 20.4. Basic results 156
- 20.5. Series 161
- 20.6. Continuity 164
- 20.7. Theorems about continuous functions 166
- 20.8. The fundamental theorem of algebra 169
- 20.9. Further reading 172
- 20.10. Problems 172

Appendix A. Equivalence Relations 173

Bibliography 175

Index 177

# Preface

Patterns have become a common theme in many fields of academic study. In programming, the book “Design Patterns” is highly influential and it has become customary to discuss programs in terms of the patterns used. The programmers generally attribute the idea of a design pattern to architecture. The fundamental idea is that each field has a collection of ways of breaking down problems into component pieces. Understanding these methodologies explicitly then leads to greater comprehension, facilitates learning and simplifies problem solving. Rather than attempting a problem cold, one first sees whether known patterns work. Even if they all fail, understanding why they fail defines the problem.

In this book, my objective is to identify and teach many of the common patterns that arise in pure mathematics. I call these “proof patterns.” The main originality in the presentation is that examples are focussed about each pattern and drawn from different areas. This differs from the usual style of teaching pure mathematics where a topic is chosen and dissected; patterns are then drawn in as needed, and they are often not explicitly mentioned. After studying enough topics the learner picks up a variety of patterns, and the difficulty of studying a new area is often determined by the degree of unfamiliarity with its patterns.

This book is intended to do a variety of things. On one level, my objective is to teach the basic patterns. On another, it is intended as a taster for pure mathematics. The reader will gain a little knowledge on a variety of topics and hopefully learn a little about what pure mathematics is. On a third level, the intention of the book is to make a case for the explicit recognition of patterns when teaching pure mathematics. On a fourth level, it is simply an enjoyable romp through topics I love.

One powerful tool of pure mathematics which I intentionally avoid is that of abstraction. I believe that patterns and concepts are best learnt via the study of concrete objects wherever possible. Whilst one must go abstract eventually to obtain the full power and generality of results, a proof or pattern that has already been understood in a concrete setting is much easier to comprehend and apply.

The target reader of this book will already be familiar with the concept of proof but need not know much more. So whilst I assume very few results from pure mathematics, the reader who does not know what a proof is will struggle. There are several excellent texts such as Eccles, Velleman and Houston for such readers to study before reading here. In particular, I regard this book as a second book on proof, and my hope is that the reader will find that the approach here eases their study of many areas of pure mathematics.

I try to build up everything from the ground up as much as possible. I therefore try to avoid the "pull a big theorem out of the hat" style of mathematics presentation. The emphasis is much more on how to prove results rather than on trying to impress with theorems whose proofs are far beyond the book's scope.

Inevitably, as with many introductory books on proof, many examples are drawn from combinatorics and elementary number theory. This reflects the fact that these areas require fewer prerequisites than most and so patterns can be discussed in simple settings. However, I also draw on a variety of areas including group theory, linear algebra, computer science, analysis, topology, Euclidean geometry, and set theory to emphasize patterns' universality.

There is little if any originality in the mathematical results in this book: the objective was to provide a different presentation rather than new results. We look at the "Four-Colour problem" at various points. Our treatment is very much inspired by Robin Wilson's excellent book "Four colours suffice" and I recommend it to any reader whose interest has been piqued. A book with some similarities to this one but requiring a little more knowledge from the reader is "Proofs from the book" by Aigner and Ziegler. The emphasis there is more on beauty in proof than on patterns and it is a good follow on for the reader who wants more. However, I do hope that any reader of this book will develop some appreciation for the beauty of mathematics.

This book is ultimately an expression of my philosophy of how to approach the teaching of mathematics. My views have been shaped by interactions with innumerable former teachers, students and colleagues and I thank them all. I particularly thank Alan Beardon and Navin Ranasinghe for their detailed comments on a former version of the text.

Mark Joshi
Melbourne 2013.

## CHAPTER 1

# Induction and complete induction

### 1.1. Introduction

Two basic proof tools are the principle of induction and the related principle of complete induction. Each of these relies on certain properties of the natural numbers which we now explore.

The idea of induction is simple. We wish to prove that a set of statements, $P(n)$, hold for all $n \in \mathbb{N}$ greater than or equal to some natural number $k$. We first check it for $P(k)$, and then we show that if it holds for one value of $n$ then it also holds for the next one. In other words, we have to prove

- $P(k)$;
- $P(n) \implies P(n+1)$.

It then follows that it holds for all values of $n$ starting with $k$. Why? It holds for $k$ so putting $n = k$, it holds for $k + 1$. Putting $n = k + 1$, it holds for $k + 2$. Repeating, it holds for $k + 3, k + 4, k + 5, \dots,$. We discuss how to show that $P(n)$ really does hold for $n \geqslant k$ in Section 1.3.

Complete induction is a closely related tool. The difference is that we are allowed to use $P(l)$ for $k \leqslant l \leqslant n$ when proving $P(n+1)$ rather than just $P(n)$. This can be advantageous – for certain types of results, it is the truth of $P(l)$ for some much smaller $l$ that is useful rather than that of $P(n)$. It holds for similar reasons to ordinary induction and we will discuss the proofs that they hold together.

### 1.2. Examples of induction

Induction often come in useful when establishing formulas for sums. Let $f(n)$ be the sum of the first $n$ natural numbers. We want to prove

$$f(n) = \frac{n(n+1)}{2}.$$

In this case, $P(n)$ is the statement that the formula is correct for $f(n)$.

We certainly have $f(1) = 1$ so $P(1)$ is true. We now assume that $P(n)$ is true and try to establish that $P(n + 1)$ is true. In this context $P(n)$ is sometimes called the *inductive hypothesis.* We have

$$f(n + 1) = f(n) + n + 1,$$

by definition. We now substitute the formula which we *assumed* to be true for $n$, to obtain

$$f(n + 1) = \frac{n(n + 1)}{2} + (n + 1) = \frac{n(n + 1) + 2(n + 1)}{2}.$$

Simplifying,

$$f(n + 1) = \frac{(n + 2)(n + 1)}{2}.$$

We have shown that $P(n)$ implies $P(n + 1)$ and the results holds for all $n$ by induction.

Whilst proof by induction is often easy and in a case like this it will generally work if the result is true, it has the disadvantage that you have to already know the formula! It is therefore more a way of proving formulas rather than of finding them. It can therefore come in useful when a result has been guessed in some non-rigorous fashion and still has to be proven.

We now give a classic example of applying the principle of complete induction. We show that any natural number bigger than one can be written as a product of prime numbers. In this case, $P(n)$ is the statement that $n$ can be written as a product of product of primes.

We start with $P(2)$. It certainly holds since 2 is prime. We now assume the inductive hypothesis that $P(2), P(3), \dots, P(n)$ are true. Consider $n + 1$. Either it is prime in which case we are done, or it is composite. In the latter case, we can write

$$n + 1 = ab$$

with $2 \leqslant a, b \leqslant n$. We have assumed that $P(a)$ and $P(b)$ hold so we can write

$$a = p_1 p_2 \dots p_\alpha, \quad b = q_1 q_2 \dots q_\beta$$

for some primes $p_j$ and $q_j$ and $\alpha, \beta \in \mathbb{N}$. So

$$n + 1 = p_1 p_2 \dots p_\alpha q_1 q_2 \dots q_\beta$$

and $P(n + 1)$ holds. It follows from complete induction, that all natural numbers bigger than $1$ can be written as a product of primes. Note that this proof used complete induction in a non-trivial way; it is not at all clear how we could prove the result using only ordinary induction.

More generally, note that whilst we have established that every natural number bigger than $1$ is a product of primes, we have not shown that the representation is unique. Indeed without some extra restrictions, it is not:

$$2 \times 2 \times 3 = 12 = 2 \times 3 \times 2.$$

This non-uniqueness can be circumvented by requiring the primes to be ascending order, and the representation is then unique. Note the general technique here: impose additional structure to remove non-uniqueness. However, one still has to prove that this amount of extra structure is enough to gain uniqueness. We will prove that it is sufficient in Section 4.5.

## 1.3. Why does induction hold?

How can we prove that induction works? One solution is to use the well-ordering of the natural numbers: every non-empty subset of the natural numbers has a smallest element. The idea is essentially that there cannot be a least element that the statement does not hold for, so the set of such elements must be empty. More formally, the argument is then let $E$ be the set of $n$ for $n > k$ for which $P(n)$ is false.

If $E$ is non-empty then it has a least element $l$ and we know $l > k$. So $l - 1 \notin E$ so $P(l-1)$ holds. By the inductive hypothesis, $P(l)$ holds. So $l$ is both in $E$ and not in $E$. We have a contradiction. So $E$ has no least element and must be empty.

We can prove the principle of complete induction in a similar fashion. In fact, we can deduce it from the principle of induction directly. Let $Q(n)$ be the statement that $P(l)$ holds for $k \leqslant l \leqslant n$. We then have that $Q(k)$ holds, and that $Q(n)$ implies $Q(n+1)$ so it follows by induction that $Q(n)$ holds for all $n \geqslant k$. Since $Q(n)$ certainly implies $P(n)$, we also have that $P(n)$ holds and we are done.

These arguments are correct; however, we have proven the principle of induction by assuming the well ordering of the natural numbers. How can we prove that well ordering holds? Unfortunately, proofs really come down to the fact that the principle of induction holds! We have to take one of the two as an axiom and use it to deduce the other.

## 1.4. Induction and binomials

The *binomial coefficient* $\begin{pmatrix} n \\ k \end{pmatrix}$ expresses the number of ways that $k$ objects can be selected from $n$ objects with $n \geqslant k$. We do not care about the order of the objects selected. We have

$$\binom{n}{k} = \frac{n!}{(n-k)!k!}.$$

Why? The number of ways we can select the first object is $n$. The second is then $n-1$ since one is gone, and $n-2$ for the one after, and so on. So we can select $k$ objects in

$$n(n-1)\dots(n-k+1) = \frac{n!}{(n-k)!}$$

different ways. However, this is an ordered selection. We do not care about the ordering so we can divide again by the number of different orderings of $k$ objects and we get

$$\frac{n!}{(n-k)!k!},$$

as desired. Note that an immediate consequence of our interpretation of this fraction is that it represents a whole number! We always have that

$$k! \mid n(n-1)\dots(n-k+1)$$

which is not obvious. Interpreting a formula in a certain way can yield non-obvious properties: *proof by observation*.

We take $0! = 1$. This can be regarded as a definition, however, it makes sense in that $k!$ expresses the number of ways you can order $k$ objects. If we have no objects then there is only one way to order them so we should have $0! = 1$. We have

$$\binom{n}{0} = 1 = \binom{n}{n};$$

we also have

$$\binom{n}{1} = n = \binom{n}{n-1}.$$

An obvious symmetry exists

$$\binom{n}{k} = \binom{n}{n-k}.$$

We can also show *Pascal's identity*

$$\binom{n}{k} + \binom{n}{k-1} = \binom{n+1}{k},$$

for $1 \leqslant k \leqslant n$. This can be proven either via algebraic manipulation or by interpretation. We use the latter approach. We want to show that the number of subsets with $k$ elements taken from $\{1, 2, \dots, n+1\}$ is the left-hand-side of the identity. We show that any such subset corresponds to either a subset with $k$ elements of $1, 2, \ldots, n$ or one with $k-1$ elements. If our subset, $E$, with $k$ elements contains the element $n+1$ then discarding $n+1$ gives a subset of $\{1, \ldots, n\}$ with $k-1$ elements. Clearly all such subsets can be obtained this way. If $E$ does not contain $n+1$ then it is a subset of $\{1, 2, \ldots, n\}$ with $k$ elements, and again we can get all such subsets in this way. We have constructed a correspondence and the identity follows.

With Pascal's identity in hand, we can now prove something using induction.

Theorem 1.1. *The binomial theorem. If $n$ is a natural number, and $x, y$ are real numbers then*

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}.$$

Proof. If $n = 0$, both sides are equal to 1. Now suppose the result holds for $n$. We write

$$(x+y)^{n+1} = (x+y)(x+y)^n = x(x+y)^n + y(x+y)^n.$$

Using the inductive hypothesis,

$$(x+y)^{n+1} = \sum_{k=0}^{n} \binom{n}{k} \left( x^{k+1} y^{n-k} + x^k y^{n-k+1} \right).$$

We need to gather terms with the same powers of $x$ and $y$ together,

$$\binom{n}{k} x^{k+1} y^{n-k} = \binom{n}{l-1} x^l y^{n+1-l}$$

where $l = k+1$. So, letting $\binom{n}{-1} = 0$,

$$(x+y)^{n+1} = \sum_{k=0}^{n} \left( \binom{n}{k} + \binom{n}{k-1} \right) x^k y^{n-k+1} + \binom{n}{n} x^{n+1}.$$

Invoking, Pascal's identity and the fact that $\binom{n}{n} = 1 = \binom{n+1}{n+1}$, we have

$$(x+y)^{n+1} = \sum_{k=0}^{n+1} \binom{n+1}{k} x^k y^{n+1-k},$$

as required and the result follows by induction. ☐

Induction is not essential for the proof of the binomial theorem. Another approach is to think about how the coefficient of $x^k y^{n-k}$ arises when we work out the expansion of $(x+y)^n$. It occurs from picking the $x$ of $(x+y)$ in $k$ places out of $n$ possible ones. It can therefore arise $\left(\begin{array}{c} n \\ k \end{array}\right)$ different times and we get the binomial theorem.

With the binomial theorem proven, we can make various proofs by observation. First,

$$2^n = \sum_{k=0}^{n} \left(\begin{array}{c} n \\ k \end{array}\right). \tag{1.4.1}$$

To prove this, just put $x = y = 1$ in the binomial theorem. Once we interpret this result, it is actually clear for other reasons. We are adding the number of ways of choosing $k$ elements from $n$ for each $k$. We therefore counting the number of subsets of $\{1, 2, \ldots, n\}$. Each number is either in a given subset or not, so each number gives us two possibilities. There are $n$ numbers so there are $2^n$ subsets and we have (1.4.1). Our alternate proof is a proof by *double counting*; we counted a collection of objects in two different ways to establish a formula.

Another immediate consequence of the binomial theorem is

$$\sum_{k=0}^{n} \left(\begin{array}{c} n \\ k \end{array}\right) (-1)^k = 0. \tag{1.4.2}$$

Just set $x = -1$ and $y = 1$. We can rewrite this as

$$\sum_{k=0}^{k \leqslant n/2} \left(\begin{array}{c} n \\ 2k \end{array}\right) = \sum_{k=0}^{k<n/2} \left(\begin{array}{c} n \\ 2k+1 \end{array}\right). \tag{1.4.3}$$

The number of different ways of choosing a subset with an even number of elements equals the number of ways of choosing a subset with an odd number. This is clear when $n$ is odd; taking the complement provides a natural bijection between the two classes of subsets. It is not so obvious when $n$ is even.

## 1.5. Triangulating polygons

We now look at an application of complete induction to a quite different area. We prove that every polygon in the plane can be triangulated. In fact, we prove a stronger result, we show that it can be done without adding any extra vertices. Before proceeding to the proof, we discuss what a triangulation is. A polygon is a closed loop in the plane consisting of a sequence of straight line segments which

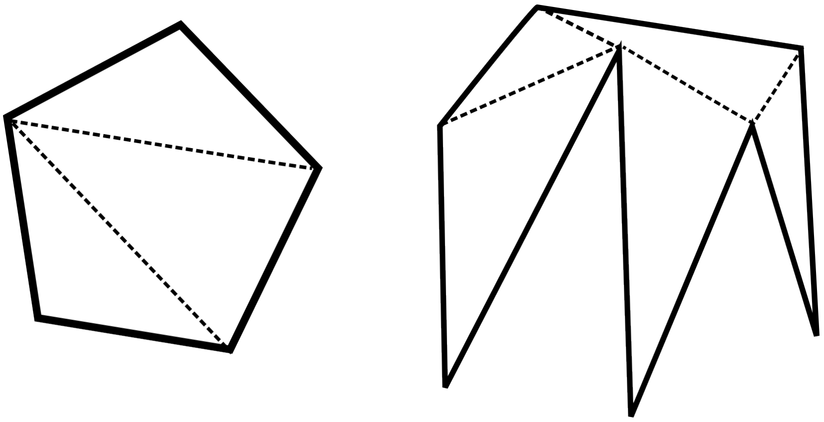

FIGURE 1.5.1. Examples of triangulations of polygons.

starts and finishes at the same point. The loop does not self intersect anywhere. To triangulate means to divide the polygon into triangles which only intersect along common sides. We give a couple of examples in Figure 1.5.1. We prove

THEOREM 1.2. *If $P$ is a polygon in the plane, it is possible to write $P$ as a union of triangles whose vertices are subsets of those of $P$ and which only intersect in common sides or vertices.*

PROOF. A polygon has at least 3 vertices. If it has 3 exactly then it is a triangle and the result is trivial. We now assume that any polygon with $3 \leqslant k < n$ sides can be triangulated. Let $P$ be a polygon with $n$ sides. It must have a vertex, $V$, where the interior angle between its two edges is less than 180 degrees. To see this observe that if the change in direction of the edge at a vertex is $x$ degrees then the interior angle is $180 - x$ degrees. Since all the changes of direction must add up to 360 degrees, at least one must have $x > 0$ and so that angle must be less than 180.

Now consider the two vertices next to $V$. Call them $A$ and $B$. We draw a line between them. See Figure 1.5.2. If this line's interior does not intersect $P$ then $ABC$ defines a triangle that we can cut off $P$. The remaining part of $P$ can be triangulated by the inductive hypothesis and so $P$ is triangulable.

If the line $AB$'s interior does intersect $P$, then we slide its endpoints along towards $V$. We do the sliding in such a way that they both reach $V$ at the same time. Since there are only a finite number of vertices in $P$, there will be a last time at which the line crosses a vertex. Call a vertex crossed at this last time $D$. (There could be more than one but this is not important.) The line from $A$ to $D$ will

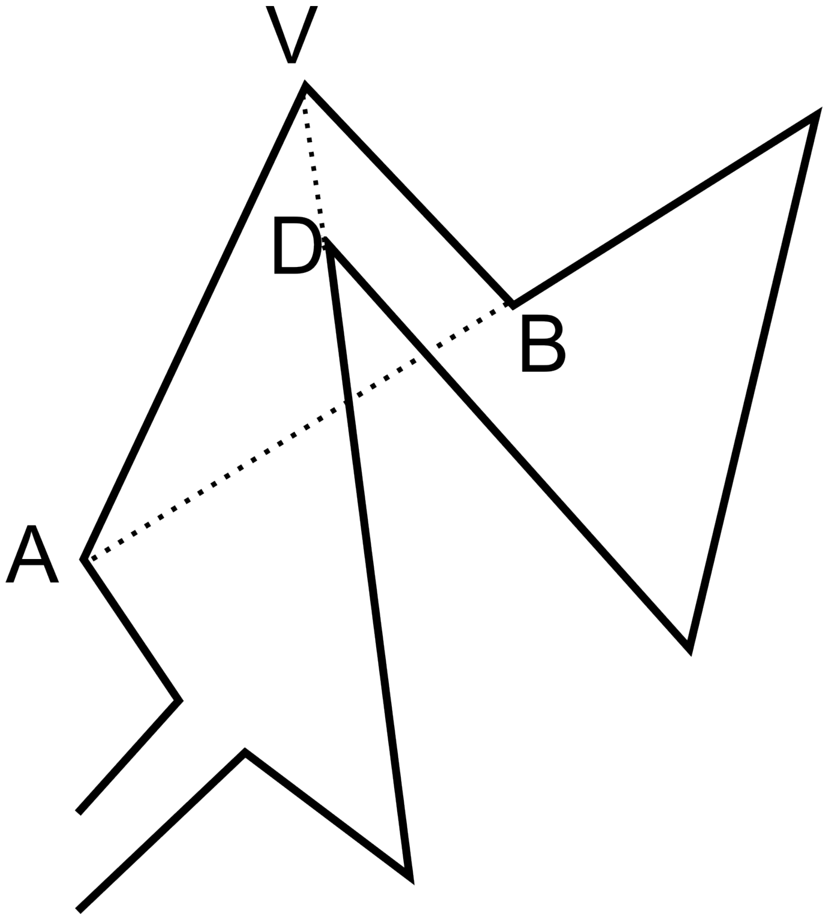

Figure 1.5.2. A proof of the triangulation of a polygon.

now not intersect $P$ and it divides $P$ into two smaller polygons. These can both be triangulated by the inductive hypothesis. The result now follows by complete induction. ☐

## 1.6. Problems

Exercise 1.1. *Develop an expression for the first* $n$ *odd numbers and prove that it holds using induction.*

EXERCISE 1.2. *Prove that for* $n \geqslant 1$,

$$\sum_{j=1}^{n} \frac{1}{\sqrt{j}} \geqslant \sqrt{n}.$$

EXERCISE 1.3. *A chess-style board is of size* $2^n \times 2^n$ *and has a single square deleted. A trimino is three squares joined together so as to have an angle; it is like a domino with a square stuck to the side. Show that for any* $n$, *the board can be covered by non-overlapping triminos.*

EXERCISE 1.4. *Prove that*

$$\sum_{j=1}^{n} j^2 = \frac{n(n+1)(2n+1)}{6}.$$

EXERCISE 1.5. *Prove that*

$$\sum_{j=1}^{n} 2^{j-1} = 2^n - 1.$$

EXERCISE 1.6. *Prove that if* $x > -1$ *then*

$$(1+x)^n \geqslant 1 + nx,$$

*for* $n \in \mathbb{N}$.

EXERCISE 1.7. *Prove that every number bigger than* $11$ *is a positive integer combination of* $4$ *and* $5$. *That is if* $k \in \mathbb{N}$, $k \geqslant 12$, *there exists* $a, b \in \mathbb{N}$ *such that*

$$k = 4a + 5b.$$

EXERCISE 1.8. *A polynomial is reducible if it can be written as a product of polynomials which are not constant. If it is not reducible, it is said to be irreducible. Prove that every non-constant polynomial is a product of irreducible polynomials.*

EXERCISE 1.9. *Show that if* $n \in \mathbb{N}$, $n^4 - n^2$ *is a multiple of* $12$.

EXERCISE 1.10. *A statement,* $P(k)$ *is defined for all* $k \in \mathbb{Z}$. *We know that* $P(l)$ *is true. We also know that for all* $n \in \mathbb{Z}$

$$\begin{aligned} P(n) &\implies P(n+1), \\ P(n) &\implies P(n-1). \end{aligned}$$

*Prove that* $P(n)$ *holds for all* $n \in \mathbb{Z}$.

CHAPTER 2

# Double Counting

## 2.1. Introduction

Often mathematicians want to develop formulas for sums of terms. As we have seen, induction is one way to establish the truth of such formulas. However, induction relies on foreknowledge of the formula which has to be derived or guessed in some other way first. Induction also does not assist with recall. An alternate way to find and derive many counting formulas is *double counting*. With this technique, we divide up a collection of objects in two different ways. The results of the two different divisions must agree and this can yield a formula for the more complicated one. Such approaches have the advantage that the technique for finding the formula is often more memorable than the formula itself.

## 2.2. Summing numbers

We wish to find the sum of the first $N$ numbers. Call this $x_N$. So

$$x_N = \sum_{j=1}^{N} j.$$

We can regard the sum as a triangle with 1 stone in the first row, two stones in the second, three in the third and so on. Now if we make a copy of the triangle and rotate it through 180 degrees, we get $N$ stones in the first row, $N-1$ in the second, $N-2$ in the third and so on. Joining the two triangles together, we have $N$ rows of $N+1$ stones. The total number of stones is $N(N+1)$. So

$$X_N = \frac{1}{2}N(N+1).$$

We could also give this proof algebraically. Before proceeding to the algebra, we look at a special case. If we have $5$ stones, then

$$\begin{aligned} X_5 =& 1+2+3+4+5, \\ X_5 =& 5+4+3+2+1, \\ 2X_5 =& (1+5)+(2+4)+(3+3)+(4+2)+(5+1). \end{aligned}$$

Now the algebra, if we reverse the order of the sum, we get

$$x_N = \sum_{j=1}^{N} (N+1-j).$$

Adding the two expressions for $x_N$ together,

$$2x_N = \sum_{j=1}^{N} (N+1) = N(N+1),$$

and the result follows.

A similar argument can be used for the sum of the first $N$ odd numbers

$$y_N = \sum_{j=1}^{N} (2j-1).$$

Either by rotating the triangle or by reversing the order of the sum, we have

$$y_N = \sum_{j=1}^{N} (2N+1-2j).$$

So

$$2y_N = \sum_{j=1}^{N} 2N.$$

We conclude that

$$y_N = N^2.$$

As well as being the sum of the first $N$ odd numbers, $N^2$ is also the sum of the first $N$ numbers plus the first $N-1$ numbers. One way to see this is take an $N \times N$ grid of stones and see how many stones lie on each upwards sloping diagonal. For the first $N$ diagonals, the $j$th diagonal has $j$ stones. These contribute the sum of the first $N$ numbers. After $N$, the length of the diagonals goes down by one each time. The second set therefore contribute the sum of the first $N-1$ numbers and the results follows.

## 2.3. Vandermonde's identity

Suppose we have $m + n$ jewels of varying sizes. There are $m$ rubies and $n$ sapphires. How many different ways can we select $r$ jewels to be placed on a bracelet? Clearly, the answer is

$$\binom{m+n}{r}.$$

Note that this since the jewels are all of different sizes, two different selections of $r$ jewels are essentially different. However, if we first think in terms of using $k$ rubies and $r - k$ sapphires, we see that the answer is also

$$\sum_{k=0}^{r} \binom{m}{k} \binom{n}{r-k}.$$

(We take the binomial coefficient to be zero when the inputs are out of their natural range. For example, if $r > m$, there are zero ways to choose $r$ rubies.) In conclusion, we have *Vandermonde's identity:*

$$\binom{m+n}{r} = \sum_{k=0}^{r} \binom{m}{k} \binom{n}{r-k}.$$

## 2.4. Fermat's little theorem

Fermat's little theorem states

THEOREM 2.1. *If $a$ is a positive integer and $p$ is prime then $p$ divides $a^p - a$.*

An elementary proof can be made using double counting.

PROOF. Consider strings of letters of length $p$. The letters are from the first $a$ letters in the alphabet. (If $a > 26$, we add extra letters to the alphabet.) How many such strings are there? Order matters, so we have $a$ choices in each slot and there are $p$ slots, so we get $a^p$ different strings.

Now consider the operation on these strings of chopping off an element at the end and reinserting it at the front. Call this $T_1$. We define $T_j$ to be the result of applying $T_1$ $j$ times. Clearly, $T_p$ is the identity map. We give some examples when

$p = 3$ and $a = 2$.

$$\begin{aligned} T_1(AAA) &= AAA, \\ T_1(ABA) &= AAB, \\ T_2(ABA) &= BAA. \end{aligned}$$

Two strings, $x$ and $y$, are said to be in the same *orbit* if there exists $j$ such that

$$T_j x = y.$$

Note that then

$$T_{p-j} y = x.$$

Note that if $x$ and $y$ are in the same orbit, and $y$ and $z$ are in the same orbit then $x$ and $z$ are too. So being in the same orbit is an equivalence relation. (See Appendix A for further discussion of equivalence relations.) This implies that every string is in exactly one orbit.

If a string is all one letter, eg “AAA”, then it is the only string in its orbit. There are $a$ such strings and so $a$ orbits of size $1$.

Now suppose a string $x$ has more than one letter in it. Consider the strings

$$x, Tx, T^2x, \dots, T^{p-1}x.$$

These will all be in the same orbit and everything in $x$’s orbit is of this form. If we keep going we just get the same strings over again since $T^p x = x$. There are at most $p$ elements in these orbits then. We show that when $p$ is prime there are exactly $p$ elements. If there were less than $p$ then for some $k < p$, we would have

$$T^k x = x.$$

This means that cutting off the last $k$ elements and sticking them at the front does not change the string. We also have

$$x = T^k x = T^{2k} x = T^{3k} x = T^{4k} x = \dots$$

This implies that $x$ is made of $p/k$ copies of the first $k$ elements. However, $p$ is prime so its only divisor are $1$ and $p$. If $k = 1$ then we have a string of elements the same which is the case we already discussed. If $k = p$ then we are just saying that $T^p x = x$ which is always true.

So we have two sorts of orbits, those with $p$ elements and those with $1$ element. We showed that there are $a$ of the second sort. Let there be $m$ of the first sort. Since there are $a^p$ strings in total, we have

$$a^p = mp + a,$$

So

$$pm = a^p - a.$$

This says precisely that $p$ divides $a^p - a$. □

This proof is due to Golomb (1956).

## 2.5. Icosahedra

Recall that an icosahedron is a Platonic solid with twenty faces. Every face is a triangle and the same number of faces meet at each vertex. How many edges does an icosahedron have? Call this number $E$. We know that there are $20$ faces and that each face is triangle. Define an edge-face pair to be face together with one of the sides of the face which is, of course, an edge of the icosahedron. There are $60$ such pairs since each face has 3 sides.

Each edge of the icosahedron lies in precisely two sides. The number of edge-face pairs is therefore double the number of edges. That is

$$2E = 60$$

and so $E = 30$.

## 2.6. Pythagoras's theorem

The reader will already be familiar with the theorem that for a right-angled triangle, the square of the hypotenuse is equal to the sum of the squares of the other two sides. So if the sides are $a$, $b$ and $c$, with $c$ the longest side, we have

$$c^2 = a^2 + b^2.$$

We can use an extension of the double counting pattern to prove this theorem. Instead of using equal numbers of objects, we use equal areas. We take the triangle and fix a square with side length $c$ to its side of that length. We then fix a copy of the triangle to each of the square's other sides. See Figure 2.6.1. At each vertex of the square, we get 3 angles, each of which is one of the angles of the triangle so these add up to $180$ degrees and make a straight line. We now have two squares: the big one has side $a + b$ and the small one $c$. The former's area is

$$(a + b)^2 = a^2 + 2ab + b^2.$$

We can also regard it as the small one plus $4$ copies of the triangle and so it has area

$$c^2 + 4 \times 0.5ab = c^2 + 2ab.$$

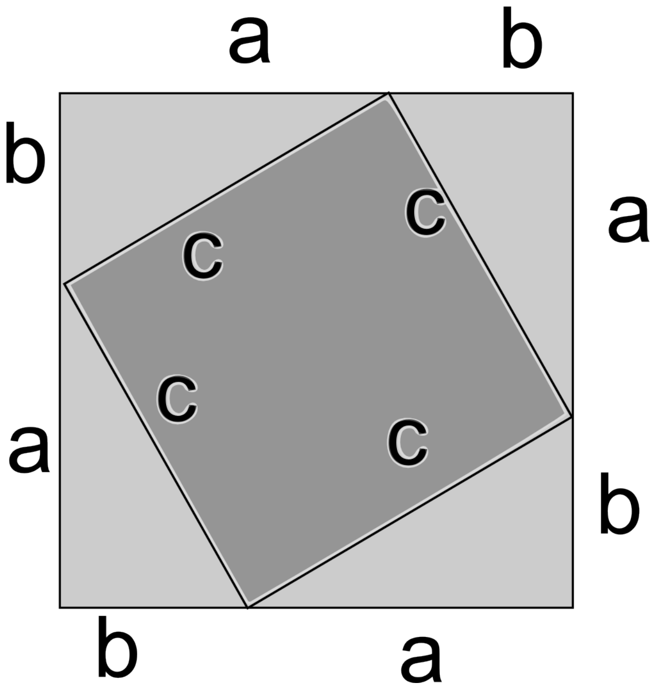

FIGURE 2.6.1. Four right-angled triangle with sides $a, b, c$, placed on a square of side $c$.

Equating these two, we get

$$a^2 + b^2 = c^2,$$

as required. Our proof is complete.

## 2.7. Problems

EXERCISE 2.1. *Let $a$ and $b$ be integers. Develop a formula for*

$$\sum_{j=1}^{n} a + jb.$$

EXERCISE 2.2. *How many vertices does an icosahedron have?*

EXERCISE 2.3. *If $p$ is a prime and not $2$, and $a$ is an integer, show that $2$ divides into $a^p - a$. (Try to construct a bijection on the set of size $p$ orbits which pairs them.)*

EXERCISE 2.4. *Suppose we have three sorts of jewels and apply the arguments for Vandermonde's identity, what formula do we find?*

CHAPTER 3

# The pigeonhole principle

## 3.1. Introduction

A pigeonhole is another name for a mailbox. If we have more pieces of mail than pigeonholes, someone's pigeonhole gets two pieces of mail. That is the pigeonhole principle. Whilst easy to state and rather obvious, it is very useful.

More formally, let $A$ and $B$ be sets such that the cardinality of $A$ is greater than that of $B$. Suppose a function $f$ maps from $A$ to $B$ then $f$ is not injective. That is there exists $x, y \in A$ such that

$$f(x) = f(y).$$

In this chapter, we look at a variety of applications from some quite different areas.

## 3.2. Rationals and decimals

The reader will be familiar with the fact that some rational numbers are not easy to represent with decimals. In particular, their expansions repeat. For example,

$$\frac{1}{3} = 0.3333333\ldots$$

and

$$\frac{1}{9} = 0.111111\ldots$$

More interestingly,

$$\frac{1}{7} = 0.142857142857142857142857\ldots$$

which repeats every 6 places.

Two obvious questions arise:

- If a decimal expansion repeats, will it always represent a rational number?
- If a number is rational, will its expansion always be finite or repeating?

The first of these is easy. Suppose the expansion of $x$ repeats every $k$ places after some point in the expansion. Then $y = 10^k x - x$ is zero after that point in the expansion. So we can make $y$ into an integer simply by multiplying it by some power of 10. We then have for some $l, m \in \mathbb{N}$

$$10^l(10^k - 1)x = m.$$

Equivalently,

$$x = \frac{m}{10^l(10^k - 1)},$$

so $x$ is rational.

For the second question, we analyze the algorithm that produces the terms in the decimal expansion. We have $q = m/n$, with $m$ and $n$ natural numbers. We take $q > 0$, since the result will follow for negative $q$ simply by putting a minus sign in front.

If we proceed using the arithmetic algorithms we learnt in school, then eventually, we get to a point in the decimal expansion computation where we take a number $r_0$ and write

$$r_0 = a_0 n + b_0,$$

with $0 \leqslant b_0 \leqslant n - 1$. The number $a_0$ is the term in the decimal expansion and $b_0$ is the remainder we use for the next term. We then put $r_1 = 10b_0$, and set

$$r_1 = a_1 n + b_1,$$

with $0 \leqslant b_1 \leqslant n - 1$. If we get a zero remainder the expansion terminates and we have a finitely long decimal. Otherwise, we keep going, setting $r_j = 10b_{j-1}$, and

$$r_j = a_j n + b_j.$$

If for some $j, k$ with $j < k$, $b_j = b_k$ then the expansion starting at place $k$ is the same as the one starting at place $j$. This is because the algorithm translated from $j$ to $k$ has the same inputs and the same operations.

However, if we consider the values

$$b_0, b_1, \ldots, b_{n-1},$$

there are $n$ of these. If any is zero, the expansion terminates. If none is zero then they take values from 1 to $n - 1$. The pigeonhole principle then guarantees that two of them are equal. So for some $j, k$ with $j < k \leqslant n - 1$,

$$b_j = b_k,$$

and the decimal repeats with period $k - j$. Note that the maximum period of repetition is, in fact, $n - 1$.

A consequence of our results is that we can characterize irrational numbers as those that have infinite non-repeating expansions. Note also that if a number has a finite or repeating expansion for one number base then it will have one of the two for all number bases.

### 3.3. Lossless compression

I once saw an advertisement for a compression program that claimed to reduce the size of any file by at least $20\%$. This did not seem very believable so I asked a colleague if this was really possible. His response was that he doubted that it could compress the same file twice!

We can use the pigeonhole principle to prove that an algorithm that losslessly compresses any file is impossible. By lossless compression, we mean that it is possible to recover the original file purely from the data in the new file. For compression, the new file has to be smaller than the old file. A file in a modern digital computer is no more and no less than a sequence of $1$s and $0$s. The amount of memory used is the length of the sequences.

A compression algorithm maps the sequences of length $m$ to a sequence of length $n$ with $m > n$. There are $2^m$ sequences of length $m$ and $2^n$ sequences of length $n$. If we fill out any sequences of length less than $n$ using zeros to be of length $m - 1$, then a lossless compression algorithm would give an injection from those of length $m$ to those of length $m - 1$. However, a map from the sequences of length $m$ to those of length $m - 1$ cannot be injective by the pigeonhole principle. In other words, it must be the case that two files get mapped to the same target file, and the decompression algorithm will not be able to say which is which. Lossless compression cannot occur.

However, most people use lossless compression programs all the time. So what is going on? In practice, most files, that have not already been compressed already, have some structure that the compression program can work with. For example, a musical recording will display quite different characteristics from a text file or an image file. By recognizing this internal structure, the compression program can exploit it and compress the data. In particular, different compression formats are tuned to different sorts of files. We use mp3 files to store our music but not for our images.

### 3.4. More irrationality

Let $x > 0$ be an irrational number. Consider the set

$$S_x = \{[nx] \text{ for } n \in \mathbb{Z}\}.$$

Here $[y]$ is the fractional part of $y$, that is the part obtained after discarding the whole number part: the bit after the decimal point if you prefer. More formally, it is the result after subtracting the largest integer less than or equal to it so it is always in range $[0, 1)$. We use the pigeonhole principle to show that $S_x$ contains a number arbitrarily close to zero: we show that if $\delta > 0$, then there exists $z \in S_x$ such that

$$0 < z < \delta.$$

All elements of $S_x$ are positive, since if there was a zero the number $x$ would be rational. So $S_x \subset (0, 1)$. Now suppose we pick $n$ such that $n\delta > 1$. If we consider the intervals $(j/n, (j+1)/n]$ for $j = 0, \dots, n-1$, there are $n$ such intervals. This means that if we take the first $n+1$ points in $S_x$, at least two must lie in one of these intervals by the pigeonhole principle. We therefore have values $l$ and $m$ with $m > l$, such that

$$|[lx] - [mx]| < 1/n < \delta.$$

In other words, there exist integers $a$ and $b$ such that

$$lx \in (a + j/n, a + (j+1)/n), \quad mx \in (b + j/n, b + (j+1)/n).$$

Now consider $(m - l)x = mx - lx$. This number is certainly positive and it is clearly in the interval

$$(b - a - 1/n, b - a + 1/n).$$

If it is bigger then $b - a$, its fractional part is less than $1/n$ and we are done. Otherwise, consider $(l - m)x$. This will lie in the range

$$(a - b, a - b + 1/n),$$

and taking its fractional part we are done. The crucial part of this argument was that the pigeonhole principle implied that eventually the sequence of numbers $[nx]$ had to bunch together.

### 3.5. Problems

EXERCISE 3.1. *Show that if a lossless compression algorithm exists for any file, then it is possible to represent any file by a single bit of information.*

EXERCISE 3.2. *Show that there are two residents of London with the same numbers of hairs on their heads.*

EXERCISE 3.3. *Suppose* $N$ *people are at a party. Some have met before and some have not. Show that there are two people who have met the same number of other people before.*

EXERCISE 3.4. *Suppose we take a set of* $101$ *different integers between* $1$ *and* $200$. *Show that there is a pair such that one divides the other.*

EXERCISE 3.5. *Represent the following decimals as ratios of integers.*

- $0.1212121212\ldots,$
- $0.123123123123123\ldots,$
- $0.456456456\ldots.$

CHAPTER 4

# Divisions

## 4.1. Introduction

In this chapter, we explore some basic results about division. These allow us to illustrate some proof techniques. The first result we want to prove is sometimes called the Division Lemma – it says that we can always write the answer as a whole number plus a remainder just as we did in primary school! To prove a result, we first have to formulate it correctly. Once that is done, we will see how it follows from more basic properties of natural numbers and, in particular, the *well-ordering principle.*

Our second task is to show that if two numbers, $m$ and $n$, have highest common factor $h$ then there exist integers $a$ and $b$ such that

$$am + bn = h.$$

We will show this via *algorithmic construction*. That is we construct an algorithm whose output is the numbers $a, b$ and $h$ and we prove that it always terminates.

## 4.2. Division and well-ordering

We have a natural number $m$ and we want to divide it by another integer $n.$ If we work solely with whole numbers, then if we proceed as we did when first learning arithmetic, we write

$$m = nq + r$$

with $q, r$ natural numbers, and $0 \leqslant r < n.$

The Division Lemma states that such a decomposition is always possible. The standard way to prove it is to deduce it from the well-ordering of the natural numbers.

Definition 4.1. *A set $E$ with an ordering $<$ is said to be well-ordered if every non-empty subset, $F$, of $E$ has a smallest element. In other words, there exists $f \in F$, such that for all $g \in F$, we have $f \leqslant g$.*

The standard ordering of the natural numbers is a well-ordering. We will take this for granted for now.

How does well-ordering help us? Let

$$F = \{m - nq \mid q \in \mathbb{Z},\ m - nq \geqslant 0\}.$$

Note that if $x \in F$ and $x \geqslant n$ then $x - n \in F$ also. So any element bigger than $n - 1$ is not the smallest element of $F$. The set $F$ is non-empty since it contains $m$, so by the well-ordering principle it has a smallest element. Let $r$ be the smallest element of $F$. It must be smaller than $n$, since any element greater than or equal to $n$ is not the smallest.

We therefore have

$$m = nq + r$$

with $0 \leqslant r < n$, as required.

### 4.3. Algorithms and highest common factors

The highest common factor of two natural numbers, $m$ and $n$, is the largest natural number, $h$, which divides into both of them with no remainders. This definition presupposes that such a number exists! It is, however, easy to show that there is such a number. First, the set of factors is non-empty since it contains $1$. Second, any factor of a number is less than or equal to it. So the largest element is less than or equal to $\min(m, n)$. We therefore have a finite set of common factors and its largest element is the highest common factors. (Note the implicit use of well-ordering here – we are using that a finite non-empty subset of the integers has a largest element. How would you prove this?)

Our proof of the existence of the highest common factor is not very constructive – it does not tell us how find it. Indeed, the implicit method is to test every number from $1$ to $\min(m, n)$ and see if it divides both $m$ and $n$. Once this is done, take the largest such number. Whilst that algorithm would work, it is not very efficient. How could we improve it? One simple improvement is simply to count downwards and stop as a soon a common divisor is reached. This is still not very efficient, however.

Fortunately, there is a much better method known as the *Euclidean algorithm*. We will write $(m, n)$ for the highest common factor from now on to simplify notation. It relies on the observation that if

$$m = qn + r$$

with $m, n, q, r$ all natural numbers then

$$(m, n) = (n, r).$$

To see this, note that we can also write

$$r = m - qn.$$

If $x$ divides $a$ and $b$ then it also divides $a + b$ and $a - b$. Setting $a = qn$ and $b = r$, any factor of $r$ and $n$ is a factor of $m$. With $a = m$ and $b = qn$, any factor of $n$ and $m$ is also a factor of $r$. Since the set of common factors of $m$ and $n$ is the same as the set of common factors of $n$ and $r$, the two pairs must have the same highest common factor.

There is an obvious choice for $r$ and $q$: the results of the Division lemma. This will make $r$ as small as possible. If $r$ is zero, then we have

$$m = qn$$

which means that the highest common factor is $n$. Otherwise, we can repeat, and

$$n = q_1 r + r_1,$$

with $r_1 < r$. We have

$$(m, n) = (n, r) = (r, r_1).$$

Letting $r_0 = r$, we can now keep going until we obtain a zero remainder.

$$(m, n) = (r_j, r_{j+1})$$

for all $j$ with $r_{j+1} > 0$.

Once we hit a point with zero remainder, the algorithm terminates. We need to show that it does. However, we always have

$$r_{j+1} < r_j,$$

so it must terminate in at most $n$ steps.

We do an example with $m = 57$ and $n = 51$.

$$\begin{aligned} 57 &= 1 \times 51 + 6, \\ 51 &= 8 \times 6 + 3, \\ 6 &= 2 \times 3 + 0. \end{aligned}$$

The highest common factor is 3.

A consequence of the Euclidean algorithm is that we can always express a highest common factor as an integer combination of the two original numbers.

$$3 = 51 - 8 \times 6 = 51 - 8 \times (57 - 51) = 9 \times 51 - 8 \times 57.$$

We simply work our way backwards through the algorithm. We can formally prove the result using complete induction. We assume $n \leqslant m$ without loss of generality since they can always be switched. First note that if $n$ is the highest common factor then the result holds

$$n = 1 \times n.$$

Otherwise, let $r_j$ be the remainder after $j + 1$ steps. For convenience, let $r_{-1} = n$. The first remainder $r_0$ is certainly an integer combination since

$$r_0 = m - qn.$$

Now assume the result holds for $r_l$ for $l \leqslant j$. We have

$$r_{j-1} = q_j r_j + r_{j+1},$$

and so

$$r_{j+1} = r_{j-1} - q_j r_j$$

Substituting the linear combinations for $r_{j-1}$ and $r_j$, we have a linear combination for $r_{j+1}$ and the result follows.

We have proven

Theorem 4.1. *If $m$ and $n$ are natural numbers then there exist integers $a$ and $b$ such that*

$$(m, n) = am + bn.$$

Our proof proceeded by using an algorithm that constructed the numbers $a$ and $b$. We also proved that the algorithm does terminate. This is an example of *algorithmic construction.*

## 4.4. Euclid's Lemma

We can now use the result of the last section, to prove a result about factors which is sometimes called Euclid's Lemma.

Lemma 4.1. *Suppose $k, m,$ and $n$ are natural numbers such that $(k, m) = 1$ and $k$ is factor of $mn$, then $k$ is a factor of $n$.*

Proof. We have that there exists integers $a$ and $b$ such that

$$am + bk = 1.$$

So

$$amn + bkn = n.$$

We have that $mn$ is a multiple of $k$ so $amn$ is. So

$$amn = \alpha k,$$

for some $\alpha$. Clearly, $bkn$ is $nb$ times $k$. So

$$n = (\alpha + bn)k$$

and $k$ is a factor of $n$. ☐

A nice consequence of this Lemma is

COROLLARY 4.1. *Let $m$ and $n$ be positive natural numbers. If $p$ is a prime or 1, and $p|mn$ then $p$ divides at least one of $m$ and $n$.*

PROOF. If $p$ is 1 it divides anything so assume that $p$ is prime. If $p$ divides $m$ then we are done. Otherwise, the highest common factor of $m$ and $p$ is 1, since the highest common factor must divide into $p$ and its only factors are $p$ and 1. Euclid's lemma then states that $p$ divides $n$ and we are done. ☐

Note many authors state this lemma with the hypothesis that $p$ is prime rather than that $p$ is prime or 1. The more general version is certainly true and it will be convenient later when we are trying to prove that a certain number is 1 to allow its possibility here.

## 4.5. The uniqueness of prime decompositions

We know from Section 1.2 that every natural number bigger than one can be written as a product of positive integer powers of prime numbers. So given $m$, there exists $p_j$ prime and $\alpha_j \in \mathbb{N}, \alpha_j > 0$, such that

$$m = p_1^{\alpha_1} p_2^{\alpha_2} \dots p_k^{\alpha_k}.$$

Rearranging if necessary, we can assume that $p_i < p_{i+1}$ for all $i$. We want to show that this representation is unique. Suppose we also have primes $q_j$ and powers $\beta_j$ with the same properties. We need to show that

$$q_j = p_j$$

for all $j$ and $\alpha_j = \beta_j$.

If $p$ and $q$ are prime numbers they are either equal or coprime. We have that for each $l$

$$q_l | p_1^{\alpha_1} p_2^{\alpha_2} \dots p_k^{\alpha_k}.$$

From Corollary 4.1, we have that $q_l$ divides at least one of

$$p_1^{\alpha_1} \text{ and } p_2^{\alpha_2} \dots p_k^{\alpha_k}.$$

Repeating the argument we see that for some $r$,

$$q_l | p_r^{\alpha_r}.$$

If $\alpha_l > 1$, we can argument in same way again that $q_l$ divides $p_r$ or $p_r^{\alpha_r - 1}$. Repeating, we have

$$q_l | p_r$$

so $p_r = q_l$.

Since the problem is symmetric in the $p$s and $q$s, for every $p_r$ there exists an $l$ such that $p_r = q_l$. In other words, we have the same sets of primes.

It remains to show that the powers are the same. We need to show that if

$$p_1^{\alpha_1} p_2^{\alpha_2} \dots p_k^{\alpha_k} = p_1^{\beta_1} p_2^{\beta_2} \dots p_k^{\beta_k},$$

with $p_j$ prime and $\alpha_l, \beta_l$ positive natural numbers then $\alpha_l = \beta_l$ for all $l$.

For each point where $\alpha_l \geqslant \beta_l$, we divide both sides by $p_l^{\beta_l}$. For any remaining values of $l$, we divide both sides by $p_l^{\alpha_l}$. We then have primes to the power $\alpha_l - \beta_l$ on the left hand side and primes to the power $\beta_l - \alpha_l$ on the right hand side. The sets of primes on each side with non-zero powers are disjoint.

If we repeat the argument above where we showed that the primes on each side must be the same, we realize that all the powers must be zero. In others $\alpha_l = \beta_l$ for all $l$ and we are done.

## 4.6. Problems

EXERCISE 4.1. *Find the highest common factors of the following number pairs.*

- $1236$ *and* $369$.
- $144$ *and* $900$.
- $99$ *and* $36$.

EXERCISE 4.2. *Show that the highest common factor of* $n$ *and* $n + 1$ *is* $1$.

EXERCISE 4.3. *Show that the highest common factor of* $n$ *and* $n^2 + 1$ *is* $1$.

EXERCISE 4.4. *What are the possible highest common factors* $n$ *and* $n^2 + k$ *if* $k < n$*?*

CHAPTER 5

# Contrapositive and contradiction

## 5.1. Introduction

Two very common proof tools are proof by contrapositive and proof by contradiction. Whilst these are related, they are distinct notions and it is useful to distinguish between them. They are both really principles of logic rather than of mathematics.

In what follows let $P$, $Q$ and $R$ be mathematical statements that may be true or false. We write $\neg P$ for "not $P$" and so on. We write $P\&Q$ for the statement that both $P$ and $Q$ hold. Proof by contrapositive states that the following two statements are logically equivalent:

$$\begin{aligned} P &\implies Q; \\ \neg Q &\implies \neg P. \end{aligned}$$

If we can prove one of them, then the other holds. If we can disprove one, then the other is false.

Proof by contradiction involves a third statement $R$. It states that if there is a statement $R$ such that

$$\begin{aligned} P\&\neg Q &\implies R, \text{ and} \\ P\&\neg Q &\implies \neg R \end{aligned}$$

then

$$P \implies Q.$$

## 5.2. An irrational example

Many proofs by contrapositive are phrased as proofs by contradiction. Namely, a statement $Q$ is assumed to be false and this is shown to lead to the fact that $P$ is false. So if $P$ is assumed to be true, then $P\&\neg Q$ implies that both $P$ and $\neg P$ are true. So if we set $R$, above, to be $P$ then we have a proof by contradiction, and we have that $P$ implies $Q$. However, if we did not actually use $P$ in the proof that $\neg Q$ implies $\neg P$ then we could equally well have invoked proof by contrapositive and not have had a contradiction.

A standard well-known example of proof by contradiction is the irrationality of the square root of 2. The argument goes as follows.

- Suppose $\sqrt{2}$ is rational then there exists $m, n$ positive natural numbers such that $2 = m^2/n^2$ where $m$ and $n$ have no common factors.
- We can write $m^2 = 2n^2$.
- So $m^2$ is even. The number $m$ is even or odd. The square of an odd number is odd so $m$ is not odd.
- So $m$ is even and $m = 2k$ for some integer $k$.
- $2n^2 = 4k^2$ so $n^2 = 2k^2$.
- So $n^2$ is even and by the same argument as above, $n$ is even.
- So 2 is a common factor of $m$ and $n$, but they were said to have no common factors so we have a contradiction.
- Our initial assumption must be false and $\sqrt{2}$ is not rational.

Some points to note about this proof. Provided we phrase correctly, we do not need $\sqrt{2}$ to exist as a real number for this proof; the statement that lead to a contradiction was

$$2 = (m/n)^2$$

not that $\sqrt{2} = m/n$. The difference being that the latter statement requires us to work in a bigger number system where $\sqrt{2}$ makes sense. Second, we used the fact that any rational number can be expressed as a ratio of two integers that have no common factors. Whilst this is true, it arguably needs to be proven.

Before considering contrapositives, it is worth thinking about what we have proven and to what extent the result is generalizable. Our main result is that

$$x^2 = 2,$$

has no rational solutions. A natural question to ask is “what is special about 2”? Of course, many things are special about 2. It is even, it is prime, and it appears twice in this equation. The crucial argument in the proof was that if 2 divided the square of a number, $m$, then 2 also divides $m$. So we can expect the same result to hold for the square root of any number $k$ that has this property. Prime numbers have this property: see Section 4.4

$$p|m^2 \implies p|m, \tag{5.2.1}$$

so it follows that the square root of any prime is also irrational, (provided we have proven (5.2.1).)

Perfect squares trivially have rational square roots. However, there are many numbers that are neither prime nor are perfect squares. The smallest such number is $6$, another one is $12$. If we use a calculator to compute their square roots, they certainly look irrational. My computer gives

$$\sqrt{6} = 2.44948974278318 \text{ and } \sqrt{12} = 3.46410161513775.$$

Remember that computer numbers are always approximations; this does not mean that there are not more digits. We certainly have $12|36$ but $12$ does not divide into $6$. So our proof does not generalize to $12$. Of course, that does not mean that $\sqrt{12}$ is not irrational, merely that we need a different method of proof.

What if we proceed via contrapositive? Our new statement is: if $q$ is rational and not an integer then $q^2$ is not an integer. If we can we prove this, then we have shown that all non-perfect squares have irrational square roots. Even more generally, we could consider the analogous statement for $k$th powers: for $k \geqslant 2$, if $q$ is rational and not an integer then $q^k$ is not an integer. Once this is proven, we will have that all $k$th roots are integers or irrational.

Now for the proof: if $q$ is rational and not an integer then $q$ can be written as $m/n$ with $n > 1$ and $n$ does not divide into $m$. If $m$ and $n$ have a highest common factor, $h$, bigger than 1 then $h$ is less than $n$ since $n$ does not divide into $m$. We can replace $m$ and $n$ by $m/h$ and $n/h$. Having done so, we have that $m$ and $n$ have highest common factor 1 and we still have $n > 1$. Now consider

$$q^k = \frac{m^k}{n^k}.$$

This will be an integer if and only if $n^k$ divides into $m^k$. We show that it cannot. We show that no number bigger than 1 divides both $n^k$ and $m^k$. This will follow if we show no prime number divides both numbers, since any number is prime or composite, and any composite number will have a prime factor. Let $p$ be a number that is either $1$ or a prime such that

$$p|m^k \text{ and } p|n^k.$$

We need a lemma:

LEMMA 5.1. *Let $l$ be a positive integer. If a number, $p$, that is $1$ or prime divides $l^k$ for $k \geqslant 2$ then $p|l$.*

Given this lemma, we have that $p$ divides $m$ and $n$, but $m$ and $n$ have highest common factor 1 so $p$ must be 1. So $q^k$ is not an integer and every root of an integer that is not an integer is irrational.

We still have to prove our lemma.

PROOF. We use Corollary 4.1. If $p$ is 1 there is nothing to prove, so assume that $p$ is prime. We use induction. If $k = 2$, then this is just Corollary 4.1. For general $k$ assume the result is known for $k$ and suppose $p|l^{k+1}$ then writing

$$l^{k+1} = l^k l,$$

we have that either $p|l$ or $p|l^k$. In the first case, we are done, and in the second, the result follows immediately from the inductive hypothesis and we are done. ☐

Although the irrationality of the square root is a standard example of the power of proof by contradiction, we have seen that is perfectly possible to prove it by contrapositive. We have also proven a much more general result: any $k$th root of a positive integer that is not an integer is irrational.

### 5.3. The infinitude of primes

There are an infinite number of primes. Before proceeding to the proof. We first establish a lemma.

LEMMA 5.2. *Every integer, $y$, bigger than 1 has a prime factor.*

PROOF. If $y$ is prime, we are done. Otherwise, $y = k_1 l_1$ with $k_1, l_1 > 1$. Note that $k_1 \leqslant y/2$. Either $k_1$ is prime and we are done, or $k_1 = k_2 l_2$, with $k_2, l_2 > 1$. We now have

$$k_2 \leqslant k_1/2 \leqslant k_1/4.$$

Either $k_2$ is prime or we can construct $k_2 = k_3 l_3$ and so on. Since $2^y \geqslant y$, this process must halt by step $y$ and we are done. One could also prove this result using complete induction – the reader is encouraged to do so! ☐

How can we prove the infinity of primes?

First, we use proof by contradiction.

PROOF. Suppose there are a finite number, $N$. Then we can label them, $p_1, p_2, \ldots, p_N$ to make a list. If we now let

$$x = p_1 p_2 p_3 \ldots p_N + 1,$$

then $x$ is either prime or composite. It cannot be prime since it is not on the list. We also have that $x$ divided by $p_j$ has remainder $1$ for each $j$ so it is not a product of the primes on our list either. However, by our lemma $x$ has a prime factor $p$, but $p$ is not on our list. We have a contradiction: assuming the list was finite and complete led to the existence of a prime not on the list. So our initial assumption that a finite list could be complete was false, and there are an infinite number of times. ☐

However, we do not really need proof by contradiction.

Proof. We show that any finite list of primes is not complete, and so the number of primes must be infinite. Let $p_1, p_2, \ldots, p_N$ be a list of primes. Let

$$x = p_1 p_2 p_3 \ldots p_N + 1.$$

From our lemma, $x$ has a prime factor, $p$. Each $p_j$ does not divide $x$, since the remainder is $1$. So $p_j \neq p$ for all $j$. We have constructed a prime not on the list, and so the list is not complete, as claimed. The number of primes is therefore infinite. ☐

The two arguments are very similar. However, in the first, we gratuitously assumed that the final result was false in order get a contradiction. In the second, by using a slightly different phrasing, a contradiction was not required.

## 5.4. More irrationalities

Suppose $x$ is irrational and $q$ is rational. What can we say about $y = x + q$? If $y$ is rational, then we have for some integers, $k, l, m, n$,

$$\frac{k}{l} = x + \frac{m}{n}.$$

It follows that

$$x = \frac{k}{l} - \frac{m}{n} = \frac{kn - lm}{ln}.$$

The right hand side is a rational number so this contradicts the irrationality of $x$. Invoking proof by contradiction, we conclude that $y$ is not rational.

Alternatively, we could employ proof by contrapositive. Our contrapositive is

Proposition 5.1. *If $q_1$ and $q_2$ are rational and for some $x \in \mathbb{R}$, we have*

$$q_1 = q_2 + x,$$

*then $x$ is rational.*