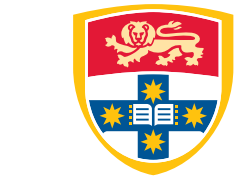
MATH2069

Discrete Mathematics and Graph Theory



THE UNIVERSITY OF SYDNEY

Anthony Henderson

Part I

Topics in Discrete Mathematics

Anthony Henderson

Acknowledgements

These lecture notes were written in 2008 (and revised in 2009) for the units MATH2069 Discrete Mathematics and Graph Theory and MATH2969 Discrete Mathematics and Graph Theory (Advanced), given at the University of Sydney. I am extremely grateful to David Easdown for allowing me to make use of his lecture notes on these topics, and for his careful correction of, and comments on, these notes. Any errors which remain are entirely my own fault.[1]

I also acknowledge a debt to the following textbooks, which students should consult for more examples and exercises.

- *Introduction to Discrete Mathematics* by K.-G. Choo and D. E. Taylor.
- *Combinatorics: An Introduction* by K. H. Wehrhahn.
- *The Art of Computer Programming* Vol. 1 by D. E. Knuth.
- *Enumerative Combinatorics* Vol. 1 by R. P. Stanley.

Anthony Henderson

[1] A few minor changes were made by Alexander Molev.

Contents

Chapter 0

Introduction

What is discrete mathematics and why is it useful? It is best to begin by considering an example of a problem in discrete mathematics.

The 'tower of Hanoi' is a popular mathematical toy (apparently invented in France, despite being named after a Vietnamese building). It consists of three pegs and a certain number of circular discs of differing sizes which sit around the pegs. You can move one disc at a time from one peg to another, subject to the rule that you can never place a larger disc on top of a smaller one. The discs start off all being on peg 1, in increasing order of size from top to bottom, and the aim is to move them all to peg 2. The mathematical question is: what is the smallest number of moves needed to achieve this, as a function of the number of discs?

We let n be the number of discs, and write h_n for the smallest number of moves required. If there are no discs, then there is nothing to do, so $h_0 = 0$. If there is one disc, then clearly one move suffices, so $h_1 = 1$. If there are two discs, then the rule that the larger disc cannot sit on top of the smaller one forces you to put the smaller one temporarily on peg 3, so $h_2 = 3$.

This illustrates a general feature: the largest disc has to move from peg 1 to peg 2 at some stage, and while that happens all the other discs have to be on peg 3. Clearly it takes at least h_{n-1} moves to get all the other discs to

peg 3, and then after moving the largest disc, it takes at least h_{n-1} moves to get them back to peg 2. This suggests a recursive procedure for solving the puzzle, which uses the smallest possible number of moves:

(1) Move all but the largest disc from peg 1 to peg 3 in the smallest possible number of moves.

(2) Move the largest disc from peg 1 to peg 2.

(3) Move all the other discs from peg 3 to peg 2 in the smallest possible number of moves.

This procedure is recursive because in steps (1) and (3) you need to go through the whole three-step procedure, but for $n-1$ discs rather than n (the change of peg numbers is immaterial); and when you then examine each of those, their own steps (1) and (3) call for the whole three-step procedure for $n-2$ discs, and so on. Returning to the number of moves required, we have derived a recurrence relation

$$h_n = 2h_{n-1} + 1,$$

which determines the answer for n in terms of the answer for $n-1$. This recurrence relation, together with the inital condition $h_0 = 0$, determines h_n for all n. For example, continuing from where we left off,

$$\begin{aligned} h_3 &= 2h_2 + 1 = 2 \times 3 + 1 = 7, \\ h_4 &= 2h_3 + 1 = 2 \times 7 + 1 = 15, \\ h_5 &= 2h_4 + 1 = 2 \times 15 + 1 = 31. \end{aligned}$$

You may notice that these values of h_n are all one less than a power of 2. In fact, the evidence we have so far supports the conjecture that $h_n = 2^n - 1$ always. To prove such a formula for a sequence defined recursively, the most natural method is mathematical induction: we know that $h_0 = 2^0 - 1$ is true, and if we assume that $h_{n-1} = 2^{n-1} - 1$ is true, then we can deduce

$$h_n = 2h_{n-1} + 1 = 2(2^{n-1} - 1) + 1 = 2^n - 2 + 1 = 2^n - 1.$$

By induction, $h_n = 2^n - 1$ for all nonnegative integers n. So the answer to the original question is that the smallest number of moves needed to complete the n-disc tower of Hanoi puzzle is $2^n - 1$.

"Discrete" is the opposite of "continuous", and discrete mathematics deals with things which are composed of separate parts rather than flowing together. What made the tower of Hanoi problem discrete was that we were trying to find a formula for something which was a function of a nonnegative integer rather than a real variable. It would make no sense at all to ask about the smallest number of moves needed to solve the puzzle for 4.673 discs; the number of discs has to be a nonnegative integer.

So in this course we will almost never be dealing with functions $f(x)$ of a real variable x, like those in calculus (e.g. $\sin x$, e^{x^2}). We will almost always be discussing functions whose domain is the set of nonnegative integers:

$$\mathbb{N} = \{0, 1, 2, 3, 4, \cdots\}.$$

(In mathematics it is usually best to start counting at 0 if possible, even if it means making some slightly contrived interpretations, such as declaring that a tower of Hanoi with 0 discs is already solved, or that an Introduction is Chapter 0.) It is perfectly acceptable to write such functions in the same way as functions of a real variable – for instance, we could have written $h(n)$ instead of h_n above – but we will tend to put the integer variable in the subscript, as say a_n. This notation goes along with the idea of writing the values of the function in a sequence:

$$a_0,\ a_1,\ a_2,\ a_3,\ \cdots .$$

There is in fact no logical difference between a function whose domain is $\mathbb{N}$ and a sequence like this; just a slight psychological shift of viewpoint. Calling the function a sequence is particularly natural in cases which are recursive like the tower of Hanoi problem: every term of the sequence (past a certain point, maybe) is determined by earlier terms, so you can imagine writing out the sequence in succession, with what you've already written determining what to write next.

In the tower of Hanoi example, the terms of the sequence were nonnegative integers. This feature is not essential for a problem to be considered discrete. For maximum flexibility, we will usually allow the terms of our sequences, i.e. the values of our functions with domain $\mathbb{N}$, to be complex numbers (though examples with non-real values will be rare).

The usefulness of discrete mathematics comes from the fact that, even though physics sometimes gives the impression that we live in a continuous universe, there are countless phenomena in nature and society which are best understood in terms of sequences (functions of a nonnegative integer variable) rather than functions of a real variable.

Obvious examples are the regular additions to or subtractions from bank accounts: if you want to calculate how much more interest you will earn if it is calculated monthly rather than yearly, or what the monthly repayments will be on a loan where interest is regularly calculated on the balance, you are automatically dealing with quantities which depend on an integer number of months (or years, or whatever) rather than anything continuous. And of course, the world of computers is discrete by construction: if you want to work out how long a program for constructing students' university timetables will take to run, all the variables are nonnegative integers – number of students, number of classes, number of bytes in the memory, etc.

It is also important to see the analogies between discrete and continuous problems, and these will be pointed out in the following chapters wherever possible. For example, solving a recurrence relation (that is, finding a closed formula for the terms of a sequence which is defined recursively) is strongly analogous to solving a differential equation. Some authors speak of "difference equations" instead of "recurrence relations", on the principle that the difference $a_{n+1} - a_n$ is the discrete analogue of the derivative of a function in calculus; but we will not explore this point of view.

There are some natural phenomena which could be modelled either discretely or continuously. For example, if you wanted to predict the temperature in Sydney tomorrow, you could think of temperature as a real-valued function of three real variables (longitude, latitude, and time), and try to understand what differential equations it satisfied. Or you could break up NSW into a finite number of square kilometre blocks, and time into a discrete sequence of days, and try to understand what recurrence relations hold: i.e. express the temperature on a given day in a given block in terms of the temperatures on the previous day in adjacent blocks. The discrete approach might well be easier, in the sense of being better suited to the computer technology you would need to use.

But it is certainly not the case that discrete mathematics is always easier than continuous mathematics; we will encounter many challenging problems which have no continuous analogues. For a first example, imagine a new, improved tower of Hanoi puzzle which has four pegs rather than three. The rules are the same as before; what is the smallest number of moves required to move the discs from peg 1 to peg 2, if there are n discs? Nobody knows the answer (at least, there is no known closed formula like $2^n - 1$).

Comments on the text

In these notes, the main results are all called "Theorem", irrespective of their difficulty or significance. Usually, the statement of the Theorem is followed (perhaps after an intervening Example or Remark) by a rigorous Proof; as is traditional, the end of each proof is marked by an open box.

To enable cross-references, the Theorems, Definitions, Examples, and Remarks are all numbered in sequence (the number always starts with the relevant chapter number). Various equations and formulas are also numbered (on their right-hand edge) in a separate sequence.

In definitions (either formally numbered or just in the main text), a word or phrase being defined is underlined. The symbol ":=" is used to mean "is defined to be" or "which is defined to be".

The text in the Examples and Remarks is *slanted*, to make them stand out on the page. Many students will understand the Theorems more easily by studying the Examples which illustrate them than by studying their proofs. Some of the Examples contain blank boxes for you to fill in. The Remarks are often side-comments, and are usually less important to understand.

The more difficult Theorems, Proofs, Examples, and Remarks are marked at the beginning with either * or **. Those marked * are at the level which MATH2069 students will have to understand in order to be sure of getting a Credit, or to have a chance of a Distinction or High Distinction. Those marked ** are really intended only for the students enrolled in the Advanced unit MATH2969, and can safely be ignored by those enrolled in MATH2069.

Chapter 1

Counting Problems

As we saw in the Introduction, the feature that characterizes discrete mathematics is the role played by the set of nonnegative integers:

$$\mathbb{N} = \{0, 1, 2, 3, 4, \cdots\}.$$

To state the obvious, the reason that $\mathbb{N}$ arises is that for every finite set X, the number of elements of X is a nonnegative integer. For instance, in the tower of Hanoi problem we had a finite set of discs, and n was the number of them; this is why the function we were computing had domain $\mathbb{N}$.

So the foundation of the whole subject is the idea of counting the number of elements in finite sets. This can be much harder than it sounds, and in its higher manifestations goes by the name of enumerative combinatorics. In this chapter we will review and extend some of the basic principles of counting.

1.1 Fundamental principles

If X is any finite set, we will write $|X|$ for its size (also known as its cardinality), i.e. the number of elements of X. For instance, if $X = \{0, 2, 4\}$, then $|X| = 3$. The most basic principle of all is:

Theorem 1.1 (Bijection Principle)**.** If two sets X and Y are in bijection, then $|X| = |Y|$.

A bijection, also known as a one-to-one correspondence, between X and Y is a way of associating to each element of X a single element of Y, such that every element of Y is associated to exactly one element of X. ("Exactly one" means not zero, and not more than one.)

Example 1.2. *If we assume that every person in the room has a single head, and every head in the room is part of exactly one person, then the set of people in the room is in bijection with the set of heads in the room. So the number of people in the room equals the number of heads in the room.*

Example 1.3. *How many numbers between* 1 *and* 999 *are divisible by* 3 *(in the sense that* 3 *divides the number exactly, with zero remainder)? This question is asking for the size of the set*

$$\begin{aligned} X &= \{n \in \mathbb{N} \,|\, 1 \leq n \leq 999,\ n \text{ is divisible by } 3\} \\ &= \{3, 6, 9, 12, \cdots, 999\}. \end{aligned}$$

The answer is easy. The elements of X are exactly 3×1, 3×2, and so on up to 3×333, so $|X| = 333$. But even this simple argument is really an application of the Bijection Principle: we are implicitly observing that X is in bijection with the set $Y = \{1, 2, \cdots, 333\}$, so $|X| = |Y| = 333$. This illustrates how the Bijection Principle is used. When it's not obvious how to count the elements of some set X, we try to translate the problem by finding another set Y, whose size we do know, which is in bijection with X.

Definition 1.4. The formal definition of a bijection between X and Y is a pair of functions, one from X to Y and one from Y to X, which are inverse to each other. Recall that a function $f : X \to Y$ is said to be injective ('one-to-one') if no function value is taken more than once, i.e. we never have $f(x_1) = f(x_2)$ for different elements $x_1, x_2 \in X$; equivalently,

$$\text{for every } y \in Y \text{ there is at most one } x \in X \text{ such that } f(x) = y.$$

(The opposite of "injective" is "not injective": to say that $f : X \to Y$ is not injective is to say that there are some elements $x_1, x_2 \in X$, with $x_1 \neq x_2$, such that $f(x_1) = f(x_2)$.) A function $f : X \to Y$ is said to be surjective ('onto') if every element of Y is one of the values taken by the function, i.e.

$$\text{for every } y \in Y \text{ there is at least one } x \in X \text{ such that } f(x) = y.$$

(The opposite of "surjective" is "not surjective", which means something completely different from "injective".) A function $f : X \to Y$ is bijective if it is both injective and surjective, i.e.

$$\text{for every } y \in Y \text{ there is exactly one } x \in X \text{ such that } f(x) = y.$$

It is only in this last situation that you can define an inverse function $f^{-1} : Y \to X$ (which you do by sending y to the unique x such that $f(x) = y$), and thus put X and Y in bijection.

The next most basic principle, again not requiring proof (since it is bound up with such elementary matters as the definition of addition), is:

Theorem 1.5 (Sum Principle)**.** If a finite set X is the disjoint union of two subsets A and B, i.e.

$$X = A \cup B \quad \text{and} \quad A \cap B = \emptyset,$$

which means that every element of X is either in A or in B but not both, then $|X| = |A| + |B|$. More generally, if X is the disjoint union of subsets $A_1, A_2, \cdots, A_n$, i.e.

$$X = A_1 \cup A_2 \cup \cdots \cup A_n \quad \text{and} \quad A_i \cap A_j \text{ for all } i \neq j,$$

then $|X| = |A_1| + |A_2| + \cdots + |A_n|$.

Example 1.6. *If every person in the room is either left-handed or right-handed, and no-one claims to be both, then the total number of people equals the number of left-handed people plus the number of right-handed people. Of course this relies on the disjointness of the two subsets: if anyone was ambidextrous and was counted in both the left-handed total and the right-handed total, it would throw out the calculation (see the section on the Inclusion/Exclusion Principle later in this chapter).*

Example 1.7. *How many numbers from* 1 *up to* 999 *are palindromic, in the sense that they read the same backwards as forwards? As soon as you start to think about this question, you have to break it up into the cases of one-digit numbers, two-digit numbers, and three-digit numbers. So, whether you choose to make it explicit or not, you are using the Sum Principle: in order to count the set*

$$X = \{n \in \mathbb{N} \,|\, 1 \leq n \leq 999,\ n \textit{ is palindromic}\},$$

you are using the fact that it is the disjoint union of subsets A_1, A_2, A_3, where

$$A_1 = \{n \in \mathbb{N} \mid 1 \leq n \leq 9,\ n \text{ is palindromic}\},$$
$$A_2 = \{n \in \mathbb{N} \mid 10 \leq n \leq 99,\ n \text{ is palindromic}\},$$
$$A_3 = \{n \in \mathbb{N} \mid 100 \leq n \leq 999,\ n \text{ is palindromic}\},$$

and then calculating $|X|$ as the sum $|A_1| + |A_2| + |A_3|$. The calculation is left for you to complete:

$$|A_1| = \boxed{} \text{ because } \boxed{}$$
$$|A_2| = \boxed{} \text{ because } \boxed{}$$
$$|A_3| = \boxed{} \text{ because } \boxed{}$$
$$|X| = \boxed{}$$

A variant form of the Sum Principle involves the complement of a subset: if A is a subset of X, its complement $X \setminus A$ is defined to be $\{x \in X \mid x \notin A\}$, the subset of X consisting of all elements which are not in A.

Theorem 1.8 (Difference Principle)**.** For any subset A of a finite set X,

$$|X \setminus A| = |X| - |A|.$$

Proof. This is just the Sum Principle rearranged, because X is the disjoint union of A and $X \setminus A$. □

Example 1.9. *How many three-digit numbers are divisible by 3? One way to answer this is to see that the set of three-digit numbers divisible by 3 can be written as $X \setminus A$, where*

$$X = \{n \in \mathbb{N} \mid 1 \leq n \leq 999,\ n \text{ is divisible by } 3\},$$
$$A = \{n \in \mathbb{N} \mid 1 \leq n \leq 99,\ n \text{ is divisible by } 3\}.$$

As seen in Example 1.3, $|X| = 333$, and similarly $|A| = 33$. By the Difference Principle, the answer to the question is $|X \setminus A| = |X| - |A| = 300$.

One situation in which a disjoint union naturally arises is if you have a function $f : X \to Y$. For any $y \in Y$, the preimage $f^{-1}(y)$ is defined by

$$f^{-1}(y) := \{x \in X \mid f(x) = y\}, \quad \text{a subset of } X.$$

Note that the condition for f to be injective is that $|f^{-1}(y)| \leq 1$ for all $y \in Y$, whereas the condition for f to be surjective is that $|f^{-1}(y)| \geq 1$ for all $y \in Y$.

Remark 1.10. *If f is bijective, and therefore has an actual inverse function $f^{-1} : Y \to X$, we have a slight notational clash. If you interpret $f^{-1}(y)$ as the image of $y \in Y$ under f^{-1}, then it is the unique $x \in X$ such that $f(x) = y$. If you interpret $f^{-1}(y)$ as the preimage, then it is the set whose single element is this x. In practice, the context always determines which of the two is meant.*

It is clear that X is the disjoint union of the subsets $f^{-1}(y)$ as y runs over Y, because for every $x \in X$ there is a unique $y \in Y$ such that $f(x) = y$ (that is what it means to have a well-defined function). So by the Sum Principle,

$$|X| = \sum_{y \in Y} |f^{-1}(y)|. \tag{1.1}$$

Example 1.11. *Suppose X is the set of students at this university, Y is the set of days in the year (including February 29), and $f : X \to Y$ is the function which associates to every student their birthday. Then (1.1) is just the obvious statement that the total number of students is the sum of the number whose birthday is January 1, the number whose birthday is January 2, and so on. It is actually possible to deduce something from this: there must be some day of the year which is the birthday of at least 126 students. The reason is that the size of X is 46054 (at least it was last year), and the size of Y is 366. If $|f^{-1}(y)| \leq 125$ for all $y \in Y$, then the right-hand side of (1.1) would be at most $366 \times 125 = 45750$, a contradiction. The magic number 126 was obtained by dividing 46054 by 366 and then rounding up.*

Definition 1.12. For any real number x, define

$$\begin{aligned} \lfloor x \rfloor &:= x \text{ rounded down to the nearest integer} \\ &= \text{largest integer } m \text{ such that } m \leq x, \\ \lceil x \rceil &:= x \text{ rounded up to the nearest integer} \\ &= \text{smallest integer } m \text{ such that } m \geq x. \end{aligned}$$

Because we are dealing mainly with functions which only make sense for integer variables, we will quite often have to carry out such roundings. We will use without comment the obvious inequalities

$$x - 1 < \lfloor x \rfloor \leq x, \qquad x \leq \lceil x \rceil < x + 1. \tag{1.2}$$

Theorem 1.13 (Pigeonhole Principle)**.** If we have a function $f : X \to Y$ between two finite (nonempty) sets, then there must be some $y \in Y$ such that

$$|f^{-1}(y)| \geq \left\lceil \frac{|X|}{|Y|} \right\rceil.$$

In particular, if $|X| > |Y|$ then there must be some $x_1 \neq x_2$ in X such that $f(x_1) = f(x_2)$, i.e. f is not injective.

Example 1.14. *The reason for the name is the imagined situation of pigeons flying into pigeonholes: if there are* 60 *pigeons and only* 25 *holes, there must be some pigeonhole that ends up with at least three pigeons in it. The "In particular" part of the statement just says that if the number of pigeons is more than the number of holes, there must be two pigeons which fly into the same hole. (Some authors call this the Pigeonhole Principle and regard Theorem 1.13 as a generalization.)*

Proof. As in Example 1.11, the proof is by contradiction. We assume the negation of what we are trying to prove, i.e. that

$$|f^{-1}(y)| < \left\lceil \frac{|X|}{|Y|} \right\rceil, \text{ for all } y \in Y.$$

Since $|f^{-1}(y)|$ is an integer, this inequality implies that $|f^{-1}(y)| < \frac{|X|}{|Y|}$. So

$$\sum_{y \in Y} |f^{-1}(y)| < \sum_{y \in Y} \frac{|X|}{|Y|} = |X|,$$

which contradicts (1.1). Therefore our assumption was wrong, which means that there is some $y \in Y$ for which $|f^{-1}(y)| \geq \left\lceil \frac{|X|}{|Y|} \right\rceil$. To deduce the "In particular" sentence, note that $|X| > |Y|$ means that $\frac{|X|}{|Y|} > 1$, so $\left\lceil \frac{|X|}{|Y|} \right\rceil \geq 2$. So we have shown that there is some $y \in Y$ for which there at least two elements of X which are taken to y by f; this gives the stated result. □

Remark 1.15. *Note that the Pigeonhole Principle asserts that something is true for some $y \in Y$, but not for all $y \in Y$, and not for a particular y specified in advance. For instance, in the example of students' birthdays, we cannot pick a particular date like June 16 and be sure that at least 126 students have that birthday. Indeed, it is possible that no students have that birthday (though admittedly that's unlikely).*

The Pigeonhole Principle is surprisingly useful even in its basic form of proving non-injectivity.

Example 1.16. *If there are 30 students in a tutorial, there must be two whose surnames begin with the same letter, because there are only 26 possible letters. This is an example of the Pigeonhole Principle: we are saying that the function from the set of students to the set of letters given by taking the first letter of the surname cannot be injective. Of course, we can't pick a particular letter like A and be sure that there are two students whose surname begins with A: there might be none.*

Example 1.17*. *If m is any positive integer, the decimal expansion of $1/m$ eventually gets into a repeating cycle: for instance,*

$$1/5 = 0.20000000\cdots,\ 1/65 = 0.0153846153846153846\cdots.$$

We can prove this using the Pigeonhole Principle, as follows. Consider the set $X = \{10^0, 10^1, \cdots, 10^m\}$. For each element of X, we can divide it by m and find the remainder, which must be in the set $Y = \{0, 1, \cdots, m-1\}$. This defines a function from X to Y. Since $|X| = m+1$ and $|Y| = m$, this function can't be injective, so there are some nonnegative integers $k_1 < k_2$ such that 10^{k_1} and 10^{k_2} have the same remainder after division by m. This means that $\frac{10^{k_1}}{m}$ and $\frac{10^{k_2}}{m}$ are the same after the decimal point (i.e. their difference is an integer). But the decimal expansion of $\frac{10^{k_1}}{m}$ is obtained from that of $\frac{1}{m}$ by shifting every digit k_1 places to the left, and similarly for k_2. So for every $k > k_1$, the kth digit after the decimal point in $\frac{1}{m}$ equals the $(k-k_1)$th digit after the decimal point in $\frac{10^{k_1}}{m}$, which equals the $(k-k_1)$th digit after the decimal point in $\frac{10^{k_2}}{m}$, which equals the $(k+k_2-k_1)$th digit after the decimal point in $\frac{1}{m}$. So once you are sufficiently far to the right of the decimal point, the digits in $\frac{1}{m}$ repeat themselves every $k_2 - k_1$ digits (maybe more often than that, but at least that often).

There is a related way of proving non-surjectivity:

Theorem 1.18 (Reverse Pigeonhole Principle)**.** If $|X| < |Y|$, then any function $f : X \to Y$ is not surjective. That is, there must always be some $y \in Y$ such that $f^{-1}(y)$ is empty.

Proof. The range of f can clearly have at most $|X|$ elements, so it cannot be the whole of Y. □

There is an old joke that a combinatorialist is someone who, to find out how many sheep there are in a flock, counts their legs and divides by 4. That may not be a practical method in the case of sheep, but it is a surprisingly useful trick in other sorts of counting problems.

Theorem 1.19 (Overcounting Principle)**.** Let X and Y be finite sets, and m a positive integer. Suppose there is a function $f : X \to Y$ with the property that for all $y \in Y$, $|f^{-1}(y)| = m$ (i.e. there are always m elements $x \in X$ such that $f(x) = y$). Then $|Y| = \frac{|X|}{m}$.

Proof. By (1.1), we have $|X| = \sum_{y \in Y} |f^{-1}(y)| = m|Y|$. □

Example 1.20. *In the spurious sheep example, X is the set of legs, Y is the set of sheep, $m = 4$, and f is the function which associates to each leg the sheep it belongs to. The assumption (which we certainly do need, to apply the method!) amounts to saying that every sheep has four legs.*

Obviously, the cases where it is useful to apply this principle are those where the set X is for some reason easier to count than the set Y (despite being bigger). It is vital that all the preimages have the same size, so that when we count X, we are overcounting Y by a known factor.

Example 1.21. *How many edges does a cube have? If you don't have pen and paper handy, you can just observe that a cube has 6 faces, and every face has 4 edges. Is the answer $6 \times 4 = 24$? No: this counts every edge twice, since every edge is on the border of two faces. So the correct answer is 12, and this argument is a case of the Overcounting Principle. To interpret it in the above terms, Y is the set of edges, X is the set of pairs $(face, edge)$ where the edge belongs to the face, and f is the function which takes such a*

pair and forgets the face. Here is a similar problem: the outside of a soccer ball is made from 32 *pieces of cloth,* 12 *of which are pentagonal (having five edges) and* 20 *of which are hexagonal (having six edges). These are stitched together along their edges, and at every vertex where an edge ends, three faces meet. Using the Overcounting Principle, you can deduce that:*

the number of edges is $\boxed{}$,

the number of vertices is $\boxed{}$.

Our final fundamental principle relates to the most important way of constructing new sets from old.

Definition 1.22. If X and Y are sets, the Cartesian product $X \times Y$ is defined to be the set of all pairs (x, y) where the first entry is an element of X and the second entry is an element of Y. In symbols:

$$X \times Y := \{(x, y) \,|\, x \in X, y \in Y\}.$$

More generally, if we have n sets $X_1, X_2, \cdots, X_n$, then their product is the set of n-tuples where the ith entry is an element of X_i:

$$X_1 \times X_2 \times \cdots \times X_n := \{(x_1, x_2, \cdots, x_n) \,|\, x_1 \in X_1, x_2 \in X_2, \cdots, x_n \in X_n\}.$$

If $X_1 = X_2 = \cdots = X_n = X$, then $X_1 \times X_2 \times \cdots \times X_n$ is simply written X^n.

Example 1.23. *Let* $X = \{\mathrm{A}, \mathrm{B}, \mathrm{C}, \cdots, \mathrm{Z}\}$ *be the set of letters in the alphabet, and let* $Y = \{0, 1, 2, \cdots, 9\}$. *Then* $X \times X \times X \times Y \times Y \times Y = X^3 \times Y^3$ *is the set of* 6*-tuples where the first three entries are letters and the last three entries are single-digit numbers, such as* $(\mathrm{L}, \mathrm{B}, \mathrm{M}, 5, 2, 3)$. *Ignoring the brackets and commas, this is the set of all strings of three letters followed by three digits, such as* LBM523*: exactly the set of possible car numberplates, under the old style. Asking how many possible numberplates there were when this was the uniform style is the same as asking for the size of* $X^3 \times Y^3$. *The answer is that there are* 26 *choices for the first letter,* 26 *choices for the second letter,* 26 *choices for the third letter, and* 10 *choices for each of the three digits, so a total of* $26^3 \times 10^3 = 17576000$ *possibilities.*

Remark 1.24. *Before we extract the last part of this argument as a general principle, a pedantic comment. According to Definition 1.22,* $X^3 \times Y^3$ *ought to be a set of pairs where each entry of the pair is a triple: its elements look like* $((\mathrm{L}, \mathrm{B}, \mathrm{M}), (5, 2, 3))$*. So this is not really the same set as* $X \times X \times X \times Y \times Y \times Y$*; however, the bijection between them is so obvious (just a matter of deleting or inserting some brackets) that it does no harm to identify them. A similar comment applies to the identification of* 6*-tuples with 'strings'.*

The following is the reason why Cartesian products are called products.

Theorem 1.25 (Product Principle)**.** If $X_1, X_2, \cdots, X_n$ are finite sets, then

$$|X_1 \times X_2 \times \cdots \times X_n| = |X_1| \times |X_2| \times \cdots \times |X_n|.$$

In particular, $|X^n| = |X|^n$.

This principle encapsulates the idea of 'independent choices': in the definition of the Cartesian product $X_1 \times X_2 \times \cdots \times X_n$, there are no restrictions which relate the different entries of the n-tuple, so no matter which of the $|X_1|$ choices you make for the first entry, there are still $|X_2|$ choices for the second entry and so on. This is why the total number of n-tuples is the product of all the individual sizes.

Remark 1.26. *Just as it is useful to define* $x^1 = x$*, one would always identify* X^1 *with* X *(without worrying about whether a '*1*-tuple' ought to have brackets around it – see Remark 1.24). Similarly, on the same principle which motivates setting* $x^0 = 1$*, it is useful to stipulate that* X^0 *always has a single element: some sort of 'empty tuple' which you could write as* $()$*. Then the rule* $|X^n| = |X|^n$ *holds for all* $n \in \mathbb{N}$*.*

Example 1.27. *How many five-digit numbers are there? Probably the simplest approach is that the five-digit numbers are those* $n \in \mathbb{N}$ *such that* $n > 9999$ *and* $n \leq 99999$*, so the answer is* $99999 - 9999 = 90000$*; this is an application of the Difference Principle, because we are implicitly viewing the set of five-digit numbers as the complement of* $\{1, 2, \cdots, 9999\}$ *in* $\{1, 2, \cdots, 99999\}$*. An alternative approach is that there are* 9 *choices for the first digit (since it can't be zero) and* 10 *choices for every other digit, so the answer is* $9 \times 10 \times 10 \times 10 \times 10 = 90000$*; this is an application of the Product Principle, because we are implicitly identifying the set of five-digit numbers*

with the set of 5-tuples where the first entry is an element of $\{1, \cdots, 9\}$ *and all the other entries are elements of* $\{0, \cdots, 9\}$*. The second approach is necessary for some variants of the question. For instance, the number of five-digit numbers which are palindromes (i.e. the digits have the form abcba) is* $9 \times 10 \times 10 = 900$*, because the choice of the first three digits determines the last two. (The Bijection Principle is also involved here, because we are effectively using the fact that the five-digit palindromes are in bijection with the strings formed by deleting their last two digits.)*

For some situations, the above statement of the Product Principle is too restrictive. Instead of counting the ways of making independent selections from pre-determined sets $X_1, \cdots, X_n$, you might have a situation where the set X_2 depends on what element you selected from X_1, and then the set X_3 depends on what elements you selected from X_1 and X_2, and so on. All you really need to be able to apply the Product Principle is that the sizes $|X_i|$ don't depend on the choices you've already made.

Example 1.28. *How many two-digit numbers have two different digits? Since there are* 90 *two-digit numbers in total, and the digits are the same in* 9 *of them (*$11, 22, \cdots, 99$*), the answer is* 81 *(by the Difference Principle). Another approach is that there are* 9 *choices for the first digit (anything from* 1 *to* 9*), and then having chosen the first digit, there are* 9 *choices for the second digit (anything except the digit you've already used), so the answer is* $9 \times 9 = 81$*. This is an application of the looser form of the Product Principle, since the set of possibilities for the second digit varies depending on what first digit you chose; what matters is that it always has* 9 *elements. What about trying to choose the digits in the other order? Since the second digit is allowed to be zero, there are* 10 *choices for it, and then there should surely be* 8 *choices for the first digit (anything except zero and the chosen second digit). But the answer isn't* 80*; what's gone wrong is that if you do choose zero as the second digit, the number of choices for the first digit is not* 8 *but* 9*. Although this example is pretty trivial, it does illustrate the need for care in applying the Product Principle.*

1.2 Ordered selection

A classic situation where the Product Principle arises is in counting the number of ways to carry out some 'ordered selection'.

Example 1.29. *Suppose there are three students A, B, and C, each of whom has to be allocated to one of 4 tutorial times, with no restrictions on how many are put at each time (in particular, we are allowed to 'repeat a selection', i.e. choose the same tutorial time for more than one student). If all the students can attend all 4 times, then the number of ways of doing the allocation is* $4^3 = 64$. *If student A can only attend 2 of the times, student B can only attend 3 of the times, and student C can only attend 1 of the times, then the number of ways is* $2 \times 3 \times 1 = 6$. *Here we are implicitly using the Product Principle to count the elements of the set* $X_1 \times X_2 \times X_3$, *where* X_1 *is the set of times that student A can attend,* X_2 *is the set of times that student B can attend, and* X_3 *is the set of times that student C can attend.*

Since it is not always easy to think of such problems in terms of a literal selection, it is helpful to have a more purely mathematical framework in which to discuss them. The use of the word "allocate" in the previous example is a hint that it is essentially about counting some functions from one set to another (namely, from the set of students to the set of tutorial times – the allocation is a function because every student is allocated a single tutorial time). We have the following general result:

Theorem 1.30. If X and Y are sets with $|X| = k$ and $|Y| = n$, then the number of functions $f : X \to Y$ is n^k.

Proof. We can name the elements of X as $x_1, x_2, \cdots, x_k$. (Another way to say this is that we can choose a bijection between X and $\{1, 2, \cdots, k\}$.) Then to specify a function $f : X \to Y$ is the same as choosing the elements $f(x_1), f(x_2), \cdots, f(x_k)$, all of which must lie in Y. There are n choices for $f(x_1)$, n choices for $f(x_2)$, and so on up to $f(x_k)$, so the answer is $n \times n \times \cdots \times n$ (k factors), which is n^k. Implicitly, we are using the Bijection Principle and the Product Principle here: the set of functions $f : X \to Y$ is in bijection with the set Y^k via the association of any function $f : X \to Y$ with the k-tuple $(f(x_1), f(x_2), \cdots, f(x_k))$, and the size $|Y^k|$ equals $|Y|^k$ by the Product Principle. □

Remark 1.31*. *To make Theorem 1.30 true when X is empty, we need the number of functions $f : \emptyset \to Y$ to be $|Y|^0 = 1$. Indeed, the convention is that there is a unique function from the empty set to every other set.*

Example 1.32. *If there are 5 different presents under a Christmas tree, and 3 greedy children grabbing at them until they are all claimed, and no present can be shared between more than one child, how many possible results are there? The answer is given by Theorem 1.30; the main mental difficulty is deciding which way round the functions go, from the set of children to the set of presents or vice versa. Here we have a giveaway phrase: "no present can be shared between more than one child". That is, every present is associated with a single child, so the "possible results" are really the functions from the set of presents to the set of children, and the answer is $3^5 = 243$. To phrase a problem which amounted to counting the functions from the set of children to the set of presents, you would have to specify that every child gets a single present and remove the ban on sharing presents and the requirement that every present is claimed. Then the answer would be $5^3 = 125$.*

An important consequence of Theorem 1.30 is:

Theorem 1.33. The number of subsets of a finite set X (including the empty set and the set itself) is $2^{|X|}$.

Example 1.34. *The set $\{1, 2, 3\}$ has the following $8 = 2^3$ subsets: $\emptyset$, $\{1\}$, $\{2\}$, $\{3\}$, $\{1, 2\}$, $\{1, 3\}$, $\{2, 3\}$, $\{1, 2, 3\}$.*

Proof. The idea of the proof is that to specify a subset is the same as specifying, for every element of the set, whether it is in or out. We can formalize this idea as saying that there is a bijection between the set of subsets of X and the set of functions $f : X \to \{0, 1\}$. To define this bijection, start with any subset A of X, and associate to this the function

$$f_A : X \to \{0, 1\} \text{ defined by } f_A(x) = \begin{cases} 1, & \text{if } x \in A, \\ 0, & \text{if } x \notin A. \end{cases}$$

This is a bijection because every function $f : X \to \{0, 1\}$ arises in this way for a unique subset A, namely the subset $\{x \in X \mid f(x) = 1\}$. Applying the Bijection Principle, the number of subsets of X is the same as the number of functions $X \to \{0, 1\}$, which is $2^{|X|}$ by Theorem 1.30. □

Many problems amount to counting the number of functions $f : X \to Y$ which satisfy the extra condition of being injective, i.e. we never have $f(x_1) = f(x_2)$ when $x_1 \neq x_2$: different inputs give different outputs. These are the problems which involve 'ordered selection with repetition not allowed'. We already saw one such problem in Example 1.28; here is another.

Example 1.35. *Suppose that car numberplates consisted of six letters of the alphabet. If there was no restriction on repeating letters, the number of possible numberplates would be*

If repetition of letters was not allowed, the number of possible numberplates would be

Theorem 1.36. If X and Y are sets with $|X| = k$ and $|Y| = n$, then the number of injective functions $f : X \to Y$ is $n(n-1)(n-2)\cdots(n-k+1)$.

Proof. First suppose that $n < k$, i.e. Y is a smaller set than X; then there cannot be any injective function from X to Y (by the Pigeonhole Principle). The formula does indeed give the answer 0 in this case, because one of the factors in the product is $n - n$.

Now suppose that $n \geq k$. Name the elements of X as $x_1, x_2, \cdots, x_k$. To specify an injective function $f : X \to Y$, we can first choose $f(x_1)$ in n ways. Then $f(x_1)$ is ruled out as a possible value for $f(x_2)$, so there are $n - 1$ options for $f(x_2)$; and so on, until when we come to choose $f(x_k)$, our $k - 1$ previous choices are ruled out, so there are $n - k + 1$ options left. By the looser form of the Product Principle, the number of injective functions is the product of $n, n-1, \cdots, n-k+1$, as claimed. □

Notice that there are k factors in the product in Theorem 1.36, which is called a falling factorial ("falling" because the factors decrease by 1 at each step). The usual notation for these is:

$$n_{(k)} := n(n-1)(n-2)\cdots(n-k+1). \tag{1.3}$$

This makes sense when n is any complex number, which will be useful later.

Remark 1.37. *A note on conventions regarding products. The definition* (1.3) *of* $n_{(k)}$ *shows the first three factors of a product, a "* $\cdots$ *", and then the final factor. This may seem to imply that there should always be at least five factors. But in interpreting this for small values of* k*, one has to follow the pattern that there are always* k *factors, rather than the literal sense. For instance,* $n_{(1)}$ *is defined to be* n*; what should be the last factor, namely* $n-1+1$*, is the same as the first, so we stop there already. Similarly,* $n_{(2)}$ *is defined to be* $n(n-1)$*. As for the case* $k=0$*, we make sense of that by the convention that a product of* 0 *things means* 1*, so* $n_{(0)}=1$ *for any* n*.*

An important special case of Theorem 1.36 is when $Y = X$. An injective function from X to itself must also be surjective, because the values of the function are all different and hence exhaust the elements of X. Thus it must be a bijection from X to itself, which we call a permutation of X. So in this special case, Theorem 1.36 says that the number of permutations of X is

$$k! := k_{(k)} = k(k-1)(k-2)\cdots 3 \times 2 \times 1, \tag{1.4}$$

which is called k factorial. According to our conventions, $0! = 1$ (there is a single permutation of the empty set). The sequence of factorial continues:

$$1! = 1,\ 2! = 2,\ 3! = 6,\ 4! = 24,\ 5! = 120,\ 6! = 720,\ 7! = 5040,\ \cdots$$

Example 1.38. *The permutations of* $\{1,2,3\}$ *can be written as strings where the first digit is* $f(1)$*, the next* $f(2)$*, and the third* $f(3)$*:*

$$123,\ 132,\ 213,\ 231,\ 312,\ 321.$$

There are indeed $3! = 6$ *of these.*

In general, $k!$ is the number of ways of ordering k objects, where "ordering" may manifest itself differently in various examples.

Example 1.39. *Here are some of the things that are counted by* $4! = 24$*:*

- *the ways that four people can line up in a queue;*
- *the four-digit numbers whose digits are* 3*,* 5*,* 7*, and* 9 *(in some order);*
- *the ways of putting four different letters in four different envelopes.*

We can express $n_{(k)}$ in terms of plain factorials, whenever n is another non-negative integer and $n \geq k$:

$$n_{(k)} = n(n-1)\cdots(n-k+1) = \frac{n(n-1)\cdots 1}{(n-k)(n-k-1)\cdots 1} = \frac{n!}{(n-k)!}. \quad (1.5)$$

But beware that $\frac{n!}{(n-k)!}$ makes no sense if n is not an integer, or if $n < k$.

Counting the surjective functions from X to Y is a more subtle problem which we will return to in the section after next. To conclude this section, here is another example along similar lines to Theorem 1.36.

Example 1.40. *In how many ways can a group of $2n$ people be split into n pairs for games of table tennis? Imagine them lining up in single file. The opponent of the first person can be any one of the other people, so there are $2n-1$ choices. Having made that choice, the first pair can both leave the line, leaving $2n-2$ people behind; there are $2n-3$ choices for the opponent of the first person, then they can both leave the line, leaving $2n-4$ people behind; and so on. So the answer is*

$$(2n-1)!! := (2n-1)(2n-3)\cdots 5\times 3\times 1, \quad (1.6)$$

the product of all the odd numbers up to $2n-1$. (The double ! is meant to indicate that the factors decrease by 2 at each step.) This can be expressed in terms of ordinary factorials:

$$\begin{aligned}(2n-1)!! &= (2n-1)(2n-3)\cdots 5\times 3\times 1\\ &= \frac{(2n)(2n-1)(2n-2)(2n-3)(2n-4)\cdots 5\times 4\times 3\times 2\times 1}{(2n)(2n-2)(2n-4)\cdots 4\times 2}\\ &= \frac{(2n)(2n-1)(2n-2)(2n-3)(2n-4)\cdots 5\times 4\times 3\times 2\times 1}{2^n\, n(n-1)(n-2)\cdots 2\times 1}\\ &= \frac{(2n)!}{2^n\, n!}.\end{aligned}$$

For instance, if we have 6 people then the answer is $15 = 5\times 3\times 1 = \frac{6!}{2^3\,3!}$. Incidentally, it will be convenient later to allow $n=0$ in the definition of $(2n-1)!!$. Applying to (1.6) the principle that a product of 0 things means 1, we get the somewhat unpleasant definition $(-1)!! := 1$.

Remark 1.41*. *There is an alternative approach to Example 1.40. We know that $(2n)!$ counts the number of ways of ordering the $2n$ people in a*

line. Any such ordering gives rise to a way of pairing them off: simply pair the first with the second, the third with the fourth, and so on. So we have a function from the set of orderings to the set of pairings. If we start with a given pairing, the number of orderings which give rise to it is $2^n\,n!$, because we can order the n pairs in $n!$ ways, and then decide which person comes first in each pair in 2^n ways. So every one of the preimages for our function has size $2^n\,n!$, and by the Overcounting Principle, there are $\frac{(2n)!}{2^n\,n!}$ pairings.

1.3 Binomial coefficients

Another application of the Overcounting Principle is in counting subsets of a prescribed size. Recall that the total number of subsets of an n-element set is 2^n.

Theorem 1.42. For $n, k \in \mathbb{N}$, the number of k-element subsets of an n-element set X is

$$\frac{n_{(k)}}{k!} = \frac{n(n-1)(n-2)\cdots(n-k+1)}{k!}.$$

Example 1.43. *The number of 2-element subsets of $\{1,2,3,4\}$ is six:*

$$\{1,2\},\ \{1,3\},\ \{1,4\},\ \{2,3\},\ \{2,4\},\ \{3,4\}.$$

This agrees with the stated formula: $\frac{4\times 3}{2\times 1} = 6$.

Proof. By Theorem 1.36, $n_{(k)}$ counts the injective functions $\{1,2,\cdots,k\} \to X$. That is, it is the number of ordered selections of k different elements $x_1, x_2, \cdots, x_k$ of X. What we want is the number of unordered selections $\{x_1, x_2, \cdots, x_k\}$ of k different elements of X. The relationship between the set of ordered selections and the set of unordered selections is exactly that of the Overcounting Principle: every ordered selection gives rise to an unordered selection by forgetting the order, and for every unordered selection, there are exactly $k!$ orderings of it. So the number of unordered selections is the number of ordered selections divided by $k!$, which is the claim. □

Remark 1.44. *Notice that if $n < k$, there are no k-element subsets of X; the formula does indeed give 0 in that case. Another notable aspect of*

Theorem 1.42 is that it proves that $k!$ always divides $n_{(k)}$ (i.e. their quotient is an integer), which is not particularly easy to see otherwise. For instance, $n(n-1)(n-2)(n-3)$ is a multiple of 24 for all $n \in \mathbb{N}$.

We will use the notation

$$\binom{n}{k} := \frac{n_{(k)}}{k!} = \frac{n(n-1)(n-2)\cdots(n-k+1)}{k!}, \tag{1.7}$$

read "n choose k". This definition makes sense when n is any complex number, which will be useful in Chapter 3. Note that

$$\binom{n}{0} = \binom{n}{n} = 1, \qquad \binom{n}{1} = \binom{n}{n-1} = n. \tag{1.8}$$

If n is an integer and $n \geq k$, then the alternative formula (1.5) for $n_{(k)}$ gives

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}. \tag{1.9}$$

The formula (1.9) is symmetric in k and $n-k$, so

$$\binom{n}{k} = \binom{n}{n-k}, \text{ for all } n, k \in \mathbb{N},\ n \geq k. \tag{1.10}$$

Remark 1.45. *Equation* (1.10) *can also be proved using the Bijection Principle: there is a bijection between the k-element subsets of an n-element set X and the $(n-k)$-element subsets, which associates to every k-element subset A its complement $X \setminus A$. This is a bijection because every $(n-k)$-element subset B of X arises in this way for exactly one k-element subset, namely $X \setminus B$.*

Example 1.46. *There are 52 playing cards in a standard deck, 13 of each suit (spades, hearts, diamonds, and clubs). If you are dealt 5 of them at random, how many possible hands can you end up with? Since the order in which you are dealt the cards doesn't matter, this is just the number of 5-element subsets of the set of cards, which is*

$$\binom{52}{5} = \frac{52 \times 51 \times 50 \times 49 \times 48}{5 \times 4 \times 3 \times 2 \times 1} = 2598960.$$

In card games such as poker, you need to know the probability of getting various kinds of hands, which means you need to count how many of the 5-element subsets satisfy certain properties. For instance, how many hands:

contain no spades? ☐

contain the king of spades? ☐

contain exactly one spade? ☐

Example 1.47. *The number of ways of forming a string of seven a's and five b's is* $\binom{12}{5} = \binom{12}{7} = \frac{12!}{7!5!} = 792$. *This is because the string must be* 12 *letters long, and to specify it is the same as choosing which seven of the* 12 *positions are occupied by an a (or equivalently, which five of the* 12 *positions are occupied by a b).*

The numbers $\binom{n}{k}$ are known as binomial coefficients, from their role in:

Theorem 1.48 (Binomial Theorem)**.** If a, b are any numbers and $n \in \mathbb{N}$,

$$\begin{aligned}(a+b)^n &= \sum_{k=0}^{n} \binom{n}{k} a^{n-k}b^k \\ &= \binom{n}{0}a^n + \binom{n}{1}a^{n-1}b + \binom{n}{2}a^{n-2}b^2 + \cdots \\ &\qquad + \binom{n}{n-2}a^2b^{n-2} + \binom{n}{n-1}ab^{n-1} + \binom{n}{n}b^n \\ &= a^n + na^{n-1}b + \frac{n(n-1)}{2}a^{n-2}b^2 + \cdots + \frac{n(n-1)}{2}a^2b^{n-2} + nab^{n-1} + b^n.\end{aligned}$$

Proof. Imagine expanding out $(a+b)^n = (a+b)(a+b)\cdots(a+b)$. There will be 2^n terms in the expansion, each of which is (initially) a string of a's and b's, of total length n; this is because you get one term for every choice of either a or b from the first factor, a or b from the second factor, and so on. For every k from 0 up to n, exactly $\binom{n}{k}$ of these strings have a in $n-k$ places and b in k places; this is because there are k ways of specifying which factors to choose b from, and then the other factors must all be ones you choose a

from (see Example 1.47). So when you collect terms together using the fact that $ab = ba$, the coefficient of $a^{n-k}b^k$ is $\binom{n}{k}$. □

Remark 1.49*. *As the proof shows, the a and b in the Binomial Theorem don't have to be numbers. They could be elements of any ring (a sort of generalized number system where addition and multiplication are defined and satisfy the distributive law), provided that $ab = ba$. So, for instance, the result is still true if applied to two square matrices A and B which commute (but would be false if A and B did not commute).*

Some special cases of the Binomial Theorem are worth pointing out. When $a = b = 1$, it says that

$$2^n = \sum_{k=0}^{n} \binom{n}{k} = \binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{n-1} + \binom{n}{n}. \tag{1.11}$$

This also follows from the Sum Principle, because the left-hand side counts all subsets of a set of size n, and the terms of the right-hand side count the number of subsets of a given size. When $a = 1$ and $b = -1$, we get

$$\sum_{k=0}^{n} (-1)^k \binom{n}{k} = 0^n = \begin{cases} 1, & \text{if } n = 0, \\ 0, & \text{otherwise.} \end{cases} \tag{1.12}$$

If n is odd this can be seen directly, because $\binom{n}{k}$ and $\binom{n}{n-k}$ occur on the left-hand side with opposite signs and hence cancel each other out.

There is a famous diagrammatic representation of the binomial coefficients called Pascal's triangle:

$$\begin{array}{ccccccccc}
1 & & & & & & & & \\
1 & 1 & & & & & & & \\
1 & 2 & 1 & & & & & & \\
1 & 3 & 3 & 1 & & & & & \\
1 & 4 & 6 & 4 & 1 & & & & \\
1 & 5 & 10 & 10 & 5 & 1 & & & \\
1 & 6 & 15 & 20 & 15 & 6 & 1 & & \\
1 & 7 & 21 & 35 & 35 & 21 & 7 & 1 & \\
\vdots & & \vdots & & \vdots & & \vdots & & \ddots
\end{array}$$

Here the rows correspond to $n = 0$, $n = 1$, $n = 2$, and so on, and the columns correspond to $k = 0$, $k = 1$, $k = 2$, and so on. (Since $\binom{n}{n-k} = \binom{n}{k}$, each row of Pascal's triangle reads the same backwards as forwards. Because of this symmetry, it is more usual to align the rows so that they are all centred on a vertical line.) Notice that each entry in the above triangle is the sum of the one above and to the left, and the one directly above; this suggests the following result.

Theorem 1.50. The binomial coefficients satisfy the recurrence relation

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}, \text{ for all } n \geq 1.$$

Proof. One can prove this from the Binomial Theorem, by expressing $(a+b)^n$ as $(a+b)(a+b)^{n-1}$. But here is a more satisfying argument using the Sum Principle. The left-hand side counts the number of k-element subsets A of $\{1, 2, \cdots, n\}$. Now a subset A either contains the element n or it doesn't. (More formally, the set of k-element subsets is the disjoint union of the set of k-element subsets which contain n and the set of k-element subsets which don't contain n.) Choosing a k-element subset A which contains n just amounts to choosing a $(k-1)$-element subset of $\{1, 2, \cdots, n-1\}$, so there are $\binom{n-1}{k-1}$ subsets which contain n (notice the implicit appeal to the Bijection Principle in the phrase "just amounts to"). Similarly, there are $\binom{n-1}{k}$ subsets which don't contain n. The recurrence relation follows. □

Here is a slight generalization of binomial coefficients.

Theorem 1.51. Let X be a set with n elements, and let $n_1, n_2, \cdots, n_m \in \mathbb{N}$ be such that $n_1 + n_2 + \cdots + n_m = n$. Let $\binom{n}{n_1, n_2, \cdots, n_m}$ be the number of ways to choose subsets $A_1, A_2, \cdots, A_m$ of X such that:

(1) $|A_i| = n_i$ for $i = 1, 2, \cdots, m$, and

(2) X is the disjoint union of the A_i's.

Then

$$\binom{n}{n_1, n_2, \cdots, n_m} = \frac{n!}{n_1! n_2! \cdots n_m!}.$$

Proof. We can choose A_1 in $\binom{n}{n_1}$ ways. Then to maintain disjointness, A_2 must be a subset of $X \setminus A_1$, so it can be chosen in $\binom{n-n_1}{n_2}$ ways. Then A_3 must be a subset of $X \setminus (A_1 \cup A_2)$, so it can be chosen in $\binom{n-n_1-n_2}{n_3}$ ways, and so forth. So we have

$$\begin{aligned}\binom{n}{n_1, n_2, \cdots, n_m} &= \binom{n}{n_1}\binom{n-n_1}{n_2}\cdots\binom{n-n_1-n_2-\cdots-n_{m-1}}{n_m} \\ &= \frac{n!}{n_1!(n-n_1)!}\frac{(n-n_1)!}{n_2!(n-n_1-n_2)!}\cdots\frac{(n-n_1-n_2-\cdots-n_{m-1})!}{n_m!(n-n_1-n_2-\cdots-n_m)!} \\ &= \frac{n!}{n_1!n_2!\cdots n_m!},\end{aligned}$$

because everything else cancels out. A slightly more elegant proof uses the Overcounting Principle: instead of choosing unordered subsets we could choose the elements of A_1 in order, then the elements of A_2 in order, and so on. The number of ways of doing this would be $n!$, since it amounts to just ordering all the elements of X; and this overcounts by a factor of $n_1!n_2!\cdots n_m!$, because we have ordered each of the subsets. □

The numbers $\binom{n}{n_1, n_2, \cdots, n_m}$ are called multinomial coefficients because of their occurrence in the generalization of the Binomial Theorem:

Theorem 1.52 (Multinomial Theorem)**.** If $a_1, a_2, \cdots, a_m$ are any numbers and $n \in \mathbb{N}$, then

$$(a_1 + a_2 + \cdots + a_m)^n = \sum_{\substack{n_1, n_2, \cdots, n_m \in \mathbb{N} \\ n_1+n_2+\cdots+n_m=n}} \binom{n}{n_1, n_2, \cdots, n_m} a_1^{n_1} a_2^{n_2} \cdots a_m^{n_m}.$$

Proof. This is similar to the proof of the Binomial Theorem. When you expand $(a_1 + a_2 + \cdots + a_m)^n$, you get a sum of terms. A term corresponds to a set of choices of which of $a_1, a_2, \cdots, a_m$ to select from each of the n factors. To get the coefficient of $a_1^{n_1} a_2^{n_2} \cdots a_m^{n_m}$, you need to count how many ways there are to select a_1 from n_1 factors, a_2 from n_2 factors, and so on. This amounts to choosing disjoint subsets of the factors, one of size n_1, one of size n_2, etc., which is what the multinomial coefficient counts. □

Remark 1.53. *A nice consequence of Theorem 1.51 is that $n_1!n_2!\cdots n_m!$ always divides $(n_1 + n_2 + \cdots + n_m)!$. Beware of a slight notational conflict: the binomial coefficient $\binom{n}{k}$ equals the multinomial coefficient $\binom{n}{k, n-k}$.*

Multinomial coefficients arise in a range of counting problems.

Example 1.54. *If there are* 5 *different presents under a Christmas tree and* 3 *virtuous children who wait to be given what they deserve, how many ways are there to distribute the presents so that child A gets two, child B gets two, and child C gets one? This amounts to choosing a disjoint union of the set of presents into three subsets of sizes* 2, 2, *and* 1, *so the answer is* $\binom{5}{2,2,1} = 30$.

Example 1.55. *How many ways are there to rearrange the letters of WOOLLOOMOOLOO? (The answer is not* 13!, *because of the repeated letters.) In other words, how many different strings are there which consist of one W, one M, three L's, and eight O's? As in the proof of the Multinomial Theorem, to specify such a string we have to express the* 13 *positions as a disjoint union of a subset of size* 1 *(the position of W), another subset of size* 1 *(the position of M), a subset of size* 3 *(the positions of the L's), and a subset of size* 8 *(the positions of the O's). So the answer is* $\binom{13}{1,1,3,8} = \frac{13!}{1!1!3!8!} = 25740$.

Example 1.56*. *How many ways are there to split* kn *people into* n *groups of* k*? The answer depends on whether you consider the groups to be ordered, in the sense of caring which group is which. If the groups are ordered (say, because you want to assign each group to a different task), then this is a special case of Theorem 1.51, and the answer is* $\binom{kn}{k,k,\cdots,k} = \frac{(kn)!}{(k!)^n}$. *If there is no ordering of the groups (say, because you want all the groups to work amongst themselves on the same task), then we can use the Overcounting Principle: every unordered grouping corresponds to exactly* $n!$ *ordered groupings, so the answer is* $\frac{(kn)!}{(k!)^n\, n!}$. *Note that in the case* $k = 2$ *this becomes the same answer that we gave in Example 1.40.*

Example 1.57*. *Suppose you have* 16 *tennis players in a tournament,* 4 *of whom are seeds. How many ways are there to construct a standard four-round knock-out draw so that no two seeds can possibly play each other before the semi-finals? The trick here is to notice that in any such draw, the players are divided into four groups of* 4, *with one seed in each group; each group plays a mini-tournament amongst itself in the first two rounds, with the winner progressing to the semi-finals. If you imagine each group as named after the seed it contains, then assigning the other twelve players to the groups amounts to splitting them into* 4 *ordered groups of* 3, *so it can be done in* $\binom{12}{3,3,3,3}$ *ways. Then within each group, there are three ways to form the two-round mini-tournament; and there are also three ways to allocate the group winners to semi-finals. So the answer is* $3^5\binom{12}{3,3,3,3} = 89812800$.

One can think of $\binom{n}{k}$ as the number of ways to make an unordered selection of k things from a total of n with repetition not allowed. It is then natural to consider what happens if you allow repetition.

Example 1.58. *If you have 5 different marbles in a bag – say red, blue, white, yellow and green – and stick your hand in to pull out 3, there are* $\binom{5}{3} = 10$ *possible outcomes. (Remember that if you draw out the marbles one by one, and you consider that the order in which you draw them out matters, there are* $5_{(3)} = 60$ *possible outcomes.) To allow repetition, suppose that, instead of just one marble of each colour, you have an unlimited supply of each colour, indistinguishable amongst themselves. Then you may get more than one marble of a particular colour when you stick your hand in the bag and pull out 3. To specify an outcome, you just need to specify the number of red marbles, the number of blue marbles, and so on; together, these numbers constitute a 5-tuple* $(k_1, k_2, k_3, k_4, k_5)$ *of nonnegative integers such that* $k_1+k_2+k_3+k_4+k_5 = 3$*. One way to count the possible outcomes is by dividing them into three disjoint subsets according to the following cases:*

(1) *all three marbles have the same colour (i.e.* $k_s = 3$ *for some* s*, and all the other* k_i*'s are zero); or*

(2) *two marbles have the same colour, and the third is a different colour (i.e.* $k_s = 2$ *and* $k_t = 1$ *for some* $s \neq t$*, and the other* k_i*'s are zero); or*

(3) *all three marbles have different colours (i.e. three of the* k_i*'s are* 1*, and the other two* k_i*'s are zero).*

You can calculate that

the number of outcomes of type (1) is ☐

the number of outcomes of type (2) is ☐

the number of outcomes of type (3) is ☐

the total number of outcomes is ☐

Remark 1.59*. *It may seem that we could count unordered selections of k from n with repetition allowed by taking the count of ordered selections of k from n with repetition allowed – which as we have seen is just n^k – and dividing by $k!$, in an application of the Overcounting Principle. But this gives the wrong answer in Example 1.58, and indeed $\frac{n^k}{k!}$ is not usually an integer. The reason the Overcounting Principle doesn't apply is that the number of orderings of one of our unordered selections is not always the same. For instance, in Example 1.58, a selection of three marbles of different colours corresponds to* 3! *ordered selections, whereas a selection of three marbles of the same colour corresponds to only one ordered selection.*

Here is the answer to the problem in general.

Theorem 1.60. Let $n, k \in \mathbb{N}$. When making an unordered selection of k things from n possibilities, with repetition allowed, the number of different outcomes is

$$\binom{n+k-1}{k} = \frac{n(n+1)(n+2)\cdots(n+k-1)}{k!}.$$

Remark 1.61. *Notice the similarity between the formulas in Theorems 1.42 and 1.60: the difference is that in Theorem 1.60, the factors in the numerator rise from n in steps of* 1*, instead of falling from n as they do in $\binom{n}{k}$. Some authors use the notation $\left(\binom{n}{k}\right)$ for this, but since it equals $\binom{n+k-1}{k}$ there doesn't seem much need.*

Proof*. We will first give the proof in the case of Example 1.58, where we are making an unordered selection of 3 marbles from 5 possible kinds. Having made a selection, we can lay the marbles out in a row, with the red marbles (if any) first, followed by the blue, the white, the yellow and the green. Imagine putting down dividing lines to mark off the different colours: one line between red and blue, one line between blue and white, one line between white and yellow, and one line between yellow and green. Then you can forget the colours, and what remains is an arrangement of 3 marbles and 4 dividing lines. For instance, the arrangement

$$|\circ\circ|\circ||$$

means that we have no red marbles, two blue marbles, one white marble, and no yellow or green marbles (the 5-tuple is $(0, 2, 1, 0, 0)$). What we have

shown is that the possible outcomes are in bijection with the different strings of three $\circ$'s and four |'s. By the Bijection Principle, the number of outcomes equals the number of such strings, which by Example 1.47 is $\binom{7}{3} = 35$. The general case is no harder: we can imagine that we are selecting k marbles from n different kinds, which amounts to considering strings of k $\circ$'s and $n-1$ |'s. The number of such strings is $\binom{k+n-1}{k} = \binom{k+n-1}{n-1}$, as claimed.

If talk of "marbles" and "dividing lines" strikes you as too informal, the real effect of this argument is to construct a bijection between n-tuples $(k_1, k_2, \cdots, k_n)$ of nonnegative integers which satisfy $k_1 + k_2 + \cdots + k_n = k$ (these are what specify the possible outcomes of the selection, with k_i recording how many times the ith possibility was selected) and $(n-1)$-element subsets of the set $\{1, 2, \cdots, n+k-1\}$ (these specify the positions of the dividing lines in the string). Starting from an n-tuple $(k_1, \cdots, k_n)$, the subset we associate to it is $\{k_1 + 1, k_1 + k_2 + 2, \cdots, k_1 + \cdots + k_{n-1} + n - 1\}$; it is easy to see that the elements here are listed in increasing order, so they are indeed $n-1$ different elements of $\{1, 2, \cdots, n+k-1\}$. For the inverse function, we start with a subset $\{i_1, i_2, \cdots, i_{n-1}\}$, which we can assume is listed in increasing order, and then associate to it the n-tuple $(i_1 - 1, i_2 - i_1 - 1, \cdots, i_{n-1} - i_{n-2} - 1, n + k - 1 - i_{n-1})$, whose elements are easily seen to be nonnegative with sum k. The verification that these maps are inverse to each other is straightforward. □

Example 1.62. *If there are* 5 *identical presents under a Christmas tree, and* 3 *greedy children grabbing at them until they are all claimed, and no present can be shared between more than one child, how many possible results are there? Notice that the only difference between this question and Example 1.32 is that the presents are now indistinguishable. Is this a question about ordered or unordered selection? Is repetition allowed or not allowed? Once again, we should think that we are selecting, for each present, the child it ends up with: thus it is an unordered selection of* 5 *things (children) from* 3 *possibilities, with repetition allowed (because more than one present is allowed to end up with the same child), and the answer is* $\binom{3+5-1}{5} = 21$. *If you find yourself confused about what is being selected, it may help to think in terms of* n*-tuples: in the present case, all that matters is how many presents child A gets, how many presents child B gets, and how many presents child C gets (because the presents are all identical). So this is the problem of counting* 3*-tuples of nonnegative integers which add up to* 5*, not the problem*

of counting 5-tuples of nonnegative integers which add up to 3 (which we solved in Example 1.58). Finally, what happens if we add the reasonable requirement that every child gets at least one present, so that we are counting 3-tuples of positive integers which add up to 5? We can solve this simply by imagining that we give every child one present to start with; then there are 2 presents left, and we are back in the original situation with 2 in place of 5, so the answer is $\binom{3+2-1}{2} = 6$. *Explicitly, the six 3-tuples are*

$$(3,1,1),\ (1,3,1),\ (1,1,3),\ (2,2,1),\ (2,1,2),\ (1,2,2).$$

Remark 1.63.** *In these problems about unordered selection, we are not counting the functions between two sets like we were in the previous section: every outcome of the allocation of presents to children corresponds to many different functions from the set of presents to the set of children, and we are blurring the distinction between these functions by refusing to distinguish between the presents. What we are really counting is the number of equivalence classes of functions* $f : X \to Y$, *where two functions* f *and* f' *are thought of as equivalent if there is some permutation* σ *of the set* X *such that* $f'(x) = f(\sigma(x))$ *for all* $x \in X$ *(in other words,* f' *takes the same values as* f *but for different inputs). Theorem 1.60 says that if* $|X| = k$ *and* $|Y| = n$, *the number of equivalence classes of functions* $f : X \to Y$ *is* $\binom{n+k-1}{k}$ *(compare Theorem 1.30, which counts the functions without imposing the concept of equivalence). We also know that the number of equivalence classes of injective functions* $f : X \to Y$ *is* $\binom{n}{k}$ *(because we just need to specify the* k *values of the function, which must all be different – this is how we counted* k*-element subsets of an* n*-element set in the first place).*

As for the number of equivalence classes of surjective functions, we can find that by the reasoning at the end of Example 1.62. If $k < n$, *there are no surjective functions* $f : X \to Y$, *by the Reverse Pigeonhole Principle. If* $k \geq n$, *then we can let* A *be any* n*-element subset of* X *(it doesn't matter which, because we are only working up to equivalence), define* f *on* A *so that it gives a bijection between* A *and* Y, *and then define* f *on* $X \setminus A$ *in an arbitrary way. Thus the number of equivalence classes of surjective functions* $f : X \to Y$ *equals the number of equivalence classes of functions* $f : X \setminus A \to Y$, *which is* $\binom{n+(k-n)-1}{k-n} = \binom{k-1}{k-n}$. *Remember that we haven't yet seen how to compute the actual number of surjective functions* $f : X \to Y$ *(without equivalence).*

1.4 Inclusion/Exclusion

Definition 1.64. Recall that if A and B are subsets of a set X, then their union, their intersection, and the difference of A from B are defined by:

$$A \cup B := \{x \in X \mid x \in A \text{ or } x \in B\},$$
$$A \cap B := \{x \in X \mid x \in A \text{ and } x \in B\},$$
$$A \setminus B := \{x \in X \mid x \in A \text{ and } x \notin B\} = A \cap (X \setminus B).$$

These operations satisfy a number of basic laws:

Theorem 1.65. If A, B, and C are subsets of X, then:

(1) $A \cup A = A \cap A = A$.

(2) $A \cup \emptyset = A$, $A \cap \emptyset = \emptyset$.

(3) $A \cup B = B \cup A$, $A \cap B = B \cap A$.

(4) $A \cup (B \cup C) = (A \cup B) \cup C$, so we just call it $A \cup B \cup C$.

(5) $A \cap (B \cap C) = (A \cap B) \cap C$, so we just call it $A \cap B \cap C$.

(6) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.

(7) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

(8) $A \setminus (B \cup C) = (A \setminus B) \cap (A \setminus C)$.

(9) $A \setminus (B \cap C) = (A \setminus B) \cup (A \setminus C)$.

(10) If $B \subseteq A$, then $A \setminus (A \setminus B) = B$.

Proof. These are all equivalent to logical statements which are more or less part of the axioms of mathematics. For example, we 'prove' (8) as follows. To say that two subsets A_1 and A_2 are equal is to say that the condition for x to lie in A_1 is equivalent to the condition for x to lie in A_2 (that is, for every choice of x in the universal set X, they are either both true or both

false). In the present case, the condition for x to lie in $A \setminus (B \cup C)$ is that $x \in A$ and $x \notin (B \cup C)$. But saying that $x \notin (B \cup C)$ is equivalent to saying that $x \notin B$ and $x \notin C$ (if it is false that either $x \in B$ or $x \in C$, then it is true that $x \notin B$ and $x \notin C$, and conversely). So $x \in (A \setminus (B \cup C))$ if and only if

$$x \in A \text{ and } x \notin B \text{ and } x \notin C. \tag{1.13}$$

On the other hand, the condition for x to lie in $(A \setminus B) \cap (A \setminus C)$ is that $x \in (A \setminus B)$ and $x \in (A \setminus C)$; that is, $x \in A$ and $x \notin B$ and $x \in A$ and $x \notin C$. Since the second "$x \in A$" is redundant, this is again equivalent to (1.13), and (8) is proved. □

Now (assuming we are dealing with finite sets) how do these operations relate to counting? As noted in our discussion of the Sum Principle, $|A \cup B| = |A| + |B|$ is true only if A and B are disjoint, i.e. $A \cap B = \emptyset$. If there are elements in the intersection $A \cap B$, they are counted twice when we take the sum $|A| + |B|$. To get the right formula for $|A \cup B|$, we need to correct this:

$$|A \cup B| = |A| + |B| - |A \cap B|. \tag{1.14}$$

Example 1.66. *Suppose that by a show of hands it is determined that of the* 80 *students in a maths class,* 30 *are studying physics,* 40 *are studying chemistry, and* 10 *are studying both physics and chemistry. How many are not studying either physics or chemistry? If we let X be the set of students in the class, A the subset of those doing physics, and B the subset of those doing chemistry, then the question asks for $|X \setminus (A \cup B)|$. We have*

$$|A \cup B| = \boxed{}$$

$$|X \setminus (A \cup B)| = \boxed{}$$

Similarly, if we want a formula for $|A \cup B \cup C|$, and start with $|A| + |B| + |C|$ as our first attempt, then we have to correct for the overcounting of the intersections. If $|A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C|$ is our second attempt, we find that it is still not quite right, because elements in the triple intersection $A \cap B \cap C$ have been added three times in $|A| + |B| + |C|$ and subtracted three times in $-|A \cap B| - |A \cap C| - |B \cap C|$. So we need to add

the triple intersection back in:

$$|A\cup B\cup C| = |A|+|B|+|C|-|A\cap B|-|A\cap C|-|B\cap C|+|A\cap B\cap C|. \quad (1.15)$$

Example 1.67. *How many numbers between* 1 *and* 1000 *are divisible by at least one of* 2, 3, *or* 5*? If we let* X *be* $\{1, 2, 3, \cdots, 1000\}$, A *the subset of numbers divisible by* 2 *(i.e. even),* B *the subset of numbers divisible by* 3*, and* C *the subset of numbers divisible by* 5*, then we want to find* $|A\cup B\cup C|$*. We can use* (1.15)*, because the sizes of these subsets and their intersections can be easily determined. The key is the observation that in the set* $\{1, 2, \cdots, n\}$*, the number of multiples of* m *is exactly* $\left\lfloor \frac{n}{m} \right\rfloor$ *(every* m*th number is divisible by* m *– we round down because all the numbers after the last multiple of* m *in the set are wasted). Moreover, a number is divisible by both* 3 *and* 5 *(say) if and only if it is divisible by* 15 *(this relies on the fact that* 3 *and* 5 *have no common divisors, so their least common multiple is just their product). So*

$$\begin{aligned}
|A| &= \left\lfloor \frac{1000}{2} \right\rfloor = 500,\\
|B| &= \left\lfloor \frac{1000}{3} \right\rfloor = 333,\\
|C| &= \left\lfloor \frac{1000}{5} \right\rfloor = 200,\\
|A\cap B| &= \left\lfloor \frac{1000}{6} \right\rfloor = 166,\\
|A\cap C| &= \left\lfloor \frac{1000}{10} \right\rfloor = 100,\\
|B\cap C| &= \left\lfloor \frac{1000}{15} \right\rfloor = 66,\\
|A\cap B\cap C| &= \left\lfloor \frac{1000}{30} \right\rfloor = 33,
\end{aligned}$$

so $|A\cup B\cup C| = 500+333+200-166-100-66+33 = 734$*. Calculations like this are used to determine how many numbers up to a certain point are prime (have no divisors except themselves and* 1*).*

Equations (1.14) and (1.15) are the $n = 2$ and $n = 3$ cases of:

Theorem 1.68 (Inclusion/Exclusion Principle)***.** If $A_1, A_2, \cdots, A_n$ are subsets of a finite set X, then

$$|A_1 \cup A_2 \cup \cdots \cup A_n| = \sum_{k=1}^{n} (-1)^{k-1} \sum_{1 \le i_1 < i_2 < \cdots < i_k \le n} |A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}|.$$

Proof*. We will prove the same property we saw in the special cases: namely, that every element x of $A_1 \cup A_2 \cup \cdots \cup A_n$, no matter which of the A_i's it's in and which it's not in, ends up being counted exactly once after all the additions and subtractions in the right-hand side. Let I be the (nonempty) subset of $\{1, 2, 3, \cdots, n\}$ consisting of those i such that $x \in A_i$. Then $A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}$ contains x if and only if $i_1, i_2, \cdots, i_k$ all belong to I. So for a fixed k, x is counted in $\binom{|I|}{k}$ terms in the inner sum on the right-hand side. Thus the total number of times x is counted on the right-hand side is $\sum_{k=1}^{|I|} (-1)^{k-1} \binom{|I|}{k}$. But by (1.12), $\sum_{k=0}^{|I|} \binom{|I|}{k} (-1)^{k-1} = 0$, so

$$\sum_{k=1}^{|I|} \binom{|I|}{k} (-1)^{k-1} = -\binom{|I|}{0} (-1)^{0-1} = 1.$$

Hence the right-hand side equals the left-hand side. □

One important application of the Inclusion/Exclusion Principle is in counting the number of derangements of a set S. A derangement is a permutation $f : S \to S$ such that $f(s) \neq s$ for all s, i.e. no element of the set is left unchanged.

Example 1.69*. *If you have four different letters and four already addressed envelopes, how many ways are there to put the letters in the envelopes so that no letter gets sent to the right address? This is equivalent to counting the derangements of a set with four elements. Let X be the set of all ways of putting the letters in the envelopes; we have seen that $|X| = 4! = 24$. Let A_1 be the subset of ways in which the first letter is in the right envelope, A_2 the same for the second letter, and similarly define A_3 and A_4. We want to calculate $|X \setminus (A_1 \cup A_2 \cup A_3 \cup A_4)|$, and we will do it using the Inclusion/Exclusion Principle. For this we need to know the sizes of these sets and their intersections. But $|A_i| = 3! = 6$ for all i, because if the ith letter is in the right envelope we are just counting the number of ways of*

arranging the others. Similarly, $|A_i \cap A_j| = 2! = 2$ *for any* $i < j$*, because if the* i*th and* j*th letters are right we are just counting the number of ways of arranging the other two; and* $|A_i \cap A_j \cap A_k| = 1! = 1$ *for any* $i < j < k$*, and* $|A_1 \cap A_2 \cap A_3 \cap A_4| = 0! = 1$ *(one way to get all the letters right). So*

$$\begin{aligned}|A_1 \cup A_2 \cup A_3 \cup A_4| &= |A_1| + |A_2| + |A_3| + |A_4| \\ &\quad - |A_1 \cap A_2| - \cdots - |A_3 \cap A_4| \\ &\quad + |A_1 \cap A_2 \cap A_3| + \cdots + |A_2 \cap A_3 \cap A_4| \\ &\quad - |A_1 \cap A_2 \cap A_3 \cap A_4| \\ &= 4 \times 3! - 6 \times 2! + 4 \times 1! - 0! \\ &= 24 - 12 + 4 - 1 = 15.\end{aligned}$$

Hence the number we want is $24 - 15 = 9$.

Theorem 1.70*. The number of derangements of an n-element set is

$$\sum_{k=0}^{n} (-1)^k \frac{n!}{k!}.$$

Proof*. We may as well assume that the set is $\{1, 2, \cdots, n\}$. As in the preceding example, let X be the set of all permutations of $\{1, 2, \cdots, n\}$, and let A_i be the subset consisting of those permutations which leave i unchanged. Then for any $1 \leq i_1 < i_2 < \cdots < i_k \leq n$,

$$|A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}| = (n-k)!,$$

because once you fix k elements you are just dealing with a permutation of the remaining $n - k$. Moreover, there are $\binom{n}{k}$ such sequences $1 \leq i_1 < i_2 < \cdots < i_k \leq n$. So the Inclusion/Exclusion Principle says that

$$|A_1 \cup A_2 \cup \cdots \cup A_n| = \sum_{k=1}^{n} (-1)^{k-1} \binom{n}{k} (n-k)! = \sum_{k=1}^{n} (-1)^{k-1} \frac{n!}{k!}.$$

Therefore the number of derangements is

$$|X \setminus (A_1 \cup A_2 \cup \cdots \cup A_n)| = n! - \sum_{k=1}^{n} (-1)^{k-1} \frac{n!}{k!} = \sum_{k=0}^{n} (-1)^k \frac{n!}{k!},$$

as claimed. □

Remark 1.71*. *As a consequence of Theorem 1.70, the probability that a randomly chosen permutation of a set of size n is a derangement is $\sum_{k=0}^{n} \frac{(-1)^k}{k!}$. As n tends to infinity, this probability converges to $e^{-1} = \frac{1}{e}$ (about 0.37).*

Another application of the Inclusion/Exclusion Principle is in counting surjective functions.

Example 1.72. *Suppose we have 5 students who need to be assigned to 3 tutors, so that every tutor gets at least one student. How many ways can this be done? Numbering the students and tutors, we see that this is equivalent to finding the number of surjective functions $f : \{1,2,3,4,5\} \to \{1,2,3\}$. Let X be the set of all functions $f : \{1,2,3,4,5\} \to \{1,2,3\}$. Recall that $|X| = 3^5 = 243$. To count the surjective functions, we need to take away those whose range is a proper subset. So let A be the set of functions whose range is contained in $\{1,2\}$ (i.e. the allocations in which the third tutor doesn't get any students), B the set of functions whose range is contained in $\{1,3\}$, and C the set of functions whose range is contained in $\{2,3\}$. Clearly*

$$|A| = |B| = |C| = 2^5 = 32,$$
$$|A \cap B| = |A \cap C| = |B \cap C| = 1,\ |A \cap B \cap C| = 0.$$

So the Inclusion/Exclusion Principle says that

$$|A \cup B \cup C| = 3 \times 32 - 3 \times 1 + 0 = 93,$$

and the number of surjective functions is $|X \setminus (A \cup B \cup C)| = 243 - 93 = 150$.

Theorem 1.73*. If X and Y are sets with $|X| = m$ and $|Y| = k$, then the number of surjective functions $f : X \to Y$ is

$$\sum_{j=0}^{k} (-1)^j \binom{k}{j} (k-j)^m = k^m - \binom{k}{1}(k-1)^m + \binom{k}{2}(k-2)^m - \cdots + (-1)^k \binom{k}{k} 0^m.$$

Proof*. As in the example, let A_y be the set of functions $f : X \to Y$ for which $y \notin f(X)$, for $y \in Y$. Then any intersection $A_{y_1} \cap \cdots \cap A_{y_j}$ consists of functions whose range lies in $Y \setminus \{y_1, \cdots, y_j\}$, and there are $(k-j)^m$ of these. The Inclusion/Exclusion Principle says that

$$\left| \bigcup_{y \in Y} A_y \right| = \sum_{j=1}^{k} (-1)^{j-1} \binom{k}{j} (k-j)^m,$$

which tells us the number of non-surjective functions. So the number of surjective functions $f : X \to Y$ is

$$k^m - \sum_{j=1}^{k} (-1)^{j-1} \binom{k}{j} (k-j)^m = \sum_{j=0}^{k} (-1)^j \binom{k}{j} (k-j)^m,$$

as claimed. □

Note that the $j = k$ term of the sum, $(-1)^k \binom{k}{k} 0^m$, vanishes unless $m = 0$.

Remark 1.74*. *This is a much less straightforward formula than the count of injective functions given by Theorem 1.36. For small values of k (and assuming $m \geq 1$, so the $j = k$ term vanishes), the formula is:*

$$\begin{aligned} k = 1: &\quad 1 \\ k = 2: &\quad 2^m - 2 \\ k = 3: &\quad 3^m - 3 \times 2^m + 3 \\ k = 4: &\quad 4^m - 4 \times 3^m + 6 \times 2^m - 4 \end{aligned}$$

Note the pattern of alternating signs, decreasing bases to the power m, and binomial coefficients. There are two unsatisfactory aspects. Firstly, the number of terms is k, so if we have a problem where k is growing, the formula gets less and less 'closed'. Secondly, the fact that there are minus signs in the sum means that the terms can potentially grow much bigger than the final answer. For example, when $m < k$ we know that the final answer must be zero by the Reverse Pigeonhole Principle, so all the terms just cancel away! To be sure of not making bigger computations than necessary, combinatorialists prefer 'positive' formulas, which do not involve minus signs.

1.5 Stirling numbers

If you experiment with the formula given in Theorem 1.73, you will notice that it evaluates to numbers which are more divisible than you would have a right to expect. We will prove in this section that, just as the number of injective functions $f : X \to Y$ (i.e. $|Y|_{(|X|)}$) is always divisible by $|X|!$, the number of surjective functions $f : X \to Y$ is always divisible by $|Y|!$.

Definition 1.75. For $n, k \in \mathbb{N}$, the Stirling number $S(n, k)$ is the number of ways of writing an n-element set as a disjoint union of k nonempty subsets. A way of writing a set as a disjoint union of nonempty subsets is usually called a partition of the set, and the subsets are called the blocks of the partition. Notice that, unlike in Theorem 1.51, there is no order on these blocks, and their sizes are not specified.

Example 1.76. *The partitions of* $\{1, 2, 3\}$ *with* 2 *blocks can be represented as* $1|23$, $2|13$, *and* $3|12$, *where* $2|13$ *means that you write* $\{1, 2, 3\}$ *as the disjoint union of* $\{2\}$ *and* $\{1, 3\}$, *and so on. So* $S(3, 2) = 3$.

Example 1.77. *The partitions of* $\{1, 2, 3, 4\}$ *with* 2 *blocks are as follows:*

$\square$ *So* $S(4, 2) =$ $\square$.

The partitions of $\{1, 2, 3, 4\}$ *with* 3 *blocks are as follows:*

$\square$ *So* $S(4, 3) =$ $\square$.

Example 1.78. *It is obvious that you can't partition an* n*-element set into more than* n *blocks, so* $S(n, k) = 0$ *when* $k > n$.

Example 1.79. *You can't partition a nonempty set into* 0 *blocks, so we have* $S(n, 0) = 0$ *when* $n > 0$. *When* $n = 0$, *the convention is to declare that there is a single partition of the empty set into* 0 *blocks, so* $S(0, 0) = 1$. *(This is the right value for subsequent formulas.)*

Example 1.80. *If* $n > 0$, *then there is only one way to partition an* n*-element set into* 1 *block (everything has to be lumped in together), and only one way to partition it into* n *blocks (everything has to be alone in its own block). So* $S(n, 1) = S(n, n) = 1$.

This is all a bit reminiscent of the binomial coefficient facts: $\binom{n}{0} = \binom{n}{n} = 1$, and $\binom{n}{k} = 0$ if $k > n$. So it seems like a good idea to display the Stirling numbers $S(n, k)$ in a triangle like Pascal's triangle, with rows corresponding to $n = 1$, $n = 2$, and so on, and columns corresponding to $k = 1$, $k = 2$, and so on. The start of the triangle would be:

$$\begin{array}{cccc} 1 & & & \\ 1 & 1 & & \\ 1 & 3 & 1 & \\ 1 & 7 & 6 & 1 \end{array}$$

We can see already from the fourth row that we are not going to have the same symmetry property as in Pascal's triangle (so there's not much point aligning the centres of the rows). But there is still a rule, like Pascal's recurrence relation, for generating the nth row from the $(n-1)$th:

Theorem 1.81. For $n, k \geq 1$,

$$S(n,k) = S(n-1,k-1) + k\,S(n-1,k).$$

Proof. The proof is similar to the bijective proof we gave of Theorem 1.50. By definition, $S(n,k)$ counts the number of partitions of the set $\{1,2,\cdots,n\}$ into k blocks. Now the element n is either in a block by itself or it's not. Choosing a partition in which n is in a block by itself amounts to choosing a partition of $\{1,2,\cdots,n-1\}$ into $k-1$ blocks, so there are $S(n-1,k-1)$ such partitions. Choosing a partition in which n is not on its own amounts to choosing a partition of $\{1,2,\cdots,n-1\}$ into k blocks and then choosing which of these k blocks to put n in. So there are $k\,S(n-1,k)$ such partitions, and the desired recurrence relation follows. □

The coefficient k is the only change from Theorem 1.50 to this, but that is enough to make a big difference. In terms of the triangle, Theorem 1.81 tells us that each entry is the sum of the one above and to the left and the product of the one directly above by the column number. So we can build more of the Stirling triangle:

1							
1	1						
1	3	1					
1	7	6	1				
1	15	25	10	1			
1	31	90	65	15	1		
1	63	301	350	140	21	1	
⋮		⋮		⋮		⋮	⋱

With more data, we can spot some patterns in the Stirling numbers.

Example 1.82. *The second column consists of the tower of Hanoi numbers, i.e. those which are one less than a power of* 2*. The reason for this is that every partition of* $\{1,2,\cdots,n\}$ *into two blocks gives you two nonempty proper*

subsets of $\{1, 2, \cdots, n\}$, which are complements of each other. ("Proper" means "not equal to the whole set".) Moreover, every such subset A features in exactly one partition with two blocks, namely the partition into A and $\{1, 2, \cdots, n\} \setminus A$. So by the Overcounting Principle, $S(n, 2)$ is half the number of nonempty proper subsets of $\{1, 2, \cdots, n\}$. Since the total number of subsets of $\{1, 2, \cdots, n\}$ is 2^n, there are $2^n - 2$ nonempty proper subsets, so $S(n, 2) = \frac{2^n - 2}{2} = 2^{n-1} - 1$ for all $n \geq 1$.

Example 1.83. *The diagonal just inside the right-hand edge coincides with one of the diagonals of Pascal's triangle. The reason for this is that in any partition of $\{1, 2, \cdots, n\}$ into $n - 1$ blocks, there must be two elements in one block, and every other element is in its own block. The two elements can be chosen in $\binom{n}{2}$ ways, so $S(n, n-1) = \binom{n}{2}$ for all $n \geq 1$.*

Although we do not have a closed formula for the Stirling numbers, Theorem 1.81 gives us a good handle on them, which we will be able to use more effectively in later chapters. To the extent that we understand the Stirling numbers, the following result is a better formula for the number of surjective functions than Theorem 1.73:

Theorem 1.84. If X and Y are sets with $|X| = m$ and $|Y| = k$, then the number of surjective functions $f : X \to Y$ is $k!\, S(m, k)$.

Proof. As observed before, every function $f : X \to Y$ gives rise to a way of writing X as a disjoint union, namely as the disjoint union of the preimages $f^{-1}(y)$ for $y \in Y$ (that is, you group together elements of X on which f takes the same value). Saying that f is surjective is the same as saying that all these preimages are nonempty, so in that case we get a partition of X into k blocks. For a fixed partition of X into k blocks, there are $k!$ surjective functions $f : X \to Y$ which give rise to it, because once you have grouped together the elements of X which get sent to the same element of Y by f, you just need to decide which block gets sent to which element of Y. So the number of surjective functions is $k!$ times the number of partitions into k blocks. □

Example 1.85. *Recall from Example 1.77 the partitions of $\{1, 2, 3, 4\}$ into three blocks. Each of these gives rise to $3! = 6$ surjective functions $f : \{1, 2, 3, 4\} \to \{1, 2, 3\}$: for instance, the partition $23|1|4$ corresponds to the six functions for which $f(2) = f(3)$, but $f(1)$ and $f(4)$ are different from this*

and from each other. There are six of these because this is the number of ways of deciding which of $1, 2, 3$ *is to be* $f(2) = f(3)$*, which is to be* $f(1)$*, and which is to be* $f(4)$*.*

Generalizing the idea of Theorem 1.84, we have a relationship between four of the key sorts of numbers we have encountered in this chapter: integer powers, factorials, binomial coefficients, and Stirling numbers.

Theorem 1.86*. For any $n, m \in \mathbb{N}$,

$$n^m = \sum_{k=0}^{m} k!\, S(m,k) \binom{n}{k}.$$

Proof*. By Theorem 1.30, n^m counts the functions $f : \{1, 2, \cdots, m\} \to \{1, 2, \cdots, n\}$. We can count these another way using the Sum Principle, dividing the functions according to the size of the range $f(\{1, 2, \cdots, m\})$. If we stipulate that $|f(\{1, 2, \cdots, m\})| = k$, then we can choose what k-element subset J of $\{1, 2, \cdots, n\}$ this range actually is in $\binom{n}{k}$ ways. Having chosen J, the choice of f amounts to the choice of a surjective function from $\{1, 2, \cdots, m\}$ to J. By Theorem 1.73, there are $k!\, S(m,k)$ of these. Putting this all together, we get the formula in the statement. □

Remark 1.87*. *The* $k = 0$ *term of the sum in Theorem 1.86 is usually zero, because* $S(m, 0) = 0$ *for* $m \geq 1$*: this is put in purely for the case when* $m = 0$*, when the theorem says that* $1 = 1 \times 1 \times 1$*.*

Chapter 2

Recursion and Induction

Recursion is a phenomenon that is widespread in discrete mathematics. It occurs when the answer to some problem which depends on a nonnegative integer n makes use of the answers to the same problem for smaller values of n. In other words, if the answer to the problem for n is a_n, then the sequence $a_0, a_1, a_2, \cdots$ satisfies a recurrence relation, which expresses a_n (for sufficiently large values of n) as a function of the earlier terms in the sequence, i.e. $a_0, a_1, a_2, \cdots, a_{n-1}$. In this chapter we will examine many such sequences, and show how to prove facts about them by induction. What we regard as the goal, usually, is to find a closed formula for a_n, one which does not involve the earlier terms (and avoids anything equally inconclusive, like an unevaluated sum of n things). This is what is meant by 'solving' a recurrence relation.

2.1 Examples of recursive sequences

A sequence $a_0, a_1, a_2, a_3, \cdots$ which satisfies a recurrence relation is said to be recursive; every term to which the recurrence relation applies is determined by the earlier ones. To get the whole thing rolling, there must be at least one term at the beginning defined separately as an initial condition.

The simplest type is when each term depends only on the previous one, i.e. a_n (for $n \geq 1$) is a function of a_{n-1}, and only a_0 needs to be given as an initial condition; this was the case in the tower of Hanoi problem, for example. But there are also important cases where earlier terms of the sequence are involved.

Example 2.1. *One of the most famous of all recursive sequences was mentioned first in 1202 by the Italian mathematician Fibonacci. He imagined a rabbit farm which on Day 1 contains only a single newborn breeding pair of rabbits. On Day 2 these rabbits mature, so that on Day 3 the female gives birth to a new pair of rabbits. On Day 4 the new pair is still maturing, but the older female gives birth to a third pair. On Day 5 the newest pair is still maturing, but the other two females give birth to new pairs, so there are now 5 pairs, and so on. If we let F_n denote the number of pairs after n days, Fibonacci's assumptions imply that $F_n - F_{n-1}$ (the number of new pairs born on Day n) equals F_{n-2} (the number of pairs existing on Day $n-2$). Formally, we define the Fibonacci sequence by*

$$F_0 = 0,\ F_1 = 1,\ F_n = F_{n-1} + F_{n-2} \text{ for } n \geq 2. \tag{2.1}$$

(Starting the sequence with a 0 makes various formulas nicer.) Notice that because the recurrence relation doesn't kick in until we come to F_2, we have to define F_0 and F_1 as initial conditions. The Fibonacci sequence begins:

$$0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233, \cdots$$

and we will prove many interesting facts about it as we go, including a surprising closed formula for F_n in Example 2.24.

Example 2.2. *Another famous sequence is the sequence of Catalan numbers c_n. These arise in an amazingly wide range of contexts: in the book 'Enumerative Combinatorics' Vol. 2 by R. P. Stanley, there is a list of 66 counting problems, to all of which the answer is the Catalan numbers. The one which is most often used as the definition is that c_n is the number of balanced strings of n left brackets and n right brackets, where "balanced" means that as you read from left to right, the number of right brackets never exceeds the number of left brackets. For instance, $c_3 = 5$ because the balanced strings of 3 left brackets and 3 right brackets are as follows:*

$$()()(),\ ()(()),\ (())(),\ (()()),\ \text{and } ((())).$$

What's more important for us, however, is the recursive definition:

$$c_0 = 1,\ c_n = c_0c_{n-1} + c_1c_{n-2} + \cdots + c_{n-2}c_1 + c_{n-1}c_0 \text{ for } n \geq 1. \tag{2.2}$$

As usual, the right-hand side of this recurrence relation is to be interpreted according to the pattern rather than the literal sense for small values of n: it always has n terms, so when $n = 1$ it is just c_0c_0, when $n = 2$ it is $c_0c_1 + c_1c_0$, etc. It would probably be better to avoid this problem of interpretation by writing the recurrence relation using sigma notation:

$$c_n = \sum_{m=0}^{n-1} c_m c_{n-1-m}.$$

The reason that the recurrence relation holds, if you define the Catalan numbers using balanced strings of brackets, is that every such string must have the form

$$(\textit{first balanced substring})\,\textit{second balanced substring},$$

i.e. there is a unique right bracket which corresponds to the initial left bracket, and it must occur in the $(2m+2)$th position for some m between 0 and $n-1$; the substring enclosed by these brackets, and the substring which follows them, must both themselves be balanced. For any fixed m, the number of ways of choosing the first balanced substring is c_m, and the number of ways of choosing the second balanced substring is c_{n-1-m}; the recurrence relation follows. The Catalan sequence begins $1, 1, 2, 5, 14, 42, 132, 429, \cdots$.

Even some familiar sequences which we wouldn't normally think of as recursive really are. The tell-tale sign is a "$\cdots$" in a formula, or a verbal instruction to carry out some operation a variable number of times.

Example 2.3. *If asked for the definition of 2^n, most people would say:*

$$2^n = 2 \times 2 \times 2 \times \cdots \times 2 \ (n \textit{ twos}).$$

In other words, 2^n is the result of starting with 1 and doubling n times. Since, in this procedure, you necessarily reach 2^{n-1} in the step before reaching 2^n, the definition may as well be expressed using a recurrence relation:

$$2^0 = 1,\ 2^n = 2 \times 2^{n-1} \text{ for } n \geq 1. \tag{2.3}$$

So you could argue that our solution of the tower of Hanoi problem was not all it was cracked up to be: we just expressed one recursive sequence, h_n, in terms of another, 2^n. But, of course, the latter is something we understand so well that expressing h_n in terms of it counts as a solution.

Example 2.4. *Similarly, the "$\cdots$" in the definition $n! = n(n-1)\cdots 2 \times 1$ is a sign that we can define factorials recursively:*

$$0! = 1,\ n! = n \times (n-1)! \text{ for } n \geq 1. \tag{2.4}$$

Although the sequence of factorials is slightly more complicated than the sequence of powers of 2, we still regard $n!$ as something known, and we are quite happy if we can express some other sequence in terms of factorials. For instance, in Example 3.45 we will prove a famous formula for the Catalan numbers: $c_n = \frac{(2n)!}{(n+1)!\,n!}$.

Example 2.5. *In many problems we have some function f of a nonnegative integer variable, and we then have to consider the function of n defined by*

$$a_n := f(0) + f(1) + f(2) + \cdots + f(n) = \sum_{m=0}^{n} f(m).$$

The sequence $a_0, a_1, a_2, \cdots$ can obviously be defined recursively:

$$a_0 = f(0),\ a_n = a_{n-1} + f(n) \text{ for } n \geq 1. \tag{2.5}$$

Conversely, any recurrence relation which is of this sum type, where each term of the sequence a_n equals the previous term plus some fixed function of n, gives rise to the corresponding sequence of sums $\sum_{m=0}^{n} f(m)$, because we can just unravel the recurrence:

$$\begin{aligned} a_n &= a_{n-1} + f(n) \\ &= a_{n-2} + f(n-1) + f(n) \\ &= a_{n-3} + f(n-2) + f(n-1) + f(n) \\ &\quad\ \vdots \\ &= a_0 + f(1) + f(2) + \cdots + f(n-1) + f(n) \\ &= f(0) + f(1) + f(2) + \cdots + f(n-1) + f(n). \end{aligned}$$

Such an unravelling procedure is often possible: the idea is that you substitute in the recursive formula for a_n the recursive formula for a_{n-1}, and

keep substituting until a pattern is clear, whereupon you jump to the point at which the terms of the sequence disappear, leaving a large formula with an inevitable "$\cdots$". This is not as helpful as you might think, because it is often just rephrasing the recurrence relation in a more cumbersome way.

Example 2.6. *Consider the sequence defined by*

$$a_0 = 1,\ a_n = \sqrt{1 + a_{n-1}} \text{ for } n \geq 1.$$

If you unravel the recurrence relation, you get

$$\begin{aligned} a_n &= \sqrt{1 + a_{n-1}} \\ &= \sqrt{1 + \sqrt{1 + a_{n-2}}} \\ &= \sqrt{1 + \sqrt{1 + \sqrt{1 + a_{n-3}}}} \\ &\ \ \vdots \\ &= \sqrt{1 + \sqrt{1 + \sqrt{1 + \cdots \sqrt{1 + 1}}}} \quad \text{(with } n \text{ } \surd \text{ signs)}. \end{aligned}$$

However, this way of writing a_n tells us nothing more than the recurrence relation did; it certainly doesn't count as a solution of the recurrence relation.

However, there are cases where the unravelling procedure helps to reveal that the problem is of a kind you already know how to solve.

Example 2.7. *The triangular number t_n is the number of dots in a triangular array with n dots on each side. Since removing one of the sides produces a smaller triangular array, this has a recursive definition of sum type:*

$$t_0 = 0,\ t_n = t_{n-1} + n \text{ for } n \geq 1.$$

Unravelling this recurrence relation gives the non-closed formula

$$t_n = \boxed{}$$

Summing the arithmetic progression gives the solution of the recurrence:

$$t_n = \boxed{}$$

This sequence begins $0, 1, 3, 6, 10, 15, 21, 28, \cdots$.

Example 2.8. *Consider the sequence defined by*

$$a_0 = 1,\ a_n = 3a_{n-1} + 1 \text{ for } n \geq 1.$$

The unravelling procedure gives:

$$\begin{aligned} a_n &= 3a_{n-1} + 1 = 3(3a_{n-2} + 1) + 1 \\ &= 3^2 a_{n-2} + 3 + 1 = 3^2(3a_{n-3} + 1) + 3 + 1 \\ &= 3^3 a_{n-3} + 3^2 + 3 + 1 \\ &\ \ \vdots \\ &= 3^n a_0 + 3^{n-1} + 3^{n-2} + \cdots + 3 + 1 \\ &= 3^n + 3^{n-1} + 3^{n-2} + \cdots + 3 + 1. \end{aligned}$$

This would be no real help, except that we recognize this as the sum of a geometric progression, and hence obtain the closed formula

$$a_n = \frac{3^{n+1} - 1}{3 - 1} = \frac{3^{n+1} - 1}{2},$$

which counts as a solution of the recurrence.

Remark 2.9*. *It is wise to be cautious about arguments, such as that in Example 2.8, in which a chain of equalities goes on for a variable number of steps. After all, if we are supposed to be writing a logical proof in which each line is deducible from the previous one, are we really allowed to say to the reader, in effect, "the next lot of lines all follow the same pattern, but I can't actually tell you how many there are"? Just as horizontal dots are a sign that some definition is really recursive, vertical dots are a sign that some proof is really a proof by induction (see Example 2.13).*

Example 2.10. *In Chapter 1 we encountered two important collections of numbers which were more naturally displayed as a triangle than as a sequence, because they depended on not one but two nonnegative integers: the binomial coefficients* $\binom{n}{k}$ *and the Stirling numbers* $S(n, k)$*. We can naturally extract sequences from these numbers by fixing* k*, which corresponds to looking at a single column of the Pascal or Stirling triangle (for instance, the* $k = 2$ *column of Pascal's triangle gives us the sequence of triangular numbers). In both cases we have a recurrence relation (Theorems 1.50 and 1.81), which expresses an element of the* k*th column in terms of the previous element of the* k*th column and also an element of the* $(k-1)$*th column. So*

we can't really consider the kth column on its own as a recursive sequence, until we have a closed formula for the elements of the $(k-1)$th column; this suggests that we should analyse the columns from left to right. This idea is rather pointless in the case of Pascal's triangle, because we already know the closed formula $\binom{n}{k} = \frac{n!}{k!(n-k)!}$. But it will contribute to our understanding of the Stirling numbers in the next chapter.

2.2 Proof by induction

The principle of mathematical induction, or induction for short, is one of the most common and most powerful methods of proof in discrete mathematics. The context in which it is used is that one wants to prove for all $n \in \mathbb{N}$ some statement $S(n)$ which depends on n. The principle of induction, in its most general form, says that if we prove the following two statements:

(1) (base case) $S(0)$ is true;

(2) (inductive step) for any $n \geq 1$, the truth of $S(n)$ follows from the truth of $S(0), S(1), S(2), \cdots, S(n-1)$,

then we have proved that $S(n)$ is true for all $n \in \mathbb{N}$.

The reason this works is that $S(0)$ implies $S(1)$, and then $S(0)$ and $S(1)$ together imply $S(2)$, and then $S(0)$, $S(1)$, and $S(2)$ together imply $S(3)$, and so on. So as long as the base case can be proved (which is often a simple check), we can assume any or all of $S(0), S(1), S(2), \cdots, S(n-1)$ in our attempts to prove $S(n)$; the assumption we make is called the induction hypothesis. This is especially useful when $S(n)$ involves the terms of a recursive sequence.

Example 2.11. *Consider the sequence defined by*

$$a_0 = 2,\ a_n = a_{n-1}^2 + 1 \text{ for } n \geq 1.$$

The sequence begins $2, 5, 26, 677, \cdots$. To give some idea of how fast it grows, we want to prove that $a_n \geq 2^{2^n}$ for all $n \in \mathbb{N}$. We do this by induction, making $S(n)$ the statement $a_n \geq 2^{2^n}$. The base case $S(0)$ asserts that $2 \geq 2^1$, which

is clearly true. Then in proving $S(n)$ for $n \geq 1$, we are allowed to assume that $S(0), S(1), \cdots, S(n-1)$ are true. Actually we only need to use the truth of $S(n-1)$ in the proof:

$$a_n = a_{n-1}^2 + 1 \geq (2^{2^{n-1}})^2 + 1 = 2^{2^n} + 1 \geq 2^{2^n}.$$

This completes the inductive step, so $a_n \geq 2^{2^n}$ is proved for all $n \in \mathbb{N}$.

Example 2.12. *Suppose that $S(n)$ is the inequality $F_n < 2^n$. To start the induction, we check that $S(0)$ is true: $S(0)$ asserts that $0 < 1$, so this is obvious. Now we want to carry out the induction step, in which we prove $S(n)$ having made the assumption that $S(0), S(1), \cdots, S(n-1)$ are true. The natural starting point is to apply the recurrence relation $F_n = F_{n-1} + F_{n-2}$. Conveniently, our inductive hypothesis includes the statements $S(n-1)$ and $S(n-2)$, which say that $F_{n-1} < 2^{n-1}$ and $F_{n-2} < 2^{n-2}$. We conclude that*

$$F_n < 2^{n-1} + 2^{n-2} = \frac{1}{2} \times 2^n + \frac{1}{4} \times 2^n = \frac{3}{4} \times 2^n < 2^n.$$

This appears to completes the inductive step. However, there is a gap in the proof: to use the Fibonacci recurrence, we need to assume that $n \geq 2$, so we haven't yet covered the $n = 1$ case. This is a dangerous gap, because in an induction proof, all the statements $S(n)$ for $n \geq 2$ rely on the truth of $S(1)$. Fortunately, $S(1)$ just says that $1 < 2$, so it is obviously true and the proof is finished. If you like, you can regard $S(1)$ as an extra base case which we should have checked along with $S(0)$; alternatively, the induction step is divided into two cases depending on whether $n = 1$ or $n \geq 2$.

Perhaps because assuming smaller versions of what you are trying to prove feels a bit like cheating, there is an aesthetic (rather than logical) feeling that the induction hypothesis should be restricted as much as possible. As the preceding examples suggest, if you are proving a statement about a recursively-defined sequence a_n where the recurrence relation expresses a_n in terms of $a_{n-1}, a_{n-2}, \cdots, a_{n-k}$ for some k, you probably only need to appeal to $S(n-1), S(n-2), \cdots, S(n-k)$ in proving $S(n)$. In particular, if a_n is a function of a_{n-1} only, then assuming $S(n-1)$ is usually enough to prove $S(n)$, as it was in Example 2.11; this form of the argument is sometimes distinguished by the special term "weak induction". However, since the proof of $S(n-1)$ relies on $S(n-2)$, whose proof relies on $S(n-3)$, and so on, all the earlier statements are still implicitly involved.

Example 2.13*. *After defining a recurrence relation of sum type in* (2.5)*, we showed using an unravelling argument that it gave the same sequence as the definition* $a_n = \sum_{m=0}^{n} f(m)$*. If you share the doubts expressed in Remark 2.9 about the unravelling, the true justification is the following proof by induction. The base case is the statement that* $a_0 = f(0)$*, which is given as the initial condition. In proving the statement for* $n \geq 1$*, we are allowed to assume that* $a_{n-1} = \sum_{m=0}^{n-1} f(m)$ *is true, and then the recurrence gives*

$$a_n = \left(\sum_{m=0}^{n-1} f(m) \right) + f(n) = \sum_{m=0}^{n} f(m),$$

completing the inductive step and finishing the proof. The ease of such induction proofs means that we usually don't worry about these matters.

Of course, the same principles apply to statements which are to be proved for all $n \geq 1$, or all $n \geq 2$, etc. You just have to start with the appropriate base case. (Actually, you could put these variants in the $n \in \mathbb{N}$ framework, by defining $S'(n)$ to be $S(n+1)$ or $S(n+2)$, etc., but that's pedantic.)

Example 2.14. *We prove by induction that* $2^n < n!$ *for all* $n \geq 4$*. The base case is that* $2^4 < 4!$*, i.e.* $16 < 24$*. Now we can assume that* $n \geq 5$*, and that* $2^{n-1} < (n-1)!$ *is true; then*

$$2^n = 2 \times 2^{n-1} < 2 \times (n-1)! < n \times (n-1)! = n!,$$

completing the induction step and finishing the proof.

Example 2.15*. *A famous example of induction is the proof that every integer* $n \geq 2$ *has a divisor which is prime. (Recall that a positive integer* n *is said to be prime if it is bigger than* 1 *and has no divisors other than* 1 *and* n*.) For the base case we need to show that* 2 *has a divisor which is prime: but clearly* 2 *itself is prime, and a number is always a divisor of itself. Now we can assume that* $n \geq 3$*, and (as our induction hypothesis) that every number from* 2 *up to* $n-1$ *has a prime divisor. If* n *itself is prime, the statement is automatically true. If* n *is not prime, then it must have some divisor* d *which is neither* 1 *nor* n*. The induction hypothesis includes the fact that* d *has a prime divisor, say* e*. Since* e *is a divisor of* d*, and* d *is a divisor of* n*,* e *is a divisor of* n*, so the statement is true in this case also. This completes the proof. You may think that this fact can be easily proved without induction, as follows: if* n *is prime we are finished; otherwise, we can*

write n as the product of two smaller numbers which divide it; if either of those is prime we are finished; otherwise, we write them in turn as products of smaller numbers; and keep going in this way until we break everything into a product of primes. There is nothing wrong with this, except that the phrase "keep going in this way until" falls foul of the same sort of objection as in Remark 2.9. Really it requires an induction argument to make it rigorous.

The availability of proof by induction is very useful in solving recurrence relations. One great advantage that recurrence relations have over the somewhat similar differential equations is that you can directly work out as many terms of the sequence as you have the patience to do. There may be a pattern which suggests a closed formula for the general term, and you can then try to prove this formula by induction.

Example 2.16. *(For this example, forget that you know how to sum an arithmetic progression.) Consider the sequence defined by*

$$a_0 = 0,\ a_n = a_{n-1} + (2n-1) \text{ for } n \geq 1.$$

This is a recurrence of the sum type, in the terminology introduced in the previous section. After unravelling, it says that a_n is the sum of the first n odd numbers:

$$a_n = 1 + 3 + 5 + \cdots + (2n-1).$$

Direct calculation shows that the sequence starts $0, 1, 4, 9, 16, 25, \cdots$, from which it is not hard to guess the formula $a_n = n^2$. So we make this our statement $S(n)$, and aim to prove it by induction. The base case $S(0)$ is given to us as the initial condition, so all that remains is to prove $S(n)$ for $n \geq 1$, assuming $S(n-1)$. This is easy:

$$a_n = a_{n-1} + (2n-1) = (n-1)^2 + 2n - 1 = n^2 - 2n + 1 + 2n - 1 = n^2.$$

Notice that what the proof boils down to is the fact that the sequence of squares satisfies the same initial condition and recurrence relation as a_n, and therefore must be the same sequence.

The last comment in Example 2.16 is a general principle, almost trivial but worth making explicit: if two sequences satisfy the same initial conditions and the same recurrence relation, they must be the same. Formally:

Theorem 2.17. Suppose that $a_0, a_1, a_2, \cdots$ and $b_0, b_1, b_2, \cdots$ are two sequences, the first of which satisfies some initial conditions:

$$a_0 = I_0,\ a_1 = I_1,\ a_2 = I_2,\ \cdots,\ a_{k-1} = I_{k-1},$$

and some recurrence relation:

$$a_n = R^{(n)}(a_0, a_1, \cdots, a_{n-1}),\ \text{for all } n \geq k,$$

where $k \geq 1$ is fixed but the function $R^{(n)}$ possibly varies with n. If the second sequence also satisfies the same initial conditions:

$$b_0 = I_0,\ b_1 = I_1,\ b_2 = I_2,\ \cdots,\ b_{k-1} = I_{k-1},$$

and the same recurrence relation:

$$b_n = R^{(n)}(b_0, b_1, \cdots, b_{n-1}),\ \text{for all } n \geq k,$$

then $a_n = b_n$ for all $n \in \mathbb{N}$.

Proof. The reason is the principle of induction: we are given that $a_n = b_n$ for all $0 \leq n \leq k-1$, which gives us all the base cases we could possibly need. Then in proving $a_n = b_n$, we can assume that $n \geq k$ and that $a_0 = b_0$, $a_1 = b_1$, and so on up to $a_{n-1} = b_{n-1}$. We then have

$$a_n = R^{(n)}(a_0, a_1, \cdots, a_{n-1}) = R^{(n)}(b_0, b_1, \cdots, b_{n-1}) = b_n,$$

completing the induction step. □

So if we have guessed a formula for a_n, say $f(n)$, and we want to prove that it really is true that $a_n = f(n)$ for all $n \in \mathbb{N}$, all we need to do is to check that $f(n)$ satisfies the same initial conditions and recurrence relation that a_n does. Although induction need not be explicitly mentioned, this will boil down to an induction proof, since Theorem 2.17 is proved by induction.

Example 2.18. *Consider the sequence defined by*

$$a_0 = 0,\ a_n = a_{n-1} + \frac{1}{n(n+1)}\ \textit{for } n \geq 1.$$

This recurrence relation is of sum type, so we could alternatively write

$$a_n = \frac{1}{1\times 2} + \frac{1}{2\times 3} + \cdots + \frac{1}{(n-1)n} + \frac{1}{n(n+1)}. \tag{2.6}$$

Working out the first few terms of the sequence, we find that

$$a_1 = \frac{1}{2},\ a_2 = \frac{2}{3},\ a_3 = \frac{3}{4},\ a_4 = \frac{4}{5},$$

from which we naturally guess that $a_n = \frac{n}{n+1}$ *for all* $n \in \mathbb{N}$*. To prove this, we just need to show that the candidate formula* $\frac{n}{n+1}$ *satisfies the initial condition and the recurrence relation. The initial condition is obvious:* $\frac{0}{1} = 0$*. For the recurrence relation we need to show that* $\frac{n}{n+1} = \frac{n-1}{n} + \frac{1}{n(n+1)}$*, which is true by a trivial calculation. So the claim is proved. In fact,* (2.6) *is an example of a "telescoping sum": since* $\frac{1}{k(k+1)} = \frac{1}{k} - \frac{1}{k+1}$*, the right-hand side becomes*

$$\frac{1}{1} - \frac{1}{2} + \frac{1}{2} - \frac{1}{3} + \cdots + \frac{1}{n-1} - \frac{1}{n} + \frac{1}{n} - \frac{1}{n+1},$$

and every term cancels out except for the first and last, giving us $a_n = 1 - \frac{1}{n+1} = \frac{n}{n+1}$*. Although such telescoping arguments are aesthetically more pleasing, they really boil down to the same thing as the induction method described above: the cancellation of an unspecified number of terms needs induction to justify it properly.*

Example 2.19. *The formulas for the sum of an arithmetic progression and the sum of a geometric progression can both be proved by this induction method. In the geometric case, we have to prove that for all* $r \neq 1$*,*

$$1 + r + r^2 + r^3 + \cdots + r^{n-1} + r^n = \frac{r^{n+1} - 1}{r - 1},\ \text{for } n \in \mathbb{N}. \tag{2.7}$$

The left-hand side can be viewed as the a_n *term of a sequence defined by a recurrence relation of 'sum type', namely* $a_n = a_{n-1} + r^n$*, with initial condition* $a_0 = 1$*. The right-hand side is our candidate formula. That it satisfies the initial condition is easy to see, and then all that remains is to prove that, for* $n \geq 1$*,*

$$\frac{r^{n+1} - 1}{r - 1} = \frac{r^n - 1}{r - 1} + r^n.$$

This is trivial, after multiplying both sides by $r-1$. Of course, most people would prove (2.7) by multiplying both sides by $r-1$ in the first place; but this results in a telescoping sum on the left-hand side, and so the principle of induction would be implicitly involved.

Example 2.20. *Consider the sequence defined by*

$$a_0 = 1,\ a_n = a_{n-1}(2^{2^{n-1}} + 1) \text{ for } n \geq 1.$$

Work out the next few terms:

$$a_1 = \boxed{}$$

$$a_2 = \boxed{}$$

$$a_3 = \boxed{}$$

Can you guess a closed formula that fits all this data?

$$a_n = \boxed{}$$

To prove your formula is correct, all that remains is to check that it satisfies the recurrence relation. Write this out to make sure:

$\boxed{}$ *for* $n \geq 1$.

Example 2.21. *One formula which is surprisingly easy to guess is for the sum of the first n positive integer cubes:*

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

Rephrasing the definition recursively:

$$a_0 = 0,\ a_n = a_{n-1} + n^3 \text{ for } n \geq 1.$$

Working out the first few terms, we find that

$$a_1 = 1,\ a_2 = 9,\ a_3 = 36,\ a_4 = 100,\ a_5 = 225,$$

and notice that all the values are squares. Moreover, the square roots seem to be the triangular numbers, so we conjecture that

$$a_n = \left(\frac{n(n+1)}{2}\right)^2 = \frac{n^2(n+1)^2}{4}.$$

This formula obviously satisfies the initial condition, so all that remains to prove is that

$$\frac{n^2(n+1)^2}{4} = \frac{(n-1)^2n^2}{4} + n^3,$$

which follows from some straightforward manipulation. To sum up, in this example we have shown that

$$\sum_{m=0}^{n} m^3 = \frac{n^2(n+1)^2}{4}, \text{ for all } n \in \mathbb{N}. \tag{2.8}$$

Example 2.22. *To show that guessing a formula is not always straightforward, consider the sequence defined by summing the first n positive squares, whose recursive definition is*

$$a_0 = 0,\ a_n = a_{n-1} + n^2 \text{ for } n \geq 1.$$

Working out the first few terms, we find that

$$a_1 = 1,\ a_2 = 5,\ a_3 = 14,\ a_4 = 30,\ a_5 = 55,\ a_6 = 91,$$

and no very obvious pattern emerges (although if you look at the divisors of these numbers, and spot that a_n seems to be often divisible by $2n+1$, you may get on the right track). We will solve this recurrence in Example 3.47.

Example 2.23*. *Consider the sequence defined by*

$$a_0 = 1,\ a_n = \frac{1}{n}(a_{n-1} + 2a_{n-2} + 3a_{n-3} + \cdots + na_0) \text{ for } n \geq 1.$$

Working out the next few terms, we find that

$$a_1 = 1,\ a_2 = \frac{3}{2},\ a_3 = \frac{13}{6},\ a_4 = \frac{73}{24},\ a_5 = \frac{501}{120}.$$

You have probably spotted that the denominators here are the factorials; this suggests the general conjecture that $n!a_n$ is a positive integer, which is easy to prove by induction. But the numerators remain a bit mysterious for now. We will return to this sequence in Example 3.51.

Example 2.24*. *Here is a formula for the Fibonacci numbers that seems at first sight impossible to guess:*

$$F_n = \frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n\right]. \tag{2.9}$$

With all these fractions and $\sqrt{5}$'s, it's even surprising that the right-hand side works out to be an integer. Nevertheless, following the principle of Theorem 2.17, we can prove this is a correct formula just by checking that it satisfies the initial conditions and recurrence relation which define the Fibonacci sequence. For the initial conditions:

$$n = 0: \qquad \frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^0 - \left(\frac{1-\sqrt{5}}{2}\right)^0\right] = \frac{1}{\sqrt{5}}[1-1] = 0,$$

$$n = 1: \qquad \frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^1 - \left(\frac{1-\sqrt{5}}{2}\right)^1\right] = \frac{\sqrt{5}}{\sqrt{5}} = 1.$$

For the recurrence relation, in which $n \geq 2$:

$$\begin{aligned}
&\frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n\right] \\
&= \frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^2\left(\frac{1+\sqrt{5}}{2}\right)^{n-2} - \left(\frac{1-\sqrt{5}}{2}\right)^2\left(\frac{1-\sqrt{5}}{2}\right)^{n-2}\right] \\
&= \frac{1}{\sqrt{5}}\left[\left(\frac{6+2\sqrt{5}}{4}\right)\left(\frac{1+\sqrt{5}}{2}\right)^{n-2} - \left(\frac{6-2\sqrt{5}}{4}\right)\left(\frac{1-\sqrt{5}}{2}\right)^{n-2}\right] \\
&= \frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}+1\right)\left(\frac{1+\sqrt{5}}{2}\right)^{n-2} - \left(\frac{1-\sqrt{5}}{2}+1\right)\left(\frac{1-\sqrt{5}}{2}\right)^{n-2}\right] \\
&= \frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^{n-1} - \left(\frac{1-\sqrt{5}}{2}\right)^{n-1}\right] \\
&\qquad + \frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^{n-2} - \left(\frac{1-\sqrt{5}}{2}\right)^{n-2}\right].
\end{aligned}$$

2.3 Homogeneous linear recurrence relations

There is a general method for solving (constant-coefficient) linear recurrence relations which is strongly reminiscent of the solution of linear ordinary differential equations. As in that theory, there are two kinds, homogeneous and non-homogeneous.

Definition 2.25. If k is a positive integer, a kth-order homogeneous linear recurrence relation for a sequence a_n is one of the form

$$a_n = r_1 a_{n-1} + r_2 a_{n-2} + \cdots + r_k a_{n-k}, \text{ for all } n \geq k,$$

where $r_1, \cdots, r_k$ are fixed numbers (independent of n), and $r_k \neq 0$.

Remember that such a recurrence relation on its own is not enough to specify the sequence: one also needs to give the first k terms $a_0, a_1, \cdots, a_{k-1}$ as initial conditions. But the crucial idea is to consider all the sequences satisfying the same recurrence relation together.

Example 2.26. *A first-order homogeneous linear recurrence relation is of the form $a_n = ra_{n-1}$ for some nonzero r. To specify the sequence, one needs to give the value of a_0 as an initial condition. If $a_0 = C$, it is clear that $a_n = Cr^n$ for all $n \in \mathbb{N}$. (Strictly speaking, we need an application of Theorem 2.17: the sequence Cr^n satisfies the right initial condition and recurrence relation, therefore it equals a_n.) So we have solved the first-order homogeneous case.*

Example 2.27. *A second-order homogeneous linear recurrence relation is of the form $a_n = ra_{n-1} + sa_{n-2}$ for some numbers r, s with $s \neq 0$ (because having $s = 0$ would make it actually just first-order). For instance, the Fibonacci sequence satisfies such a recurrence relation, with $r = s = 1$. The philosophy now is that we should consider the Fibonacci sequence together with all other solutions of $a_n = a_{n-1} + a_{n-2}$ for which the initial conditions are different. An example of another such is the Lucas sequence defined by*

$$L_0 = 2,\ L_1 = 1,\ L_n = L_{n-1} + L_{n-2} \text{ for } n \geq 2. \tag{2.10}$$

This sequence begins $2, 1, 3, 4, 7, 11, 18, 29, \cdots$.

Example 2.28. *Many of the recurrence relations we encountered in previous sections are not covered by the above definition. For instance, the Catalan recurrence relation* $c_n = c_0c_{n-1} + \cdots + c_{n-1}c_0$ *is ruled out, because it involves products of two terms of the sequence; the recurrence relation in Example 2.23,* $a_n = \frac{1}{n}(a_{n-1} + 2a_{n-2} + 3a_{n-3} + \cdots + na_0)$*, is ruled out because the coefficients, and the number of terms, depend on* n*.*

The main reason that linear recurrence relations are nicer is that if you have two solutions, a linear combination of them is also a solution:

Theorem 2.29. If $a_0, a_1, a_2, \cdots$ and $b_0, b_1, b_2, \cdots$ are two sequences satisfying the same kth-order homogeneous linear recurrence relation:

$$\begin{aligned} a_n &= r_1a_{n-1} + r_2a_{n-2} + \cdots + r_ka_{n-k}, \text{ for all } n \geq k, \text{ and} \\ b_n &= r_1b_{n-1} + r_2b_{n-2} + \cdots + r_kb_{n-k}, \text{ for all } n \geq k, \end{aligned}$$

then for any constants C_1 and C_2, the sequence

$$C_1a_0 + C_2b_0, C_1a_1 + C_2b_1, C_1a_2 + C_2b_2, \cdots$$

also satisfies this recurrence relation.

Proof. This is trivial: taking C_1 times the first equation plus C_2 times the second equation immediately gives us

$$\begin{aligned} C_1a_n + C_2b_n = r_1(C_1a_{n-1} + C_2b_{n-1}) + r_2(C_1a_{n-2} + C_2b_{n-2}) \\ + \cdots + r_k(C_1a_{n-k} + C_2b_{n-k}), \text{ for all } n \geq k, \end{aligned}$$

which is the desired recurrence relation. □

Exactly the same principle applies to a linear combination of three or more solutions.

Remark 2.30*. *In the terminology of linear algebra, what we have actually shown here is that the set of all sequences which satisfy a given homogeneous linear recurrence relation forms a vector space (in fact, a vector subspace of the vector space of all sequences).*

Thanks to Theorem 2.29, we can hope to build up a general solution by finding special solutions and taking linear combinations of them. Inspired

by the first-order case, we look for special solutions of the form $a_n = \lambda^n$ and slight modifications thereof, and the next result tells us which λ to take.

Definition 2.31. The characteristic polynomial of the recurrence relation $a_n = r_1 a_{n-1} + r_2 a_{n-2} + \cdots + r_k a_{n-k}$ is

$$x^k - r_1 x^{k-1} - r_2 x^{k-2} - \cdots - r_{k-1} x - r_k.$$

This is a polynomial of degree k in the indeterminate x.

Theorem 2.32. Let λ be a (possibly complex) root of the characteristic polynomial.

(1) The sequence $a_n = \lambda^n$ satisfies the recurrence relation

$$a_n = r_1 a_{n-1} + r_2 a_{n-2} + \cdots + r_k a_{n-k}, \text{ for all } n \geq k.$$

(2) If λ is a repeated root of the characteristic polynomial of multiplicity m, then $a_n = n^s \lambda^n$ also satisfies this recurrence relation for $1 \leq s \leq m-1$.

Proof*. The fact that λ is a root of the characteristic polynomial means that

$$\lambda^k - r_1 \lambda^{k-1} - r_2 \lambda^{k-2} - \cdots - r_{k-1} \lambda - r_k = 0. \tag{2.11}$$

Moving the terms with the $-$ sign to the other side and multiplying both sides by λ^{n-k}, we find that

$$\lambda^n = r_1 \lambda^{n-1} + r_2 \lambda^{n-2} + \cdots + r_k \lambda^{n-k}, \text{ for all } n \geq k, \tag{2.12}$$

which is exactly what part (1) claims. The proof of part (2) is slightly harder. We need to show that λ is a root of the polynomial

$$n^s x^n - r_1 (n-1)^s x^{n-1} - r_2 (n-2)^s x^{n-2} - \cdots - r_k (n-k)^s x^{n-k}, \tag{2.13}$$

for all $n \geq k$ and $1 \leq s \leq m-1$. What we know is that $(x-\lambda)^m$ divides the characteristic polynomial, and hence also divides

$$x^n - r_1 x^{n-1} - r_2 x^{n-2} - \cdots - r_k x^{n-k}. \tag{2.14}$$

But the polynomial (2.13) can be obtained from (2.14) by applying, s times, the operation of differentiating and then multiplying by x. Each time we apply this operation can only reduce by 1 the power of $x - \lambda$ which divides the polynomial, so at the end it remains divisible by $(x-\lambda)^{m-s}$, and therefore λ is a root of (2.13). □

Example 2.33. *We can use Theorems 2.29 and 2.32 to solve the following second-order recurrence relation:*

$$a_0 = 2,\ a_1 = 5,\ a_n = 7a_{n-1} - 12a_{n-2},\ \text{for } n \geq 2.$$

The characteristic polynomial is $x^2-7x+12$*, which factorizes as* $(x-3)(x-4)$*; so its roots are* 3 *and* 4*. By Theorem 2.32, both the sequence* 3^n *and the sequence* 4^n *satisfy our recurrence relation (although, clearly, neither of them satisfies our initial conditions). So by Theorem 2.29, any sequence with formula* $C_1 3^n + C_2 4^n$*, where* C_1 *and* C_2 *are constants (i.e. independent of* n*), also satisfies this recurrence relation. It so happens that we can find constants* C_1 *and* C_2 *such that the initial conditions hold also. For this we need to have the following:*

$$C_1 3^0 + C_2 4^0 = 2, \text{ i.e. } C_1 + C_2 = 2, \text{ and}$$
$$C_1 3^1 + C_2 4^1 = 5, \text{ i.e. } 3C_1 + 4C_2 = 5.$$

This is a system of two linear equations in the two unknowns C_1 *and* C_2*, and it is straightforward to find the unique solution:* $C_1 = 3$*,* $C_2 = -1$*. (For example, the second equation minus three times the first equation tells us that* $C_2 = -1$*, and then* $C_1 = 3$ *follows by substituting this in the first equation.) Since* $3 \times 3^n + (-1) \times 4^n$ *satisfies the same recurrence relation and initial conditions as* a_n*, we must have*

$$a_n = 3 \times 3^n + (-1) \times 4^n = 3^{n+1} - 4^n.$$

So we have solved the original recurrence relation.

Example 2.34. *Consider the sequence defined by*

$$a_0 = 1,\ a_1 = 4,\ a_n = 4a_{n-1} - 4a_{n-2} \text{ for } n \geq 2.$$

The characteristic polynomial is $\square$ *which has repeated root* $\square$*.*

By Theorems 2.32 and 2.29, any sequence a_n *which the form* $a_n = \square$ *for constants* C_1 *and* C_2 *must satisfy* $a_n = 4a_{n-1} - 4a_{n-2}$*. To find constants which make the initial conditions satisfied also, we rewrite* $a_0 = 1$ *and* $a_1 = 4$ *as linear equations in* C_1 *and* C_2*:*

$\square$

The unique solution of this system of linear equations is $C_1 =$ $\boxed{}$, $C_2 =$ $\boxed{}$. *Hence the solution of the original recurrence relation is*

$$a_n = \boxed{}$$

Example 2.35*. *Consider the sequence defined by*

$$a_0 = 2,\ a_1 = 2,\ a_n = 2a_{n-1} - 2a_{n-2} \text{ for } n \geq 2.$$

The characteristic polynomial is $x^2 - 2x + 2$, *whose roots are complex:* $1 \pm i$. *Thus both* $(1+i)^n$ *and* $(1-i)^n$ *satisfy the recurrence relation, and hence so does* $C_1(1+i)^n + C_2(1-i)^n$ *for any constants* C_1 *and* C_2. *The latter satisfies the initial conditions if and only if*

$$\begin{aligned} C_1 + C_2 &= 2, \\ (1+i)C_1 + (1-i)C_2 &= 2, \end{aligned}$$

which has unique solution $C_1 = C_2 = 1$. *So we have* $a_n = (1+i)^n + (1-i)^n$. *We can rewrite this in a slightly better way, which removes the need for complex numbers, if we recall the polar forms* $1 \pm i = \sqrt{2}\, e^{\pm i\frac{\pi}{4}}$:

$$a_n = (1+i)^n + (1-i)^n = \sqrt{2}^{\,n}(e^{i\frac{n\pi}{4}} + e^{-i\frac{n\pi}{4}}) = 2\sqrt{2}^{\,n}\cos(\frac{n\pi}{4}).$$

Thus the sequence a_n *is almost periodic with period* 8; *whenever the argument of* cos *recurs, the value gets multiplied by* $\sqrt{2}^{\,8} = 16$. *If we had worked out the first few terms of the sequence, the pattern would have emerged:*

$$2, 2, 0, -4, -8, -8, 0, 16, 32, 32, 0, -64, -128, -128, \cdots.$$

In all of these examples we were able to find a linear combination of the solutions provided by Theorem 2.32 which satisfied the right initial conditions to be the sequence we were looking for. This was no coincidence, because it always happens:

Theorem 2.36. The general solution of the kth-order homogeneous linear recurrence relation

$$a_n = r_1 a_{n-1} + r_2 a_{n-2} + \cdots + r_k a_{n-k}, \text{ for all } n \geq k,$$

is a linear combination of the k solutions arising from the roots of the characteristic polynomials via parts (1) and (2) of Theorem 2.32. In particular, the general solution of the second order recurrence relation

$$a_n = ra_{n-1} + sa_{n-2}, \text{ for all } n \geq 2,$$

is as follows:

(1) if $x^2 - rx - s = (x - \lambda_1)(x - \lambda_2)$ where $\lambda_1 \neq \lambda_2$ (distinct roots),

$$a_n = C_1\lambda_1^n + C_2\lambda_2^n \text{ for some constants } C_1 \text{ and } C_2;$$

(2) if $x^2 - rx - s = (x - \lambda)^2$ (repeated root),

$$a_n = C_1\lambda^n + C_2 n\lambda^n \text{ for some constants } C_1 \text{ and } C_2.$$

Proof.** First note that we do indeed always get k solutions to the recurrence relation out of parts (1) and (2) of Theorem 2.32: this is because every root λ gives rise to the m solutions $n^s\lambda^n$, $0 \leq s \leq m-1$, where m is the multiplicity (which is 1 if λ is not a repeated root), and the sum of the multiplicities of all the roots equals the degree of the polynomial, which is k. (This relies on the fact that we are allowing the roots to be complex numbers, so the polynomial factorizes completely.)

What remains to be proved is that no matter what initial conditions $a_0 = I_0, a_1 = I_1, \cdots, a_{k-1} = I_{k-1}$ we impose, there is always a linear combination of these k solutions which satisfies them. As in the examples, this amounts to showing that there is a solution to a system of k linear equations in the k unknown coefficients. By a basic theorem in linear algebra, such a system has a solution if and only if a certain determinant is nonzero. Proving this in general requires some knowledge of Vandermonde determinants, which would take us too far off the main topic, so we will just go through the proof in the case when $k = 2$.

In this case the initial conditions are $a_0 = I_0$ and $a_1 = I_1$, the recurrence relation is $a_n = ra_{n-1} + sa_{n-2}$, and the characteristic polynomial is the quadratic $x^2 - rx - s$. First suppose that this has distinct roots λ_1 and λ_2;

then our linear combination is $C_1\lambda_1^n + C_2\lambda_2^n$, and we are trying to show that there is always a choice of C_1 and C_2 which satisfies the initial conditions. The system of linear equations this amounts to is:

$$\begin{aligned} C_1 + C_2 &= I_0, \\ \lambda_1 C_1 + \lambda_2 C_2 &= I_1. \end{aligned}$$

Because the determinant of the matrix $\begin{pmatrix} 1 & 1 \\ \lambda_1 & \lambda_2 \end{pmatrix}$ is $\lambda_2 - \lambda_1$, which we are assuming is nonzero, there is a unique solution. In fact, that solution is:

$$C_1 = \frac{\lambda_2 I_0 - I_1}{\lambda_2 - \lambda_1}, \quad C_2 = \frac{I_1 - \lambda_1 I_0}{\lambda_2 - \lambda_1}. \tag{2.15}$$

(It's probably not worth memorizing these formulas, given how easy it is to solve any particular 2×2 linear system.) Now suppose that the characteristic polynomial has a repeated root, i.e. $x^2 - rx - s = (x - \lambda)^2$. Notice that we have $r = 2\lambda$, $s = \lambda^2$. Our linear combination is $C_1\lambda^n + C_2 n\lambda^n$, and the equations we need to solve are

$$\begin{aligned} C_1 &= I_0, \\ \lambda C_1 + \lambda C_2 &= I_1. \end{aligned}$$

Because the determinant of the matrix $\begin{pmatrix} 1 & 0 \\ \lambda & \lambda \end{pmatrix}$ is λ, and λ must be nonzero because $s \neq 0$ (or else we wouldn't be dealing with a second-order recurrence relation at all), there is a unique solution, namely:

$$C_1 = I_0, \quad C_2 = \frac{I_1 - \lambda I_0}{\lambda}. \tag{2.16}$$

So in either case the sequence a_n must be expressible in the right form. □

Remark 2.37. *Notice that the first-order case explained in Example 2.26 is also covered by Theorem 2.36: in this case the characteristic polynomial is the degree-1 polynomial $x - r$, whose sole root is r.*

Example 2.38. *We can now explain the strange formula for the Fibonacci numbers that we proved in Example 2.24. The characteristic polynomial for the Fibonacci recurrence is $x^2 - x - 1$, which has distinct roots $\frac{1+\sqrt{5}}{2}$ and*

$\frac{1-\sqrt{5}}{2}$. *By Theorem 2.36, this is enough to prove that*

$$F_n = C_1 \left(\frac{1+\sqrt{5}}{2}\right)^n + C_2 \left(\frac{1-\sqrt{5}}{2}\right)^n, \text{ for all } n \in \mathbb{N},$$

for some constants C_1 *and* C_2. *The initial conditions* $F_0 = 0$ *and* $F_1 = 1$ *give*

$$C_1 + C_2 = 0,$$
$$\left(\frac{1+\sqrt{5}}{2}\right) C_1 + \left(\frac{1-\sqrt{5}}{2}\right) C_2 = 1,$$

a system whose solution is $C_1 = \frac{1}{\sqrt{5}}$, $C_2 = -\frac{1}{\sqrt{5}}$. *Hence we derive* (2.9).

Example 2.39. *Recall from* (2.10) *that the Lucas numbers satisfy the same recurrence relation as the Fibonacci numbers, but different initial conditions. So they must also be of the same form:*

$$L_n = C_1 \left(\frac{1+\sqrt{5}}{2}\right)^n + C_2 \left(\frac{1-\sqrt{5}}{2}\right)^n, \text{ for all } n \in \mathbb{N},$$

for some constants C_1 *and* C_2. *The initial condition* $L_0 = 2$ *and* $L_1 = 1$ *give*

$$C_1 + C_2 = 2,$$
$$\left(\frac{1+\sqrt{5}}{2}\right) C_1 + \left(\frac{1-\sqrt{5}}{2}\right) C_2 = 1,$$

a system whose solution is $C_1 = C_2 = 1$. *Hence we get a formula for the Lucas numbers:*

$$L_n = \left(\frac{1+\sqrt{5}}{2}\right)^n + \left(\frac{1-\sqrt{5}}{2}\right)^n. \tag{2.17}$$

As with the Fibonacci formula, if you just saw the right-hand side, it would be quite surprising that it always worked out to be a positive integer.

Example 2.40. *Consider the 'stupid' second-order recurrence relation*

$$a_0 = I_0,\ a_1 = I_1,\ a_n = a_{n-2} \text{ for } n \geq 2,$$

where I_0 *and* I_1 *are some initial values. Obviously the solution is that* a_n *is either* I_0 *or* I_1 *depending on whether* n *is even or odd; how does this fit*

the framework of Theorem 2.36? The characteristic polynomial is $x^2 - 1 = (x-1)(x+1)$*, so the general solution is* $a_n = C_1 1^n + C_2(-1)^n = C_1 + C_2(-1)^n$*. To fit the initial conditions we must have* $C_1 = \frac{I_0+I_1}{2}$ *and* $C_2 = \frac{I_0-I_1}{2}$*, so*

$$a_n = \frac{I_0 + I_1}{2} + (-1)^n \frac{I_0 - I_1}{2} = \begin{cases} I_0, & \text{if } n \text{ is even,} \\ I_1, & \text{if } n \text{ is odd,} \end{cases}$$

as expected.

2.4 Non-homogeneous linear recurrence relations

Definition 2.41. A kth-order non-homogeneous linear recurrence relation for a sequence a_n is one of the form

$$a_n = r_1 a_{n-1} + r_2 a_{n-2} + \cdots + r_k a_{n-k} + f(n), \text{ for all } n \geq k,$$

where $r_1, \cdots, r_k$ are fixed numbers (independent of n), $r_k \neq 0$, and $f(n)$ is some nonzero function of n.

What makes this "non-homogeneous" is the extra term $f(n)$, which is not of the same kind as all the others.

Example 2.42. *Recall the recurrence relation from the tower of Hanoi problem in the Introduction:*

$$h_n = 2h_{n-1} + 1, \text{ for all } n \geq 1.$$

This can be viewed as a first-order non-homogeneous recurrence relation, with the 1 *being the extra term.*

Example 2.43. *Suppose we want to find a closed formula for* a_n*, defined recursively by*

$$a_0 = 3,\ a_1 = 5,\ a_n = 2a_{n-1} + 3a_{n-2} + 8n - 4 \text{ for } n \geq 2.$$

The recurrence relation here is non-homogeneous because of the extra term $8n - 4$*. So we can't directly apply the characteristic-polynomial method.*

Example 2.44*. *A general first-order non-homogeneous recurrence relation has the form*

$$a_n = ra_{n-1} + f(n), \text{ for all } n \geq 1.$$

Unravelling this, we obtain

$$\begin{aligned} a_n &= ra_{n-1} + f(n) \\ &= r^2 a_{n-2} + rf(n-1) + f(n) \\ &= r^3 a_{n-3} + r^2 f(n-2) + rf(n-1) + f(n) \\ &\quad \vdots \\ &= r^n a_0 + r^{n-1} f(1) + r^{n-2} f(2) + \cdots + rf(n-1) + f(n), \end{aligned}$$

so solving the recurrence relation amounts to finding a closed formula for the sum $\sum_{m=1}^{n} r^{n-m} f(m)$. *Currently we only know how to do this in special cases such as when* $f(n)$ *is a constant function, where we can use the formula* (2.7) *for the sum of a geometric progression.*

The key to tackling the non-homogeneous case is a very simple observation:

Theorem 2.45. If $a_0, a_1, a_2, \cdots$ and $p_0, p_1, p_2, \cdots$ are two sequences satisfying the same kth-order non-homogeneous linear recurrence relation:

$$\begin{aligned} a_n &= r_1 a_{n-1} + r_2 a_{n-2} + \cdots + r_k a_{n-k} + f(n), \text{ for all } n \geq k, \text{ and} \\ p_n &= r_1 p_{n-1} + r_2 p_{n-2} + \cdots + r_k p_{n-k} + f(n), \text{ for all } n \geq k, \end{aligned}$$

then the sequence $a_n - p_n$ satisfies the corresponding homogeneous recurrence relation (with the extra term omitted). So the general solution of the non-homogeneous relation has the form

$$a_n \;=\; b_n + p_n,$$

where b_n is an arbitrary solution of the homogeneous recurrence relation and p_n is one particular solution of the non-homogeneous recurrence relation.

Proof. Subtracting the two equations, we get

$$a_n - p_n = r_1(a_{n-1} - p_{n-1}) + r_2(a_{n-2} - p_{n-2}) + \cdots + r_k(a_{n-k} - p_{n-k}),$$

because the $f(n)$ term cancels out. This shows that $a_n - p_n$ is a solution of the homogeneous recurrence relation; letting it be b_n, we have $a_n = b_n + p_n$

as claimed. Conversely, given any solution b_n of the homogeneous recurrence relation, the sequence a_n defined by $a_n = b_n + p_n$ does satisfy the non-homogeneous recurrence relation, because p_n does and the subtraction of the two equations is true. □

Remark 2.46*. *This theorem implies that the set of all solutions of the non-homogeneous recurrence relation forms an affine subspace of the vector space of all sequences (a vector subspace translated so that it doesn't pass through the origin).*

In the parallel theory of differential equations, b_n is called the 'complementary function'. We know the general form of b_n from Theorem 2.36, so we can substitute that in the equation $a_n = b_n + p_n$ to get the general form of a solution of the non-homogeneous recurrence relation. Then we can once again use the initial conditions to determine the undetermined constants.

The problem, of course, is that we still have to find one particular solution p_n. Indeed, we may seem not to have made any progress: in trying to find the solution a_n of the non-homogeneous recurrence relation, we are reduced to finding a solution p_n of the non-homogeneous recurrence relation. But the point is that p_n doesn't have to satisfy the initial conditions that a_n does, which gives us more scope. Indeed, p_n can often be found by something approaching guesswork. If the extra term $f(n)$ in the recurrence is a polynomial in n of degree d, it is a good idea to try for a solution p_n which is also a polynomial in n of degree d; let the coefficients be unknowns, and see what equations they need to satisfy.

Example 2.47. *Return to the sequence from Example 2.43:*

$$a_0 = 3,\ a_1 = 5,\ a_n = 2a_{n-1} + 3a_{n-2} + 8n - 4 \text{ for } n \geq 2.$$

To find a closed formula for a_n, we first have to find the general solution of the recurrence relation $a_n = 2a_{n-1} + 3a_{n-2} + 8n - 4$. So we forget about the initial conditions $a_0 = 3$ and $a_1 = 5$ for now.

According to Theorem 2.45, $a_n = b_n + p_n$ where b_n is a solution of the homogeneous recurrence relation $b_n = 2b_{n-1} + 3b_{n-2}$ and p_n is a particular solution of the non-homogeneous recurrence relation. Since the characteristic polynomial $x^2 - 2x - 3 = (x - 3)(x + 1)$, the general form of b_n is $C_1 3^n + C_2(-1)^n$, where C_1 and C_2 are constants.

Since the extra term $8n - 4$ is a degree-1 polynomial in n, we try for a particular solution of the form $p_n = C_3 n + C_4$, where C_3 and C_4 are two more constants. For this to satisfy the non-homogeneous recurrence relation, we need to have the following equation for all $n \geq 2$:

$$\begin{aligned} C_3 n + C_4 &= 2(C_3(n-1) + C_4) + 3(C_3(n-2) + C_4) + 8n - 4 \\ &= (2C_3 + 3C_3 + 8)n + (-2C_3 + 2C_4 - 6C_3 + 3C_4 - 4) \\ &= (5C_3 + 8)n + (-8C_3 + 5C_4 - 4). \end{aligned}$$

For this to hold for all $n \geq 2$, it has to be true that the coefficients of $n^1 = n$ and $n^0 = 1$ on both sides are the same. So we want C_3 and C_4 to satisfy $C_3 = 5C_3 + 8$ and $C_4 = -8C_3 + 5C_4 - 4$. The unique solution is $C_3 = -2$ and $C_4 = -3$, so we have found our particular solution: $p_n = -2n - 3$.

Putting all this together, we have found the general solution of the non-homogeneous recurrence relation $a_n = 2a_{n-1} + 3a_{n-2} + 8n - 4$:

$$a_n = C_1 3^n + C_2(-1)^n - 2n - 3, \text{ for some constants } C_1, C_2. \tag{2.18}$$

Only now, at the end, do we remember the initial conditions we want, $a_0 = 3$ and $a_1 = 5$. These are equivalent to the following equations for C_1 and C_2:

$$\begin{aligned} C_1 + C_2 - 0 - 3 = 3, \text{ i.e. } C_1 + C_2 = 6, \text{ and} \\ 3C_1 - C_2 - 2 - 3 = 5, \text{ i.e. } 3C_1 - C_2 = 10. \end{aligned}$$

Again this is a system of linear equations with a unique solution, namely $C_1 = 4$ and $C_2 = 2$. (In general, the left-hand sides are the same as if we were solving the homogeneous recurrence relation, so there will always be a unique solution, for the same reason as in Theorem 2.36.) We have finally found our closed formula for the original sequence:

$$a_n = 4 \times 3^n + 2 \times (-1)^n - 2n - 3. \tag{2.19}$$

Example 2.48. *We can now find the general solution of the tower of Hanoi recurrence relation $a_n = 2a_{n-1} + 1$. The corresponding homogeneous recurrence relation $b_n = 2b_{n-1}$ has general solution $b_n = \boxed{}$, where C is a constant. Since the extra term in the non-homogenous recurrence is also a constant (if you like, a polynomial in n of degree 0), we look for a solution $p_n = C'$ where C' is a constant. The equation we need C' to satisfy*

is $\boxed{}$ *which means that* $C' = \boxed{}$. *So by Theorem 2.45, the general solution of* $a_n = 2a_{n-1} + 1$ *is*

$$a_n = \boxed{} \text{ for some constant } C.$$

The constant C *is determined by the initial condition, i.e. the value of* a_0. *In the tower of Hanoi example we had* $h_0 = 0$, *so* $h_n = 2^n - 1$ *as seen before.*

Example 2.49*. *Recall from Example 1.82 that we have a closed formula for the elements of the* $k = 2$ *column of the Stirling triangle:* $S(n, 2) = 2^{n-1} - 1$ *for all* $n \geq 1$. *So Theorem 1.81 gives us a recurrence relation for the elements of the* $k = 3$ *column:*

$$S(1, 3) = 0,\ S(n, 3) = 3S(n - 1, 3) + 2^{n-2} - 1 \text{ for } n \geq 2.$$

This is a first-order non-homogeneous recurrence. According to Theorem 2.45, the general solution of $a_n = 3a_{n-1} + 2^{n-2} - 1$ *has the form* $a_n = b_n + p_n$, *where* b_n *satisfies the first-order homogeneous relation* $b_n = 3b_{n-1}$, *and* p_n *is a particular solution of the recurrence relation* $p_n = 3p_{n-1} + 2^{n-2} - 1$. *We have* $b_n = C3^n$ *for some constant* C, *so we just need to find* p_n. *This time the extra term is not a polynomial in* n, *but applying the same principle of trying things of the same form as the extra term, we look for a solution of the form* $p_n = C_1 2^n + C_2$ *where* C_1 *and* C_2 *are further constants. We need to have the following equation for all* $n \geq 2$:

$$\begin{aligned} C_1 2^n + C_2 &= 3(C_1 2^{n-1} + C_2) + 2^{n-2} - 1 \\ &= \left(\frac{3C_1}{2} + \frac{1}{4}\right) 2^n + (3C_2 - 1). \end{aligned}$$

For this to hold for all $n \geq 2$, *it has to be true that* $C_1 = \frac{3C_1}{2} + \frac{1}{4}$ *and* $C_2 = 3C_2 - 1$. *Hence* $C_1 = -\frac{1}{2}$ *and* $C_2 = \frac{1}{2}$, *and our particular solution is* $\frac{-2^n+1}{2}$, *which gives us the general solution:*

$$a_n = C3^n + \frac{-2^n + 1}{2},\ \text{for some constant } C. \tag{2.20}$$

To achieve the initial condition $S(1, 3) = 0$ *we need to have* $3C - \frac{1}{2} = 0$, *so* $C = \frac{1}{6}$. *This gives our desired formula for the* $k = 3$ *Stirling numbers:*

$$S(n, 3) = \frac{3^{n-1} - 2^n + 1}{2},\ \text{for all } n \geq 1. \tag{2.21}$$

Unfortunately, this is not really anything new: it is the same as what you get by combining Theorem 1.84 with the Inclusion/Exclusion formula for the number of surjective functions, Theorem 1.73.

Example 2.50.** *To show that finding a particular solution can be a tricky business, consider the sequence defined by*

$$a_0 = a_1 = 1,\ a_n = 4a_{n-1} - 4a_{n-2} + 2^n \text{ for } n \geq 2.$$

We saw in Example 2.34 that the general solution of the homogeneous recurrence has the form $b_n = C_1 2^n + C_2 n 2^n$, but how to find a particular solution of $p_n = 4p_{n-1} - 4p_{n-2} + 2^n$? On the pattern of the previous examples, it would be natural to look for one of the form $p_n = C_3 2^n$, but then we would have to have the following equation for all $n \geq 2$:

$$C_3 2^n = 4C_3 2^{n-1} - 4C_3 2^{n-2} + 2^n = (C_3 + 1)2^n,$$

which is clearly impossible. The problem is that a constant multiple of 2^n is already part of b_n, i.e. it satisfies the homogeneous relation, so it can't satisfy the non-homogeneous relation as well. The same objection applies to a constant multiple of $n2^n$. How about if we try $p_n = C_3 n^2 2^n$? Then we need to have, for all $n \geq 2$:

$$C_3 n^2 2^n = 4C_3(n-1)^2 2^{n-1} - 4C_3(n-2)^2 2^{n-2} + 2^n,$$

which (dividing both sides by 2^n) is equivalent to

$$C_3 n^2 = 2C_3(n-1)^2 - C_3(n-2)^2 + 1 = C_3 n^2 - 2C_3 + 1.$$

This is indeed satisfied when $C_3 = \frac{1}{2}$, so we get

$$a_n = C_1 2^n + C_2 n 2^n + \frac{1}{2} n^2 2^n, \text{ for some constants } C_1, C_2. \tag{2.22}$$

The initial condition $a_0 = 1$ forces $C_1 = 1$, and the initial condition $a_1 = 1$ forces $C_2 = -1$. So the solution of the original recurrence is

$$a_n = (1 - n + \frac{1}{2} n^2)2^n. \tag{2.23}$$

We will re-derive this solution with less guesswork in Example 3.42.

We have the following general rule to find particular solutions in the case where the sequence $f(n)$ is *quasi-polynomial*; that is, $f(n) = q(n)\mu^n$ for some polynomial $q(n)$ and some constant μ. This constant is called the *exponent* of the quasi-polynomial. Note that if $\mu = 1$ then $f(n)$ is a genuine polynomial. Consider the kth-order non-homogeneous linear recurrence relation

$$a_n = r_1 a_{n-1} + r_2 a_{n-2} + \cdots + r_k a_{n-k} + q(n)\mu^n, \text{ for all } n \geq k.$$

Theorem 2.51. Suppose that $q(n)$ is a polynomial of degree d and suppose that μ is a root of the characteristic polynomial of multiplicity l. Then a particular solution of the non-homogeneous recurrence relation can be found in the form $p_n = Q(n)n^l\mu^n$, where $Q(n)$ is a polynomial of degree d.

Note that "μ is a root of multiplicity 0" just means that μ is *not* a root of the characteristic polynomial.

By the same argument as for the proof of Theorem 2.45 one verifies that if $f(n)$ is a sum of quasi-polynomials then a particular solution of the non-homogeneous linear recurrence relation can be found as the corresponding sum of particular solutions provided by Theorem 2.51.

Returning to Example 2.50, note that $f(n) = 2^n$ can be regarded as a quasi-polynomial with $q(n) = 1$ and the exponent $\mu = 2$. The exponent is a root of the characteristic polynomial $x^2 - 4x + 4$ of multiplicity 2 and so, by Theorem 2.51, a particular solution can indeed be found in the form $p_n = Cn^2 2^n$ as we saw in the example.

Chapter 3

Generating Functions

Generating functions are a powerful device for capturing a whole sequence in a single entity. Manipulations of generating functions can be used to solve many types of recurrence relations, including the linear recurrence relations we examined in the previous chapter.

Definition 3.1. The generating function of the sequence $a_0, a_1, a_2, a_3, \cdots$ is the formal power series

$$a_0 + a_1 z + a_2 z^2 + a_3 z^3 + \cdots$$

in the indeterminate z. This is often denoted $A(z)$, where in general A should be replaced by (the capital form of) whatever letter denotes the terms of the sequence.

Example 3.2. *The sequence* $1, 2, 4, 8, \cdots$ *has generating function*

$$1 + 2z + 4z^2 + 8z^3 + \cdots .$$

The sequence $0, 1, 2, 3, \cdots$ *has generating function*

$$z + 2z^2 + 3z^3 + \cdots$$

(notice that we leave off the $0+$ *at the beginning, and write simply* z *instead of* $1z$*). The Fibonacci sequence and Catalan sequence have generating functions*

$$\begin{aligned} F(z) &= z + z^2 + 2z^3 + 3z^4 + 5z^5 + 8z^6 + \cdots , \\ C(z) &= 1 + z + 2z^2 + 5z^3 + 14z^4 + \cdots . \end{aligned}$$

3.1 Formal power series

The first question you should ask about the definition of generating function is: what is a "formal power series"? This is a very useful idea which generalizes the concept of a polynomial.

You are familiar with the terminology that something like $2+z+4z^2$ is called a polynomial in the indeterminate (or variable) z. This means that it is a sum of terms, each of which consists of a different power of z multiplied by a coefficient which is just a number. (The constant term is the term in which the power of z is z^0, and we usually don't write the z^0, just the coefficient.) It is an important part of the concept that there are only finitely many terms – although you could imagine, if you like, that all the other powers of z are present but not visible because their coefficient is 0. As you know, there are natural ways to define addition, subtraction, multiplication, and even long division of polynomials. The other important thing you can do with a polynomial in the indeterminate z is to substitute a number for z: e.g. if you substitute $z = 3$ in $2 + z + 4z^2$, you get the result $2 + 3 + 4 \times 3^2 = 41$. This means that you can think of polynomials as just a special kind of function.

A formal power series is a 'polynomial with infinitely many terms'. That is, it is something like $1+2z+4z^2+8z^3+\cdots$, in which there is a term for every power of z. It is allowed for some of the coefficients to be 0, however; indeed, a polynomial like $2+z+4z^2$ can be viewed as a formal power series in which the coefficients of $z^3, z^4, \cdots$ happen to be 0, and even a number like 2 can be viewed as a formal power series in which the coefficients of all the positive powers of z happen to be 0. But having, in general, infinitely many terms makes the crucial difference that you can't just substitute any number for z any more. For example, if you tried to substitute $z = 3$ in $1+2z+4z^2+8z^3+\cdots$, you would get the series (infinite sum) $1 + 2 \times 3 + 4 \times 3^2 + 8 \times 3^3 + \cdots$, which makes no sense.

Remark 3.3. *Of course, this statement is too strong. Power series such as this arise in many parts of mathematics – you would have seen Taylor series in calculus, for example – and one of the main aims of analysis is to make sense of them. But even in the context of analysis,* $1 + 2 \times 3 + 4 \times 3^2 + 8 \times 3^3 + \cdots$ *is a divergent series, which cannot be given a numerical value. (You could*

say it equalled ∞, but that is really just a shorthand way of expressing the divergence, not a true evaluation: whatever ∞ is, it's not a number.) MATH2962 includes some fascinating results about when such a power series is convergent and when it is divergent. These results are most interesting and powerful in the context of the complex numbers, and that is actually the reason we are using the letter z for the indeterminate, rather than, say, x. Somewhere at the back of our minds is the idea that it would be interesting to substitute a complex number for z, although in this course we will not go down that path.

The fact that we can't (or at least won't) substitute a number for z means that we can't think of a formal power series as a function, as we could a polynomial (despite the term "generating function"). This is what is meant by the adjective "formal": $a_0 + a_1 z + a_2 z^2 + a_3 z^3 + \cdots$ is not a function of z, it's just an object in its own right, in which the letter z and its 'powers' play the role of keeping the terms of the sequence in their correct order. In fact, you could think of the generating function $a_0 + a_1 z + a_2 z^2 + a_3 z^3 + \cdots$ as just a different notation for the sequence $a_0, a_1, a_2, a_3, \cdots$, using + signs and powers of z instead of commas. To get back to the sequence, you just have to extract the coefficients of the various powers of z.

Of course, this would be a bizarre choice of notation if there were nothing more to it. The point is that we can add and multiply formal power series, by rules which look completely natural in this notation, if you imagine collecting terms and multiplying powers of z in the obvious way.

Definition 3.4. If we have two formal power series,

$$A(z) = a_0 + a_1 z + a_2 z^2 + a_3 z^3 + \cdots \text{ and } B(z) = b_0 + b_1 z + b_2 z^2 + b_3 z^3 + \cdots ,$$

then

$$\begin{aligned} A(z) + B(z) &:= (a_0 + b_0) + (a_1 + b_1)z + (a_2 + b_2)z^2 + (a_3 + b_3)z^3 + \cdots , \\ A(z)B(z) &:= a_0 b_0 + (a_0 b_1 + a_1 b_0)z + (a_0 b_2 + a_1 b_1 + a_2 b_0)z^2 \\ &\qquad + (a_0 b_3 + a_1 b_2 + a_2 b_1 + a_3 b_0)z^3 + \cdots . \end{aligned}$$

The justification for this multiplication rule is that to get a z^n term in the product, you need to multiply the $a_m z^m$ term of $A(z)$ with the $b_{n-m} z^{n-m}$ term of $B(z)$, for some m. So the coefficient of z^n is $\sum_{m=0}^{n} a_m b_{n-m}$.

Example 3.5. *According to these rules, the sum of*

$F(z) = z + z^2 + 2z^3 + 3z^4 + \cdots$ *and* $C(z) = 1 + z + 2z^2 + 5z^3 + 14z^4 + \cdots$

is just obtained by adding the coefficients term-by-term:

$$F(z) + C(z) = 1 + 2z + 3z^2 + 7z^3 + 17z^4 + \cdots .$$

We obtain each coefficient of the product by adding appropriate products of one coefficient from $F(z)$ and one coefficient from $C(z)$:

$$\begin{aligned}
\textit{coefficient of } z^0 \textit{ in } F(z)C(z) &= \boxed{} \\
\textit{coefficient of } z^1 \textit{ in } F(z)C(z) &= \boxed{} \\
\textit{coefficient of } z^2 \textit{ in } F(z)C(z) &= \boxed{} \\
\textit{coefficient of } z^3 \textit{ in } F(z)C(z) &= \boxed{} \\
\textit{coefficient of } z^4 \textit{ in } F(z)C(z) &= \boxed{} \\
\textit{so } F(z)C(z) &= \boxed{}
\end{aligned}$$

It is often useful to write formal power series in sigma notation as follows.

Definition 3.6. The symbol

$$\sum_{n=0}^{\infty} a_n\, z^n \quad \text{means} \quad a_0 + a_1 z + a_2 z^2 + a_3 z^3 + \cdots .$$

The lower limit of the sum can be varied in the obvious way: for instance, $\sum_{n=3}^{\infty} b_n\, z^n$ means $b_3 z^3 + b_4 z^4 + b_5 z^5 + \cdots$. The summation variable n can of course be replaced with any other letter.

In this notation the rules for addition and multiplication become:

$$\begin{aligned}
\left(\sum_{n=0}^{\infty} a_n\, z^n\right) + \left(\sum_{n=0}^{\infty} b_n\, z^n\right) &= \sum_{n=0}^{\infty} (a_n + b_n)\, z^n, \\
\left(\sum_{n=0}^{\infty} a_n\, z^n\right)\left(\sum_{n=0}^{\infty} b_n\, z^n\right) &= \sum_{n=0}^{\infty} \left(\sum_{m=0}^{n} a_m b_{n-m}\right) z^n.
\end{aligned} \tag{3.1}$$

A special case of the multiplication rule says that multiplying $a_0+a_1z+a_2z^2+\cdots$ by a number b_0 just multiplies every coefficient by that number. So if $A(z)$ is the generating function of the sequence $a_0, a_1, a_2, \cdots$, then $3A(z)$, for instance, is the generating function of the sequence $3a_0, 3a_1, 3a_2, \cdots$.

Another important special case of the multiplication rule is when $B(z) = z^m$ for some m. When you multiply $a_0 + a_1z + a_2z^2 + \cdots$ by z^m, you get $a_0z^m + a_1z^{m-1} + a_2z^{m+2} + \cdots$, as you would expect. Notice that the coefficients of $z^0, z^1, \cdots, z^{m-1}$ in the result are all zero, and the previous coefficients $a_0, a_1, a_2, \cdots$ have all been 'shifted m places to the right', so that the coefficient of z^n is a_{n-m} for all $n \geq m$. Thus in sigma notation, the rule for multiplying by a power of z is

$$z^m(\sum_{n=0}^{\infty} a_n\, z^n) = \sum_{n=m}^{\infty} a_{n-m}\, z^n. \tag{3.2}$$

In terms of sequences, this means that if $A(z)$ is the generating function of $a_0, a_1, a_2, \cdots$, then

$zA(z)$ is the generating function of $0, a_0, a_1, a_2, a_3, \cdots$,
$z^2A(z)$ is the generating function of $0, 0, a_0, a_1, a_2, \cdots$,
$z^3A(z)$ is the generating function of $0, 0, 0, a_0, a_1, \cdots$,
and so on.

Example 3.7. *Let $G(z) = \sum_{n=0}^{\infty} z^n = 1 + z + z^2 + z^3 + \cdots$ be the geometric series, the generating function of the sequence $1, 1, 1, 1, \cdots$. We have*

$$zG(z) = \sum_{n=1}^{\infty} z^n = z + z^2 + z^3 + z^4 + \cdots = G(z) - 1,$$

so $(1-z)G(z) = 1$, which is commonly rewritten as $G(z) = \frac{1}{1-z}$. In summary:

$$1 + z + z^2 + z^3 + \cdots \;=\; \frac{1}{1-z}. \tag{3.3}$$

As you may know, if you substitute for z a complex number α inside the unit circle, the left-hand side of (3.3) becomes a series which converges to the number $\frac{1}{1-\alpha}$. On the other hand, if you substitute for z a complex number α outside the unit circle, the left-hand side becomes a divergent series which has

nothing to do with $\frac{1}{1-\alpha}$*. The equation* (3.3) *is not meant to assert anything about what happens when you substitute an actual number for* z*. It means solely that* $1+z+z^2+z^3+\cdots$ *is a formal power series which has the property that when you multiply it by* $1-z$*, you get* 1*.*

To make sense of fractions of formal power series, we need the following result. (Here 0 means the power series with all coefficients 0.)

Theorem 3.8. (1) If $A(z)B(z) = 0$, then either $A(z)$ or $B(z)$ must be 0.

(2) If $A(z)F(z) = A(z)G(z)$, and $A(z) \neq 0$, then $F(z) = G(z)$.

Proof*. We will prove (1) in the 'contrapositive' form, which says that if both $A(z)$ and $B(z)$ are nonzero, then so is $A(z)B(z)$. Since they are nonzero formal power series, they each have a 'starting term'; in other words,

$$\begin{aligned} A(z) &= a_p z^p + a_{p+1} z^{p+1} + a_{p+2} z^{p+2} + \cdots, \\ B(z) &= b_q z^q + a_{q+1} z^{q+1} + a_{q+2} z^{q+2} + \cdots, \end{aligned}$$

where a_p is nonzero and all the earlier coefficients $a_{p'}$ for $p' < p$ are zero, and similarly b_q is nonzero and all the earlier coefficients $b_{q'}$ for $q' < q$ are zero. From the multiplication rule we see that the coefficient of z^{p+q} in $A(z)B(z)$ is $a_p b_q$, because every other term you would expect to be in the coefficient involves either some $a_{p'}$ for $p' < p$ or some $b_{q'}$ for $q' < q$. Since $a_p b_q \neq 0$, this shows that $A(z)B(z)$ has at least one nonzero coefficient, so it is not the zero power series. Part (2) follows by applying part (1) to the equation $A(z)(F(z) - G(z)) = 0$. □

Remark 3.9. *Part (2) of Theorem 3.8 implies that there cannot be two different power series* $F(z)$ *and* $G(z)$ *which satisfy* $(1-z)F(z) = (1-z)G(z) = 1$*, so our use of the notation* $\frac{1}{1-z}$ *for the geometric series is unambiguous. Similarly, from now on, whenever we state that* $F(z) = \frac{B(z)}{A(z)}$ *(a quotient of formal power series), it just means that* $A(z)F(z) = B(z)$*; this notation is possible because there can be at most one formal power series* $F(z)$ *which satisfies this equation. These fractions can be manipulated in the usual way.*

Remark 3.10*. *It is not always the case that such a quotient formal power series exists: there is no formal power series* $F(z)$ *which satisfies* $zF(z) = 1+z$*, for instance. But if* $A(z)$ *has nonzero constant term, then one can find*

an inverse $\frac{1}{A(z)}$*, i.e. a formal power series* $F(z)$ *such that* $A(z)F(z) = 1$*; the coefficients of the powers of* z^n *in* $F(z)$ *can be determined recursively. So in this case,* $\frac{B(z)}{A(z)} = B(z)\frac{1}{A(z)}$ *always makes sense.*

3.2 Manipulating formal power series

In order to use generating functions effectively, we will need more operations on formal power series than just addition and multiplication. Some of the basic results involve an operation of apparently dubious validity, namely differentiating with respect to z.

Example 3.11. *Let us 'differentiate both sides of* (3.3) *with respect to* z*':*

$$1 + 2z + 3z^2 + 4z^3 + \cdots \;=\; \frac{1}{(1-z)^2}.$$

This purports to be a formula for the generating function of the sequence $1, 2, 3, 4, \cdots$*. As explained in Remark 3.9, the interpretation of this is that* $1 + 2z + 3z + 4z^3 + \cdots$ *is a formal power series with the property that when you multiply it by* $(1-z)^2$*, you get* 1*. We can easily prove this rigorously, because from the multiplication rule we get*

$$(1 + z + z^2 + z^3 + \cdots)^2 \;=\; 1 + 2z + 3z^2 + 4z^3 + \cdots, \tag{3.4}$$

from which it follows that

$$(1-z)^2(1+2z+3x^2+4z^3+\cdots) = ((1-z)(1+z+z^2+z^3+\cdots))^2 = 1^2 = 1.$$

So we do indeed have

$$\sum_{n=0}^{\infty}(n+1)\,z^n \;=\; \frac{1}{(1-z)^2}. \tag{3.5}$$

Incidentally, applying the shifting principle (3.2) *to* (3.5)*, we find*

$$\sum_{n=1}^{\infty} n\,z^n \;=\; z + 2z^2 + 3z^3 + \cdots \;=\; \frac{z}{(1-z)^2}, \tag{3.6}$$

which is a formula for the generating function of the sequence $0, 1, 2, 3, \cdots$.

Clearly this dubious differentiation is worth defining properly.

Definition 3.12. If $A(z) = \sum_{n=0}^{\infty} a_n\, z^n = a_0 + a_1 z + a_2 z^2 + a_3 z^3 + \cdots$ is a formal power series, we define its derivative to be the formal power series

$$A'(z) := \sum_{n=0}^{\infty} (n+1) a_{n+1}\, z^n = a_1 + 2a_2 z + 3a_3 z^2 + 4a_4 z^3 + \cdots .$$

Remark 3.13. *It would also be correct to write $A'(z) = \sum_{n=1}^{\infty} n a_n\, z^{n-1}$, which is what you would expect to get from differentiating $\sum_{n=0}^{\infty} a_n\, z^n$ term-by-term (killing the constant term a_0). In Definition 3.12 we have shifted the summation variable so that it equals the power of z, which gives a more convenient form.*

We can also find 'integrals' of formal power series:

Theorem 3.14. For any formal power series $A(z)$, there exists a formal power series $B(z)$ such that $B'(z) = A(z)$. Any other solution of this equation differs from $B(z)$ only in the constant term; there is a unique solution which has zero constant term.

Proof. If $A(z) = \sum_{n=0}^{\infty} a_n\, z^n$ and $B(z) = \sum_{n=0}^{\infty} b_n\, z^n$, then the equation $B'(z) = A(z)$ is equivalent to the statement that $a_n = (n+1) b_{n+1}$ for all $n \geq 0$. So b_n must be $\frac{a_{n-1}}{n}$ for all $n \geq 1$, and the value of b_0 is irrelevant. The result follows. □

We can easily prove some familiar-looking differentiation rules.

Theorem 3.15. If $A(z) = \sum_{n=0}^{\infty} a_n\, z^n$ and $B(z) = \sum_{n=0}^{\infty} b_n\, z^n$, then:

(1) (Sum rule) $(A(z) + B(z))' = A'(z) + B'(z)$.

(2) (Product rule) $(A(z)B(z))' = A'(z)B(z) + A(z)B'(z)$.

(3) (Power rule) $(A(z)^k)' = kA(z)^{k-1}A'(z)$, for all $k \geq 1$.

(4) (Quotient rule) If $B(z)F(z) = A(z)$ for some formal power series $F(z)$, then $B(z)^2 F'(z) = A'(z)B(z) - A(z)B'(z)$. In short,

$$\left(\frac{A(z)}{B(z)}\right)' = \frac{A'(z)B(z) - A(z)B'(z)}{B(z)^2}.$$

Proof*. To prove an equality of formal power series as in (1), it is enough to prove that the coefficient of z^n is the same on both sides, for all $n \geq 0$. By Definition 3.12, the coefficient of z^n in $(A(z)+B(z))'$ is $(n+1)$ times the coefficient of z^{n+1} in $A(z)+B(z)$, i.e. $(n+1)(a_{n+1}+b_{n+1})$. The coefficient of z^n in $A'(z)+B'(z)$ is $(n+1)a_{n+1}+(n+1)b_{n+1}$, so the equality is obvious. The proof of the product rule is slightly harder. The coefficient of z^n in $(A(z)B(z))'$ is $(n+1)\sum_{m=0}^{n+1} a_m b_{n+1-m}$, while in $A'(z)B(z)+A(z)B'(z)$ it is

$$\begin{aligned}
&\sum_{m=0}^{n}(m+1)a_{m+1}b_{n-m} + \sum_{m=0}^{n} a_m(n-m+1)b_{n-m+1} \\
&\qquad = \sum_{m=1}^{n+1} m a_m b_{n+1-m} + \sum_{m=0}^{n}(n-m+1)a_m b_{n+1-m} \\
&\qquad = \sum_{m=0}^{n+1} m a_m b_{n+1-m} + \sum_{m=0}^{n+1}(n-m+1)a_m b_{n+1-m} \\
&\qquad = \sum_{m=0}^{n+1}(n+1)a_m b_{n+1-m},
\end{aligned}$$

as required. Here the second-last step used the common trick of adding an extra term, which works out to be zero, at the top or bottom of a sum. We prove the power rule by induction on k. The $k=1$ case is obvious, so assume that $k \geq 2$ and that the result is known for $k-1$. Then using the product rule, we have

$$\begin{aligned}
(A(z)^k)' &= (A(z)^{k-1}A(z))' \\
&= (A(z)^{k-1})'A(z) + A(z)^{k-1}A'(z) \\
&= (k-1)A(z)^{k-2}A'(z)A(z) + A(z)^{k-1}A'(z) \\
&= kA(z)^{k-1}A'(z),
\end{aligned}$$

completing the induction step. To prove the quotient rule, we differentiate both sides of $B(z)F(z) = A(z)$ and use the product rule on the left-hand side (the same way the quotient rule is proved in calculus, incidentally). This gives

$$B'(z)F(z) + B(z)F'(z) = A'(z),$$

and multiplying both sides by $B(z)$, using $B(z)F(z) = A(z)$, and rearranging gives the desired equation $B(z)^2F'(z) = A'(z)B(z) - A(z)B'(z)$. □

Example 3.16. *Our initial proof of* (3.5) *can now be justified: we obtain it from* (3.3) *by differentiating both sides, and using the quotient rule on the right-hand side. If we differentiate both sides of* (3.5) *again, we obtain:*

$$\sum_{n=0}^{\infty}(n+1)(n+2)\,z^n \;=\; \frac{-2(1-z)(-1)}{(1-z)^4} \;=\; \frac{2}{(1-z)^3}.$$

Dividing both sides by 2*, this says that*

$$\sum_{n=0}^{\infty}\binom{n+2}{2}z^n \;=\; \frac{1}{(1-z)^3}, \tag{3.7}$$

which is a formula for the generating function of the $k=2$ column of Pascal's triangle (i.e. the triangular numbers), starting from $\binom{2}{2}=1$.

In general, the generating function of each column of Pascal's triangle has a very simple formula:

Theorem 3.17. For any $k \geq 0$,

$$\sum_{n=0}^{\infty}\binom{n+k}{k}z^n \;=\; \frac{1}{(1-z)^{k+1}}.$$

Proof*. We proceed by induction on k. The $k=0$ base case is (3.3). Assume that $k \geq 1$, and that we know the result for $k-1$, i.e.

$$\sum_{n=0}^{\infty}\binom{n+k-1}{k-1}z^n \;=\; \frac{1}{(1-z)^{k}}.$$

Taking derivative of both sides (and using the quotient and power rules on the right-hand side), we get

$$\sum_{n=0}^{\infty}(n+1)\binom{n+k}{k-1}z^n \;=\; \frac{0-k(1-z)^{k-1}(-1)}{(1-z)^{2k}} \;=\; \frac{k}{(1-z)^{k+1}}.$$

Since

$$\frac{n+1}{k}\binom{n+k}{k-1} = \frac{(n+1)(n+k)(n+k-1)\cdots(n+2)}{k\,(k-1)!} = \binom{n+k}{k},$$

we obtain the result. □

Remark 3.18*. *Theorem 3.17 is equivalent to saying that for all $s \geq 1$,*

$$(1 + z + z^2 + z^3 + \cdots)^s \;=\; \sum_{n=0}^{\infty} \binom{n+s-1}{n} z^n. \tag{3.8}$$

To get a z^n term on the left-hand side of (3.8), one must select from the various factors the $z^{n_1}, z^{n_2}, \cdots, z^{n_s}$ terms where $n_1, n_2, \cdots, n_s$ are nonnegative integers such that $n_1 + n_2 + \cdots + n_s = n$. So (3.8) is equivalent to the statement that the number of s-tuples $(n_1, n_2, \cdots, n_s) \in \mathbb{N}^s$ such that $n_1 + n_2 + \cdots + n_s = n$ is $\binom{n+s-1}{n}$. Thus Theorem 3.17 is just a disguised form of Theorem 1.60.

Multiplying Theorem 3.17 on both sides by z^k gives a generalization of (3.6):

$$\sum_{n=0}^{\infty} \binom{n}{k} z^n \;=\; \frac{z^k}{(1-z)^{k+1}}. \tag{3.9}$$

(You would normally change the summation so that it started from $n = k$, but in this case it makes no difference because $\binom{n}{k} = 0$ for $n = 0, 1, \cdots, k-1$.) Since we can express n^m as a linear combination of binomial coefficients $\binom{n}{k}$ for various k by Theorem 1.86, we can use (3.9) to get a closed formula for the generating function of the sequence $0^m, 1^m, 2^m, \cdots$ for any m, and hence of any sequence where the nth term is a polynomial function of n.

Example 3.19. *To find a closed formula for $0^2 + 1^2 z + 2^2 z^2 + 3^2 z^3 + \cdots$, we first express n^2 as an integer linear combination of binomial coefficients:*

$$n^2 = \boxed{} \binom{n}{1} + \boxed{} \binom{n}{2}.$$

Hence

$$\sum_{n=0}^{\infty} n^2 z^n = \sum_{n=0}^{\infty} \boxed{} \binom{n}{1} z^n + \sum_{n=0}^{\infty} \boxed{} \binom{n}{2} z^n$$

$$= \boxed{}$$

Another seemingly dubious operation which is often useful is to take a formal power series $A(z)$ and substitute for z another formal power series $F(z)$, expanding out each $F(z)^n$ and collecting terms; the result is denoted $A(F(z))$.

Example 3.20. *The simplest and most commonly encountered example is when $F(z) = cz$ for some complex number c; then we are just changing $A(z) = a_0 + a_1 z + a_2 z^2 + \cdots$ to $A(cz) = a_0 + a_1 cz + a_2 c^2 z^2 + \cdots$, i.e. multiplying the coefficient of z^n by the number c^n, for all n.*

Example 3.21. *A more complicated example is when $F(z) = z + \frac{1}{2}z^2$; then the result of the substitution is*

$$\begin{aligned} &a_0 + a_1(z + \frac{1}{2}z^2) + a_2(z + \frac{1}{2}z^2)^2 + a_3(z + \frac{1}{2}z^2)^3 + \cdots \\ &\quad = a_0 + a_1(z + \frac{1}{2}z^2) + a_2(z^2 + z^3 + \frac{1}{4}z^4) + a_3(z^3 + \frac{3}{2}z^4 + \frac{3}{4}z^5 + \frac{1}{8}z^6) + \cdots \\ &\quad = a_0 + a_1 z + (\frac{1}{2}a_1 + a_2)z^2 + (a_2 + a_3)z^3 + \cdots . \end{aligned}$$

It is important to notice in this last step that there cannot be any other contributions to the coefficients of $1, z, z^2, z^3$ from the "$\cdots$" on the previous line, because that "$\cdots$" involves fourth and higher powers of $(z + \frac{1}{2}z^2)$, which will expand to fourth and higher powers of z, which we are sweeping into the "$\cdots$" on the last line. In general, to work out the coefficient of z^n in $A(z + \frac{1}{2}z^2)$, we only need to know the coefficients of $1, z, z^2, \cdots, z^n$ in $A(z)$.

This finiteness principle continues to work, and therefore the substitution continues to make sense, if $F(z)$ is any formal power series with zero constant term; but not if $F(z)$ has nonzero constant term.

Example 3.22*. *If you try to substitute $1 + z$ for z in $A(z)$, you get the expression $a_0 + a_1(1+z) + a_2(1+2z+z^2) + a_3(1+3z+3z^2+z^3) + \cdots$. This is not a valid formal power series, because each power of z gets contributions from infinitely many terms in the sum.*

Theorem 3.23. If $A(z)$, $B(z)$, and $F(z)$ are formal power series, and $F(z)$ has zero constant term, then:

(1) If $S(z) = A(z) + B(z)$, then $S(F(z)) = A(F(z)) + B(F(z))$.

(2) If $P(z) = A(z)B(z)$, then $P(F(z)) = A(F(z))B(F(z))$.

(3) If $G(z) = \frac{A(z)}{B(z)}$, then $G(F(z)) = \frac{A(F(z))}{B(F(z))}$.

(4) (Chain rule) $A(F(z))' = A'(F(z))F'(z)$.

Proof*. We will only give the proof in the simple case that $F(z) = cz$, since this is enough for many of our later purposes. As seen above, substituting cz for z in a formal power series just multiplies the coefficient of z^n by c^n, for all $n \geq 0$. Part (1) is therefore easy: the coefficient of z^n in $S(F(z))$ is $c^n(a_n + b_n)$, whereas that in the right-hand side is $c^n a_n + c^n b_n$. Similarly, for part (2), the coefficient of z^n in $P(F(z))$ is $c^n \sum_{m=0}^{n} a_m b_{n-m}$, whereas that in the right-hand side is $\sum_{m=0}^{n} (c^m a_m)(c^{n-m} b_{n-m})$, which is clearly the same. Part (3) follows immediately from part (2), if you rearrange the equations so that they don't involve fractions. Part (4) in the special case $F(z) = cz$ says that $A(cz)' = cA'(cz)$; the coefficient of z^n in $A(cz)'$ is $(n+1)c^{n+1}a_{n+1}$, and that in $A'(cz)$ is $c^n(n+1)a_{n+1}$, so this is true. □

Example 3.24. *Applying part (3) in the case $F(z) = cz$ to* (3.3)*, we find*

$$\sum_{n=0}^{\infty} c^n \, z^n \;=\; 1 + cz + c^2 z^2 + c^3 z^3 + \cdots \;=\; \frac{1}{1 - cz}. \tag{3.10}$$

For instance, the generating function of the sequence $1, 2, 4, 8, \cdots$ is $\frac{1}{1-2z}$. Similarly, (3.5) *implies that*

$$\sum_{n=0}^{\infty} (n+1) \, c^n \, z^n \;=\; \frac{1}{(1 - cz)^2}. \tag{3.11}$$

More generally, Theorem 3.17 implies that for all $k \geq 0$,

$$\sum_{n=0}^{\infty} \binom{n+k}{k} \, c^n \, z^n \;=\; \frac{1}{(1 - cz)^{k+1}}. \tag{3.12}$$

At this point we know enough to extract the coefficient of z^n from any formal power series of the form $\frac{F(z)}{G(z)}$ where $F(z)$ and $G(z)$ are polynomials. We may as well assume that the degree of $F(z)$ is less than that of $G(z)$, since long division of polynomials reduces the general case to this. Then the method is to factorize $G(z)$ into powers $(1 - cz)^m$, and rewrite $\frac{F(z)}{G(z)}$ in partial-fractions form, i.e. as a sum of terms of the form $\frac{b}{(1-cz)^m}$ where b and c are constants (i.e. just numbers, not involving z). The coefficient of z^n in each term can then be read off from (3.10), (3.11), or in general (3.12).

Example 3.25. *To find the coefficient of z^n in $\frac{2-9z}{1-7z+12z^2}$, we need to write it in partial-fractions form. The denominator factorizes as $(1-3z)(1-4z)$, so this amounts to finding constants C_1 and C_2 such that*

$$\frac{2-9z}{(1-3z)(1-4z)} = \frac{C_1}{1-3z} + \frac{C_2}{1-4z}.$$

Multiplying both sides by $(1-3z)(1-4z)$, this equation becomes

$$\boxed{}$$

Equating coefficients, we get two linear equations for the two unknowns of which the unique solution is $C_1 =$ $\boxed{}$ and $C_2 =$ $\boxed{}$. So the desired partial-fractions expression is:

$$\frac{2-9z}{1-7z+12z^2} = \boxed{}$$

We can use (3.10) to read off that

$$\text{the coefficient of } z^n \text{ in } \frac{2-9z}{1-7z+12z^2} \text{ is } \boxed{}$$

Example 3.26. *Let us find the coefficient of z^n in $\frac{5-9z}{1-6z+9z^2}$. The denominator factorizes as $(1-3z)^2$; when we have a repeated factor like this, the partial-fractions form should include a term for every power of the factor up to the multiplicity, which in this case means*

$$\frac{5-9z}{(1-3z)^2} = \frac{C_1}{1-3z} + \frac{C_2}{(1-3z)^2}.$$

Clearing the denominators, this becomes

$$5-9z = C_1(1-3z) + C_2 = (C_1+C_2) - 3C_1z,$$

which has the unique solution $C_1 = 3$, $C_2 = 2$. So

$$\frac{5-9z}{1-6z+9z^2} = \frac{3}{1-3z} + \frac{2}{(1-3z)^2}.$$

By (3.10) and (3.11), the coefficient of z^n is $3\times 3^n + 2(n+1)3^n = (2n+5)3^n$.

Example 3.27*. *Let us find the coefficient of z^n in $\frac{2z}{(1-z)(1-2z)^2}$. The partial-fractions form is*

$$\frac{2z}{(1-z)(1-2z)^2} = \frac{C_1}{1-z} + \frac{C_2}{1-2z} + \frac{C_3}{(1-2z)^2}.$$

Clearing the denominators, this becomes

$$\begin{aligned} 2z &= C_1(1-2z)^2 + C_2(1-z)(1-2z) + C_3(1-z) \\ &= (C_1 + C_2 + C_3) - (4C_1 + 3C_2 + C_3)z + (4C_1 + 2C_2)z^2. \end{aligned}$$

So we must solve the following system of three linear equations:

$$\begin{aligned} C_1 + C_2 + C_3 &= 0, \\ 4C_1 + 3C_2 + C_3 &= -2, \\ 4C_1 + 2C_2 \qquad &= 0. \end{aligned}$$

The first equation minus the second plus the third gives $C_1 = 2$, and then one deduces $C_2 = -4$ and $C_3 = 2$. So

$$\frac{2z}{(1-z)(1-2z)^2} = \frac{2}{1-z} - \frac{4}{1-2z} + \frac{2}{(1-2z)^2},$$

and the coefficient of z^n is $2 - 4 \times 2^n + 2(n+1)2^n = (n-1)2^{n+1} + 2$.

It is sometimes possible to find square roots, cube roots, etc. of formal power series. By the Binomial Theorem, for any nonnegative integer a we have the following equation of polynomials in z:

$$(1+z)^a = \sum_{n=0}^{a} \binom{a}{n} z^n. \tag{3.13}$$

Definition 3.28. For any complex number α, we define the formal power series $(1+z)^\alpha$ by:

$$(1+z)^\alpha := \sum_{n=0}^{\infty} \binom{\alpha}{n} z^n.$$

(Recall that $\binom{\alpha}{n} = \frac{\alpha(\alpha-1)\cdots(\alpha-n+1)}{n!}$ is defined for any $\alpha \in \mathbb{C}$.)

In the special case when α is a nonnegative integer a, this definition does indeed give the polynomial $(1+z)^a$ that we expect: the reason is that every term where $n > a$ vanishes, because then $\binom{a}{n} = 0$.

Theorem 3.29. $(1+z)^\alpha(1+z)^\beta = (1+z)^{\alpha+\beta}$, for any $\alpha, \beta \in \mathbb{C}$.

Proof.** We just need to show that the coefficient of z^n is the same on both sides, for all $n \geq 0$. That is, we must prove that

$$\begin{aligned}\sum_{m=0}^{n} \frac{\alpha(\alpha-1)\cdots(\alpha-m+1)}{m!} \frac{\beta(\beta-1)\cdots(\beta-(n-m)+1)}{(n-m)!} \\ = \frac{(\alpha+\beta)(\alpha+\beta-1)\cdots(\alpha+\beta-n+1)}{n!}.\end{aligned} \tag{3.14}$$

One thing we certainly know is that $(1+z)^a(1+z)^b = (1+z)^{a+b}$ for every nonnegative integers a and b, so (3.14) must hold whenever $\alpha, \beta \in \mathbb{N}$. There is a clever way to deduce the general case from this, using the fact that a nonzero polynomial can have only finitely many roots (this is because every root λ of a polynomial corresponds to a factor $x - \lambda$ in its factorization). First fix α to be some nonnegative integer a. Then both sides of (3.14) are polynomial functions of β, so their difference is a polynomial function of β, which has infinitely many roots because it vanishes whenever β is a nonnegative integer. The only way this is possible is if their difference is the zero polynomial. So with α fixed to equal a, (3.14) holds for all values of β. Now we can use the same argument turned around: for any fixed value of β, the two sides of (3.14) are polynomial functions of α which agree whenever α is a nonnegative integer, so their difference must be zero. □

Example 3.30. *Directly from Theorem 3.29 we see that* $(1+z)^{1/2}$ *is a square root of* $1+z$*, in the sense that* $(1+z)^{1/2}(1+z)^{1/2} = (1+z)^1 = 1+z$*. Writing out the definition in a slightly nicer way:*

$$\begin{aligned}(1+z)^{1/2} &= \sum_{n=0}^{\infty} \binom{\frac{1}{2}}{n} z^n = \sum_{n=0}^{\infty} \frac{\frac{1}{2}(\frac{1}{2}-1)(\frac{1}{2}-2)\cdots(\frac{1}{2}-n+1)}{n!} z^n \\ &= 1 + \sum_{n=1}^{\infty} \frac{1(-1)(-3)\cdots(3-2n)}{2^n\, n!} z^n \\ &= 1 + \sum_{n=1}^{\infty} \frac{(-1)^{n-1}(2n-3)!!}{2^n\, n!} z^n.\end{aligned}$$

Here $(2n-3)!!$ *is defined as in Example 1.40 (including the convention that* $(-1)!! = 1$*). Similarly,* $(1+z)^{1/3}$ *is a cube root of* $1+z$*, and so forth.*

Example 3.31*. *More generally, if $F(z)$ is any formal power series with zero constant term, then we can find a square root of $1+F(z)$ by substituting $F(z)$ into $(1+z)^{1/2}$; however, it is not often that you can get any nice formula for the coefficients of the result. We will only use the case of the simple substitution of cz for z:*

$$(1+cz)^{1/2} = 1 + \sum_{n=1}^{\infty} \frac{(-1)^{n-1}(2n-3)!!\, c^n}{2^n\, n!}\, z^n. \tag{3.15}$$

You may like to ponder which formal power series have a square root: it is pretty clear that z does not, for instance. All we will need is the easy observation that a square root, if it does exist, is unique up to sign:

Theorem 3.32. If $F(z)$ and $G(z)$ are both square roots of $A(z)$, then either $F(z) = G(z)$ or $F(z) = -G(z)$.

Proof. The equation $F(z)^2 = G(z)^2$ can be rewritten as

$$(F(z) - G(z))(F(z) + G(z)) = 0,$$

and part (1) of Theorem 3.8 forces one of the two factors to be zero. □

Example 3.33*. *Notice that if $\alpha = -a$ is a negative integer, then $(1+z)^{-a}$ must be $\frac{1}{(1+z)^a}$, since when multiplied by $(1+z)^a$ it gives $(1+z)^0 = 1$. So*

$$\begin{aligned}
\frac{1}{(1+z)^a} &= \sum_{n=0}^{\infty} \binom{-a}{n} z^n \\
&= \sum_{n=0}^{\infty} \frac{(-a)(-a-1)(-a-2)\cdots(-a-n+1)}{n!}\, z^n \\
&= \sum_{n=0}^{\infty} \frac{(-1)^n a(a+1)(a+2)\cdots(a+n-1)}{n!}\, z^n,
\end{aligned}$$

which is a rearranged form of the $c = -1$ special case of (3.12).

Finally, we can even solve some 'differential equations' involving formal power series. In the usual theory of (linear ordinary) differential equations, the

exponential function plays a crucial role, being the unique function e^x which satisfies the differential equation $\frac{d}{dx}e^x = e^x$ subject to $e^0 = 1$. So let us try to find a formal power series $\exp(z) = \sum_{n=0}^{\infty} e_n z^n$ such that

$$\exp'(z) = \exp(z), \quad \text{subject to the condition that } \ e_0 = 1. \tag{3.16}$$

Taking the coefficient of z^n on either side of $\exp'(z) = \exp(z)$ gives

$$(n+1)e_{n+1} = e_n, \text{ or equivalently } e_{n+1} = \frac{1}{n+1}e_n.$$

It is easy to see that the solution of this recurrence relation, given the initial condition $e_0 = 1$, is $e_n = \frac{1}{n!}$, so

$$\exp(z) = \sum_{n=0}^{\infty} \frac{1}{n!}z^n = 1 + z + \frac{1}{2}z^2 + \frac{1}{6}z^3 + \frac{1}{24}z^4 + \cdots \tag{3.17}$$

is the unique solution of our 'differential equation' (3.16). Of course, this is not remotely surprising, because this is exactly the Taylor series of the exponential function (and, as a power series of a complex variable, it converges to the complex exponential function everywhere).

We can now use this exponential series to solve 'differential equations' of a simple type:

Theorem 3.34*. Let $F(z)$ be a formal power series, and let $G(z)$ be the unique formal power series with zero constant term such that $G'(z) = F(z)$. Then $A'(z) = F(z)A(z)$ if and only if $A(z) = a_0 \exp(G(z))$, where a_0 is the constant term of $A(z)$.

Remark 3.35*. *This is the analogue of the fact that the solution of the ordinary differential equation* $\frac{d}{dx}a(x) = f(x)a(x)$ *is* $a(x) = a(0)\, e^{\int_0^x f(t)dt}$.

Proof.** The "if" direction, which just says that $A(z) = a_0 \exp(G(z))$ is indeed a solution of the equation $A'(z) = F(z)A(z)$, can be proved using the chain rule:

$$(a_0 \exp(G(z)))' = a_0 \exp(G(z))G'(z) = a_0 \exp(G(z))F(z).$$

Since the constant term of $\exp(G(z))$ is 1, the constant term of $a_0 \exp(G(z))$ is indeed a_0. The "only if" direction says that $a_0 \exp(G(z))$ is the unique

solution of the equation with constant term a_0. So it is enough to prove that, given any number a_0 and any power series $F(z) = \sum_{n=0}^{\infty} b_n z^n$, there is a unique way to choose $a_1, a_2, a_3, \cdots$ so that the power series $A(z) = \sum_{n=0}^{\infty} a_n z^n$ satisfies the equation $A'(z) = F(z)A(z)$. Extracting the z^n term from both sides of this equation, we see that it is equivalent to

$$(n+1)a_{n+1} = \sum_{m=0}^{n} b_m a_{n-m}, \text{ for all } n \geq 0,$$

which is a recurrence relation determining a_{n+1} in terms of $a_0, a_1, \cdots, a_n$ and the fixed numbers b_m. So it is indeed true that $a_1, a_2, a_3, \cdots$ are uniquely determined. □

Remark 3.36*. *Although Theorem 3.34 enables us to express the solution of the 'differential equation' neatly as* $a_0 \exp(G(z))$*, it may not be easy to extract the coefficients of the powers of* z *after substituting* $G(z)$ *in* $\exp(z)$*.*

3.3 Generating functions and recursion

Now that we are familiar with the basic operations involving formal power series, we explore how generating functions can be used to say something about various kinds of recursive sequences $a_0, a_1, a_2, \cdots$. The main idea is that a recurrence relation expressing a_n in terms of earlier terms in the sequence should give rise to an equation expressing $A(z)$ in terms of itself (or its derivatives, powers, etc.). If this equation can be solved, we have a formula for the generating function $A(z)$. Ideally, we would then be able to extract the coefficient of each z^n, giving a closed formula for a_n (and thus solving the recurrence relation). Even when this is not possible, a formula for the generating function is a concise way of summarizing the sequence, and there are standard methods for deducing qualitative information from it.

In our first class of examples, everything works perfectly: generating functions provide an alternative way of solving the linear recurrence relations discussed in the previous chapter. In fact, this is the way that de Moivre originally discovered the characteristic-polynomial rule, Theorem 2.36.

Example 3.37. *Recall from Example 2.33 the sequence defined by*

$$a_0 = 2,\ a_1 = 5,\ a_n = 7a_{n-1} - 12a_{n-2} \text{ for } n \geq 2.$$

Let $A(z)$ denote the generating function of this sequence. We use the recurrence relation to express $A(z)$ in terms of itself:

$$\begin{aligned} A(z) &= a_0 + a_1 z + a_2 z^2 + a_3 z^3 + a_4 z^4 + \cdots \\ &= 2 + 5z + (7a_1 - 12a_0)z^2 + (7a_2 - 12a_1)z^3 + (7a_3 - 12a_2)z^4 + \cdots \\ &= 2 + 5z + 7(a_1 z^2 + a_2 z^3 + a_3 z^4 + \cdots) - 12(a_0 z^2 + a_1 z^3 + a_2 z^4 + \cdots) \\ &= 2 + 5z + 7z(A(z) - 2) - 12z^2 A(z). \end{aligned}$$

Rearranging this gives $(1 - 7z + 12z^2)A(z) = 2 - 9z$, i.e. $A(z) = \frac{2-9z}{1-7z+12z^2}$. We found the coefficient of z^n in the latter expression in Example 3.25, so we can conclude that $a_n = 3^{n+1} - 4^n$, recovering the result of Example 2.33.

Assuming sufficient facility with the shift operation (3.2), one can write such calculations more concisely using sigma notation rather than spelling out the series with a "$\cdots$".

Example 3.38. *Recall from Example 2.34 the sequence defined by*

$$a_0 = 1,\ a_1 = 4,\ a_n = 4a_{n-1} - 4a_{n-2} \text{ for } n \geq 2.$$

Let $A(z) = \sum_{n=0}^{\infty} a_n z^n$ be the generating function. Then

$$\begin{aligned} A(z) &= 1 + 4z + \sum_{n=2}^{\infty} (4a_{n-1} - 4a_{n-2})\, z^n \\ &= 1 + 4z + 4\sum_{n=2}^{\infty} a_{n-1}\, z^n - 4\sum_{n=2}^{\infty} a_{n-2}\, z^n \\ &= 1 + 4z + 4z(A(z) - 1) - 4z^2 A(z), \end{aligned}$$

which after rearrangement becomes $A(z) = \frac{1}{1-4z+4z^2} = \frac{1}{(1-2z)^2}$. Using (3.11), we can write down straight away that $a_n = (n+1)\,2^n$, as in Example 2.34.

Remark 3.39.** *Following the pattern of these examples, one can give a generating-function proof of Theorem 2.36.*

Slightly more useful than solving homogeneous linear relations, which we know how to do anyway, is tackling the non-homogeneous case, where generating functions can avoid the problem of finding a particular solution.

Example 3.40. *Recall the tower of Hanoi sequence satisfying $h_0 = 0$ and the recurrence relation $h_n = 2h_{n-1} + 1$ for all $n \geq 1$. If $H(z)$ denotes the generating function, we have the following equation for $H(z)$ in terms of itself and the geometric series:*

$$\begin{aligned} H(z) &= h_0 + h_1 z + h_2 z^2 + h_3 z^3 + \cdots \\ &= 0 + (2h_0 + 1)z + (2h_1 + 1)z^2 + (2h_2 + 1)z^3 + \cdots \\ &= \boxed{} \end{aligned}$$

which shows that

$$H(z) = \boxed{}$$

From this we deduce the formula $h_n = 2^n - 1$ immediately.

Example 3.41*. *Recall from Example 2.47 the sequence defined by*

$$a_0 = 3,\ a_1 = 5,\ a_n = 2a_{n-1} + 3a_{n-2} + 8n - 4 \text{ for } n \geq 2.$$

Let $A(z) = \sum_{n=0}^{\infty} a_n z^n$ be the generating function. Then

$$\begin{aligned} A(z) &= a_0 + a_1 z + \sum_{n=2}^{\infty} a_n z^n \\ &= 3 + 5z + \sum_{n=2}^{\infty} (2a_{n-1} + 3a_{n-2} + 8n - 4) z^n \\ &= 3 + 5z + 2\sum_{n=2}^{\infty} a_{n-1} z^n + 3\sum_{n=2}^{\infty} a_{n-2} z^n + 8\sum_{n=2}^{\infty} n z^n - 4\sum_{n=2}^{\infty} z^n \\ &= 3 + 5z + 2z\sum_{n=1}^{\infty} a_n z^n + 3z^2\sum_{n=0}^{\infty} a_n z^n + 8\sum_{n=2}^{\infty} n z^n - 4\sum_{n=2}^{\infty} z^n \\ &= 3 + 5z + 2z(A(z) - 3) + 3z^2 A(z) \\ &\qquad + 8\left(\frac{z}{(1-z)^2} - z\right) - 4\left(\frac{1}{1-z} - 1 - z\right), \end{aligned}$$

where the last line uses (3.3) and (3.6). Tidying this up gives

$$(1 - 2z - 3z^2)A(z) = \frac{3 - 7z + 17z^2 - 5z^3}{(1-z)^2},$$

so

$$A(z) = \frac{3 - 7z + 17z^2 - 5z^3}{(1-z)^2(1-3z)(1+z)}. \tag{3.18}$$

We now want to rewrite the right-hand side using partial fractions:

$$\frac{3 - 7z + 17z^2 - 5z^3}{(1-z)^2(1-3z)(1+z)} = \frac{C_1}{1-3z} + \frac{C_2}{1+z} + \frac{C_3}{1-z} + \frac{C_4}{(1-z)^2},$$

where C_1, C_2, C_3, C_4 are some constants. Clearing the denominator, we get

$$\begin{aligned} 3 - 7z + 17z^2 - 5z^3 &= C_1(1-z)^2(1+z) + C_2(1-z)^2(1-3z) \\ &\quad + C_3(1-z)(1-3z)(1+z) + C_4(1-3z)(1+z) \\ &= C_1(1 - z - z^2 + z^3) + C_2(1 - 5z + 7z^2 - 3z^3) \\ &\quad + C_3(1 - 3z - z^2 + 3z^3) + C_4(1 - 2z - 3z^2) \\ &= (C_1 + C_2 + C_3 + C_4) - (C_1 + 5C_2 + 3C_3 + 2C_4)z \\ &\quad + (-C_1 + 7C_2 - C_3 - 3C_4)z^2 - (-C_1 + 3C_2 - 3C_3)z^3, \end{aligned}$$

and equating coefficients gives us a system of four linear equations for the four unknowns:

$$\begin{aligned} C_1 + C_2 + C_3 + C_4 &= 3 \\ C_1 + 5C_2 + 3C_3 + 2C_4 &= 7 \\ -C_1 + 7C_2 - C_3 - 3C_4 &= 17 \\ -C_1 + 3C_2 - 3C_3 &= 5 \end{aligned}$$

Using linear algebra methods, we can solve this to obtain

$$C_1 = 4,\ C_2 = 2,\ C_3 = -1, C_4 = -2,$$

and thus we get the partial-fractions formula for the generating function:

$$A(z) = \frac{4}{1-3z} + \frac{2}{1+z} - \frac{1}{1-z} - \frac{2}{(1-z)^2}. \tag{3.19}$$

From this we can read off the answer:

$$a_n = 4 \times 3^n + 2 \times (-1)^n - 1 - 2(n+1),$$

which is of course the same as in Example 2.47. For this example it is questionable whether the generating-function approach is any better: it is more motivated and direct, but the system of linear equations involved is larger than anything we had to deal with in the other method.

Example 3.42*. *Recall from Example 2.50 the sequence defined by*

$$a_0 = a_1 = 1,\ a_n = 4a_{n-1} - 4a_{n-2} + 2^n \text{ for } n \geq 2.$$

Let $A(z) = \sum_{n=0}^{\infty} a_n\, z^n$ *be the generating function. Then*

$$\begin{aligned} A(z) &= a_0 + a_1 z + \sum_{n=2}^{\infty} a_n\, z^n \\ &= 1 + z + \sum_{n=2}^{\infty} (4a_{n-1} - 4a_{n-2} + 2^n)\, z^n \\ &= 1 + z + 4z \sum_{n=1}^{\infty} a_n\, z^n - 4z^2 \sum_{n=0}^{\infty} a_n\, z^n + \left(\frac{1}{1-2z} - 1 - 2z \right) \\ &= -z + 4z(A(z) - 1) - 4z^2 A(z) + \frac{1}{1-2z}. \end{aligned}$$

Rearranging gives

$$(1 - 4z + 4z^2)A(z) = \frac{1 - 5z + 10z^2}{1 - 2z},$$

so

$$A(z) = \frac{1 - 5z + 10z^2}{(1-2z)^3}. \tag{3.20}$$

We now need to rewrite the right-hand side using partial fractions:

$$\frac{1 - 5z + 10z^2}{(1-2z)^3} = \frac{C_1}{1-2z} + \frac{C_2}{(1-2z)^2} + \frac{C_3}{(1-2z)^3},$$

where C_1, C_2, C_3 *are some numbers. Clearing the denominator, we get*

$$\begin{aligned} 1 - 5z + 10z^2 &= C_1(1-2z)^2 + C_2(1-2z) + C_3 \\ &= (C_1 + C_2 + C_3) - (4C_1 + 2C_2)z + 4C_1 z^2, \end{aligned}$$

and after equating coefficients we find $C_1 = 5/2$, $C_2 = -5/2$, $C_3 = 1$. *So*

$$A(z) = \frac{5/2}{1-2z} - \frac{5/2}{(1-2z)^2} + \frac{1}{(1-2z)^3}. \tag{3.21}$$

Recalling (3.12), *we can extract the coefficient of* z^n *from the right-hand side:*

$$a_n = \frac{5}{2}2^n - \frac{5}{2}(n+1)2^n + \binom{n+2}{2}2^n,$$

which simplifies to the answer found in Example 2.50.

Along the same lines, we can find a formula for the generating function of the kth column of the Stirling triangle (compare this with Theorem 3.17):

Theorem 3.43*. For any $k \geq 1$,

$$\sum_{n=0}^{\infty} S(n+k,k)\, z^n \;=\; \frac{1}{(1-z)(1-2z)\cdots(1-kz)}.$$

Proof*. We prove this by induction. The $k = 1$ base case is (3.3). Assume that $k \geq 2$, and that the result is known for $k-1$. Define $G_k(z) := \sum_{n=0}^{\infty} S(n+k,k)\, z^n$. We must convert the recurrence relation for Stirling numbers, Theorem 1.81, into an equation relating $G_k(z)$ and $G_{k-1}(z)$:

$$\begin{aligned} G_k(z) &= \sum_{n=0}^{\infty} (S(n+k-1,k-1) + k\, S(n+k-1,k))\, z^n \\ &= \sum_{n=0}^{\infty} S(n+k-1,k-1)\, z^n + k \sum_{n=1}^{\infty} S(n+k-1,k) z^n \\ &= G_{k-1}(z) + kzG_k(z). \end{aligned}$$

Rearranging, this becomes $G_k(z) = \frac{G_{k-1}(z)}{1-kz}$, and the result follows from the induction hypothesis. □

Remark 3.44.** *To deduce a formula for an individual Stirling number $S(n+k,k)$, we need to find the coefficient of z^n in $\frac{1}{(1-z)(1-2z)\cdots(1-kz)}$, which is the product*

$$(1 + z + z^2 + \cdots)(1 + 2z + 2^2z^2 + \cdots)\cdots(1 + kz + k^2z^2 + \cdots).$$

To get a z^n term from the expansion of the product, one must select the z^{n_1} term from the first factor, the z^{n_2} term from the second factor, etc., up to the z^{n_k} term from the kth factor, where $n_1 + n_2 + \cdots + n_k = n$. The coefficient obtained from this selection is $1^{n_1}2^{n_2}\cdots k^{n_k}$, so we find

$$S(n+k,k) = \sum_{\substack{n_1,n_2,\cdots,n_k \in \mathbb{N} \\ n_1+n_2+\cdots+n_k=n}} 1^{n_1}2^{n_2}\cdots k^{n_k}. \tag{3.22}$$

This is a positive formula, but with the major drawback that the sum has $\binom{n+k-1}{k-1}$ terms. So finding a formula for the generating function has not

actually allowed us to solve the recurrence relation. Incidentally, (3.22) *can be proved directly from the definition of* $S(n+k,k)$*: the term indexed by* $n_1, n_2, \cdots, n_k$ *counts those partitions of* $\{1, 2, \cdots, n+k\}$ *in which one block has largest element* $n_k + 1$*, another has largest element* $n_{k-1} + n_k + 2$*, and so on until the last block has largest element* $n_1 + \cdots + n_k + k = n + k$*.*

Perhaps more importantly, generating functions can be used to solve other types of recurrence relations which are not linear.

Example 3.45*. *The classic example is the Catalan sequence* $c_0, c_1, c_2, \cdots$*, whose generating function we are calling* $C(z)$*. Recall that the Catalan numbers satisfy:*

$$c_0 = 1,\ c_n = c_0 c_{n-1} + c_1 c_{n-2} + \cdots + c_{n-2} c_1 + c_{n-1} c_0 \text{ for } n \geq 1.$$

The form of the recurrence relation suggests generating functions straight away, because the right-hand side $\sum_{m=0}^{n-1} c_m c_{n-1-m}$ *is exactly the coefficient of* z^{n-1} *in* $C(z)^2$*, by the multiplication rule. Thus we have*

$$C(z) = 1 + \sum_{n=1}^{\infty} c_n\, z^n = 1 + zC(z)^2.$$

So $C(z)$ *is a solution of the quadratic equation* $zC(z)^2 - C(z) + 1 = 0$*, to which it is tempting to apply the quadratic formula. If that seems a leap too far, we can certainly do some completion of the square (which is how the quadratic formula is proved anyway):*

$$0 = 4z - 4zC(z) + 4z^2 C(z)^2 = (1 - 2zC(z))^2 - 1 + 4z,$$

which shows that $1 - 2zC(z)$ *is a square root of* $1 - 4z$*. By Theorem 3.32, the only square roots of* $1 - 4z$ *are* $(1-4z)^{1/2}$ *and its negative; since* $1 - 2zC(z)$ *has constant term* 1*, it must equal* $(1-4z)^{1/2}$*. Setting* $c = -4$ *in* (3.15)*, we deduce that*

$$\begin{aligned} 1 - 2zC(z) &= 1 + \sum_{n=1}^{\infty} \frac{(-1)^{n-1}(2n-3)!!\,(-4)^n}{2^n\, n!}\, z^n \\ &= 1 - \sum_{n=1}^{\infty} \frac{2^n (2n-3)!!}{n!} z^n \\ &= 1 - 2z \sum_{n=0}^{\infty} \frac{2^n (2n-1)!!}{(n+1)!} z^n, \end{aligned}$$

which implies that

$$C(z) = \sum_{n=0}^{\infty} \frac{2^n(2n-1)!!}{(n+1)!} z^n. \tag{3.23}$$

From this we can read off our desired formula for the Catalan numbers:

$$c_n \;=\; \frac{2^n(2n-1)!!}{(n+1)!} \;=\; \frac{(2n)!}{(n+1)!n!}, \tag{3.24}$$

where the second equality uses $(2n-1)!! = \frac{(2n)!}{2^n\, n!}$ *(see Example 1.40). Note that* c_n *is not quite a binomial coefficient, but* $c_n = \frac{1}{n+1}\binom{2n}{n} = \frac{1}{2n+1}\binom{2n+1}{n}$.

Generating functions can be very helpful in finding closed formulas for sums $f(0)+f(1)+\cdots+f(n)$. The reason is that if $A(z)$ is the generating function of $a_0, a_1, a_2, \cdots$, then by (3.3) and the multiplication rule,

$$\frac{A(z)}{1-z} = \sum_{n=0}^{\infty} (a_0 + a_1 + \cdots + a_n)\, z^n, \tag{3.25}$$

so $\frac{A(z)}{1-z}$ is the generating function of the sequence of partial sums of $a_0, a_1, \cdots$.

Example 3.46. *To find a closed formula for* $0\times 2^0 + 1\times 2^1 + \cdots + n\times 2^n$*, we first need a formula for the generating function of the sequence* $0\times 2^0, 1\times 2^1, \cdots$*: by* (3.11)*, this is*

$$\sum_{n=0}^{\infty} n2^n\, z^n = \boxed{}$$

Now divide both sides by $1-z$*, and use* (3.25)*:*

$$0\times 2^0 + 1\times 2^1 + \cdots + n\times 2^n = \text{ coefficient of } z^n \text{ in } \boxed{}$$

We already found this coefficient in Example 3.27: it is $(n-1)2^{n+1}+2$.

Example 3.47. *In Example 2.22 we were unable to guess a closed formula for the sum of squares* $1^2+2^2+\cdots+n^2$*. To solve this problem using generating functions, we first need a formula for the generating function of the sequence of squares itself, i.e.* $0^2 + 1^2 z + 2^2 z^2 + \cdots$*. We found this in Example 3.19:*

$$\sum_{n=0}^{\infty} n^2 z^n = \frac{z}{(1-z)^2} + \frac{2z^2}{(1-z)^3}.$$

(It is convenient not to combine the two terms.) Thus (3.25) *tells us that*

$$\sum_{n=0}^{\infty} \left(0^2 + 1^2 + \cdots + n^2\right) z^n = \frac{z}{(1-z)^3} + \frac{2z^2}{(1-z)^4}$$
$$= \sum_{n=0}^{\infty} \binom{n+1}{2} z^n + 2\sum_{n=0}^{\infty} \binom{n+1}{3} z^n.$$

(As in (3.9)*, we can let the summations start from $n = 0$, because the early terms are zero anyway.) From this we read off the formula*

$$1^2 + 2^2 + \cdots + n^2 = \binom{n+1}{2} + 2\binom{n+1}{3} = \frac{(n+1)n(2n+1)}{6}. \qquad (3.26)$$

The generalization of Example 3.47 is the following:

Theorem 3.48*. For any positive integer a,

$$1^a + 2^a + \cdots + n^a = \sum_{k=1}^{a} k!\, S(a,k) \binom{n+1}{k+1}.$$

Remark 3.49*. *Expressing one sum as another sum may not seem like progress, but the sum on the right-hand side has only a terms, and we are imagining a as being fixed while n varies.*

Proof*. Using Theorem 1.86 and (3.9), we get a formula for the generating function of the sequence of ath powers:

$$\sum_{n=0}^{\infty} n^a z^n = \sum_{k=1}^{a} k!\, S(a,k) \sum_{n=0}^{\infty} \binom{n}{k} z^n = \sum_{k=1}^{a} k!\, S(a,k) \frac{z^k}{(1-z)^{k+1}}.$$

Then (3.25) tells us that

$$\sum_{n=0}^{\infty} \left(\sum_{m=0}^{n} m^a \right) z^n = \sum_{k=1}^{a} k!\, S(a,k) \frac{z^k}{(1-z)^{k+2}}.$$

Since the coefficient of z^n in $\frac{z^k}{(1-z)^{k+2}}$ is $\binom{n+1}{k+1}$, the result follows. □

To conclude, here are a couple of examples which use our (rudimentary) knowledge of differential equations of formal power series.

Example 3.50.** *Consider the sequence defined by*

$$a_0 = a_1 = 1,\ a_n = \frac{1}{n}(a_{n-1} + a_{n-2}) \text{ for } n \geq 2.$$

Let $A(z) = \sum_{n=0}^{\infty} a_n\, z^n$ *be the generating function. The fact that* na_n *is involved in the recurrence relation prompts us to look at the derivative:*

$$\begin{aligned} A'(z) &= \sum_{n=0}^{\infty}(n+1)a_{n+1}\, z^n \\ &= 1 + \sum_{n=1}^{\infty}(a_n + a_{n-1})\, z^n \\ &= 1 + \sum_{n=1}^{\infty} a_n\, z^n + \sum_{n=1}^{\infty} a_{n-1}\, z^n \\ &= 1 + (A(z) - 1) + zA(z) \\ &= (1+z)A(z). \end{aligned}$$

Theorem 3.34 says that the solution of this (subject to $a_0 = 1$*) is*

$$A(z) = \exp\left(z + \frac{1}{2}z^2\right).$$

So the generating function is very concisely expressed. To get a formula for a_n *itself, we need a bit more work:*

$$\begin{aligned} A(z) &= \sum_{m=0}^{\infty} \frac{1}{m!}\left(z + \frac{1}{2}z^2\right)^m \\ &= \sum_{m=0}^{\infty} \frac{1}{m!} z^m \left(1 + \frac{1}{2}z\right)^m \\ &= \sum_{m=0}^{\infty} \frac{1}{m!} z^m \sum_{k=0}^{m} \frac{\binom{m}{k}}{2^k}\, z^k, \end{aligned}$$

and extracting the coefficient of z^n *gives*

$$a_n = \sum_{m=0}^{n} \frac{\binom{m}{n-m}}{2^{n-m}\, m!} = \sum_{m=\lceil \frac{n}{2} \rceil}^{n} \frac{1}{2^{n-m}(n-m)!(2m-n)!}.$$

However, this is not a closed formula, since the number of terms in the sum grows as n grows. With sequences such as this, we often have to be satisfied with giving a formula for the generating function.

Example 3.51.** *Recall from Example 2.23 the sequence defined by*

$$a_0 = 1,\ a_n = \frac{1}{n}(a_{n-1} + 2a_{n-2} + 3a_{n-3} + \cdots + na_0) \text{ for } n \geq 1.$$

Letting $A(z) = \sum_{n=0}^{\infty} a_n\, z^n$ be the generating function, we have

$$\begin{aligned} A'(z) &= \sum_{n=0}^{\infty}(n+1)a_{n+1}\, z^n \\ &= \sum_{n=0}^{\infty}(a_n + 2a_{n-1} + 3a_{n-2} + \cdots + (n+1)a_0)\, z^n \\ &= (1 + 2z + 3z^2 + 4z^3 + \cdots)A(z) \\ &= \frac{1}{(1-z)^2}A(z). \end{aligned}$$

Again, this is a differential equation of the type we know how to solve: the required 'integral' of $\frac{1}{(1-z)^2}$ is $\frac{z}{1-z} = z + z^2 + z^3 + \cdots$, since we need it to have zero constant term. So $A(z) = \exp\left(\frac{z}{1-z}\right)$. As in the previous example, this does not lead to a closed formula for a_n.

Part II

Introduction to Graph Theory

Anthony Henderson

Acknowledgements

These lecture notes were written in 2009 for the units MATH2069 Discrete Mathematics and Graph Theory and MATH2969 Discrete Mathematics and Graph Theory (Advanced), given at the University of Sydney. I am extremely grateful to the previous lecturer Bill Palmer for allowing me to make use of his lecture notes on graph theory.

I also acknowledge a debt to the following textbooks, which students should consult for more examples and exercises.

- *A First Look at Graph Theory* by J. Clark and D. A. Holton
- *Introduction to Graph Theory* by R. J. Wilson
- *Graph Theory* by R. Diestel

Anthony Henderson

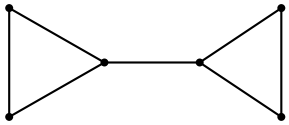
Contents

Chapter 0

Introduction

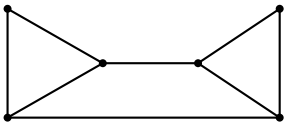
Graph theory is one of the most important and useful areas of discrete mathematics. To get a preliminary idea of what it involves, let us consider some of the real-world problems which can be translated in graph-theoretic terms.

In the modern age of telecommunications, networks and their properties are hot topics of study. Every time you download data from the internet, the request passes from your computer to the computer where the data is actually stored via a whole chain of intermediaries, each connected to the next. The speed and efficiency of the process depends vitally on how well connected the whole network is. One way to measure this connectedness is to determine how many computers would need to crash before the other computers would become unable to communicate with each other. For a simple example, imagine six computers connected as follows.



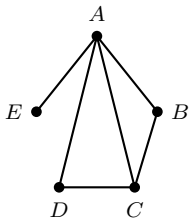
If either of the middle two computers crashes, the others would be split into two groups which could no longer communicate. To fix this problem, suppose

you add one more connection to the network.



Then the loss of any single computer would not be fatal: the remaining computers would still be linked to each other. However, if you examine the new network, you will be able to find two computers such that if they both crashed simultaneously, the remaining four computers would be split into two groups; if this represented an unacceptable risk, you would have to add another connection. This sort of thinking about the nodes of a network and their connections is quintessentially graph-theoretic.

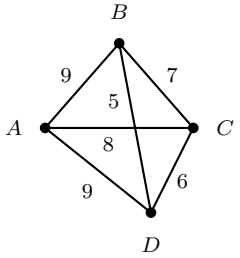
Another widespread use of graph theory is in what are called scheduling problems. Suppose there are various university courses which need to be timetabled so that no students have a clash. To work out how many timetable slots are required, you need to know which pairs of courses have a student in common (because they then can't be put on simultaneously). For example, suppose there are five courses named A, B, C, D, E, and the pairs which have a student in common are joined in the following picture.



Then clearly A, B, and C need to be allocated three different time-slots, but D could be scheduled concurrently with B and E with C. Again the problem amounts to considering abstract properties of the picture, and which courses are connected to which. A similar problem confronts map-makers when they have to choose a colour for each region of the map so that no two adjacent regions have the same colour; again, the only information which is relevant is the pattern of adjacencies between regions, which could be displayed in the same sort of picture as above.

One of the most famous problems in graph theory is the Travelling Salesman Problem. The original context was that a salesman wanted to visit a certain

number of towns, returning to his starting point, and having as short a total distance to travel as possible. The information necessary to solve the problem is not only which towns are connected to which, but how far apart they are (considering only the shortest way of travelling between each pair of towns). For example, there might be four towns A, B, C, D, where A is the salesman's home town, with distances in kilometres given in the following picture.



One way to solve the problem would be to list all the possible routes and calculate their distances: for instance, the route A–B–C–D–A has distance $9+7+6+9 = 31$, whereas the route A–B–D–C–A has distance $9+5+6+8 = 28$. The disadvantage of this approach is that if there are n towns, there are $(n-1)!$ possible routes, and as n increases, the growth rate of $(n-1)!$ is faster than exponential. There are many smarter algorithms which cut down the amount of calculation involved, but it is a major open question whether there is any algorithm for solving the Travelling Salesman Problem whose running time has polynomial growth rate. The ramifications of such an algorithm would be huge, because there are many problems in all sorts of areas which can be rephrased in terms of finding the shortest paths through a diagram.

Comments on the text

In these notes, the main results are all called "Theorem", irrespective of their difficulty or significance. Usually, the statement of the Theorem is followed (perhaps after an intervening Example or Remark) by a rigorous Proof; as is traditional, the end of each proof is marked by an open box.

To enable cross-references, the Theorems, Definitions, Examples, and Remarks are all numbered in sequence (the number always starts with the relevant chapter number). Various equations and formulas are also numbered (on their right-hand edge) in a separate sequence.

In definitions (either formally numbered or just in the main text), a word or phrase being defined is underlined. The symbol ":=" is used to mean "is defined to be" or "which is defined to be".

The text in the Examples and Remarks is *slanted*, to make them stand out on the page. Many students will understand the Theorems more easily by studying the Examples which illustrate them than by studying their proofs. Some of the Examples contain blank boxes for you to fill in. The Remarks are often side-comments, and are usually less important to understand.

The more difficult Theorems, Proofs, Examples, and Remarks are marked at the beginning with either * or **. Those marked * are at the level which MATH2069 students will have to understand in order to be sure of getting a Credit, or to have a chance of a Distinction or High Distinction. Those marked ** are really intended only for the students enrolled in the Advanced unit MATH2969, and can safely be ignored by those enrolled in MATH2069.

Chapter 1

First Properties of Graphs

What the problems in the introduction have in common is that they are all concerned with a collection of objects (computers, courses, towns) some of which are linked together (in a literal physical sense, or by some property such as two courses having a student in common). The abstract definition of a graph is meant to capture this basic idea. (Note that the word "graph" is used here in a different sense from the graphs of functions in calculus.)

1.1 Basic definitions

Here is the definition of graph we will use in this course.

Definition 1.1. A graph G is a pair (V, E) where:

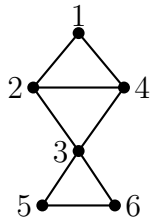
- V is a finite set whose elements are called the vertices of G,
- E is a set of unordered pairs $\{v, w\}$ where $v, w \in V$ and $v \neq w$. The elements of E are called the edges of G.

We say that $v, w \in V$ are adjacent in G if $\{v, w\}$ is one of the edges of G. If $e = \{v, w\} \in E$, we say that v and w are the ends of the edge e.

In applications, the vertices of the graphs may be computers and the edges may be connections between them, or the vertices may be towns and the edges may be roads, and so forth. But for the purpose of developing the theory, we will usually use positive integers or letters of the alphabet as the vertices. Wherever possible, n will denote the number of vertices.

In drawing pictures of graphs, we will represent the vertices as little dots, labelled by the appropriate number or symbol, and the edges as lines or curves running from one end to the other. The positions of the dots have no significance; nor do the shapes of the lines or the places where they may happen to cross in the picture (other than at the dots). All that matters is which vertices are adjacent to which.

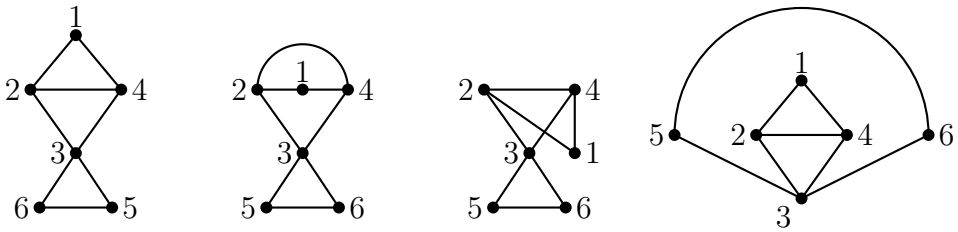
Example 1.2. *The following picture represents a graph with vertex set* $\{1, 2, 3, 4, 5, 6\}$.



The edges in this graph are as follows:

$$\{1,2\},\ \{1,4\},\ \{2,3\},\ \{2,4\},\ \{3,4\},\ \{3,5\},\ \{3,6\},\ \{5,6\}.$$

So 1 *is adjacent to* 2 *and to* 4 *but not to* 3, 5, *or* 6*;* 3 *is adjacent to all the other vertices except for* 1*; and so forth. Any of the following pictures represents the same graph.*



Remark 1.3. *Many authors allow more freedom in the definition of graph. Specifically, their graphs can have multiple edges (more than one edge joining the same two vertices) or loops (edges which have the same vertex for both ends) – both of these are ruled out in our definition, where an edge is just a set of* 2 *different vertices (those other authors would call our graphs*

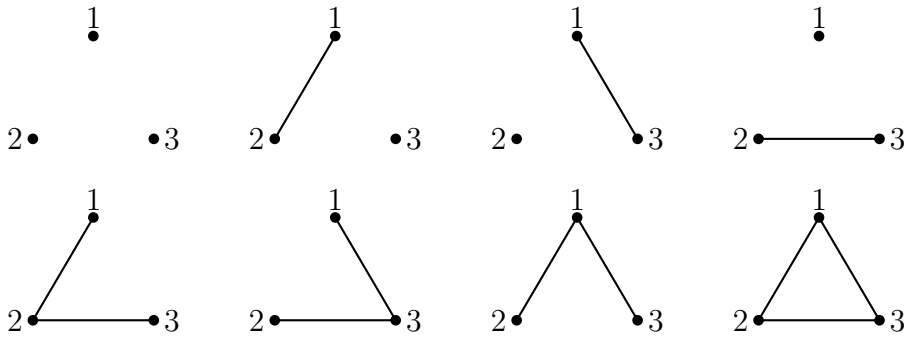
simple graphs). On the rare occasions where we need to discuss graphs with multiple edges or loops, we will call them multigraphs. There is also an important concept of directed graph, where the edges are ordered pairs rather than unordered pairs, and can hence be thought of as having a definite direction from one end to the other. We will ignore all these variants for now.

Our first result about graphs is an illustration of basic counting principles.

Theorem 1.4. If $|V| = n$, then the number of graphs with vertex set V is $2^{\binom{n}{2}}$. The number of edges in such a graph could be anything from 0 to $\binom{n}{2}$. For fixed k, the number of graphs with vertex set V and k edges is $\binom{\binom{n}{2}}{k}$.

Proof. If you fix the vertex set V, then the graph is determined by specifying the edge set E. By definition, E is a subset of the set X of two-element subsets of V. We know that if $|V| = n$, then $|X| = \binom{n}{2}$. So the size of a subset of X can be anything from 0 to $\binom{n}{2}$; the number of subsets of X of size k is $\binom{\binom{n}{2}}{k}$; and the total number of subsets of X is $2^{\binom{n}{2}}$. (Recall the reason for this last part: specifying a subset of X is the same as deciding, for each element of X, whether it is in or out. In our situation, we have $\binom{n}{2}$ pairs of vertices which are 'potential edges', and we have to decide for each pair of vertices whether to join them or not.) □

Example 1.5. *Since* $\binom{3}{2} = 3$, *there are* $8 = 2^3$ *graphs with vertex set* $\{1, 2, 3\}$*:*



Of these graphs, $\binom{3}{0} = 1$ *has no edges,* $\binom{3}{1} = 3$ *have one edge,* $\binom{3}{2} = 3$ *have two edges, and* $\binom{3}{3} = 1$ *has three edges.*

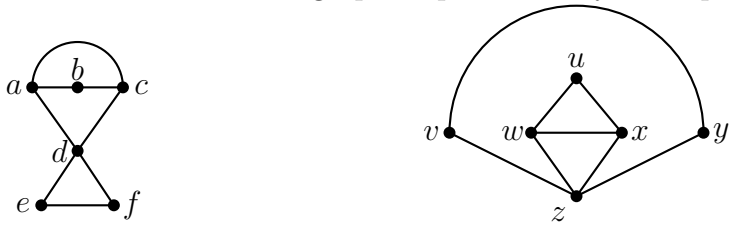
There is a natural sense in which there are 'really' only four graphs involved in Example 1.5: all the graphs with one edge have 'the same form', if you forget about the labels of the vertices, and so have all the graphs with two edges. This looser kind of sameness is called isomorphism (from the Greek for "same form"). It is formalized in the following definition.

Definition 1.6. A graph $G = (V, E)$ is said to be isomorphic to a graph $G' = (V', E')$ if there is a bijection between their vertex sets under which their edge sets correspond; that is, a bijective function $f : V \to V'$ such that

$$E' = \{\{f(v), f(w)\} \mid \{v, w\} \in E\},$$

i.e. $f(v), f(w)$ are adjacent in G' if and only if v, w are adjacent in G.

Example 1.7. *Consider the two graphs represented by these pictures.*



The first graph is (V, E) where $V = \{a, b, c, d, e, f\}$ and

$$E = \{\{a, b\}, \{a, c\}, \{a, d\}, \{b, c\}, \{c, d\}, \{d, e\}, \{d, f\}, \{e, f\}\},$$

and the second graph is (V', E') where $V' = \{u, v, w, x, y, z\}$ and

$$E' = \{\{u, w\}, \{u, x\}, \{v, y\}, \{v, z\}, \{w, x\}, \{w, z\}, \{x, z\}, \{y, z\}\}.$$

Although their pictures look superficially different, (V, E) is isomorphic to (V', E') via the following bijection:

$$a \leftrightarrow w,\ b \leftrightarrow u,\ c \leftrightarrow x,\ d \leftrightarrow z,\ e \leftrightarrow v,\ f \leftrightarrow y.$$

You can check from the above listings that the edge sets correspond; a more visual way to see the isomorphism is to re-draw the second graph so that it obviously has the same form as the first graph.

Note that the bijection given above is not the only one we could have used: it would be just as good to let e correspond to y and f to v, for instance.

Remark 1.8. *To avoid confusion, the two graphs in Example 1.7 had different vertex sets. But it is also possible for two graphs with the same vertex set to be isomorphic. In the special case that $V' = V$, Definition 1.6 says that (V, E) is isomorphic to (V, E') if and only if there is some permutation of the vertices which makes the edge sets correspond; in other words, a picture of (V, E') can be obtained from a picture of (V, E) by permuting the labels of the vertices while leaving the edges where they are.*

It is important to prove that isomorphism of graphs is an equivalence relation, which means the following.

Theorem 1.9. (1) Any graph G is isomorphic to itself.

(2) If G is isomorphic to G', then G' is isomorphic to G.

(3) If G is isomorphic to G' and G' to G'', then G is isomorphic to G''.

Proof. Let $G = (V, E)$, $G' = (V', E')$, $G'' = (V'', E'')$. To prove (1), we can use the most obvious bijective function from V to itself, namely the identity map which leaves every vertex unchanged. To prove (2), if $f : V \to V'$ is a bijection under which the edge sets of the two graphs correspond, then its inverse $f^{-1} : V' \to V$ is another such bijection. To prove (3), let $f : V \to V'$ and $g : V' \to V''$ be bijections such that

$$v, w \text{ adjacent in } G \iff f(v), f(w) \text{ adjacent in } G' \qquad \text{and}$$
$$v', w' \text{ adjacent in } G' \iff g(v'), g(w') \text{ adjacent in } G''.$$

Then the composition $g \circ f : V \to V''$ is also a bijection, and from the above two equivalences we deduce that

$$v, w \text{ adjacent in } G \iff g(f(v)), g(f(w)) \text{ adjacent in } G''.$$

So G is isomorphic to G'' as required. □

This means that it makes sense to classify graphs into isomorphism classes, where two graphs are in the same isomorphism class if and only if they are isomorphic to each other; in visual terms, if and only if they can be represented by the same picture with (possibly) different labellings of the vertices. We represent isomorphism classes of graphs by pictures with unlabelled vertices.

Remark 1.10. *Here are some elementary classification principles. To show that two graphs are isomorphic, one needs to find a bijection between their vertex sets under which the edges correspond, as in Example 1.7; in principle, this requires a laborious search, although there could well be short-cuts. But to show that two graphs are not isomorphic, one just needs to find one point of 'essential difference' between them, one discrepancy which could not occur if they were isomorphic. As Example 1.7 showed, the actual names of the vertices may well differ from one isomorphic graph to another (that is an 'inessential difference'); but the number of vertices clearly cannot differ, nor can the number of edges. So the classification immediately breaks down into a lot of sub-problems, such as "classify graphs with* 7 *vertices and* 15 *edges". We will see many more properties of graphs, and they will all be of this intrinsic kind which is unchanged under isomorphism. So every time we define a new property of graphs, we have a new tool for showing that two graphs are not isomorphic, and we can refine the classification further.*

I should confess straight away that we will never actually complete any general classification of graphs; in fact, our focus will shift away from this problem in later chapters. But it is useful to examine some small examples of such classification, to get some familiarity with basic concepts.

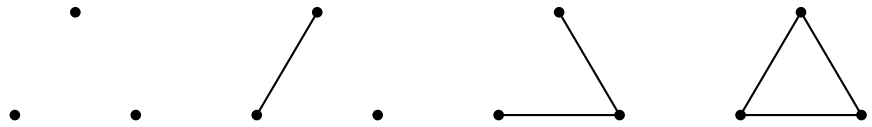
Example 1.11. *There is only one graph with* 0 *vertices (the vertex set and the edge set both have to be the empty set). By contrast, there are arguably infinitely many graphs with* 1 *vertex, because that vertex could be anything you like (the edge set has to be empty). However, what really matters is that there is only one isomorphism class of graphs with* 1 *vertex: they all have the 'same form', since they just consist of a single vertex and no edges.*

Example 1.12. *If a graph has two vertices, it can either have* 0 *edges or* 1 *edge. Clearly all the graphs with two vertices and* 0 *edges are isomorphic to each other, as are all the graphs with two vertices and* 1 *edge. Hence there are two isomorphism classes of graphs with two vertices, which can be represented by the following pictures.*

Remember that there is no point labelling the vertices, because the names of the vertices are irrelevant to the isomorphism class of a graph.

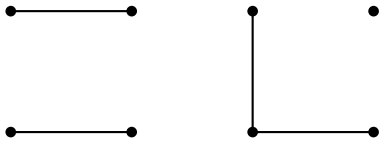
Example 1.13. *As foreshadowed after Example 1.5, there are four isomorphism classes of graphs with three vertices, one for each of the possible num-*

bers of edges (from 0 *to* 3*):*

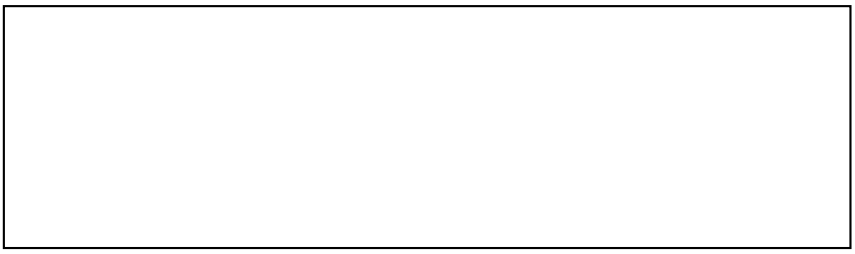


For instance, we are claiming here that every graph with three vertices and two edges must have a vertex where the two edges meet; this is clear, because if the two edges had no ends in common there would have to be at least $2 \times 2 = 4$ *vertices.*

Example 1.14. *There is only one isomorphism class of graphs with four vertices and no edges, and similarly for four vertices and one edge. If a graph has four vertices and two edges, the only essential information remaining is whether the edges meet at a vertex or not. This gives rise to two isomorphism classes:*

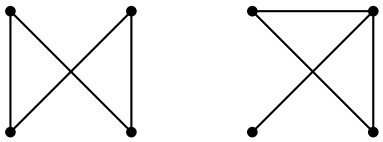


Classifying graphs with four vertices and three edges is a little more interesting. One way to approach it is to fix the specific vertex set $\{1, 2, 3, 4\}$*, and draw pictures of all the graphs with this vertex set and three edges; by Theorem 1.4, there are* $\binom{\binom{4}{2}}{3} = 20$ *of these. Examining the pictures, we can easily classify them into the following isomorphism classes:*



We will develop more systematic ways of doing this classification later. If you start considering the possible configurations of four edges, you will quickly realize that it is easier to classify these graphs according to the two 'non-edges', which must either meet at a vertex or not; this gives two isomorphism

classes:



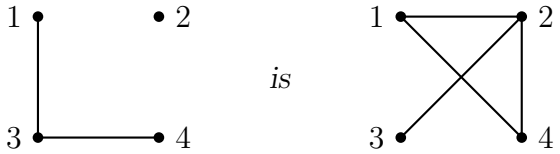
By the same reasoning, there is only one isomorphism class of graphs with four vertices and five edges (and one non-edge), and clearly there is only one isomorphism class of graphs with four vertices and six edges (where every possible edge is included).

We can formalize the idea of considering the 'non-edges' of a graph as follows.

Definition 1.15. The complement of the graph $G = (V, E)$ is the graph $\overline{G} = (V, \overline{E})$ where $\overline{E}$ is the complement of E in the set of 2-element subsets of V. In other words, $\overline{G}$ has the same vertex set V as G, and for $v \neq w \in V$,

$$v, w \text{ are adjacent in } \overline{G} \text{ if and only if } v, w \text{ are not adjacent in } G.$$

Example 1.16. *The complement of*



Theorem 1.17. Two graphs G and G' are isomorphic if and only if their complements $\overline{G}$ and $\overline{G'}$ are isomorphic.

Proof. Let $G = (V, E)$, $G' = (V', E')$. The condition for G and G' to be isomorphic is that there is a bijective function $f : V \to V'$ such that for $v \neq w \in V$,

$$v, w \text{ adjacent in } G \iff f(v), f(w) \text{ adjacent in } G'.$$

But by basic logic, this condition is unchanged if "adjacent" is changed to "not adjacent" on both sides, in which case it becomes the condition for $\overline{G}$ and $\overline{G'}$ to be isomorphic. □

Note that if G has n vertices and k edges, then $\overline{G}$ has n vertices and $\binom{n}{2} - k$ edges. So classifying graphs with n vertices and $\binom{n}{2} - k$ edges is essentially the same as classifying graphs with n vertices and k edges.

1.2 Connectedness and subgraphs

One glaring difference between some of the graphs we classified in the previous section was that sometimes there was a way to get from every vertex to every other vertex along the edges, and sometimes there wasn't: if the edges represent lines of communication, then sometimes all the vertices could communicate with each other and sometimes they split into more than one group.

Definition 1.18. Let $G = (V, E)$ be a graph. If $v, w \in V$, a walk from v to w in the graph G is a sequence of vertices

$$v_0, v_1, v_2, \cdots, v_\ell \in V, \text{ with } v_0 = v \text{ and } v_\ell = w,$$

such that v_i and v_{i+1} are adjacent for all $i = 0, 1, \cdots, \ell - 1$, or in other words the following are all in E:

$$\{v_0, v_1\},\ \{v_1, v_2\},\ \cdots,\ \{v_{\ell-1}, v_\ell\}.$$

The length of such a walk is the number of steps, i.e. ℓ. We say that v is linked to w in G if there exists a walk from v to w in G.

Theorem 1.19. In any graph $G = (V, E)$, the relation of being linked is an equivalence relation on the vertex set V. In other words:

(1) Every vertex $v \in V$ is linked to itself.

(2) If v is linked to w, then w is linked to v.

(3) If v is linked to w and w is linked to x, then v is linked to x.

Proof. For (1), we can use the walk of length 0 which consists solely of v. For (2), if $v_0, v_1, \cdots, v_\ell$ is a walk from v to w, then its reversal $v_\ell, \cdots, v_0$ is a walk from w to v. The proof of part (3) uses the fact that we can concatenate walks: if $v_0, v_1, \cdots, v_\ell$ is a walk from v to w and $w_0, w_1, \cdots, w_k$ is a walk from w to x, then $v_0, v_1, \cdots, v_\ell, w_1, \cdots, w_k$ is a walk from v to x, since $v_\ell = w = w_0$. □

As a consequence, the vertex set V is the disjoint union of the equivalence classes for the equivalence relation of being linked.

Definition 1.20. For any vertex v of a graph G, the connected component of G containing v is the graph whose vertex set consists of all vertices of G which are linked to v, where two vertices are adjacent if and only if they are adjacent in G. We say that G is connected if it has one connected component, i.e. the vertex set is nonempty and every vertex is linked to every other vertex.

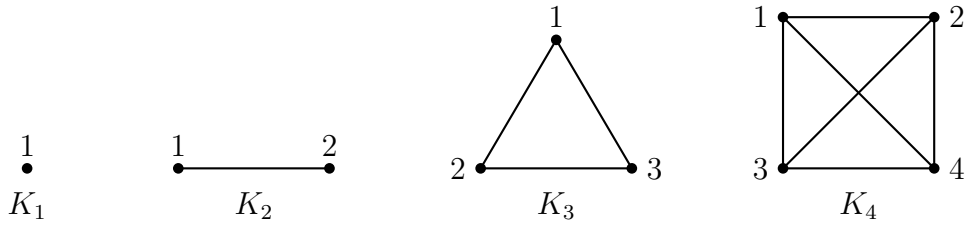
Example 1.21. *Of the two graphs in Example 1.16, the first breaks into two connected components: the vertices* $1, 3, 4$ *are all linked to each other, while the vertex* 2 *is not linked to any of these. The second graph is connected.*

At the extremes of connectedness we have the following cases.

Definition 1.22. For a fixed vertex set V, the null graph is the graph $(V, \emptyset)$ in which no two vertices are adjacent; the complete graph is the graph (V, X) where X is the set of all 2-element subsets of V, or in other words any two vertices are adjacent (this is the complement of the null graph). We write N_n and K_n for the null graph and complete graph with vertex set $\{1, 2, \cdots, n\}$.

Note that N_n has 0 edges whereas K_n has $\binom{n}{2}$, and N_n has n connected components (every vertex is a component on its own) whereas K_n is connected.

Example 1.23. *Here are the pictures of* K_n *for* $n \leq 4$*:*



The following result means that the classification of graphs up to isomorphism reduces to the case of connected graphs.

Theorem 1.24*. Let G and G' be graphs.

(1) Suppose that G and G' are isomorphic. Then G is connected if and only if G' is connected. More generally, the connected components of G must be isomorphic to corresponding connected components of G': that is, if the connected components of G are $G_1, \cdots, G_s$, it must be possible to number the connected components of G' as $G'_1, \cdots, G'_s$ in such a way that G_i is isomorphic to G'_i for all i.

(2) Conversely, if the connected components of G are $G_1, \cdots, G_s$ and those of G' are $G'_1, \cdots, G'_s$, and G_i is isomorphic to G'_i for all i, then G is isomorphic to G'.

Proof*. Let $G = (V, E)$, $G' = (V', E')$. In part (1) we are assuming the existence of a bijection $f : V \to V'$ which respects adjacency; it follows easily that v, w are linked in G if and only if $f(v), f(w)$ are linked in G'. Hence the connected components of G' are just obtained from the connected components $G_1, \cdots, G_s$ of G by applying f to each vertex set, and (1) follows. In part (2), it is clear that we can combine the individual bijections between the vertices of G_i and those of G'_i for each i into an overall bijection $f : V \to V'$. By definition there are no edges in G or G' other than those within a connected component, so f respects adjacency as required. □

Remark 1.25. *Theorem 1.24 is fiddly to state and to prove (and indeed we cheated slightly in the proof by omitting details which would have required too much notation). Yet it really expresses nothing more than the obvious idea that connected components are an intrinsic property of a graph and don't depend on the particular labelling of the vertices. There are several basic results in graph theory which are obvious once one gets the idea, yet tricky to prove rigorously. From now on we will often omit the proofs of such results.*

Connected components of a graph are a special case of the following idea.

Definition 1.26. If $G = (V, E)$ is a graph, a subgraph of G is a graph $H = (W, F)$ where W is a subset of V and F is a subset of E. The second of these statements means that for $v, w \in W$,

$$v, w \text{ are adjacent in } H \implies v, w \text{ are adjacent in } G.$$

Note that the converse is not required to hold. A subgraph with the same vertex set as the whole graph is called a spanning subgraph.

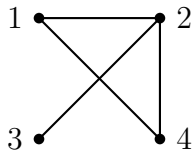
A less formal way to express this definition is to say that a subgraph is something obtained from the graph by deleting various vertices and edges, and a spanning subgraph is obtained by deleting edges only.

Example 1.27. *If e is an edge of G, you can form a spanning subgraph of G by removing e from the set of edges and leaving everything else unchanged: this subgraph will be called $G - e$. Conversely, if e is an edge of the complement $\overline{G}$ (i.e. it is a pair of non-adjacent vertices of G), then $G + e$ denotes the graph obtained by adding e to the set of edges of G; so G is a spanning subgraph of $G + e$.*

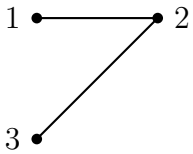
Example 1.28. *Any graph G with vertex set $\{1, 2, \cdots, n\}$ is a spanning subgraph of K_n; you can start with K_n and delete all the edges in the complement $\overline{G}$ to obtain G.*

Example 1.29. *If v is a vertex of G, the subgraph $G - v$ is obtained from G by removing v from the set of vertices and also (as is then necessary) removing all edges which ended at v. For instance,*

if G is

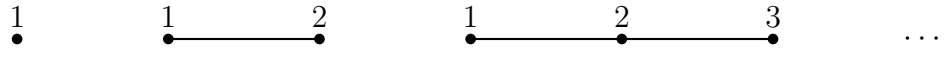


then $G - 4$ is



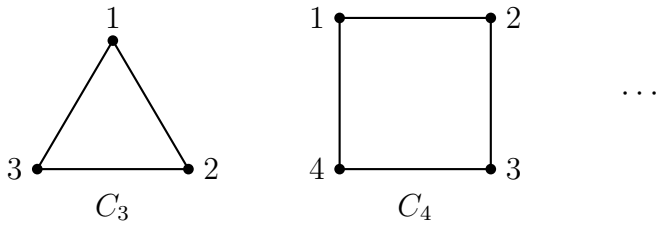
It is clear that if G is isomorphic to G', then every subgraph of G is isomorphic to a corresponding subgraph of G'. So a natural way to approach the classification of graphs is to consider the subgraphs which belong to various isomorphism classes. For example, if G has a subgraph isomorphic to K_4 (i.e. it contains four vertices which are all adjacent to each other) and G' doesn't, then G and G' are definitely not isomorphic. Here are two particularly useful kinds of subgraph to consider.

Definition 1.30. The path graph P_n has vertex set $\{1, 2, \cdots, n\}$ and edges $\{1, 2\}, \{2, 3\}, \cdots, \{n - 1, n\}$:



P_1 P_2 P_3

For $n \geq 3$, the cycle graph C_n is obtained from P_n by adding the edge $\{n, 1\}$:



C_3 C_4

Definition 1.31. For any graph G, a path in G is a subgraph of G which is isomorphic to P_n for some n: that is, a collection of distinct vertices $\{v_1, v_2, \cdots, v_n\}$ such that v_i is adjacent to v_{i+1} for $i = 1, \cdots, n-1$, together with the edges $\{v_i, v_{i+1}\}$. The length of such a path is $n-1$; its end-vertices are v_1 and v_n. A cycle in G is a subgraph of G which is isomorphic to C_n for some $n \geq 3$: that is, a collection of distinct vertices $\{v_1, v_2, \cdots, v_n\}$ such that v_i is adjacent to v_{i+1} for $i = 1, \cdots, n-1$ and v_n is adjacent to v_1, together with the edges $\{v_i, v_{i+1}\}$ and $\{v_n, v_1\}$. The length of such a cycle is n, and a cycle of length n is called an n-cycle for short.

Remark 1.32. *Notice the differences between the definition of a path or a cycle and that of a walk. Firstly, paths and cycles in a graph contain edges as well as vertices, whereas a walk as defined in Definition 1.18 consists of a sequence of vertices only (although the edges between them are involved in the definition). Secondly, the vertices of a path or cycle must all be different, whereas a walk is allowed to use a vertex more than once. Thirdly, there is no specific ordering on the vertices of a path or cycle beyond what is implied by the edges. Thus the path with vertices $v_1, \cdots, v_n$ referred to in Definition 1.31 is the same as the path with vertices $v_n, \cdots, v_1$, although there are two different walks of length $n-1$ 'along' this path, one from v_1 to v_n and one from v_n to v_1. Similarly, the cycle with vertices $v_1, \cdots, v_n$ referred to in Definition 1.31 is the same as the cycle with vertices $v_2, \cdots, v_n, v_1$, and the same as the cycle with vertices $v_3, \cdots, v_n, v_1, v_2$, and so forth. For any one of these vertices v_i, there are two different walks of length n 'along' the cycle from v_i back to itself, namely $v_i, v_{i+1}, \cdots, v_n, v_1, \cdots, v_{i-1}, v_i$ and its reversal.*

Example 1.33. *In the complete graph K_4, the number of:*

walks of length 3 *from vertex* 1 *to vertex* 4 *is* ☐

paths of length 3 *with end-vertices* 1 *and* 4 *is* ☐

walks of length 3 *from vertex* 1 *to itself is* ☐

cycles of length 3 *containing vertex* 1 *is* ☐

The numbers of paths or cycles of various lengths in a graph are obviously isomorphism-invariant, so potentially useful for classification.

There is a relationship between paths, cycles and connectedness.

Theorem 1.34. Let G be a graph, and v, w distinct vertices of G.

(1) v and w are linked in G if and only if there is a path in G with end-vertices v and w.

(2) If $e = \{v, w\}$ is an edge of G, then v and w are linked in $G - e$ if and only if there is a cycle in G containing e.

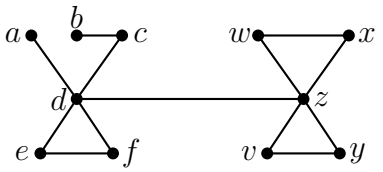
Proof*. For part (1), recall that v and w are said to be linked in G if there is a walk in G from v to w. So the "if" direction is clear: if there is a path with end-vertices v and w, then there is a walk along that path from v to w. For the "only if" direction, we assume that there is a walk from v to w in G, say $v = v_0, v_1, v_2, \cdots, v_\ell = w$. We need to produce a path with end-vertices v and w; the problem is that there may be repeated vertices in the walk. But if $v_i = v_j$ for some $i < j$, then the part of the walk between v_i and v_j is a needless detour; we can cut it out to obtain the shorter walk $v_0, \cdots, v_i, v_{j+1}, \cdots, v_\ell$. Continuing in this way, we must eventually obtain a walk from v to w with no repeated vertices, which (together with the edges between consecutive vertices) constitutes a path as required.

In part (2), if there is a cycle containing e, we may as well number its vertices $v = v_1, v_2, \cdots, v_n = w$, and then the walk $v_1, v_2, \cdots, v_n$ shows that v and w are linked without using the edge e. Conversely, suppose that v and w are linked in $G - e$. By part (1), there is a path in $G - e$ with vertices $v = v_1, v_2, \cdots, v_n = w$, and we can add the edge e to this path to form a cycle. (We cannot have $n = 2$, because then the path would contain e.) □

Definition 1.35. An edge $e = \{v, w\}$ in a graph G is said to be a bridge if v and w are not linked in $G - e$, i.e. e is not contained in any cycle. In particular, if G is connected, then e is a bridge if and only if $G - e$ is not connected, i.e. removing e disconnects the graph.

The reason for the name is that such an edge e is the only way to get from v to w, as if they were separated by water and e was the only bridge.

Example 1.36. *In the graph with the following picture*



the bridges are $\{a, d\}$, $\{b, c\}$, $\{c, d\}$, *and* $\{d, z\}$.

We can now prove a result bounding the number of edges in a graph.

Theorem 1.37. Let G be a graph with n vertices and k edges.

(1) If G is connected, then $n - 1 \leq k \leq \binom{n}{2}$.

(2) If G has s connected components, then $n - s \leq k \leq \binom{n-s+1}{2}$.

Proof*. In part (1), we already know the upper bound $k \leq \binom{n}{2}$, so the only new thing to prove is that connectedness requires at least $n - 1$ edges. We can prove this by induction on n; the $n = 1$ base case is obvious. Assume that $n \geq 2$ and that we know the result for graphs with fewer than n vertices. Suppose that G contains a bridge $e = \{v, w\}$. Since G is connected, every vertex in $G - e$ must be linked to either v or w; that is, $G - e$ has two connected components, the component G_1 containing v and the component G_2 containing w. Suppose that G_i has n_i vertices and k_i edges, for $i = 1, 2$. Then $n_1 + n_2 = n$, so $n_1, n_2 < n$; thus G_1 and G_2 are graphs to which the induction hypothesis applies, and we conclude that $k_1 \geq n_1 - 1$ and $k_2 \geq n_2 - 1$. Hence

$$k = k_1 + k_2 + 1 \geq (n_1 - 1) + (n_2 - 1) + 1 = n_1 + n_2 - 1 = n - 1.$$

On the other hand, if G does not contain a bridge, then if we delete any edge it remains connected; if we continue to delete edges, we must eventually reach a spanning subgraph which does contain a bridge. Since the number of edges in this subgraph is at least $n - 1$, the number of edges in G is even more. So in either case, $k \geq n - 1$ and the inductive step is complete.

In part (2), let $G_1, \cdots, G_s$ be the connected components of G, and suppose that G_i has n_i vertices and k_i edges, so that $n = n_1 + \cdots + n_s$ and $k = k_1 + \cdots + k_s$. By part (1) we have $k_i \geq n_i - 1$ for all i, so

$$k \geq (n_1 - 1) + \cdots + (n_s - 1) = (n_1 + \cdots + n_s) - s = n - s,$$

proving the required lower bound. We also have $k_i \leq \binom{n_i}{2}$ for all i, so

$$\begin{aligned} k &\leq \binom{n_1}{2} + \cdots + \binom{n_s}{2} \\ &= \frac{1}{2}(n_1(n_1-1) + \cdots + n_s(n_s-1)) \\ &\leq \frac{1}{2}(n-s+1)((n_1-1) + \cdots + (n_s-1)) \quad (\text{since } n_i \leq n-s+1 \text{ for all } i) \\ &= \frac{1}{2}(n-s+1)(n-s) \\ &= \binom{n-s+1}{2}, \end{aligned}$$

proving the required upper bound. □

Remark 1.38. *The upper bound in part (1) is attained exactly when G is a complete graph, isomorphic to K_n. If you examine the proof of (2), you will see that the upper bound there is attained exactly when one of the connected components of G is a complete graph and the others consist of single points. We will explore the cases when the lower bounds are attained in a later chapter.*

1.3 Degrees of vertices

One way to analyse a graph is to narrow one's focus to what is happening in the neighbourhood of each vertex. This requires some definitions.

Definition 1.39. Let G be a graph. For any vertex v of G, its degree $\deg_G(v)$ or just $\deg(v)$ is the number of edges of G which have v as an end; equivalently, the number of vertices of G which are adjacent to v. The degree sequence of G consists of the degrees of the vertices of G arranged in

weakly increasing order: if the vertices of G are numbered $v_1, v_2, \cdots, v_n$ in such a way that $\deg(v_1) \leq \deg(v_2) \leq \cdots \leq \deg(v_n)$, then the degree sequence of G is $(\deg(v_1), \deg(v_2), \cdots, \deg(v_n))$. We use some special notation for the first and last entries of this sequence:

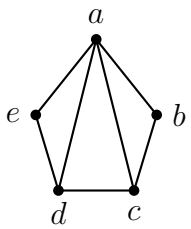
$$\begin{aligned} \delta(G) &= \text{ the smallest degree of any vertex of } G, \\ \Delta(G) &= \text{ the largest degree of any vertex of } G. \end{aligned}$$

If $\delta(G) = \Delta(G)$, i.e. every vertex has the same degree d, we say that G is regular of degree d.

It is obvious that if G has n vertices, then

$$0 \leq \delta(G) \leq \Delta(G) \leq n - 1. \tag{1.1}$$

Example 1.40. *If G is the graph with the following picture:*



then the degrees of the vertices are

$$\deg(a) = 4,\ \deg(b) = 2,\ \deg(c) = 3,\ \deg(d) = 3,\ \deg(e) = 2.$$

So the degree sequence of G is $(2, 2, 3, 3, 4)$ (note the conventional increasing order), and $\delta(G) = 2$, $\Delta(G) = 4$.

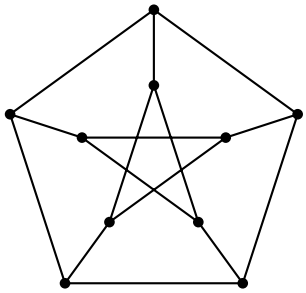
Example 1.41. *Some of our families of examples are regular:*

the null graph N_n is regular of degree $\boxed{}$

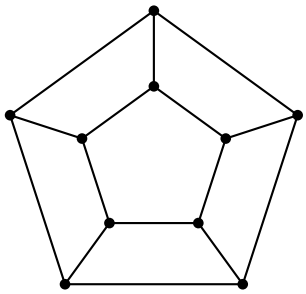
the complete graph K_n is regular of degree $\boxed{}$

the cycle graph C_n is regular of degree $\boxed{}$

Example 1.42. *There is a famous graph called the Petersen graph which has* 10 *vertices and is regular of degree* 3*:*



This is really an isomorphism class of graphs, because there is no particularly privileged way to label the vertices. In fact, the vertices play completely symmetrical roles. In the picture, the inner 5 *vertices are joined in a pentagram pattern, but that is just another way of drawing a* 5*-cycle. It is crucial that the vertices in the inner* 5*-cycle are joined to the vertices in the outer* 5*-cycle in a way which is "out of step" with the cycles. In other words, the Petersen graph is not isomorphic to the graph(s) with the following picture:*



In the latter graph there are only two 5*-cycles, the inner ring and the outer ring; but in the Petersen graph, there are many ways other than the most visible way to separate the vertices into two* 5*-cycles.*

It is obvious that isomorphic graphs have the same degree sequence, so considering the degree sequence can be a useful way to distinguish between non-isomorphic graphs. (Unfortunately, the preceding example shows that two non-isomorphic graphs can have the same degree sequence, so it is not a panacea.) The degree sequence contains some of the pieces of information we have previously used for classification: the number of vertices is the number

of terms in the degree sequence, and the number of edges can be derived from the sum of the terms, by the following result.

Theorem 1.43 (Hand-shaking Lemma)**.** For any graph, the number of edges is half the sum of the degrees of all the vertices.

Proof. This is a simple application of the Overcounting Principle: since every edge has two ends, when you add up the numbers of edges at each vertex you count every edge exactly twice. □

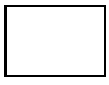
Example 1.44. *The name comes from imagining the context of a party where various people shake hands upon meeting. If you ask every one at the end of the night how many hands they shook, add up those numbers and divide by two, you get the total number of hand-shakes. (Here the vertices are the party guests, two of whom are adjacent if they shook hands.)*

Example 1.45. *Purely from the degree sequence* $(2, 2, 3, 3, 4)$ *of the graph in Example 1.40, we can see that the number of edges is* $\frac{2+2+3+3+4}{2} = 7$.

Example 1.46. *The number of edges in the Petersen graph is* $\frac{10\times 3}{2} = 15$.

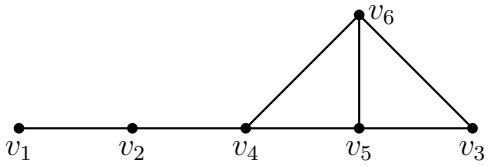
To use degree sequences in the classification of graphs, we would want a way of deciding, given a sequence $(d_1, d_2, \cdots, d_n)$ of nonnegative integers in weakly increasing order, whether there exists a graph with this as its degree sequence; for brevity, we say that the sequence is graphic if there does exist such a graph. There are some obvious requirements that $(d_1, d_2, \cdots, d_n)$ must satisfy in order to be graphic. For instance, the Hand-shaking Lemma shows that $d_1 + \cdots + d_n$ must be even, which is equivalent to saying that the number of odd d_i's is even. Another obvious requirement is that if d_n is nonzero, there must be at least d_n other d_i's which are nonzero; because if a vertex is to have degree d_n, there must be d_n other vertices which are adjacent to it. There are other such constraints.

Example 1.47. *Is the sequence* $(1, 1, 3, 3)$ *graphic? That is, does there exist a graph* G *with vertices* v_1, v_2, v_3, v_4 *such that* $\deg_G(v_1) = 1$, $\deg_G(v_2) = 1$, $\deg_G(v_3) = 3$, $\deg_G(v_4) = 3$*? The answer is*



Example 1.48. *Is the sequence* $(1, 2, 2, 3, 3, 3)$ *graphic? Let's try to actually construct a graph* G *with vertices* $v_1, v_2, \cdots, v_6$ *such that* $\deg_G(v_1) = 1$,

$\deg_G(v_2) = 2$, $\deg_G(v_3) = 2$, $\deg_G(v_4) = 3$, $\deg_G(v_5) = 3$, $\deg_G(v_6) = 3$. *We need to make v_6 adjacent to three other vertices. Crossing our fingers, we arbitrarily specify that these should be v_3, v_4, and v_5 (because it seems that we will give ourselves the best chance by using vertices whose degrees are as large as possible). With that assumption, consider the graph $G - v_6$ obtained by removing the vertex v_6 and all the edges ending at it. We have* $\deg_{G-v_6}(v_1) = 1$, $\deg_{G-v_6}(v_2) = 2$, $\deg_{G-v_6}(v_3) = 1$, $\deg_{G-v_6}(v_4) = 2$, $\deg_{G-v_6}(v_5) = 2$. *So the degree sequence of the hypothetical graph $G - v_6$ is $(1, 1, 2, 2, 2)$. But the latter sequence is definitely graphic: it is the degree sequence of the path graph P_5. So we can join the extra vertex v_6 to the appropriate vertices of P_5, and we have succeeded in constructing G:*



So the answer to the original question is that $(1, 2, 2, 3, 3, 3)$ is graphic.

To make such arguments into a hard-and-fast rule, we need the following result, which shows that the arbitrary stipulation made in Example 1.48 was not so arbitrary after all.

Theorem 1.49 (Havel-Hakimi Theorem)**.** Let $(d_1, d_2, \cdots, d_n)$ be a weakly increasing sequence of nonnegative integers. If $(d_1, d_2, \cdots, d_n)$ is graphic, then there is a graph G with vertices $v_1, v_2, \cdots, v_n$ such that $\deg(v_i) = d_i$ for $i = 1, 2, \cdots, n$ and the vertices adjacent to v_n are $v_{n-1}, v_{n-2}, \cdots, v_{n-d_n}$.

Proof*. By the definition of "graphic", there is a graph G with vertices $v_1, v_2, \cdots, v_n$ such that $\deg(v_i) = d_i$ for $i = 1, 2, \cdots, n$; all we don't know is which of the vertices $v_1, \cdots, v_{n-1}$ are adjacent to v_n. Let W be the set of vertices adjacent to v_n; thus $|W| = d_n$. Let d be the sum of the degrees (in the graph G) of all the vertices in W. Since the sequence $(d_1, d_2, \cdots, d_n)$ is weakly increasing,

$$d \leq d_{n-1} + d_{n-2} + \cdots + d_{n-d_n}. \tag{1.2}$$

If equality holds in (1.2), then the vertices in W have the same collection of degrees as $v_{n-1}, v_{n-2}, \cdots, v_{n-d_n}$; so after possibly renumbering some vertices with the same degree, we can arrange that $W = \{v_{n-1}, v_{n-2}, \cdots, v_{n-d_n}\}$ as

required. On the other hand, if the inequality in (1.2) is strict, there must be some vertex w in W whose degree is strictly less than that of some vertex $x \in \{v_1, \cdots, v_{n-1}\} \setminus W$. The inequality $\deg(w) < \deg(x)$ implies that there is some vertex $y \neq w$ which is adjacent to x but not to w. Thus v_n, w, x, y are four distinct vertices of G such that $\{w, v_n\}$ and $\{x, y\}$ are edges of G, but $\{x, v_n\}$ and $\{w, y\}$ are not. Now we modify G by deleting the edges $\{w, v_n\}$ and $\{x, y\}$, and adding the edges $\{x, v_n\}$ and $\{w, y\}$. Since each of v_n, w, x, y is involved in one of the deleted edges and one of the added edges, this modification has not changed any of the degrees; however, since w has been replaced by x in the set W, the quantity d has strictly increased. After repeating this step a finite number of times, we must reach a graph for which equality holds in (1.2), and then the result follows as seen before. □

We can deduce the following recursive rule for determining whether a sequence is graphic.

Theorem 1.50. Let $(d_1, d_2, \cdots, d_n)$ be a weakly increasing sequence of non-negative integers. We have three cases.

(1) If all the d_i's are zero, the sequence is graphic.

(2) If the number of nonzero d_i's is between 1 and d_n, the sequence is not graphic.

(3) If the number of nonzero d_i's is at least $d_n + 1$ (i.e. $d_{n-d_n} \geq 1$), let $(e_1, \cdots, e_{n-1})$ be the sequence obtained from $(d_1, d_2, \cdots, d_n)$ by carrying out the following steps:

   (a) remove the last term of the sequence, d_n;

   (b) subtract 1 from each of the next d_n terms from the end, i.e. $d_{n-1}, \cdots, d_{n-d_n}$;

   (c) if necessary, rearrange in weakly increasing order.

   Then $(d_1, d_2, \cdots, d_n)$ is graphic if and only if $(e_1, \cdots, e_{n-1})$ is graphic.

Proof*. Case (1) is obvious, because the null graph N_n has degree sequence $(0, 0, \cdots, 0)$. We have already seen the reason for case (2): a vertex of

degree d_n requires d_n other vertices of nonzero degree to be adjacent to it. So the main content of the result is the "if and only if" statement in case (3). We prove the "only if" direction first, supposing that $(d_1, d_2, \cdots, d_n)$ is graphic. By the Havel–Hakimi Theorem, there is a graph G with vertices $v_1, v_2, \cdots, v_n$ such that $\deg(v_i) = d_i$ and the vertices adjacent to v_n are $v_{n-1}, v_{n-2}, \cdots, v_{n-d_n}$. By contruction, $(e_1, \cdots, e_{n-1})$ is the degree sequence of $G - v_n$, and it is hence graphic. To prove the converse direction, we can reverse the construction: starting from a graph H with degree sequence $(e_1, \cdots, e_{n-1})$, we construct a graph G with degree sequence $(d_1, \cdots, d_n)$ by adding a new vertex which is adjacent to the appropriate vertices of H. □

In case (3), Theorem 1.50 does not immediately decide whether $(d_1, \cdots, d_n)$ is graphic or not, but it reduces the question to the same question for the shorter sequence $(e_1, \cdots, e_{n-1})$; we can then apply Theorem 1.50 again to this shorter sequence and so on, and since the sequence can't go on getting shorter indefinitely, we must eventually land in either case (1) or case (2). (Of course, we can stop before that point if the sequence becomes obviously graphic or not graphic by some other reasoning.)

Example 1.51. *The sequence* $(1, 1, 1, 1, 4, 4)$ *falls into case (3). To obtain the shorter sequence, we apply the three steps prescribed in Theorem 1.50:*

$$(1, 1, 1, 1, 4, 4) \longrightarrow (1, 1, 1, 1, 4) \longrightarrow (1, 0, 0, 0, 3) \longrightarrow (0, 0, 0, 1, 3).$$

The sequence $(0, 0, 0, 1, 3)$ *falls into case (2), so it is not graphic; hence the original sequence* $(1, 1, 1, 1, 4, 4)$ *is not graphic.*

Remark 1.52. *Since Havel and Hakimi provided the solution to the basic question of whether a sequence is the degree sequence of a graph, there has been much research on more refined questions: for instance, replacing "graph" by "connected graph". Unfortunately we will have to omit these later developments.*

Chapter 2

Special Walks in Graphs

The origin of graph theory was a question posed to the great 18th-century mathematician Euler by the citizens of Königsberg. Their city was built around two islands in the river Pregel, and included seven bridges joining the islands to the banks and to each other; they wanted to know whether it was possible to walk through the city crossing every bridge exactly once. To decide this, Euler was led to introduce the concept of a graph. Many subsequent applications of graph theory also involve special kinds of walks.

2.1 Eulerian graphs

Definition 2.1. An Eulerian circuit in a connected graph G is a walk in G, say $v_0, v_1, v_2, \cdots, v_\ell$, which:

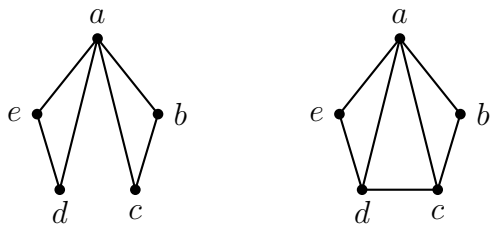
(1) uses every edge exactly once, i.e. for every $\{v, w\}$ in V there is exactly one i such that $\{v_i, v_{i+1}\} = \{v, w\}$;

(2) returns to its starting point, i.e. $v_\ell = v_0$.

A connected graph G is said to be Eulerian if it has an Eulerian circuit.

It is clear that if an Eulerian circuit does exist, you can make it start (and therefore also finish) at any vertex you want.

Example 2.2. *Consider the following two graphs.*



The first is clearly Eulerian: an example of an Eulerian circuit is the walk a, b, c, a, d, e, a. *But if you try to find an Eulerian circuit in the second graph, you will get stuck: the extra edge* $\{c, d\}$ *can't be fitted into the walk without repeating some other edge. The best you can do is a walk such as* d, e, a, d, c, a, b, c, *which uses every edge exactly once but doesn't return to its starting point. So the second graph is not Eulerian.*

Example 2.3. *The cycle graph* C_n *is obviously Eulerian: a walk around the cycle is an Eulerian circuit.*

Remark 2.4. *We restricted to connected graphs in Definition 2.1, because if a graph has more than one connected component which contains an edge, there clearly cannot be a walk which uses all the edges of the graph.*

Remark 2.5*. *We may seem to be avoiding some interesting examples of this problem by not allowing multiple edges in our graphs. Indeed, the bridges of Königsberg formed a multigraph and not a graph, because some of the landmasses were joined by more than one bridge. But given any multigraph, the question of whether it has an Eulerian circuit can be rephrased in terms of a graph, namely the one where you sub-divide every multiple edge into two edges by putting a new vertex in the middle of it.*

The reason that the first graph in Example 2.2 is so obviously Eulerian is that its edges can be split up into two 3-cycles. Euler realized that this property is crucial to the existence of an Eulerian circuit, and that it depends on the evenness of the vertex degrees; what is wrong with the vertices c and d in the second graph in Example 2.2 is that their degrees are odd.

Theorem 2.6. A connected graph G is Eulerian if and only if $\deg(v)$ is even for every vertex v of G.

Proof. We first prove the "only if" direction, that in an Eulerian graph all the degrees are even. Let $v_0, v_1, \cdots, v_\ell$ be an Eulerian circuit. A vertex v may appear more than once in this circuit, but every time it does appear, two of the edges ending at v are used: namely, $\{v_{i-1}, v_i\}$ and $\{v_i, v_{i+1}\}$ if $v = v_i$, or $\{v_0, v_1\}$ and $\{v_{\ell-1}, v_\ell\}$ if $v = v_0 = v_\ell$. Since every edge is used exactly once, the total number of edges ending at v must be even.

To prove the "if" direction, we use induction on the number of edges. The base case of this induction is the case where there are no edges at all, i.e. the graph consists of a single vertex; this graph is trivially Eulerian (the Eulerian circuit has length 0). So we assume that G does have some edges, and that all the vertex degrees are even. Our induction hypothesis says that every connected graph with fewer edges than G and all degrees even is Eulerian, and we want to prove that G is also. We start by showing that G contains a cycle. Pick any vertex v_0; since G is connected, there must be some other vertex v_1 which is adjacent to v_0; since $\deg(v_1)$ can't be 1, there must be some vertex $v_2 \neq v_0$ which is adjacent to v_1; since $\deg(v_2)$ can't be 1, there must be some vertex $v_3 \neq v_1$ which is adjacent to v_2; and we can continue in this way indefinitely. Because there are only finitely many vertices, there must be some $k < \ell$ such that $v_k = v_\ell$, and we can assume that $v_k, v_{k+1}, \cdots, v_{\ell-1}$ are all distinct. The subgraph C which consists of these distinct vertices and the edges between them (including the edge $\{v_{\ell-1}, v_k\}$) is a cycle.

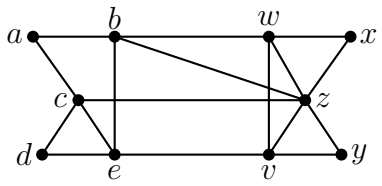
Now let H be the subgraph of G obtained by deleting all the edges in C (but keeping all the vertices). The vertex degrees in H are the same as those in G, except that 2 is subtracted from the degree of every vertex in C; so the vertex degrees in H are all even. H need not be connected, but if we let $H_1, H_2, \cdots, H_s$ be the connected components of H, then each H_i is Eulerian by the induction hypothesis. Since G is connected, every H_i must contain at least one of the vertices $v_k, \cdots, v_{\ell-1}$ of C; let k_i be such that v_{k_i} lies in H_i, and renumber if necessary so that $k_1 < k_2 < \cdots < k_s$. Then we can construct an Eulerian circuit in G as follows:

$$v_k, \overline{\cdots}, v_{k_1}, \widehat{\cdots}, v_{k_1}, \overline{\cdots}, v_{k_2}, \widehat{\cdots}, v_{k_2}, \cdots, v_{k_s}, \widehat{\cdots}, v_{k_s}, \overline{\cdots}, v_{\ell-1}, v_k,$$

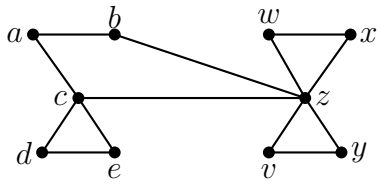
where $\overline{\cdots}$ indicates a walk around the appropriate part of the cycle C, and $\widehat{\cdots}$ indicates an Eulerian circuit in the appropriate H_i starting and finishing at v_{k_i}. This completes the induction step. □

Theorem 2.6 not only gives us a simple criterion for the existence of an Eulerian circuit; its proof supplies a recursive way of finding one.

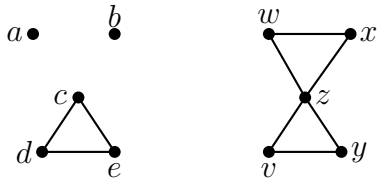
Example 2.7. *Let G be the following graph, in which all degrees are even.*



Theorem 2.6 guarantees that an Eulerian circuit exists. To use the construction in the proof, we need to choose a cycle C, say the 4-cycle with vertices b, w, v, e. After removing the edges of C, we obtain the graph H:



We now need to find an Eulerian circuit in H; again, the first step is to remove a cycle, say the 4-cycle with vertices a, b, z, c.



This leaves four connected components, and Eulerian circuits in these are obvious: the length-0 walks a and b in those components, the walk c, d, e, c, and the walk z, v, y, z, w, x, z. We now patch these into the appropriate places in the cycle a, b, z, c to form an Eulerian circuit in H:

$$a, b, z, v, y, z, w, x, z, c, d, e, c, a.$$

Finally, we make this circuit start at b instead, and patch it into the original cycle b, w, v, e to form an Eulerian circuit in G:

$$b, z, v, y, z, w, x, z, c, d, e, c, a, b, w, v, e, b.$$

Of course, we made many choices along the way, so the result is not unique.

There is an easy variant of Theorem 2.6 for the case of an Eulerian trail, which is a walk that uses every edge exactly once but finishes at a different vertex from where it started. (We saw such a walk in Example 2.6.)

Theorem 2.8. A connected graph G has an Eulerian trail if and only if it has exactly two vertices of odd degree. Moreover, if this the case, the trail must start at one of these vertices of odd degree and finish at the other.

Proof. The "only if" direction, that a graph with an Eulerian trail must have exactly two vertices of odd degree, works in the same way as in the proof of Theorem 2.6. The difference is that the first vertex and last vertex of the trail are no longer the same, so their appearances at the beginning and end of the trail use up only 1 of their edges, and their total number of edges ends up odd. Notice that this proves the second sentence in the statement: the first and last vertex of the trail must be the two odd-degree vertices.

For the "if" direction, suppose that G is connected and has exactly two vertices of odd degree, say v and w. If v and w are not adjacent, then we can add the edge $\{v, w\}$ to form the graph $G + \{v, w\}$; since this increases the degrees of v and w by 1, all vertices in $G + \{v, w\}$ have even degree. By Theorem 2.6, there is an Eulerian circuit in $G + \{v, w\}$. Since this circuit includes the edge $\{v, w\}$ at some point, we may as well suppose that it starts $v, w, \cdots, v$; we then obtain an Eulerian trail $w, \cdots, v$ in G by deleting $\{v, w\}$ from the circuit. If v and w are adjacent in G, the argument is similar, but instead we add a new vertex x as well as the edges $\{v, x\}$ and $\{w, x\}$. (This is to avoid having multiple edges.) Since all the degrees in the new graph are even, it has an Eulerian circuit. Since this circuit visits x only once, we may as well suppose that it starts $v, x, w, \cdots, v$; again, deleting the superfluous edges gives an Eulerian trail in G. □

Example 2.9. *What is the condition on n for the graph K_n to have:*

an Eulerian circuit? ☐

an Eulerian trail? ☐

neither? ☐

2.2 Hamiltonian graphs

In many real-world applications of graph theory, the vertices of the graph represent tasks which have to be carried out, and the edges represent the possible transitions between them. The Eulerian circuit concept is then not so relevant, because carrying out all the tasks is more important than making all the possible transitions. Here is a concept which is more relevant to such interpretations.

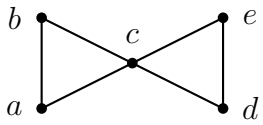
Definition 2.10. A connected graph G with ≥ 3 vertices is Hamiltonian if it has a walk which visits every vertex exactly once and then returns to its starting point.

Since such a walk never repeats a vertex except for returning to where it began, it can't possibly repeat an edge (here we need the assumption that there are more than 2 vertices). So the vertices and edges involved in the walk form a cycle. Thus we can rephrase the definition more simply: G is Hamiltonian if and only if it contains a spanning cycle (recall that "spanning" just means "using all the vertices of G"). In other words, G is Hamiltonian if and only if you can delete some of the edges of G (without deleting any of the vertices) and obtain a graph isomorphic to C_n, where n is the number of vertices of G. More precisely, "some of the edges" could read "all but n of the edges", since C_n has n edges. Despite the fact that this condition is as easy to state as the existence of an Eulerian circuit, it is significantly more difficult to give a criterion for when it holds.

One obvious principle is that if you add more edges to a Hamiltonian graph, it remains Hamiltonian. Heuristically speaking, the more edges G has, the more likely it is that n of them form a cycle.

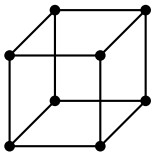
Example 2.11. *For any $n \geq 3$, the complete graph K_n is certainly Hamiltonian. In fact, since all $\binom{n}{2}$ possible edges are present, you can write down the vertices in any order you like, and walk from the first to the second to the third and so on to the last and back to the first again. So there are $n!$ walks which satisfy the Hamiltonian condition. (The number of spanning cycles is $\frac{n!}{2n} = \frac{(n-1)!}{2}$, because every spanning cycle gives rise to $2n$ different Hamiltonian walks, as explained in Remark 1.32.)*

Example 2.12. *The following graph is not Hamiltonian:*

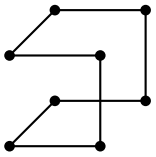


Indeed, the only two cycles in this graph are the visible 3-cycles, so there is no cycle containing all the vertices. (There is an obvious Eulerian circuit a, b, c, d, e, c, a, but that is not a cycle because it repeats the vertex c.)

Example 2.13. *The graph of vertices and edges of a cube is Hamiltonian:*

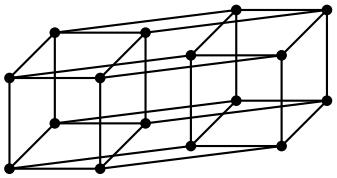


contains the spanning cycle



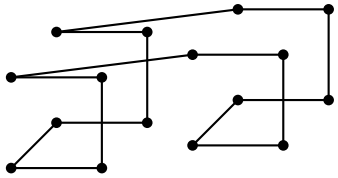
The same holds for the other regular polyhedra: tetrahedron, octahedron, dodecahedron, and icosahedron. The origin of the term "Hamiltonian" was a puzzle invented by the Irish mathematician Hamilton, in which you had to find a spanning cycle in the vertices and edges of a dodecahedron.

Example 2.14*. *Generalizing the ordinary 3-dimensional cube, there is an m-dimensional cube for any positive integer m, which sits in $\mathbb{R}^m$. Its vertices are the m-tuples $(x_1, x_2, \cdots, x_m)$ where every x_i is either 0 or 1 (hence there are 2^m vertices). Two such m-tuples are adjacent if and only if they differ in exactly one coordinate. Forgetting about the m-dimensional object itself, the vertices and edges form a graph called the cube graph Q_m. These cube graphs have a natural recursive structure: the vertices where the last coordinate is 0 form a subgraph isomorphic to Q_{m-1}, as do the vertices where the last coordinate is 1. So we can construct Q_m by taking two copies of Q_{m-1} and joining corresponding vertices. When $m = 3$, this is the standard way to draw a perspective picture of a cube: draw two squares and join the corresponding vertices. Here is a picture of Q_4 constructed in the same way:*

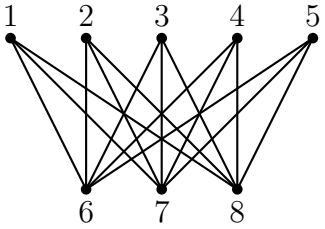


We can prove by induction that Q_m is Hamiltonian for all $m \geq 2$. The base case is clear, because Q_2 (the graph of vertices and edges of a square) is

isomorphic to C_4. Assume that $m \geq 3$ and that Q_{m-1} is Hamiltonian. Then in Q_m, there is a cycle containing all the vertices with last coordinate 0, and a corresponding cycle containing all the vertices with last coordinate 1. We then build a spanning cycle by removing an edge from the first cycle and the corresponding edge from the second cycle, and joining the cycles together along the edges between the corresponding affected vertices. For example, in the above picture of Q_4 we can find the following spanning cycle.



Example 2.15. *The complete bipartite graph $K_{p,q}$, for positive integers p and q, is the graph with vertex set $\{1, 2, \cdots, p+q\}$ where any vertex $m \leq p$ is adjacent to any vertex $m' > q$, but no two vertices $\leq p$ are adjacent, nor are any two vertices $> p$. Thus the vertices fall into two parts (hence the name "bipartite"), $\{1, 2, \cdots, p\}$ and $\{p+1, p+2, \cdots, p+q\}$, where there are no edges between vertices in the same part but every possible edge between vertices in different parts. For instance, here is a picture of $K_{5,3}$.*



In every cycle in $K_{p,q}$, the vertices must alternate between one part and the other. So it is impossible for there to be a spanning cycle unless there are the same number of vertices in both parts, i.e. $p = q$. This shows that when $p \neq q$, $K_{p,q}$ and all its spanning subgraphs are not Hamiltonian. Up to isomorphism, a spanning subgraph of $K_{p,q}$ is any graph where the vertices can be divided into two parts of sizes p and q in such a way that no two vertices in the same part are adjacent; such graphs are called bipartite. So any bipartite graph where the two parts have different numbers of vertices is not Hamiltonian. (This doesn't mean, by the way, that a bipartite graph where the two parts have the same number of vertices is Hamiltonian; for $p \geq 2$, the complete bipartite graph $K_{p,p}$ is obviously Hamiltonian, but subgraphs of it need not be.)

Here is a necessary condition for a graph to be Hamiltonian.

Theorem 2.16. Let G be a Hamiltonian graph.

(1) For any vertex v of G, the graph $G - v$ is connected.

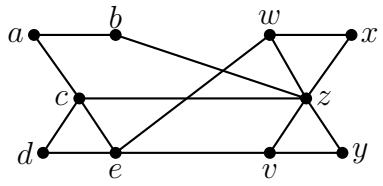
(2) More generally, for any nonempty subset S of the set of vertices of G, the graph $G - S$ obtained by deleting all the vertices in S and their edges has at most $|S|$ connected components.

Proof. Since G is Hamiltonian, it contains a spanning cycle C. For any vertex v of G, v is also a vertex of C, and the graph $C - v$ is a path and hence connected. But then $G - v$, which has all the edges of $C - v$ and more, is also connected; this proves part (1).

The proof of part (2) is similar: instead of the fact that $C - v$ is connected, we use the fact that $C - S$ has at most $|S|$ connected components, which is clear from a picture of a cycle; this implies that $G - S$ must have at most $|S|$ connected components also. □

Unfortunately, part (2) is not a sufficient condition for a graph to be Hamiltonian. So this result can be used to prove that a graph is not Hamiltonian (if it fails to satisfy (2)), but not to prove that a graph is Hamiltonian.

Example 2.17. *Let G be the following graph.*



A nonempty set S of vertices such that $G - S$ has more than $|S|$ connected components is [ ]. *Hence G is not Hamiltonian.*

Example 2.18*. *The Petersen graph of Example 1.42 does satisfy the condition that if you delete any m vertices, the result has at most m connected components. This is clear from the way the vertices are divided into two*

linked 5-cycles: if you delete only vertices in the outer 5-cycle or only vertices in the inner 5-cycle the graph stays connected; and if you delete $m_1 \geq 1$ outer vertices and $m_2 \geq 1$ inner vertices, then the outer cycle breaks into at most m_1 components and the inner cycle breaks into at most m_2 components, so the graph as a whole has at most $m_1 + m_2$ components. But the Petersen graph is not Hamiltonian, as may be seen as follows. Suppose there is a cycle C in the Petersen graph using all 10 vertices, and hence having 10 edges. At each vertex, C must use two of the three edges: in particular, at each outer vertex, C must use at least one of the two edges in the outer cycle. So C must contain either 4 of the edges of the outer cycle, or 3 of the edges (two meeting at a vertex and the other disjoint from these). The choice of which edges in the outer cycle belong to C determines which outer-inner edges belong to C, and hence which edges in the inner cycle belong to C; it is easy to check that both cases result in contradictions.

In the other direction, here is a sufficient, but not necessary, condition for a graph to be Hamiltonian.

Theorem 2.19 (Ore's Theorem)**.** Let G be a connected graph with n vertices where $n \geq 3$. If every non-adjacent pair of vertices $\{v, w\}$ satisfies $\deg(v) + \deg(w) \geq n$, then G is Hamiltonian.

Proof*. The proof is by contradiction, so assume that G is not Hamiltonian. Then G certainly can't be a complete graph, so there is some non-edge e which we could add to form the graph $G + e$. If this too is not Hamiltonian, there must be some other non-edge we can add, and so forth; there must be some point at which adding another edge changes the graph from non-Hamiltonian to Hamiltonian, because the complete graph is Hamiltonian. Since adding edges can only increase the degrees, any graph obtained from G by adding edges still has the property that every non-adjacent pair of vertices $\{v, w\}$ satisfies $\deg(v) + \deg(w) \geq n$. So we can rename our graphs if necessary so that G itself has the property that when you add some non-edge it becomes Hamiltonian.

Our situation now is that G is not Hamiltonian, but for some non-adjacent pair of vertices $\{v, w\}$, $G + \{v, w\}$ is Hamiltonian. Let C be a spanning cycle of $G + \{v, w\}$; since C is not contained in G, it must use the edge $\{v, w\}$. So $C - \{v, w\}$ is a spanning path in G with end-vertices v and w.

Let $v_1 = v, v_2, v_3, \cdots, v_n = w$ be the vertices of G in the order in which they occur along this path; thus v_i is adjacent to v_{i+1} for $i = 1, 2, \cdots, n-1$. We also know that v_1 is not adjacent to v_n, and that $\deg(v_1) + \deg(v_n) \geq n$. We want to use this information to show that there is a spanning cycle in G, which contradicts the fact that G is not Hamiltonian.

The trick is to consider the sets

$$A = \{i \,|\, 2 \leq i \leq n-2,\ v_1 \text{ is adjacent to } v_{i+1}\},$$
$$B = \{i \,|\, 2 \leq i \leq n-2,\ v_n \text{ is adjacent to } v_i\}.$$

Since all the vertices adjacent to v_1 except v_2 are of the form v_{i+1} for some $i \in A$, $|A| = \deg(v_1) - 1$. Similarly all the vertices adjacent to v_n except v_{n-1} are of the form v_i for some $i \in B$, so $|B| = \deg(v_n) - 1$. Moreover, $A \cup B \subseteq \{2, 3, \cdots, n-2\}$, so $|A \cup B| \leq n-3$. Hence

$$|A \cap B| = \deg(v_1) + \deg(v_n) - 2 - |A \cup B| \geq n - 2 - (n-3) = 1.$$

That is, there is some i such that $i \in A$ and $i \in B$, i.e. $2 \leq i \leq n-2$, v_1 is adjacent to v_{i+1}, and v_n is adjacent to v_i. But then the walk

$$v_1, v_{i+1}, v_{i+2}, \cdots, v_n, v_i, v_{i-1}, \cdots, v_1$$

traverses a spanning cycle in G, as required. □

Example 2.20. *The condition in Ore's Theorem is certainly satisfied if $\delta(G) \geq \frac{n}{2}$ (that is, every vertex has degree at least $\frac{n}{2}$). The statement that graphs with this property are Hamiltonian is known as Dirac's Theorem (it was proved earlier than the more general result of Ore).*

Example 2.21. *To see that the condition in Ore's Theorem is not necessary for a graph to be Hamiltonian, consider C_5: it is clearly Hamiltonian, but the sum of the degrees of any two non-adjacent vertices is* 4.

2.3 Minimal walks in weighted graphs

So far we have been treating all the edges of a graph on the same footing, but in many applications you need to weight the edges to take account of their

cost, or distance, or whatever factor is important in the problem at hand. Then you would want to find special walks in the graph which had as small a total weight as possible.

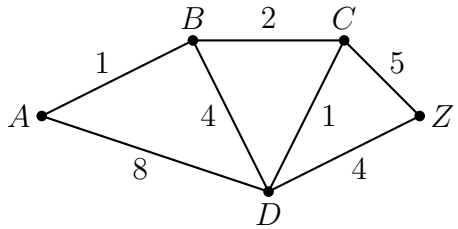
Definition 2.22. A weighted graph is a triple (V, E, f) where (V, E) is a graph and f is a function from E to the positive real numbers. For every $e \in E$, $f(e)$ is called the weight of e. Then the weight of a subgraph H of G is the sum of the weights of the edges of H; the weight of a walk $v_0, v_1, \cdots, v_\ell$ in G is the sum of the weights of the edges $\{v_i, v_{i+1}\}$. For any vertices $v, w \in V$, the distance $d(v, w)$ from v to w is defined to be the minimum weight of a walk in G from v to w (or ∞ if v and w are not linked). A walk which achieves this minimum weight is said to be a minimal walk from v to w.

Note that $d(v, v) = 0$ (the length-0 walk from v to itself has weight 0). It is clear that any minimal walk from v to w must be a walk along a path, since visiting a vertex twice would increase the weight needlessly. So $d(v, w)$ could also be defined as the minimal weight of a path with end-vertices v and w.

Remark 2.23. *An ordinary un-weighted graph can be viewed as a weighted graph by saying that all edges have weight* 1. *Then the weight of a walk is just its length, and the distance $d(v, w)$ is just the smallest number of edges you have to traverse to get from v to w. The term "distance" should arguably be restricted to this un-weighted context; its use in the weighted context is influenced by those applications where the weights of the edges are literally distances (between cities, and so forth).*

We will represent weighted graphs by labelling each edge with its weight (with no attempt to make the visual length of the edge proportional to it).

Example 2.24. *Consider the following weighted graph.*



To find $d(A, D)$, we need to consider walks from A to D (and we need not consider walks which repeat a vertex). The walk along the edge $\{A, D\}$ itself

has weight 8, which is bettered by the walk A, B, D with weight $1 + 4 = 5$ and the walk A, B, C, D with weight $1 + 2 + 1 = 4$. It is easy to see that no other walk from A to D has weight ≤ 4, so $d(A, D) = 4$. Similarly, $d(A, C) = 1 + 2 = 3$, with the walk A, B, C being the unique minimal walk from A to C. We can use this information to deduce $d(A, Z)$ as follows: every walk from A to Z has to have either C or D as its second-last vertex, since those are the only vertices adjacent to Z. Thus

the minimum weight of a walk $A, \cdots, C, Z$ is ☐

the minimum weight of a walk $A, \cdots, D, Z$ is ☐

the distance $d(A, Z)$ is ☐

the number of minimal walks from A to Z is ☐

Remark 2.25. *Notice that in Example 2.24, since the edge $\{A, D\}$ does not give a minimal walk from A to D, it can never occur in any minimal walk: it is always preferable to get from A to D via B and C. So for the purposes of minimal walks, the edge $\{A, D\}$ may as well be deleted. From the opposite point of view, you could imagine that we are always dealing with complete graphs, where some of the edges have such large weights that they are guaranteed never to be used.*

When dealing with a large graph (presumably stored in computer memory rather than pictured), it is impractical to search all possible walks from one vertex to another in order to find a minimal walk: if the graph is complete with n vertices, there are more than $n!$ possible walks from one vertex to another, and $n!$ grows super-exponentially as n increases. A better approach is to do what we did in Example 2.24 to find minimal walks from A to Z: first find minimal walks from A to vertices 'closer' to A than Z is. Here is an algorithm which uses this idea to find $d(A, v)$ for every vertex v.

DIJKSTRA'S ALGORITHM. Given a weighted graph (V, E, f) and a vertex A, the algorithm assigns temporary labels to every vertex, which are possibly changed in the course of the algorithm until they are declared to be permanent. The steps are:

(1) Give A the permanent label 0.

(2) If every $v \in V$ has a permanent label, stop.

(3) Let v be the vertex which has most recently acquired a permanent label, and let ℓ be that label.

(4) For every vertex w which is adjacent to v, do the following:

   (a) if w has a permanent label, leave it unchanged;

   (b) if w has no label, give it the temporary label $\ell + f(\{v, w\})$;

   (c) if w has a temporary label k, compare k with $\ell + f(\{v, w\})$, and change the label to $\ell + f(\{v, w\})$ if that is smaller than k.

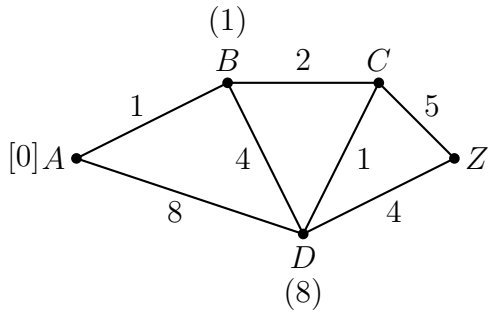
(5) Of all the vertices which have temporary labels, choose one whose label is smallest, and make that the permanent label of that vertex.

(6) Return to Step (2).

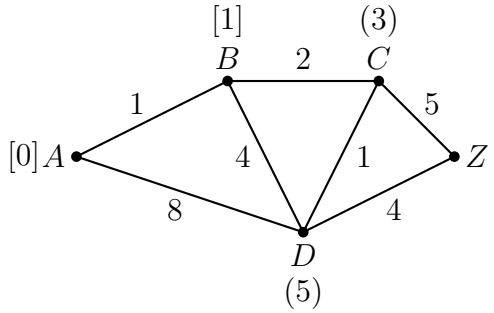
At the end of this algorithm, the permanent label of every vertex v is guaranteed to be $d(A, v)$. We will not give a rigorous proof, but here is the idea. At any point in the algorithm, the label of a vertex w is the minimum weight of all walks from A to w which you have 'considered so far' (only implicitly considered, because the point of having the algorithm is that you don't actually need to consider all these walks). The label becomes permanent at a point where you are certain that there are no other walks with smaller weights. In Step (4), you are 'considering' walking from A to w by first doing a minimal walk from A to v and then using the edge $\{v, w\}$. This walk has weight $\ell + f(\{v, w\})$, so you have to decide whether that is smaller than the previously known minimum k; if it is, it becomes the new minimum.

Example 2.26. *Let us apply the algorithm to the graph in Example 2.24. (Of course, there is no need for the algorithm in such a small example, but it will illustrate the general procedure.) We will put temporary labels in parentheses and permanent labels in square brackets. The first time we come to Step (4), v is A itself which has the label* [0]*; neither of the adjacent*

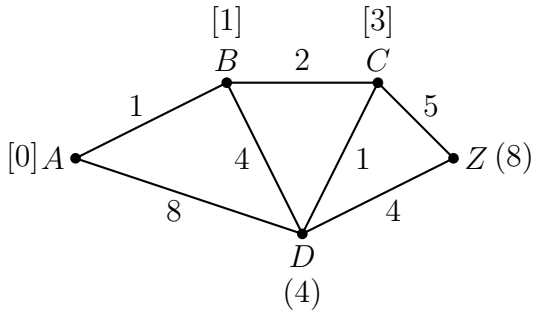
vertices has a label yet, so they both get temporary labels under case (b).



Step (5) then makes the label of B permanent, since it is the smallest temporary label. Thus the next time we come to Step (4), v is B. Of the vertices adjacent to B, vertex A is left alone under case (a); vertex C now gets a temporary label under case (b); and vertex D gets a new temporary label under case (c), because $1 + 4$ is smaller than 8.



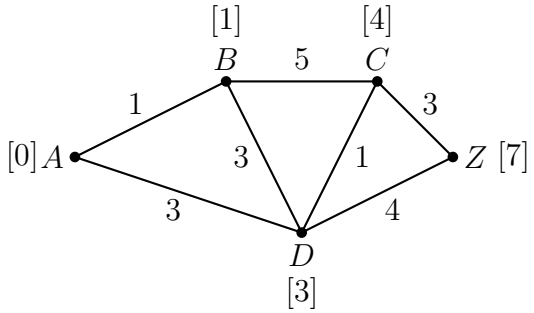
Step (5) then makes the label of C permanent, so it is the new v; the next pass through Step (4) gives Z a label, and changes the label of D again.



Step (5) then makes the label of D permanent; the next pass through Step (4) leaves the label of Z unchanged, and the final Step (5) makes that permanent also. The labels now display the distances from A to each vertex.

A good feature of Dijkstra's algorithm is that it not only tells you the minimum weight of a walk from A to Z, it allows you to find all walks which attain this minimum. The reason is that every minimal walk from A to Z must be of the form $A, \cdots, Y, Z$ where $A, \cdots, Y$ is a minimal walk from A to Y. Moreover, the only possible vertices Y which can occur in this way are the vertices which are adjacent to Z and satisfy $d(A, Y) + f(\{Y, Z\}) = d(A, Z)$. We can then use the same principle to find the minimal walks from A to Y; there is no chance of getting into a loop in this recursion, because $d(A, Y)$ is less than $d(A, Z)$.

Example 2.27. *Consider the following weighted graph, to which Dijkstra's algorithm has already been applied.*



To find all the minimal walks from A to Z, we note that the second-last vertex Y could be either C or D, because both of these satisfy $d(A, Y)+f(\{Y, Z\}) = d(A, Z)$. In the minimal walks from A to C, the second-last vertex can only be D, because

$$d(A, B) + f(\{B, C\}) = 1 + 4 > 4,\ d(A, Z) + f(\{Z, C\}) = 7 + 3 = 10 > 4.$$

Similarly, in the minimal walks from A to D the second-last vertex can only be A itself. Thus

the unique minimal walk from A to D is A, D;

the unique minimal walk from A to C is A, D, C;

so the two minimal walks from A to Z are A, D, Z and A, D, C, Z.

Note that this method finds the walks 'backwards'. It is tempting to try to find minimal walks 'forwards' by starting at A and following the smallest-weight edges; but Example 2.27 shows that this wouldn't always work. (To make it workable we would need to know the distances $d(B, Z)$ for all vertices B, i.e. we would need to have run Dijkstra's algorithm for Z instead.)

Remark 2.28. *We will not go into the details of running times for these algorithms, but it is easy enough to see that if n is the number of vertices, both Dijkstra's algorithm and the above method of finding minimal walks have running times which grow no faster than n^2. The reason is that they both essentially consist of a loop which is executed no more than n times, and in each pass through the loop, any vertex is considered at most once.*

Here are two much-studied problems which combine the minimal-weight property with the Eulerian and Hamiltonian properties.

Chinese Postman Problem. Given a connected weighted graph, find a walk which uses every edge, returns to its starting point, and has minimum weight subject to these two conditions.

Travelling Salesman Problem. Given a connected weighted graph, find a walk which visits every vertex, returns to its starting point, and has minimum weight subject to these two conditions.

The names come from the imagined situations of a postman doing the rounds of various streets, and a salesman visiting various potential clients. (The postman is Chinese in honour of the graph theorist Mei-Ko Kuan.) As with the Eulerian and Hamiltonian problems, the starting point can be arbitrary.

Notice that in the Chinese Postman Problem, the walk has to use every edge of the graph, but is allowed to use an edge more than once. It is obvious that the weight of such a walk is at least the sum of the weights of all the edges. So if the graph is Eulerian, the solutions of the Chinese Postman Problem are exactly the Eulerian circuits (which we already have a way of finding). If the graph is not Eulerian, i.e. it has some vertices of odd degree, then there will inevitably be some back-tracking in the walk. The idea of the general solution is to find the distances between the odd-degree vertices by Dijkstra's algorithm, and use that information to decide what is the most economical back-tracking to do. Here is a precise result in the case of graphs which have an Eulerian trail (see Theorem 2.8).

Theorem 2.29. Let G be a connected graph with exactly two vertices of odd degree, say v and w. Then an Eulerian trail from v to w, followed by a minimal walk from w to v, is a solution of the Chinese Postman Problem.

Proof.** Let E be the set of edges of G; then the proposed walk has weight $\sum_{e \in E} f(e) + d(v, w)$. We need to prove that any walk in G which returns to its starting point and uses every edge has weight at least as large as this. The proof is easier if we allow ourselves to talk about multigraphs. Imagine a duplicate copy of the set of vertices in G, on which we construct a multigraph G' by following the walk in G and drawing an edge in G' as the corresponding edge in G is used – thus edges in G which are used more than once correspond to multiple edges in G'. We also give each edge in G' the same weight as the corresponding edge in G. Our walk in G then corresponds to an Eulerian circuit in the multigraph G', and its weight is the weight of G'. For the same reason as for ordinary graphs, the degrees of an Eulerian multigraph are all even, so every degree in G' is even (we continue to define the degree of u as the number of edges ending at u). Now we form a new (possibly not connected) multigraph G'' by deleting from G' one copy of every edge of G. The weight of G'' expresses the amount by which the weight of our walk exceeds $\sum_{e \in E} f(e)$, so we need to prove that the weight of G'' is at least $d(v, w)$. It clearly suffices to prove that there is a walk from v to w in G''. Since $\deg_{G''}(u) = \deg_{G'}(u) - \deg_G(u)$ for every vertex u, we know that $\deg_{G''}(v)$ and $\deg_{G''}(w)$ are odd while $\deg_{G''}(u)$ is even for every other vertex u. So we construct our new walk in G'' as follows. Starting from v, we use any edge from v to an adjacent vertex v_1 in G'' (there must be such an edge since $\deg_{G''}(v)$ is odd); if the vertex v_1 is not w, then we can imagine deleting the edge we used, which makes $\deg_{G''}(v)$ even and $\deg_{G''}(v_1)$ odd, and then we can continue the walk with any edge from v_1 to an adjacent vertex v_2, and so on; we must eventually reach w, because we can't go on deleting edges indefinitely. □

Example 2.30. *Return to the graph from Example 2.27, in which B and C are the odd-degree vertices.*

An Eulerian trail from B to C is 

A minimal walk from C to B is 

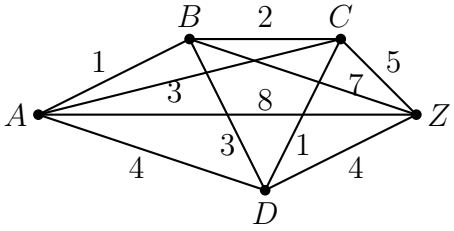
A Chinese Postman solution is 

Its weight is 

Remark 2.31*. *If a graph G with n vertices satisfied the condition of Theorem 2.29, how long would it take to actually find a solution of the Chinese Postman Problem, for large n? The time needed to determine which vertices have odd degree, if that is included, grows no faster than n^2, since the maximum number of edges is $\binom{n}{2}$. As noted in Remark 2.28, a minimal walk from w to v can be found in a time which also grows no faster than n^2. The time needed to find an Eulerian trail from v to w is probably the most significant part: if you do it by decomposing the graph $G + \{v, w\}$ into cycles, then the time needed conceivably has cubic growth, since the time taken to find each cycle could have the growth rate of n^1, and the number of cycles could have the growth rate of n^2. So the growth rate of the total time is at worst cubic.*

The Travelling Salesman Problem is considerably harder, but we can make a few worthwhile comments. Firstly, we can reduce to a problem about complete graphs, as follows. If G is the original connected weighted graph, construct a complete graph K with the same vertex set as G, where every edge $\{v, w\}$ has weight $d(v, w)$ (the distance from v to w in G, which then equals the distance from v to w in K). Think of every edge of K as representing a minimal walk between those vertices in G, which we know how to find. Then a solution to the Travelling Salesman Problem in K can be translated into a solution to the Travelling Salesman Problem in G, with the same weight.

Example 2.32*. *If G is the weighted graph of Example 2.24, then K is:*



A walk in K that starts and finishes at A and visits every other vertex must have weight at least 16, because the part of the walk from A to Z must have weight at least 8, and the part of the walk from Z back to A must have weight at least 8 also. Therefore the walk A, B, D, Z, C, A, which does have weight 16, is a solution of the Travelling Salesman Problem for K. To translate this into a solution of the Travelling Salesman Problem for G, we replace every edge in K with a minimal walk between those vertices in G, to obtain A, B, C, D, Z, C, B, A.

In the complete graph K there is no reason to visit a vertex more than once, so the Travelling Salesman Problem in K amounts to finding a minimum-weight spanning cycle. The brute-force approach would be to consider every spanning cycle. Unfortunately, as seen in Example 2.11, the number of spanning cycles in K_n is $\frac{(n-1)!}{2}$, which grows super-exponentially as the number of vertices n increases. We are thus a long way from the situation of the Chinese Postman Problem, where there is a solution with cubic growth (see Remark 2.31). In fact, it is widely believed that there is no algorithm for solving the Travelling Salesman Problem whose running time has fixed-power growth: if one was discovered, it would revolutionize mathematics and computing. We will say a little more about this problem in the next chapter.

Chapter 3

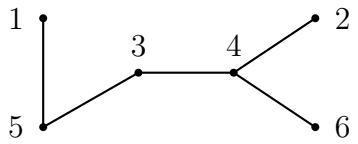
Trees

One of the earliest uses of graph theory outside mathematics was determining the possible molecular structures of various hydrocarbons; one of the most recent has been determining the possible evolutionary relationships between species based on their genetic closeness. In both cases, graphs without cycles are particularly important.

3.1 Trees and Cayley's Formula

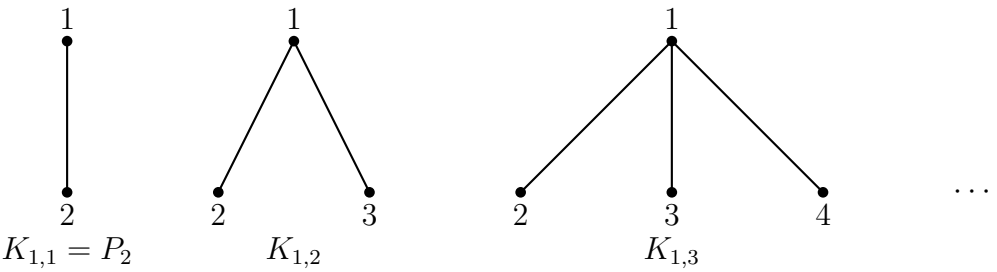
Definition 3.1. A tree is a connected graph containing no cycles. A forest is a general graph containing no cycles (so every connected component of a forest is a tree). A leaf of a tree or forest is a vertex of degree 1.

Example 3.2. *The graph with the following picture is a tree:*



Its leaves are the vertices 1, 2, *and* 6.

Example 3.3. *Of the graphs we have seen previously, the path graph P_n is a tree for all n. The complete bipartite graph $K_{1,q}$ is also a tree for all q:*



$K_{1,1} = P_2$ $K_{1,2}$ $K_{1,3}$

The cycle C_n and the complete graph K_n for $n \geq 3$ are obviously not trees.

Example 3.4. *The list in Example 1.14 shows that any tree with 4 vertices either is a path (hence is isomorphic to P_4) or has a vertex which is adjacent to all the others (hence is isomorphic to $K_{1,3}$).*

One reason that trees are so nice to work with is that if you remove a leaf from a tree (and its single edge), what you have left is still a tree; this means that trees are ideally suited for proofs by induction on the number of vertices. Before we can use this, however, we need to ensure that leaves always exist.

Theorem 3.5. Every tree T with ≥ 2 vertices has at least two leaves.

Proof. Among all the paths in the graph T, there must be one which is of maximal length; so if the vertices of the path are labelled $v_1, v_2, \cdots, v_m$ in order, with v_1 and v_m being the end-vertices, there is no path with more than m vertices. Since T is not just a single vertex, it is clear that the maximal m is at least 2. Now we claim that v_1 is a leaf. Since we know that v_1 is adjacent to v_2, this amounts to saying that v_1 is not adjacent to any other vertex of T. But v_1 can't be adjacent to any v_k with $k \geq 3$, because then we would have a cycle with vertices $v_1, v_2, \cdots, v_k$, and T contains no cycles. Also v_1 can't be adjacent to any vertex v which is not one of the v_i's, because then we would have a path with vertices $v, v_1, v_2, \cdots, v_m$, contradicting the maximality of m. So v_1 is a leaf, as claimed; the same argument shows that v_m is a leaf, so T has at least two leaves. □

We can now prove that trees and forests are exactly the graphs which attain the minimum numbers of edges found in Theorem 1.37.

Theorem 3.6. Let G be a graph with n vertices, k edges and s connected components. Then G is a forest if and only if $k = n - s$. In particular, if G is connected, then G is a tree if and only if $k = n - 1$.

Proof. We first prove that if G is a tree, it has $n-1$ edges. As foreshadowed above, we use induction on n. The $n = 1$ case is when G has a single vertex; it does indeed then have 0 edges. Assume that $n \geq 2$ and that the result is known for trees with fewer vertices. By Theorem 3.5, G has a leaf v. The graph $G - v$ obtained by deleting v and its sole edge is a tree with $n - 1$ vertices, which has $n - 2$ edges by the induction hypothesis. So G itself has $n - 1$ edges, and the induction is complete.

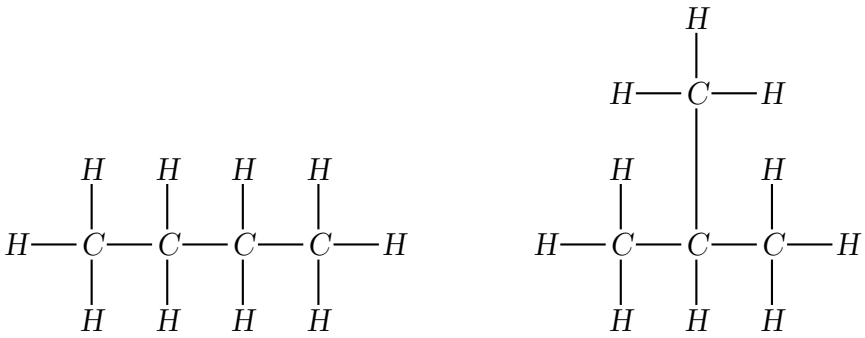
We can deduce the "only if" direction of the general result. If G is a forest, then its connected components $G_1, G_2, \cdots, G_s$ are all trees. So if G_i has n_i vertices, it must have $n_i - 1$ edges. Thus the number of edges in G is $(n_1 - 1) + \cdots + (n_s - 1) = (n_1 + \cdots + n_s) - s = n - s$.

Now we prove the "if" direction. Suppose that $k = n - s$, and let $\{v, w\}$ be any edge of G (if G has no edges, it is clearly a forest). If we delete this edge, the graph $G - \{v, w\}$ has $n - s - 1$ edges. According to Theorem 1.37, this is less than the minimum number of edges a graph with s connected components can have, so $G - \{v, w\}$ must have more connected components than G. The only way this is possible is if v and w are not linked in $G - \{v, w\}$, i.e. $\{v, w\}$ is a bridge. So we have shown that every edge of G is a bridge. We saw in Theorem 1.34 that this is equivalent to saying that no edge of G is contained in a cycle, so G contains no cycles and is a forest. □

Theorem 3.6 means that classifying trees with n vertices is the same as classifying connected graphs with n vertices and $n - 1$ edges.

Example 3.7. *One of the major problems in 19th century chemistry was to find the structure of various molecules, knowing only the molecular formula, i.e. how many atoms of various elements the molecule contained. Molecules were thought of as connected graphs, where the atoms were the vertices and the bonds between atoms were the edges. Each element had a known valency, which in graph-theoretic terms meant that the degree of all atoms of that element was the same. For instance, in hydrocarbons every vertex is either a carbon atom (C) of degree* 4 *or a hydrogen atom (H) of degree* 1*. Suppose*

we know that a molecule has formula C_kH_{2k+2}, for some positive integer k; this means that the graph has k vertices of degree 4 and $2k+2$ vertices of degree 1. We can then work out the number of edges by the Hand-shaking Lemma: it is $\frac{4k+2k+2}{2} = 3k+1$. Since this is exactly 1 less than the number of vertices, Theorem 3.6 implies that the graph is a tree. (This rules out the possibility of cycles of carbon atoms, which give rise to special chemical properties.) The hydrogen atoms are the leaves of the tree, so if you delete them all and consider just the carbon atoms you still have a tree. In fact, if you know the tree formed by the carbon atoms, you can reconstruct the whole molecule by joining hydrogen atoms where necessary to make the C degrees equal 4. For instance, there are only two possible structures for a molecule with formula C_4H_{10}:



Both stuctures do exist: the molecules are called butane and isobutane.

Motivated by such applications, the British mathematician Arthur Cayley started a systematic enumeration of trees. He found that for large n, it was easier to count all the trees with a particular vertex set, such as $\{1, 2, \cdots, n\}$, rather than to count the isomorphism classes of trees with n vertices (in other words, to count labelled rather than unlabelled trees).

Example 3.8. *We have seen that there are $\binom{6}{3} = 20$ graphs with vertex set $\{1, 2, 3, 4\}$ which have 3 edges. Of these, the only ones which are not trees are the ones with two connected components, an isolated vertex and a 3-cycle; there are 4 of these. So there are 16 trees with vertex set $\{1, 2, 3, 4\}$ (of which, incidentally, 12 are isomorphic to P_4 and 4 are isomorphic to $K_{1,3}$).*

The general result, striking in its simplicity, is:

Theorem 3.9 (Cayley's Formula)**.** For any $n \geq 2$, the number of trees with vertex set $\{1, 2, \cdots, n\}$ is n^{n-2}.

We will prove Cayley's Formula by the Bijection Principle, finding a bijection between the set of trees with vertex set $\{1, 2, \cdots, n\}$ and a set of sequences which has size n^{n-2}. The sequence we attach to a tree is called its Prüfer sequence, which is defined by the following recursive algorithm.

PRÜFER SEQUENCE. The input is a tree T with $n \geq 2$ vertices, where the vertex set is a subset of the positive integers. The output is a sequence $(p_1, p_2, \cdots, p_{n-2})$ with $n-2$ terms, all in the vertex set. The steps are:

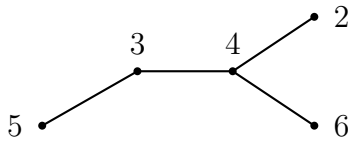
(1) If $n = 2$, return the empty sequence () and stop.

(2) (To arrive here, we must have $n \geq 3$.) Let ℓ be the smallest (in numerical order) of all the leaves of T. Define the first term p_1 to be the unique vertex of T which is adjacent to ℓ.

(3) Recursively call PRÜFER SEQUENCE to find the Prüfer sequence of the smaller tree $T - \ell$, and let $(p_2, \cdots, p_{n-2})$ be this sequence.

In words, the algorithm progressively deletes the smallest leaf of whatever is remaining, and records in the sequence not the leaf itself but the vertex which was adjacent to it.

Example 3.10. *Let T be the tree in Example 3.2. The smallest leaf is 1 which is adjacent to 5. So the first term of the Prüfer sequence is 5, and the rest is obtained by considering the tree $T - 1$:*



The smallest leaf in $T - 1$ is 2 which is adjacent to 4, so the next term of the Prüfer sequence is 4, and the rest is obtained by considering $(T-1)-2$. Continuing in this way, you can find the full Prüfer sequence of T (which should have $6 - 2 = 4$ terms):

Here is a useful property of the Prüfer sequence:

Theorem 3.11. If T is a tree with $n \geq 2$ vertices, then each vertex v occurs $\deg(v) - 1$ times in the Prüfer sequence of T. In particular, the leaves of T are exactly the vertices which don't occur in the sequence.

Proof. Since the definition of Prüfer sequence is recursive, the obvious method of proof is induction on n. The $n = 2$ base case is clear: both vertices are leaves, and the sequence is accordingly empty. Assume that $n \geq 3$ and that the result is known for smaller trees. Let ℓ be the smallest leaf of T. The Prüfer sequence of T is $(p_1, p_2, \cdots, p_{n-2})$, where p_1 is the vertex adjacent to ℓ and $(p_2, \cdots, p_{n-2})$ is the Prüfer sequence of $T - \ell$. The claim is certainly correct when $v = \ell$: ℓ does not occur in the Prüfer sequence. The claim is also correct when $v = p_1$: we have $\deg_{T-\ell}(p_1) = \deg_T(p_1) - 1$, so the induction hypothesis tells us that p_1 occurs $\deg_T(p_1) - 2$ times in $(p_2, \cdots, p_{n-2})$, and hence it occurs $\deg_T(p_1) - 1$ times in $(p_1, \cdots, p_{n-2})$. If v is any other vertex of T, then its degree is the same in $T - \ell$ as in T. So the induction hypothesis says that v occurs $\deg(v) - 1$ times in $(p_2, \cdots, p_{n-2})$ and hence also in $(p_1, \cdots, p_{n-2})$. The induction step is complete. □

The crucial point is that we can reconstruct a tree uniquely from its Prüfer sequence. The easiest way to see this is to write down a recursive algorithm going in the opposite direction to the above algorithm.

REVERSE PRÜFER. The input is a subset V of the positive integers with $n \geq 2$ elements, and a sequence $(p_1, p_2, \cdots, p_{n-2})$ whose terms belong to V. The output is a tree T with vertex set V. The steps are:

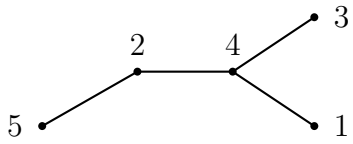
(1) If $n = 2$, output the tree with vertex set V and a single edge joining the two vertices.

(2) (To arrive here, we must have $n \geq 3$.) Let ℓ be the smallest element of V which does not occur in the sequence.

(3) Recursively call REVERSE PRÜFER with the smaller set $V \setminus \{\ell\}$ and the shorter sequence $(p_2, \cdots, p_{n-2})$; let T' be the resulting tree.

(4) Form a tree T with vertex set V by adding to T' the new vertex ℓ and a single new edge joining ℓ to p_1.

Example 3.12. *Suppose that* $V = \{1, 2, 3, 4, 5\}$ *and the sequence is* $(4, 4, 2)$. *The algorithm tells us that our tree* T *is obtained by attaching the leaf* 1 *to the vertex* 4 *of the tree* T', *where* T' *is the output of the algorithm applied to the set* $\{2, 3, 4, 5\}$ *and the sequence* $(4, 2)$. *In turn,* T' *is obtained by attaching the leaf* 3 *to the vertex* 4 *of the tree* T'', *where* T'' *is the output of the algorithm applied to the set* $\{2, 4, 5\}$ *and the sequence* (2). *Finally,* T'' *is obtained by attaching the leaf* 4 *to the vertex* 2 *of the unique tree with vertex set* $\{2, 5\}$. *So* T *is:*



We can now prove Cayley's Formula.

Proof. Let X be the set of all trees with vertex set $\{1, 2, \cdots, n\}$, and let Y be the set of all sequences $(p_1, p_2, \cdots, p_{n-2})$ where each p_i belongs to $\{1, 2, \cdots, n\}$. The PRÜFER SEQUENCE algorithm gives us a function $f : X \to Y$, taking each tree to its Prüfer sequence. The REVERSE PRÜFER algorithm gives us a function $g : Y \to X$, taking a sequence (and the set $\{1, 2, \cdots, n\}$) and constructing a tree. From the definitions of the algorithms, it is clear that these functions are inverse to each other, i.e. if you apply REVERSE PRÜFER to the output of PRÜFER SEQUENCE you do recover the original tree, and if you apply PRÜFER SEQUENCE to the output of REVERSE PRÜFER you do recover the original sequence. (You could prove this formally by induction.) So X and Y are in bijection, and $|X| = |Y|$. But clearly $|Y| = n^{n-2}$, because there are n choices for each of the $n - 2$ terms. Cayley's Formula follows. □

3.2 Spanning trees

If the edges of a connected graph represent lines of communication in some sense, it may be important to consider the ways in which you could maintain as few of them as possible and still have all the vertices linked to each other.

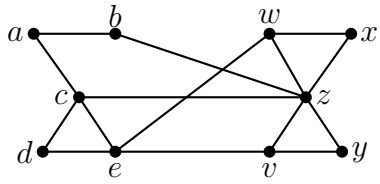
As we have seen, a graph with the smallest possible number of edges subject to being connected is a tree. So for a general connected graph G, we consider the spanning trees of G, i.e. the subgraphs which use all the vertices of G, are themselves connected, and contain no cycles.

Theorem 3.13. Every connected graph G contains a spanning tree.

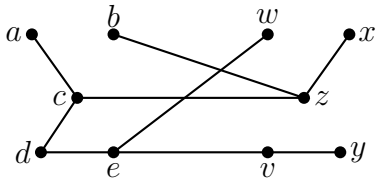
Proof. If every edge of G is a bridge, then G is itself a tree. Otherwise, we can delete some edge belonging to a cycle, and the graph remains connected. If we keep repeating this procedure, we will eventually reach a tree (at the point where the number of edges remaining is 1 less than the number of vertices); this is a spanning tree of the original graph. □

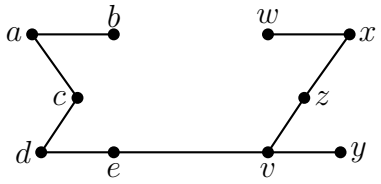
Note that you can't find a spanning tree by simply deleting all edges of G which belong to cycles in G: after all, it is perfectly possible that every edge belongs to a cycle. What the above proof does is different: it deletes an edge of a cycle, and then re-considers the cycles in that smaller graph (some of the cycles in G will have been removed by the deletion), and so on. At any stage, the choices for which edge you can delete depend on which edges you have deleted so far, so there can be many different spanning trees.

Example 3.14. *Suppose that G is the graph*



Two of the many spanning trees of G are:





Most of the important algorithms for finding spanning trees take the reverse approach: they start with a single vertex and add edges. As long as every added edge joins a vertex already in the tree with a new vertex, it is obvious that no cycles are created.

BREADTH-FIRST SEARCH (BFS). The input is a connected graph G with n vertices, and a chosen vertex v_1. The output is an ordering $v_1, v_2, \cdots, v_n$ of the vertices of G, and a spanning tree T of G. The steps are:

(1) Let T be the graph with the single vertex v_1.

(2) If v_n has been defined, then stop.

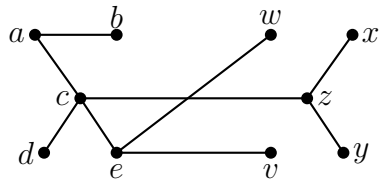
(3) Let m be the largest number for which v_m has been defined. Let $\ell \leq m$ be the smallest number such that there are vertices adjacent to v_ℓ which are not in T. If there are s such vertices, name them $v_{m+1}, \cdots, v_{m+s}$ in an arbitrary order.

(4) Add to T the vertices $v_{m+1}, \cdots, v_{m+s}$, and the edges joining these vertices to v_ℓ.

(5) Return to Step (2).

The reason for the name "search" is the idea that this algorithm searches for the vertices of G: first finding all the vertices adjacent to v_1, then finding all the unfound vertices adjacent to those, and so on until all vertices have been found. Each time Step (4) is executed, the new vertices are added to T along with the edges which record 'through which earlier vertex they were found'. Note that because of the choice in ordering the new vertices, the resulting spanning tree is not unique.

Example 3.15. *Let us apply this algorithm to the graph G in Example 3.14, with $v_1 = a$. The first pass through Step (3) defines $v_2 = c$ and $v_3 = b$, say. The order of these two vertices is arbitrarily chosen, but it affects what happens next, because on the next pass through Step (3) it is the vertices adjacent to v_2 which get added, say $v_4 = d$, $v_5 = e$, $v_6 = z$. The earliest v_i which is adjacent to an unused vertex is now v_5, so $v_7 = v$, $v_8 = w$. Then v_6 is responsible for naming the last two vertices $v_9 = x$ and $v_{10} = y$. Thus the order of the vertices is $a, c, b, d, e, z, v, w, x, y$, and the spanning tree is:*



One application of breadth-first search is to find the shortest possible path from v_1 to another vertex. (Since we are currently dealing with unweighted graphs, "shortest" simply means "with the smallest number of edges".) First note the following general result:

Theorem 3.16. If v and w are vertices of a tree T, there is a unique path in T between v and w.

Proof. We already know by Theorem 1.34 that there is a path between v and w, so the content of this result is the uniqueness. The idea is simple: if there were two different paths, you could patch appropriate pieces of them together to form a cycle, contradicting the fact that the graph is a tree. However, the notation involved in writing this out is a bit messy, so we will disguise the idea by giving an induction proof. Let n be the number of vertices; if $n = 1$ then v and w are the same and there is nothing to prove. If $n \geq 2$ and the result is known for smaller trees, then let ℓ be any leaf of T. If $v = w = \ell$ there is obviously a unique path between v and w, namely the path with no edges and single vertex ℓ. If neither v nor w equals ℓ, then we know by the induction hypothesis that there is a unique path in $T - \ell$ with end-vertices v and w. But clearly no path with end-vertices v and w can include ℓ as one of its other vertices, because $\deg(\ell) = 1$; so there is a unique path in T between v and w. The remaining cases are $v = \ell$, $w \neq \ell$ and $v \neq \ell$, $w = \ell$; these are symmetric, so we can assume $v = \ell$, $w \neq \ell$. If u is the unique vertex adjacent to v, then any path whose first vertex is v must have u as its second vertex. By the induction hypothesis, there is a unique path in $T - \ell$ with end-vertices u and w, so there is a unique path in T between v and w. The induction step is complete. □

In particular, in the spanning tree produced by the BFS algorithm there is a unique path between v_1 and v for any vertex v; this path records 'how the algorithm found v'. Note that in the progress of the algorithm, these paths from v_1 in T are built evenly, in the sense that no paths are ever present while there are shorter paths not yet present. This allows us to prove:

Theorem 3.17. Let T be a spanning tree of the graph G obtained by applying the BFS algorithm with the initial vertex v_1. Then for any vertex v, the unique path in T between v_1 and v has the shortest possible length of any path in G between these vertices.

Proof*. Let $d_T(v_1, v)$ denote the length (i.e. number of edges) of the unique path between v_1 and v in T, and let $d_G(v_1, v)$ denote the length of the shortest possible path between v_1 and v in G. It is obvious that $d_G(v_1, v) \leq d_T(v_1, v)$ for all v. We suppose for a contradiction that $d_G(v_1, v) < d_T(v_1, v)$ for some v; we can choose v so that $d_G(v_1, v)$ is minimal, subject to this inequality holding. Since $d_T(v_1, v) \geq 1$, v is not v_1. Let w be the vertex 'through which' v was added to the tree T; then w is adjacent to v on the path in T between v_1 and v, so $d_T(v_1, w) = d_T(v_1, v) - 1$. Let u be the vertex adjacent to v on some shortest-possible path in G between v_1 and v. Then the part of the path from v_1 to u must also be as short as possible, so $d_G(v_1, u) = d_G(v_1, v) - 1$. By our minimality assumption, we have

$$d_T(v_1, u) = d_G(v_1, u) = d_G(v_1, v) - 1 < d_T(v_1, v) - 1 = d_T(v_1, w).$$

This implies that in the progress of the BFS algorithm, u was added to the tree T before w; but then it is impossible for the vertex v to have been added through w, because it would have already been added through u if not through some other vertex. □

Breadth-first search is so called by comparison with depth-first search, where you follow one path as long as you can before retracing your steps to take a new one (in a somewhat similar way to the line of succession to the throne in a European monarchy).

DEPTH-FIRST SEARCH (DFS). The input is a connected graph G with n vertices, and a chosen vertex v_1. The output is an ordering $v_1, v_2, \cdots, v_n$ of the vertices of G, and a spanning tree T of G. The steps are:

(1) Let T be the graph with the single vertex v_1.

(2) If v_n has been defined, then stop.

(3) Let m be the largest number for which v_m has been defined. Let $\ell \leq m$ be the largest number such that there are vertices adjacent to v_ℓ which are not in T. Choose one of these vertices and call it v_{m+1}.

(4) Add to T the vertex v_{m+1}, and the edge joining this vertex to v_ℓ.

(5) Return to Step (2).

Example 3.18. *If you carry out a depth-first search on the graph G in Example 3.14 with $v_1 = a$, and whenever you have a choice of vertices choose the one which is latest in the alphabet, then the order of the vertices is*

and the spanning tree T is

One useful feature of a DFS spanning tree is that it contains some information about all the edges of the graph.

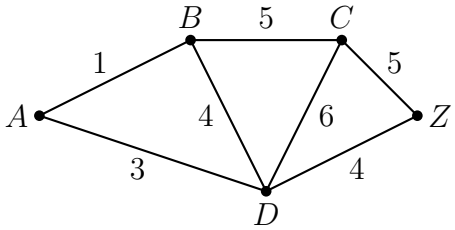
Theorem 3.19. Let T be a spanning tree of the graph G obtained by applying the DFS algorithm, and let $v_1, v_2, \cdots, v_n$ be the resulting ordering of the vertices. If v_i is adjacent to v_j in G, with $i < j$, then v_i lies on the path between v_1 and v_j in T.

Proof*. Let $v_{i_1}, v_{i_2}, \cdots, v_{i_s}$ be the successive vertices of the path from v_1 to v_i, with $i_1 = 1$ and $i_s = i$; from the construction it is clear that $i_1 < i_2 < \cdots < i_s$. Similarly, let $v_{j_1}, v_{j_2}, \cdots, v_{j_t}$ be the successive vertices of the path from v_1 to v_j, with $j_1 = 1$, $j_t = j$, and $j_1 < j_2 < \cdots < j_t$. Let k be maximal so that $i_k = j_k$; the vertex v_{i_k} is the 'latest common ancestor' of v_i and v_j, if you think of T as a family tree of descendants of v_1. We want to prove that, given the adjacency of v_i and v_j in G, this latest common ancestor is v_i itself, i.e. $k = s$. Suppose for a contradiction that $k < s$. It is impossible for k to be t, because that would imply that $j < i$, so the two paths diverge and continue through the different vertices $v_{i_{k+1}}$ and $v_{j_{k+1}}$. From the fact that the descendant v_i of $v_{i_{k+1}}$ was added to the tree before the descendant v_j of $v_{j_{k+1}}$, we deduce that $i_{k+1} \leq i < j_{k+1}$. But then consider the pass through Step (3) of the algorithm which gave $v_{j_{k+1}}$ its name. At this point of the algorithm, the value of m must have been $j_{k+1} - 1$, so v_i was already in T;

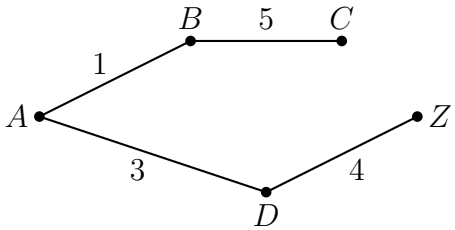
and the value of ℓ must have been i_k, because v_{i_k} was the vertex to which the new vertex $v_{j_{k+1}}$ became attached in T. But that contradicts the maximality in the choice of ℓ, because $i > i_k$ and v_i was also adjacent to a vertex not yet in T, namely v_j. □

In the context of weighted graphs, one would prefer a minimal spanning tree, i.e. one whose weight (the sum of the weights of the edges) is as small as possible. The naive approach is to build the tree by starting with a single vertex and adding one new edge at a time, where the edge has the smallest weight of all those you can possibly add to the tree.

Example 3.20. *Consider the weighted graph*



If we start with vertex A and want to add another vertex to our tree in a way which minimizes the total weight of the edges, we should obviously choose B rather than D. Having done that, the possible edges we can add in the next step are $\{A, D\}$, $\{B, C\}$, and $\{B, D\}$; since $\{A, D\}$ has the smallest weight, we choose that. We cannot now add the edge $\{B, D\}$, because that would form a cycle: to ensure that we still have a tree, we have to add a new vertex along with each new edge. The possible edges are $\{B, C\}$, $\{C, D\}$, and $\{D, Z\}$; the last of these has the smallest weight, so that is the one we add. Finally, we choose between the three edges ending at C: since $\{B, C\}$ and $\{C, Z\}$ have the same weight, we arbitrarily choose $\{B, C\}$ to reach a spanning tree of weight 13.



It is easy to see that there is no spanning tree of weight less than 13*, so we have found a minimal spanning tree for our weighted graph.*

Somewhat remarkably, this naive method always works, and is worth describing formally.

PRIM'S ALGORITHM. The input is a connected weighted graph G with n vertices, and a chosen vertex v_1. The output is an ordering $v_1, v_2, \cdots, v_n$ of the vertices of G, and a spanning tree T of G. The steps are:

(1) Let T be the graph with the single vertex v_1.

(2) If v_n has been defined, then stop.

(3) Let m be the largest number for which v_m has been defined. Consider all the edges $\{v_i, w\}$ in G where $i \leq m$ and w is not in T; choose one of minimal weight, and let v_{m+1} be w.

(4) Add to T the vertex v_{m+1}, and the edge joining this vertex to v_i.

(5) Return to Step (2).

Theorem 3.21. A spanning tree produced by Prim's Algorithm is minimal.

Proof.** Let T be a spanning tree of the weighted graph G produced by Prim's Algorithm, and let $v_1, v_2, \cdots, v_n$ be the vertices in the order in which they were added. Thus the edges of T can be listed as $\{v_{a_2}, v_2\}, \cdots, \{v_{a_n}, v_n\}$ for some $a_2, \cdots, a_n \in \{1, \cdots, n\}$ such that $a_j < j$ for $j = 2, \cdots, n$.

Let S be a minimal spanning tree of G. If $S = T$ we are finished, so we can assume there is some j such that the edge $\{v_{a_j}, v_j\}$ of T does not belong to S; in fact, we can define j to be minimal with this property. Since S is a tree, there is a unique path between v_{a_j} and v_j in S: let its successive vertices be $v_{i_1}, v_{i_2}, \cdots, v_{i_s}$ where $i_1 = a_j$ and $i_s = j$. We know that $s \geq 3$, because the edge $\{v_{a_j}, v_j\}$ is not in S. Now let $t \leq s$ be minimal such that $i_t \geq j$; then $t \geq 2$ and $i_{t-1} < j$. Since the edge $\{v_{i_{t-1}}, v_{i_t}\}$ was passed over for inclusion in T at the stage of the algorithm where $\{v_{a_j}, v_j\}$ was added, its weight cannot be less than that of $\{v_{a_j}, v_j\}$. So if we modify the spanning tree S by adding the edge $\{v_{a_j}, v_j\}$ (temporarily forming a cycle with the existing path) and then deleting the edge $\{v_{i_{t-1}}, v_{i_t}\}$ to return to $n-1$ edges, we obtain another minimal spanning tree S'.

If $S' = T$ we are finished. Otherwise, since the edge we deleted from S was not any of $\{v_{a_2}, v_2\}, \cdots, \{v_{a_{j-1}}, v_{j-1}\}$, the quantity j' defined for S' in the same way as j was for S is strictly larger. So if we repeat this procedure, modifying S' to obtain a new minimal spanning tree S'' and so on, we must eventually obtain T. □

With Prim's Algorithm to help us find minimal spanning trees, we can say a little more about the Travelling Salesman Problem. Remember that we had reduced the problem to finding minimal spanning cycles in a Hamiltonian weighted graph (in fact, we saw that we could assume the graph is complete, but it doesn't help much).

Theorem 3.22. Let H be a Hamiltonian weighted graph, and v a vertex of H. Then the weight of a minimal spanning cycle of H is at least equal to:

the weight of a minimal spanning tree of $H - v$
+ the sum of the two smallest weights of edges of H at v

Proof. Let C be any spanning cycle in H. Then the weight of C equals the weight of $C - v$ plus the sum of the weights of the two edges of C at v. Since $C - v$ is a path and contains all the vertices of $H - v$, it is a spanning tree of $H - v$, so its weight is at least equal to the weight of a minimal spanning tree of $H - v$. Trivially, the sum of the weights of the two edges of C at v is at least equal to the sum of the two smallest weights of edges of H at v. The result follows. □

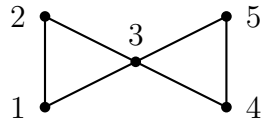
Theorem 3.22 gives us a lower bound for the weight of a minimal spanning cycle, or rather one lower bound for each vertex, which can be computed efficiently. However, there is no guarantee that equality holds.

Example 3.23. *If K is the complete weighted graph of Example 2.32, then the weight of a minimal spanning tree in $K - A$ is* 7*, and the two smallest weights of edges at A are* 1 *and* 4*, so the lower bound provided by Theorem 3.22 is* 12*. Similarly, the lower bounds obtained by considering the vertices B, C, D, and Z are* 11*,* 11*,* 12*, and* 13 *respectively; all well short of the actual weight of a minimal spanning cycle of K, which we saw is* 16*.*

3.3 Kirchhoff's Matrix–Tree Theorem

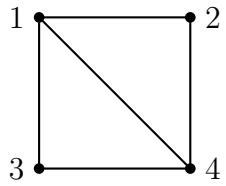
In the previous section we saw several different ways to construct a spanning tree of a connected graph. This raises the interesting combinatorial question of how many spanning trees a given connected graph has.

Example 3.24. *Let G be the graph*



Since G has five vertices, any spanning tree must have four edges, so it must be obtained from G by deleting two edges. Since a spanning tree cannot contain cycles, we need to delete one edge from each of the two 3-cycles. Clearly we can choose any edge from the left-hand 3-cycle and any edge from the right-hand 3-cycle, so G has $3 \times 3 = 9$ spanning trees. (These fall into three isomorphism classes, but we are not concerned with isomorphism classes at the moment.)

Example 3.25. *Let G be the graph*



To find a spanning tree, we need to delete two edges; of the $\binom{5}{2} = 10$ possible pairs of edges, the only cases which would leave a cycle remaining are if we deleted $\{1, 2\}$ and $\{2, 4\}$, or $\{1, 3\}$ and $\{3, 4\}$. So there are $10 - 2 = 8$ spanning trees.

Example 3.26. *If G is a tree, then of course the unique spanning tree of G is G itself.*

Example 3.27. *If $G = C_n$, then we obtain a spanning tree (specifically, a path) by deleting any edge, so there are n spanning trees.*

Example 3.28. *Another general class for which we have already answered the question is that of the complete graphs K_n, for $n \geq 2$. A spanning tree of K_n is just the same thing as a tree with vertex set $\{1, 2, \cdots, n\}$, and by Cayley's Formula there are n^{n-2} of these.*

The physicist Kirchhoff (who was interested in graphs through his study of electrical circuits) found a general formula for the number of spanning trees, which reveals an unexpected connection with the determinants of matrices.

Definition 3.29. Let G be a graph with vertex set $\{1, 2, \cdots, n\}$. The Laplacian matrix of G is the $n \times n$ matrix M whose (i, j)-entry is defined by

$$m_{ij} = \begin{cases} \deg(i) & \text{if } i = j, \\ -1 & \text{if } i \neq j \text{ and } \{i, j\} \text{ is an edge of } G, \\ 0 & \text{if } i \neq j \text{ and } \{i, j\} \text{ is not an edge of } G. \end{cases}$$

That is, the diagonal entries give the degrees of the vertices, and the off-diagonal entries are -1 in positions corresponding to an edge and 0 elsewhere.

Example 3.30. *The Laplacian matrix of the graph in Example 3.24 is*

$$\begin{pmatrix} 2 & -1 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 \\ -1 & -1 & 4 & -1 & -1 \\ 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & -1 & 2 \end{pmatrix}.$$

Here are some properties of the Laplacian matrix.

Theorem 3.31. Let M be the Laplacian matrix of a graph G as above.

(1) M is symmetric (its (j, i)-entry equals its (i, j)-entry for all i, j).

(2) The sum of the entries in any row or column of M is zero.

(3) The determinant $\det(M)$ is zero.

(4) For any $k, \ell \in \{1, 2, \cdots, n\}$, let $M^{\hat{k},\hat{\ell}}$ denote the $(n-1) \times (n-1)$ matrix obtained from M by deleting the kth row and the ℓth column. Then $(-1)^{k+\ell} \det(M^{\hat{k},\hat{\ell}})$ is the same for all values of k, ℓ.

Proof*. Part (1) is obvious from the definition. Because M is symmetric, its rows are the same as its columns, so we only need to prove part (2) for a row,

say the ith row. By definition, the nonzero entries in the ith row are a single entry of $\deg(i)$, and an entry of -1 for every vertex adjacent to i; since there are $\deg(i)$ adjacent vertices, these entries cancel out and give zero, proving part (2). Now suppose we multiply the matrix M by the $n \times 1$ column vector $\mathbf{v}$ whose entries are all 1. By the definition of matrix multiplication, $M\mathbf{v}$ is another column vector whose ith entry is the sum of the entries in the ith row of M; we have just seen that this sum is always zero, so $M\mathbf{v}$ is the zero vector. In other words, $\mathbf{v}$ belongs to the null space of the matrix M. A standard linear algebra result says that a square matrix has a nontrivial null space if and only if its determinant is zero, so part (3) follows.

To prove part (4), it is enough to show that

$$(-1)^{k+\ell} \det(M^{\hat{k},\hat{\ell}}) = (-1)^{k'+\ell} \det(M^{\hat{k}',\hat{\ell}}) \text{ for all } k, k', \ell, \tag{3.1}$$

since then the symmetry of M will imply that we also have

$$(-1)^{k+\ell} \det(M^{\hat{k},\hat{\ell}}) = (-1)^{k+\ell'} \det(M^{\hat{k},\hat{\ell}'}) \text{ for all } k, \ell, \ell', \tag{3.2}$$

and the result clearly follows from (3.1) and (3.2). Moreover, it suffices to prove the case of (3.1) where $k' = k + 1$. In that case, the only difference between the matrices $M^{\hat{k},\hat{\ell}}$ and $M^{\hat{k}',\hat{\ell}}$ is in their kth rows: the kth row of $M^{\hat{k}',\hat{\ell}}$ is the original kth row of M without its ℓth entry, which is the very row that was deleted from M to form $M^{\hat{k},\hat{\ell}}$. Since the sum of the entries in each column of M is zero, the sum of all the rows of M is the zero row vector, so the kth row of M is the negative of the sum of all the other rows. Thus $M^{\hat{k}',\hat{\ell}}$ is exactly what you get when you apply to the matrix $M^{\hat{k},\hat{\ell}}$ the row operation $R_k \leftarrow -R_1 - R_2 - \cdots - R_{n-1}$. This row operation changes the sign of the determinant, so $\det(M^{\hat{k}',\hat{\ell}}) = -\det(M^{\hat{k},\hat{\ell}})$, which gives the required case of (3.1). □

The quantity $(-1)^{k+\ell} \det(M^{\hat{k},\hat{\ell}})$ is called the (k,ℓ)-cofactor of the matrix M.

Remark 3.32. *These cofactors are what you multiply the matrix entries by when calculating the determinant by expanding along a row or column. For instance, the rule for calculating* $\det(M)$ *by expanding along the first row is*

$$\det(M) = m_{11} \det(M^{\hat{1},\hat{1}}) - m_{12} \det(M^{\hat{1},\hat{2}}) + \cdots + m_{1n}(-1)^{1+n} \det(M^{\hat{1},\hat{n}}).$$

In our case there is no need to use this rule, since part (3) of Theorem 3.31 tells us that $\det(M)$ *is zero.*

Theorem 3.33 (Kirchhoff's Matrix–Tree Theorem)**.** Let M be the Laplacian matrix of a graph G with vertex set $\{1, 2, \cdots, n\}$, where $n \geq 2$. Then for all $k, \ell \in \{1, 2, \cdots, n\}$, the (k, ℓ)-cofactor $(-1)^{k+\ell} \det(M^{\hat{k},\hat{\ell}})$ equals the number of spanning trees of G. (In particular, if G is not connected, all the cofactors of M are zero.)

Before we give the proof, here are some examples of how to use the Matrix–Tree Theorem to count spanning trees.

Example 3.34. *Let us apply the Matrix–Tree Theorem to find the number of spanning trees in the graph G of Example 3.24, whose Laplacian matrix M was found in Example 3.30. Since all the cofactors of M are guaranteed to be the same, we may as well choose to delete the row and column with the most nonzero entries; the $(3, 3)$-cofactor is*

$$(-1)^{3+3} \det(M^{\hat{3},\hat{3}}) = \det \begin{pmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & 0 & 0 \\ 0 & 0 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{pmatrix}.$$

When a matrix has block-diagonal form like this, its determinant is the product of the determinants of the blocks. Since

$$\det \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} = 2 \times 2 - (-1) \times (-1) = 3,$$

the $(3, 3)$-cofactor of M is $3 \times 3 = 9$. This of course agrees with the number of spanning trees found in Example 3.24.

Example 3.35. *If G is as in Example 3.25, the Laplacian matrix M is*

$$\begin{pmatrix} & & & \\ & & & \\ & & & \\ & & & \end{pmatrix}.$$

Its $(1, 1)$-cofactor is

$$\det \begin{pmatrix} & & \\ & & \\ & & \end{pmatrix} = \square.$$

Example 3.36*. *We can use the Matrix–Tree Theorem to give another proof of Cayley's Formula for the number of spanning trees of K_n. The Laplacian matrix M of K_n is the $n \times n$ matrix where all the diagonal entries are $n-1$ and all the off-diagonal entries are -1. The $(1,1)$-cofactor is the determinant of an $(n-1) \times (n-1)$ matrix with the same rule for the entries:*

$$\begin{pmatrix} n-1 & -1 & -1 & \cdots & -1 \\ -1 & n-1 & -1 & \cdots & -1 \\ -1 & -1 & n-1 & \cdots & -1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & -1 & \cdots & n-1 \end{pmatrix}.$$

To put this matrix in a form where the determinant is easier to compute, we apply the row operation $R_1 \leftarrow R_1 + R_2 + \cdots + R_{n-1}$ (which does not change the determinant):

$$\begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ -1 & n-1 & -1 & \cdots & -1 \\ -1 & -1 & n-1 & \cdots & -1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & -1 & \cdots & n-1 \end{pmatrix},$$

and then the row operations $R_i \leftarrow R_1 + R_i$ for $i = 2, \cdots, n-1$:

$$\begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 0 & n & 0 & \cdots & 0 \\ 0 & 0 & n & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & n \end{pmatrix}.$$

For an upper-triangular matrix like this, the determinant is the product of the diagonal entries, which are a 1 and $n-2$ copies of n. So we recover the answer n^{n-2}, re-proving Cayley's Formula.

The proof of the Matrix–Tree Theorem uses a fact about determinants which you may not have seen. Recall that determinants are only defined for square matrices, and that they are multiplicative in the sense that

$$\det(AB) = \det(A)\det(B), \tag{3.3}$$

whenever A and B are square matrices of the same size. There is a generalization of this multiplicativity property for possibly non-square matrices:

Theorem 3.37 (Cauchy–Binet Formula)***.** Let A be an $n \times m$ matrix and B an $m \times n$ matrix. Then

$$\det(AB) = \sum_{\substack{J \subseteq \{1,2,\cdots,m\} \\ |J|=n}} \det(A^{-,J}) \det(B^{J,-}),$$

where $A^{-,J}$ is the $n \times n$ matrix obtained from A by deleting all the columns except those whose number lies in J, and $B^{J,-}$ is the $n \times n$ matrix obtained from B by deleting all the rows except those whose number lies in J. (The $-$ in the superscripts is to indicate that there is no deletion of any rows in forming $A^{-,J}$ or columns in forming $B^{J,-}$.)

We will omit the proof. Note that if $m = n$, the only choice for J is all of $\{1, \cdots, n\}$, and for this J we have $A^{-,J} = A$ and $B^{J,-} = B$, which brings the formula in line with (3.3). In general, there are $\binom{m}{n}$ terms on the right-hand side of the Cauchy–Binet Formula (in particular, if $n > m$ the statement is that $\det(AB) = 0$).

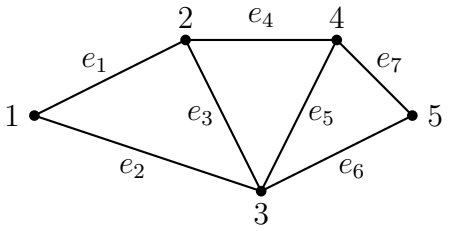
The reason we are interested in non-square matrices is the following.

Definition 3.38. Let G be a graph with vertex set $\{1, 2, \cdots, n\}$ and edges numbered $e_1, e_2, \cdots, e_k$ in some fixed order (we assume that $G \neq N_n$). The incidence matrix F of G is the $n \times k$ matrix whose (i, j)-entry is defined by

$$f_{ij} = \begin{cases} 1 & \text{if } e_j = \{i, i'\} \text{ where } i < i', \\ -1 & \text{if } e_j = \{i, i'\} \text{ where } i > i', \\ 0 & \text{if } i \text{ is not an end of } e_j. \end{cases}$$

So the jth column contains a 1 in the row corresponding to the smaller end of e_j, a -1 in the row corresponding to the larger end, and zeroes elsewhere.

Example 3.39. *If G is the following graph with edges numbered as shown:*



then the incidence matrix F is the following 5×7 matrix:

$$\begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & -1 & -1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & -1 & -1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & -1 & -1 \end{pmatrix}.$$

Theorem 3.40*. Let F be the incidence matrix of a graph G as above.

(1) The sum of the entries in any column of F is zero.

(2) Let I be a subset of $\{1, 2, \cdots, n\}$ such that $|I| = k$. Let $F^{I,-}$ be the $k \times k$ matrix obtained from F by deleting all the rows except those whose number lies in I. If I contains all the vertices of some connected component of G, then $\det(F^{I,-}) = 0$.

(3) Suppose that G is a tree; hence $k = n - 1$. For any $i \in \{1, \cdots, n\}$, let $F^{\hat{i},-}$ be the $k \times k$ matrix obtained from F by deleting the ith row. Then $\det(F^{\hat{i},-}) = \pm 1$.

Proof.** Part (1) is obvious, because the entries of each column of F are a 1, a -1, and $n - 2$ zeroes. For part (2), let I' be a subset of I consisting of all the vertices of some connected component of G. For convenience, we continue to number the rows of the matrix $F^{I,-}$ by their row numbers from F (that is, the elements of I) rather than changing their numbers to $1, \cdots, k$. Let $\mathbf{v}$ be a row vector whose columns are indexed by the elements of I, in which the ith entry is 1 if $i \in I'$ and 0 otherwise. Then $\mathbf{v}F^{I,-}$ is a $1 \times k$ row vector whose jth entry is the sum, as i runs over I', of the ith entry of the jth column of $F^{I,-}$. If the edge e_j does not belong to the connected component with vertex set I', all these entries are zero; if the edge e_j does belong to this connected component, the entries consist of a 1, a -1, and the rest zeroes. In either case the sum is zero, so $\mathbf{v}F^{I,-}$ is the zero row vector. Thus the left null space of $F^{I,-}$ is nontrivial, proving that $\det(F^{I,-}) = 0$ as required.

We prove part (3) by induction on k. The $k = 1$ base case is clear, because then the incidence matrix F is $\binom{1}{-1}$. Assume that $k \geq 2$, and that the result

is known for trees with fewer edges. Now by the same argument as in the proof of Theorem 3.31, the fact that every column of F sums to zero implies that all the determinants $\det(F^{\hat{i},-})$ are the same up to sign, so it is enough to show that just one of them equals ± 1. We can therefore assume that vertex i is a leaf of the tree G. Let i' be the unique vertex which is adjacent to i, and suppose that the edge $\{i, i'\}$ is e_j. Then the jth column of $F^{\hat{i},-}$ has a single nonzero entry, namely a ± 1 in row i'. By expanding along this column, we conclude that

$$\det(F^{\hat{i},-}) = \pm \det(F^{\widehat{i,i'},\hat{j}}),$$

where the meaning of the superscripts on the right-hand side is hopefully clear. But $F^{\hat{i},\hat{j}}$ is the incidence matrix of the tree $G - i$, so by the induction hypothesis, the right-hand side is ± 1 as required. □

The connection between the Laplacian and incidence matrices is:

Theorem 3.41. Let G be a graph with vertex set $\{1, 2, \cdots, n\}$ and edges numbered $e_1, e_2, \cdots, e_k$ in some fixed order. Let M be the Laplacian matrix of G, let F be the incidence matrix of G, and let $F^{\mathbf{t}}$ be the transpose of F. Then $M = FF^{\mathbf{t}}$.

Proof*. By definition of matrix multiplication and transpose, the (i, i')-entry of $FF^{\mathbf{t}}$ is $\sum_{j=1}^{k} f_{ij} f_{i'j}$. Since f_{ij} is zero unless i is an end of e_j, the only nonzero terms in this sum are those where i and i' are ends of e_j. If $i \neq i'$ and $\{i, i'\}$ is not an edge, there are no such nonzero terms, so the entry is zero. If $i \neq i'$ and $\{i, i'\}$ is an edge, then the unique nonzero term is $1 \times (-1) = -1$, so the entry is -1. If $i = i'$, there is a nonzero term of $f_{ij}^2 = 1$ for every j such that i is an end of e_j, so the entry is $\deg(i)$. Thus in every case the (i, i')-entry of $FF^{\mathbf{t}}$ agrees with the (i, i')-entry of M. □

We can now prove the Matrix–Tree Theorem.

Proof.** Number the edges of G as $e_1, \cdots, e_k$ and form the incidence matrix F. Since all cofactors of M are equal by part (4) of Theorem 3.31, it is enough to show that the (n, n)-cofactor $\det(M^{\hat{n},\hat{n}})$ equals the number of spanning trees of G. It follows from Theorem 3.41 that $M^{\hat{n},\hat{n}} = (F^{\hat{n},-})(F^{\hat{n},-})^{\mathbf{t}}$, where $F^{\hat{n},-}$ is the $(n-1) \times k$ matrix obtained from F by deleting the nth row. Note that since G is connected, $k \geq n - 1$. Applying the Cauchy–Binet Formula,

we obtain

$$\begin{aligned}\det(M^{\hat{n},\hat{n}}) &= \det((F^{\hat{n},-})(F^{\hat{n},-})^{\mathbf{t}}) \\ &= \sum_{\substack{J\subseteq\{1,2,\cdots,k\} \\ |J|=n-1}} \det(F^{\hat{n},J})\det((F^{\hat{n},J})^{\mathbf{t}}) \\ &= \sum_{\substack{J\subseteq\{1,2,\cdots,k\} \\ |J|=n-1}} \det(F^{\hat{n},J})^2,\end{aligned}$$

where the meaning of $F^{\hat{n},J}$ should by now be clear, and the last step uses the fact that the transpose of a square matrix has the same determinant. Now $F^{-,J}$ is the incidence matrix of a spanning subgraph G_J of G, namely the one with edges e_j for $j \in J$. Since G_J has n vertices and $n-1$ edges, it is either a tree or has more than one connected component. If G_J has more than one connected component, then there is a connected component which does not include the vertex n, and part (2) of Theorem 3.40 implies that $\det(F^{\hat{n},J}) = 0$. If G_J is a tree, then part (3) of Theorem 3.40 implies that $\det(F^{\hat{n},J}) = \pm 1$. So the nonzero terms in the above sum are all 1, and the number of them is the number of spanning trees of G, as required. □

Chapter 4

Colourings of Graphs

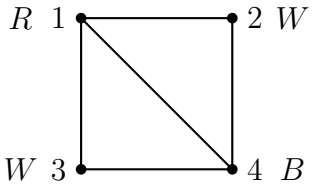
As mentioned in the introduction, one application of graph theory is to scheduling problems. The common feature of such problems is that there are a number of tasks which all have to be carried out in as short a time as possible, with the constraint that some of the tasks cannot be done simultaneously: computer processes which need to access the same database, or exams which have a candidate in common, or flights which need to land on the same runway, or whatever. These problems can be translated into the abstract framework of colourings of graphs.

4.1 Vertex colourings and chromatic number

Definition 4.1. Let G be a graph with vertex set V. A vertex colouring of G is a function $c : V \to C$, where C is a finite set whose elements are called colours, with the property that no two adjacent vertices are assigned the same colour: in other words, $c(v) \neq c(w)$ for every edge $\{v, w\}$ of G. We say that G is t-colourable if there exists a vertex colouring of G where $|C| = t$. The minimum number of colours required in such a vertex colouring, i.e. the smallest t such that G is t-colourable, is called the chromatic number of G and written $\chi(G)$.

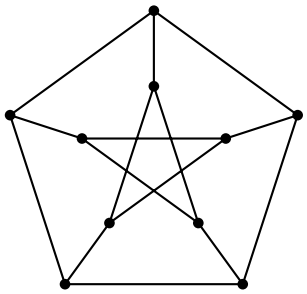
In applications to scheduling, the vertices represent the different tasks, the colours represent the available times for them, and the edges record which pairs of tasks are not allowed to be assigned the same time. From a mathematical point of view, the colours could always just be numbers, but it is convenient to stick with the metaphor and use white, black, red, blue, etc. We will sometimes represent vertex colourings of a graph by labelling the vertices with abbreviations for these colour words, in addition to the labels which give the actual names of the vertices – although the latter may be omitted, since we often care only about the isomorphism class of the graph.

Example 4.2. *Let G be the graph in Example 3.25. It is obvious that G is 4-colourable, because if we have four colours we can give every vertex a different colour, and then the condition that no two adjacent vertices have the same colour is trivially satisfied. Almost as obvious is that G is not 2-colourable, because the vertices 1, 2, and 4 are all adjacent to each other, so there would be a contradiction if we had only 2 colours. Suppose we want a vertex colouring with 3 colours: red, white, and blue. By the previous remark, 1, 2, and 4 have to get one colour each; the remaining constraint is that 3 cannot have the same colour as 1 or 4, so it must get the same colour as 2. Thus a possible vertex colouring of G with three colours is:*



We have shown that the chromatic number $\chi(G)$ is 3. Since the colours red, white, and blue could be shuffled arbitrarily, there are actually $3! = 6$ different vertex colourings with these colours.

Example 4.3. *The chromatic number of the Petersen graph is* $\square$.



Here are some fairly obvious principles concerning colourability.

Theorem 4.4. Let G be a graph with n vertices.

(1) G is 1-colourable if and only if it has no edges. So if G has edges, $\chi(G) \geq 2$.

(2) If G is a complete graph, then $\chi(G) = n$. If G is not a complete graph, then $\chi(G) \leq n - 1$.

(3) If H is a subgraph of G and G is t-colourable, then H is t-colourable. Hence $\chi(H) \leq \chi(G)$.

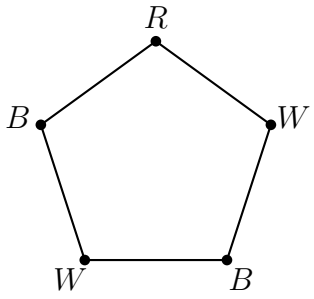
(4) If G has connected components $G_1, G_2, \cdots, G_s$, then G is t-colourable if and only if every G_i is t-colourable. Hence $\chi(G) = \max\{\chi(G_i)\}$.

Proof. If G has a vertex colouring with 1 colour, then no two vertices can be adjacent; conversely, if no two vertices are adjacent we can obviously colour all vertices the same colour, which proves part (1). Now it is clear that if we have n colours we can give every vertex a different colour. (In scheduling terms, this corresponds to the easy but potentially uneconomic option of having no simultaneous processes.) If G is a complete graph, i.e. any two vertices are adjacent, then we are in fact forced to give every vertex a different colour, so we cannot have a vertex colouring with fewer than n colours. If G is not a complete graph, there must be two vertices v and w which are not adjacent, so we can construct a vertex colouring with $n - 1$ colours by giving v and w the same colour and using the other $n - 2$ colours for the other $n - 2$ vertices, one each. This proves part (2).

For part (3), any vertex colouring of G clearly restricts to give a vertex colouring of the subgraph H, which implies the statement. The reason that we have an inequality $\chi(H) \leq \chi(G)$ is that there could be vertex colourings of H which cannot be extended to a vertex colouring of G: for instance, because vertices which are not adjacent in H may be adjacent in the larger graph G. However, in the special case where H is a connected component of G, H and the rest of the graph are completely independent of each other; if you have a vertex colouring of every connected component of G, then taken together they form a vertex colouring of G. Part (4) follows. □

As a consequence of parts (2) and (3) of Theorem 4.4, if G has a subgraph isomorphic to the complete graph K_m, then $\chi(G) \geq m$. In other words, if G contains m vertices which are all adjacent to each other, then G has no vertex colouring with fewer than m colours. (We used this principle in Example 4.2 to show that G was not 2-colourable.) However, it is not the case that G always contains $\chi(G)$ vertices which are all adjacent to each other: the Petersen graph illustrates this.

Example 4.5. *Consider the cycle graph C_n, for $n \geq 3$. If we are to have a vertex colouring with two colours, say white and black, then the colours have to alternate as you move around the cycle. Clearly this is possible if n is even. But if n is odd, you would have a contradiction when you arrived back at the first vertex; thus you need to colour one of the vertices with a third colour, say red, as in the following picture of C_5.*



Thus $\chi(C_n) = 2$ if n is even, and $\chi(C_n) = 3$ if n is odd.

In general, a graph is 2-colourable if and only if it is bipartite in the sense mentioned in Example 2.15: the vertices can be divided into two parts – white vertices and black vertices, to use the colour terminology – such that there are no edges between vertices of the same colour. Example 4.5 showed that even cycles (that is, cycles with an even number of vertices) are bipartite and odd cycles are not. There is a general theoretical criterion:

Theorem 4.6. A graph G is bipartite if and only if it contains no odd cycles.

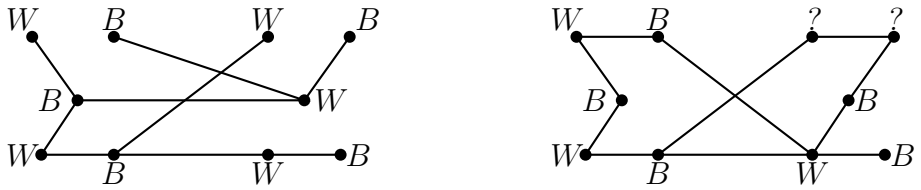
Proof. If G is bipartite (i.e. 2-colourable), then every subgraph of G must also be bipartite; since odd cycles are not bipartite, G cannot contain any odd cycles. We must now prove the converse: if G contains no odd cycles, then G is bipartite. Using part (4) of Theorem 4.4, we reduce to the case that G is connected. Choose any vertex v. To colour the vertices, we divide

the vertex set V into disjoint subsets $V_0, V_1, V_2, \cdots$ as follows. Firstly, V_0 consists of v alone, which we colour white. Then V_1 consists of those vertices adjacent to v, which we are forced to colour black. Then V_2 consists of those vertices which are adjacent to some vertex in V_1 but which are not in V_0 or V_1; we are forced to colour these vertices white. Similarly, V_3 consists of those vertices which are adjacent to some vertex in V_2 but which are not in V_0, V_1, or V_2; we are forced to colour the vertices in V_3 black. Continuing in this way, we eventually reach all the vertices, because the graph is connected. In fact, V_k consists of those vertices w such that the shortest walk from v to w has length k. Note that we have coloured w white if $w \in V_k$ for k even, and we have coloured w black if $w \in V_k$ for k odd.

The only question is whether this is a proper vertex colouring: we need to prove that we never have $w \in V_k$ being adjacent to $w' \in V_\ell$, where k and ℓ are both odd or both even. Suppose for a contradiction that such w, w' do exist. Let $v = v_0, v_1, \cdots, v_k = w$ and $v = v'_0, v'_1, \cdots, v'_\ell = w'$ be minimal-length walks from v to w and to w' respectively. Since $v_i, v'_i \in V_i$ for all i, any vertices in common between the two walks must have the same position in both walks. Let m be maximal such that $v_m = v'_m$ (i.e. this is the last vertex common to both walks). Then the vertices $v'_\ell, v'_{\ell-1}, \cdots, v'_{m+1}, v_m, v_{m+1}, \cdots, v_k$, together with the edges between successive vertices and between first and last, form a cycle in G with $(\ell - m) + (k - m) + 1$ vertices. Since k and ℓ are both odd or both even, this number is odd, and we have our contradiction. □

In particular, every tree is bipartite, because a tree contains no cycles at all. The colouring procedure in the proof of Theorem 4.6 gives a practical way to either find a 2-colouring or prove that none exists.

Example 4.7. *The graph on the left is a tree; starting by colouring the top-left vertex white, we construct the rest of the 2-colouring as shown. If we try to do the same with the graph on the right, we end up wanting to colour both vertices marked "?" white, a contradiction. So the graph on the left is bipartite and the graph on the right is not.*



If a graph is not bipartite, there is no particularly good algorithm for finding its chromatic number or constructing a vertex colouring with that many colours. There are various algorithms for constructing vertex colourings where the number of colours used is not necessarily minimal; here is one.

WELSH–POWELL ALGORITHM. The input is a graph G with n vertices and a set of colours. (We don't know exactly how many colours will be needed, but no more than n.) The output is a vertex colouring of G.

(1) Order the vertices $v_1, v_2, \cdots, v_n$ in order of decreasing degree (for vertices with the same degree, order them arbitrarily).

(2) If all the vertices have been coloured, stop.

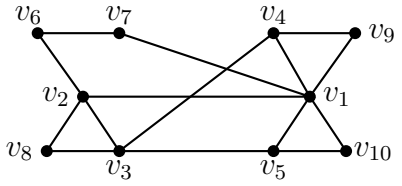
(3) Let m_1 be minimal such that v_{m_1} has not been coloured. Let $m_2 > m_1$ be minimal such that v_{m_2} has not been coloured and v_{m_2} is not adjacent to v_{m_1}. Let $m_3 > m_2$ be minimal such that v_{m_3} has not been coloured and v_{m_3} is not adjacent to v_{m_1} or v_{m_2}. Continue in this way to define $m_1, \cdots, m_s$.

(4) Colour $v_{m_1}, v_{m_2}, \cdots, v_{m_s}$ with a colour which has not yet been used.

(5) Return to Step (2).

Note that in this algorithm you colour all the red vertices (say) at once and then never use the colour red again: to determine which vertices to colour red, you choose the first uncoloured vertex and then whatever later vertices you can allowably give the same colour to. The ordering by decreasing degree is not necessary, but tends to reduce the number of colours which are needed.

Example 4.8. *The vertices in the following graph have been ordered by decreasing degree. Use the Welsh–Powell Algorithm to give a vertex colouring.*



Hence the chromatic number of the graph is $\square$.

There is an upper bound for the chromatic number of G in terms of $\Delta(G)$, the maximum of the vertex degrees. (We assume that the vertex set is nonempty, so $\Delta(G)$ make sense.) We start with a weak form of the result.

Theorem 4.9. Any graph G is $(\Delta(G)+1)$-colourable.

Proof. The proof is by induction on n, the number of vertices of G. In the case where G is a single vertex we have $\Delta(G) = 0$, and G is indeed 1-colourable. So we can assume that $n \geq 2$ and that the result is true for graphs with fewer than n vertices. Let v be any vertex of G. It is clear that $\Delta(G-v) \leq \Delta(G)$, so the induction hypothesis implies that $G-v$ is $(\Delta(G)+1)$-colourable; choose a vertex colouring of $G-v$ where the colour set has size $\Delta(G)+1$. Now the number of vertices of G adjacent to v is $\deg(v)$, which is at most $\Delta(G)$ by definition. So at most $\Delta(G)$ different colours are used among the vertices adjacent to v. Therefore there is a colour which is not used among these vertices, and we can colour v with this colour to obtain a vertex colouring of G. The induction step is complete. □

Another way to state this result is that $\chi(G) \leq \Delta(G)+1$. Note that equality holds in the cases where $G = K_n$ (since $\chi(K_n) = n$ and $\Delta(K_n) = n-1$) or $G = C_n$ for $n \geq 3$ odd (since $\chi(C_n) = 3$ and $\Delta(C_n) = 2$). It turns out that these are the only connected graphs for which equality holds.

Theorem 4.10 (Brooks)**.** Let G be a connected graph which is neither complete nor an odd cycle. Then G is $\Delta(G)$-colourable, so $\chi(G) \leq \Delta(G)$.

Proof.** The proof is again by induction on the number of vertices of G. The smallest case allowed by the hypotheses is when G has three vertices but is not complete (hence $G \cong P_3$); in this case $\chi(G) = \Delta(G) = 2$. So we can assume that G has at least four vertices, and that the result is known for graphs with fewer vertices. Let v be a vertex of G of minimal degree $\delta(G)$, and let d denote $\Delta(G)$. Let $G_1, \cdots, G_s$ be the connected components of $G-v$ (possibly $s = 1$); we claim that each G_i is d-colourable. Since $\Delta(G_i) \leq d$, this is part of the induction hypothesis unless G_i is complete or an odd cycle; but in these two cases we have $\Delta(G_i) < d$ (because G_i is regular and some vertex of G_i is adjacent to v), so the claim follows from Theorem 4.9. So $G-v$ is d-colourable; choose a vertex colouring of $G-v$ where the colour

set is $\{1, 2, \cdots, d\}$. We want to extend this vertex colouring to the whole of G (possibly after some modification). If G is not regular of degree d, then $\deg(v) = \delta(G) < d$, so there is a colour not used among the vertices adjacent to v, and we can colour v with this colour to obtain a vertex colouring of G as required. So henceforth we may assume that G is regular of degree d. If $d = 2$ this forces G to be a cycle (an even cycle, by assumption), in which case we know the result. So we may also assume that $d \geq 3$.

We now have various cases for the vertex colouring of $G - v$. To save words, we will omit to say in the description of each case that the previous cases do not hold, but that should always be understood.

Case 1: the vertices adjacent to v do not all have distinct colours. In this case there is a colour not used among these vertices, so as before we can extend the vertex colouring to G.

In the remaining cases, every colour is used exactly once among the vertices adjacent to v; we let v_i denote the unique vertex adjacent to v with colour i.

Case 2: for some i, the vertices in $G - v$ adjacent to v_i do not all have distinct colours. Since there are $d - 1$ such vertices, this means that there is some colour other than i which they do not use. We can then change the colour of v_i to this other colour, putting us back in Case 1.

In the remaining cases, we let v_{ij} denote the unique vertex in $G - v$ which is adjacent to v_i and has colour j, for all $i \neq j$.

Case 3: for some $i \neq j$, there is no walk in $G - v$ from v_i to v_j which uses only vertices coloured i and j. Let H be the subgraph of $G - v$ formed by the vertices with colours i and j, and whatever edges of G join such vertices. Our assumption means that v_i and v_j belong to different connected components of H. We can then swap the colours i and j throughout the component containing v_j, and still have a vertex colouring. After this change, v_j and v_i both have colour i, putting us back in Case 1.

In the remaining cases, we let H_{ij} denote the unique connected component of the subgraph of $G - v$ formed by vertices coloured i and j which contains the particular vertices v_i and v_j.

Case 4: for some $i \neq j$, H_{ij} is not a path between v_i and v_j. Hence as you walk in H_{ij} from v_i (without back-tracking), there is some vertex w where you first have a choice about which vertex to walk to next. Note that w is not v_i itself, because the first step of the walk must be to v_{ij} (similarly, $w \neq v_j$). Thus the degree of w in H_{ij} is at least 3, so there are $\leq d-3$ vertices adjacent to w with colours other than i and j, and there must be some third colour not used among these vertices. If we change w to this third colour it has the effect of disconnecting H_{ij}, putting us back in Case 3.

Case 5: for some distinct i, j, k, H_{ij} and H_{ik} have a vertex in common other than v_i. Let this vertex be u; its colour must clearly be i. Since u has two neighbours coloured j on the path H_{ij} and two neighbours coloured k on the path H_{ik}, we must have $d \geq 4$; moreover, there must be a fourth colour not used among the vertices adjacent to u. If we change u to this fourth colour we disconnect H_{ij} and H_{ik}, putting us back in Case 3.

Case 6: for some $i \neq j$, v_i is not adjacent to v_j. In particular, this implies that $v_{ij} \neq v_j$. Let k be any third colour. We can swap the colours i and k throughout the path H_{ik}, and still have a vertex colouring of $G - v$. We claim that this new vertex colouring falls into one of the previous cases; assume for a contradiction that it does not. One of the changes made in the new colouring is that the old v_i is the new v_k and vice versa. So the old v_{ij} (whose colour is still j, and which is still not equal to v_j) is the new v_{kj}. On the other hand, the old v_{ij} was joined to v_j by the path $H_{ij} - v_i$, and none of the vertices on that path had their colours changed (because none of them belonged to H_{ik}, by virtue of us not having been in Case 5). So the new v_{kj} is joined to v_j by a path whose vertices have colours i and j, and it thus belongs to both the new H_{ij} and the new H_{kj}, contradicting the assumption that our new colouring did not fall into Case 5.

In the only case remaining, we know that all the v_i's are adjacent to each other. So the d neighbours of v_i in G are exactly v and the other v_j's; this shows that G consists of just the vertices $v, v_1, \cdots, v_d$, and is a complete graph. This contradicts our assumption, so the last case vanishes and the induction step (remember that this was all inside the induction step!) is finished. □

If you think the proof of Brooks' Theorem was long, you'll be relieved that we are only going to mention the most famous result about vertex colourings, whose proof runs to hundreds of pages.

Theorem 4.11 (Four-Colour Theorem)**.** Every planar graph is 4-colourable.

A graph is said to be planar if you can draw a picture of it (on an ordinary piece of paper) in such a way that the edges don't cross at non-vertex points. (For instance, if you try to draw such a picture of K_5, you will find it can't be done; which is just as well for the sake of the Four-Colour Theorem, because we know that K_5 is not 4-colourable.) The notion of planarity is substantially different from any property of graphs we have considered in this course, because it brings in ideas of topology: it focuses on the spatial properties of a picture of a graph, rather than the abstract adjacencies which the picture represents. These topological properties of graphs are explored further in MATH3061 Geometry and Topology.

4.2 The chromatic polynomial

In studying colourability of a graph, it is natural to ask how many different vertex colourings a graph has with a given list of colours.

Definition 4.12. If G is a graph and $t \in \mathbb{N}$, the chromatic polynomial $P_G(t)$ is the number of vertex colourings of G with a fixed colour set of size t.

Thus G is t-colourable if and only if $P_G(t) \neq 0$. The name discloses one of the unexpected facts about $P_G(t)$, namely that it is a polynomial function of t; we will prove this later.

Example 4.13. *Let $G = K_3$ (which is the same as C_3). We know that K_3 has no vertex colourings with fewer than 3 colours, so $P_{K_3}(0) = P_{K_3}(1) = P_{K_3}(2) = 0$. If our colour set has three elements – say red, white, and blue – then to get a vertex colouring, we must assign a different colour to each vertex; the number of ways of doing this is $3! = 6$, so $P_{K_3}(3) = 6$. In fact, for a colour set with t elements, the number of vertex colourings of K_3 is*

just the number of ways of making an ordered selection of 3 colours from t possibilities with repetition not allowed, i.e. $P_{K_3}(t) = t_{(3)} = t(t-1)(t-2)$. Recall the reason for this formula: there are t choices for the colour of vertex 1, then $t-1$ choices for the colour of vertex 2 (because it can't be the same as that of vertex 2), then $t-2$ choices for the colour of vertex 3 (because it can't be the same as either of the two colours already used). Notice that the formula $t(t-1)(t-2)$ gives the right answer 0 for $t = 0, 1, 2$ also; so $P_{K_3}(t)$ is indeed a polynomial in t.

Example 4.14. *Let G be the path graph P_3. Given a colour set with t colours, the number of vertex colourings of P_3 is $t(t-1)(t-1) = t(t-1)^2$, by similar reasoning: there are t choices for the colour of vertex 1, then $t-1$ choices for the colour of vertex 2, then $t-1$ choices for the colour of vertex 3 (which can be anything except that colour of vertex 2; notice that there is no reason you can't re-use the colour of vertex 1 for vertex 3). The fact that 0 and 1 are roots of the polynomial $t(t-1)^2$, while 2 is not, is another way of saying that $\chi(P_3) = 2$.*

These examples can be generalized as follows.

Theorem 4.15. Let G be a graph with $n \geq 1$ vertices, and $t \in \mathbb{N}$.

(1) If G has connected components $G_1, \cdots, G_s$, $P_G(t) = P_{G_1}(t) \cdots P_{G_s}(t)$.

(2) If G is a complete graph, $P_G(t) = t(t-1)(t-2)\cdots(t-n+1)$.

(3) If G is a tree, $P_G(t) = t(t-1)^{n-1}$.

Proof. Part (1) follows from the fact that choosing a vertex colouring of G is the same as independently choosing vertex colourings of all the G_i's. If G is complete, then any vertex colouring must use different colours for all the vertices. So the number of vertex colourings is the number of ordered selections of n colours from t possibilities with repetition not allowed, which is given by $t_{(n)}$ as stated in part (2).

We prove part (3) by induction on n. If $n = 1$ (i.e. G consists of a single vertex), then obviously $P_G(t) = t$ as claimed. Assume that $n \geq 2$ and that the result is known for trees with fewer vertices. Let v be a leaf of G, and w

the unique vertex it is adjacent to. Any vertex colouring of G is obtained by making a vertex colouring of $G - v$ and then choosing a colour for v which is not the same as that of w. By the induction hypothesis, the number of vertex colourings of $G-v$ is $t(t-1)^{n-1}$, and no matter what vertex colouring of $G - v$ is chosen, the number of ways of choosing the colour for v is $t - 1$. So the number of vertex colourings of G is $t(t-1)^{n-2}(t-1) = t(t-1)^{n-1}$, completing the induction step. □

Remark 4.16. *For obvious reasons, we are not vitally concerned with the graph whose vertex set is empty. But according to the standard conventions, this graph has a single vertex colouring for any colour set, so its chromatic polynomial is* $1 = t^0$ *(and its chromatic number is* 0*).*

Chromatic polynomials in general can sometimes be computed simply by using the Product Principle, as in Theorem 4.15.

Example 4.17. *Again let* G *be the graph in Example 3.25. We have a subgraph isomorphic to* K_3 *formed by the vertices* 1, 2, *and* 4*; the number of vertex colourings of this subgraph is* $t(t-1)(t-2)$*, as seen above. Given any such vertex colouring, the number of ways to extend it to a vertex colouring of* G *is* $t - 2$*, because the colour of vertex* 3 *can be anything except the colours of vertices* 1 *and* 4 *(which are definitely different from each other). So* $P_G(t) = t(t-1)(t-2)^2$*.*

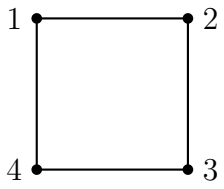
The generalization is:

Theorem 4.18. Let G be a graph and t a nonnegative integer. If v is a vertex of G such that the vertices adjacent to v are all adjacent to each other, then $P_G(t) = P_{G-v}(t)(t - \deg(v))$.

Proof. In any vertex colouring of $G - v$, the $\deg(v)$ vertices adjacent to v must all have different colours, since by assumption they are all adjacent to each other. So if $t < \deg(v)$, there are no vertex colourings of $G - v$, and therefore none of G; so both sides are zero. If $t \geq \deg(v)$, then for any vertex colouring of $G-v$ there are $t-\deg(v)$ ways to choose a colour for the vertex v, so as to extend it to a vertex colouring of G. The result follows. □

However, there need not be any vertex which satisfies the condition in Theorem 4.18. In such cases we may need the Sum Principle as well.

Example 4.19. *Consider the cycle graph C_4:*



If we try to use the same sort of Product Principle arguments to compute $P_{C_4}(t)$, we run into a slight problem. If we first choose the colours of vertices 1, 2, and 3 (in t, $t-1$, and $t-1$ ways respectively), then when we come to choose the colour of vertex 4 there could be either $t-1$ or $t-2$ possibilities, depending on whether the colours we have chosen for vertices 1 and 3 are the same or not. We need to consider each case separately:

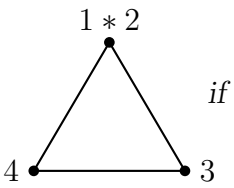
$$\begin{array}{rcl} \textit{no. of ways to colour } 1,2,3 \\ \textit{so that } 1 \textit{ and } 3 \textit{ have the same colour} & = & \boxed{} \\ \textit{no. of ways to colour } 1,2,3 \\ \textit{so that } 1 \textit{ and } 3 \textit{ have different colours} & = & \boxed{} \end{array}$$

So $P_{C_4}(t) = \boxed{}$.

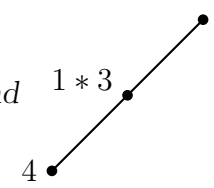
The logic of Example 4.19 can be applied more generally using the following definition.

Definition 4.20. Let G be a graph, and let v, w be distinct vertices of G. The graph obtained from G by fusing v and w, written $G[v, w]$, is the graph whose vertices are the same as those of G except that v and w are replaced by a single vertex $v * w$. The edges between vertices other than $v * w$ are the same as in G, and $v * w$ is adjacent to another vertex u if and only if either $\{u, v\}$ or $\{u, w\}$ is an edge of G (or both are).

Example 4.21. *If you start with the graph C_4, you obtain:*



if you fuse 1 and 2, and



if you fuse 1 and 3.

Theorem 4.22. Let G be a graph and t a nonnegative integer.

(1) If v and w are non-adjacent vertices of G, then

$$P_G(t) = P_{G+\{v,w\}}(t) + P_{G[v,w]}(t).$$

(2) If v and w are adjacent vertices of G, then

$$P_G(t) = P_{G-\{v,w\}}(t) - P_{G[v,w]}(t).$$

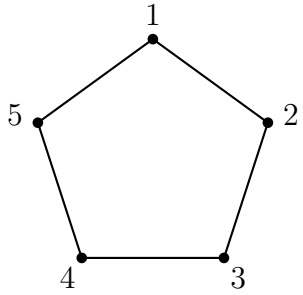
Proof. If v and w are not adjacent, then in a vertex colouring of G there is no constraint on whether the colours of v and w are different or not. Thus the set of all vertex colourings of G (with a fixed colour set of size t) is the disjoint union of the set of vertex colourings of G in which the colours of v and w are different, and the set of vertex colourings of G in which the colours of v and w are the same. The vertex colourings of G in which the colours of v and w are different are obviously in bijection with the vertex colourings of $G+\{v,w\}$, because the extra edge $\{v,w\}$ imposes exactly the condition that the colours of v and w are different. The vertex colourings of G in which the colours of v and w are the same are obviously in bijection with the vertex colourings of $G[v,w]$, because if their colours have to be the same, the two vertices may as well be fused into one. Part (1) follows.

For part (2), we apply part (1) to $G-\{v,w\}$, obtaining

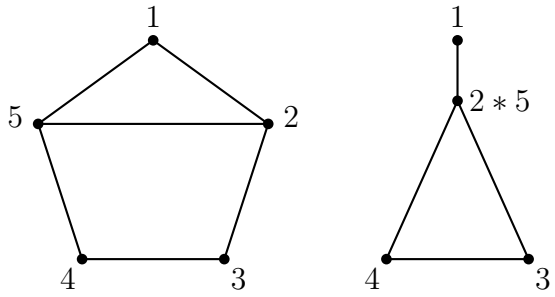
$$P_{G-\{v,w\}}(t) = P_G(t) + P_{G[v,w]}(t),$$

because it makes no difference to the fused graph $G[v,w]$ whether v and w are adjacent in G or not. Rearranging this equation gives the result. □

Example 4.23. *Consider the graph C_5.*



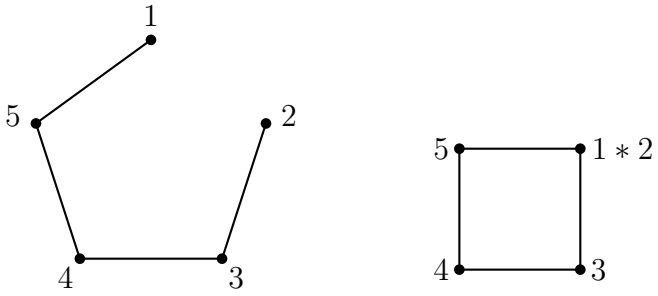
Applying part (1) of Theorem 4.22 with the vertices 2 *and* 5*, we see that* $P_{C_5}(t)$ *is the sum of the chromatic polynomials of the following graphs:*



In both of these graphs, Theorem 4.18 applies to vertex 1*. Hence their chromatic polynomials are* $t(t-1)(t^2-3t+3)(t-2)$ *and* $t(t-1)(t-2)(t-1)$ *respectively, and we thus find that*

$$P_{C_5}(t) = t(t-1)(t-2)[(t^2-3t+3)+(t-1)] = t(t-1)(t-2)(t^2-2t+2).$$

An alternative method is to apply part (2) of Theorem 4.22 to the vertices 1 *and* 2*. This tells us that* $P_{C_5}(t)$ *is the difference between the chromatic polynomials of the following graphs:*



The first of these being a tree, and the second isomorphic to C_4*, their chromatic polynomials are* $t(t-1)^4$ *and* $t(t-1)(t^2-3t+3)$ *respectively, so*

$$P_{C_5}(t) = t(t-1)[(t^3-3t^2+3t-1)-(t^2-3t+3)] = t(t-1)(t^3-4t^2+6t-4),$$

which is the same polynomial as found above.

If necessary, we can apply Theorem 4.22 recursively to the graphs appearing on the right-hand side. Using part (2) repeatedly, we must eventually reduce to graphs whose chromatic polynomials we know, such as trees or forests, because the number of edges keeps decreasing. Alternatively, using part (1)

repeatedly, we must eventually reach complete graphs, because the number of non-edges keeps decreasing. As a theoretical consequence of this, we can finally justify the terminology "chromatic polynomial".

Theorem 4.24*. Let G be a graph with $n \geq 1$ vertices. There are unique nonnegative integers $a_1, \cdots, a_{n-1}$, depending on G, such that for any $t \in \mathbb{N}$,

$$P_G(t) = t^n - a_{n-1}t^{n-1} + a_{n-2}t^{n-2} - \cdots + (-1)^{n-1}a_1 t.$$

Moreover, a_{n-1} is the number of edges of G.

Proof*. The proof is by induction on the number of edges of G. If G has no edges, then G has n connected components, all single vertices, and its chromatic polynomial is clearly t^n; the result holds with all $a_i = 0$. So we can assume that G has at least one edge, and that the result is known for graphs with fewer edges. If $\{v, w\}$ is an edge of G, then $G - \{v, w\}$ and $G[v, w]$ both have fewer edges than G, so by the induction hypothesis,

$$\begin{aligned} P_{G-\{v,w\}}(t) &= t^n - b_{n-1}t^{n-1} + b_{n-2}t^{n-2} - \cdots + (-1)^{n-1}b_1 t, \text{ and} \\ P_{G[v,w]}(t) &= \qquad\qquad t^{n-1} - c_{n-2}t^{n-2} + \cdots + (-1)^{n-2}c_1 t, \end{aligned}$$

for some nonnegative integers $b_1, \cdots, b_{n-1}$ and $c_1, \cdots, c_{n-2}$ (note that $G[v, w]$ has only $n - 1$ vertices). By part (2) of Theorem 4.22, we get

$$P_G(t) = t^n - (b_{n-1} + 1)t^{n-1} + (b_{n-2} + c_{n-2})t^{n-2} - \cdots + (-1)^{n-1}(b_1 + c_1)t.$$

So the desired equation holds with $a_i = b_i + c_i$ for $i = 1, \cdots, n-2$ and $a_{n-1} = b_{n-1} + 1$. Moreover, the induction hypothesis includes the fact that b_{n-1} is the number of edges of $G - \{v, w\}$, so a_{n-1} is the number of edges of G. Finally, the a_i's must be unique, because two polynomials in the variable t which agree for infinitely many values of t (namely, all nonnegative integers) must have the same coefficients. □

Remark 4.25. *Remember that $P_G(t) = 0$ for all nonnegative integers $t < \chi(G)$, so the polynomial in Theorem 4.24 has $\chi(G)$ known roots. In particular, if G has any edges then 1 must be a root of the polynomial, so*

$$1 - a_{n-1} + a_{n-2} - \cdots + (-1)^{n-1}a_1 = 0.$$

It is not known in general which sequences of nonnegative integers arise as the sequence $(a_1, \cdots, a_{n-1})$ for some graph G.

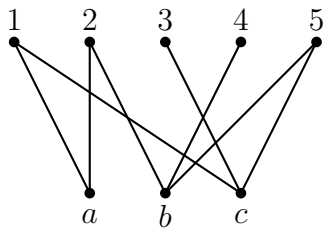
4.3 Edge colourings

An important class of scheduling problems involves 'appointments'. Imagine a number of students who need to have one-on-one meetings with various lecturers, where each lecturer may be required for several meetings. If we are given the list of which lecturers each student needs to meet with, how can we schedule the meetings so that the whole thing takes as little time as possible? If we adopted our previous approach, we would construct a graph which has a vertex for each meeting that has to take place (i.e. each pair of a student and a lecturer who need to meet), and make two vertices adjacent (thus ensuring that they are not scheduled simultaneously) if they have either a student in common or a lecturer in common. We then need to find the chromatic number of this graph, which may not be easy.

A better alternative is to consider the graph which is naturally suggested by the problem: namely, where there are vertices corresponding to the students, other vertices corresponding to the lecturers, and edges joining each student with every lecturer whom they need to meet. Note that this graph is bipartite: there are no edges joining a student to another student, or a lecturer to another lecturer. Since it is now the edges, not the vertices, which represent the meetings that need to be scheduled, our question amounts to finding the edge chromatic number of the graph, in the following sense.

Definition 4.26. Let G be a graph with edge set E. An edge colouring of G is a function $c : E \to C$, where C is a finite set whose elements are called colours, with the property that no two edges with the same colour have an end in common. We say that G is t-edge-colourable if there exists an edge colouring of G where $|C| = t$. The smallest t such that G is t-edge-colourable is called the edge chromatic number of G and written $\chi'(G)$.

Example 4.27. *Suppose that students* $1, 2, 3, 4, 5$ *need to have meetings with lecturers* a, b, c *according to the edges of the following graph:*



Since there are three students who need to meet lecturer b, there will have to be at least three meeting timeslots. Can all the meetings be scheduled in only three timeslots? Try to find an edge colouring of the graph with three colours (say red, white, and blue). The conclusion is that the edge chromatic number of this graph is $\boxed{}$.

Here are some obvious principles about edge colourings.

Theorem 4.28. Let G be a graph and t a nonnegative integer.

(1) If $t < \Delta(G)$, then G is not t-edge-colourable. Hence $\chi'(G) \geq \Delta(G)$.

(2) If H is a subgraph of G and G is t-edge-colourable, then H is t-edge-colourable. Hence $\chi'(H) \leq \chi'(G)$.

(3) If G has connected components $G_1, \cdots, G_s$, then G is t-edge-colourable if and only if each G_i is t-edge-colourable. Hence $\chi'(G) = \max\{\chi'(G_i)\}$.

Proof. By definition of $\Delta(G)$, there is some vertex where $\Delta(G)$ edges all end; these edges must have different colours in any edge colouring, which implies part (1). Parts (2) and (3) follow from the same reasoning as in the case of vertex colourings (see parts (3) and (4) of Theorem 4.4). □

A graph G has edge chromatic number 0 if and only if it has no edges, i.e. $\Delta(G) = 0$. If $\Delta(G) = 1$, i.e. the graph does have edges but no two edges ever end at the same vertex, then obviously the edges can all be given the same colour, so the edge chromatic number is 1. In these cases, the inequality in part (1) of Theorem 4.28 is actually equality. But there are examples where equality fails, i.e. graphs G which are not $\Delta(G)$-edge-colourable.

Example 4.29. *Consider the cycle graph C_n for $n \geq 3$. An edge colouring of C_n requires at least $\Delta(C_n) = 2$ colours. In any edge colouring with 2 colours, the colours must alernate as you go around the cycle; if n is odd, this results in a contradiction, exactly as with vertex colourings of C_n. So the edge chromatic number is actually the same as the (vertex) chromatic number of C_n, namely 2 if n is even and 3 if n is odd.*

We can find the edge chromatic numbers of complete graphs as follows.

Theorem 4.30. For any $n \geq 2$,

$$\chi'(K_n) = \begin{cases} n & \text{if } n \text{ is odd,} \\ n-1 & \text{if } n \text{ is even.} \end{cases}$$

Proof*. First consider the case of odd $n \geq 3$. We have $\Delta(K_n) = n-1$. However, we can prove that K_n is not $(n-1)$-edge-colourable as follows. Suppose for a contradiction that we had an edge colouring of K_n with $n-1$ colours. Since there are $\binom{n}{2} = \frac{n(n-1)}{2}$ edges, the Pigeonhole Principle tells us that there must be some colour for which there are $\lceil \frac{n}{2} \rceil = \frac{n+1}{2}$ edges with that colour. But then the $n+1$ ends of those $\frac{n+1}{2}$ edges must all be different, which is a contradiction since K_n has n vertices.

We can prove that K_n is n-edge-colourable by a direct construction. We let the colour set be $\{1, 2, \cdots, n\}$, the same as the vertex set. Picture the vertices as the vertices of a regular n-gon, numbered clockwise from 1 to n. We extend the numbering of vertices cyclically, so that vertex 0 is vertex n, vertex -1 is vertex $n-1$, vertex $n+1$ is vertex 1, and so forth. Then the edges which we colour with colour i are those where the ends are equidistant from i: namely, $\{i-1, i+1\}$, $\{i-2, i+2\}$, and so on until we reach $\{i-\frac{n-1}{2}, i+\frac{n-1}{2}\}$, which is the edge of the n-gon directly opposite i. These edges clearly have no ends in common (in fact, in the picture they are all parallel). Moreover, each edge of K_n gets a unique colour in this way, because for any two vertices there is a unique third vertex from which they are equidistant (note that this would not hold if n was even). So K_n is n-edge-colourable, and we have shown that $\chi'(K_n) = n$ for n odd.

Now consider the case of even n. It is clear that $\chi'(K_2) = 1$, so we can assume that $n \geq 4$. Since $\Delta(K_n) = n-1$, we just need to prove that K_n is $(n-1)$-edge-colourable. But if we delete the vertex n, the result is the complete graph K_{n-1}, and we have just seen how to construct an edge colouring of K_{n-1} with colours $1, \cdots, n-1$. It is clear from our construction that none of the edges coloured with colour i ends at the vertex i. So when we add the vertex n, we can colour the edge $\{i, n\}$ with colour i for $i = 1, \cdots, n-1$, and the result is an edge colouring of K_n with colours $1, \cdots, n-1$. The proof is finished. □

Remark 4.31*. *As hinted at the beginning of this section, the problem of finding edge colourings of G can be rephrased in terms of vertex colourings. Let $\widehat{G}$ denote a graph whose vertices correspond to the edges of G, where two vertices are adjacent if and only if the corresponding edges of G have an end in common. Then the edge colourings of G correspond to the vertex colourings of $\widehat{G}$, and $\chi'(G) = \chi(\widehat{G})$. For instance, $\widehat{C_n}$ happens to be isomorphic to C_n, which is why the computation of $\chi'(C_n)$ in Example 4.29 was essentially the same as the computation of $\chi(C_n)$. The reason for not converting everything into vertex colourings is just that our general results about (vertex) chromatic numbers are a bit weak. For instance, $\widehat{K_n}$ for $n \geq 4$ is a graph with $\binom{n}{2}$ vertices, which is regular of degree $2n-4$ (every edge of K_n has one of its ends in common with $n-2$ other edges, and the other end in common with $n-2$ further edges). The best general result for such a graph, Brooks' Theorem, would tell us only that $\chi(\widehat{K_n}) \leq 2n-4$, whereas Theorem 4.30 determines $\chi'(K_n)$ exactly.*

The main result about edge colourings is surprisingly conclusive.

Theorem 4.32 (Vizing)**.** Any graph G is $(\Delta(G)+1)$-edge-colourable. So $\chi'(G)$ is always either $\Delta(G)$ or $\Delta(G)+1$.

Thus the whole universe of graphs is divide into two types, those for which $\chi'(G) = \Delta(G)$ and those for which $\chi'(G) = \Delta(G)+1$. For instance, Example 4.29 says that even cycles are the first type whereas odd cycles are the second type; Theorem 4.30 says that complete graphs with an even number of vertices are the first type, whereas complete graphs with an odd number of vertices (at least 3) are the second type.

We will not give the general proof of Vizing's Theorem, but some idea of the argument is provided by the following result, which shows that all bipartite graphs (including all trees and forests) are of the first type.

Theorem 4.33. Any bipartite graph G is $\Delta(G)$-edge-colourable; hence its edge chromatic number is $\Delta(G)$.

Proof*. The proof is by induction on the number of edges of G; we know the result if there are no edges, so assume that G has some edges and that

the result holds for bipartite graphs with fewer edges than G. Let $\{v, w\}$ be an edge of G. By the induction hypothesis, the graph $G - \{v, w\}$ (which is clearly still bipartite) has an edge colouring with $\Delta(G - \{v, w\})$ colours, so it has an edge colouring with $\Delta(G)$ colours. We choose such an edge colouring of $G - \{v, w\}$ where the colour set is $\{1, 2, \cdots, \Delta(G)\}$; we want to extend this edge colouring to the whole of G (possibly after some modification).

Since the degree of v in $G - \{v, w\}$ is at most $\Delta(G) - 1$, there must be at least one colour not used among the edges ending at v; similarly, there must be at least one colour not used among the edges ending at w. If there is any colour which is not used either among the edges ending at v or among the edges ending at w, we can simply colour the edge $\{v, w\}$ with that colour, and we are finished. So we can assume that there is a colour i which is used among the edges ending at v but not among the edges ending at w, and another colour j which is used among the edges ending at w but not among the edges ending at v.

Now starting from the vertex $v = v_0$, we can build a path in $G - \{v, w\}$ where the edges alternate between the colours i and j: the first edge $\{v_0, v_1\}$ is the unique edge ending at v of colour i, the next edge $\{v_1, v_2\}$ is the unique edge ending at v_1 of colour j (if such an edge exists), the next edge $\{v_2, v_3\}$ is the unique edge ending at v_2 of colour i (if such an edge exists), and so on until the next required edge does not exist. This is indeed a path, because if any vertex was ever repeated, we would have three edges ending at a vertex with only two colours between them, contradicting the edge colouring property (or if v_0 was the repeated vertex, we would have two edges ending there with the colour i, because we assumed the colour j is not used by any edge ending there). If w belonged to this path, it would have to be the other end-vertex, coming after an edge of colour j (there being no edge of colour i with which to continue the path). But this would mean that $w = v_m$ for some even integer m, and then we could add the edge $\{v, w\}$ to form a cycle in G with an odd number of vertices (namely, $m + 1$); since G is bipartite, it contains no odd cycles, so this cannot be the case. So w does not belong to the path. We can now exchange the colours i and j throughout the path; this clearly still gives an edge colouring of $G - \{v, w\}$, in which the colour i is not used among the edges ending at v or the edges ending at w. We can then colour $\{v, w\}$ with colour i, and we are finished. □

Note that this proof gives a concrete recursive procedure for finding an edge colouring of G with $\Delta(G)$ colours.

Returning to the situation of students and lecturers imagined at the start of the section, we can now conclude that the minimum number of timeslots required for all the meetings is the obvious lower bound, namely the largest number of meetings that any single student or lecturer is involved in (as seen in Example 4.27).

Example 4.34. *A special case of Theorem 4.33 is that the complete bipartite graph $K_{p,p}$ has an edge colouring with p colours. We can display such an edge colouring in a $p \times p$ array, where the entry in row i and column j is the colour of the edge between vertices i and $p + j$. The fact that no two edges ending at vertex i have the same colour means that no two entries in row i are the same; the fact that no two edges ending at vertex $p + j$ have the same colour means that no two entries in column j are the same. So every row of the array contains each of the p colours exactly once, and so does every column; such an array is called a 'Latin square', and we have proved that Latin squares of all sizes exist. The $p = 9$ case is familiar to anyone who has ever done a Sudoku puzzle!*

Since it is traditional for books to end with a wedding, we will use Theorem 4.33 to prove a special case of what is called the Marriage Theorem.

Example 4.35. *Suppose there are n men and n women, and every woman has a list of which of the men she would be happy to marry. Every woman's list has d men on it, and every man occurs on d lists, where $d \geq 1$. Is there a way to marry the men off to the women so that every woman gets one of the men on her list? If we construct a graph G where the vertices are the people and there is an edge between a man and a woman if the man is on the woman's list, then G is bipartite and regular of degree d. By Theorem 4.33, there is an edge colouring of G with d colours. Let i be any colour; since the edges coloured i cannot have any men or women in common, there are at most n of them. But the total number of edges is nd, so in fact there must be exactly n edges of colour i for every i. Thus for every i, the edges of colour i join n men to n women, and thus provide a way to marry them off. So the answer is that there is not just one way; you can actually find d different ways which have no weddings in common.*