Combinatorics

Course Notes

July 7, 2024

Manuel Bodirsky, Institut für Algebra, TU Dresden

Disclaimer: this is a draft and probably contains many typos and mistakes.
Please report them to manuel.bodirsky@tu-dresden.de.

Contents

Preface

These are course notes for a course offered at TU Dresden for bachelor students of mathematics in their third year and computer science diploma students who study mathematics as their second subject. Combinatorics is a vast field and there is no canonical selection of material. When selecting material, I have focussed on *methods* rather than attempting some encyclopaedic approach. For each method, I try to identify prototypical results, which I then prove in the simplest possible form that still reveals the *idea*. In many applications that the students might encounter in the future, stronger forms of the results will be needed; for such situations, I always give references to more advanced textbooks.

The text starts with a short introduction to *graphs*, since graphs are a great playground on which we may test our methods. Each of the remaining chapters is about one particular method in combinatorics. Our second chapter is about *duality*, with the linear programming duality theorem at its core. This is a topic that is often taught in optimisation or in theoretical computer science because of its computational consequences. Our focus is on the mathematical consequences of LP duality: many other results, e.g. in graph theory, can be derived from it in one way or the other. I illustrate this with results about matchings in graphs, or the minimax theorem of von Neumann for zero-sum games, or finding the winner in Condon's simple stochastic games.

The next chapter is about the probabilistic method in combinatorics and its many stunning applications. Our warm-up application is an extremely simple probabilistic existence proof for so-called *k-paradoxical tournaments*. A more advanced application is the proof of Erdős that for all $k, \ell \in \mathbb{N}$ there exists a finite graph that has chromatic number k but no cycles of length at most ℓ. This result itself has many (generalisations and) applications. Interestingly, the simplest known proof of this theorem is via the probabilistic method (despite the fact that there is no probability in the formulation of the result). The probabilistic proof requires random graphs with a carefully chosen edge probability. If you are exposed for the first time to this proof the choice of the edge probability might 'fall from heaven'. Therefore, before proving Erdős's theorem, we first give an introduction to random graphs to gain some intuition. This is why the chapter first treats the landscape of the so-called Erdős-Rényi evolution of the random graph $G(n, p)$ with n vertices and edge probability p. Here, p is a function that depends on n! Depending on the asymptotic growth of p the random graph $G(n, p)$ may have various properties that hold almost surely, for n tending to infinity. Many of these properties have a sharp *threshold*: the probability that these properties hold for large n jumps from 0 to 1. When discussing these properties, we concentrate on two types of arguments, the first moments method and the second moment method (there are many more advanced techniques but they are out of the scope of this short introduction).

The chapter on Ramsey theory contains another application of the probabilistic method, for proving Ramsey lower bounds. But Ramsey theory itself is an extremely useful tool in many other areas of combinatorics and mathematics in general so that

it might also qualify as a method and has its own chapter. We focus on two results: Ramsey's theorem, which I prove via its infinite version. The finite version is then derived by a typical compactness argument. To keep things simple, I only give the proof of the graph version of Ramsey's theorem, but I state (and use!) the theorem in its general version later. Two applications of Ramsey's theorem, in number theory and in geometry, are discussed. The second big result of the chapter is the theorem of Hales-Jewett, which plays a central role in Ramsey theory. Historically, it has been motivated by positional games, which is a beautiful topic in combinatorics that also creates some links with what we have earlier learned about matchings. I present two more applications of Hales-Jewett, namely a proof of the famous van der Waerden theorem about the existence of monochromatic arithmetic progressions, and another Ramsey-type theorem for graphs (and many more applications can be found in the exercises).

The fourth method that I present are generating functions, used in enumerative combinatorics. Generating functions can be viewed as formal power series and treated algebraically, or they can be viewed as functions over the complex numbers and treated analytically. Both perspectives have their advantages. In almost all the applications that we present (including more advanced topics, such as the Lagrange inversion formula) we have found purely algebraic treatments (those tend to be more elementary). However, I do not neglect the links to the analytic approach since these links can provide intuition and also connect well to other courses that the students have followed or will follow. With the generating function method, I can then treat very efficiently much of the material that would take central space in many combinatorics courses, namely counting

- words of a given length in a regular language (Chomsky-Schützenberger),
- binary trees of a given size (Catalan),
- equivalence relations on a given set of elements (Bell numbers and Dobiński's formula),
- labelled connected graphs with a given set of vertices,
- rooted labelled trees with a given set of vertices (Cayley's formula), and
- permutations of n elements (Stirling's formula).

I thank the participants of the course in the winter semester 2018/19 for their feedback, in particular Benedikt Bartsch, Janik Fechtelpeter, Paul Senf, Manuel Thieme, and Tony Zorman, and in particular Andrés Aranda who was teaching assistant. The course notes contain some definitions and fundamental facts that were not covered in detail in class because the participants had already seen these concepts in other courses. I have still added them to the notes because our goal was to keep the notes self-contained (for the convenience of the reader; it is always easy to skip these parts!). Thus, please also let me know if there are gaps, missing definitions and missing explanations, I am happy to fill them. The text contains 94 exercises; the ones with a star are harder.

Dresden, July 7, 2024
Manuel Bodirsky

CHAPTER 1

Graphs

There are *directed* und *undirected* graphs. We start with undirected graphs.

1.1. Undirected Graphs

We mostly follow the notation of the text book "Graph Theory" of Reinhard Diestel [**12**], which has both an English and a German version. For a set S we write $\binom{S}{k}$ for the set of all subsets of M with k elements (also called *k-subsets of S*). The notation is motivated by the identity

$$\left|\binom{S}{k}\right| = \binom{|S|}{k}.$$

DEFINITION 1.1.1. A (*simple*[1], *undirected*) *graph* G is a pair (V, E) where $V = V(G)$ is a set, called the *vertex set* and where $E = E(G) \subseteq \binom{V}{2}$ is called the *edge set*.

The elements of the vertex set of a graph G are also called the *vertices* or the *nodes* of G. A graph G is called *finite* if $V(G)$ is finite. If $\{u, v\} \in E(G)$ then u and v are called *adjacent*, and v is called a *neighbour* of u.

The number of neighbours of x in G is called the *degree* of x. We give a couple of examples of fundamental graphs along with their names. Let $n \in \mathbb{N}$ be a positive natural number.

- K_n denotes the graph (V, E) with $V := \{1, 2, \ldots, n\}$ and $E := \binom{V}{2}$. This graph is called the n-element *clique* (or the *complete* graph).
- I_n denotes the graph $(\{1, 2, \ldots, n\}, \emptyset)$ and is called *independet set* (or *stable set*) of size n.
- P_n, for $n \geq 2$, denotes the *path of length* n, that is, the graph (V, E) with $V := \{1, \ldots, n\}$ and $E := \{\{1, 2\}, \{2, 3\}, \ldots, \{n-1, n\}\}$. Warning: in some books and articles P_n denotes the graph with n edges, and not, as here, the path with n nodes.
- C_n, for $n \geq 3$, denotes the graph
$$\big(\{0, 1, \ldots, n-1\}, \big\{\{i, j\} \,|\, (i - j) \equiv 1 \pmod n\big\}\big)$$
called *cycle* (with n nodes und n edges).

The *complement* of a graph $G = (V, E)$ is the graph $\overline{G} = (V, \binom{V}{2} \setminus E)$. For instance the complement of I_n is K_n. Obviously, $\overline{(\overline{G})} = G$.

DEFINITION 1.1.2 (Isomorphism). Two graphs G and H are called *isomorphic* if there exists a bijection $f\colon V(G) \to V(H)$ such that $\{u, v\} \in E(G)$ if and only if $\{f(u), f(v)\} \in E(H)$.

For example the complement of C_5 is isomorphic to C_5.

[1]For the moment, our graphs also don't have *loops*; graphs without multiple edges or loops (whatever this is) are called *simple*.

DEFINITION 1.1.3 (subgraph). A graph H is called a *subgraph of* G if $V(H) \subseteq V(G)$ and $E(H) \subseteq E(G) \cap \binom{V(H)}{2}$. An *induced subgraph of* G is a graph H with $V(H) \subseteq V(G)$, and $E(H) = E(G) \cap \binom{V(H)}{2}$.

For $V \subset V(G)$ we write $G[V]$ for the (uniquely determined) induced subgraph of G with vertex set V, and call $G[V]$ the *subgraph of* G *induced by* V. A sequence $(u_1, u_2, \ldots, u_l)$ of nodes of a graph G is called a *walk* from u_1 to u_l in G if $\{u_i, u_{i+1}\} \in E(G)$ for all $i \in \{1, \ldots, l-1\}$. We allow the case $l = 1$; in this case the walk has only one vertex and no edges. A walk $(u_1, u_2, \ldots, u_l)$ is called *closed* if $u_1 = u_l$, and otherwise *open*. A walk $(u_1, u_2, \ldots, u_l)$ is called a *path* from u_1 to u_l if $u_i \neq u_j$ for distinct $i, j \in \{1, \ldots, l\}$. Note that if there is a walk from u to v then clearly there is also a path from u to v.

A *cycle* is a walk $(u_0, u_1, \ldots, u_{l-1}, u_l)$ with $u_0 = u_l$, $l \geq 3$ and $u_i \neq u_j$ for all distinct $i, j \in \{1, \ldots, l-1\}$. Note that a graph contains a cycle if and only if it contains a subgraph isomorphic to C_n for some $n \geq 3$.

Exercises.

(1) Let G be a graph, and for $u \in V(G)$ let d_u be the degree of u. Show that
$$|E(G)| = \frac{1}{2} \sum_{u \in V(G)} d_u.$$
(2) Suppose that a graph has 12 edges and 6 vertices, each of which has degree 3 or 5. How many vertices are there of each degree?
(3) Show that if a graph G is not connected, then its complement is connected.

DEFINITION 1.1.4. A *directed graph* (short *digraph*) is a pair (V, E) where $V = V(G)$ is a vertex set and where $E = E(G) \subseteq V^2$ is a set of (directed) edges.

Edges of the form (u, u) for $u \in V$ are called *loops*. A directed graph is called *symmetric* if for every $(u, v) \in E$ we have $(v, u) \in E$. There is an natural bijection between symmetric directed graphs without loops, and undirected graphs, which is why directed graphs may be viewed as a generalisation of undirected graphs. Directed (and therefore also undirected) graphs may be represented by its *adjacency matrix* $A = (a_{i,j})_{i,j \in V}$ where for all vertices $u, v \in V$ the entry $a_{u,v}$ is one if there is an edge from u to v, and zero otherwise.

1.2. Connectivity

A graph G is called *connected* if for all $s, t \in V(G)$ there is a walk from s to t in G. Let $G = (U, E)$ and $H = (V, F)$ be two graphs with disjoint vertex sets. Then $G \uplus H$ denotes the graph $(U \cup V, E \cup F)$, called the *disjoint union* of G and H. The following is easy to see.

LEMMA 1.2.1. *A graph G is connected if and only if it cannot be written as $H_1 \uplus H_2$ for graphs H_1, H_2 with at least one vertex.*

A *connected component* of a graph G is a connected induced subgraph C of G such that for any $v \notin C$, the graph $G[C \cup \{v\}]$ is not connected. Clearly, every graph can be written as a disjoint union of its connected components.

More generally, a graph $G = (V, E)$ is k-connected, for $k \geq 1$, if for any $S \subseteq V$ with $|S| = k-1$ the graph $G[V \setminus S]$ is connected. So we see that a graph is connected if and only if it is 1-connected. We state the following without proof; three different proofs can be found in [**12**]. Two paths $(u_1, \ldots, u_k)$ and $(v_1, \ldots, v_l)$ from $a = u_1 = v_1$ to $b = u_k = v_l$ are called *independent* if $\{u_1, \ldots, u_k\} \cap \{v_1, \ldots, v_l\} = \{a, b\}$.

THEOREM 1.2.2 (Menger's theorem). *A finite graph $G = (V, E)$ is k-connected, for $k \geq 1$, if and only if for all $a, b \in V$ there are at least k pairwise independent paths from a to b.*

Exercises.

(4) If a graph G has p vertices, and the degree of every vertex is at least $\lceil \frac{p-1}{2} \rceil$, then G is connected.

1.3. Colorability

A *k-colouring* of a graph G is a function

$$f\colon V(G) \to \{0, 1, \ldots, k-1\}$$

such that $f(u) \neq f(v)$ for all $\{u, v\} \in E(G)$. A graph G is called *k-colorable* (or *k-partite*) if there *exists* a k-colouring of G. When is a graph 2-colorable?

PROPOSITION 1.3.1. *A finite graph G is 2-colorable if and only if it contains no* odd cycles *(i.e., cycles of odd size).*

PROOF. Odd cycles are certainly not 2-colourable, and neither are graphs that contain odd cycles. So suppose that $G = (V, E)$ has no odd cycles. Note that G is 2-colourable if and only if all its connected components are 2-colourable. We color a connected component C of G as follows:

(1) Select an arbitrary vertex u in C and define $f(u) := 0$.
(2) For all $v \in N(u)$ define $f(v) := 1$.
(3) If $f(v)$ is defined for all $v \in C$ then f is the desired colouring.
(4) Otherwise, suppose that $f(w') = i \in \{0, 1\}$ and $w \in N(w')$.
Define $f(w) := 1 - i$.
(5) Continue with step 3.

Since C is finite, this procedure terminates after finitely many steps, and we have found the desired colouring. □

Note that the proof shows that there exists an efficient algorithm (which performs at most linearly many operations in $n+m$) that determines for a given graph whether it is 2-colorable.

Two-colorable graphs are also called *bipartite*. In other words, a graph G is bipartite if its vertex set can be partitioned into two independent sets A and B. The set $\{A, B\}$ is called a *bipartition* of G, and A und B are called *partition classes* (or *colour classes*).

EXAMPLE 1. $K_{n,m}$ denotes the *complete bipartite Graph* with partition classes $P_1 := \{1, \ldots, n\}$ and $P_2 := \{n+1, \ldots, m+n\}$, that is,

$$K_{n,m} = (P_1 \cup P_2, \big\{\{u, v\} \mid u \in P_1, v \in P_2\big\}).$$

△

When is a graph 3-colourable? For this we do not have a similarly elegant description as in Proposition 1.3.1. It is an (important) open problem (in fact, it is one of the *Millenium Problems* of the Clay Mathematics Institute; `http://www.claymath.org/millennium-problems/p-vs-np-problem`) whether there exists an efficient algorithm that tests for a given graph whether it is 3-colourable (the problem is *NP-complete*; see Appendix B).

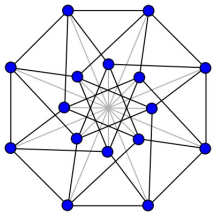

FIGURE 1.1. The Clebsch graph (see Exercise 5).

Exercises.

(5) Show that the Clebsch graph (see Figure 1.1) is 4-colourable, but not 3-colourable.

1.4. Trees

A *tree* is a connected graph (V, E) with at least one vertex and without cycles. More generally, a graph without cycles is called a *forest*. Clearly, from what we have learned in Section 1.2, every forest is a disjoint union of trees (so the degenerate graph without any vertex is considered to be a forest: it contains no cycles, and it is the union of 0 trees). Proposition 1.3.1 implies that trees (and forests) are 2-colourable.

A node in a tree with degree one is called a *leaf*. The following statements are easy to prove and left for the reader.

LEMMA 1.4.1. *Every finite tree with at least two nodes has a leaf.*

We write $G - x$ for the subgraph of G induced by $V(G) \setminus \{x\}$.

LEMMA 1.4.2. *Let $G = (V, E)$ be a finite connected graph with at least one node. Then the following are equivalent:*

(1) G is a tree.
(2) $|E| = |V| - 1$.
(3) $|E| \leq |V| - 1$.

LEMMA 1.4.3. *Let G be a graph. Then the following are equivalent.*

(1) G is a tree.
(2) G has maximally many edges without containing a cycle.
(3) G has minimally many edges with the property of being connected.
(4) Between any two nodes in G there exists a unique path.

1.5. Matchings

Let G be a graph. A *matching* (in G) is a subset M of $E(G)$ of pairwise disjoint edges: that is, for all $u, v \in M$ we have $u \cap v = \emptyset$. A *perfect matching* is a matching M with $2|M| = |V|$. If $\{x, y\} \in M$ then y is called the *partner* of x. If $S \subseteq V(G)$ then M is a *matching of* S if every element of S appears in an edge of M.

How can we find in G a matching of maximal size? Let M be any matching in G. A path in G whose edges alternate between edges from $E \setminus M$ and edges from M is called an *alternating path*. An alternating path P is called *augmenting* with respect to M if both the first and the last vertex of P have no partner in M. Augmenting paths can be used to obtained larger matchings than M. To formalise this we need the notion of a *symmetric difference* of two sets A and B: this is the set

$$A \Delta B \ := \ \{x \in A \cup B \mid x \notin A \cap B\} \ .$$

LEMMA 1.5.1. *Let M_1 and M_2 be matchings in $G = (V, E)$. Then the graph $(V, M_1 \Delta M_2)$ consists of a disjoint union of cycles of even length and of paths.*

Let M' be the symmetric difference of M and the augmenting path P. Then M' is again a matching and $|M'| > |M|$.

LEMMA 1.5.2 (Lemma of Berge). *Let G be a finite graph. A matching M in G is of maximal size if and only if there is no augmenting path with respect to M in G.*

PROOF. We have already seen that a matching with an augmenting path cannot be of maximal size. To prove the converse, suppose that G has a matching M' in G with $|M'| > |M|$. Since $|M'| > |M|$ the graph $(V, M'\Delta M)$ has a component with more edges from M' than from M. By Lemma 1.5.1 such a component must be an augmenting path with respect to M. □

If $S \subseteq V$ we write $N(S)$ for $\{n \in V \mid \{n, s\} \in E, \text{ for some } s \in S\}$, the *neighborhood* of S in G. An obvious necessary condition for the existence of a matching of $S \subseteq V$ in G is $|N(S)| \geq |S|$. This condition is in general not sufficient (example?). Theorem 1.5.3 presents a necessary *and* sufficient condition for the existence of a matching of S in G in *bipartite* (i.e., two-colorable) graphs.

In the following, let $G = (V, E)$ be a bipartite graph with a fixed bipartition A, B, so that $A \cup B = V$ and A, B are independent sets.

THEOREM 1.5.3 (Hall's marriage theorem). *A bipartite graph G has a matching of $A \subseteq V(G)$ if and only if $|N(S)| \geq |S|$ for all $S \subseteq A$.*

PROOF. Let M be a matching of G such that $a_0 \in A$ remains without partner. We will construct an augmenting path with respect to M. Let $a_0, b_1, a_1, b_2, a_2, \ldots$ be a sequence of maximal length of distinct vertices $a_i \in A$ and $b_i \in B$ such that

(1) $\{b_i, a_i\} \in M$, and
(2) b_i has an edge to a vertex $a_{f(i)} \in \{a_0, \ldots, a_{i-1}\}$.

Note that the i nodes $a_0, \ldots, a_{i-1}$ together have at least i neighbours in B, so we can always find an edge $\{a, b\} \in E$ such that $a \in \{a_0, \ldots, a_{i-1}\}$ and $b \in B \setminus \{b_1, \ldots, b_{i-1}\}$. So the maximality of the sequence implies that the sequence cannot end in a vertex from A. Let $b_k \in B$ be the last vertex of this sequence. Because of the two properties (1) and (2)

$$P := b_k a_{f(k)} b_{f(k)} a_{f^2(k)} b_{f^2(k)} \ldots a_{f^r(k)}$$

with $f^r(k) = 0$ is an alternating path.

We claim that b_k has no partner in M. If a would be a partner of b_k and $a = a_i$ for an $i \in \{1, \ldots, k-1\}$, then $b_k = b_i$, since M is a matching, contradiction. If $a \neq a_i$ for all $i \in \{1, \ldots, k-1\}$, then $a_k := a$ would prolong our sequence, contradicting its maximality. Hence, b_k is without partner and P is an augmenting path. □

Exercises.

(6) The goal of this exercise is to show that the marriage theorem is false for infinite bipartite graphs. Let G be the graph with vertex set $\mathbb{Z}$ and edge set

$$\{\{a, -a\} \mid a \in \mathbb{Z} \setminus \{0\}\} \cup \{\{0, a\} \mid a \in \mathbb{N} \setminus \{0\}\}.$$

Show that the conditions in Hall's marriage theorem are satisfied for $A := \{a \in \mathbb{Z} \mid a \leq 0\}$, but that A has no matching in G.

(7) Let G be a finite bipartite graph with partition classes A and B, and let $A' \subseteq A$ and $B' \subseteq B$. Suppose that A' has a matching in G and B' has a matching in G. Prove that there exists a matching of $A' \cup B'$ in G. Does this statement hold if G is infinite? Does it hold if G is not bipartite?

DEFINITION 1.5.4. A graph G is called *k-regular* if every node in G has degree k.

The marriage theorem has the following elegant consequence.

COROLLARY 1.5.5. *For $k \geq 1$ any bipartite k-regular graph has a perfect matching.*

PROOF. Every subset $S \subseteq A$ has exactly $k|S|$ edges into $N(S)$. Together there are $k|N(S)|$ edges to vertices in $N(S)$. Therefore $k|S| \leq k|N(S)|$ and $|S| \leq |N(S)|$. The Marriage Theorem gives us a matching of A in G. In regular bipartite graphs we clearly have $|A| = |B|$. Hence we have found a perfect matching for G. □

Another consequence of the marriage theorem is the important theorem of Kőnig. (In fact, it is also possible to derive the marriage theorem from Kőnig's theorem; this is Exercise 8.) Let $G = (V, E)$ be a graph. A set $U \subseteq V$ is called a *covering* of G if every edge of G contains a vertex from U.

THEOREM 1.5.6 (Kőnig). *Let G be a finite bipartite graph. Then the maximal size of a matching of G equals the minimum size of a covering of G.*

PROOF. Let U be a covering of G of minimal size. Let M be a matching of G. We need at least $|M|$ nodes to cover M. Hence, $|U| \geq |M|$. We will prove that there exists a matching of size $|U|$ in G. Let $U_1 := U \cap A$ and $U_2 := U \cap B$. To verify the marriage condition for U_1 in the bipartite graph $G_1 := G[U_1 \cup B \setminus U_2]$, we have to show for an arbitrary $S \subseteq U_1$ that $|S| \leq |N(S)|$. Otherwise, we could have replaced the set S in U by the smaller set $N(S)$, and would still have a cover of G, contradicting the minimality of U. The Marriage Theorem (Theorem 1.5.3) gives us a matching M_1 of U_1 in G_1. Analogously, we obtain a matching M_2 of U_2 in the graph $G_2 := G[U_2 \cup A \setminus U_1]$. Then $M_1 \cup M_2$ is a matching in G, and $|M_1 \cup M_2| = |U_1| + |U_2| = |U|$. □

Another useful combinatorial presentation of the marriage theorem is as follows. Let $\mathcal{F}$ be a finite family of finite subsets of a finite set X where the members of $\mathcal{F}$ are counted with multiplicity (i.e., the same set might appear several times in $\mathcal{F}$). A *transversal* (or *system of distinct representatives*) *for* $\mathcal{F}$ is the image of an injective function f from $\mathcal{F}$ to X such that $f(S) \in S$ for every $S \in \mathcal{F}$. In other words, f selects one representative from each set in $\mathcal{F}$ in such a way that no two sets from S get the same representative. We are interested in characterising the situation when $\mathcal{F}$ has such a transversal.

The collection $\mathcal{F}$ satisfies the *marriage condition* if for each subfamily $\mathcal{S} \subseteq \mathcal{F}$

$$|\mathcal{S}| \leq |\bigcup_{A \in \mathcal{S}} A|.$$

Clearly, if the marriage condition fails then there cannot be a transversal f of $\mathcal{F}$. The following is basically a reformulation of Hall's marriage theorem.

THEOREM 1.5.7. *Let $\mathcal{F}$ be a family of finite subsets of a set X. Then $\mathcal{F}$ has a transversal if and only if $\mathcal{F}$ satisfies the marriage condition.*

PROOF. Let $\mathcal{F} = \{A_1, A_2, \ldots, A_n\}$. Let G be the bipartite graph with colour classes $\mathcal{F}$ and X and with and edge between $A_i \in \mathcal{F}$ and $y \in X$ if $y \in A_i$. The marriage condition for $\mathcal{F}$ implies that for any $\mathcal{S} \subseteq \mathcal{F}$ we have $|N(\mathcal{S})| \geq |\mathcal{S}|$, so Hall's marriage theorem for bipartite graphs implies that $\mathcal{F}$ has a a matching of $\mathcal{F}$ in G. The matching is the graph of a transversal of $\mathcal{F}$, proving the statement. □

Exercises.

(8) Deduce Theorem 1.5.3 (Hall's marriage theorem finite bipartite graphs) from Theorem 1.5.7.

(9) A *Latin square of order* n is matrix $A \in \{1, \dots, n\}^{n \times n}$ such that each symbol appears exactly once in each row and each column. If $A \in \{1, \dots, n\}^{m \times n}$ for $m < n$ is such that each symbol appears at most once in every row and every column then A is called a *Latin rectangle*. Use Hall's marriage theorem to prove that any $m \times n$ Latin rectangle can be extended to an $n \times n$ Latin square.

(10) A *partially ordered set* is a pair $(P, \leq)$ where P is a set and $\leq$ is a binary relation on P that satisfies for all $p, q, r \in P$:
   (a) $p \leq p$,
   (b) if $p \leq q$ and $q \leq p$, then $p = q$, and
   (c) if $p \leq q$ and $q \leq r$, then $p \leq r$.

   Two elements p, q of P are called *comparable* if $p \leq q$ or $q \leq p$, and *incomparable* otherwise. A subset S of P is called
   - a *chain* if all pairs of elements of S are comparable.
   - an *antichain* if distinct elements of S are incomparable.

   Use Kőnig's theorem to prove Dilworth's theorem: if $(P, \leq)$ is a finite partial order, then the size of the largest antichain of P equals the minimal k such that there are chains $C_1, \dots, C_k$ in P that *cover* P, i.e., $P = C_1 \cup \dots \cup C_k$.

(11) (from [**12**]) Find a partially ordered set that has no infinite antichain but cannot be covered by finitely many chains.

(12) Consider the following two-person game played on an undirected graph $G = (V, E)$. There are two players, A and B, that play alternatingly. At step 0 the first player chooses an arbitrary edge. Then each player in turn chooses a previously unchosen edge such that the set of chosen edges forms a simple path. The first player who is unable to make a legal move looses. Prove: if G has a perfect matching, then the first player has a winning strategy.

(13) Consider the following variant of the game of the previous exercise. This time, the players alternatingly pick a vertex. At the first move, the choice can be arbitrary. After that, the picked vertex must always be adjacent to the previously picked vertex. Prove that the *second* player has a winning strategy if and only if every connected component of the graph has a perfect matching.

(14) ($*$) Show that every 3-regular graph $(V; E)$ contains a subgraph $(V; E')$ all of whose vertices have degree 1 or 2.

(15) (from [**12**]) Let k be an integer. Show that any two partitions of a finite set into k-sets admit a common choice of representatives.

CHAPTER 2

Duality

Proposition 1.3.1 is an example of a *duality*: for every graph G, either it is 2-colourable, which is easy to verify once we are given the 2-colouring $f\colon V(G) \to \{0,1\}$, or it contains an odd cycle, which is easy to verify, too. Even Lemma 1.2.1 can be viewed as a (baby-) duality: either a graph has a decomposition as a disjoint union of non-trivial smaller graphs, or for every pair of vertices $u, v \in V(G)$ there exists a path from u to v in G. The marriage theorem of Hall and the theorem of Kőnig are further examples of dualities.

One way to formalise the similarities in these statements has been proposed by Jack Edmonds. A class of finite graphs $\mathcal{C}$ (or a class of more general mathematical structures) is said to have a *good characterisation* if

- $\mathcal{C}$ is in the complexity class NP (i.e., there exists a polynomial-time nondeterministic Turing machine that decides membership in $\mathcal{C}$, and
- the complement of $\mathcal{C}$ is in the complexity class NP, too, i.e., $\mathcal{C}$ is in the complexity class coNP.

Typically, if a class is in NP $\cap$ coNP, it is also in the complexity class P, i.e., it can be solved in polynomial time. Hence, a good characterisation for $\mathcal{C}$ can be taken as an indication that an efficient algorithm for membership in $\mathcal{C}$ exists.

In this section we first revisit a well-known duality from linear algebra. We then find a common generalisation of Kőnig's theorem for matchings and linear algebra duality, namely linear programming duality. In Section 2.4 we will see many applications of this principle with many algorithmic consequences.

2.1. Duality in Linear Algebra

Let $A \in \mathbb{Q}^{m\times n}$ and $b \in \mathbb{Q}^m$. If $Ax = b$ has a solution, then this can be shown by simply presenting a solution from $\mathbb{Q}^n$ for the vector of unknowns. There also simple proofs for *unsatisfiability* of $Ax = b$ (even simpler than performing Gaussian elimination):

THEOREM 2.1.1 (Duality). *Let $A \in \mathbb{Q}^{m\times n}$, $x = (x_1, \dots, x_n)$ an n-tuple of variables, and $b \in \mathbb{Q}^m$. Then $Ax = b$ is unsatisfiable if and only if the system*

$$(A|b)^\top y = (0, \dots, 0, 1)^\top \tag{1}$$

is satisfiable.

PROOF. Let $z_1, \dots, z_m$ be the rows of $A|b$. The back direction is the easier direction. Suppose that $(A|b)^\top y = (0, \dots, 0, 1)^\top$ has a solution $y \in \mathbb{Q}^m$. Then $y_1 z_1 + \cdots + y_m z_m = (0, \dots, 0, 1)$, i.e., $(0, \dots, 0, 1) \in \langle z_1, \dots, z_m\rangle$. So we can derive the row $(0, \dots, 0, 1)$ from $A|b$ by elementary row transformations. This means that $Ax = b$ implies $0x_1 + \cdots + 0x_n = 1$, which is unsatisfiable. Hence, $Ax = b$ must be unsatisfiable.

The direction $\Rightarrow$ is also easy to show using the row echelon form. Use row transformations to bring the matrix $(A|b)$ into row echelon form $(C|d)$. If $Ax = b$ is unsatisfiable, then $Cx = d$ is unsatisfiable, and $r := \operatorname{rank}(C) < \operatorname{rank}(C|d)$ by

well-known linear algebra. In particular $z_{r+1} = (0, \ldots, 0, d_{r+1})$ with $d_{r+1} \in \mathbb{Q} \setminus \{0\}$. Replace z_{r+1} by $d_{r+1}^{-1} z_{r+1} = (0, \ldots, 0, 1)$. Since this row has been derived from $(A|b)$ by row transformations, we have $(0, \ldots, 0, 1) \in \langle z_1, \ldots, z_m \rangle$. Hence, the system $(A|b)^\top y = (0, \ldots, 0, 1)^\top$ has a solution. □

2.2. Weighted Matchings

We would like to find a common generalisation of dualities for matchings and the duality of linear algebra that we have just seen in the previous section. In our first step from matchings towards algebra we consider in this section the *weighted matching problem*. In this problem, we are given a bipartite graph G with partition classes A and B where each edge $e \in E(G)$ is decorated by a weight $w_e \in \mathbb{Q}$. We are interested in finding a matching M of A in G whose *weight*

$$w(M) := \sum_{e \in M} w_e$$

is maximal. This problem has numerous applications.

2.2.1. Maximum matching as an integer linear program. We will reformulate the weighted matching problem for G as a linear optimisation problem over a system of linear inequalities. We introduce a variable x_e for each edge $e \in E(G)$; the variable x_e can attain values 0 or 1. They encode the desired matching M, where $x_e = 1$ means $e \in M$ and $x_e = 0$ means $e \notin M$. Then $\sum_{e \in M} w_e$ can be written as

$$w(x) := \sum_{e \in E(G)} w_e x_e$$

where x is a tuple listing all the variables; $w(x)$ will be called the *objective function*. The requirement that a vertex $v \in V$ appears in at most one edge from M can be expressed by

$$\sum_{e \in E(G), v \in e} x_e \leq 1.$$

More generally, a set of linear inequalities over the integers together with a linear objective function is called an *Integer Linear Program (ILP)*. Unfortunately, there is in general no efficient algorithm known that decides whether a given ILP has a feasible solution (this problem is another example of an *NP-hard problem*; see Appendix B). However, if the ILP comes from a weighted matching problem as described above, we are lucky and can always find optimal *integer* solutions, as we will see in the next subsection.

2.2.2. A relaxation. If we leave out the integrality conditions, i.e., if we allow each variable x_e to attain all values in the interval $[0, 1]$, we obtain the following:

$$\begin{aligned} &\text{Maximize} \sum_{e \in E} w_e x_e \\ &\text{subject to} \sum_{e \in E, v \in e} x_e \leq 1 \text{ for each } v \in V \text{ and} \\ &\qquad 0 \leq x_e \leq 1 \text{ for each } e \in E. \end{aligned} \tag{2}$$

Such an optimisation problem is called a *linear program*; they can be solved efficiently, and are extremely important in optimisation and theoretical computer science. There is a notion of a *dual* of a linear program, and this notion of duality has many applications in combinatorics, as we will see in later sections. For more background on

linear programming, we recommend [**28**], which we have freely used to prepare the following subsections.

Linear programs that are obtained from integer linear programs as described above are called *LP relaxations*. A linear program might not have any solution at all (if the set of linear inequalities is unsatisfiable); if this happens for an LP relaxation, then the original integer linear problem does not have a solution as well. For example, this might happen if we consider the integer linear program for a bipartite graph which does not have a perfect matching.

Let us now assume that the LP relaxation has an optimal solution $s \in \mathbb{Q}^n$, i.e., a solution where the value of the objective function is largest possible. Certainly s provides an *upper bound* for the objective function of the original integer program. This is because every feasible solution of the integer program is also a feasible solution of the LP relaxation. In the case that we started with the integer program for the weighted matching problem is that this upper bound is also tight: it provides an optimal solution of the original problem!

THEOREM 2.2.1. *Let $G = (V, E)$ be a finite bipartite graph with rational edge weights w_e. Then the LP relaxation (2) has an integral optimal solution. This solution is an optimal solution for the integer program as well, and hence provides a maximum weight matching of G.*

PROOF. Clearly, the LP relaxation has a feasible solution since we may set all variables to 0. Let s be an optimal solution of the LP relaxation, and let $w(s) = \sum_{e \in E} w_e s_e$ be the value of the objective function at s. Let

$$E' := \{e \in E \mid 0 < s(e) < 1\}.$$

We show the statement by induction on E'. If $|E'| = 0$ there is nothing to be shown. Otherwise, consider the case that E' contains a cycle $C = (c_0, c_1, \ldots, c_{l-1})$ (which must have even length). Let a be the minimum of s_e over all edges e on C, and let b be the maximum of s_e over all edges e on C. Define ϵ as the minimum of a and $1 - b$, and define s'_e for $e \in E$ as follows:

$$s'_e := \begin{cases} s_{\{c_i, c_{i+1}\}} - \epsilon & \text{for } e = \{c_i, c_{i+1}\} \text{ and } i \text{ even} \\ s_{\{c_i, c_{i+1}\}} + \epsilon & \text{for } e = \{c_i, c_{i+1}\} \text{ and } i \text{ odd} \\ s_e & \text{otherwise.} \end{cases}$$

Then s' satisfies for every $v \in V$ the condition $\sum_{e \in E, v \in e} s'_e = 1$. The objective function evaluated at s' is

$$w(s') = \sum_{e \in E} w_e s'_e = w(s) + \epsilon \sum_{i=0}^{l-1} (-1)^i w_{\{c_i, c_{i+1}\}}.$$

Note that s' is still feasible, and that s'_e is integral for at least one edge on C. Since s is optimal, we must have $\Delta := \sum_{i=0}^{l-1} (-1)^i w_{\{c_i, c_{i+1}\}} = 0$ since otherwise we could achieve $w(s') > w(s)$ by choosing $\epsilon > 0$ for $\Delta > 0$ and by choosing $\epsilon < 0$ for $\Delta < 0$. This means that s' is another feasible optimal solution with strictly more integer values than s. Similarly, if E' is acylic, we can increase the number of integer values by picking a maximal path in (V, E') and proceeding similarly. □

Exercises.

(16) Show that Theorem 2.2.1 is false for general finite (not necessarily bipartite) graphs.

2.3. The Duality Theorem

This section presents an extremely powerful and useful duality result that implies many of the dualities that we have seen previously, and many more results in combinatorics.

2.3.1. Example first. Let us start with a concrete example. Consider the linear program (see Figure 2.1)

$$\begin{aligned} \text{maximize } & x_1 + x_2 \\ \text{subject to } & 3x_1 - x_2 \leq 0 \\ & -x_1 + x_2 \leq 4 \\ & 2x_1 + 4x_2 \leq 40 \\ & x_1, x_2 \geq 0 \end{aligned} \tag{3}$$

Since $x_1, x_2 \geq 0$ we obtain that

$$x_1 + x_2 \leq 2x_1 + 4x_2 \leq 40$$

so the optimum is bounded by 40. We can obtain a better bound by first dividing the third inequality by two:

$$x_1 + x_2 \leq x_1 + 2x_2 \leq 20.$$

We can do even better by adding the second and two times the third inequality:

$$x_1 + x_2 = (3x_1 - x_2) + (-2x_1 + 2x_2) \leq 8.$$

More generally, from the constraints we are trying to derive an inequality of the form $c_1x_1 + c_2x_2 \leq h$ where $c_1, c_2 \geq 0$ and h is as small as possible. We derive inequalities by choosing nonnegative coefficients y_1, y_2, y_3, obtaining

$$y_1(3x_1 - x_2) + y_2(-x_1 + x_2) + y_3(2x_1 + 4x_2) \leq y_2 4 + y_3 40$$

which can be rewritten to

$$(3y_1 - y_2 + 2y_3)x_1 + (-y_1 + y_2 + 4y_3)x_2 \leq 4y_2 + 40y_3$$

and thus $c_1 = 3y_1 - y_2 + 2y_3$, $c_2 = -y_1 + y_2 + 4y_3$, and $h = 4y_2 + 40y_3$. Finding such y_1, y_2, y_3 is again a linear program, namely

$$\begin{aligned} \text{minimize } & 4y_2 + 40y_3 \\ \text{subject to } & 3y_1 - y_2 + 2y_3 \geq 1 \\ & -y_1 + y_2 + 4y_3 \geq 1 \\ & y_1, y_2, y_3 \geq 0 \end{aligned} \tag{4}$$

Clearly, the optimum of the new linear program, which is called the *dual linear program*, provides an upper bound for the optimum of the original linear program. Note that in the dual LP, we have one variable for each constraint of the original LP, and one constraint for each variable of the original LP.

In fact, the upper bound from the dual is tight! This can be seen from the observation that the maximum of the original linear program is at least 8, because 4 is attained for $x_1 = 2$ and $x_2 = 6$. The minimum of the dual is at most 8, because 8 is attained for $y_1 = 1$, $y_2 = 2$, and $y_3 = 0$. So the maximum of the original linear program equals the minimum of the dual linear program.

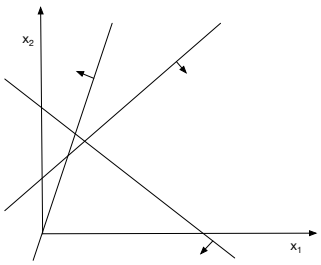

FIGURE 2.1. An illustration of the linear program 3.

2.3.2. The dual linear program in general. Let A be a matrix with m rows and n columns and entries from $\mathbb{Q}$ (the same results hold for $\mathbb{R}$ instead of $\mathbb{Q}$). Consider the linear program

$$\text{maximize } c^\top x \text{ subject to } Ax \leq b \text{ and } x \geq 0 \qquad \text{(P)}$$

which we call the *primal linear program* in the following. Similarly as in Section 2.1, we are trying to combine the m inequalities of the system $Ax \leq b$ with some nonnegative coefficients $y_1, \ldots, y_m$ so that

- the resulting inequality has the j-th coefficient at least c_j, and
- the right-hand side is as small as possible.

This leads to the *dual linear program*

$$\text{minimize } b^\top y \text{ subject to } A^\top y \geq c \text{ and } y \geq 0. \qquad \text{(D)}$$

The following is clear from the way we constructed the dual linear program:

PROPOSITION 2.3.1. *For each feasible solution t of the dual linear program* (D) *the value $b^\top t$ provides an upper bound on the maximum of the objective function of the primal* (P).

Note that this implies that if (P) is unbounded (from below), then (D) must be infeasible, and if (D) is unbounded (from above) then (P) is infeasible. The following, on the other hand, requires proof.

THEOREM 2.3.2 (LP duality). *Exactly one of the following possibilities occurs.*

(1) Neither (P) *nor* (D) *has a solution.*
(2) (P) *is unbounded and* (D) *has no solution.*
(3) (P) *has no solution and* (D) *is unbounded.*
(4) Both (P) *and* (D) *have a solution, and if s is an optimal solution to* (P) *and t is an optimal solution to* (D) *then*

$$c^\top s = b^\top t.$$

EXAMPLE 2. The problem of finding small vertex covers can also be formulated as an integer linear program, namely as

$$\begin{aligned} &\text{minimize } \sum_{v \in V} y_v \\ &\text{subject to } y_v \geq 0 && \text{for all } v \in V \\ &\qquad \sum_{v \in e} y_v \geq 1 && \text{for all } e \in E. \end{aligned} \qquad (5)$$

Note that the linear program for matchings that we have already seen for the weighted case

$$\begin{aligned} \text{maximize } & \sum_{e \in E} x_e & \\ \text{subject to } & x_e \geq 0 & \text{for all } e \in E \\ & \sum_{e \in E, v \in e} x_e \leq 1 & \text{for all } v \in V \end{aligned}$$

is precisely the dual of the linear program for the vertex cover problem! Moreover, similarly as for matchings it turns out that for bipartite graphs the LP relaxation (5) provides the optimum also for the integer linear program. To see this, let t be an optimal solution to the LP relaxation (5). If for some $v \in V$ we have $t_v > 1$ then we may set t_v to 1 while fixing all other values of t, which still meets the boundary conditions but reduces the objective function, contrary to the optimality assumption on t. Let V' be the set of vertices v such that t_v is strictly between 0 and 1. We show the statement by induction on $|V'|$. If $|V'| = 0$ then we are done. Define

$$\epsilon := \min\{t_v, 1 - t_v \mid v \in V'\}$$

and note that $\epsilon > 0$ by our assumptions. Let V_1, V_2 be the color classes of the bipartite graph G such that $|V_1| \leq |V_2|$. Define t' as follows:

$$t'_v := \begin{cases} t_v + \epsilon & \text{if } v \in V_1 \cap V', \\ t_v - \epsilon & \text{if } v \in V_2 \cap V', \\ t_v & \text{otherwise.} \end{cases}$$

Then t' is again a feasible solution to (5):

- $t'_v \geq 0$ holds for all $v \in V$ by the choice of ϵ;
- $\sum_{v \in e} t'_v \geq 1$ holds for all $e \in E$ because if $u = \{u, v\}$ and $t_v = 1$ then $t'_v = 1$ and the statement holds, and if $t_v = 0$ then $t_u = 1 = t'_u$ and the statement holds, so we may suppose that $u \in V_1 \cap V'$ and $v \in V_2 \cap V'$, or $v \in V_1 \cap V'$ and $u \in V_2 \cap V'$; in both cases, ϵ cancels out and hence $t'_u + t'_v \geq 1$.

Moreover, $\sum_{v \in V} t'_v \leq \sum_{v \in V} t_v$ by our choice of t' because $|V_1| \leq |V_2|$. Finally, there is at least one $v \in V$ where t_v is fractional and t'_v is integral, so the statement follows from the inductive assumption.

Therefore, linear programming duality implies Kőnig's theorem! △

Exercises.

(17) Suppose that $G = (V, E)$ is a bipartite graph with a rational edge weight w_e for every $e \in E$. Write down a linear program for the maximum weight *perfect* matching in G. Does the LP always have a feasible solution? Is it true that if it has a feasible solution, then there is also an integer solution? What is the relation between the value of the LP and the size of the maximal matching?

(18) A serious dietician wishes to design a minimal-cost diet to meet some minimum daily requirements, and draws the following table.

	Potatoes	Oats	Requirements per dish
Vitamin C [mg/kg]	150	10	15 mg
Fibre [g/kg]	20	100	4 mg
Protein [g/kg]	20	120	20 mg
Cost [Euro/kg]	1	3	

Formulate the problem of finding a minimum-cost diet as a linear program. Find a good lower bound to the minimum cost. What is the meaning of optimal solutions to the dual problem?

(19) (Learning linear classifiers from data.) Suppose we are given a finite set of black points $B \subseteq \mathbb{Q}^n$ and a finite set of while points $W \subseteq \mathbb{Q}^n$ (the *learning data*), and we would like to know whether there exists a linear halfspace of $\mathbb{Q}^n$ (the *classifier*) that contains all the while points from W but no black point from B. Show that this task can be modelled by linear programming.

(20) You want to install a circular irrigation system on your piece of land. Your land has the shape of a convex n-gon. What is the maximum radius that you can choose for your irrigation system? Model this task as a linear program.

(21) Consider the following 2-player game played on a directed graph (V, E). We assume that V is finite and partitioned in two subsets V_1 and V_2, and $E \subseteq V_1 \times V_2 \cup V_2 \times V_1$. Player i, for $i \in \{1, 2\}$, chooses a probability distribution p_i on V_i.

2.3.3. Optimality via feasibility. In this section we use the duality theorem to reduce the task of deciding whether a given linear program (P) has an *optimal* solution to the task of deciding whether a set of linear inequalities has *any* solution. We simply combine the constraints from (P), the constraints from (D), and add an inequality between the objective functions, obtaining the following system of linear inequalities:

$$\begin{aligned} Ax &\leq b \\ A^\top y &\geq c \\ c^\top x &\geq b^\top y \\ x, y &\geq 0 \end{aligned}$$

For each feasible solution (s, t) for the variables (x, y) of this system, s is an optimal solution of the linear program (P).

The satisfiability problem for systems of linear equalities (see Section 2.3.3) can be solved in polynomial time. The historically first algorithm for this problem is due to Khachiyan [**25**]. Another method is the simplex algorithm, which works quite well in practise, but has an exponential running time in general. We refer to [**19**] and [**38**] for a more detailed treatment of the subject.

Exercises.

(22) A function $f\colon \mathbb{Q}^n \to \mathbb{Q}$ is called *convex* if for all $x, y \in \mathbb{Q}^n$ and $\alpha \in [0,1]$ we have that $f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y)$. Show that functions of the form $x \mapsto \max(f_1(x), \ldots, f_k(x))$, for linear functions $f_1, \ldots, f_k$, are convex.

(23) Prove that for given linear functions $f_1, \ldots, f_k$ there is a polynomial-time algorithm for deciding whether

$$\inf_{x \in \mathbb{Q}^n} (\max(f_1(x), \ldots, f_k(x))$$

equals $-\infty$. Also show that if $\inf(\max(f_1(x), \ldots, f_k(x))$ is finite, then we can compute in polynomial time a $x_0 \in \mathbb{Q}^n$ such that

$$\inf_{x \in \mathbb{Q}^n} (\max(f_1(x), \ldots, f_k(x)) = \max(f_1(x_0), \ldots, f_k(x_0)).$$

(24) The computational complexity of deciding whether two given finite graphs G and H are isomorphic is a famous open problem in theoretical computer science. Let G and H be two finite graphs with $V := V(G) = V(H)$ and adjacency matrices A and B, respectively. Then G and H are called *fractionally isomorphic* if there exists a *doubly stochastic matrix* D such that $AD = DB$. A quadratic matrix $D \in \mathbb{R}^{n \times n}$ is called *doubly stochastic* if all entries are non-negative and the sum of the entries in each row and column is equal to 1.
- Show that fractional isomorphism defines an equivalence relation on finite graphs.
- Show that two graphs that are isomorphic are also fractionally isomorphic.
- (*) Find an example that shows that the converse of the previous statement is false.
- Show that fractional isomorphism can be decided in polynomial time.

2.3.4. Fourier-Motzkin elimination. Our proof of LP duality uses the lemma of Farkas, which in turn we prove using Fourier-Motzkin elimination. Fourier-Motzkin elimination is a systematic procedure for eliminating variables from systems of linear inequalities. If we eliminate all variables from the system, we might end up with an inequality of the form $a \leq b$ for a value a which is strictly larger than b, a contradiction. In this case we know that the original system was infeasible, and otherwise the original system was feasible. Geometrically speaking, we compute in each step a system that describes the projection of the solution space of the system to a subset of the variables. The way this is done is very similar in spirit to Gaussian elimination that we already used in Section 2.1. We first look at an example:

$$\begin{aligned} x - y &\leq 0 \\ -x - 3y &\leq -6 \\ y + x &\leq 2 \\ -x + 3y &\leq 0 \end{aligned}$$

To eliminate the variable y, we collect lower bounds on y in terms of x, and upper bounds on y in terms of x, so we rewrite the equation system into:

$$y \geq x \tag{6}$$
$$y \geq 2 - \frac{1}{3}x \tag{7}$$
$$y \leq 2 - x \tag{8}$$
$$y \leq \frac{1}{3}x \tag{9}$$

Note that each upper bound must be larger than each lower bound, so the system implies that

$$\begin{aligned} x &\leq 2 - x && \text{(combining (6) and (8))} \\ x &\leq \frac{1}{3}x && \text{(combining (6) and (9))} \\ 2 - \frac{1}{3}x &\leq 2 - x && \text{(combining (7) and (8))} \\ 2 - \frac{1}{3}x &\leq \frac{1}{3}x && \text{(combining (7) and (9))} \end{aligned}$$

Rewriting, we obtain

$$\begin{aligned} x &\leq 1 \\ x &\leq 0 \\ x &\leq 0 \\ x &\geq 3 \end{aligned}$$

Again combining lower with upper bounds, we obtain a contradiction, so the original system was unsatisfiable. Motzkin's theorem states that if we cannot derive a contradiction by this procedure, then the original system was satisfiable. This simply follows from the observation that at each step of the procedure, any solution to the new system can be extended to a solution to the old system (by picking any value that lies between the lower and the upper bounds).

The Fourier-Motzkin procedure is not very efficient: in each step, the number of inequalities can grow quadratically, which might lead to an exponential growth in general. However, our current goal is theoretical: we want to prove the duality theorem, and do not care about efficiency. Because of the simplicity of the procedure, it is a very good starting point for proving LP duality.

We mention that the results in this section have straightforward generalisations to the situation where some of the inequalities are strict.

2.3.5. The Farkas lemma. We use Fourier-Motzkin Elimination to prove the following important lemma, the *lemma of Farkas*. There are numerous variants of this lemma; we give two. They are easily seen to be equivalent. The first variant is the more natural one to be proven using Fourier-Motzkin elimination. The second variant is the one needed in our proof of the LP duality theorem.

LEMMA 2.3.3 (Lemma of Farkas in two variants). *Let $A \in \mathbb{Q}^{m\times n}$ and let $b \in \mathbb{Q}^m$.*

(1) The system $Ax \leq b$ has a solution if and only if every nonnegative $y \in \mathbb{Q}^m$ with $y^\top A = 0^\top$ also satisfies $y^\top b \geq 0$.
(2) The system $Ax \leq b$ has a nonnegative solution if and only if every nonnegative $y \in \mathbb{Q}^m$ with $y^\top A \geq 0$ also satisfies $y^\top b \geq 0$.

PROOF. We prove variant (1) using Motzkin, and then derive (2) from (1).

First we prove the easy direction of variant (1). If $Ax \leq b$ has some solution $\tilde{x}$, and $y \geq 0$ satisfies $y^\top A = 0^\top$, we get $y^\top b \geq y^\top A\tilde{x} = 0^\top \tilde{x} = 0$. For the interesting direction of (1) we assume that $Ax \leq b$ has no solution. Our task is to construct a vector $y \geq 0$ satisfying $y^\top A = 0^\top$ and $y^\top b < 0$. We find such a witness of infeasibility by induction on the number of variables. In the base case the system $Ax \leq b$ has no variables, so it is of the form $0 \leq b$ with $b_i < 0$ for some $i \leq m$. Then $y = e_i$ (the i-th unit vector) clearly satisfies the requirements for y. If $x \leq b$ has at least one variable, we perform a step of the Fourier-Motzkin elimination. This yields an infeasible system $A'x' \leq b'$ with one variable less. So inductively we find an unfeasibility witness y' for it. Recall that all inequalities of $A'x' \leq b'$ are positive linear combinations of original inequalities; equivalently, there is an $m \times m$ matrix M with all entries nonnegative and $(0|A') = MA$, $b' = Mb$. We claim that $y = M^\top y'$ is a witness of infeasibility for the original system $Ax \leq b$. Indeed, we have $y^\top A = y'^\top MA = y'^\top (0|A') = 0^\top$ and $y^\top b = y'^\top Mb = y'^\top b' < 0$ since y' is a witness of infeasibility for $A'x' \leq b'$. The condition $y \geq 0$ follows from $y' \geq 0$ by the nonnegativity of M.

To prove that (1) implies (2) we have to find an equivalent condition for $Ax \leq b$ having a nonnegative solution. Let $\bar{A} = \binom{A}{-I_n}$, where I_n is the $n \times n$-unit matrix, and $\bar{b} = \binom{b}{0}$. Note that $Ax \leq b$ has a nonnegative solution if and only if $\bar{A}x \leq \bar{b}$ has any solution. The latter is equivalent, by (1), to the condition that all $\bar{y} \geq 0$ with $\bar{y}^\top \bar{A} = 0^\top$ satisfy $\bar{y}^\top \bar{b} \geq 0$. Writing $\bar{y} = \binom{y}{y'}$ where y is a vector with m components, we have

$$\bar{y} \geq 0 \text{ and } \bar{y}^\top \bar{A} = 0^\top \text{ if and only if } y \geq 0 \text{ and } y'^\top = y^\top A \geq 0^\top.$$

Moreover, $\bar{y}^\top \bar{b} = y^\top b$. Hence, $Ax \leq b$ has a nonnegative solution if and only if all $y \geq 0$ with $y^\top A \geq 0^\top$ satisfy $y^\top b \geq 0$, which is what we had to show. □

Exercises.

(25) Let $A \in \mathbb{Q}^{n\times p}$, $B \in \mathbb{Q}^{n\times q}$, $b \in \mathbb{Q}^p$, and $c \in \mathbb{Q}^q$. The following are equivalent (*Motzkin's transposition theorem* [**38**, Corollary 7.1k]):
- There exists $x \in \mathbb{Q}^n$ with $Ax < b$ and $Bx \leq c$.
- For every $y \in \mathbb{Q}^p$ and $z \in \mathbb{Q}^q$

$$\begin{aligned} &\text{if } yA + zB = 0 \text{ then } yb + zc \geq 0, \text{ and} \\ &\text{if } yA + zb = 0 \text{ and } y \neq 0 \text{ then } yb + zc > 0. \end{aligned}$$

Hint: first identify and prove the instructive easy direction.

(26) Prove the following (Lemma 6.1 in [**7**]). Let $n, s, t \in \mathbb{N}$, $(a_{i,j}) \in \mathbb{Q}^{n\times(s+t)}$, and $(b_j) \in \mathbb{Q}^{s+t}$. Then exactly one of the following holds.
- Either there exists $(x_i) \in \mathbb{Q}^n$ and $d \in \mathbb{Q}$ such that

$$\begin{aligned} x_i &\geq 0 && \text{for every } i \in \{1,\dots,n\} \\ \sum_{i=1}^{n} a_{i,j}x_i &\geq b_j + d && \text{for every } j \in \{1,\dots,s\} \\ \sum_{i=1}^{n} a_{i,j}x_i &= b_j + d && \text{for every } j \in \{s+1,\dots,s+t\}. \end{aligned}$$

- Or else there exists $(y_j) \in \mathbb{Q}^{r+s}$ such that

$$\sum_{j\in\{1,\dots,s+t\}} y_j = 0 \tag{10}$$

$$y_j \geq 0 \qquad \text{for every } j \in \{1,\dots,s\} \tag{11}$$

$$\sum_{j\in\{1,\dots,s+t\}} y_j a_{i,j} \leq 0 \qquad \text{for every } i \in \{1,\dots,n\} \tag{12}$$

$$\sum_{j\in\{1,\dots,s+t\}} y_j b_j > 0. \tag{13}$$

2.3.6. Proving the duality theorem.

Proof of Theorem 2.3.2. Let us assume that the linear program (P) has an optimal solution x^*. We show that the dual (D) has an optimal solution as well, and that the optimum values of both programs coincide. Let $\gamma = c^\top x^*$ be the optimum value of (P). Then we know that the system of inequalities

$$Ax \leq b \text{ and } c^\top x \geq \gamma \tag{14}$$

has a nonnegative solution, but for any $\epsilon > 0$, the system

$$Ax \leq b, c^\top x \geq \gamma + \epsilon \tag{15}$$

has no nonnegative solution. If we define an $((m+1)\times n)$-matrix $\hat{A}$ and a vector $\hat{b}_\epsilon \in \mathbb{Q}^{m+1}$ by

$$\hat{A} := \begin{pmatrix} A \\ -c^\top \end{pmatrix} \text{ and } \hat{b}_\epsilon := \begin{pmatrix} b \\ -\gamma - \epsilon \end{pmatrix}$$

then (14) is equivalent to $\hat{A}x \leq \hat{b}_0$ and (15) is equivalent to $\hat{A}x \leq \bar{b}_\epsilon$.

We now apply variant (2) of the Farkas lemma and conclude that there is a nonnegative vector $\hat{y} = (u, z) \in \mathbb{Q}^{m+1}$ such that $\hat{y}^\top \hat{A} \geq 0^\top$ but $\hat{y}^\top \hat{b}_\epsilon < 0$. These conditions boil down to

$$A^\top u \geq zc \text{ and } b^\top u < z(\gamma + \epsilon). \tag{16}$$

Applying the Farkas lemma for the case $\epsilon = 0$ we see that the very same vector $\hat{y}$ must satisfy $\hat{y}^\top \hat{b}_0 \geq 0$, and this is equivalent to

$$b^\top u \geq z\gamma.$$

It follows that $z > 0$, since $z = 0$ would contradict the strict inequality in (16). But then we may set $v := \frac{1}{z}u \geq 0$, and (16) gives

$$A^\top v \geq c \text{ and } b^\top v < \gamma + \epsilon.$$

In other words, v is a feasible solution of (D), with the value of the objective function smaller than $\gamma + \epsilon$.

We have already observed that every feasible solution of (D) has value of the objective function at least γ. Hence (D) is a feasible and bounded linear program, and so we know that it has an optimal value y^*. Its value $b^\top y^*$ is between γ and $\gamma + \epsilon$ for every $\epsilon > 0$, and thus it equals γ. □

Corollary 2.3.4 (Complementary Slackness). *Let x be a feasible solution to an LP and y a feasible solution to its dual. A necessary and sufficient condition for x*

and y to be both optimal is that for all $i \in \{1, \dots, n\}$ and $j \in \{1, \dots, m\}$

$$x_i > 0 \Rightarrow \sum_{k \in \{1,\dots,n\}} A_{i,k} y_k = b_i$$

$$\sum_{k \in \{1,\dots,n\}} A_{i,k} y_k < b_i \Rightarrow x_i = 0$$

$$y_j > 0 \Rightarrow \sum_{k \in \{1,\dots,m\}} A_{k,j} x_k = c_j$$

$$\sum_{k \in \{1,\dots,m\}} A_{k,j} x_k < c_j \Rightarrow y_j = 0$$

Exercises.

(27) Prove Corollary 2.3.4.

2.3.7. The dualization recipe. The linear program (P) in the duality theorem, Theorem 2.1.1, has a very particular shape: all variables must be nonnegative, and we only have inequality conditions, no equalities. We can easily transform a general linear program into a program of this shape: for each variable x, we introduce two new x^+ and x^-, add the constraints $x^+ \geq 0$ and $x^- \geq 0$, and we substitute x by $x^+ - x^-$ everywhere. Moreover, an equality $a^\top x = b$ can be rewritten as a conjunction of two inequalities, $a^\top x \leq b$ and $a^\top x \geq b$. Finally, inequalities of the form $a^\top x \geq b$ can be turned by using $-a^\top x \leq -b$ instead. So we can transform every linear program into one that has the shape as in the duality theorem, and then apply the duality theorem. However, it is always possible to read off the dual LP directly from the original LP, without doing the transformation. For this, we use the following recipe (details can be worked by the readers themselves):

	Primal	Dual
Variables	$x_1, \dots, x_n$	$y_1, \dots, y_m$
Matrix	A	$A^\top$
Right-hand side	b	c
Objective function	$\max c^\top x$	$\min b^\top y$
Constraints	i-th constraint $\leq$	$y_i \geq 0$
	i-th constraint $\geq$	$y_i \leq 0$
	i-th constraint $=$	$y_i \in \mathbb{Q}$
	$x_j \geq 0$	j-th constraint $\geq$
	$x_j \leq 0$	j-th constraint $\leq$
	$x_j \in \mathbb{Q}$	j-th constraint $=$

A minimisation problem can be turned into a maximisation problem by changing the sign of the objective function. Hence, we can compute the dual of the dual if the dual is phrased as a maximisation problem. It is then an easy observation that the dual of the dual equals the primal!

Exercises.

(28) Let $(a_{i,j,k})_{i,j,k \in \{1,\dots,n\}}$ be rational numbers and $c \in \mathbb{Q}^n$, $b \in \mathbb{Q}^{n \times n}$. Compute the dual of the linear program

$$\min c^\top x \text{ subject to } \sum_{k \in \{1,\dots,n\}} a_{i,j,k} x_k \leq b_{i,j} \text{ for all } i, j \in \{1, \dots, n\}.$$

(29) Let $(a_{i,j,k})_{i,j,k \in \{1,\dots,n\}}$ be rational numbers and $c \in \mathbb{Q}^n$, $b \in \mathbb{Q}^{n \times n}$. Compute the dual of the linear program (in comparison to the previous exercise just

one index changed)

$$\min c^\top x \text{ subject to } \sum_{k\in\{1,\dots,n\}} a_{i,j,k}x_i \le b_{i,j} \text{ for all } i,j \in \{1,\dots,n\}.$$

2.4. Applications

Linear programming has many applications in combinatorics and the theory and practise of computations; besides matchings, we have already seen applications in economy (Exercise 18), machine learning (Exercise 19), and discrete geometry (Exercise 20). in this section we present applications for flows in networks, Markov decision processes, zero-sum games, and stochastic games. Further applications will be covered in the exercises.

2.4.1. Flows in networks. A *network* (V, E, s, t, c) consists of

- a set of nodes V;
- a set of *directed edges* $E \subseteq V^2$;
- a *source* $s \in V$ (there are no incoming edges $(u, s) \in E$);
- a *sink* $t \in V$ (there are no outgoing edges $(t, u) \in E$);
- a non-negative *capacity function*

$$c\colon E \to \mathbb{Q}_{\ge 0}\ .$$

Definition 2.4.1 (Flows). A *flow* in a network (V, E, s, t, c) is a non-negative function

$$f\colon E \to \mathbb{R}_{\ge 0}$$

such that for every $v \in V \setminus \{s, t\}$:

$$\sum_{(u,v)\in E} f(u, v) = \sum_{(v,u)\in E} f(v, u) \qquad \text{('What flows in needs to flow out')}$$

A flow is *admissible* if $f(u, v) \le c(u, v)$ for every $(u, v) \in E$. If f is a flow and $U \subseteq V \setminus \{s, t\}$, then summing over the elements in U yields

$$\sum_{(u,v)\in E, u\notin U, v\in U} f(u, v) = \sum_{(v,u)\in E, v\in U, u\notin U} f(v, u) \qquad \textit{('Flow preservation').}$$

Choosing $U = V \setminus \{s, t\}$ we obtain in particular that what leaves the source is what enters the sink:

$$\sum_{(s,u)\in E} f(s, u) = \sum_{(u,t)\in E} f(u, t)\ .$$

This amount is written $||f||$ and called the *strength* of the flow f.

Definition 2.4.2. A *cut* in a network (V, E, s, t, c) is a set $S \subseteq E$ such that the directed graph $(V, E \setminus S)$ does not have a directed walk from s to t.

In other words, S is a cut if every walk from the source to the sink contains at least one edge from S. The *capacity* of a cut S is defined as

$$c(S) := \sum_{e\in S} c(e)\ .$$

Lemma 2.4.3. *Let S be a cut in a finite network (V, E, s, t, w) and f an admissible flow, then $||f|| \le c(S)$.*

Theorem 2.4.4 (Ford and Fulkerson; Max Flow = Min Cut). *Let (V, E, s, t, w) be a finite network. Then*

$$\max_{f \text{ admissible flow}} ||f|| = \min_{S \text{ cut}} c(S)$$

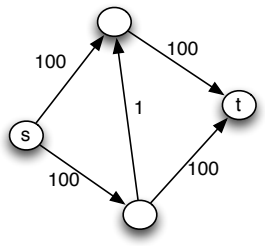

FIGURE 2.2. A small network.

In words: the strength of the biggest admissible flow equals the capacity of the smallest cut.

Exercises.

(30) Write the max-flow problem as an ILP.
(31) Write the min-cut problem as an ILP.
(32) Show that the LP for the max-flow problem is the dual of the LP relaxation of an LP for the min-cut problem.

The translation of the max-flow problem into linear programming is very robust in the sense that it can be adapted to also capture generalisations of the flow problem, for instance the generalisation where each edge $e \in E$ does not only have a capacity $c(e)$, but also a *payoff* $p(e)$; the payoff of the flow is then defined to be $\sum_{e \in E} p(e)f(e)$. We want to find a flow with maximum payoff (rather than a flow with maximum strength $\sum_{e \in E, s \in e} f(e)$). The LP program for the flow problem can be easily adapted to this problem.

2.4.2. The easychair problem. Suppose you are the chair of a scientific conference with peer reviewed submissions of papers; you are leading a program committee whose task is to select 80 papers from the submissions that will be admitted for presentation at the conference. Suppose that 400 papers have been submitted. Each paper will be assigned to at least 3 program committee (PC) members for peer review. Your program committee consists of 60 experts, so that each expert has to write 20 reports (all these numbers are quite realistic). The PC members can select for each paper one of the following responses.

(0) No: I don't want to review this paper (I don't feel qualified).
(1) Maybe: I might review this paper.
(2) Yes: I would like to review this paper.

(This is how things actually happen e.g. within the easychair system). We would like to find an assignment of at least 3 PC members to each paper such that the the number of papers that are assigned to a PC member who voted 'Maybe' or 'Yes' for that paper is maximised. If there are several optimal solutions, the number of papers assigned to a PC member who voted 'Yes' should be maximised among all these optimal solutions.

To turn this into a flow problem we create the following network N where each edge has a payoff value as described at the end of the previous section (taken from `http://corner.mimuw.edu.pl/?p=811`). Besides the source s and the sink t we have a node for each paper and for each PC member. The edges in E are defined as follows.

- The source s is connected with each PC member with an edge of capacity 20 and payoff 0;
- Each PC member is connected with each paper with an edge of capacity 1 and payoff 1 for 'No', payoff 10000 for 'Maybe', and payoff 10001 for 'Yes'.
- Each paper is connected with the sink t with an edge of capacity 3 and payoff 0.

We are interested in a flow with maximum payoff (again, see the remarks at the end of the previous section for the variant of the maximum flow problem with payoffs). An integral flow of size 1 from a PC member to a paper means that the PC member has to write a report for the paper. The payoffs are chosen so that the maximum flow assigns as many papers as possible to PC members who chose 'Yes' or 'Maybe' for that paper. If there are several flows that are equally good with respect to this condition, it prefers flows that have more papers assigned to PC member that chose 'Yes' rather than 'Maybe'.

Note that just optimising the flow for the (global) payoff can lead to very *unfair* assignments: some PC members might receive many papers that were labelled by 'No', while others have none. This can be addressed as well; we refer to the discussion in `http://corner.mimuw.edu.pl/?p=811`.

2.4.3. The Markov Decision Problem. A *Markov decision process (MDP)* is a discrete-time stochastic control process that can be used to model certain optimisation problems where outcomes are partly random and partly under the control of a decision maker. They are sometimes viewed as "1.5 player game" since the controller is viewed as one player and nature (randomness) is viewed as half a player.

Let S be a finite set. A *probability distribution on* S is a function $P\colon S \to [0,1] \subset \mathbb{R}$ such that $\sum_{s\in S} P(s) = 1$ for all $s \in S$.

Definition 2.4.5. A *Markov decision process* is a tuple $(S, A, P_{s,a}, R_{s,a})$ where

- S is a finite set, called the *states*,
- A is a finite set, called the *actions*,
- $P_{s,a}$ is a probability distribution on S for each $s \in S$ and $a \in A$, and
- $R_{s,a} \in \mathbb{Q}$ is called the *reward* for each $s \in S$ and $a \in A$.

At each time step, the process is in some state $s \in S$ and the decision maker may choose any action $a \in A$. The process responds at the next time step by randomly moving into a new state $s' \in S$ according to the probability distribution $P_{s,a}$, and giving the decision maker a corresponding reward $R_{s,a}$. A *policy function* π is a (potentially probabilistic) mapping from S to A. A fixed policy π and a start state s_0 gives rise to a so-called *Markov chain*: to each sequence $s_0, s_2, \ldots, s_n$ we can associate a probability that this sequences arises if for every $i \in \{0, \ldots, n-1\}$ the state s_{i+1} is chosen independently at random according to the distribution $P_{s_i,\pi(s_i)}$. We are interested in finding a policy π for the decision maker that maximises the expected total reward, i.e., if the process starts in vertex s_0 the goal is to maximise the expectation of

$$\sum_{t=0}^{\infty} R_{s_t,\pi(s_t)}. \tag{17}$$

However, note that in general this expectation might not be finite. In this text, we work in a setting with so-called *discounting*, i.e., future rewards are discounted according to some constant factor $\beta \in [0,1)$. The motivation is that discounting

(1) applies directly to many economic problems,
(2) has an elegant theory, and

(3) allows for efficient computation of optimal policies using linear programming.

The *expected discounted payoff* for some discount factor $\beta \in [0,1)$, start state s_0, and policy π is defined to be the expectation of

$$v_\pi^\beta(s_0) := \sum_{t=0}^{\infty} \beta^t R_{s_t,\pi(s_t)}. \tag{18}$$

We say that a policy π^* is *β-discount optimal*, for $\beta \in [0,1)$, if $v_{\pi^*}^\beta(s) \geq v_\pi^\beta(s)$ for every $s \in S$ and every policy π. Using LP duality, we will show that there always exists such an optimal policy. The task to compute such an optimal policy is sometimes also referred to as the *Markov decision problem (with discounting)*.

REMARK 2.4.6. Note that we may also consider $R_{s,a}$ as a *cost* instead of a reward; in this case, we are interested in *minimizing* the quantity in (18); clearly, by negating all the values in R we can computationally translate between the two settings.

REMARK 2.4.7. The discount factor $\beta < 1$ is necessary for formulating the Markov decision problem as an LP. The reason is that some policies might lead to Markov chains that are not *ergodic*, and in such a situation the expected discounted payoff cannot be described so easily via a linear program. The idea that linear programming can be used in the discounting case is due to d'Epenoux [**11**].

To present alternative descriptions of the values of v_π^β we need the following observation.

LEMMA 2.4.8. *Let $P \in [0,1]^{n\times n}$ be a square matrix and $\beta \in [0,1)$. Then the matrix $I - \beta P$ is invertible and*

$$(I - \beta P)^{-1} = \sum_{t=0}^{\infty} \beta^t P^t.$$

PROOF. Note that for each $t \in \mathbb{N}$ we have that

$$(I - \beta P)(I + \beta P + \cdots + (\beta P)^t) = I - (\beta P)^{t+1}$$

and $\lim_{t\to\infty}(\beta P)^t = 0$; this implies the statement of the lemma. □

Let π be a policy for a given MDP. We view v_π^β as a vector whose entries are indexed by S. Let P_π be the square matrix whose entry at row s and column s' equals $P_{s,\pi(s)}(s') = \sum_{a\in A} P_{s,a}(s')\pi(s,a)$. Likewise, R_π denotes the vector whose entry at position s equals $R_{s,\pi(s)} = \sum_{a\in A} R_{s,a}\pi(s,a)$. Then v_π^β can be written as

$$\begin{aligned} v_\pi^\beta &= \sum_{t=0}^{\infty} \beta^t P_\pi^t R_\pi \\ &= (I - \beta P_\pi)^{-1} R_\pi \qquad \text{(by Lemma 2.4.8).} \end{aligned}$$

DEFINITION 2.4.9. The *β-discounted value vector* of the MDP is defined to be

$$v^\beta := \sup_\pi v_\pi^\beta.$$

So a policy π^* is β-discount optimal if $v_{\pi^*}^\beta = v^\beta$. A central role in the theory of discounted MDPs is a certain optimality equation, which is also called the *Bellman equation* (Equation (19)).

PROPOSITION 2.4.10. *If π^* is β-discount optimal, then*

$$v^\beta = R_{\pi^*} + \beta P_{\pi^*} v^\beta \tag{19}$$

PROOF.

$$v^\beta = v^\beta_{\pi^*} = \sum_{t=0}^{\infty} \beta^t P^t_{\pi^*} R_{\pi^*} = R_{\pi^*} + \beta P_{\pi^*} v^\beta_{\pi^*} = R_{\pi^*} + \beta P_{\pi^*} v^\beta. \qquad \square$$

The following can be seen as a converse to Proposition 2.4.10.

PROPOSITION 2.4.11. *Let $v\colon S \to \mathbb{R}$ be such that*

$$v(s) \geq R_{s,a} + \beta \sum_{s' \in S} P_{s,a}(s') v(s') \tag{20}$$

holds for all $s \in S$ and $a \in A$. Then for every policy π we have $v \geq v^\beta_\pi$.

PROOF. Multiplying each inequality 20 by $\pi(s,a)$ and summing them over all $a \in A$ yields

$$v(s) \geq R_\pi(s) + \beta \sum_{s' \in S, a \in A} P_{s,a}(s') \pi(s,a) = R_\pi(s) + \beta \sum_{s' \in S} P_{s,\pi(s)}$$

which can be written in matrix form as follows.

$$v \geq R_\pi + \beta P_\pi v \tag{21}$$

Note that (21) applied k times shows that

$$v \geq R_\pi + \beta P_\pi R_\pi + \beta^2 P^2_\pi R_\pi + \cdots + \beta^k P^k_\pi v$$

holds for all $k \in \mathbb{N}$, which implies that

$$v \geq \sum_{t \in \mathbb{N}} \beta^t P^t_\pi R_\pi = v^\beta_\pi.$$

MORE DETAIL NEEDED! $\square$

Exercises.

(33) Let $\beta \in [0,1)$ be a discount factor. Show how to transform an arbitrary MDP into a new MDP such that the expected total payoff of the new MDP equals the expected β-discounted payoff of the original MDP. This is meant in the following sense: the new MDP contains the state space S of the old MDP, and
- for any policy π for the old MDP there exists a policy ρ of the new MDP such that $v^\beta_\pi(s)$ equals the expected total payoff of the new MDP for policy ρ, for every start state $s \in S$, and conversely
- for every strategy ρ of the new MDP there exists a strategy π for the old MDP such that $v^\beta_\pi(s)$ equals the expected total payoff for ρ, for every start state $s \in S$.

The optimality equation motivates the following linear program.

$$\begin{aligned} &\text{Minimise } \sum_{s \in S} \frac{v_s}{n} && (22)\\ &\text{subject to } v_s \geq R_{s,a} + \beta \sum_{s' \in S} P_{s,a}(s') v_{s'} && \text{for all } s \in S \text{ and } a \in A. \end{aligned}$$

The dual program is as follows (following the dualisation recipe from Section 2.3.7).

$$\begin{aligned} &\text{Maximize} && \sum_{s\in S, a\in A} R_{s,a}x_{s,a} && && (23)\\ &\text{subject to} && \sum_{a\in A} x_{s',a} - \sum_{s\in S,a\in A} \beta P_{s,a}(s')x_{s,a} = 1/n && \text{for every } s'\in S\\ & && x_{s,a}\geq 0 && \text{for all } s\in S, a\in A. \end{aligned}$$

Let x^0 be a solution to (23). For $s\in S$, define $x^0_s := \sum_{a\in A} x^0_{s,a}$. Note that $x^0_s > 0$ and that for every $s\in S$ and $a\in A$ we have $x^0_{s,a}/x^0_s \in [0,1]$. Therefore, the map π that sends s to a with probability $x^0_{s,a}/x^0_s$ may be viewed as a probabilistic policy for the MDP. The following is taken from [**15**].

THEOREM 2.4.12. *For a given MDP, the expected discounted payoff of an optimal policy equals the optimal value of the LP* (22).

PROOF. We first show that the given LP has a feasible solution with a finite value. Define $m := \min_{s\in S,a\in A} R_{s,a}$ and $M := \max_{s\in S,a\in A} R_{s,a}$. Note that setting v_s to $\frac{M}{1-\beta}$ satisfies all the constraints in (22):

$$\big(1-\beta\sum_{s'\in S} P_{s,a}(s')\big)\frac{M}{1-\beta} = M \geq R_{s,a}.$$

We show that the value of every feasible solution $(v_s)_{s\in S}$ to (22) is bounded below by $\frac{m}{1-\beta}$. Let $\tilde{s}\in S$ be such that $v_{\tilde{s}} \leq v_{s'}$ for all $s'\in S$. Then for all $a\in A$

$$v_{\tilde{s}} \geq R_{\tilde{s},a} + \beta\sum_{s'\in S} P_{\tilde{s},a}(s')v_{\tilde{s}} = R_{\tilde{s},a} + \beta v_{\tilde{s}}$$

and hence $v_{\tilde{s}} \geq \frac{R_{\tilde{s},a}}{1-\beta}$ and

$$v_s \geq v_{\tilde{s}} \geq \frac{1}{1-\beta}R_{\tilde{s},a} \geq \frac{m}{1-\beta}.$$

It follows from LP duality (Theorem 2.3.2) that both the LP (22) and its dual (23) have a (finite, optimal) solution. Moreover, by complementary slackness (Corollary 2.3.4), if $(v^*_s)_{s\in S}$ is an optimal solution to (22) and $(x^*_{s,a})_{s\in S,a\in A}$ is an optimal solution to (23), then

$$v^*_s = R_{s,a} + \beta\sum_{s'\in S} P_{s,a}(s')v^*_{s'} \tag{24}$$

for all $s\in S$ and $a\in A$ such that $x^*_{s,a} > 0$. From the constraints of the dual LP we obtain that for every $s'\in S$

$$\sum_{a\in A} x^*_{s,a} = 1/n + \beta\sum_{s\in S,a\in A} P_{s,a}(s')x^*_{s,a} > 0.$$

Let π^* be the policy of the MDP defined from x^* as explained earlier. Then multiplying (24) by $x^*_{s,a}/\sum_{a\in A} x^*_{s,a} \geq 0$ and summing over all $a\in A$ yields, in vector notation,

$$v^* = R_{\pi^*} + \beta\sum_{s'\in S} P_{x^*}(s')v^*_{s'} = R_{\pi^*} + \beta P_{\pi^*}v^*. \tag{25}$$

Hence, Proposition 2.4.11 implies that

$$v^* = v^\beta_{\pi^*} = v^\beta. \qquad \square$$

Exercises.

(34) Show that for $\gamma = 1$ the LP (22) is infeasible.

(35) Show that (23) is indeed the dual of (22).
(36) Modify the definitions and statements in this section so that the reward function $R(s, a)$ not only depends on the present state s and the action a which was taken, but additionally also on the next state s' (which is chosen from S with probability $p(a, s, s')$).
(37) Show that the modified setting where in each state $s \in S$ a potentially different finite set of actions $A(s)$ is available reduces to the present setting.
(38) (Freely following and simplifying [**11**]) Let us consider an enterprise which produces, stocks, and sells a single item. The quantities demanded per day are independent random variables with the same known probability distribution (see Section 3.3.3 for definitions of basic concepts from probability theory); those demands which are not immediately satisfied are lost permanently. The existing relationship between cost and output is taken for granted; in other words, we ignore the possibility of adapting the company's equipment more closely to the actual environment. The production rate is adjusted daily to the current situation. One wants to find the optimal sequential decision rule, defined as the one which minimizes the expected future costs of the enterprise. Discuss how to formally model the task.

2.4.4. Von Neumann Minimax Theorem. In this section we consider games with two players; each player has a finite set of strategies. The *payoff* for each of the players is determined by the strategies chosen by both players. We focus on *zero sum games* which are games in which the payoff to the second player is the negative payoff to the first, so the sum of their payoffs is zero. We also refer to the first player as the *row player* and to the second player as the *column player*.

Example 3. The Paper-Scissor-Stone game is the game which is given by the following payoff table for the first player.

		Column Player		
		Paper	Scissor	Stone
Row Player	Paper	0	-1	1
	Scissor	1	0	-1
	Stone	-1	1	0

△

If the row player plays strategy i, and the column player plays strategy j, the payoff to the row player is $a_{i,j}$. If the row player plays first, she can obtain the profit

$$\max_i \min_j a_{i,j}.$$

If the row player plays last, she can obtain the profit

$$\min_j \max_i a_{i,j}.$$

In our concrete game Paper-Scissor-Stone, we have that for all strategies j for the column player

$$\max_i a_{i,j} = 1$$

and for all strategies i for the row player we have

$$\min_j a_{i,j} = -1$$

So it is a big advantage to play second in the above game.

Now we change the game. Each of the players has to expose a probability distribution $\Delta = \{x \in \mathbb{R}^3 \mid x \geq 0, \sum x_i = 1\}$ on the strategies; these are called *mixed*

strategies. Is it still an advantage to play second in the game? If the row player plays first, her expected profit is

$$P_1 := \max_{x\in\Delta} \min_{y\in\Delta} \sum_{i,j} x_i y_j a_{i,j}.$$

If the row player plays second, her expected profit is

$$P_2 := \min_{y\in\Delta} \max_{x\in\Delta} \sum_{i,j} x_i y_j a_{i,j}.$$

Clearly, $P_1 \leq P_2$.

EXAMPLE 4. In the Paper-Scissor-Stone game, the row player can play the mixed strategy that assigns probability $1/3$ to each of the three options, and hence we have

$$P_1 \geq \min_{y\in\Delta} \sum_{i,j} \frac{y_i}{3} = 0.$$

On the other hand, if the row player assigns, for example, to stone the probability $1/2$ and to paper and scissors the probability $1/4$, then column player may assign probability 1 to paper. In this case, the payoff for the row player is -1 with probability p, it is 0 with probability $1/4$, and it is 1 with probability $1/4$. Hence, the expected payoff is $-1/4$. It is easy to see that the expected payoff is smaller than 0 whenever the row player assigns to one of the options a probability that is larger than $1/3$. Hence, we have $P_1 = 0$. A similar argument, with the role of the players exchanged, applies if the row player plays second, so we have $P_2 = 0$ as well. △

We will show that $P_1 = P_2$ holds in general. For compact notation, we write $A = (a_{i,j})$ for the payoff matrix, so that the expected payoff $\sum_{i,j} x_i y_j a_{i,j}$ can be written as $x^\top Ay$.

THEOREM 2.4.13 (Von Neumann Min-Max Principle). *Let $A = (a_{i,j}) \in \mathbb{Q}^{m\times n}$ be a two-person zero-sum game. Let Γ, Δ be the set of all mixed strategies for row and column player, respectively. Then there are $\tilde{x} \in \Gamma$ and $\tilde{y} \in \Delta$ such that*

$$\max_{x\in\Gamma} \min_{y\in\Delta} x^\top Ay = \min_{y\in\Delta} \max_{x\in\Gamma} x^\top Ay = \tilde{x}^\top A\tilde{y}.$$

The following terminology is important and will help us to give a clear presentation of the proof. For simplicity, we call player one *Alice* (the row player) and player two *Bob* (the column player). The *worst-case payoff* for a mixed strategy $x \in \Gamma$ for Alice is defined to be

$$\alpha(x) := \min_{y\in\Delta} x^\top Ay$$

and likewise the worst-case payoff for a mixed strategy $y \in \Delta$ for Bob is

$$\beta(y) := \min_{x\in\Gamma} x^\top Ay.$$

These are well-defined functions since Δ and Γ are compact sets (and this is why we chose $\mathbb{R}$ instead of $\mathbb{Q}$ in the definition of mixed strategies; we will see below that both settings are equivalent in the sense that the probabilities in the mixed strategies $\tilde{x}$ and $\tilde{y}$ in Theorem 2.4.13 can be chosen to be rational, as we will see in the proof). A pair $(\tilde{x}, \tilde{y})$ such that

$$\alpha(\tilde{x}) = \tilde{x}^\top A\tilde{y} = \beta(\tilde{y})$$

is called a *mixed Nash equilibrium*. Alice's mixed strategy $\tilde{x}$ is called *worst-case optimal* if $\alpha(\tilde{x}) = \max_{x\in\Gamma} \alpha(x)$, and we make the analogous definition for a mixed strategy of Bob.

PROOF. We first show how worst-case optimal mixed strategies $\tilde{x}$ for Alice and $\tilde{y}$ for Bob can be found by linear programming. Then we prove that $\alpha(\tilde{x}) = \beta(\tilde{y})$ holds.

First notice that Bob's best response to a *fixed* mixed strategy x of Alice can be found by solving a linear program. That is, $\alpha(x)$, with x a concrete vector of m numbers, is the optimal value of the following linear program in the variables $y_1, \ldots, y_n$:

$$\begin{aligned} &\text{Minimize } x^\top A y \\ &\text{subject to } \sum_{j=1}^{n} y_j = 1 \\ &y \geq 0 \end{aligned}$$

In particular, it follows that if x is rational then $\alpha(x)$ is rational, too. Unfortunately, $\alpha(x)$ is not a linear function, so we cannot directly formulate the maximisation of $\alpha(x)$ as a linear program. Fortunately, we can circumvent this issue by using LP duality. The dual of the above LP can be computed via the dualization recipe from Section 2.3.7: first we have to write the primal as a maximisation problem by changing the sign of the objective function. Also replacing $\sum_{j=1}^{n} y_j = 1$ by $-\sum_{j=1}^{n} y_j = -1$, we obtain for the dual

$$\begin{aligned} &\text{Minimize } -x_0 \\ &\text{subject to } -x_0\mathbf{1} \geq -(x^\top A)^\top. \end{aligned}$$

Note that this LP has just one variable! It can be rewritten to

$$\begin{aligned} &\text{Maximize } x_0 \\ &\text{subject to } A^\top x \geq \mathbf{1}x_0. \end{aligned}$$

By the duality theorem, the optimal value of the dual LP equals $\alpha(x)$. In order to maximise $\alpha(x)$ over all mixed strategies x of Alice, we derive a new LP from the dual in which $x_1, \ldots, x_m$ are now regarded as variables.

$$\begin{aligned} &\text{Maximize } x_0 \\ &\text{subject to } A^\top x \geq x_0\mathbf{1} \\ &\sum_{i=1}^{m} x_i = 1 \\ &x \geq 0 \end{aligned} \tag{26}$$

Clearly, there exist feasible solutions to this LP. If $(\tilde{x}_0, \tilde{x})$ denotes an optimal solution, we have by construction that

$$\tilde{x}_0 = \alpha(\tilde{x}) = \max_{x \in \Gamma} \alpha(x).$$

Symmetrically, we can construct an LP for computing a worst-case optimal mixed strategy $\tilde{y}$ for Bob:

$$\begin{aligned} &\text{Minimize } y_0 \\ &\text{subject to } A^\top y \geq y_0\mathbf{1} \\ &\sum_{j=1}^{n} y_j = 1 \\ &y \geq 0 \end{aligned} \tag{27}$$

If $(\tilde{y}_0, \tilde{y})$ denotes an optimal solution to this LP, then $\tilde{y}_0 = \beta(\tilde{y}) = \min_{y \in \Delta} \beta(y)$. Now observe that the two linear programs (26) and (27) are dual to each other! By LP duality we obtain that $\tilde{x}_0 = \tilde{y}_0$ and hence $\alpha(\tilde{x}) = \beta(\tilde{y})$, as required. □

A general (not necessarily zero-sum) 2-player game is given by *two* matrixes A and B, one for the payoff for the first player, and one for the payoff for the second player. Zero-sum games are then the special case where $A = -B$.

REMARK 2.4.14. Nash equilibria also exist for general 2-player games [**29**], but this is outside the scope of this course. Finding a Nash equilibrium for such game is a very interesting problem which is not known to be in P, but believed not to be NP-hard (since if it were NP-hard, then NP=coNP). On the other hand, there is also some evidence that the problem might not be in P; see [**10**] and the references therein.

Exercises.

(39) Find a mixed Nash equilibrium for the game "Papers-Scissors-Stone-Well" which is the modification of Papers-Scissors-Stone where an additional pure strategy "Well" has been added, which wins against Stone and Scissors, but looses against paper.

(40) Where in the proof of Theorem 2.4.13 did we use the assumption that A is a zero-sum game? Where does the same argument fail for general 2-player games?

(41) Consider a three-player zero-sum game, given by three reward matrices $A, B, C \in \mathbb{Q}^{n_1 \times n_2 \times n_3}$ in which the rewards of the three players always sum to zero, $A + B + C = 0$. Show that finding a Nash equilibrium in such a game is at least as hard as the same problem for general (not necessarily zero-sum) two-player games.

(42) Consider the following 2-player game. We are given a directed graph $(A \cup B, E)$ where $A \cap B = \emptyset$ and $E \subseteq (A \times B) \cup (B \times A)$. First player chooses a probability distribution p on A, second player independently chooses a probability distribution q on B. The first player wins if

$$\sum_{(a,b) \in E \cap A \times B} p(a)q(b) - \sum_{(b,a) \in E \cap B \times A} q(b)p(a) > 0;$$

otherwise second player wins. Show how to determine in polynomial time in the size of a given digraph which player has a winning strategy.

2.4.5. Simple stochastic games. A *simple stochastic game (SGG)* is a special case of a *stochastic game* as introduced by Shapley in 1953 (a grad school friend of Nash from the previous section). They are played on a directed graph $G = (V, E)$ whose vertices are partitioned into three disjoint sets $V_{\max}$, $V_{\min}$, V_{stoch}, called max-, min-, and stochastic vertices, respectively. Moreover, there is a distinguished start vertex s, and two distinguished terminal vertices $t_{\max}$ and $t_{\min}$. Each vertex has at least one outgoing edge, except for $t_{\max}$ and $t_{\min}$, which have no outgoing edge (they are *sinks*).

The game is played by two players, called the max player and the min player. At the start of the game, a token is placed on the start vertex. In each round, the token is moved from a vertex v along some of the outgoing edges at v. When the token is positioned on a max vertex, then player max decides along which edge the token is moved, and when the token is on a min vertex then player min decides. When the token is on a stochastic vertex, then each outgoing edge is chosen with equal probability (so in some sense there is a third player, randomness; for this reason, this type of game is also called a $2\frac{1}{2}$-player game). The game ends when the token

reaches $t_{\max}$ or $t_{\min}$; in the first case, player max wins, and in the second case, player min wins. If the play continues forever then player min wins. See Figure 2.3 for an example.

A *(positional) strategy* $\sigma\colon V_{\max} \to E$ for player max is a function that selects for each vertex $u \in V_{\max}$ one outgoing edge (u, w). Corresponding to a strategy σ is a subgraph G_σ of G obtained from G by removing from each max vertex the outgoing edges of vertices in $V_{\max}$ that are not selected by σ. Strategies $\tau\colon V_{\min} \to E$ for player min are defined analogously. If σ is a strategy for max and τ is a strategy for min, then $G_{\sigma,\tau}$ is the subgraph of G where for each vertex in $V_{\max} \cup V_{\min}$ the graph only contains the outgoing edge selected by the strategies. Note that the game for $G_{\sigma,\tau}$ can be viewed as a Markov chain. We say that the SGG *halts with probability 1* if for all pairs of strategies σ, τ every vertex in $G_{\sigma,\tau}$ has a path to a sink vertex. It can be shown that if a player can win, it can win following such a positional strategy (see [**9**]).

The *value* $v_{\sigma,\tau}(u)$ of $u \in V$ with respect to σ and τ is the probability that the pebble reaches $t_{\max}$ starting from u in the Markov chain given by $G_{\sigma,\tau}$. The *optimal value* $v(u)$ of a vertex $u \in V$ is defined to be

$$\max_{\sigma} \min_{\tau} v_{\sigma,\tau}(u).$$

The *value of the game* is defined to be $v(s)$. The following result does not follow from the Minimax theorem since the players have to play pure strategies.

THEOREM 2.4.15 (of [**39**] and [**9**]). *Let G be a simple stochastic game that halts with probability one. Then we have*

$$\max_{\sigma} \min_{\tau} v_{\sigma,\tau}(s) = \min_{\tau} \max_{\sigma} v_{\sigma,\tau}(s).$$

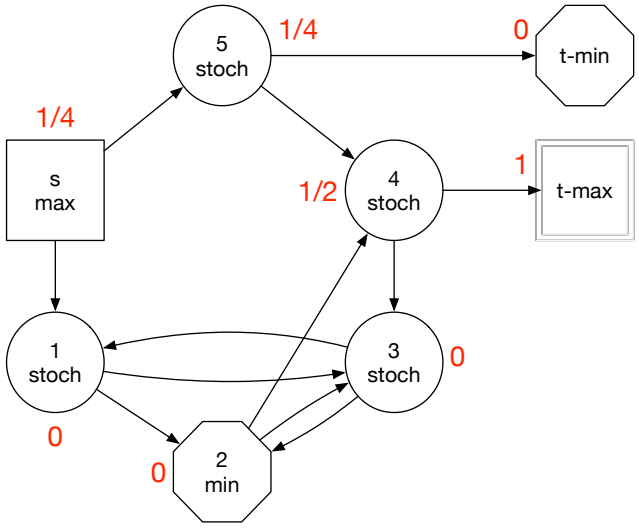

FIGURE 2.3. An example of a simple stochastic game with the value of each vertex in red.

The primary question about a simple stochastic game is the question: what is its value? There is no polynomial-time algorithm known that solves this problem. In particular, we will be interested in the decision problem whether the value of the game is at least 1/2 (i.e., we decide whether min has a greater winning probability than max if min plays optimally).

DEFINITION 2.4.16. A *solution* to a SSG is an assignment $\tilde{x}\colon V \to [0,1]$ that satisfies

- $\tilde{x}(t_{\max}) = 1$, $\tilde{x}(t_{\min}) = 0$,
- $\tilde{x}(u) = \max_{(u,w)\in E} \tilde{x}(w)$ for $u \in V_{\max}$,
- $\tilde{x}(u) = \min_{(u,w)\in E} \tilde{x}(w)$ for $u \in V_{\min}$, and
- $\tilde{x}(u) = \frac{\tilde{x}(w_1)+\cdots+\tilde{x}(w_k)}{k}$ for $u \in V_{\text{stoch}}$ with out-neighbours $w_1, \ldots, w_k$.

SSGs might in general not have unique solutions.

Exercises.

(43) Present a SGG with infinitely many solutions.

(44) Show that the assumption that G is stopping is necessary in Theorem 2.4.15.

However, it can be shown that computing the value of an SGG can be (efficiently) reduced to the case where the solutions of the SGG are unique; this is technical and out of the scope of this text. We refer to [**9**] for details. If a game has a unique solution and contains no max vertices, then the solution can be found by the following linear program.

$$\begin{aligned}
\text{minimise } & \sum_{u\in V} x_u \\
\text{subject to } & x_u \le x_w && \text{if } u \in V_{\min} \text{ and } (u,w) \in E \\
& x_u \le \frac{x_{w_1} + \cdots + x_{w_k}}{k} && \text{if } u \in V_{\text{stoch}} \\
& && \text{and } w_1, \ldots, w_k \text{ are the out-neighbours of } u \\
& x_u \le 1 && u \in V \\
& x_{t_{\max}} = 1, x_{t_{\min}} = 0.
\end{aligned}$$

THEOREM 2.4.17. *The LP above has an optimal solution, $(x^*_u)_{u\in V}$, and $x^*_u = v(u)$ as defined above.*

PROOF. It is clear that $x_u := v(u)$, for all $u \in V$, is a solution to the game. It is also clear that every solution to the game gives a valid solution to the LP. Since all variables are upper bounded by 1 and the objective is to maximize $\sum_{u\in V} x_u$ it follows that the LP has an optimal solution x^*.

We claim that every optimal solution x^* to the LP gives a solution to the game. Suppose otherwise that for some $u \in V_{\min}$ we have $x^*_u < x^*_w$ for all out-neighbours w of u. Then we construct a better solution x' to the LP as follows: $x'(u) := \min_{(u,w)\in E}(x^*_w)$ and $x'(w) := x^*(w)$ for all $w \in V \setminus \{u\}$. The new solution satisfies all the constraints but has a strictly larger objective function, contradicting the maximality of x^*. Now suppose for contradiction that for some $u \in V_{\text{stoch}}$ with out-neighbours $w_1, \ldots, w_k$ we have $x^*_u < \frac{x^*_{w_1}+\cdots+x^*_{w_k}}{k}$. Then similarly as above we can construct a solution with a strictly larger objective function. It follows that if the game has a unique solution, then $x^* = v(u)$. □

This shows that deciding whether the value of a game is larger than a given threshold is in the complexity class NP: we simply guess an outgoing edge for each $v \in V_{\max}$, remove all other outgoing edges from v, and turn v into a min vertex. This corresponds to selecting one strategy for max. The resulting game has no more max vertices and we can find an optimal counter-strategy for min by the LP above.

We now want to argue that the problem is also in coNP. Theorem 2.4.15 shows that in order to compute $v(s)$ we can therefore swap the roles of the players, and solve a very similar linear program for in the case that the game has no min-nodes:

$$\begin{aligned}
&\text{maximise } \sum_{u \in V} x_u \\
&\text{subject to } x_u \geq x_w && \text{if } u \in V_{\max} \text{ and } (u,w) \in E \\
&\qquad x_u \geq \frac{x_{w_1} + \cdots + x_{w_k}}{k} && \text{if } u \in V_{\text{stoch}} \\
&\qquad && \text{and } w_1, \ldots, w_k \text{ are the out-neighbours of } u \\
&\qquad x_u \geq 0 && u \in V \\
&\qquad x_{t_{\max}} = 1, x_{t_{\min}} = 0
\end{aligned}$$

This shows that deciding the winner in a simple stochstic game is in NP $\cap$ coNP (a result of Condon [**9**]).

CHAPTER 3

The Probabilistic Method

For further reading with advanced material, we recommend [**3**]. For more introductory texts, see [**23**, **27**].

3.1. Tournaments

In this section we present one of the historically first applications of the probabilistic method, due to Erdős [**13**], solving a problem of the logician Schütte.

A *tournament* is a directed graph (V, E) such that for any two distinct vertices $x, y \in V$ we have either $(x, y) \in E$ or $(y, x) \in E$ (but not both). We can imagine this as a soccer tournament with teams V where every team plays once against every other team; each game has precisely one winner, and there is an edge $(x, y) \in E$ if x wins against y.

Clearly, there are tournaments such that for every team x there exists another team y such that $(y, x) \in E$; we call such tournaments 1-*paradoxical*. More generally, a tournament (V, E) is called k-*paradoxical* if for all vertices $x_1, \ldots, x_k \in V$ there exists a vertex $y \in V$ such that $(y, x_1), \ldots, (y, x_k) \in E$. Our question is: does there exist for every $k \geq 1$ a k-paradoxical tournament? The construction of a 1-paradoxical tournament is trivial, and a 2-paradoxical tournament is still easy to construct by hand.

Exercises.

(43) Construct such a tournament for $k = 2$.

For $k \geq 3$, it seems quite difficult to explicitly construct finite k-paradoxical tournaments – but using the probabilistic method, it will be quite easy to show their existence.

THEOREM 3.1.1. *For every k there exists a k-paradoxical finite tournament.*

PROOF. For notational simplicity, we present the proof only for $k = 3$, but the general case can be shown analogously.

Let us consider a *random tournament*, i.e. we imagine that between any two distinct vertices $x, y \in V$ we toss a fair coin to decide whether we add the edge (x, y) or the edge (y, x).

Let $x, y, z \in V$. The probability that another vertex $w \in V \setminus \{x, y, z\}$ wins against all three of them is $2^{-3} = \frac{1}{8}$. Thus, the probability that w loses against one of them is $1 - \frac{1}{8} = \frac{7}{8}$. The probability that each of the $n - 3$ other players loses against at least one of x, y, z is $\left(\frac{7}{8}\right)^{n-3}$ because the results of all the involved matches with x, y, z are mutually independent.

The set $\{x, y, z\}$ can be selected in $\binom{n}{3}$ many ways. Now we need the elementary fact that the probability of a sum of events is *at most* the sum of the probabilities of the events. Equality occurs only when the events are disjoint, which will not be the case in our situation. The computation of the probability of the union can be quite complicated (involving for example so-called 'inclusion-exclusion arguments') but surprisingly often the crude bound above will suffice. We obtain the probability that for

at least one of those sets no player beats all three elements x, y, z simultaneously is at most

$$\binom{n}{3}\left(\frac{7}{8}\right)^{n-3}.$$

With a standard computer calculator one can compute that for $n = 90$ this bound is still larger than 1, but for $n = 91$ it is $121485 \cdot 0.00000788331\ldots = 0.957704\ldots$ Therefore, there exists at least one tournament with 91 players in which any 3 players are simultaneously beaten by some other player. □

We see that the key point in the proof of this theorem is that the exponential decay of $\frac{7}{8}^{n-3}$ wins against the polynomial growth of $\binom{n}{3}$. In fact, we have shown something stronger: instead of the *existence* of k-paradoxical tournaments, we have shown that for large n almost all tournaments with n vertices are k-paradoxical![1]

In this next section we introduce powerful notation to keep more complex arguments about asymptotic growth of functions notationally manageable.

What is the smallest k-paradoxical tournament? Clearly, we need 3 vertices for $k = 1$. For $k = 2$, you already have some bound because you solved Exercise (43) above. For $k = 3$, we have seen that 91 vertices suffice for $k = 3$, but it seems clear from the crude union bound in the proof that most likely smaller tournaments exist. In fact, there is the following lower bound (which to the best of our knowledge could be sharp!), which is not hard to show by induction on k.

PROPOSITION 3.1.2. *If (V, E) is a k-paradoxical tournament, then $|V| \geq 2^{k+1} - 1$.*

3.2. Asymptotic Growth

In the application of the probabilistic method that we have seen in the previous section, the key point in the proof was that some probability was tending to zero because it was a product of polynomial with a function with exponential decay. We can already guess that *asymptotic growth* plays an important role in this chapter. We introduce very useful notation to compare functions with respect to their asymptotic growth, going back to Bachman and Landau. Then we recall some basic estimate that will be useful later.

3.2.1. O-notation. The letters o and O stand for the *order* of growth of the function. The *big-O notation* is used to express asymptotic upper bounds, and the *little-o notation* to express that functions are asymptotically negligible when compared to other functions. We mention that there exists related notation to describe other kinds of bounds on asymptotic growth, e.g., Θ, Ω, ω, of which we only need Θ in this text, so we skip the definitions of the others (in particular since there are competing definitions for Ω, one from number theory and one from complexity theory).

Let $g\colon \mathbb{R} \to \mathbb{R}$ (we use $\mathbb{R}$ for convenience; the same definition applies to other domains such that as $\mathbb{N}$ and $\mathbb{Q}$, etc). Then $O(g)$ is the set of all functions $f\colon \mathbb{R} \to \mathbb{R}$ such that there exists $c, x_0 \in \mathbb{R}$ such that $|f(x)| \leq c|g(x)|$ for all $x \geq x_0$. Note that

$$f \in O(g) \Leftrightarrow \limsup_{x\to\infty} \left|\frac{f(x)}{g(x)}\right| < \infty. \tag{28}$$

We mention that similar definitions are used if ∞ is replaced by some $a \in \mathbb{R}$. In typical usage, the formal definition of $O(g)$ is not used directly; rather, we first use the following simplification rules:

[1]This shows that not only finding a needle in a haystack, but also that finding hay in a haystack can be difficult [**4**].

- if $g(x)$ is a sum of several terms, if there is one with largest growth rate, then we drop all other terms;
- if $g(x) = c \cdot f(x)$ and c is a constant that does not depend on x, then c can be omitted.

When we write $O(g)$, we typically choose g to be as simple as possible. O-notation can also be used within arithmetic terms. For example, $h + O(g)$ denotes the set of functions of the form $h + f$ for $f \in O(g)$. In other words, $k \in h + O(g)$ is equivalent to $k - h \in O(G)$.

We write $o(g)$ for the set of all functions $f\colon \mathbb{R} \to \mathbb{R}$ such that for every $\epsilon \in \mathbb{R}_{>0}$ there exists $x_0 \in \mathbb{R}$ such that $|f(x)| \leq \epsilon|g(x)|$ for all $x \geq x_0$. Informally, $f \in o(g)$ means that asymptotically, the growth of f is negligible compared to the growth of g. For example, $x \mapsto 2x$ is in $o(x \mapsto x^2)$, and $x \mapsto 1/x$ is in $o(1)$. Note that $o(g) \subseteq O(g)$, and that

$$f \in o(g) \Leftrightarrow \lim_{x\to\infty} \frac{f(x)}{g(x)} = 0. \tag{29}$$

Similarly as in the case of the O-notation we may use the o-notation in arithmetic expressions. Note that if $f \in o(g)$ and c is a constant, then $cf \in o(g)$. Frequent notation is to write $f \ll g$ (or $g \gg f$) if $f \in o(g)$.

We write $\Theta(g)$ for the set of all functions f such that there are constants $c, C > 0$ and $x_0 \in \mathbb{R}$ such that $cg(x) \leq |f(x)| \leq C|g(x)|$ for every $x \geq x_0$. In other words, $f \in \Theta(g)$ if $f \in O(g)$ and $g \in O(f)$.

Finally, we write $f \sim g$ if

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = 1$$

and we say that f and g are *asymptotically equivalent* (for $x \to \infty$). Clearly, if $f \sim g$ then $f \in O(g)$ and $g \in O(f)$, so $\Theta(g) = \Theta(f)$.

Exercises.

(44) Prove the statement in (28) and in (29).
(45) Show that if $f_1 \in O(f_2)$ and $f_2 \in O(f_3)$, then $f_1 \in O(f_3)$.
(46) Show that if $f_1 \in o(f_2)$ and $f_2 \in o(f_3)$, then $f_1 \in o(f_3)$.
(47) Show that if $a < b$, then $a^x \in o(b^x)$.
(48) Find an example of functions f and g such that $\Theta(g) = \Theta(h)$, but not $f \sim g$.

3.2.2. The exponential function. The following bound is very crude, but nonetheless it often turns out to be sufficient for approximately comparing the growth of functions. We need the definition of the exponential function $\exp\colon \mathbb{R} \to \mathbb{R}$ given by

$$\exp(x) := \sum_{k=0}^{\infty} \frac{x^k}{k!} = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \cdots$$

Recall that for all $x, y \in \mathbb{R}$

$$\exp(x + y) = \exp(x) \cdot \exp(y).$$

So with $e := \exp(1) \in \mathbb{R}$ we see that for $n \in \mathbb{N}$

$$\exp(n) = \exp(\underbrace{1 + \cdots + 1}_{n \text{ times}}) = e^n$$

and from calculus we know that the identity $\exp(x) = e^x$ holds for arbitrary $x \in \mathbb{R}$. In particular, for every $x \in \mathbb{R}$

$$e^x \geq 1 + x. \tag{30}$$

This is obvious for non-negative x and for $x \leq -1$, and for $-1 < x < 0$ it follows from the definition of exp by grouping and bounding the terms in pairs. Geometrically this can be explained as follows. The line defined by $y = x + 1$ is the tangent to $y = e^x$ for $x = 0$; since e^x is convex, it always remains above its tangent lines. Phrased differently (and this is how it will be applied frequently):

$$1 - x \leq e^{-x}. \tag{31}$$

Also note that if we view x as a function from $o(1)$, then

$$1 - x \in e^{-x+O(x^2)}. \tag{32}$$

Exercises.

(49) Show that $\log_b(x) \sim \ln(x)$ for every $b \in \mathbb{R}$ (where $\ln(x) := \log_e(x)$ and $\log_b(x)$ is the unique $c \in \mathbb{R}$ such that $b^c = x$).

(50) Most complexity functions $f(n)$ that we encounter in practice are *monotonic*, i.e., they satisfy $f(n) \leq f(n+1)$ for all $n \in \mathbb{N}$. Show that the following pair of monotonic functions are *incomparable* with respect to asymptotic growth.

$$f(n) = (2\lfloor (n+1)/2 \rfloor - 1)!$$
$$g(n) = (2\lfloor n/2 \rfloor)!$$

The functions are constructed so that they cross each other periodically in such a way that neither is asymptotically bounded by the other.

n	1	2	3	4	5
$f(n)$	1	1	6	6	120
$g(n)$	1	2	2	24	24

(51) Show that the class of functions constructed from the identity function and constants using addition, subtraction, multiplication, division, logarithm, and exponentiation is totally ordered by asymptotic growth (i.e., for any two functions in this class, $f \in O(g)$, or $g \in O(f)$, or both).

3.3. Random Graphs

In the proof of Theorem 3.1.1, we used *random tournaments* to prove the existence of tournaments with some very specific combinatorial property. We would like to pick up this idea in the context of simple graphs as introduced in Chapter 1. To that end, we will introduce a famous fundamental model for random graphs, namely $G(n,p)$. The presentation is inspired by the text-book *The strange logic of random graphs* by Joel Spencer [**40**] which I would recommend to master students.

3.3.1. Introducing random graphs. On an intuitive level, $G(n,p)$ is a graph with vertex set $V := \{1, \ldots, n\}$ obtained as follows: for each two-element subset $\{u, v\} \in \binom{V}{2}$ we decide whether to make u and v adjacent in the graph based on flipping a biased coin: the coin comes up heads with probability p, and in this case we add the edge $\{u, v\}$, and otherwise there is no edge between u and v.

More formally, $G(n,p)$ is the probability space whose elements are all the $2^{\binom{n}{2}}$ graphs with vertex set V. As usual, subsets of the probability space are called *events*. The probability distribution is determined by the requirement that for all $u, v \in V$, the probability that u and v are adjacent is p, i.e.,

$$\Pr(\{(V,E) \mid \{u,v\} \in E) = p,$$

and that these events are for all pairs of vertices mutually independent, that is, if $e_1, e_2 \in \binom{V}{2}$ then

$$\Pr(\{(V,E) \mid e_1 \in E \text{ and } e_2 \in E\}) = \Pr(\{(V,E) \mid e_1 \in E\}) \cdot \Pr(\{(V,E) \mid e_2 \in E\})$$

Alternatively, the probability space can be defined as follows. If $H = (V, E)$ with $|E| = m$, then the probability of the event $\{H\}$ is $\Pr(\{H\}) = p^m(1-p)^{\binom{n}{2}-m}$. However, the first description is the better one when working with $G(n,p)$ since it stresses the independence of the edge probabilities. Luckily, our probability space is finite and hence every set has a measure; no measure theory is needed in what follows.

We will be interested in *properties* of random graphs: e.g., what is the probability that a graph from $G(n,p)$ is connected, that it contains K_3 as a subgraph, etc. Formally, and in the language of probability theory, a property $\mathcal{A}$ is a set of graphs (the set of all graphs that *satisfy* the property). If a graph G is in $\mathcal{A}$ we say that G *has property* $\mathcal{A}$. The probability that a graph from $G(n,p)$ has the property $\mathcal{A}$ is denoted by $\Pr(G(n,p) \in \mathcal{A})$. If $p = 1/2$, then $\Pr(G(n,p) \in \mathcal{A})$ is simply the fraction of all graphs $(V, E) \in \mathcal{A}$ in the set of all graphs with vertex set V.

DEFINITION 3.3.1. A property $\mathcal{A}$ holds in $G(n,p)$ *asymptotically almost surely (a.a.s.)* if

$$\lim_{n\to\infty} \Pr(G(n,p) \in \mathcal{A}) = 1.$$

Instead of a.a.s., many scholars use *whp.* (for *with high probability*). For $G(n,p)$, the situation where $p = 0$ or $p = 1$ are trivial. Otherwise, if p is a constant strictly between 0 and 1, it turns out that $G(n,p)$ asymptotically almost surely satisfies a strong property, called the *extension property*.

DEFINITION 3.3.2. Let $r, s \in \mathbb{N}$. The r,s-extension property, denoted by $\mathcal{A}_{r,s}$, is the property that for all disjoint sets of vertices $\{x_1, \ldots, x_r\}$ and $\{y_1, \ldots, y_s\}$ there exists a distinct vertex z that is adjacent to all of $x_1, \ldots, x_r$ and to none of $y_1, \ldots, y_s$.

THEOREM 3.3.3. *For all $r, s \in \mathbb{N}$ and $0 < p < 1$ the property $\mathcal{A}_{r,s}$ holds in $G(n,p)$ a.a.s.*

PROOF. For given vertices $x_1, \ldots, x_r, y_1, \ldots, y_s$ let $\mathrm{noz}(\bar{x}, \bar{y})$ be the event that there is no valid witness z for the extension property.

Claim. $\Pr(\mathrm{noz}(\bar{x}, \bar{y})) = (1 - p^r(1-p)^s)^{n-r-s}$. There are $n - r - s$ potential witnesses for a valid extension z. Each has probability $p^r(1-p)^s$ of being a witness since $r+s$ coin tosses must come up in a particular way. The events 'z is not a witness' are mutually independent of the z's as they involve disjoint sets of coin tosses. Thus the probability that no z is a witness is $(1 - p^r(1-p)^s)^{n-r-s}$.

There are a $\binom{n}{r}\binom{n-r}{s}$ many choices for $\bar{x}$ and $\bar{y}$. The probability that $G(n,p)$ does not satisfy $\mathrm{noz}(\bar{x}, \bar{y})$ for *some* $\bar{x}, \bar{y}$ is bounded by

$$\binom{n}{r}\binom{n-r}{s}(1 - p^r(1-p)^s)^{n-r-s},$$

which is in $o(1)$. Hence, $\mathcal{A}_{r,s}$ holds a.a.s. □

Exercises.

(52) Show that for every constant $p \in \mathbb{R}$, $0 < p < 1$, the random graph $G(n,p)$ a.a.s. contains a triangle.
(53) Show that for every constant $p \in \mathbb{R}$, $0 < p < 1$, the random graph $G(n,p)$ a.a.s. is connected.
(54) Is there a single graph that satisfies $\mathcal{A}_{r,s}$ for every r, s?
(55) Prove that for every constant $p \in \mathbb{R}$, $0 < p < 1$ and every $k \in \mathbb{N}$, the random graph $G(n,p)$ is a.a.s. *k-connected*.
(56) Discuss: if $f(p)$ denotes the probability of some event in $G(n,p)$, does the expression '$f \in o(1)$ holds a.a.s.' make sense? Is this expression really defined unambiguously?

3.3.2. The Erdős-Rényi evolution. As p goes from 0 to 1, the random graph $G(n,p)$ evolves from the empty graph to K_n. We will study properties that hold a.a.s. in $G(n,p)$ as p increases. Paul Erdős and Alfred Rényi started the area of random graphs with the discovery that for many natural graph properties $\mathcal{A}$ there exists a narrow range for p where the probability that $\mathcal{A}$ holds in $G(n,p)$ a.a.s. moves from 0 to 1. That range is typically not a constant, but a function that depends on n. A graph property $\mathcal{A}$ is called *monotone* if for every $H \in \mathcal{A}$, all subgraphs of H are also in $\mathcal{A}$.

DEFINITION 3.3.4. Let $\mathcal{A}$ be a monotone graph property and let $t, p\colon \mathbb{N} \to [0,1]$. Then t is called a *threshold function for* $\mathcal{A}$ if

$$\lim_{n\to\infty} \Pr[G(n,p) \in \mathcal{A}] = \begin{cases} 1 & \text{if } t \in o(p) \\ 0 & \text{if } p \in o(t). \end{cases}$$

EXAMPLE 5. Let $\mathcal{A}$ be the graph property *"containing a triangle"*. This property is clearly monotone. There are $\binom{n}{3}$ potential triangles and each has probability p^3 of being a triangle, so the expected number of triangles is $\binom{n}{3}p^3$. When $p(n) \ll \frac{1}{n}$ (for example, if $p = n^{-1.001}$) then the expected number of triangles is in $o(1)$ (i.e., it tends to 0). In this case, it follows from the so-called *first moment method* (presented in the next section) that the probability that $G(n,p)$ a.a.s. does *not* contain a K_3. On the other hand, if $p(n) \gg \frac{1}{n}$ (for example, if $p = n^{0.99}$) then the expected number of triangles goes to infinity. It follows from the so-called *second moment method* (presented in Section 3.3.4) that $G(n,p)$ a.a.s. *does* contain a K_3. Jointly, these two observations prove that the threshold function for containing a K_3 is $1/n$. △

3.3.3. The first moment method. One of the most basic probabilistic notions is the expected value of a random variable X. Intuitively, it is the value that we would expect to obtain if we repeated a random experiment many times and took the average of the outcomes of X. More formally, a random variable is a function that maps every element of the probability space S to a value, typically from $\mathbb{R}$. For $i \in \mathbb{R}$ we write $X = i$ as a shortcut for $\{s \in S \mid X(s) = i\}$; the shortcuts $X \geq i$, $X \leq i$, $X > i$, $X < i$ are defined analogously. Two random variables X and Y are called *independent* if the events $X \leq i$ and $Y \leq y$ are independent.

The *expectation of* X is defined as

$$E[X] := \sum_i i \cdot \Pr(X = i)$$

where the sum is over all values i in the range of X. The following properties of expectation can be easily verified; for a full introduction to probability theory, we refer to Schilling [**37**]. The first is *linearity of expectation*. Let X_1 and X_2 be random variables. Then

$$E[X_1 + X_2] = E[X_1] + E[X_2].$$

If X_1 and X_2 are *independent*, i.e., if for every i and j the events $X \leq i$ and $Y \leq j$ are independent, then

$$E[X_1 \cdot X_2] = E[X_1] \cdot E[X_2].$$

Exercises.

(57) Let π be a permutation of $\{1,\dots,n\}$. An element $k \in \{1,\dots,k\}$ is called a *fixed point of* π if $\pi(k) = k$. What is the expected number of fixed points of π if π is drawn uniformly at random from all permutations on $\{1,\dots,n\}$?

(58) A *Hamiltonian path* in a directed graph is a directed path that visits every vertex exactly once. Show that for every $n \in \mathbb{N}$ there is a tournament with n vertices and at least $n!/2^{n-1}$ Hamiltonian paths.

(59) Show that every tournament contains a Hamiltonian path.

The use of expectation in existence proofs is based on the following *pigeonhole property* of expectation: a random variable cannot always be smaller (or always greater) than its expectation.

THEOREM 3.3.5 (Markov's Inequality). *Let X be a random variable such that $\Pr(X \geq 0) = 1$ and let $t \in \mathbb{R}$ be positive. Then*

$$\Pr(X \geq t) \leq \frac{E[X]}{t}.$$

PROOF.

$$\begin{aligned} E[X] &= \sum_x x \cdot \Pr(X = x) = \sum_{x \geq 0} x \cdot \Pr(X = x) \\ &\geq \sum_{x \geq t} t \cdot \Pr(X = x) = t \cdot \Pr(X \geq t) \end{aligned}$$ □

If X is integer, we even have the following consequence

$$\Pr(X > 0) \leq E[X]. \tag{33}$$

We refer to applications of (33) as the *first moment method*. Here is an example:

PROPOSITION 3.3.6. *Let $p(n) \ll 1/n$. Then $G(n,p)$ a.a.s. does not contain a K_3.*

PROOF. Let X be the number of triangles in $G(n,p)$. As we have seen above $E[X] \in o(1)$. Equation 33 implies that

$$\lim_{n\to\infty} \Pr[X = 0] = \lim_{n\to\infty} (1 - \Pr[X > 0]) = 1.$$ □

3.3.4. The second moment method. The second moment property provides a condition which implies that X "almost always" equals $E[X]$, i.e., the values of X are concentrated around its expectation.

Let X be a random variable. The *variance of X* is defined as

$$\mathrm{Var}[X] := E[(X - E[X])^2].$$

Note that

$$\mathrm{Var}[X] = E(X^2 - 2XE[X] + E[X]) = E[X^2] - E[X]^2.$$

THEOREM 3.3.7 (Chebyshev's Inequality). *Let X be a random variable. Then for any positive $t \in \mathbb{R}$*

$$\Pr(|X - E[X]| \geq t) \leq \frac{\mathrm{Var}[X]}{t^2}.$$

PROOF. Let Y be the random variable defined by $Y = (X - E[X])^2$ and apply Markov's inequality:

$$\Pr(|X - E[X]|) \geq t) \leq \Pr(Y \geq t^2) \leq \frac{E[Y]}{t^2} = \frac{\mathrm{Var}[X]}{t^2}$$ □

If $E[X] > 0$, then by setting $t = E[X]$, Chebychev's inequality implies

$$\Pr(X = 0) \leq \Pr\big(|X - E[X]| \geq E[X]\big) \leq \frac{\mathrm{Var}[X]}{E[X]^2}. \tag{34}$$

Applying (34) is often called the *second moment method*. Clearly, to apply (34) we need good tools to compute or bound the variance of a random variable X. If X is a sum of random variables, then the variance can be computed using the *covariance* (see formula 36 below), which is defined as

$$\mathrm{Cov}[X,Y] := E[(X - E[X])(Y - E[Y])].$$

The expression for the covariance can be simplified.

$$\begin{aligned}
\operatorname{Cov}[X,Y] &= E[(X-E[X])(Y-E[Y])] \\
&= E[XY - XE[Y] - E[X]Y + E[X]E[Y]] \\
&= E[XY] - E[XE[Y]] - E[E[X]Y] + E[E[X]E[Y]] \quad \text{(linearity of } E[.]) \\
&= E[XY] - E[X]E[Y] - E[X]E[Y] + E[X]E[Y] \qquad (E[\text{const}] = \text{const}) \\
&= E[XY] - E[X]E[Y] \qquad (35)
\end{aligned}$$

In particular, if X and Y are independent, then $\operatorname{Cov}[X,Y] = 0$, since we then have

$$\begin{aligned}
E[XY] &= \sum_i \big(i \cdot \Pr(XY = i)\big) \\
&= \sum_{k,l} \big(kl \cdot \Pr(X = k \text{ and } Y = l)\big) \\
&= \sum_{k,l} \big(kl \cdot \Pr(X = k)\Pr(Y = l)\big) \qquad (X \text{ and } Y \text{ are independent}) \\
&= E[X]E[Y].
\end{aligned}$$

Note that $\operatorname{Var}[X] = \operatorname{Cov}[X,X] = E[X^2] - E[X]^2$. Also note that

$$\begin{aligned}
\operatorname{Var}[X+Y] &= E[(X+Y-E[X+Y])^2] \\
&= E[((X-E[X]) + (Y-E[Y]))^2] \\
&= \operatorname{Cov}[X,X] + \operatorname{Cov}[Y,Y] + 2\operatorname{Cov}[X,Y].
\end{aligned}$$

More generally, if $X = \sum_{i\in\{1,\dots,n\}} X_i$ for indicator variables $X_1, \dots, X_n$ then we have the formula

$$\begin{aligned}
\operatorname{Var}[X] &= \sum_{i,j\in\{1,\dots,n\}} \operatorname{Cov}[X_i, X_j] \qquad (36) \\
&= \sum_{i=1}^{n} \operatorname{Var}[X_i] + \sum_{i\neq j} \operatorname{Cov}[X_i, X_j] \\
&\leq E[X] + \sum_{i\neq j} \operatorname{Cov}[X_i, X_j] \qquad (37)
\end{aligned}$$

where the inequality holds because $\operatorname{Var}[X_i] = E[X_i^2] - E[X_i]^2 \leq E[X_i]$ since X_i is an indicator variable and hence $X_i^2 = X_i$. Our first application of the second-moment method is the following.

PROPOSITION 3.3.8. *Let $p(n) \gg 1/n$. Then $G(n,p)$ a.a.s. contains a K_3.*

PROOF. For a subset S of three vertices, let $\mathcal{A}_S$ be the event that S induces a K_3 in $G(n,p)$, let X_S be the indicator variable of $\mathcal{A}_S$, and let $X := \sum_{S\in\binom{V}{3}} X_S$. Clearly, $\Pr(X_S = 1) = p^3$ and $E[X] \in O(n^3p^3)$.

Our goal is to show that $\Pr(X = 0) \in o(1)$. By the second moment method,

$$\Pr(X = 0) \leq \frac{\operatorname{Var}[X]}{E[X]^2}$$

so we only need a good upper bound for the variance. By (37), the variance can be bounded by

$$\operatorname{Var}[X] = \sum_S E[X_S] + \sum_{S\neq T} \operatorname{Cov}[X_S, X_T]. \qquad (38)$$

To estimate the second sum we analyse the contribution of a pair (S,T) of distinct subsets. If $\mathcal{A}_S$ and $\mathcal{A}_T$ are independent, then $\mathrm{Cov}[X_S, X_T] = 0$, and this pair contributes nothing. The events $\mathcal{A}_S$ and $\mathcal{A}_T$ are dependent if and only if S and T share common pairs of vertices, that is, if and only if $|S \cap T| = 2$. There are $O(n^4)$ pairs S, T with $|S \cap T| = 2$ and for each of these

$$\mathrm{Cov}[X_S, X_T] \leq E[X_S X_T] = p^5$$

because there are five edges that have to be present. So the total contribution of these pairs is in $O(n^4 p^5) \in o(1)$. Putting this together we obtain

$$\mathrm{Var}[X] \in O(n^3 p^3 + n^4 p^5) = O(n^3 p^3) \in o(E[X]^2).$$

Therefore, $\Pr(X = 0) \in o(1)$ and hence $G(n,p)$ a.a.s. contains a K_3. □

3.3.5. The void. Suppose that $p \ll n^{-2}$. Then asymptotically almost surely there are no edges.

3.3.6. The k-th day. When $p(n)$ reaches $\Theta(n^{-2})$ then edges appear. They form a matching until $p(n)$ reaches $\Theta(n^{-3/2})$. Let $1 \leq k \in \mathbb{N}$.

- When $p(n)$ reaches $\Theta(n^{-1-1/k})$ then a.a.s. trees on $k+1$ vertices appear.
- When $p(n)$ reaches $\Theta(n^{-1})$ then a.a.s. cycles appear.

We can zoom a bit further on the night between day number k and day number $k+1$. For the remainder of this section we assume that

$$n^{-\frac{k+1}{k}} = n^{-1-\frac{1}{k}} \ll p(n) \ll n^{-1-\frac{1}{k+1}} = n^{-\frac{k+2}{k+1}}.$$

LEMMA 3.3.9. *$G(n,p)$ has a.a.s. no components with $k+2$ (or more) vertices.*

PROOF. There are $O(n^{k+2})$ choices of $k+2$ vertices and $O(1)$ choices of a tree on those vertices. With probability p^{k+1} we have those edges. So the expected number of such trees is $O(n^{k+2}p^{k+1})$ which is in $o(1)$ since $p \in o(n^{-\frac{k+2}{k+1}})$ (check!). Again, the statement follows from the first moment method. □

LEMMA 3.3.10. *$G(n,p)$ has a.a.s. no cycles.*

PROOF. From the previous lemma it follows that $G(n,p)$ has a.a.s. no cycles with more than $k+2$ vertices. There are $O(n^l)$ choices of $l \leq k+2$ vertices and $O(1)$ choices for a cycle, which appears with probability p^l. So the expected number of l-cycles is $O((np)^l)$ which is in $o(1)$ as $p \ll n^{-1}$. The statement follows from the first moment method. □

The first moment method is not sufficient to prove the following; here we need the second moment method.

LEMMA 3.3.11. *For every r, every tree T on at most $k+1$ vertices appears a.a.s. at least r times as a component of $G(n,p)$.*

PROOF. There are $\Theta(n^r)$ choices of r vertices, at least one way of placing a given tree T on those vertices and probability p^{r-1} of having the tree edges. The probability that there are no further edges involving those r vertices is bounded by $(1-p)^{r(n-1)} \leq e^{-rp(n-1)}$ which is asymptotically one because $p(n) \ll n^{-1}$. So the expected number of tree components T is in $\Theta(n^r p^{r-1})$. One can use the second-moment method to prove that almost surely there are at least r components T for any fixed r.

For $S \in \binom{V}{k+1}$, let X_S the the indicator random variable that S induces the tree T in $G(n,p)$. Then the total number of trees is $X := \sum_{S \in \binom{V}{k+1}} X_S$ and we need to bound its variance. We have

$$\mathrm{Var}[X] = \sum_{S \in \binom{V}{k+1}} \sum_{S' \subseteq S} \sum_{T \in V \setminus S k+1-|S'|} \mathrm{Cov}[X_S, X_{S' \cup T}]$$

We have $\mathrm{Cov}[X_S, X_{S' \cup T}] = 0$ if $S' \leq 1$. If $S' > 1$, then

$$\mathrm{Cov}[X_S, X_{S' \cup T}] \leq E[X_S S_{S' \cup T}] \leq p^{2k-(|S'|-1)}.$$

Hence,

$$\mathrm{Var}[X] \in O\left(\sum_{l>1} n^{2k+2-\ell} p^{2k+1-\ell}\right) \subseteq o\left((E[X]+r)^2\right).$$

By Chebychev's Inequality (Theorem 3.3.7) we have

$$\Pr(X < r) \leq \frac{\mathrm{Var}[X]}{(E[X]+r)^2} \in o(1). \qquad \square$$

3.3.7. Day ω. Suppose now that $p(n) \gg n^{-1-\epsilon}$ for every $\epsilon > 0$ but $p(n) \ll n^{-1}$. This includes functions such as $p(n) = \frac{1}{n \log n}$. For such functions, the following properties hold a.a.s.:

(1) there are no cycles.
(2) for every r, every finite tree T occurs at least r times as a component.

The proofs are similar to the ones for day k.

Exercises.

(60) Show that if $p(n) \ll n^{-1}$ then $G(n,p)$ is a.a.s. not connected.

3.3.8. The double jump. The random graph undergoes a critical transition at $p(n) \in \Theta(n^{-1})$. We do not prove the claims in this section since they require methods that are out of the scope of this course (see [**22**]); however, we want to state the result since it is an important part of the overall picture of the Erdös-Rényi evolution. Suppose that $p(n) = c/n$.

- $c < 1$: the largest component has $\log n$ vertices.
- $c = 1$: the size of the largest component jumps to approximately $n^{2/3}$.
- $c > 1$: the size of the largest component approaches n.

Thus, $p = 1/n$ is the threshold for a dramatic double-jump in the size of the largest component.

3.3.9. Past the double jump. In this section we assume

$$\frac{1}{n} \ll p(n) \ll \frac{\ln n}{n}.$$

At this stage cycles have appeared, but still small subgraphs only have at most one cycle. The following properties hold a.a.s:

(1) For every k there are no k vertices with at least $k+1$ edges.
(2) For every r and every $k > 3$ there are at least r cycles of size k.
(3) For every s, d and every $k \geq 3$ there does not exist a cycle of size k and a vertex of degree d at distance s from the cycle.
(4) For every r and every finite tree T there are at least r components isomorphic to T.

The first property can be proved by the first moment method, similarly as in 3.3.10 but using that $p(n) \ll \frac{\ln n}{n}$. The second property can be shown by the second moment method, similarly as in Proposition 3.3.8, using that $1/n \ll p(n)$. To see the third property, set $u := k+s+d-1$. There are $O(n^u)$ choices for selecting u vertices, $O(1)$ ways to place a k-cycle, a path of length s from it to a vertex v, and $d-1$ further neighbours of v. These edges are present with probability p^u. The probability that v is not adjacent to any other vertex is $(1-p)^{n-1-d}$. The expected number of such configurations is

$$\begin{aligned} n^u p^u (1-p)^{n-1-d} &\leq (np)^u e^{-p(n-1-d)} && \text{(using (31))} \\ &\in O((np)^u e^{-np}) \subseteq o(1) && \text{(since } np \to \infty\text{).} \end{aligned}$$

The fourth property can be shown as in Lemma 3.3.11.

3.3.10. Connectivity. In this section we prove that $\frac{\ln n}{n}$ is a threshold function for the connectivity of $G(n,p)$. This happens to be also a threshold function for loosing all isolated vertices. This fact play an important role in the proof. The intuition is that large components are more likely to merge than smaller ones. Let X_k be the random variable for the number of connected components of size exactly k. So X_1 is the number of isolated vertices.

THEOREM 3.3.12. *$\frac{\ln n}{n}$ is a threshold for the existence of isolated vertices in $G(n,p)$.*

PROOF. It will be convenient to prove something stronger. Instead of considering the two cases that $p \ll \frac{\ln n}{n}$ and that $\frac{\ln n}{n} \ll p$, we write $p = \frac{\ln n + c(n)}{n}$ and distinguish the cases that $c(n) \to \infty$ and $c(n) \to -\infty$. In the first case, we prove that $G(n,p)$ has a.a.s. no isolated vertices, and in the second case we prove that a.a.s. it has. Since having no isolated vertices is a property that is stable under the addition of edges, we may assume that $|c(n)| \ll \ln n$.

Let Z_i be the indicator variable for $i \in \{1,\dots,n\}$ being isolated. Then we may write the random variable for the number of isolated vertices as $X_1 = \sum_{i\in\{1,\dots,n\}} Z_i$ and by the linearity of expectation we have

$$\begin{aligned} E[X_1] = \sum_{i\in\{1,\dots,n\}} E[Z_i] &= n(1-p)^{n-1} \\ &= \exp(\ln n + n\ln(1-p) - \ln(1-p)) \\ &= \exp(\ln n - np + p + O(np^2)) && \text{(using (32))} \\ &\in \exp(-c(n) + p + O(np^2)) \\ &\subseteq (1+o(1))\exp(-c(n)) && \text{(since } |c(n)| \ll \log n\text{).} \end{aligned}$$

It follows that if $c(n) \to \infty$ then $E[X_1] \to 0$, and by the first moment method we obtain that in this case $G(n,p)$ a.a.s. does not contain isolated vertices.

If $c(n) \to -\infty$ then $E[X_1] \to \infty$. To prove the existence of isolated vertices we need to apply the second moment method. For $i \neq j$, the random variables Z_i and Z_j are not independent, but we can compute the covariance using (35):

$$\begin{aligned} \operatorname{Cov}[Z_i, Z_j] &= \Pr(Z_i = Z_j = 1) - \Pr(Z_i = 1)\Pr(Z_j = 1) \\ &= (1-p)^{2n-3} - (1-p)^{2n-2} \\ &= p(1-p)^{2n-3}. \end{aligned}$$

We obtain

$$\begin{aligned}\frac{\operatorname{Var}[X_1]}{E[X_1]^2} &\leq \frac{E[X_1]}{E[X_1]^2} + \frac{n(n-1)p(1-p)^{2n-3}}{2n^2(1-p)^{2(n-1)}} \\ &\leq \frac{1}{E[X_1]} + \frac{p}{1-p} \in o(1)\end{aligned}$$

since p tends to 0 and $E[X_1]$ tends to ∞. By (34) it follows that $X_1 > 0$ a.a.s. □

Clearly, if a graph has isolated vertices, then it is not connected. In Lemma 3.3.14 below we show that a.a.s. $G(n, \frac{\ln n}{n})$ has no connected components of 'intermediate' size. Again, a first moment argument suffices.

LEMMA 3.3.13. $E(X_k) \leq \binom{n}{k} k^{k-2} p^{k-1} (1-p)^{k(n-k)}$.

PROOF SKETCH. Use the fact that there are k^{k-2} trees with vertex set $\{1, \ldots, k\}$ (Theorem 5.8.3) and the union bound. □

LEMMA 3.3.14. *If $p = \log n/n$ then*

$$\sum_{k=2}^{\lfloor n/2 \rfloor} \Pr(X_k > 0) \in o(1).$$

PROOF. Using Markov's inequality we get

$$\begin{aligned}\sum_{k=2}^{\lfloor n/2 \rfloor} \Pr(X_k > 0) &\leq \sum_{k=2}^{\lfloor n/2 \rfloor} E[X_k] && \text{(Theorem 3.3.5)} \\ &\leq \sum_{k=2}^{\lfloor n/2 \rfloor} \binom{n}{k} k^{k-2} p^{k-1} (1-p)^{k(n-k)} && \text{(by Lemma 3.3.13).} \quad (39)\end{aligned}$$

With the bounds $\binom{n}{k} \leq (\frac{en}{k})^k$ (see Example 25) and $1 - p \leq e^{-p}$ (see (31)) we obtain

$$\begin{aligned}\binom{n}{k} k^{k-2} p^{k-1} (1-p)^{k(n-k)} &\leq \left(\frac{en}{k}\right)^k k^{k-2} \left(\frac{\ln n}{n}\right)^{k-1} e^{-pk(n-k)} \\ &\leq n e^k (\ln n)^{k-1} e^{k^2 \ln n/n - k \ln n} \\ &= \exp(\ln n + k + (k-1) \ln \ln n + k^2 \ln n/n - k \ln n) \\ &\in O(\exp((\ln \ln n - \ln n)(k-1)) \\ &\subseteq O\left(\left(\frac{\ln n}{n}\right)^{k-1}\right) \subseteq o(1).\end{aligned}$$

This expression tends to 0 even if we sum over finitely many k. But the bound is not strong enough to show that (39) tends to 0. However, for $k \leq n/2$ we may also obtain another bound. Observe that $k \leq n/2$ implies $k(n-k) = kn - k^2 \geq kn/2$. We obtain that

$$\begin{aligned}\binom{n}{k} k^{k-2} p^{k-1} (1-p)^{k(n-k)} &\leq n e^k (\ln n)^{k-1} e^{-pkn/2} \\ &\in O\left(n \left(\frac{\ln n}{\sqrt{n}}\right)^k\right) \subseteq O\left(n \left(\frac{\ln n}{\sqrt{n}}\right)^5\right).\end{aligned}$$

The advantage of this bound is that it tends to 0 even if we multiply it with n, and hence can be used to bound (39):

$$\sum_{k=2}^{\lfloor n/2\rfloor} \Pr(X_k > 0) = \sum_{k=2}^{4} \Pr(X_k > 0) + \sum_{k=5}^{\lfloor n/2\rfloor} \Pr(X_k > 0)$$
$$\in O\left(\left(\frac{\ln n}{n}\right)^{k-1} + n^2\left(\frac{\ln n}{\sqrt{n}}\right)^5\right) \subseteq o(1). \qquad \square$$

THEOREM 3.3.15. *$\frac{\ln n}{n}$ is a threshold for connectivity of $G(n,p)$.*

PROOF. If $p \ll \ln n/n$ then a.a.s. $G(n,p)$ has isolated vertices by Theorem 3.3.12 and hence is not connected.

Now suppose that $\ln n/n \ll p$. Then a.a.s. there are no isolated vertices by Theorem 3.3.12. With Lemma 3.3.14 we obtain that $\sum_{k=1}^{\lfloor n/2\rfloor} \Pr(X_k > 0) \in o(1)$ and hence $G(n,p)$ is a.a.s. connected. $\square$

3.3.11. Beyond connectivity. Assume that $p(n)$ is such that for every $\epsilon > 0$

$$\frac{\ln n}{n} \ll p(n) \ll n^{-1+\epsilon}.$$

Then the following holds:

(1) For every k, there are a.a.s. no k vertices with at least $k+1$ edges.
(2) For every r and every $k \geq 3$ there exist a.a.s. (at least) r cycles of size k.
(3) For every d a.a.s. all vertices have at least d neighbours (in particular, there are no trees left).

Again, the first property can be shown with the first moment method and the second property with the second moment method. For the third property, the expected number of vertices of degree precisely i is

$$n\binom{n-1}{i}p^i(1-p)^{n-1-i} \leq n(np)^i e^{-np} \qquad \text{(using (31))}.$$

With $np \gg \ln n$ the term e^{-np} dominates and it follows that the expected number is in $o(1)$.

Exercises.

(61) Prove Claim (1) in Section 3.3.11.
(62) Prove Claim (3) in Section 3.3.11.
(63) For $c < 1$, give an algorithm that (always) computes the chromatic number of $G(n, c/n)$ and has an expected polynomial running time.[2]
(64) Prove that if a graph has for all k no k-element subgraph with at least $k+1$ edges, then every component of the graph has at most one cycle.
(65) Please argue why the following statements are *not* in contradiction:
- $G(n,p)$ is a.a.s. connected, and
- $G(n,p)$ contains a.a.s at least two cycles, and
- For every connected graph H with at least two cycles, $G(n,p)$ does not contain H as a subgraph a.a.s.

[2]This has been pushed much further; see [**8**].

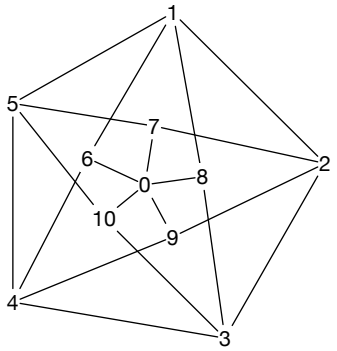

FIGURE 3.1. The Grötzsch graph: a triangle-free graph which is not 3-colorable.

3.3.12. Powers of n. So far, all properties of the random graph were a.a.s. either true or false. When the edge probability is a rational power of n, for instance $p = n^{-1/3}$, then there are very natural properties of $G(n,p)$ whose probability does neither tend to 0 nor to 1 as n tends to infinity. Take a graph H with n vertices and m edges such that $n/m = 1/3$, for example, take $H = K_7$. The number X of copies of H in $G(n,p)$ has the expected value

$$\binom{n}{7} p^{21} = \frac{n(n-1)\cdots(n-6)}{7!} n^{-7}$$

which tends to $1/7!$ as n tends to infinity. It can be shown using the second moment method that $\Pr(X = 0)$ tends to $e^{-1/7!}$, which is certainly neither 0 nor 1. Irrational powers behave better in this respect; we again refer to [**40**].

3.4. High Girth and High Chromatic Number

Let $G = (V, E)$ be a graph. The *girth* of G is the size of its shortest cycle. We write

- $\alpha(G)$ for the size of the largest stable set in G;
- $\chi(G)$ for the *chromatic number of* G, i.e., the smallest number k such that G is k-colourable.

The smallest graph of girth 4 with chromatic number 4 is shown in Figure 3.1.

Let $c\colon V \to \{1,\dots,k\}$ be a k-colouring of G. Since the pre-image of a colour under c is a stable set, we have that $k \cdot \alpha(G) \geq |V|$. It follows that $\chi(G) \geq \frac{|V|}{\alpha(G)}$. Therefore, for a fixed number of vertices we can obtain a lower bound for $\chi(G)$ by obtaining an upper bound for $\alpha(G)$.

The probability that $S \subseteq V = \{1,\dots,n\}$ is a stable set in $G(n,p)$ is $(1-p)^{\binom{|S|}{2}}$. The probability that G has a stable set of size x is at most

$$\binom{n}{x} (1-p)^{\binom{x}{2}}.$$

THEOREM 3.4.1 (Erdős). *For all $k, l \in \mathbb{N}$ there exists a graph with girth at least ℓ and $\chi(G) \geq k$.*

PROOF. Fix $0 < \epsilon < 1/\ell$ and let G be drawn from $G(n,p)$ with $p = n^{-1+\epsilon}$. Let X be the number of cycles of size at most ℓ. Then

$$\begin{aligned} E[X] &= \sum_{i=3}^{\ell} \frac{n(n-1)\cdots(n-i+1)}{2i} p^i \\ &\leq \sum_{i=3}^{\ell} \frac{n^{\epsilon i}}{2i} \in o(n) \qquad \text{(since } \epsilon\ell < 1\text{)}. \end{aligned}$$

By Markov's inequality (Theorem 3.3.5)

$$\Pr(X \geq n/2) \in o(1). \tag{40}$$

The probability to have a stable set of size at least x can be bounded as follows.

$$\begin{aligned} \Pr(\alpha(G) \geq x) \leq \binom{n}{x}(1-p)^{\binom{x}{2}} &\leq n^x e^{-p\frac{x-1}{2}x} \qquad \text{(using (31))} \\ &\leq \left(ne^{-p\frac{x-1}{2}}\right)^x. \end{aligned}$$

Hence, setting $x := \lceil \frac{3}{p} \ln n \rceil$ we obtain

$$\Pr(\alpha(G) \geq x) \in o(1). \tag{41}$$

Let n be sufficiently large so that both events (40) and (41) have probability less than $1/2$. Then there exists a specific G with less than $n/2$ cycles of length at most ℓ and with $\alpha(G) < x \leq \lceil 3n^{1-\epsilon} \ln n \rceil$. Remove from G a vertex from each cycle of length at most ℓ. This gives a graph G^* with at least $n/2$ vertices, girth greater than ℓ, and $\alpha(G^*) \leq \alpha(G)$. Thus

$$\chi(G^*) \geq \frac{|V(G^*)|}{\alpha(G^*)} \geq \frac{n/2}{\lceil 3n^{1-\epsilon} \ln n \rceil} \geq \frac{n^\epsilon}{6 \ln n}.$$

To complete the proof, let n be sufficiently large so that this value is at least k. □

REMARK 3.4.2. Lovász was the first who gave deterministic constructions of graphs with arbitrarily large girth and chromatic number; another construction was given by Nešetřil and Rödl [**31**]. A recent reference is [**2**]. A powerful generalisation of Theorem 3.4.1 in the world of graph homomorphisms can be found in [**32**], and is called the *sparse incomparability lemma*; a further generalisation concerns homomorphism problems between general relational structures, also known as constraint satisfaction problems (see Theorem 5 in [**14**]).

3.5. Extremal Graph Theory

In the previous section, we have seen an application of the probabilistic method to prove that certain finite graphs *exist*. The probabilistic method can also be applied to show that *every* graph must have certain properties. We present an example of such an application for proving a complement version of *Turan's theorem*, which is one of the cornerstones of a research area called *extremal graph theory*. We will deduce this theorem from another theorem which provides a lower bound on the largest stable set in a graph in terms of the degree sequence of the graph.

THEOREM 3.5.1. *Let $G = (\{1,\ldots,n\}, E)$ be a graph and for $i \in V(G)$ let d_i denote the degree of i in G. Then*

$$\alpha(G) \geq \sum_{i=1}^{n} \frac{1}{d_i + 1}.$$

EXAMPLE 6. For example, if G is 1-regular (i.e., a perfect matching) then $\alpha(G) \geq \frac{n}{2}$, if it is 2-regular (i.e., a disjoint union of cycles) then $\alpha(G) \geq \frac{n}{3}$, and more generally if G is k-regular then $\alpha(G) \geq \frac{n}{k+1}$. Also note that all these bounds are tight, as demonstrated by a disjoint union of $(k+1)$-cliques. △

PROOF. Let $\pi\colon V(G) \to V(G)$ be a random permutation chosen from the uniform distribution, i.e., each permutation is drawn with probability $1/n!$. For $i \in V(G)$, let A_i be the event that $\pi(i) < \pi(j)$ for all d_i neighbours of i. There are $\binom{n}{d_i+1}$ possibilities to choose a set $S \subseteq V(G)$ with $d_i + 1$ elements that are the π-images of i and all its d_i neighbours. Moreover, there are $d_i!$ many possibilities to arrange the images of the neighbours of i under π, and $(n - d_i - 1)!$ possibilities to arrange the vertices outside S. Thus,

$$\Pr(A_i) = \binom{n}{d_i+1} \frac{d_i!(n-d_i-1)!}{n!} = \frac{1}{d_i+1}.$$

Let U be the set of all vertices i such that A_i holds. By linearity of expectation,

$$E[|U|] = \sum_{i=1}^{n} \Pr(A_i) = \sum_{i=1}^{n} \frac{1}{d_i+1}.$$

Thus, for some permutation π we have $|U| \geq \sum_{i=1}^{n} \frac{1}{d_i+1}$. Finally, note that U is a stable set: Let $\{i,j\} \in E$. If $\pi(i) < \pi(j)$ then $j \notin U$, and if $\pi(j) < \pi(i)$ then $i \notin U$. □

THEOREM 3.5.2 (Turan's Theorem, complement version). *Let G be a graph with n vertices and at most $nk/2$ edges. Then $\alpha(G) \geq \frac{n}{k+1}$.*

EXAMPLE 7. Again, the bound in theorem 3.5.3 is tight: if $(k+1)|n$ then the disjoint union G of $\frac{n}{k+1}$ many $(k+1)$-cliques has n vertices, $\alpha(G) = \frac{n}{k+1}$, and the number of edges is

$$\frac{n}{k+1}(k+1)k/2 = nk/2.$$ △

PROOF. Clearly, we may assume that the number of edges is exactly $nk/2$. The intuition for using Theorem 3.5.1 is that the sum $\sum_{i=1}^{n} \frac{1}{d_i+1}$ is minimised if the d_i's are equal. Formally, we apply the Cauchy-Schwarz inequality (Lemma A.2.2) setting $x_i := \sqrt{d_i+1}$ and $y_i := 1/x_i$ and obtain

$$\begin{aligned} n^2 &= \left(\sum_{i=1}^{n} x_i y_i\right)^2 \\ &\leq \left(\sum_{i=1}^{n} x_i^2\right)\left(\sum_{i=1}^{n} y_i^2\right) = \left(\sum_{i=1}^{n} (d_i+1)\right)\left(\sum_{i=1}^{n} \frac{1}{d_i+1}\right). \end{aligned}$$

Since $\sum_{i=1}^{n} d_i$ equals twice the total number of edges (Exercise 1),

$$\sum_{i=1}^{n} \frac{1}{d_i+1} \geq n^2/(n+nk) = n/(k+1).$$

We then conclude

$$\begin{aligned} \alpha(G) &\geq \sum_{i=1}^{n} \frac{1}{d_i+1} && \text{Theorem 3.5.1} \\ &\geq \frac{n}{k+1}. \end{aligned}$$ □

We state without proof the stronger ('primal') version of Turan's theorem.

THEOREM 3.5.3 (Turan's theorem). *Let G be a graph with n vertices and more than $(1-\frac{1}{k})n^2/2$ edges. Then G must contain an $(k+1)$-clique.*

Exercises.

(66) Construct a graph that shows that Theorem 3.5.3 is tight. I.e., construct a graph without $(k+1)$-cliques and $(1-\frac{1}{k})n^2/2$ edges. First treat the case that $(k+1)|n$, then the general case.
(67) Derive Theorem 3.5.2 from Theorem 3.5.3.
(68) Derive Mantel's theorem: if a graph on $2n$ vertices contains n^2+1 edges, then G contains a triangle.
(69) Let $G=(V,E)$ be a graph with $|V|=10$ such that for any $S \in \binom{V}{3}$ we have $E \cap \binom{S}{2} \neq \emptyset$. What is the best lower bound that you can prove on $|E|$?
(70) Can you also derive Theorem 3.5.3 from Theorem 3.5.2?

CHAPTER 4

Ramsey Theory

Ramsey theory seeks regularity within disorder: general conditions for the existence of substructures with regular properties. Many questions in Ramsey theory are of the form: "how many elements must a structure have to guarantee that a particular property holds?"

4.1. The Pigeonhole Principle

If n pigeons fly to fewer than n holes, there must be one hole that got more than one pigeon. There is an important infinite version of the statement: if infinitely many pigeons fly to finitely many holes, one hole must have gotten infinitely many pigeons. This triviality will be used (and later greatly generalised) in the next tools that we present. We mention that the first moment method can be seen as a probabilistic version of this principle: if the pigeons fly randomly to the holes, then not all holes can have more pigeons than the expected number of pigeons for that hole.

Exercises.

(71) Prove that every finite graph with at least two vertices contains two vertices of the same degree.

4.2. Kőnig's Tree Lemma

Kőnig's tree lemma is a simple but important tool to translate between statements in finite combinatorics and statements in infinite combinatorics. Often, a nice feature of the statements in infinite combinatorics is that they may have fewer quantifiers than their finite counterparts; so sometimes it is easier to prove a statement in infinite combinatorics and to derive the finite counterpart via Kőnig's lemma (we see such an application in Section 4.3).

LEMMA 4.2.1 (Kőnig's Tree Lemma). *Let (V, E) be a tree such that V is infinite and every vertex in V has finite degree. Then (V, E) contains an infinite path.*

PROOF. Arbitrarily choose $v_0 \in V$. Since the degree of v_0 is finite, there exists a neighbour v_1 of v_0 such that the connected component of v_1 in $(V, E)-v_0$ is infinite (by the infinite pigeonhole principle). We construct the infinitely long path by induction. Suppose we have already found a path $v_0, v_1, \ldots, v_i$ in (V, E) such that the connected component T of v_i in $(V, E) - \{v_0, v_1, \ldots, v_{i-1}\}$ is infinite. Note that T is a tree, and since the degree of v_i is finite, v_i must have a neighbour v_{i+1} in T which lies in an infinite connected component of $T - v_i$. In this way, we define an infinitely long path $v_0, v_1, v_2, \ldots$ in (V, E). □

Note that even if a tree contains arbitrarily long finite paths, it might not contain infinitely long paths, as can be seen from Figure 4.1. The following proposition illustrates one of the many uses of Kőnig's tree lemma.

PROPOSITION 4.2.2. *A countably infinite graph G is k-colourable if and only if every finite subgraph of G is k-colourable.*

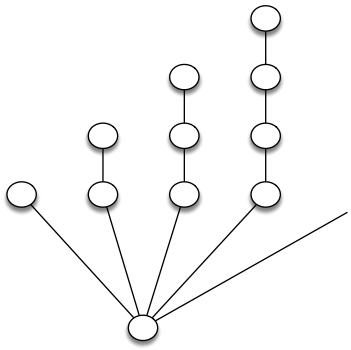

FIGURE 4.1. A tree with arbitrarily long paths, but no infinite paths.

PROOF. To prove the non-trivial direction of the statement we assume that every finite subgraph of G is k-colourable. Let $u_1, u_2, \dots$ be an enumeration of $V(G)$. For $n \in \mathbb{N}$, let X_n be the set of all proper k-colourings of the subgraph induced by $\{u_1, \dots, u_n\}$ in G. We consider the following tree with vertex set $X := \bigcup_{n\in\mathbb{N}} X_n$. We connect $x, y \in X$ if y is the extension of x by one element. Clearly, this defines a tree which has infinitely many vertices by assumption. Moreover, every vertex of this graph has finite degree since there are only finitely many maps $\{u_1, \dots, u_{n+1}\} \to [k]$. By Kőnig's tree lemma there exists an infinite path $v_0, v_1, v_1, \dots$ with $v_i \in X_i$ for all $i \in \mathbb{N}$. Define $f\colon V(G) \to [k]$ by $f(u_i) := v_i(u_i)$ for all $i \geq 1$. Then f is a k-colouring of G. □

We mention that the proposition also holds for graphs with uncountably many vertices, but then Kőnig's tree lemma is not enough for the proof. Sometimes, proofs as the proof of the proposition above are called *compactness arguments*, and indeed there is an associated topology which is compact. We refer to courses on general topology or model theory for more about this topic.

4.3. Ramsey's Theorem

The set $\{0, \dots, n-1\}$ is sometimes denoted by $[n]$. We refer to mappings $\chi\colon \binom{M}{s} \to [c]$ as a *colouring* of $\binom{M}{s}$ (with the *colours* $[c]$). In Ramsey theory, one writes

$$L \to (m)^s_c$$

if for every $\chi\colon \binom{L}{s} \to [c]$ there exists an $M \in \binom{L}{m}$ which is χ-*monochromatic*, i.e., χ is constant on $\binom{M}{s}$. Note the following.

- For all $c \in \mathbb{N}$ we have $[c+1] \to (2)^1_c$: this is the pigeonhole principle.
- For all $c \in \mathbb{N}$ we have $\mathbb{N} \to (\mathbb{N})^1_c$: this is the infinite pigeonhole principle.

THEOREM 4.3.1 (Ramsey's theorem). *For all $c, m, s \in \mathbb{N}$ there is an $l \in \mathbb{N}$ such that $[l] \to (m)^s_c$.*

For $c, m, s \in \mathbb{N}$, the smallest l such that $[l] \to (m)^s_c$ is denoted by $R^s_c(m)$. We first prove a variant of Ramsey's theorem; we write $\aleph_0$ for the cardinality of $\mathbb{N}$.

THEOREM 4.3.2. $\mathbb{N} \to (\aleph_0)^2_2$.

This statement has the following interpretation in terms of undirected graphs: every countably infinite undirected graph either contains an infinite *clique* (a complete subgraph) or an infinite independent set.

PROOF. Let $\chi\colon \binom{\mathbb{N}}{2} \to [2]$ be a 2-colouring of $\binom{\mathbb{N}}{2}$. We define an infinite sequence $x_0, x_1, \dots$ of numbers from $\mathbb{N}$ and an infinite sequence $V_0 \supseteq V_1 \supseteq \cdots$ of infinite subsets of $\mathbb{N}$. Start with $V_0 := \mathbb{N}$ and $x_0 = 0$. By the infinite pigeonhole principle, there is a $c_0 \in [2]$ such that $\{v \in V_0 \mid \chi(x_0, v) = c_0\} =: V_1$ is infinite. We now repeat this procedure with any $x_1 \in V_1$ and V_1 instead of V_0. Continuing like this, we obtain sequences $(c_i)_{i\in\mathbb{N}}$, $(x_i)_{i\in\mathbb{N}}$, $(V_i)_{i\in\mathbb{N}}$.

Again by the infinite pigeonhole principle, there exists $c \in [2]$ such that $c_i = c$ for infinitely many $i \in \mathbb{N}$. Then $P := \{x_i \mid c_i = c\}$ has the desired property. To see this, let $i < j$ be such that $x_i, x_j \in P$. Then $\chi(\{x_i, x_j\}) = c_i = c$. □

We now state the infinite version of Ramsey's theorem; the proof is similar to the proof of Theorem 4.3.2 shown above.

THEOREM 4.3.3 (Ramsey's theorem). *Let $s, c \in \mathbb{N}$. Then $\mathbb{N} \to (\aleph_0)^s_c$.*

A proof of Theorem 4.3.3 can be found in [**21**] (Theorem 5.6.1); for a broader introduction to Ramsey theory see [**18**]. Ramsey's theorem is a consequence of the infinite version via Kőnig's tree lemma, similarly as in Proposition 4.2.2.

PROOF OF THEOREM 4.3.1. Let $(*)_{l,\chi}$ be the property that for all m-subsets M of $[l]$ the mapping χ is not constant on $\binom{M}{s}$. Our proof is by contradiction: suppose that there are positive integers c, m, s such that for all $l \in \mathbb{N}$ there is a $\chi\colon \binom{[l]}{s} \to [c]$ such that $(*)_{l,\chi}$.

We construct a tree as follows. The vertices are the maps $\chi\colon \binom{[l]}{s} \to [c]$ that satisfy $(*)_{l,\chi}$. We make the vertex $\chi\colon \binom{[l]}{s} \to [c]$ adjacent to $\chi'\colon \binom{[l+1]}{s} \to [c]$ if χ is a restriction of χ'. Clearly, every vertex in the tree has finite degree. By assumption, there are arbitrarily long paths that start in the vertex χ_0 where χ_0 is the map with the empty domain. By Lemma 4.2.1, the tree contains an infinite path $\chi_0, \chi_1, \dots$ We use this to define a map $\chi_{\mathbb{N}}\colon \binom{\mathbb{N}}{s} \to [c]$ as follows. For every $x \in \mathbb{N}$, there exists a $c_0 \in [c]$ and an $i_0 \in \mathbb{N}$ such that $\chi_i(x) = c_0$ for all $i \geq i_0$. Define $\chi_{\mathbb{N}}(x) := c_0$. Then $\chi_{\mathbb{N}}$ satisfies $(*)_{\mathbb{N},\chi}$, a contradiction to Theorem 4.3.3. □

Exercises.

(72) Let $R(k, \ell)$ denote the smallest number n such that if we colour the edges of K_n in red and blue, the resulting edge-coloured graph always contains a red K_k or a blue K_ℓ. Show that

$$R(k, \ell) \leq R(k, \ell - 1) + R(k - 1, \ell)$$

and use $R(\ell, 2) = R(2, \ell)$ to prove that

$$R(k, \ell) \leq \binom{k + \ell - 1}{k - 1}.$$

4.4. A Probabilistic Lower Bound

The proof of Ramsey's theorem in Section 4.3 was non-constructive: it derived the statement from the infinite version via an application of Kőnig's tree lemma; in particular, we did not learn anything about the size of $R^s_c(m)$, the *minimal* $l \in \mathbb{N}$ such that $[l] \to (m)^s_c$. We now present a probabilistic proof of a lower bound (for simplicity again only for $c = s = 2$; the general case is similar).

THEOREM 4.4.1. *For any $m \geq 3$ we have $R^2_2(m) > \lfloor 2^{m/2} \rfloor$.*

PROOF. We claim that if $\ell := \lfloor 2^{m/2} \rfloor$ then with positive probability $G(\ell, 1/2)$ has no clique of size m and no stable set of size m. For $S \in \binom{V}{m}$, let $\mathcal{A}_S$ be the event that S forms a clique or a stable set in $G(\ell, 1/2)$. We have $\Pr(\mathcal{A}_S) = 2(\frac{1}{2})^{\binom{m}{2}} = 2^{1-\binom{m}{2}}$. Since there are $\binom{\ell}{m}$ possible choices for S, the probability that at least one of the events $\mathcal{A}_S$ occurs is at most

$$\begin{aligned}\binom{\ell}{m} 2^{1-\binom{m}{2}} &< \frac{\ell^m}{m!} 2^{1-\frac{m^2}{2}+\frac{m}{2}} \\ &= \frac{2^{1+\frac{m}{2}}}{m!} \cdot \frac{\ell^m}{2^{m^2/2}} < 1 \qquad \text{(since } \ell = \lfloor 2^{m/2} \rfloor\text{).}\end{aligned}$$

□

4.5. Applications

Ramsey theory has numerous applications in many fields, e.g. in topological dynamics [**24**], set theory, model theory [**21**], just to name a few. Further applications in number theory, harmonic analysis, geometry, and theoretical computer science are surveyed in [**35**]. Here, we only give some of the very basic applications of Ramsey's theorem.

4.5.1. Number Theory. We start with a result motivated by number theory which in fact pre-dates Ramsey's theorem.

THEOREM 4.5.1 (Schur's theorem, 1917). *Let c be a positive integer. Then there exists an $s = s(c) \in \mathbb{N}$ such that for every colouring $f\colon \{1, \dots, s\} \to [c]$ there exist $x, y, z \in \{1, \dots, s\}$ such that*

- *$x + y = z$;*
- *x, y, z is monochromatic (i.e., f is constant on x, y, z).*

EXAMPLE 8. For the map $f\colon \{1, 2, 3, 4\} \to \{0, 1\}$ given by $f(1) = f(4) = 0$ and $f(2) = f(3) = 1$ there are no elements $x, y, z \in \{1, \dots, 4\}$ such that $f(x) = f(y) = f(z)$ and $x + y = z$.

Let $f\colon \{1, \dots, 5\} \to \{0, 1\}$. Suppose for contradiction that there are no $x, y, z \in \{1, \dots, 5\}$ such that $f(x) = f(y) = f(z)$ and $x + y = z$. Without loss of generality we may suppose that $f(1) = 0$. Since $1 + 1 = 2$ we must have $f(2) = 1$. Since $2 + 2 = 4$ we must have $f(4) = 0$. Since $1 + 3 = 4$ and $f(1) = 0$ we must have $f(3) = 1$. Since $3 + 2 = 5$ and $f(2) = 1$ we must have $f(5) = 0$. But $4 + 1 = 5$ and $f(4) = f(1) = 0$ implies that $f(5) = 1$, a contradiction. △

PROOF. Let $s = s(c)$ be $R^2_{c-1}(3)$. Let $f\colon \{1, \dots, s\} \to [c]$. From f, we define a $(c-1)$-colouring h of the edges of K_s:

$$h(\{a, b\}) := f(|a - b|) \in \{1, \dots, c - 1\}.$$

By the choice of s, there is an h-monochromatic 3-element subset $\{u, v, w\}$ of $V(K_s)$. Without loss of generality, $u > v > w$. We claim that $x := u - v$, $y := v - w$, and $z := u - w$ gives the desired subset of $\{1, \dots, s\}$:

- $f(x) = h(\{u, v\}) = h(\{v, w\}) = f(y)$, since $\{u, v, w\}$ is h-monochromatic;
- likewise, $f(x) = f(z)$;
- $x + y = u - w = z$. □

The famous equation of Fermat is

$$x^n + y^n = z^n.$$

Here we study a 'local' version of Fermat's problem, i.e., the question when Fermat's equation has a *non-trivial* solutions modulo a prime p. A solution is *trivial* if $x = z \mod p$ (and $y = 0 \mod p$) or $y = z \mod p$ (and $x = 0 \mod p$).

COROLLARY 4.5.2. *Let $n \in \mathbb{N}$. Then there exists $q \in \mathbb{N}$ such that for all primes $p \geq q$, the equation $x^n + y^n = z^n$ has a non-trivial solution in $\mathbb{Z}_p$.*

PROOF. Take $q := s(n)+1$, where $s(n)$ is taken from Schur's theorem. Let $p \geq q$. Then $G_n := \{x^n \mid x \in \mathbb{Z}_p^*\}$ is a subgroup of the multiplicative group $\mathbb{Z}_p^* := \mathbb{Z}_p \setminus \{0\}$ of $\mathbb{Z}_p$, so we can partition $\mathbb{Z}_p^*$ into cosets $a_1G_n, \ldots, a_rG_n$. Note that $r \leq n$: this follows from the fact that $x \mapsto x^n$ is a homomorphism from $\mathbb{Z}_p^*$ to G_n. Since r equals the size of the kernel of this homomorphism, it is the number of roots of the polynomial $x^n - 1$ in the field $\mathbb{Z}_p$, hence $r \leq n$.

Colour the elements of a_iG_n with colour i. Since $r \leq n$ and since

$$|\mathbb{Z}_p^*| = p - 1 \geq q - 1 = s(n),$$

Schur's theorem implies the existence of a monochromatic triple $(x, y, z) \in (\mathbb{Z}_p^*)^3$ with $x + y = z$. Hence, there are $i \leq r$ and $\tilde{x}, \tilde{y}, \tilde{z} \in \mathbb{Z}_p^*$ such that

$$a_i\tilde{x}^n + a_i\tilde{y}^n = a_i\tilde{z}^n \mod p.$$

Since a_i is not divisible by p it follows that $\tilde{x}^n + \tilde{y}^n = \tilde{z}^n \mod p$. The solution $\tilde{x}, \tilde{y}, \tilde{z}$ is non-trivial since $\tilde{x}$, $\tilde{y}$, and $\tilde{z}$ are pairwise incongruent modulo p. □

4.5.2. Geometry. Now a geometric application. We say that ℓ points $p_1, \ldots, p_\ell$ in $\mathbb{R}^2$ are in *general position* if no three of them are collinear. A subset S of $\mathbb{R}^2$ is called *convex* if for all $x, y \in S$

$$\{\alpha x + \beta y \mid \alpha, \beta \in \mathbb{R}, \alpha + \beta = 1, \alpha, \beta \geq 0\} \subseteq S.$$

The *convex hull* of S is the smallest convex subset of $\mathbb{R}^2$ that contains S. A *convex n-gon* is a set of n points in $\mathbb{R}^2$ such that none of the points lies in the convex hull of the other points. A *convex quadrilateral in* $\mathbb{R}^2$ is a convex 4-gon. The happy end problem[1] is the following:

PROPOSITION 4.5.3. *Let S be a set of 5 points in general position in $\mathbb{R}^2$. Then S contains four points which form a convex quadrilateral.*

PROOF. If two points a, b are in the convex hull of the three other points c, d, e, then two out of these three points must lie on the same side as the line connecting a and b. Then these two points together with a, b forms a convex quadrilateral. Otherwise, there are at least four of the points on the boundary of the convex hull, and hence form a convex quadrilateral. □

So far, no Ramsey theory is involved. This changes with the next statement. We will present two proofs, both based on Ramsey theory.

THEOREM 4.5.4 (Erdős-Szekeres 1935). *For every $n \in \mathbb{N}$ there exists an $L \in \mathbb{N}$ such that any set of L points in general position in $\mathbb{R}^2$ contains the vertices of a convex n-gon.*

ORIGINAL PROOF OF ERDŐS AND SZEKERES. The statement is trivial for $n \leq 3$. Let $L := R_2^4(n)$, so that $[L] \to (n)_2^4$. Let $V = \{s_0, \ldots, s_{L-1}\}$ be a set of L points in general position in $\mathbb{R}^2$. Let $S \in \binom{V}{4}$, and colour S blue if the points in S are in convex position, and red otherwise. By Theorem 4.3.1, there is a monochromatic subset M of V of size n. We have already seen in the happy ending problem that not all sets in $\binom{M}{4}$ can be red. Hence, all four-element subsets of M are in convex position, and it is easy to see that then all points in M must be in convex position. □

[1] A name given by Erdős since it lead to the marriage of George Szekeres and Esther Klein.

ANOTHER PROOF. Let $L := R_2^3(n)$, so that $[L] \to (n)_2^3$. Let $V = \{s_0, \dots, s_{L-1}\}$ be a set of ℓ points in general position in $\mathbb{R}^2$. Let $\{s_i, s_j, s_k\} \in \binom{V}{3}$ and assume without loss of generality that $i < j < k$. Colour this set according to the following rule:

- blue if going from s_i to s_j to s_k is a clockwise movement, and
- red otherwise.

By Theorem 4.3.1, there is a monochromatic subset S of V of size n, say of colour blue (the other case is symmetric). Suppose for contradiction that $s_\ell \in S$ lies in the convex hull of $\{s_i, s_j, s_k\} \in \binom{S\setminus\{s_\ell\}}{3}$. Without loss of generality, suppose that $i < j < k$ so that $\{s_i, s_j, s_k\}$ is coloured blue. Then $\{s_i, s_\ell, s_k\}$ is blue, so we have $i < \ell < k$. If $\ell < j$ then $i < \ell < j$ and the set $\{s_i, s_\ell, s_j\}$ should be coloured red, and if $\ell > j$ then $j < \ell < k$ and the set $\{s_j, s_\ell, s_k\}$ should be coloured red. In both cases we reached a contradiction. □

Exercises.

(73) Formulate and prove an infinite version of Schur's theorem. Show that for every $\chi\colon \mathbb{N} \to [c]$ there exist infinitely many triples $a, b, c \in \mathbb{N}$ such that $a + b = c$ and $\chi(a) = \chi(b) = \chi(c)$.

(74) Derive Schur's theorem from the statement in the previous exercise.

(75) Let $(P, \leq)$ be a countably infinite partially ordered set. Then $(P; \leq)$ either contains an infinite chain or an infinite antichain (see Exercise (10)).

(76) A generalisation of the previous exercise: prove that for any infinite directed graph there exists an infinite subset of the vertices that induces one of the following:
- a clique,
- a clique where additionally all vertices have loops,
- an independent set,
- loops at each vertex and otherwise no edges,
- a strict linear order,
- a weak linear order.

(77) Prove that for every $n \in \mathbb{N}$ there exists $\ell \in \mathbb{N}$ such that any directed graph without loops of size at least ℓ contains a clique, a stable set, or a linear order with n vertices as an induced subgraph.

(78) (Meta-mathematical exercise) What constitutes a *constructive proof*? Is the proof of Theorem 4.3.1 constructive? Is the proof of Theorem 4.4.1 constructive? Are constructive proofs more satisfactory than non-constructive proofs?

4.6. The Theorem of Hales-Jewett

One of the original motivations of the Hales-Jewett theorem [**20**] is an application for so-called *positional games* to show that certain games cannot end in a draw. The theorem became one of the most useful theorems in Ramsey theory. Graham, Rothschild, and Spencer in their classical textbook [**18**] write: *"(The Hales-Jewett theorem) is a focal point from which many results can be derived and acts as a cornerstone for much of the more advanced work. Without this result,* Ramsey theory *would more properly be called* Ramseyian theorems*."*

4.6.1. Positional games. A *hypergraph* is a pair (V, H) where

- V is a set (again, the elements of V are called the *vertices*), and
- $H \subseteq \mathcal{P}(V)$ is a set of subsets of V, called the *(hyper-)edges*.

A *(strong) positional game* (or a *maker-maker game*) is given by a hypergraph (V, H) and played by two players, called W und S (there are obvious generalisations to positional n-player games). Player W starts the game and colours one of the vertices white. Then player S colours one of the other vertices black. They alternatingly continue to colour previously uncoloured vertices with their respective color until

- all of the vertices of one of the hyperedges are coloured with white, in which case W wins;
- all of the vertices of one of the hyperedges are coloured with black, in which case B wins;
- all vertices are coloured without a monochromatic hyperedge, in which case the game ends in a draw.

This game is a finite perfect information zero-sum 2-player game, so either one of the players has a winning strategy, or both players can force a draw. The most famous of such positional games is Tic-Tac-Toe, where the hypergraph has the vertices $[3]^2$ (a three by three grid) and the hyperedges are the 3 rows, the 3 columns, and the two 2 diagonals.

The following famous idea has already been used by Nash around 1940, and is called *strategy stealing argument*. We write $W(x)$ (and $B(x)$) if x has been coloured white (black) during the game.

THEOREM 4.6.1. *Let $(V; H)$ be a strong positional game. Then the player B who plays second cannot have a winning strategy.*

PROOF. Suppose for contradiction that B has a winning strategy σ. We now specify a strategy for W. Arbitrarily select $x \in V$ (but don't colour it yet!). Let $y \in V$ be the response of B according to σ if W *would* have played x. Then player W plays y. By assumption, B can force a win for the game where $W(x)$ and $B(y)$ and where it is W's turn. Symmetrically, W can force a win if $B(x)$, $W(y)$, and if it is B's turn. But then, W can force a win if $W(y)$ and if it is B's turn, even if x has not been coloured (since additionally coloured vertices can only help B to win). This contradicts that fact that B has a winning strategy for the game W has just started against σ (where $W(x)$ and it is B's turn). □

We have here a real proof by contradiction; note that from the proof we gain no insight how B should play in order to force a draw or to win. In the case of Tic-Tac-Toe, it can be checked by an elementary case distinction that both players can force a draw. We will not go into the details here but rather discuss the obvious generalisation of Tic-Tac-Toe where we play on a n by n board, rather than a 3 by 3 board. Formally, the game $n \times n$ is the positional game whose vertices are the grid $[n] \times [n]$ and where we have n hyperedges of size n for the n rows, the n hyperedges of size n for the n columns, and two diagonals.

To decide whether B can force a draw there is a powerful condition that sometimes works. The idea even has a (Japanese) name in the world of go: B tries to cover each hyperedge with a pair of vertices that are *miai*.

DEFINITION 4.6.2. A *pairing strategy* for $(V; H)$ is given by a matching M on the complete graph with vertices V such that every hyperedge contains at least one edge from M.

For example, the following represents a pairing strategy for the game 5×5: entries with the same number are joint by a matching edge.

$$\begin{pmatrix} 1 & 2 & 3 & 2 & 4 \\ 5 & 7 & 7 & 8 & 9 \\ 10 & 6 & * & 8 & 10 \\ 5 & 6 & 11 & 11 & 9 \\ 4 & 12 & 3 & 12 & 1 \end{pmatrix}$$

PROPOSITION 4.6.3. *If $(V;H)$ has a pairing strategy, then B can force a draw.*

PROOF. If W plays one of the endpoints of a matching edge, then B responds with the other end. Since every edge is covered by some hyperedge, none of the hyperedges will be coloured monochromatically, and the game ends in a draw. □

Note that there is no pairing strategy for the game 4×4: there are 10 hyperedges, but only 16 vertices; hence, at most 8 hyperedges can be covered by some matching.

Exercises.

(79) Prove that B can force a draw in the game 4×4.
Hint. Check that for each of the possible first moves of W there is a response for B such that B has a pairing strategy.

(80) The game *Sim* is played by the two players white and black on a hypergraph (V, H) where
- $V = \binom{[6]}{2}$.
- $H = \big\{\{\{a,b\},\{b,c\},\{a,c\}\} \mid a,b,c \in [6]\big\}$.

White begins. The players alternatingly color the elements of V by their color. If one player creates a monochromatic hyperedge, she looses. Prove that one of the players has a winning strategy. Determine which of the players has a winning strategy.

4.6.2. The $[n]^d$ game. In the following, we will be interested in a d-dimensional generalisation of the game $n\times n$, which is called the n^d *game*. The vertex set in this this game is $[n]^d$. The hyperedges are of size n and defined as follows: $\{\alpha_1,\dots,\alpha_n\}$ with $\alpha_i = (a_{i,1},\dots,a_{i,d}) \in [n]^d$ is a hyperedge if for each $j \le d$ the tuple $(a_{1,j},\dots,a_{n,j})$ has one of the following forms:

$$\begin{aligned} &(1,\dots,1)\\ &(2,\dots,2)\\ &\dots\\ &(n,\dots,n)\\ &(n,n-1,\dots,1)\\ &(1,2,\dots,n) \end{aligned}$$

If $n = 4, d = 3$ then there are for instance the hyperedges

$$\begin{aligned} &\{113,112,111,110\},\\ &\{020,121,222,323\}, \text{ and}\\ &\{031,131,231,331\}. \end{aligned}$$

Sometimes, pairing strategies are guaranteed to exist because of the marriage theorem of Hall, in the form of Theorem 1.5.7. For example, we have the following.

LEMMA 4.6.4. *Let $(V;H)$ be a hypergraph and let s be the size of the smallest hyperedge in H. Let $g := \max_{x\in V} |\{h \in H \mid x \in h\}|$. If $s \ge 2g$ then B has a pairing strategy in the game $(V;H)$.*

PROOF. Let $\mathcal{F}$ be the family of finite subsets where we add each hyperedge from H twice. Note that pairing strategies in the game on (V, H) are in one-to-one correspondence to transversals of $\mathcal{F}$: by construction, each hyperedge will have two distinct representatives a, b in a transversal, and $\{a, b\}$ will be part of the desired matching for the pairing strategy. Theorem 1.5.7 states that is suffices to verify the marriage condition for $\mathcal{F}$. Let $\mathcal{S} \subseteq \mathcal{F}$.

$$\begin{aligned} |N(\mathcal{S})| &= \left| \bigcup_{A \in \mathcal{S}} A \right| && \text{(by the definition of } G) \\ &\overset{(*)}{\geq} \frac{s|\mathcal{S}|}{2g} \\ &\geq |\mathcal{S}| && \text{(since } s \geq 2g). \end{aligned}$$

The inequality $(*)$ holds since we have at least $s|\mathcal{S}|$ edges leaving $\mathcal{S}$, and each vertex in $\bigcup_{A \in \mathcal{S}} A$ has at most $2g$ neighbours in $\mathcal{S}$. □

COROLLARY 4.6.5. *For $n \geq 3^d - 1$ both players can force a draw in the game n^d.*

PROOF. Note that any point is contained in at most $(3^d - 1)/2$ many hyperedges (and this bound is achieved for odd k at the center point): for each coordinate, we have to decide whether the entries are constant, increasing, or decreasing. Not all hyperedges can be constant, so we have to subtract one. Finally, if we flip all increasing coordinates to decreasing ones, and vice versa, we obtain the same hyperedge, so we divide by 2. Hence, if $n \geq 3^d - 1$, then Lemma 4.6.4 applies and implies the statement. □

The result in the following section shows that for every $n \in \mathbb{N}$ there exists a $d \in \mathbb{N}$ so that $[n]^d$ cannot end in a draw, and hence, by Theorem 4.6.1, W has a winning strategy.

4.6.3. The Hales-Jewett Theorem. In the Hales-Jewett theorem, we color the elements of $[m]^d$ with c colours, and we look for a monochromatic *combinatorial line*.

DEFINITION 4.6.6. A *combinatorial line* in $[m]^d$ is a set of points $\{\alpha_1, \ldots, \alpha_m\} \subseteq [m]^d$ with $\alpha_i = (a_{i,1}, \ldots, a_{i,d}) \in [m]^d$ for all $i \leq m$, such that for each $j \leq d$ we have that either

$$a_{1,j} = \cdots = a_{m,j}$$

or

$$a_{i,j} = i.$$

Hence, combinatorial lines are defined as the hyperedges $\{\alpha_1, \ldots, \alpha_m\}$ of the game m^d with the exception that we no longer have the option that $\alpha_{1,j} = m + 1 - j$. So there are fewer combinatorial lines; for example in the game Tic-Tac-Toe, we have the hyperedge $\{(0, 2), (1, 1), (2, 0)\}$ which is *not* a combinatorial line.

Combinatorial lines can be represented as words over the alphabet $[m] \cup \{*\}$ with at least one occurrence of $*$. The letter i at the j-th position indicates that in the combinatorial line

$$\{(a_{1,1}, \ldots, a_{1,d}), \ldots, (a_{m,1}, \ldots, a_{m,d})\}$$

we have $a_{1,j} = \cdots = a_{m,j} = i$, and the letter $*$ at the j-th position indicates that $a_{i,j} = i$ for all $i \in \{1, \ldots, d\}$. It follows that the number of combinatorial lines in $[m]^d$ is $(m + 1)^d - m^d$.

THEOREM 4.6.7. *For all $c, m \in \mathbb{N}$ there exists $d = HJ(c, m) \in \mathbb{N}$ such that for all colourings $f \colon [m]^d \to [c]$ there exists a monochromatic combinatorial line in $[m]^d$.*

For example, if $c = 2$ and $m = 2$ then $HJ(c, m) = 2$ is the smallest $d \in \mathbb{N}$ that satisfies the statement of the theorem.

COROLLARY 4.6.8. *For every $n \in \mathbb{N}$ there exists a $d \in \mathbb{N}$ such that player W wins the game n^d.*

PROOF. Choosing $d = HJ(c, n)$, the game n^d cannot end in a draw since every combinatorial line in $[n]^d$ is a hyperedge in the game n^d, and if we have a monochromatic hyperedge then one of the two players wins. Since player S cannot have a winning strategy by Theorem 4.6.1, player W must have a winning strategy. □

4.6.4. Application: van der Waerden's theorem. An *arithmetic progression of length m (m-AP)* is a sequence $(a_0, a_1, \ldots, a_{m-1})$ of integers of the form

$$a_0, a_0 + b, a_0 + 2b, \ldots, a_0 + (m-1)b.$$

THEOREM 4.6.9 (van der Waerden, 1927). *For all $m, c \in \mathbb{N}$ there is an $w = W(m, c) \in \mathbb{N}$ such that for every colouring*

$$\chi \colon \{1, \ldots, w\} \to [c]$$

there exists a monochromatic arithmetic progression of length m.

PROOF. Let $d = HJ(m, c)$, and define $w := (m-1)d + 1$. Let $f \colon [m]^d \to \{1, \ldots, w\}$ be the map that sends $(p_1, \ldots, p_d)$ to $p_1 + \cdots + p_d + 1$. Given a c-colouring χ of $\{1, \ldots, w\}$, define a c-colouring χ' of $[m]^d$ by setting $\chi'(p) := \chi(f(p))$ for $p \in [m]^d$. By the theorem of Hales-Jewett (Theorem 4.6.7), there exists a monochromatic line $q = (q_0, \ldots, q_{m-1})$ in $[m]^d$. We claim that $f(q_0), \ldots, f(q_{m-1})$ is an m-AP. Let $a_0 := f(q_0)$ and let b be the number of $*$'s in the word that corresponds to the line q. Then $f(q_i) = a_0 + ib$ for $i \in [m]$. So we found the monochromatic m-AP

$$a_0, a_0 + b, \ldots, a_0 + (m-1)b.$$ □

4.6.5. Application: monochromatic copies of graphs.

THEOREM 4.6.10. *For every $c \in \mathbb{N}$ and every finite graph H there exists a graph G such that for every colouring $\chi \colon V(G) \to [c]$ there exists a monochromatic subgraph H' of G which is isomorphic to H.*

PROOF. Let $m := |V(H)|$ and choose $d = HJ(c, m)$. Let G be the graph with vertex set $[m]^d$ where the edges are defined in such a way that for each combinatorial line $\{(\alpha_{1,1}, \ldots, \alpha_{1,d}), \ldots, (\alpha_{m,1}, \ldots, \alpha_{m,d})\}$ in $[m]^d$ induces a copy of H: this is well-defined since any two combinatorial lines intersect in at most one element of $[m]^d$. Clearly, monochromatic lines in $[m]^d$ then correspond to monochromatic copies of H in G. □

The following exercises follow pretty much the same idea.

Exercises.

(81) Prove that for every $c \in \mathbb{N}$ and every finite metric space M there exists a metric space L such that for every colouring $\chi \colon L \to [c]$ there exists a monochromatic subspace of L which is isometric to M.

(82) Prove that for every $c \in \mathbb{N}$ and every finite partially ordered set $(M, \leq)$ there exists a partially ordered set $(L, \leq)$ such that for every colouring $\chi \colon L \to [c]$ there exists a monochromatic subposet of $(L, \leq)$ which is isomorphic to $(M, \leq)$.

(83) Prove that for every $c \in \mathbb{N}$ and every finite tournament S there exists a finite tournament T such that for every colouring $\chi \colon V(T) \to [c]$ there exists a monochromatic subtournament of T which is isomorphic to S.

Remark. One of the important directions into which Ramsey theory has developed is that instead of just colouring single elements as in the theorem of Hales-Jewett and its applications that we have seen above, we colour entire copies of of some fixed ('small') structure S in some ('large') structure L. The goal is to find conditions that imply that no matter how the copies of S in L are coloured, we find a copy of some ('medium-size') structure M in S such that all copies of S in M have the same colour. This is similar to the statement of Ramsey's theorem, except that for Ramsey's theorem there were just sets and subsets, rather than structures and substructures. For more on this topic, we refer to [**30**]. Another important topic in Ramsey theory is the question how big structures need to be so that *infinite* monochromatic objects can be found.

CHAPTER 5

Generating Functions

In combinatorics, we often want to *count* the objects of size n in a given set of combinatorial structures. For example, we want to count the number of graphs, trees, etc. that have certain properties, as a function in the number of vertices. Or we would like to count the number of words in some language depending on the number of letters in the word. There are numerous reasons why we might want to count. One important application is the probabilistic method where probability theory is mostly a fancy (and powerful!) language for arguments that are essentially counting arguments. Another application is in theoretical computer science, where we are sometimes interested in *typical properties* of large random objects (the large random objects might actually be the input to your computer program). So we sometimes might want to efficiently sample large combinatorial objects from some distribution; efficient counting is very important in this context. The corresponding research field is called *enumerative combinatorics*.

This section does not attempt to give an overview of enumerative combinatorics, which would be a challenging task, but rather to highlight one very powerful method in enumerative combinatorics, namely the usage of *generating functions*. This method has strong links with algebra, but also with analysis of complex functions (in German *Funktionentheorie*); the respective research area is often called *analytic combinatorics*. This section is inspired by the famous introduction of Wilf with the title *generatingfunctionology* [**41**] (which is suited for mathematics undergraduate students) and the monumental book of Flajolet and Sedgewick [**16**]. However, we also try to be self-contained in what concerns the facts concerning the fundamental link between power series and complex-valued functions.

5.1. Motivating Generating Functions

Suppose we have a problem whose answer is a sequence of numbers, $a_0, a_1, a_2, \ldots$ Ideally, we would like to obtain a simple formula for the n-th number, something like $a_n = 2n^2 + 7n - 2$, for example. But sometimes, such a simple formula might simply not exist. In such situations we would still like to know whether the n-th number can be somehow computed (as efficiently as possible!), or we might want to know the asymptotic growth of the a_n, or good and simple bounds on the asymptotic growth. Or you might want to prove that two sequences are equal. For all these possible tasks (and many more) generating functions might be the right method.

EXAMPLE 9. The *Fibonacci numbers* $f_0, f_1, f_2, \ldots$ are inductively defined as follows: $f_0 = 0$, $f_1 = 1$, and for $n \geq 1$ we have

$$f_{n+1} := f_n + f_{n-1}.$$

The sequence begins with

$$0, 1, 1, 2, 3, 5, 8, 13, 21, \ldots$$

Using generating functions, one can prove that for every $n \in \mathbb{N}$

$$f_n = \frac{1}{\sqrt{5}}(r_+^n - r_-^n). \tag{42}$$

where $r_+ := (1+\sqrt{5})/2$ is the *golden ratio* (*sectio aurea*, *proportio divina*) and $r_- := (1-\sqrt{5})/2$. By any means, this is a formula that is easy to compute (by hand, or efficiently even for gigantic numbers n by a computer). Note that for large n, since $|r_-| < 1$, the second term $r_-^n/\sqrt{5}$ in (42) will be tiny when compared to the first. So, in particular,

$$f_n \sim \left(\frac{1+\sqrt{5}}{2}\right)^n. \qquad \triangle$$

5.2. The Idea

We start with a very simple example to illustrate the idea. The Fibonacci numbers and Catalan numbers will be more interesting examples that follow later. A formal treatment of the involved tool, namely formal power series, and more advanced analytic techniques follow later.

Let $a_0, a_1, \ldots$ be the integer sequence that is given by $a_0 := 0$, and recursively for all $n \geq 0$

$$a_{n+1} := 2a_n + 1. \tag{43}$$

The sequence starts with $0, 1, 3, 7, 15, 31, \ldots$ In this example it is still possible to somehow guess an explicit formula for the sequence and then prove it by induction on n. This is *not* our approach (the guess and check approach might be quite tough in more advanced examples – would you have guessed the answer in Example 9?).

Step 1: turn the sequence into a power series $A(x)$. The starting point of the approach via generating functions is to turn the sequence $(a_n)_{n\in\mathbb{N}}$ into a single object, namely the *(formal) power series*

$$\sum_{n\in\mathbb{N}} a_n x^n \tag{44}$$

where x is a variable. This object has two possible interpretations:

- The *formal/syntactic* perspective, where (5.5) is just viewed as a formal power series. More on that in Section 5.3.
- A function from a subset of $\mathbb{C}$ to $\mathbb{C}$, mapping x to $\sum_{n\in\mathbb{N}} a_n x^n$. More on that in Section 5.5.

Step 2: translate the given data into statements about $A(x)$. Our next step is the interpretation of the recurrence relation (43) in terms of (44). Multiplying both sides of the recurrence relation by x^n, and then summing over all $n \in \mathbb{N}$ (for all of them the recurrence relation is true) we obtain

$$\sum_{n\geq 0} a_{n+1} x^n = \sum_{n\geq 0} (2a_n + 1)x^n. \tag{45}$$

On the left-hand-side of (45) we have

$$\sum_{n\geq 0} a_{n+1} x^n = \frac{\sum_{n\geq 0} a_{n+1} x^{n+1}}{x} + \underbrace{\frac{a_0 x^0}{x}}_{=0} = \frac{A(x)}{x}.$$

To simplify the right-hand-side, recall that (this will be revisited in Section 5.3)

$$\frac{1}{1-x} = 1 + x + x^2 + \cdots = \sum_{n\in\mathbb{N}} x^n. \tag{46}$$

It is the result of performing the (infinite) polynomial division

$$\begin{array}{llll} 1 & & : & (1-x) = 1 + x + x^2 + \cdots \\ 1 & -x & & \\ & x & -x^2 & \\ & & x^2 & -x^3 \\ & & & \cdots \end{array}$$

Hence, the right-hand-side of (45) can be rewritten as

$$\sum_{n\geq 0} 2a_n x^n + \sum_{n\geq 0} x^n = 2A(x) + \frac{1}{1-x}. \tag{47}$$

Step 3: obtain a simple expression for $A(x)$ by algebraic manipulations. If we now put the left-hand side expression and the right hand side expression together and obtain

$$\frac{A(x)}{x} = 2A(x) + \frac{1}{1-x}$$

Solving for $A(x)$ we obtain

$$A(x) = \frac{x}{(1-x)(1-2x)}.$$

Step 4: read off the coefficients. In general, if $B(x)$ is a power series, then we write $[x^n]B(x)$ for the coefficient of x^n. In our case, we have $[x^n]A(x) = a_n$. The value of a_n for a given specific value of n can be computed by performing a polynomial division for n steps until we know a_n.

To obtain a formula for a_n, we first compute the *partial fraction decomposition* of $A(x)$. That is, we want to write the fraction $\frac{x}{(1-x)(1-2x)}$ as

$$\frac{P}{1-x} + \frac{Q}{1-2x}$$

for appropriate polynomials P and Q. In our case, it is possible to guess $P = -x$ and $Q = 2x$ (a general method to find P and Q will be revisited in Section 5.3.5):

$$\frac{-x}{1-x} + \frac{2x}{1-2x} = x\frac{-(1-2x)+2(1-x)}{(1-x)(1-2x)} = \frac{x}{(1-x)(1-2x)}.$$

We have already seen in (46) that $[x^n](\frac{1}{1-x}) = 1$ for all $n \in \mathbb{N}$, and similarly

$$[x^n]\frac{1}{1-2x} = 2^n. \tag{48}$$

Putting these together, we obtain that

$$a_n = [x^n]A(x) = [x^{n-1}]\left(\frac{2}{1-2x} - \frac{1}{1-x}\right) = 2\cdot 2^{n-1} - 1 = 2^n - 1.$$

5.3. Formal Power Series

To introduce formal power series, we need a commutative ring R (such as $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$) and a formal variable, x. We first define the *ring of formal power series* $R[[x]]$ by defining the elements of $R[[x]]$ and how to add and how to multiply elements of $R[[x]]$.

5.3.1. Defining power series. We define $R[[x]]$ as the set of infinite sequences of elements of R, indexed by the natural numbers $\mathbb{N}$ (including 0; otherwise, we write $\mathbb{N}^+$). Let $(a_i)_{i\in\mathbb{N}}$ and $(b_i)_{i\in\mathbb{N}}$ be two elements of $R[[x]]$. We define

$$(a_i)_{i\in\mathbb{N}} + (b_i)_{i\in\mathbb{N}} := (a_i + b_i)_{i\in\mathbb{N}}$$

$$(a_i)_{i\in\mathbb{N}} \cdot (b_i)_{i\in\mathbb{N}} := \sum_{k=0}^{n} (a_k b_{n-k})_{n\in\mathbb{N}}$$

It is straightforward to verify that $R[[x]]$ with these operations forms a ring. Clearly, the multiplicative unit is the identity series $1 := (1, 0, 0, \dots)$.

The definition of the product is called the *Cauchy product* of $(a_i)_{i\in\mathbb{N}}$ and $(b_i)_{i\in\mathbb{N}}$, and it is clearly useful for combinatorial enumeration, for the following reason: often, in order to construct an object of size n in some class of combinatorial objects $\mathcal{C}$, we have to take an object of size i from some class $\mathcal{A}$ and an object of size $n-i$ from some other class $\mathcal{B}$ and put them together in some unique way (a concrete example of such a class is given in Section 5.6). So the number of all objects of size n equals $\sum_{i=0}^{n} a_i b_{n-i}$, where a_i, b_i is the number of objects of size i in $\mathcal{A}$ and $\mathcal{B}$, respectively. And this sum is precisely the sum that appears in the definition of the Cauchy product.

We view the set $R[x]$ of polynomials over R with one variable x as a subset of $R[[x]]$, identifying $(a_0, a_1, \dots, a_n, 0, 0, \dots)$ with $\sum_{i=0}^{n} a_i x^i$. We adopt a similar notation for the general case: instead of $(a_i)_{i\in\mathbb{N}}$ we write $\sum_{n=0}^{\infty} a_n x^n$. If $A(x) = \sum_{n=0}^{\infty} a_n x^n \in R[[x]]$ we also write $[x^n]A(x)$ instead of a_n.

REMARK 5.3.1. Of course, we can consider formal power series $S[[x]]$ where S itself is a ring of formal power series, $S = R[[y]]$. If we iterate this n times with the formal variables $x_1, \dots, x_n$, starting with the ring R, we directly use the obvious notation $R[[x_1, \dots, x_n]]$ for the resulting object.

5.3.2. The reciprocal power series. A formal power series $B(x)$ is a *reciprocal* of $A(x)$ if $A(x)B(x) = 1$; that is, $B(x)$ is the multiplicative inverse of $A(x)$ in the ring $R[[x]]$. However, we prefer the term 'reciprocal' since the 'inverse power series' is used for something else in Section 5.3.4. For an example of a power series without a reciprocal, take $R = \mathbb{Z}$ and consider $2 \in R[x] \subseteq R[[x]]$.

LEMMA 5.3.2. *A series $A(x) \in R[[x]]$ has a reciprocal in $R[[x]]$ if and only if $[x^0]A(x)$ is invertible in R. In this case the reciprocal is unique.*

PROOF. If $A(x) = \sum_{n=0}^{\infty} a_n x^n$ has the reciprocal $B(x) = \sum_{n=0}^{\infty} b_n x^n \in R[[x]]$, then

$$1 = [x^0](1) = [x^0](A(x) \cdot B(x)) = a_0 b_0$$

and hence a_0 has the inverse b_0 in R. Moreover, from the definition of $A(x)B(x) = 1$ it follows that for $n \geq 1$ we have $\sum_{k=0}^{n} a_k b_{n-k} = 0$, which implies that

$$b_n = -\frac{1}{a_0} \sum_{k\geq 1} a_k b_{n-k} \tag{49}$$

which, by induction, defines $b_1, b_2, \dots$ uniquely.

Conversely, if a_0 has the inverse b_0 in R, then we can determine $b_1, b_2, \dots$ from (49) and the series $B(x) := \sum_{n\in\mathbb{N}} b_n x^n$ is the reciprocal series for $A(x)$. □

If $A(x)$ has a reciprocal $B(x)$, then we also write $A(x)^{-1}$ for $B(x)$. An important special case is the identity

$$(1 - cx)^{-1} = \sum_{n\in\mathbb{N}} c^n x^n \tag{50}$$

which we have already encountered in (46) for $c = 1$ and later for $c = 2$ in (48). Here,

$$\begin{aligned}[x^n](1-cx)^{-1} &= -\sum_{i=1}^{n} a_i b_{n-i} && \text{(by (49))}\\ &= -a_1 b_{n-1} && \text{(since } a_i = 0 \text{ for } i \geq 2)\\ &= c^n && \text{(since } a_1 = -c \text{ and inductively } b_{n-1} = c^{n-1}).\end{aligned}$$

It is important to note that for Equation (50) we do not care about convergence issues; x is just a formal variable, and not a number. On both sides we have a formal power series, and they are equal.

5.3.3. The derived power series. The derivative of a formal power series $A(x) = \sum_{n\in\mathbb{N}} a_n x^n$ is the series

$$A'(x) := \sum_{n\in\mathbb{N}^+} n a_n x^{n-1}.$$

The usual rules for the derivative also hold in the formal setting:

$$\begin{aligned}(A(x)+B(x))' &= A'(x)+B'(x)\\ (A(x)B(x))' &= A'(x)B(x)+A(x)B'(x) \qquad \text{(Leibniz rule)}\end{aligned} \tag{51}$$

PROPOSITION 5.3.3. *Let $A \in R[[x]]$ be such that $A'(x) = 0$. Then $A(x) = c$ for some $c \in R$.*

PROOF. $A'(x) = 0$ means precisely that $[x^n]A'(x) = 0$ for all $n \in \mathbb{N}^+$. □

Recall from Section 3.2.2 that $\exp(x)$ was defined as the formal power series

$$\sum_{n\in\mathbb{N}} \frac{x^n}{n!};$$

note that this definition makes sense for every ring R.

PROPOSITION 5.3.4. *Let $A \in R[[x]]$ be such that $A'(x) = A(x)$. Then $A(x) = c\exp(x)$ for some $c \in R$.*

PROOF. If $A' = A$ then $a_n := [x^n]A = [x^n]A'$ and hence $(n+1)a_{n+1} = a_n$, for all $n \geq 0$. Therefore, $a_{n+1} = \frac{a_n}{n+1}$, and by induction on n we obtain that $a_n = \frac{a_0}{n!}$, so $A(x) = a_0 \exp(x)$. □

Integration is defined analogously to derivation.

Exercises.

(84) Prove the Leibniz rule (51) for the derivative of products of formal power series.

5.3.4. Composing power series. Let $A(x) = \sum_{k=0}^{n} a_k x^k$ be a polynomial over the ring R. As with polynomials $B(x) \in R[x]$, we can define the *composition* of $A(x)$ with a formal power series $B(x) \in R[[x]]$ by

$$A(B(x)) := \sum_{k=0}^{n} a_k (B(x))^k.$$

EXAMPLE 10. Clearly, $A(0) := \sum_{k=0}^{n} a_k 0^k = a_0$. More generally, we can recover a_n as $A^{(n)}(0)$ using the formal derivative from the previous section and composition with 0 as follows. To denote higher derivatives, we define $A^{(0)}(x) := A(x)$ and for $n \in \mathbb{N}$

$$A^{(n+1)}(x) := (A^n(x))'.$$

It is now easy to check that

$$a_n = \frac{A^{(n)}(0)}{n!}. \tag{52}$$

△

Now suppose that $A(x)$ is in $R[[x]]$ instead of $R[x]$. We would like to use the same definition, but we need that $[x^0]B(x) = 0$ because then

$$A(B(x)) = \sum_{n\in\mathbb{N}} a_n (B(x))^n$$

can be given formal meaning by defining

$$[x^n]A(B(x)) := \sum_{k\in\mathbb{N}, j_1,\ldots,j_k\in\mathbb{N}^+, j_1+\cdots+j_k=n} a_k b_{j_1}\cdots b_{j_k}. \tag{53}$$

We also write $(A\circ B)(x)$ instead of $A(B(x))$.

Example 11. The expression $\exp(\exp(x)-1)$ is a well-defined formal power series since $[x^0](\exp(x)-1) = [x^0]\sum_{n\in\mathbb{N}} \frac{x^n}{n!} - 1 = 1-1 = 0$. On the other hand, $\exp(\exp(x))$ is not defined, at least not by the general definition of composition of formal power series. △

Remark 5.3.5. So we now have defined composition in two different situations: polynomials $A \in K[x]$ with arbitrary power series $B \in K[[x]$, or arbitrary power series $A \in K[[x]]$ with power series $B \in K[[x]]$ having constant term 0. We mention that using the notion of *summable* power series allows for a common generalisation of the two situations [**36**].

Composition of power series is of great use in combinatorial enumeration: often, an object of size n in a class of combinatorial objects $\mathcal{C}$ is obtained in a unique way by taking a structure with k elements from one class $\mathcal{A}$, and by replacing each element by an non-empty object O_i from some other class $\mathcal{B}$ such that the sizes of $O_1,\ldots,O_k$ add up to n. Hence, the number of objects of size n in $\mathcal{C}$ equals

$$\sum_{k\in\mathbb{N}, j_1,\ldots,j_k\in\mathbb{N}^+, j_1+\cdots+j_k=n} a_k b_{j_1}\cdots b_{j_k}.$$

As in calculus, we have the *chain rule* for computing the derivative of the composition $A(B(x))$:

$$A(B(x))' = A'(B(x))B'(x). \tag{54}$$

An *inverse* of a series $A(x)$, if it exists, is a power series $B(x)$ such that

$$A(B(x)) = B(A(x)) = x.$$

Proposition 5.3.6. *Let $A \in R[[x]]$ be such that $[x^0]A(x) = 0$. Then $A(x)$ has an inverse if and only if $[x^1]A(x)$ has a (multiplicative) inverse in R. In this case the inverse is unique.*

Proof. Suppose that $A(x) = \sum_{n\in\mathbb{N}} a_n x^n$ has the inverse $B(x) = \sum_{n\in\mathbb{N}} b_n x^n$ in $R[[x]]$. Let $r, s \in \mathbb{N}$ be smallest such that $a_r \neq 0$ and $b_s \neq 0$. Then the smallest n such that $[x^n]A(B(x)) \neq 0$ equals rs, and since $A(B(x)) = x$ we must have that $rs = n = 1$, and hence $r = s = 1$. Moreover, $a_1 b_1 = 1$, and hence $[x^1]A(x) = a_1$ is invertible. Moreover, $b_1 \in R$ is uniquely determined. The identity $A(B(x)) = x$ and (53) implies that

$$[x^2]A(B(x)) = a_1 b_2 + a_2 b_1^2 = 0$$

and hence b_2 is uniquely determined in terms of the $A(x)$ and b_1. Likewise,

$$[x^3]A(B(x)) = a_1b_3 + 2a_2b_1b_2 + a_3b_1^3 = 0$$

and hence b_3 is uniquely determined in terms of the $A(x)$ and b_1, b_2. Similarly, by induction, b_n is uniquely determined for all $n \in \mathbb{N}$.

Conversely, if a_1 has the inverse b_1 in R, then we choose $b_0 = 0$ and $b_2, b_3, \ldots$ as above and then the series $B(x) := \sum_{n\in\mathbb{N}} b_n x^n$ is the inverse series for $A(x)$. □

Note that the proof also shows that if $A(B(x)) = x$, then this implies that $B(A(x)) = x$. If R is a field of characteristic 0 one can use the so-called *Lagrange inversion formula* to compute the b_i's explicitly. This will be treated in Section 5.8.2.

Exercises.

(85) Prove the chain rule (54).

5.3.5. The partial fraction decomposition. Quotients $\frac{p(x)}{q(x)}$ of two polynomials $p, q \in R[x]$ are called *rational functions*. A *partial fraction decomposition* of a rational function is an expression of the form

$$\frac{p(x)}{q(x)} = \sum_{j=1}^{k} \frac{p_j(x)}{q_j(x)}$$

where $q_1, \ldots, q_k$ are factors of q that have smaller degree than q. In the full decomposition (formally defined in Theorem 5.3.7) each of the q_j is irreducible.

For getting the idea why such a decomposition exists, suppose that $q = q_1q_2 \in K[x]$ for some field K such that q_1 and q_2 are coprime. By Bézout's theorem, which applies since $K[x]$ is a principal ideal domain (German *'Hauptidealring'*), there are $c, d \in K[x]$ such that $cq_1 + dq_2 = 1$ (recall that if K is the rational numbers then c and d can be found efficiently using the extended Euclidean algorithm). Thus,

$$\frac{1}{q} = \frac{cq_1 + dq_2}{q_1q_2} = \frac{c}{q_2} + \frac{d}{q_1}.$$

Hence,

$$\frac{p}{q} = \frac{cp}{q_2} + \frac{dp}{q_1}.$$

Following these ideas, one can show the following.

THEOREM 5.3.7. *Let K be a field and let $p, q \in K[x]$. Write q as a product of powers of distinct irreducible polynomials, $q = \prod_{i=1}^{k} q_i^{n_i}$. Then there are unique polynomials $b, a_{1,1}, \ldots, a_{1,n_1}, \ldots, a_{k,1}, \ldots, a_{k,n_k} \in K[x]$ such that the degree of $a_{i,j}$ is smaller than the degree of p_i and*

$$\frac{p}{q} = b + \sum_{i=1}^{k} \sum_{j=1}^{n_i} \frac{a_{i,j}}{q_i^j}$$

which is called the full partial fraction decomposition *of $\frac{p}{q}$. If the degree of f is smaller than the degree of g, then $b = 0$.*

5.3.6. The Fibonacci numbers. We now apply the generating function method to the Fibonacci numbers (Example 9). The **first step** is automatic: we turn the sequence $f_0, f_1, f_2, \ldots$ into the generating function

$$F(x) := \sum_{n\in\mathbb{N}} f_n x^n.$$

The **second step** is to express the recurrence relation

$$f_{n+1} = f_n + f_{n-1}$$

in terms of the generating function. Multiplying the left hand side by x^n and summing over $n \geq 1$ we obtain (using that $f_0 = 0$ and $f_1 = 1$):

$$\begin{aligned}\sum_{n\in\mathbb{N}^+} f_{n+1}x^n &= \frac{\sum_{n\in\mathbb{N}^+} f_{n+1}x^{n+1}}{x} \\ &= \frac{\sum_{n\geq 2} f_n x^n + x - x}{x} = \frac{F(x) - x}{x}.\end{aligned}$$

Doing the same on the right hand side we get (using that $f_0 = 0$)

$$\begin{aligned}\sum_{n\in\mathbb{N}^+} f_n x^n + \sum_{n\in\mathbb{N}^+} f_{n-1}x^n &= F(x) + x\sum_{n\in\mathbb{N}^+} f_{n-1}x^{n-1} \\ &= F(x) + x\sum_{n\in\mathbb{N}} f_n x^n \\ &= F(x) + xF(x).\end{aligned}$$

It follows that $\frac{F(x)-x}{x} = F(x) + xF(x)$. In the **third step** we solve the equation for $F(x)$, and obtain

$$F(x) = \frac{x}{1 - x - x^2}.$$

In the **fourth step** we try to get a formula for the coefficients of $F(x)$. As in our initial example in Section 5.2, we compute the partial fraction decomposition of $F(x)$. First, we factor the denominator: $1 - x - x^2$ has the roots

$$r_\pm := \frac{1}{2} \pm \sqrt{\frac{1}{4} + 1} = (1 \pm \sqrt{5})/2$$

and we obtain

$$(1 - r_- x)(1 - r_+ x) = 1 - (r_+ + r_-)x + (r_- r_+)x^2 = 1 - x - x^2.$$

To find the partial fraction expansion, we have to find $a, b \in K[x]$ such that

$$\frac{x}{(1 - r_- x)(1 - r_+ x)} = \frac{a}{1 - r_- x} + \frac{b}{1 - r_+ x}.$$

To find a, b we rewrite

$$\begin{aligned}\frac{a}{1 - r_- x} + \frac{b}{1 - r_+ x} &= \frac{(1 - r_+ x)a + (1 - r_- x)b}{(1 - r_+ x)(1 - r_- x)} \\ &= \frac{(a + b - (r_- b - r_+ a)x}{(1 - r_+ x)(1 - r_- x)};\end{aligned}$$

so we must have $a + b = 0$ and $r_- b - r_+ a = 1$. Substituting the first equation into the second we obtain $r_- a - r_+(-a) = 1$ and hence $b = \frac{1}{r_+ - r_-}$ and $b = -a$. Therefore,

$$\begin{aligned}\frac{x}{1 - x - x^2} = \frac{x}{(1 - r_+ x)(1 - r_- x)} &= \frac{ax}{1 - r_+ x} + \frac{bx}{1 - r_- x} \\ &= \frac{1}{r_+ - r_-}\left(\frac{1}{1 - r_+ x} - \frac{1}{1 - r_- x}\right) \\ &= \frac{1}{\sqrt{5}}\left(\sum_{n\in\mathbb{N}} r_+^n x^n - \sum_{n\in\mathbb{N}} r_-^n x^n\right).\end{aligned}$$

Therefore, $f_n = \frac{1}{\sqrt{5}}(r_+^n - r_-^n)$ as announced earlier in (42).

5.4. Regular Languages

In this section we study how to compute the number of words of length n in a given regular language. For example: how many ASCII texts of length 400 are there that contain the word 'combinatorics', but not the word 'generating function'? The answer to this question and other questions of this type can be easily computed.

5.4.1. Deterministic finite automata. A *deterministic finite automaton (DFA)* A is a 5-tuple

$$A = (Q, \Sigma, \delta, s, F)$$

consisting of

- a finite set of *states* Q;
- a finite set of input symbols called the *alphabet* Σ;
- a *transition function* $\delta\colon Q \times \Sigma \to Q$;
- an *initial* or *start state* $s \in Q$;
- a set of *accepting states* $F \subseteq Q$.

Let $w = (w_1, \ldots, w_n) \in \Sigma^n$ be a word (also called *string*; see Appendix B) of length n over the alphabet Σ. Then A accepts w if there exists a sequence of states $s = s_0, s_1, s_2, \ldots, s_n$ such that $s_{i+1} = \delta(s_i, w_{i+1})$ and $s_n \in F$.

Example 12. It is easy to specify a finite automaton A that accepts precisely the ASCII texts that contain the word 'combinatorics'. It is also easy to specify an automaton B that accepts precisely the ASCII texts that do *not* contain the word 'generating function'. It is a classic fact that for any two finite automata A and B there exists an automaton $A \times B$ (the so-called *product automaton*) that accepts a word w precisely when A accepts w *and* B accepts w. Hence, there exists a finite automaton that accepts precisely the ASCII texts that contain the word 'combinatorics' but that do not contain the word 'generating function'. The automaton would be too huge to be drawn here, so we present the automaton for two shorter words than *'combinatorics'* and *'generating function'* in the next example. △

Example 13. Following the strategy in the previous example, we explicitly construct an automaton that accepts all words that contain the subword 'abb', but not the subword 'aa'; see Figure 5.1. △

In this section we study the question: how many words of length n are accepted by a finite automaton A? Let us write a_n for this number. We first describe how to translate A into a so-called *regular expression*.

5.4.2. Regular expressions. A *regular expression* (over the alphabet Σ) is an expression that is defined inductively:

(1) $\emptyset$, ϵ, and all symbols from Σ are regular expressions.
(2) if e is a regular expression, then so is e^* (the *Kleene star*);
(3) if e_1, e_2 are regular expressions, then so are e_1e_2 (concatenation) and $e_1 \cup e_2$ (alternation).

A regular expression e over the alphabet Σ describes a formal language $L(e) \subseteq \Sigma^*$:

(1) $L(\emptyset) = \emptyset$ (the empty language).
(2) $L(\epsilon) = \{\epsilon\}$ (the language that just contains the empty word).
(3) $L(s) = \{s\}$ for any $s \in \Sigma$.
(4) $L(e^*) := \{w_1 \ldots w_n \mid w_1, \ldots, w_n \in L(e), n \in \mathbb{N}\}$.
(5) $L(e_1e_2) := \{w_1w_2 \mid w_1 \in L(e_1), w_2 \in L(e_2)\}$.
(6) $L(e_1 \cup e_e) := L(e_1) \cup L(e_2)$.

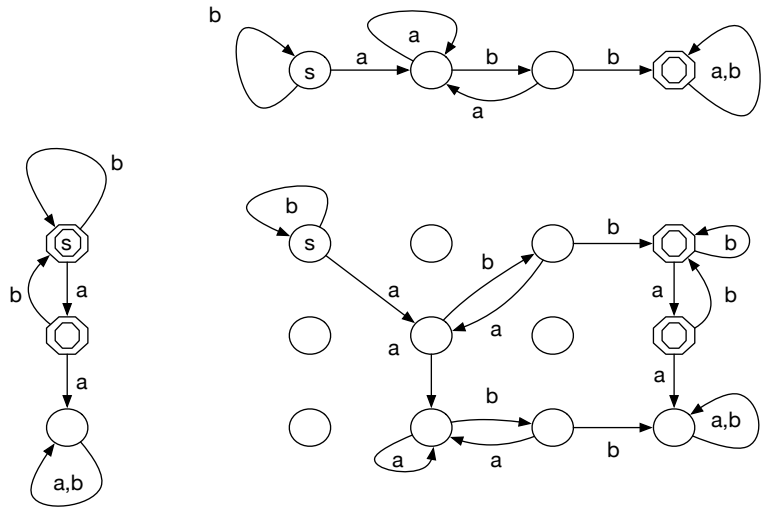

FIGURE 5.1. An illustration of the construction of an automaton accepting all words that contain 'abb', but not the subword 'aa'. Edges from states that can never be reached from the start vertex are not drawn since they don't matter anyway.

We also write $L_n(e)$ instead of $L(e) \cap \Sigma^n$. In general, a regular expression e might be *ambiguous* in the sense the same word $w \in L(e)$ can be derived in many different ways. For instance, if $e = (a \cup aa)^*$ then the word $aaaa \in L(e)$ can be 'parsed' in five different ways, indicated by the brackets as follows:

$$aaaa, aa(aa), (aa)aa, a(aa)a, (aa)(aa)$$

If every word in $L(e)$ has exactly one parse with respect to e, then we say that e is an *unambiguous regular expression*. For example, $e = (ab \cup abb)^*$ is unambiguous. Ambiguity for regular expressions can be defined formally by induction as follows[1]:

(1) $\emptyset$, ϵ, and all symbols in Σ are unambiguous.
(2) if e_1 and e_2 are unambiguous, and $|L_n(e_1 \cup e_2)| = |L_n(e_1)| + |L_n(e_2)|$ for all $n \in \mathbb{N}$, then $e_1 \cup e_2$ is unambiguous.
(3) if e_1 and e_2 are unambiguous, and $|L_n(e_1 e_2)| = \sum_{k=1}^{n-1} |L_k(e_1)| \cdot |L_{n-k}(e_2)|$, then $e_1 \cup e_2$ is unambiguous.
(4) if e is unambiguous, and $|L_n(e^*)| = 1 + |L_{k_1}(e)| + |L_{k_2}(ee)| + |L_{k_3}(eee)| + \cdots$ such that $k_1 + k_2 + \cdots = n$, then e^* is unambiguous.

PROPOSITION 5.4.1. *For every deterministic finite automaton A there exists an unambiguous regular expression e such that $w \in L(e)$ if and only if A accepts w.*

PROOF. Let $\{1, \ldots, n\}$ be the state space of A. We define an unambiguous regular expression $R(i, j, k)$ describing all words w that take state i to state j while using intermediate states 1 to k only:

- $R(i, j, 0)$ is $x_1 \cup x_2 \cup \ldots$ where $x_1, x_2, \ldots$ are the symbols x such that $\delta(i, x) = j$. Clearly, this expression is unambiguous.
- $R(i, i, 0)$ is defined similarly but including ϵ in the union.
- for $k > 0$ we define $R(i, j, k)$ to be

$$R(i, j, k-1) + R(i, k, k-1)R(k, k, k-1)^* R(k, j, k-1).$$

[1]Thanks to Florian Starke for pointing this out.

That is, any string that takes state i to state j using intermediate states up to k either goes from i to j without going through k, or can be divide into a word that goes from i to k, a word that goes from k back to itself, and then a word that goes from k to j. Again, it is easy to see that $R(i,j,k)$ is an unambiguous regular expression.

The entire language accepted by A can then be described as $\bigcup_{i\in F} R(1,i,n)$, which is again an unambiguous regular expression since any word w accepted by the deterministic automaton there is exactly one i such that $w \in L(R(1,i,n))$. □

We mention that conversely, it holds that for *every* regular expression e there exists a DFA that accepts precisely the words in $L(e)$ (unfortunately, in general the size of the DFA might be exponentially large). It follows that for every regular expression there exists an unambiguous regular expression that describes the same language.

5.4.3. The generating function of a regular language. Coming back to our original task, namely the task of finding a formula for the sequence $(a_i)_{i\in\mathbb{N}}$, we describe how to translate an unambiguous regular expression e into a generating function $A_e(x)$.

(1) $A_\emptyset(x) := 0$.
(2) $A_\epsilon(x) = 1$.
(3) If $s \in \Sigma$ then $A_s(x) := x$.
(4) $A_{e_1\cup e_2}(x) := A_{e_1}(x) + A_{e_2}(x)$.
(5) $A_{e_1e_2}(x) := A_{e_1}(x) \cdot A_{e_2}(x)$.
(6) if $e = e^*$ then $A_{e^*}(x) := \frac{1}{1-A_e}$.

LEMMA 5.4.2. *Let e be an unambiguous regular expression. Then*

$$A_e(x) = \sum_{n\in\mathbb{N}} |L_n(e)|x^n.$$

PROOF. The proof is more or less obvious from the definition of unambiguity. The most interesting translation step is perhaps item (6). To see that this is correct, recall that $\frac{1}{1-x} = 1 + x + x^2 + x^3 + \cdots$. Moreover, using composition of formal power series, we get $\frac{1}{1-A_e(x)} = 1 + A_e(x) + A_e(x)^2 + A_e(x)^3 + \cdots$. Since $A_e(x)^k$ is the generating function for words composed of k words in e, unambiguity gives us that $[x^n]\frac{1}{1-A_e(x)} = |L_n(e^*)|$. □

EXAMPLE 14. Let $e = (a^*b)^*$. Then we have

$$A_e = \frac{1}{1-A_{a^*b}(x)} = \frac{1}{1-\frac{x}{1-x}} = \frac{1-x}{1-2x} = \frac{1}{2} + \frac{1}{2(1-2x)}$$

and hence $a_n = 1$ if $n = 0$ and $a_n = 2^{n-1}$ for $n > 0$. △

In general, to obtain an explicit formula for a_n we can compute the partial fraction expansion (see Section 5.3.5).

EXAMPLE 15. Let $e = (aa|bb)^*$. Then we have

$$A_e = \frac{1}{1-A_{aa|bb}(x)} = \frac{1}{1-2x^2}.$$

The polynomial $1-2x^2$ can be factored as $(1-\sqrt{2}x)(1+\sqrt{2}x)$. For a partial fraction expansion, we are looking for a, b such that

$$(1+\sqrt{2}x)a + (1-\sqrt{2}x)b = 1$$

and thus $a+b=1$ and $a-b=0$. Therefore, $a=b=1/2$. We then have

$$A_e = \frac{1}{2}\left(\frac{1}{1-\sqrt{2}x} + \frac{1}{1+\sqrt{2}x}\right)$$

and hence

$$[x^n]A_e = \frac{1}{2}\left(\sqrt{2}^n - \sqrt{2}^n\right) = \begin{cases} 2^{n/2} & \text{for even } n \\ 0 & \text{for odd } n. \end{cases} \qquad \triangle$$

REMARK 5.4.3. It is straightforward to adapt Lemma 5.4.2 to obtain a generating function in several variables (also see Remark 5.3.1) that takes into account the number of a's, b's, etc. in the word.

5.5. Analytic Combinatorics

For certain values of $x \in \mathbb{C}$ a given power series $A(x) = \sum_{n\in\mathbb{N}} a_n x^n$ (say, with coefficients a_i in $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, or even $\mathbb{C}$) might *converge*; the set of all $x \in \mathbb{C}$ where $A(x)$ converges might tell us a lot about the asymptotic growth of the a_n. We recommend the following video for motivation of the use of complex numbers in combinatorics

https://www.youtube.com/watch?v=bOXCLR3Wric.

Some of the necessary concepts from calculus are introduced in Appendix A.

5.5.1. From formal power series to functions: convergence.

DEFINITION 5.5.1. Let $A(x) = \sum_{n\in\mathbb{N}} a_n x^n \in \mathbb{C}[[x]]$ and let $z \in \mathbb{C}$. We say that

- $A(z)$ *converges* if the limit $\lim_{n\to\infty} \sum_{i=0}^{n} a_i z^i \in \mathbb{C}$ exists; in this case, we also use the notation $\sum_{n\in\mathbb{N}} a_i z^i$ for the limit;
- $A(z)$ *diverges*, otherwise;
- $A(z)$ *converges absolutely* if $\lim_{n\to\infty} \sum_{i=0}^{n} |a_i| z^i \in \mathbb{C}$ exists.

Hence, if $S \subseteq \mathbb{C}$ is such that $A(z)$ converges for all $z \in S$, then $A(x)$ defines a function $S \to \mathbb{C}$ mapping $z \in S$ to $A(z)$. Absolute convergence is important for the study of infinite series because it occurs often but behaves nicely; in particular, rearrangements do not change the value of the limit. This is not true for convergent series in general:

$$\sum_{n\in\mathbb{N}} \frac{(-1)^n}{n+1} = 1 - \tfrac{1}{2} + \tfrac{1}{3} - \tfrac{1}{4} + \tfrac{1}{5} - \tfrac{1}{6} + \cdots = \ln 2 \qquad \text{(see (61) below)}$$

which is smaller than 1. However, the rearrangement[2]

$$1 + (-\tfrac{1}{2} + \tfrac{1}{3} + \tfrac{1}{5}) + (-\tfrac{1}{4} + \tfrac{1}{7} + \tfrac{1}{9}) + (-\tfrac{1}{6} + \tfrac{1}{11} + \tfrac{1}{13}) + \cdots$$

is larger than 1 since each of the three-term sums is positive.

We say that a series $\sum_{n\in\mathbb{N}} a_n$ is *unconditionally convergent* if all rearrangements of the series converge to the same value, i.e., for every permutation $\pi\colon \mathbb{N} \to \mathbb{N}$ we have that $\sum_{n\in\mathbb{N}} a_{\pi(n)}$ converges. Since the topology of $\mathbb{C}$ is complete, absolute convergence implies unconditional convergence. In fact, we even have equivalence, by the *Riemann rearrangement theorem*, but we do not need this in the further course. However, what *is* heavily used is the following.

LEMMA 5.5.2 (Weierstrass). *Any absolutely convergent series $\sum_{n\in\mathbb{N}} a_n = a$ in $\mathbb{C}$ is unconditionally convergent.*

[2]The recipe to form the triples is: enumerate the even numbers and the odd numbers in parallel; the first entry of the triple is $-\frac{1}{n}$ for the next even number, and the second and third entries of a triple are $\frac{1}{m}$ and $\frac{1}{m+2}$ for the next two odd numbers m and $m+2$. Clearly, this gives a bijection to the terms in (61).

PROOF. Let $\pi\colon \mathbb{N} \to \mathbb{N}$ be a permutation and let $\epsilon > 0$. By the absolute convergence of $\sum_{n\in\mathbb{N}} a_n$ we can find a $k \in \mathbb{N}$ such that $\sum_{n=k}^{\infty} |a_n| < \epsilon$. Let $m \in \mathbb{N}$ be such that $\{a_0, \dots, a_{k-1}\} \subseteq \{a_{\pi(1)}, \dots, a_{\pi(m)}\}$. Then for all $\ell \geq m$ we have

$$\begin{aligned}
\left| -\sum_{n=0}^{\ell} a_{\pi(n)} + \sum_{n\in\mathbb{N}} a_n \right| &\leq \left| \sum_{n=k}^{\infty} a_n \right| && (*)\\
&\leq \sum_{n=k}^{\infty} |a_n| && \text{(triangle inequality)}\\
&\leq \epsilon && \text{(by the choice of } k\text{)}.
\end{aligned}$$

To see why the inequality in line $(*)$ holds, observe that the terms $a_0, \dots, a_{k-1}$ appear in both $\sum_{n=0}^{\ell} a_{\pi(n)}$ and $\sum_{n\in\mathbb{N}} a_n$ and thus get cancelled (some more terms might get cancelled so we might not have equality). □

We recall one of the basic divergence and convergence tests.

LEMMA 5.5.3 ((d'Alembert's) ratio test). *Let $(a_n)_{n\in\mathbb{N}}$ be a sequence of real numbers and let*

$$\ell := \lim_{n\to\infty} \left| \frac{a_{n+1}}{a_n} \right|.$$

Then $\sum_{n\in\mathbb{N}} a_n$ converges absolutely if $\ell < 1$, and diverges if $\ell > 1$.

PROOF. Suppose that $\ell < 1$. Let $r := \frac{\ell+1}{2}$. Then $\ell < r < 1$ and there exists an $m \in \mathbb{N}$ such that $|a_{n+1}| < r|a_n|$ for every $n \geq m$. We conclude

$$\sum_{i\in\mathbb{N}} |a_{i+m}| < \sum_{i\in\mathbb{N}} r^i |a_m| = |a_m| \sum_{i\in\mathbb{N}} r^i = |a_m| \frac{1}{1-r} < \infty.$$

Now suppose that $\ell > 1$. Then there exists an $m \in \mathbb{N}$ so that $|a_{n+1}| > |a_n|$ for all $n \geq m$, and hence the series diverges by the term test. □

EXAMPLE 16. The ratio test implies that $\exp(x) = \sum_{n\in\mathbb{N}} \frac{1}{n!} x^n$ converges absolutely for all $z \in \mathbb{C}$, because

$$\lim_{n\to\infty} \left| \frac{n! z^{n+1}}{(n+1)! z^n} \right| = \lim_{n\to\infty} \left| \frac{z}{n+1} \right| = 0 < 1. \qquad \triangle$$

THEOREM 5.5.4 (Cauchy-Hadamard). *Let $A(x) \in \mathbb{C}[[x]]$. Then there exists a non-negative number $r \in \mathbb{R} \cup \{+\infty\}$, called the* radius of convergence, *such that the series $A(z)$ converges for all $z \in \mathbb{C}$ with $|z| < r$ and diverges for all $z \in \mathbb{C}$ with $|z| > r$. Moreover,*

$$\limsup_{n\to\infty} |a_n|^{1/n} = \frac{1}{r}$$

where $\frac{1}{+\infty}$ is meant to be 0 and $\frac{1}{0}$ is meant to be $+\infty$. In particular, $a_n \in O\big((1/r)^n\big)$.

PROOF. Let $z \in \mathbb{C}$ and

$$r := \frac{1}{\limsup_{n\to\infty} |a_n|^{1/n}}.$$

First suppose that $|z| < r$. We have to show that $A(z)$ converges. Pick $\epsilon > 0$ such that

$$|z| < \frac{r}{1+\epsilon r}$$

which is the case if and only if

$$\alpha := |z| \cdot (1/r + \epsilon) < 1.$$

Since $\limsup_{n\to\infty} |a_n|^{1/n} = 1/r$ there exists an $n \in \mathbb{N}$ such that for all $m > n$

$$|a_m|^{1/m} < 1/r + \epsilon.$$

Hence,

$$|a_m| \cdot |z|^m < \big(|z| \cdot (1/r + \epsilon)\big)^m = \alpha^m.$$

We have that $\sum_{m\in\mathbb{N}} \alpha^m$ converges absolutely by the ratio test (Lemma 5.5.3), and hence $A(z) = \sum_{n\in\mathbb{N}} a_n z^n$ converges absolutely, too.

Now suppose that $|z| > r$. We will show that $a_n z^n$ for $n \to \infty$ does not converge to 0, which implies in particular that $\sum_{n\in\mathbb{N}} a_n z^n$ diverges. Choose $\epsilon > 0$ such that $\theta := |z/r - \epsilon z| > 1$. Since $\limsup_{n\to\infty} |a_n|^{1/r} = \frac{1}{r}$ there exists an n such that for all $m \geq n$

$$|a_m|^{1/m} > \frac{1}{r} - \epsilon.$$

Hence,

$$|a_m z^m| > \left| \left(\frac{1}{r} - \epsilon \right) z \right|^m = \theta^m$$

which tends to ∞ for $m \to \infty$ since $\theta > 0$. The cases where $r \in \{0, \infty\}$ are similar and left to the reader. □

Theorem 5.5.4 shows how to obtain some asymptotic information about a_n from the radius of convergence of the corresponding series $A(x)$. Much more detailed information can be obtained via a method called *singularity analysis* which is beyond the scope of this lecture. An introduction to this technique can be found in [**16**] and Sections 5.7.5 and 5.7.6 give illustrations of what can be shown with it.

Example 17. The formal power series $\sum_{n\in\mathbb{N}} n! x^n \in \mathbb{Z}[[x]]$ diverges for every $x \in \mathbb{C} \setminus \{0\}$, and hence its radius of convergence is 0. Hence, Theorem 5.5.4 does not provide any interesting bound for the growth of the coefficients. An approach to analyse sequences that have such a fast growth will be presented in Section 5.7. △

For $r \in \mathbb{R}^+$ we write D_r for the set $\{x \in \mathbb{C} : |x| < r\}$, the open disc centered at 0.

Definition 5.5.5. Let $A \in \mathbb{C}[[x]]$ be a formal power series with radius of convergence $r \in \mathbb{R}^+$. Then $f_A \colon D_r \to \mathbb{C}$ defined by $z \mapsto A(z)$ is called the *function described by* A.

Recall that absolute convergence implies unconditional convergence (Lemma 5.5.2), so we can rearrange the order of summation in the proof of the following proposition.

Proposition 5.5.6. *Let $A, B \in \mathbb{C}[[x]]$ be power series with convergence radius $r \in \mathbb{R}^+$ and $s \in \mathbb{R}^+$, respectively; suppose that $r \leq s$. Then*

1. *$A(x)+B(x)$ has convergence radius at least r and describes the function $f_A + f_B$;*
2. *$A(x)\cdot B(x)$ has convergence radius at least r and describes the function $f_A f_B$.*

Proof. Concerning (1), let $z \in D_r$, and compute:

$$\begin{aligned} f_A(z) f_B(z) &= \sum_{k\in\mathbb{N}} a_k z^\ell + \sum_{\ell\in\mathbb{N}} b_\ell z^\ell && \text{(absolute convergence)} \\ &= \sum_{n\in\mathbb{N}} (a_n + b_n) z^n && \text{(Lemma 5.5.2))} \\ &= f_{A+B}(z) && \text{(by the definition of } A+B). \end{aligned}$$

Concerning (2), let $z \in D_r$, and compute:

$$\begin{aligned} f_A(z)f_B(z) &= f_A(z)\sum_{\ell\in\mathbb{N}} b_\ell z^\ell \\ &= \sum_{\ell\in\mathbb{N}} f_A(z) b_\ell z^\ell \\ &= \sum_{\ell\in\mathbb{N}}\sum_{k\in\mathbb{N}} a_k b_\ell z^{k\ell} && \text{(have absolute convergence)} \\ &= \sum_{n\in\mathbb{N}} \left(\sum_{k+l=n} a_k b_\ell\right) z^n && \text{(Lemma 5.5.2))} \\ &= f_{A\cdot B}(z) && \text{(by the definition of } A\cdot B). \end{aligned}$$ □

We finally also treat composition of power series analytically. Not surprisingly, composition of functions matches the definition of composition of formal power series in the following sense.

PROPOSITION 5.5.7. *Let $A, B \in \mathbb{C}[[x]]$ be power series with convergence radius $r \in \mathbb{R}^+$ and $s \in \mathbb{R}^+$, respectively; suppose that $r \leq s$. If $[x^0]B(x) = 0$ so that $A(B(x))$ is defined, then $A(B(x))$ has a positive radius of convergence $t \leq r$ and describes the function $f_A \circ f_B \colon D_t \to \mathbb{C}$.*

PROOF. Let $S := \{z \in D_s \mid B(z) \in D_r\}$. Note that this set is non-empty since $0 \in S$. Since f_B is continuous by Lemma 5.5.8, we have that S is the intersection of the open set D_s with the open pre-image of D_r under f_B, and hence open, and therefore contains D_t for some $t \in \mathbb{R}^+$. Let $z \in D_t$. Then

$$\begin{aligned} &f_A(f_B(z)) \\ &= \sum_{n\in\mathbb{N}} a_n \left(\sum_{k\in\mathbb{N}} b_k z^k\right)^n && \text{(have absolute convergence)} \\ &= \sum_{n\in\mathbb{N}} a_n \sum_{\substack{k\in\mathbb{N}, m_1,\ldots,m_k\in\mathbb{N}^+,\\ m_1+\cdots+m_k=n}} |b_{m_1}\cdots b_{m_k}| z^n && \text{(use Prop. 5.5.6 (2) several times)} \\ &= f_{A(B(x))}(z) && \text{(by definition of } A(B(x))). \end{aligned}$$ □

A function $f \colon D_r \to \mathbb{C}$ is called *holomorphic* if f is (complex-) differentiable at every point $z \in D_r$, i.e.,

$$f'(z) := \lim_{a\to z} \frac{f(a) - f(z)}{a - z}$$

exists (by purpose, we use the same notation as for the formal derivative in (5.3.3)). The next lemma shows that if $A \in \mathbb{K}[[x]]$ has a positive radius of convergence $r \in \mathbb{R}^+$, then the function $f_A \colon D_r \to \mathbb{C}$ is holomorphic (and in particular continuous). The derivative f'_A is given by the formal derivative A'; in other words, the derivative can be computed 'term by term'. The proof here is slightly more difficult than the easy facts from Proposition 5.5.6.

LEMMA 5.5.8. *Suppose that $A = \sum_{n\in\mathbb{N}} a_n x^n \in \mathbb{C}[[x]]$ has convergence radius $r \in \mathbb{R}$ and let $f_A \colon D_r \to \mathbb{C}$ be given by $z \mapsto A(z)$. Then $A'(x)$ has radius of convergence r and describes f'_A.*

PROOF. To prove that $A'(z)$ converges for $z \in D_r$, pick $s \in \mathbb{R}$ such that

$$|z| < s < r.$$

Since $A(s)$ converges, the terms $|a_n s^n|$ are bounded above, say by b. Then

$$n|a_n|z^n = n|a_n s^n|(z/s)^n \leq nb(z/s)^n.$$

The series $\sum_{n\in\mathbb{N}} nb(z/s)^n$ converges absolutely by the ratio test. Therefore, by the comparison test, the series $A'(z) = \sum_{n\in\mathbb{N}^+} n|a_n|z^{n-1}$ converges absolutely. It follows that $A'(x)$ has the same radius of convergence as $A(x)$.

To prove that $f'_A(z) = A'(z)$ we use the identity

$$\begin{aligned}(p-q)(p^{n-1}+p^{n-2}q+\cdots+pq^{n-2}+q^{n-1}) &= p^n - p^{n-1}q + p^{n-1}q - \cdots + q^n \\ &= p^n - q^n \end{aligned} \tag{55}$$

to compute

$$\begin{aligned} f'_A(z) = \lim_{a\to z}\frac{f(a)-f(z)}{a-z} &= \lim_{a\to z}\sum_{n\in\mathbb{N}}\frac{a_n(a^n-z^n)}{a-z} \\ &= \lim_{a\to z}\sum_{n\in\mathbb{N}} a_n(a^{n-1}+a^{n-2}z+\cdots+az^{n-2}+z^{n-1}) \end{aligned} \tag{56}$$

This already looks quite close to $\sum_{n\in\mathbb{N}} a_n n z^{n-1}$, but in general we cannot just exchange the order of a limit and an infinite sum. So we choose the following approach: it suffices to prove that

$$\lim_{a\to z}\frac{f_A(a)-f_A(z)}{a-z} - f_{A'}(z) = 0 \tag{57}$$

the advantage being that the left hand side can be broken into three parts that can be analysed separately. Write $S_k(x)$ for the polynomial $\sum_{n=1}^{k} a_n x^n$ and $E_k(x)$ for the series $\sum_{n=k+1}^{\infty} a_n x^n$ so that $A(x) = S_k(x) + E_k(x)$ (and we think of $S_k(x)$ as an approximation to $A(x)$ and the $E_k(x)$ as the error we make with the approximation). The expression in (57) can now be rewritten:

$$\frac{E_k(a)-E_k(z)}{a-z} + \left(\frac{S_k(a)-S_k(z)}{a-z} - S'_k(z)\right) + (S'_k(z) - f_{A'}(z)). \tag{58}$$

To show that this term goes to 0 as a tends to z, let $\epsilon > 0$ be given. The first term in (58) can be analysed as above: since $|z| < s$ and $a \to \infty$ the triangle inequality will give us eventually

$$\left|\frac{E_k(a)-E_k(z)}{a-z}\right| = \left|\sum_{n=k+1}^{\infty} a_n(a^{n-1}+a^{n-2}z+\cdots+z^{n-1})\right| \leq \sum_{n=k+1}^{\infty}|a_n|ns^{n-1}. \tag{59}$$

Note that this is just the tail $A'(s)$ which converges (Lemma A.1.4), so this term must approach 0 for $k \to \infty$. That is, we can find a $k_1 \in \mathbb{N}$ and a $\delta_1 > 0$ such that for all $k > k_1$ and for all a such that $|a - z| < \delta_1$ we have that (59) $\leq \epsilon/3$.

The other two terms are even easier to bound: $S_k(x)$ is a polynomial, so

$$\lim_{a\to z}\frac{S_k(a)-S_k(z)}{a-z} - S'_k(z) = 0.$$

In other words, we can find a $\delta_2 > 0$ such that for all a with $|a - z| < \delta_2$ the second term is smaller than $\epsilon/3$.

For the third term, we know that $S'_k(z) \to f_{A'}(z)$ as $k \to \infty$ because $A'(z)$ converges absolutely for $|z| < r$ and $S'_k(z)$ is just the k-th partial sum of this power series. Hence, we can find a $k_3 \in \mathbb{N}$ and a $\delta_3 > 0$ such that for all $k > k_3$ and a with $|a - z| < \delta_3$ we have $S'_k(z) - f_{A'}(z) < \epsilon/3$. Now select any $k_0 > \max(k_1, k_3)$ and $\delta := \min(\delta_1, \delta_2, \delta_3)$ and the triangle inequality gives us that the expression in (58) is smaller than ϵ, which is what we wanted to show. □

REMARK 5.5.9. A posteriori, the deeper reason why the limit and the infinite sum in (56) can be exchanged is that the polynomials $S_k(x)$ converge *uniformly* against $A(x)$ within D_s; we do not further elaborate on this, but refer to more advanced calculus classes.

5.5.2. From functions to power series: Taylor expansion. The correspondence between power series and functions is a two-way correspondence; when we talk about generating functions, we always have this double perspective in mind. The following definition is the central notion to produce a power series from a function.

DEFINITION 5.5.10. Let f be a real-valued or complex-valued function that is infinitely differentiable at a real or complex number a. Then the *Taylor expansion of f at a* is the power series

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n.$$

In the following we work mostly in a setting where the correspondence between functions and power series is simplest to describe, namely for functions over the complex numbers. One of the major theorems in complex analysis says that every holomorphic function f is *analytic*, i.e., can be expressed as a converging power series (with a positive radius of convergence). We do not need this result in this course (since in analytic combinatorics we *start* with the power series and analyse it using holomorphic functions, not the other way round) but we want to mention the theorem here to give the full picture.

THEOREM 5.5.11. *Let $r \in \mathbb{R}^+$ and let $f\colon D_r \to \mathbb{C}$ be a holomorphic function. Then for all $z \in D_r$*

$$f(z) = \sum_{n\in\mathbb{N}} \frac{f^{(n)}(0)}{n!} z^n$$

that is, $f(z)$ equals the Taylor series of f at 0 evaluated at z.

PROOF. This is outside the scope of this course, and publicity for the complex analysis courses offered at TU Dresden ... □

Why do we need to work over the complex numbers? Well, a naive variant of the statement for the real numbers is simply wrong, as the following example shows.

EXAMPLE 18. There are non-constant functions $f\colon \mathbb{R} \to \mathbb{R}$ that are infinitely often differentiable at each $z \in D_1$, but their Taylor expansion at 0 is the series that is identically 0. Namely, we consider the function given by

$$z \mapsto \begin{cases} 0 & \text{if } z = 0 \\ e^{-\frac{1}{z^2}} & \text{otherwise.} \end{cases}$$

Note that if we consider the function defined by the same expression over the complex numbers instead, then by choosing $a \in \mathbb{R}^+$ sufficiently small the value of $f(ai) = e^{\frac{1}{a^2}}$ can be made arbitrarily large, and hence $\lim_{a\to 0} \frac{f(ai)-f(0)}{ai-0}$ does not exist. Therefore, $f\colon \mathbb{C} \to \mathbb{C}$ is not holomorphic. △

Lemma 5.5.12 below implies that, informally, when we start from a formal power series $A(x)$ with positive radius of convergence r and turn it into a function $f_A\colon D_r \to \mathbb{C}$, then the Taylor expansion of f_A gives back $A(x)$.

LEMMA 5.5.12. *Let $A(x) = \sum_{n\in\mathbb{N}} a_n x^n \in \mathbb{C}[[x]]$ be a formal power series with a positive radius of convergence $r \in \mathbb{R}^+ \cup \{+\infty\}$. Then the function $f_A \colon D_r \to \mathbb{C}$ given by $z \mapsto A(z)$ is holomorphic, and*

$$a_n = \frac{f_A^{(n)}(0)}{n!}. \tag{60}$$

PROOF.

$$\begin{aligned} a_n &= \frac{A^{(n)}(0)}{n!} && \text{(see (52))} \\ &= \frac{f_A^{(n)}(0)}{n!} && \text{(Lemma 5.5.8)} \end{aligned}$$

and hence A is the Taylor expansion of f_A at 0. □

COROLLARY 5.5.13. *Let $r \in \mathbb{R}^+$, let $f, g \colon D_r \to \mathbb{C}$ be holomorphic, and let $A, B \in \mathbb{C}[[x]]$ be the Taylor expansions of f and g at 0. Then*

(1) $f + g$ is holomorphic and has the Taylor expansion $A(x) + B(x)$ at 0.
(2) fg is holomorphic and has the Taylor expansion $A(x) \cdot B(x)$ at 0.
(3) If $g(0) = 0$, then $f \circ g$ is holomorphic with Taylor expansion $A(B(x))$ at 0.

PROOF. The first fact can easily be proven directly: it is easy to verify that $f+g$ is holomorphic, and by Theorem 5.5.11 it equals its Taylor expansion at 0, which is

$$\sum_{n\in\mathbb{N}} \frac{(f+g)^{(n)}(0)}{n!} x^n = \sum_{n\in\mathbb{N}} \frac{f^{(n)}(0) + g^{(n)}(0)}{n!} x^n = A(x) + B(x).$$

Here is an alternative argument: we have $f + g = f_A + f_B = f_{A+B}$ by Proposition 5.5.6, and this function is holomorphic and has the Taylor expansion $A(x)+B(x)$ by Lemma 5.5.8. The advantage of this argument is that the second statement can be shown in complete analogy (while it seems more tedious to me to determine the Taylor expansion of $f \cdot g$). Also the third statement can be shown in the same way: we have $f \circ g = f_A \circ f_B = f_{A\circ B}$ by Proposition 5.5.7, which is holomorphic with Taylor expansion $A(B(x))$ by Lemma 5.5.8. □

We will illustrate the use of the correspondence between power series and real-valued functions by Newton's theorem (Theorem 5.5.16 below). The following fact about polynomials is well-known to the reader.

THEOREM 5.5.14 (Binomial theorem). *Let $z \in \mathbb{N}$ and $x \in \mathbb{C}$ (or any other field instead of $\mathbb{C}$). Then*

$$(1+x)^z = \sum_{k=0}^{z} \binom{z}{k} x^k \,.$$

We would like to generalise the binomial theorem to the situation where the *exponent* z is a real number. Note that for $n \in \mathbb{N}$ the expression $(1+x)^n$ is a well-defined formal power series (even a polynomial), but for non-integer z we have not defined $(1+x)^z$ as a formal power series. However, $x \mapsto (1+x)^z$ is certainly for every $z \in \mathbb{R}$ a well-defined function from $D_1 \to \mathbb{C}$.

We recall the definition of b^z for $b, z \in \mathbb{C}$, $b > 0$ for the reader:

- For $z = -n$ with $n \in \mathbb{N}$, we define $b^z := (b^n)^{-1}$.
- For $z = \frac{p}{q}$ we define $b^z := \left(b^{\frac{1}{q}}\right)^p$ where $b^{\frac{1}{q}}$ is the unique real r such that $r^q = b$.

- For $z \in \mathbb{R}$ and positive b, we choose an arbitrary sequence $e_1, e_2, \dots$ of rational numbers with $\lim_n e_n = z$, and define
$$b^z := \lim_{n\to\infty} b^{e_n}.$$

We also generalise the binomial coefficients.

Definition 5.5.15. Let z be a real (or complex) number, and $k \in \mathbb{N}$. Define
$$\binom{z}{k} := \frac{z(z-1)(z-2)\cdots(z-k+1)}{k!}\,.$$
In particular, $\binom{r}{0} = 1$.

Theorem 5.5.16 (Newton). *Let $x \in \mathbb{C}$ with $|x| < 1$ and $z \in \mathbb{R}$. Then*
$$(1+x)^z = \sum_{n\in\mathbb{N}} \binom{z}{n} x^n\,.$$

Proof. The function $f\colon \mathbb{C} \to \mathbb{C}$ given by $x \mapsto (1+x)^z$ is holomorphic, and from calculus courses you know that the derivative is $z(1+x)^{z-1}$ (the proof for holomorphic functions is analogous to the corresponding proof for real-valued functions). The n-th derivative of f is
$$z(z-1)(z-2)\cdots(z-n+1)(1+x)^{z-n}.$$
Hence, the Taylor expansion of f at 0 is
$$\sum_{n\in\mathbb{N}} \frac{f^{(n)}(0)}{n!} x^n = \sum_{n\in\mathbb{N}} \binom{z}{n} x^n.$$
The statement now follows from Theorem 5.5.11. □

Example 19. The Taylor expansion of $\sqrt{1-x}$ is
$$\sum_{n\in\mathbb{N}} \binom{1/2}{n} (-x)^n = 1 - \frac{1}{2}x - \frac{1}{8}x^2 - \cdots \qquad \triangle$$

The *natural logarithm* is the function $\ln\colon \mathbb{R}^+ \to \mathbb{R}$ which maps $x > 0$ to the unique element $y \in \mathbb{R}$ such that $e^y = x$. Applying the chain rule from calculus to $\ln(e^y) = y$ gives $\ln'(e^y)e^y = 1$, and substituting e^y by x gives $\ln'(x) = \frac{1}{x}$.

The function $(-1,1) \to \mathbb{R}\colon x \mapsto \ln(1+x)$ is infinitely often differentiable in 0 and we can compute its Taylor expansion: we have
$$\ln(1+x)' = \frac{1}{1+x} = \sum_{n\in\mathbb{N}} (-1)^n x^n$$
$$\text{and } \ln(1+x)^{(n)}(0) = (-1)^{n-1}(n-1)! \qquad \text{for } n \geq 1$$
and therefore the Taylor expansion of $\ln(1+x)$ at 0 is
$$\ln(1+x) = \sum_{n\in\mathbb{N}} \frac{\ln(1+x)^{(n)}(0)}{n!} x^n = \sum_{n\in\mathbb{N}^+} \frac{(-1)^{n-1}}{n} x^n \tag{61}$$
$$= x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots$$
This series converges for all $z \in \mathbb{C}$ with $|z| < 1$ and diverges for $z = -1$, so it has radius of convergence 1. The function from $D_r \to \mathbb{C}$ described by the series is holomorphic (Lemma 5.5.12) and extends the function $\ln(1+x) : (-1,1) \to \mathbb{R}$.

We will now show that exponentiation and logarithm are also inverses of each other as formal power series. First, note that the composition $\exp(\ln(1+x))$ is well-defined as a formal power series since $[x^0]\ln(1+x) = 0$.

PROPOSITION 5.5.17. $\exp(\ln(1+x)) = 1+x$.

PROOF SKETCH. We use the chain rule (54) to compute the derivative of the left-hand side of the equation:

$$\exp(\ln(1+x))' = \exp(\ln(1+x)) \cdot \frac{1}{1+x}.$$

So $\exp(\ln(1+x))$ is a solution for H in the differential equation

$$H' = \frac{H}{1+x}$$

with the initial condition $H(0) = \exp(\ln(1)) = 1$. Clearly, $H(x) = 1+x$ is a solution to the same differential equation. There is a formal version of the Picard-Lindelöf (aka Cauchy-Lipschitz) theorem which asserts the uniqueness of solutions to such ordinary differential equations, and we conclude that $\exp(\ln(1+x)) = 1+x$. □

How about the converse, i.e., inverting $\exp(x)$ with the help of a logarithm? Note that we have not defined $\ln(x)$ as a formal power series, but we may define $\ln(\exp(x))$ as the composition $\ln(\exp(x)) := \ln(1 + (\exp(x) - 1))$ which is well-defined.

PROPOSITION 5.5.18. $\log(\exp(x)) = \ln(1 + (\exp(x) - 1)) = x$.

PROOF. We proceed similarly as in Corollary 5.5.17.

$$\begin{aligned}\ln(1 + (\exp(x) - 1))' &= \ln'(1 + (\exp(x) - 1))(\exp(x) + 1)' \quad \text{(by the chain rule (54))} \\ &= \frac{\exp'(x)}{\exp(x)} = 1\end{aligned}$$

So $\log(\exp(x))$ is a solution to the ordinary differential equation $H(x)' = 1$ with the initial condition $H(0) = \ln(\exp(0)) = \ln(1) = 0$. The obvious unique solution to this differential equation is $H(x) = x$ and so we must have $\ln(\exp(x)) = x$. □

5.6. The Catalan Numbers

A *binary tree* is defined recursively:

- the empty set is a binary tree, and
- if B_1 and B_2 are binary trees, then (B_1, B_2) is a binary tree.

Let B be a binary tree. Then the *size* $|B|$ of B is also defined recursively: the binary tree $\emptyset$ has size 0, and (B_1, B_2) has size $|B_1| + |B_2| + 1$. We can associate in the obvious way to each binary tree B a tree in the sense of graph theory (Section 1.4): if B has size n, then this tree has $2n+1$ vertices, and each vertex has either degree 1, 2, or 3. Instead of formally defining the translation, we simply illustrate this simple connection by a picture in Figure 5.2.

There is at most one vertex which can have degree two, and this vertex is called the *root* of the tree. Such a degree-2 vertex does not exist precisely if the tree has only one vertex, and in this special case the root is defined to be this vertex. Conversely, every rooted tree with the properties above can be naturally associated to a binary tree.

5.6.1. A recursion formula. Directly from the definition of binary trees we obtain a formula for the number $c(n)$ of all binary trees of size $n \in \mathbb{N}$:

$$c(0) = 1 \tag{62}$$

$$c(n+1) = \sum_{i=0}^{n} c(i)c(n-i) \tag{63}$$

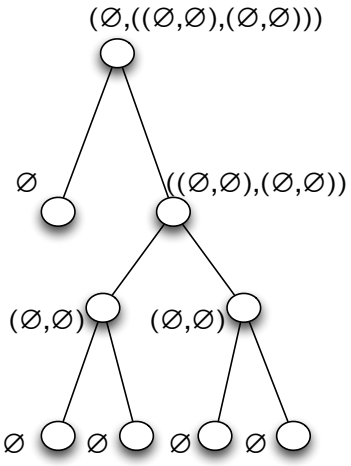

FIGURE 5.2. A binary tree and the corresponding rooted tree.

In this way, the values of $c(n)$ can be computed relatively quickly. We show the values for $n \in \{0, \ldots, 9\}$.

n	0	1	2	3	4	5	6	7	8	9
$c(n)$	1	1	2	5	14	42	132	429	1430	4862

These initial terms suffice to match our sequence with sequence A000108 in the *Online Encyclopedia of Integer Sequences*, the so-called *Catalan Numbers*.

5.6.2. Generating trees uniformly at random. The recursion formula (69) can be used to generate a binary tree of a given size n *uniformly at random*, that is, every binary tree of size n should be generated with equal probability. Random generation of combinatorial objects have several applications. For example, you might empirically test a conjecture concerning typical properties of binary trees (such as the depth of the tree, etc), by generating many of those objects for large n and verify the property of interest.

To generate a binary tree of size n uniformly at random, we follow the recursive definition of binary trees:

- if we have to generate a binary tree of size 0, this tree must be the binary tree $\emptyset$, and there is no randomness involved: we simply output $\emptyset$.
- If we have to generate a binary tree of size $n \geq 1$, then this tree must have the shape (B_1, B_2). We want to first generate B_1 and B_2 recursively. To do this, we have to make a decision: what should be the size i of B_1? The size of B_2 is fixed by this decision: it must be $n - i - 1$.

There are exactly $c_i c_{n-i}$ binary trees (B_1, B_2) of size $n+1$ where $|B_1| = i$, for $i \in \{0, \ldots, n\}$. The fraction of binary trees within the set of all binary trees of size $n+1$ is therefore exactly

$$p := \frac{c_i c_{n-i}}{c_{n+1}} .$$

Our algorithm chooses this particular i with probability p (this is a primitive operation in most programming languages). Since the Catalan numbers can be computed efficiently, the value of p can be computed efficiently as well. To summarise, we

Procedure BTree.
Input: $n \in \mathbb{N}$.
Output: a binary tree of size n uniformly at random.
If $n = 0$ return $\emptyset$ aus.
If $n > 0$, compute $c_1, \ldots, c_n$.
Choose an integer from $\{1, \ldots, c_n\}$ uniformly at random.
Set $b := 0$.
Loop over $i = 0$ to $n - 1$:
    If $p \in \{b, \ldots, b + c_i c_{n-i-1}\}$
        Return $(\mathrm{BTree}(i), \mathrm{BTree}(n - i - 1))$.
    $b := b + c_i c_{n-i-1}$.

FIGURE 5.3. Generation of a binary tree of size n, drawn uniformly at random, with the recursive method.

show in Figure 5.3 the entire algorithm in pseudocode. This method works for many combinatorial classes, and is called the *recursive method*.

5.6.3. A closed expression. An even faster method to compute the values of c_n is the following theorem.

THEOREM 5.6.1 (Euler 1751). *For all* $n \in \mathbb{N}$

$$c_n = \frac{1}{n+1}\binom{2n}{n}.$$

Theorem 5.6.1 can be proved by specifying an appropriate bijection (between certain words) and binary trees; see [**17**]. This bijection is also the key to more efficient methods to generate binary trees uniformly at random. Here we will present a different, algebraic proof (historically the first) of Theorem 5.6.1.

PROOF OF THEOREM 5.6.1. **Step 1.** We consider the power series

$$C(x) := \sum_{n=0}^{\infty} c_n x^n.$$

Step 2. Multiplying the left hand side of (69) by x^n and summing over all $n \in \mathbb{N}$ we obtain

$$\sum_{n\in\mathbb{N}} c_{n+1}x^n = \frac{\sum_{n\in\mathbb{N}} c_{n+1}x^{n+1} + 1 - 1}{x} = \frac{\sum_{n\in\mathbb{N}} c_n x^n - 1}{x} = \frac{C(x) - 1}{x}$$

Doing the same with the right hand side of (69) yields

$$\sum_{n=0}^{\infty}\sum_{i=0}^{n} c_i c_{n-i} x^n = C(x)^2$$

by the very definition of the Cauchy product $C(x)^2 = C(x)C(x)$. Hence,

$$\frac{C(x) - 1}{x} = C(x)^2$$

and reformulating we obtain

$$xC(x)^2 - C(x) + 1 = 0 \tag{64}$$

Step 3. We can now solve for $C(x)$:

$$C(x) = \frac{1 \pm \sqrt{1-4x}}{2x}. \tag{65}$$

If x tends to 0, then $C(x)$ tends to 1, but the expression $\frac{1+\sqrt{1-4x}}{2x}$ tends to infinity. So we choose the solution

$$\frac{1 - \sqrt{1-4x}}{2x}.$$

This is already sufficient to determine the asymptotic growth of the sequence $(c_n)_{n\in\mathbb{N}}$: $C(x)$ converges for $|x| < \frac{1}{4}$ and diverges for $x = \frac{1}{4}$, so Theorem 5.5.4 implies that $c_n \in O(4^n)$.

Step 4. The obtain a precise formula for c_n, we continue as follows.

$$\begin{aligned}
C(x) &= \frac{1-\sqrt{1-4x}}{2x} \\
&= \frac{1}{2x} - \frac{1}{2x}(1-4x)^{1/2}\big) && \text{(rewriting)} \\
&= \frac{1}{2x} - \frac{1}{2x}\sum_{n\in\mathbb{N}} \binom{1/2}{n}(-4x)^n && \text{(Newton, Theorem 5.5.16)} \\
&= \frac{1}{2x} + \sum_{n\in\mathbb{N}^+} \binom{1/2}{n}\frac{(-4)^n}{-2}x^{n-1} + \binom{1/2}{0}\frac{(-4)^0}{-2}x^{-1} && \text{(rewriting)} \\
&= \sum_{n\in\mathbb{N}} \binom{1/2}{n+1}\frac{(-4)^{n+1}}{-2}x^n && \text{(simplifying, index change).}
\end{aligned}$$

We now read off the coefficients and obtain that

$$\begin{aligned}
c_n &= \binom{1/2}{n+1}\frac{(-4)^{n+1}}{-2} \\
&= \frac{\frac{1}{2}(\frac{1}{2}-1)(\frac{1}{2}-2)\cdots(\frac{1}{2}-n)}{(n+1)!}\cdot\frac{(-4)^{n+1}}{-2} && \text{(Definition 5.5.15)} \\
&= \frac{(2\cdot 1-1)\cdot(2\cdot 2-1)\cdots(2(n-1)-1)\cdot(2n-1)}{(n+1)!}\cdot\frac{2^{n+1}}{2} && \text{(rewriting)} \\
&= \frac{1}{n+1}\frac{(2n-1)\cdot(2n-3)\cdots 3\cdot 1}{n!}\cdot 2^n && \text{(rewriting)} \\
&= \frac{1}{n+1}\frac{2n(2n-1)(2n-2)\cdots 3\cdot 2\cdot 1}{n!n!} && \text{(rewriting)} \\
&= \frac{1}{n+1}\binom{2n}{n} && \text{(by definition).} \quad \square
\end{aligned}$$

Note that we can use Stirling's formula (Section 5.7.6) to obtain formulas for the *asymptotic growth* of the Catalan numbers. Another proof of Theorem 5.6.1 based on the so-called Lagrange inversion formula can be found in Section 5.8.4.

5.6.4. Further correspondences. There are many classes of combinatorial objects where the objects with n elements are in bijective correspondence with binary trees of size n.

- We have defined convex n-gons already in Section 4.5.2. Every convex n-gon $\{p_1, \ldots, p_n\}$ has exactly $c(n)$ partitions of the convex hull of $p_1, \ldots, p_n$ in triangles with endpoints from $p_1, \ldots, p_n$ (also called *triangulations*). In Figure 5.4 are depicted $c(3) = 5$ partitions of a convex 5-gon.

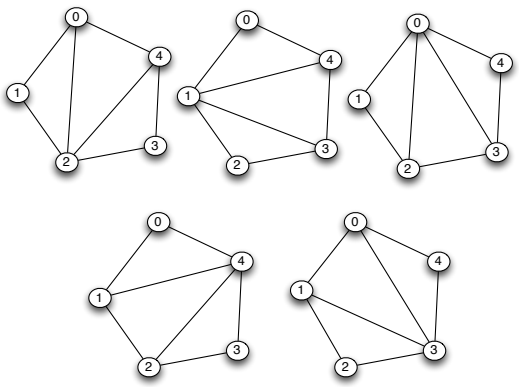

FIGURE 5.4. The five possible triangulations of a convex 5-gon.

- There are $c(n)$ possible bracketings of a product with $n+1$ factors (and n multiplications). For example, we have $c(3) = 5$ such bracketings of the factors $x_1, \ldots, x_4$:

$$(x_1x_2)(x_3x_4),$$
$$(x_1(x_2x_3))x_4,$$
$$x_1((x_2x_3)x_4),$$
$$((x_1x_2)x_3)x_4,$$
$$\text{and } x_1(x_2(x_3x_4))$$

- There are $c(n)$ sequences $a_1, \ldots, a_{2n}$ with $a_i \in \{1, -1\}$ such that $\sum_{i=1}^{2n} a_i = 0$ and $\sum_{i=1}^{k} a_i \geq 0$ for all $k \leq n$. Here we have for $n = 3$ precisely the following five sequences

$$(1, 1, 1, -1, -1, -1)$$
$$(1, 1, -1, 1, -1, -1)$$
$$(1, 1, -1, -1, 1, -1)$$
$$(1, -1, 1, 1, -1, -1)$$
$$(1, -1, 1, -1, 1, -1).$$

Exercises.

(86) Find a formula for the number of ways to write a non-commutative non-associative product of n-terms. For example, there are 12 ways to write a product of 3 terms, namely

$$(ab)c, (ac)b, (ba)c, (bc)a, (ca)b, (cb)a, a(bc), a(cb), b(ac), b(ca), c(ab), c(ba).$$

(87) Prove that there are $(2n-3)!!$ ways to write a commutative non-associative product of n-terms. That is, we identify $(ab)c$ with $(ba)c$, $c(ab)$, and $c(ba)$. Hence, for $n = 1$ and $n = 2$ there is only one way, for $n = 3$ there are three ways, and for $n = 4$ there are 15 ways.

5.7. Exponential Generating Functions

The generating functions that we have seen so far are also called *ordinary generating functions*, to distinguish them from other generating functions that exist, such

as *exponential generating functions* or *Dirichlet generating functions*. Depending on the enumeration problem that we want to study, those might be more appropriate. To give you an idea how to vary what we have learned about ordinary generating functions, we discuss the equally important exponential generating functions and their applications in this section.

5.7.1. Labelled enumeration. How many graphs are there with n vertices? This question needs to be formulated more carefully, since we have not yet specified the vertex set of the trees, so phrased like this the answer would be: infinitely many. There are two different ways to turn this question into a meaningful question.

The first is: how many graphs with n vertices are there *up to isomorphism*? That is, we want to count the number of equivalence classes of the equivalence relation induced by graph isomorphism (Definition 1.1.2) on the class of all trees. Clearly, for each n this number u_n will be finite. We also say that u_n is the number of *unlabelled graphs* with n vertices.

The second way is to count the number of graphs with the vertex set $[n]$. This number can be bigger than u_n and will be denoted by l_n. For instance, for $n = 3$ we have $u_3 = 4$, but $l_3 = 8$. We say that l_n is the number of *labelled graphs* with n vertices.

The same distinction can be made for counting the structures of size n in various classes of combinatorial objects, like hypergraphs, directed graphs, etc. There is no general simple way to translate between the two settings. For illustration, we describe two extreme situations.

Example 20. There is precisely one labelled independent set of size 1, and there is precisely one independent set of size 1 up to isomorphism; so for counting the number of independent sets, there is no difference between the labelled and the unlabelled count. △

Example 21. Up to isomorphism, there is precisely one linear order with n elements. However, there are $n!$ many linear orders on the set $[n]$. So, for linear orders we have the maximal possible difference between the labelled and the unlabelled count. △

Typically, for labelled enumeration problems the *exponential generating function* is more appropriate.

5.7.2. The exponential generating function. The *exponential generating function (EGF)* of a sequence $(a_n)_{n\in\mathbb{N}}$ is the formal power series

$$A(x) = \sum_{n\in\mathbb{N}} a_n \frac{x^n}{n!}.$$

We can recover the coefficients of an EGF $A(x)$ in the obvious way, by the formula

$$a_n = n! \cdot [x^n]A(x).$$

Example 22. For every $n \in \mathbb{N}$, the number of linear orders on $[n]$ is $n!$, and the corresponding EGF is

$$\sum_{n\in\mathbb{N}} \frac{n!}{n!} x^n = \frac{1}{1-x}.$$ △

Example 23. For every $n \in \mathbb{N}$, the number of directed cycles on $[n]$ is $(n-1)!$: we choose a first element on the cycle, then from the remaining elements we choose

the next, and so on. In this way we have counted each directed cycle n times since we do not care where the cycle starts. The EGF for these numbers is

$$\sum_{n\in\mathbb{N}} \frac{(n-1)!}{n!}x^n = \sum_{n\in\mathbb{N}} \frac{1}{n}x^n = \ln\frac{1}{1-x}.$$

Concerning the last equation, observe that $\ln\frac{1}{1-x} = \ln(1) - \ln(1-x) = -\ln(1-x)$ and recall from (61) that $\ln(1+x) = \sum_{n\in\mathbb{N}^+} \frac{(-1)^{n-1}}{n}x^n$, so

$$\ln(1-x) = \sum_{n\in\mathbb{N}^+} \frac{-1}{n}x^n$$

and the equation follows. △

Example 24. For every $n \in \mathbb{N}$, the number of cliques on $[n]$ is 1. The EGF for the sequence $1 = a_0 = a_1 = \cdots$ is

$$\sum_{n\in\mathbb{N}} \frac{1}{n!}x^n = \exp(x).$$ △

5.7.3. Dictionary for labelled combinatorial constructions. Clearly, if $A(x)$ is the EGF for $(a_n)_{n\in\mathbb{N}}$ and $B(x)$ is the EGF for $(b_n)_{n\in\mathbb{N}}$, then $A(x)+B(x)$ is the EGF for $(a_n+b_n)_{n\in\mathbb{N}}$. This formula is useful when we want to count classes of combinatorial objects that are composed from two *disjoint* classes of structures.

5.7.3.1. *Multiplication.* Recall that

$$[x^n]A(x)B(x) = \sum_{k\in\{0,\dots,n\}} \frac{a_k}{k!}\frac{b_{n-k}}{(n-k)!}$$

which shows that $A(x)B(x)$ is the EGF for

$$\left(\sum_{k\in\{0,\dots,n\}} \binom{n}{k} a_k b_{n-k}\right)_{n\in\mathbb{N}}. \tag{66}$$

As we will see, in many contexts this sequence has an interesting combinatorial interpretation. Instead of formulating a general lemma we will illustrate this by examples later.

5.7.3.2. *Differentiation.* Note that

$$A'(x) = \sum_{n\in\mathbb{N}^+} na_n\frac{x^{n-1}}{n!} = \sum_{n\in\mathbb{N}^+} a_n\frac{x^{n-1}}{(n-1)!} = \sum_{n\in\mathbb{N}} a_{n+1}\frac{x^n}{n!}$$

is the EGF for $(a_{n+1})_{n\in\mathbb{N}}$. Compare this with the ordinary generating function of $(a_{n+1})_{n\in\mathbb{N}}$, which is

$$\frac{A(x)-A(0)}{x}.$$

5.7.3.3. *Composition.* Note that

$$[x^n]A(B(x)) = \sum_{k\in\mathbb{N}} \frac{a_k}{k!} \sum_{j_1,\dots,j_k\in\mathbb{N}^+, j_1+\cdots+j_k=n} \frac{b_{j_1}}{j_1!}\cdots\frac{b_{j_k}}{j_k!}$$

which shows that $A(B(x))$ is the EGF for

$$\left(\sum_{k\in\mathbb{N}} \frac{a_k}{k!} \sum_{j_1,\dots,j_k\in\mathbb{N}^+, j_1+\cdots+j_k=n} \binom{n}{j_1,\dots,j_k} b_{j_1}\cdots b_{j_k}\right)_{n\in\mathbb{N}}$$

where

$$\binom{n}{j_1,\dots,j_k} := \frac{n!}{j_1!\cdots j_k!}$$

is the so-called *multinomial coefficient* counting the number of partitions of $[n]$ into sets of size $j_1,\dots,j_k$.

5.7.4. The Bell numbers. Let b_n be the number of partitions of $[n]$ (equivalently, the number of equivalence relations on the set $[n]$); these numbers are called the *Bell numbers*. They satisfy the following recurrence formula:

$$b_{n+1} = \sum_{k\in\{0,\dots,n\}} \binom{n}{k} b_k. \tag{67}$$

The reason is that in order to form a partition, we choose for some k the k elements that do not lie in the equivalence class containing the largest element, and multiply with the number of elements to form a partition with those elements.

Let $B(x)$ be the exponential generating function for $(b_n)_{n\in\mathbb{N}}$. In Theorem 5.7.1 we present a stunningly simple description of $B(x)$ with two proofs:

- the first proof translates the recurrence for b_n into an differential equation and solves it;
- the second proof uses the general dictionary for labelled combinatorial constructions.

THEOREM 5.7.1.

$$B(x) = \exp(\exp(x) - 1)$$

PROOF NUMBER 1. We multiply both sides of the recurrence formula (67) with $\frac{x^n}{n!}$ and sum over all $n \in \mathbb{N}$. On the left hand side, we obtain $B'(x)$. On the right hand side, we obtain $\exp(x)B(x)$ since $\exp(x)$ is the exponential generating function for the constant-one sequence (recall (66)). Hence, we get the differential equation

$$B'(x) = \exp(x)B(x)$$

which has the unique solution $B(x) = c\exp(\exp(x))$. Since $B(0) = 1$ we must have $c = e^{-1}$, so $B(x) = \exp(\exp(x) - 1)$. □

PROOF NUMBER 2. What is the series $\exp(\exp(x) - 1)$? It is a composition of $\exp(x)$ with $\exp(x) - 1$, so it can be interpreted according to the dictionary for composition of labelled objects: $\exp(x)-1$ is the series for cliques with at least one vertex, and hence $n![x^n]\exp(\exp(x) - 1)$ counts the number of disjoint unions of cliques with vertex set $[n]$, i.e., the number of ways in which we can partition $[n]$. □

Now, b_n can be computed as follows (Dobiński's formula). Let $g\colon \mathbb{R} \to \mathbb{R}$ be the function $x \mapsto e^{e^x}$. Note that $g'(x) = e^x e^{e^x}$. Then

$$\begin{aligned} b_n &= n![x^n]\exp(\exp(x) - 1) \\ &= \frac{1}{e} g^{(n)}(0) \\ &= \frac{1}{e} \sum_{k\in\mathbb{N}} \left(\frac{e^{kx}}{k!}\right)^{(n)} (0) \\ &= \frac{1}{e} \sum_{k\in\mathbb{N}} \frac{k^n}{k!}. \end{aligned}$$

5.7.5. 2-regular graphs. A graph is 2-regular if all its vertices have degree two; note that such a tree is a disjoint union of cycles. In this section we want to count the number of labelled 2-regular graphs with n vertices. For $n \leq 2$ there are no undirected cycles at all. For $n \geq 3$, there are $(n-1)!/2$ many cycles with vertex set $[n]$: we choose a first vertex, under the remaining vertices we choose a second, etc. This gives us $n!$, but we have over-counted since we do not care where to start the cycle, and we do not care about the direction of the cycle, so we have to divide by $2n$ which gives the formula.

Therefore, the exponential generating function $C(x)$ for labelled cycles is

$$\begin{aligned} C(x) &= \sum_{n\geq 3} \frac{(n-1)!}{2n!} x^n \\ &= \frac{1}{2} \sum_{n\geq 3} \frac{1}{n} x^n = \frac{1}{2}(\ln \frac{1}{1-x} - x - \frac{x^2}{2}). \end{aligned}$$

The exponential generating function for 2-regular graphs is

$$\begin{aligned} \exp(C(x)) &= \exp\left(\frac{1}{2}(\ln \frac{1}{1-x} - x - \frac{x^2}{2})\right) \\ &= \left(\frac{1}{1-x} \cdot e^{-x} \cdot e^{-x^2/2}\right)^{1/2} = \frac{e^{-x/2-x^2/4}}{\sqrt{1-x}}. \end{aligned}$$

The radius of convergence is $r = 1$ (we see that the function is singular at $x = 1$ where it has a branch point). By a technique called *singularity analysis* (we refer to [**16**] for this more advanced topic) one can show that

$$[x^n]C(x) = \frac{e^{-\frac{3}{4}}}{\sqrt{\pi n}} - \frac{5e^{-3/4}}{8\sqrt{\pi n^3}} + O\left(\frac{1}{n^{5/2}}\right).$$

5.7.6. Permutations and Stirling's formula. A permutation of $[n]$ is a bijection between $[n]$ and itself; there are $n!$ many permutations of $[n]$. Permutations of $[n]$ can be thought of as a directed graph with vertex set $[n]$ where each vertex has indegree one (precisely one predecessor) and outdegree one (precisely one successor). Clearly, such digraphs are disjoint unions of directed cycles (loops are allowed). Labelled directed cycles have been counted already in Example 23, and they have the EGF $\ln(\frac{1}{1-x})$. We conclude that the EGF for the number of permutations is

$$\exp\big(\ln\left(\frac{1}{1-x}\right)\big) = \frac{1}{1-x} = \sum_{n\in\mathbb{N}} x^n$$

as we have seen already in Example 22.

Stirling's formula is an approximation for $n!$, and in one of the formulations states that

$$n! \in \left(\frac{n}{e}\right)^n \sqrt{2\pi n}\left(1 + O\left(\frac{1}{n}\right)\right)$$

or, phrased differently,

$$n! \sim \left(\frac{n}{e}\right)^n \sqrt{2\pi n}.$$

This approximation has an excellent quality, as the following table shows; the error is below 1% already for $n = 10$.

n	1	2	5	10	100	1000
$\frac{n!}{n^n e^{-n}\sqrt{2\pi n}}$	1.084437	1.042207	1.016783	1.008365	1.000833	1.000083

We give a simple proof of a weaker version, namely that for all $n \geq 2$ we have

$$n \ln n - n < \ln(n!) < n \ln n$$

and therefore

$$\ln(n!) \sim n \ln n.$$

The inequality $\ln(n!) < n \ln n$ is a consequence of the trivial inequality $n! < n^n$. For the other inequality, we use the power series expansion of e^x,

$$e^x = \sum_{n \in \mathbb{N}} \frac{x^n}{n!}$$

Comparing e^n to the n-th term in the series gives $\frac{n^n}{n!} < e^n$, so $\frac{n^n}{e^n} < n!$. Therefore $n \ln n - n < \ln(n!)$. Dividing the inequalities $n \ln n$ we obtain

$$1 - \frac{1}{\ln n} < \frac{\ln(n!)}{n \ln n} < 1$$

and hence $\ln(n!) \sim n \ln n$. This formula is already sufficient for many applications where we need an asymptotic bound for the factorial.

The known proofs of Stirling's formula in the stronger formulation above are substantially more involved; several very different proofs (one of them using singularity analysis) can be found in Flajolet and Sedgewick [**16**].

Via the binomial coefficients, Stirling's formula for the factorial enters in many asymptotic estimations.

Example 25. By Stirling's formula

$$\binom{n}{k} \leq \frac{n^k}{k!} \leq \frac{n^k}{(k/e)^k} = \left(\frac{en}{k}\right)^k. \qquad \triangle$$

Example 26. Recall that the Catalan numbers c_n are given by

$$c_n = \frac{1}{n+1}\binom{2n}{n} = \frac{1}{n+1}\frac{(2n)!}{n!n!}.$$

We can now use Stirling's formula to obtain

$$c_n \sim \frac{1}{n}\frac{(2n)^{2n}e^{-2n}\sqrt{4\pi n}}{n^{2n}e^{-2n}2\pi n} \sim \frac{4^n}{\sqrt{\pi n^3}}$$

(and the error is for $n \geq 100$ below 1%). $\triangle$

Exercises.

(88) Prove by induction that

$$\left(\frac{n}{e}\right)^n \leq n! \leq en\left(\frac{n}{e}\right)^n$$

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

(89) A *derangement* is a permutation with no fixed points. Let $D(x)$ be the exponential generating function of the number of derangements. Show that

$$\frac{1}{1-x} = \exp(D(x)).$$

(90) Let a_n be the number of (not necessarily perfect) matchings of K_n; in other words, we count the number of graphs with maximum degree 1 on the vertex set $[n]$.

- Find a rekursive formula for a_n.
- Determine the convergence radius of $\sum_{n \in \mathbb{N}} a_n x^n$.

5.7.7. Labelled graphs and labelled connected graphs. Clearly, there are $2^{\binom{n}{2}}$ labelled graphs with n vertices. How many *connected* labelled graphs with n vertices are there? Let $G(x)$ be the exponential generating function for the number d_n of labelled graphs on n vertices, and let $C(x)$ be the exponential generating function for the number c_n of labelled connected graphs with n vertices. Note that $G(x)$ does not converge for any $x \neq 0$ (d_n simply grows too rapidly) so we cannot apply analytic techniques here, but $G(x)$ and $C(x)$ are perfectly well-defined formal power series. We have

$$G(x) = \exp(C(x))$$

since every graph is uniquely given as the disjoint union of connected graphs.

It follows that the number of connected labeled graphs with n vertices satisfies

$$n2^{\binom{n}{2}} = \sum_{k\in\mathbb{N}} \binom{n}{k} k c_k 2^{\binom{n-k}{2}}.$$

Solving this equation for c_n in terms of c_k for $k < n$, we obtain a recurrence formula from which we can compute c_n for small values of n: the sequence starts with

$$1, 1, 4, 28, 728, 26704, 1866256, 251548592, \ldots$$

5.8. The Lagrange Inversion Formula

The Lagrange inversion formula can be used to access the coefficients of generating functions that are implicitly given. We will use it to prove Cayley's formula for the number of labelled trees with n vertices. In its simplest form, the Lagrange inversion formula expresses the n-th coefficients of the inverse of a power series $A(x)$ in terms of the reciprocal of an n-th power of $A(x)$. If $A(x)$ is invertible then $[x^0]A(x) = 0$ and hence $A(x)$ and powers of $A(x)$ do not have a reciprocal in $R[[x]]$ even if R is a field. However, $A(x)$ has a multiplicative inverse in the larger ring $R((x))$ of *formal Laurent series* that will be introduced in the next section.

5.8.1. Laurent series. Informally, the ring $R((x))$ of formal Laurent series compares to the ring $R[[x]]$ of formal power series in the same way as the ring $R(x)$ of rational functions compares to the ring $R[x]$ of polynomials. Formally, a *Laurent series* is a formal expression of the form

$$A(x) = \sum_{i\in\mathbb{Z}} a_i x^i$$

where $a_i = 0$ for all but finitely many negative i. We define $[x^i]A(x) := a_i$ for all $i \in \mathbb{Z}$. Addition and multiplication, differentiation, and composition can be defined similarly as for formal power series and behave as expected. For example, for $A, B \in R((x))$ the product $A \cdot B$ is defined by putting, for $k \in \mathbb{Z}$,

$$[x^k]A(x)B(x) := \sum_{i\in\mathbb{Z}} [x^i]A[x^{k-i}]B.$$

For $A \neq 0$ we write $\operatorname{ord}(A)$ for the smallest $i \in \mathbb{Z}$ such that $a_i \neq 0$. Note that every Laurent series $A(x) \in R((x))$ can be written as $p(x^{-1}) + B(x)$ where $B \in R[[x]]$ and p is a polynomial (of degree $\operatorname{ord}(A)$).

The coefficient of x^{-1} in a Laurent series $A(x)$ is of particular interest, as we see in the following lemma; it is called the *formal residue of* $A(x)$.

LEMMA 5.8.1. *Let K be a field of characteristic 0 and let $A, B \in K((x))$.*

(1) $[x^{-1}]A'(x) = 0$.
(2) $[x^{-1}]A \cdot B' = -[x^{-1}]A' \cdot B$.
(3) $[x^{-1}]A'/A = \operatorname{ord}(A)$ *if* $A \neq 0$.

(4) $[x^{-1}](A \circ B) \cdot B' = [x^{-1}]A$ *if* $\operatorname{ord}(B) > 0$.

PROOF. (1) is clear from the definition of differentiation. To see (2), note that $0 = [x^{-1}](A \cdot B)' = [x^{-1}](AB' + A'B)$ and the statement follows.

We show (3). Any non-zero $A \in R((x))$ can be written as $A = x^m C$ for $m := \operatorname{ord}(A) \in \mathbb{Z}$ and some $C \in R((x))$ where $\operatorname{ord}(C) = 0$. Then

$$\frac{A'}{A} = \frac{(x^m C)'}{x^m C} = \frac{m x^{m-1} C + x^m C'}{x^m C} = \frac{m}{x} + \frac{C'}{C}.$$

Since $\operatorname{ord}(C) = 0$ we have

$$[x^{-1}]\frac{C'}{C} = \sum_{i \in \mathbb{Z}} [x^i] C' [x^{-1-i}] C^{-1} = 0$$

and it follows that

$$[x^{-1}]\frac{A'}{A} = \underbrace{[x^{-1}] m x^{-1}}_{=m} + \underbrace{[x^{-1}]\frac{C'}{C}}_{=0} = m = \operatorname{ord}(A).$$

We show (4). Let $a_{-1} := [x^{-1}]A(x)$. First note that $A = a_{-1}x^{-1} + C'$ for some $C \in K[[x]]$. Hence,

$$\begin{aligned}
&[x^{-1}](A \circ B)B' \\
&= [x^{-1}]\big(a_{-1}(x^{-1} \circ B)B' + [x^{-1}](C' \circ B) \cdot B'\big) \\
&= a_{-1}[x^{-1}]B^{-1}B' + [x^{-1}](C \circ B)' \\
&= a_{-1}[x^{-1}]B'/B && \text{(using (1))} \\
&= a_{-1} && \text{(using (3), } \operatorname{ord}(B) = 1). \qquad \square
\end{aligned}$$

5.8.2. Lagrange inversion. We now prove the Lagrange inversion formula for obtaining the coefficients of the inverse $B(x)$ of $A(x)$ (there exist more general formulations of the Lagrange inversion formula). The typical proof heavily relies on results from complex analysis (Funktionentheorie); we prefer an algebraic treatment here since it is entirely elementary.

THEOREM 5.8.2. *Suppose that* $A \in K[[x]]$ *has an inverse* B. *Then*

$$[x^n]B = \frac{1}{n}[x^{-1}]A^{-n}.$$

PROOF.

$$\begin{aligned}
[x^n]B &= [x^{-1}]B \cdot x^{-n-1} \\
&= [x^{-1}]((B \cdot x^{-n-1}) \circ A) \cdot A' && \text{(Lemma 5.8.1, item (4))} \\
&= [x^{-1}](B \circ A) \cdot A^{-n-1} \cdot A' \\
&= [x^{-1}]x A^{-n-1} A' && \text{(since } B(A(x)) = x) \\
&= -\frac{1}{n}[x^{-1}]x(A^{-n})' && \text{(definition of derivation)} \\
&= \frac{1}{n}[x^{-1}]A^{-n} && \text{(Lemma 5.8.1, item (2)).} \qquad \square
\end{aligned}$$

5.8.3. Labelled trees. For $n \in \mathbb{N}$, let t_n be the number of trees with vertex set $[n]$. Since trees are required to have at least one vertex, we have $t_0 = 0$, and clearly $t_1 = 1$. Using the Lagrange inversion formula, we will prove Cayley's formula; there are now different, more direct combinatorial proofs of the formula, e.g. via Prüfer codes (from 1918). Let us point out that the symbolic proof via generating functions was found first!

THEOREM 5.8.3 (Cayley 1889). *For every* $n \in \mathbb{N}^+$

$$t_n = n^{n-2}.$$

To prove this result we use an important idea in combinatorial enumeration: the idea to *root* objects. A *rooted tree* is a tree with one distinguished vertex, the *root*. A finite rooted tree can be obtained recursively as follows: we recursively construct a set of smaller rooted trees $T_1, \dots, T_k$ with vertex sets $[n_1], \dots, [n_k]$, respectively, such that $n_1 + \cdots + n_k = n - 1$. Note that $k = 0$ is allowed; this will be the base case of the recursive description. We then rename the vertices of $T_1, \dots, T_k$ such that $V(T_1) \cup \cdots \cup T_k = [n-1]$, and create a tree T with vertex set $[n]$ and root n by joining n with each of the roots of $T_1, \dots, T_k$. Let r_n be the number of rooted trees with vertex set $[n]$; clearly, $r_n = n \cdot t_n$ for all $n \in \mathbb{N}$.

Let $R(x) := \sum_{n\in\mathbb{N}} \frac{r_n}{n!} x^n$ be the exponential generating function for $(r_n)_{n\in\mathbb{N}}$. With a bit of practice, it can be seen directly from the recursive description above that

$$R(x) = x \exp(R(x)) \tag{68}$$

(recall the interpretation of composition presented in Section 5.7.3.3, and the fact that $\exp(x)$ is the exponential generating function for sets; alternatively, use the description above to write down a recurrence formula for r_n, multiply with x^n, and sum over all $n \in \mathbb{N}$ to then obtain the equation, as we have practiced numerous times by now).

The central idea to use the Lagrange inversion formula here is the observation that $R(x)$ can be viewed as the inverse of $S(x) := \frac{x}{\exp(x)}$ since then the identity $S(R(x)) = x$ is equivalent to (68). To access the coefficients of $R(x)$, we can therefore use Lagrange inversion (Theorem 5.8.2):

$$\begin{aligned}
[x^n]R(x) = \frac{1}{n}[x^{-1}]S(x)^{-n} &= \frac{1}{n}[x^{-1}]\left(\frac{e^x}{x}\right)^n \\
&= \frac{1}{n}[x^{n-1}]e^{nx} \\
&= \frac{1}{n}[x^{n-1}]\sum_{k\in\mathbb{N}} \frac{(nx)^k}{k!} \\
&= \frac{1}{n}\frac{n^{n-1}}{(n-1)!} \\
&= \frac{n^{n-1}}{n!}.
\end{aligned}$$

Hence, we have

$$r_n = n![x^n]R(x) = n^{n-1}$$

and

$$t_n = \frac{r_n}{n} = n^{n-2}$$

which concludes the proof of Theorem 5.8.3.

Exercises.

(91) Let $E(x)$ be the exponential generating function for the number of functions from $[n] \to [n]$. Show that $E(x) = \frac{1}{1-R(x)}$ where $R(x)$ is the EGF for rooted trees. **Hint:** View a function from $[n] \to [n]$ as a digraph. How does the resulting digraph look like? Also see Example 22.

(92) Let $V(x)$ be the exponential generating function of labelled trees with *two* (not necessarily distinct) distinguished vertices. Show that

$$V(x) = \frac{1}{1 - R(x)} - 1.$$

Hint: Find a bijection with non-empty permutations of rooted trees. Again see Example 22.

(93) Use the the previous two exercises and an explicit formula for the number of functions from $[n] \to [n]$ to derive a new proof of Cayley's formula.

(94) Find an algorithm that outputs for a given $n \in \mathbb{N}^+$ a labelled tree with vertex set $[n]$ uniformly at random and whose running time is polynomial in n.

5.8.4. Binary trees revisited. The Lagrange inversion formula can also be applied for unlabelled enumeration and ordinary generating functions. We present a second proof that binary trees of size n are counted by the Catalan numbers, $c_n = \frac{1}{n+1}\binom{2n}{2}$ (Theorem 5.6.1). Recall from (64) that the (ordinary) generating function $C(x) := \sum_{n\in\mathbb{N}} c_n x^n$ satisfies

$$C(x) = 1 + xC(x)^2. \tag{69}$$

Note that a binary tree of size n is constructed from n pairs of brackets applied to $n+1$ leaves, so has $2n+1$ vertices in total if we draw the binary tree as in Figure 5.2. Let b_n be the number of binary trees with n vertices in total; so $c_n = b_{2n+1}$. Let $B(z) = \sum_{n\in\mathbb{N}} b_n z^n$ be the corresponding ordinary generating function; the reason we consider $B(z)$ instead of $C(x)$ is that the computations come out more easily when applied to $B(z)$. Equation (69) implies that

$$B(z) = z + zB(z)^2 \tag{70}$$

which has an immediate combinatorial explanation: a binary tree either consists of a single vertex, or otherwise is constructed from a left binary tree and a right binary tree that jointly have one vertex less (to account for the internal vertex created for combining the two subtrees). To compute the coefficients of $B(x)$ we use the Lagrange inversion formula.

Let $D(u) := \frac{u}{1+u^2}$. Note that $[x^0]D(u) = 0$ and that

$$\begin{aligned} D(B(z)) &= \frac{B(z)}{1 + B(z)^2} = \frac{B(z)}{1 + \frac{B(z)-z}{z}} && \text{(by (70))} \\ &= z \end{aligned}$$

and hence D is the inverse of B. We can therefore apply the Lagrange inversion formula and obtain

$$\begin{aligned} b_n &= \frac{1}{n}[u^{-1}]D(u)^{-n} \\ &= \frac{1}{n}[u^{-1}]\frac{(1+u^2)^n}{u^n} \\ &= \frac{1}{n}[u^{n-1}]\sum_{k=0}^{n}\binom{n}{k}u^{2k} \\ &= \begin{cases} 0 & \text{if } n \text{ is even} \\ \frac{1}{n}\binom{n}{(n-1)/2} & \text{if } n \text{ is odd.} \end{cases} \end{aligned}$$

Hence,

$$c_n = b_{2n+1} = \frac{1}{2n+1}\binom{2n+1}{n} = \frac{1}{2n+1}\frac{(2n+1)!}{(n+1)!n!} = \frac{1}{n+1}\binom{2n}{n}.$$

5.9. Unlabelled Enumeration

For labelled enumeration, we have seen a powerful dictionary between combinatorial construction principles and basic operations on exponential generating functions, such as addition, Cauchy product, derivation, and composition (Section 5.7.3). Is there a similar dictionary for unlabelled enumeration and ordinary generating functions (OGFs)? This is true for the formation of disjoint unions, which still corresponds to addition of OGFs, and for the formation of ordered pairs, which still corresponds to the product of OGFs. However, the composition operation poses problems, due to potential symmetries that the composed object might have. For the same reason, taking the derivative and multiplying with the formal variable no longer corresponds to rooting the object.[3] This deficiency can be solved by Polya theory and cycle index sums. Our main application will be a proof of a formula for the number of unlabelled trees with n vertices. There is a great variety of classes of combinatorial structures that can be treated similarly.

5.9.1. Relational structures. We work with the general concept of *(relational) structures* which generalises directed graphs, but also hypergraphs, directed graphs with several types of edges, or with distinguished subsets of the vertices, etc. Even if we are only interested in enumerating unlabelled graph classes (such as unlabelled trees), we still need a more powerful concept than graphs to develop a general theory of unlabelled enumeration. Another approach is to use the concept of so-called *compinatorial species*, which has been developed by Joyal and is based on concepts from category theory [**5**].

A *relational signature* is a set of *relation symbols* R, each equipped with an arity $\mathrm{ar}(R) \in \mathbb{N}$. A *$\tau$-structure* $\underline{A}$ consists of a set A, called the *domain* of $\underline{A}$, and a relation $R^{\underline{A}} \subseteq A^{\mathrm{ar}(k)}$ for each relation symbol $R \in \tau$. A structure is called *finite* if its domain is finite. Relational structures are often written like $(A; R_1^{\underline{A}}, R_2^{\underline{A}_2}, \dots)$, with the obvious interpretation; for example, $(\mathbb{Q}; <)$ denotes the structure whose domain is the set of rational numbers $\mathbb{Q}$ and which carries a single binary relation $<$ which denotes the usual strict linear order o the rationals. Following common practice, we sometimes do not distinguish between the symbol R for a relation and the relation $R^{\underline{A}}$ itself, in situations where this does not lead to confusion. We *do* allow structures with an empty domain.[4] Directed graphs are viewed as τ-structures for the signature $\tau = \{E\}$, where E is a binary relation symbol that denotes the edge relation of a digraph.

5.9.2. Cycle index sums. Cycle index sums of classes of combinatorial objects are formal power series that incorporate information about the *symmetries* of the objects in the class. Symmetry is formalised by the concept of an *automorphism*, which we define now.

Let τ be a relational signature and let $\underline{A}$ and $\underline{B}$ be two τ-structures. A function $f\colon A \to B$ is a *homomorphism* if for every $R \in \tau$ of arity k and for every $(a_1, \dots, a_k) \in R^{\underline{A}}$ it holds that $(f(a_1), \dots, f(a_k)) \in R^{\underline{B}}$. An *isomorphism* i from $\underline{A}$ to $\underline{B}$ is a bijective

[3]The reason that this issue did not show up for binary trees, which can be viewed as unlabelled objects, is that binary trees do not have non-trivial automorphisms; and indeed, the recursive description of the corresponding ordinary generating function only involves addition and multiplication.

[4]Warning: in particular in many (but not all) logic text books the authors use the convention that structures have non-empty domains).

homomorphism from $\underline{A}$ to $\underline{B}$ such that i^{-1} is a homomorphism as well. A permutation α of the domain of a structure $\underline{A}$ is called an *automorphism* if it is an isomorphism between $\underline{A}$ and $\underline{A}$.

Clearly, the set of all automorphisms of a structure $\underline{A}$ contains the identity and is closed under composition and taking the inverse, and hence forms a group with respect to composition (a *permutation group*).

DEFINITION 5.9.1. Let α be a permutation of a set of size n. The *weight* of α is defined as

$$w_\alpha := \frac{1}{n!}\prod_{i=1}^{n} s_i^{c_i(\alpha)}$$

where s_i is a formal variable and $c_i(\alpha)$ is the number of cycles of α of length i.

Let $\mathcal{A}$ be a class of finite structures which is closed under isomorphisms, i.e., if $\underline{A} \in \mathcal{A}$ and $\underline{B}$ is isomorphic to $\underline{A}$, then $\underline{B} \in \mathcal{A}$. We write $\mathcal{A}_n$ for the set of all structures in $\mathcal{A}$ with domain $\{1, \ldots, n\}$. The following definition goes back to Polya [**34**].

DEFINITION 5.9.2 (Cycle Index Sum). The *cycle index sum of* $\mathcal{A}$, denoted by $Z_{\mathcal{A}}(s_1, s_2, \ldots)$, is the formal power series

$$\sum_{n\in\mathbb{N}} \left(\sum_{\underline{A}\in\mathcal{A}_n} \left(\sum_{\alpha\in \mathrm{Aut}(\underline{A})} w_\alpha \right) \right).$$

We present several important examples, all over the signature $\tau = \{E\}$ of directed graphs.

EXAMPLE 27. Let $\mathcal{L}$ be the class of all finite τ-structures where E denotes a linear order. Note that every structure $\underline{A} \in \mathcal{L}$ has just a single automorphism, namely the identity map. If $|A| = n$ then the identity map has weight $\frac{s_1^n}{n!}$. Note that $\mathcal{L}_n$ has exactly $n!$ elements, and hence the cycle index sum of $\mathcal{L}$ is

$$Z_{\mathcal{L}} = \sum_{n\in\mathbb{N}} s_1^n = \frac{1}{1-s_1}. \qquad \triangle$$

EXAMPLE 28. Let $\mathcal{K}$ be the class that consists of all finite complete digraphs (including the digraph with an empty vertex set). Note that $\mathrm{Aut}(K_n)$ is the full symmetric group on n elements. Hence, for example for $n = 3$ we have

$$Z_{\mathcal{K}_3} = \frac{1}{6}(s_1^3 + 3s_1s_2 + 2s_3).$$

In general, the cycle index sum for $\mathcal{K}$ is simply the exponential generating series for all permutations where each cycle of length r is marked by the formal variable s_r.

$$Z_{\mathcal{K}} = \sum_{n\in\mathbb{N}} \sum_{\alpha\in\mathrm{Aut}(K_n)} w_\alpha = \sum_{n\in\mathbb{N}} \frac{1}{n!}\left(\sum_{r\geq 1} \frac{s_r}{r}\right)^n = \exp\left(\sum_{r\geq 1} \frac{s_r}{r}\right)$$

Recall that $\sum_{r\geq 1} \frac{x^r}{r}$ is the exponential generating function for labelled directed cycles (Example 23) and that $\exp(\sum_{r\geq 1} \frac{x^r}{r})$ is the exponential generating function for labelled permutations (Section 5.7.6). $\triangle$

EXAMPLE 29. Let $\mathcal{C}$ be the class of all finite directed cycles (we assume that they contain at least one element). If $(v_0, \ldots, v_{n-1})$ is a directed cycle $\underline{C}_n$ in $\mathcal{C}_n$, for $n \geq 1$, then the map that sends v_i to v_{i+m}, where $m \in \{0, \ldots, n-1\}$ and indices are considered modulo n, is an automorphism of $\underline{C}_n$. Every automorphism of $\underline{C}_n$ is of this form. Such an automorphism consists of n/r cycles of length r, where r is

the order of m in $\mathbb{Z}_n$. For each divisor r of n, there are $\phi(r)$ elements of order r in $\mathbb{Z}_n$. Here, ϕ is the Euler totient function, which returns for given r the number of elements in $\{1,\dots,n-1\}$ that are pairwise prime with r. Since there are $(n-1)!$ cycles with the vertices $\{1,\dots,n\}$, the cycle index sum of $\mathcal{C}_n$ is

$$Z_{\mathcal{C}_n} = (n-1)! \sum_{\alpha\in \mathrm{Aut}(\underline{C}_n)} w_\alpha = \frac{1}{n}\sum_{r|n}\phi(r)s_r^{n/r}$$

and the cycle index sum of $\mathcal{C}$ is

$$\begin{aligned} Z_{\mathcal{C}} &= \sum_{n\in\mathbb{N}} Z_{\mathcal{C}_n} = \sum_{r\geq 1}\sum_{n\geq 1,\ r|n} \frac{\phi(r)}{n} s_r^{n/r} = \sum_{r\geq 1}\frac{\phi(r)}{r}\sum_{m\geq 1}\frac{s_r^m}{m} \\ &= \sum_{r\geq 1}\frac{\phi(r)}{r}\log\frac{1}{1-s_r}. \end{aligned}$$ △

Let a_n be the number of structures with n vertices in $\mathcal{A}$, up to isomorphism, and let $A(x) := \sum_{n\in\mathbb{N}} a_n x^n$ be the corresponding ordinary generating function. The following lemma shows that the cycle index sum of $\mathcal{A}$ incorporates all the information of $A(x)$.

LEMMA 5.9.3. $Z_{\mathcal{A}}(x, x^2, x^3, \dots) = A(x)$.

To prove this fundamental fact, we need Burnside's lemma from permutation group theory presented in the next section.

5.9.3. Basics of permutation groups. Let G be a permutation group on a set X. For $\alpha \in G$, the set of fixed points of α is defined to be

$$\mathrm{Fix}(\alpha) := \{x \in X \mid \alpha(x) = x\}.$$

For $x \in X$, the *stabiliser of x in G* is the subgroup of G defined as

$$G_x := \{\alpha \in G \mid \alpha(x) = x\}.$$

The *orbit* of x in G is the set

$$G.x := \{\alpha(x) \mid \alpha \in G\}$$

Note that the orbits of G partition $\{1,\dots,n\}$. We write X/G for the set of all orbits of G.

LEMMA 5.9.4 (Orbit-Stabiliser Lemma). *Let G be a permutation group on $\{1,\dots,n\}$ and $x \in \{1,\dots,n\}$. Then $|G| = |G_x| \cdot |G.x|$.*

LEMMA 5.9.5 (Burnside's orbit-counting lemma). *Let G be a permutation group on a finite set X. Then*

$$|X/G| = \frac{1}{|G|}\sum_{\alpha\in G} |\,\mathrm{Fix}(\alpha)|.$$

In other words, the number of orbits of G equals the average number of fixed points of the elements of G.

PROOF. Burnside's lemma is a consequence of the orbit-stabiliser lemma via an easy double counting argument:

$$\begin{aligned}\sum_{\alpha\in G} |\operatorname{Fix}(\alpha)| &= |\{(\alpha,x) \mid \alpha\in G, x\in X, \alpha(x)=x\}| \\ &= \sum_{x\in X} |G_x| \\ &= \sum_{x\in X} \frac{|G|}{G.x} && \text{(Lemma 5.9.4)} \\ &= |G| \sum_{O\in X/G} \sum_{x\in O} \frac{1}{|O|} \\ &= |G| \sum_{O\in X/G} 1 = |G|\cdot|X/G|\end{aligned}$$

and the statement follows. □

We can now prove that the cycle index sum indeed specialises to the ordinary generating function.

PROOF OF LEMMA 5.9.3. To prove that $[x^n]Z_{\mathcal{A}}(x,x^2,x^3,\dots) = a_n$ we apply Burnside's lemma with respect to the permutation group G defined as follows. The domain of G is the finite set $\mathcal{A}_n$ which consists of all structures in $\mathcal{A}$ with vertex set $\{1,\dots,n\}$. For every permutation α of $\{1,\dots,n\}$ the group G contains the permutation of $\mathcal{A}_n$ which maps $\underline{S}\in\mathcal{A}_n$ to the isomorphic copy of $\underline{S}$ in $\mathcal{A}_n$ obtained by renaming $u\in S$ with $\alpha(u)$. Note that a_n equals the number of isomorphism classes of structures in $\mathcal{A}$, hence, equals the number of orbits of G. Also note that $|G| = n!$. We have

$$w_\alpha(x,x^2,\dots,x^n) = \frac{1}{n!}\prod_{i=1}^{n} x^{i\cdot c_i(\alpha)} = \frac{1}{n!}x^{c_1(\alpha)+2c_2(\alpha)+\cdots+nc_n(\alpha)} = \frac{x^n}{n!}.$$

Hence,

$$\begin{aligned}\sum_{\underline{S}\in\mathcal{A}_n}\left(\sum_{\alpha\in\operatorname{Aut}(\underline{S})} w_\alpha\right)(x,x^2,x^3,\dots,x^n) &= \sum_{\underline{S}\in\mathcal{A}_n} |\operatorname{Aut}(\underline{S})|\cdot\frac{x^n}{n!} \\ &= \frac{1}{n!}\sum_{\alpha\in G}|\operatorname{Fix}(\alpha)|\cdot x^n \\ &= |\mathcal{A}_n/G|\cdot x^n && \text{(by Lemma 5.9.5).}\end{aligned}$$

It follows that $[x^n]Z_{\mathcal{A}}(x,x^2,x^3,\dots) = |\mathcal{A}_n/G| = a_n$. □

5.9.4. Combinatorial constructions and cycle index sums. In this section develop we a dictionary between combinatorial constructions and the corresponding algebraic operations on cycle index sums.

Let τ be a relational signature and let $\underline{A}$ be a τ-structure. We start by defining some general basic terminology for relational structures.

DEFINITION 5.9.6 (Reduct and Expansion). If $\tau'\subseteq\tau$ and $\underline{A}'$ is a τ'-structure with the same domain as A and $R^{\underline{A}} = R^{\underline{A}'}$ for all $R\in\tau'$, then $\underline{A}'$ is called the τ'-*reduct* (or simply *reduct*) of $\underline{A}$, and $\underline{A}$ is called a τ-expansion (or simply *expansion*) of $\underline{A}'$.

DEFINITION 5.9.7 (Substructure and Extension). A τ-structure $\underline{B}$ is is called a *substructure* of $\underline{A}$ if $B \subseteq A$ and for every $R \in \tau$ of arity k we have that $R^{\underline{B}} = R^{\underline{A}} \cap B^k$. In this case, we say that $\underline{A}$ is an *extension* of $\underline{B}$. We also say that $\underline{B}$ is *the substructure induced by* $\underline{A}$ *on* B.

5.9.4.1. *Disjoint union.* Let $\mathcal{A}$ and $\mathcal{B}$ be two disjoint classes of τ-structures. Then

$$Z_{\mathcal{A}\cup\mathcal{B}} = Z_{\mathcal{A}} + Z_{\mathcal{B}}.$$

5.9.4.2. *Product.* Let $\mathcal{A}$ and $\mathcal{B}$ be two classes of structures with disjoint relational signatures τ_1 and τ_2. The *(partitional) product* $\mathcal{A} \cdot \mathcal{B}$ consists of all $\tau_1 \cup \tau_2 \cup \{P\}$-structures $\underline{S}$, where P is a new unary relation symbol, such that the τ_1-reduct $\underline{A}$ of the substructure of $\underline{S}$ with domain $P^{\underline{S}}$ is from $\mathcal{A}$, and the τ_2-reduct $\underline{B}$ of the substructure of $\underline{\mathrm{S}}$ with domain $S \setminus P^{\underline{S}}$ is from $\mathcal{B}$. Note that if γ is an automorphism of $\underline{S}$, then the restriction of γ to $P^{\underline{S}}$ is an automorphism of $\underline{A}$ and the restriction of γ to $S \setminus P^{\underline{S}}$ is an automorphism of $\underline{B}$. Conversely, if α is an automorphism of $\underline{A}$ and β is an automorphism of $\underline{B}$, then the permutation of S that agrees with α on $P^{\underline{S}}$ and with β on $S \setminus P^{\underline{S}}$ is an automorphism of $\underline{S}$. This shows the following.

PROPOSITION 5.9.8.

$$Z_{\mathcal{A}\cdot\mathcal{B}} = Z_{\mathcal{A}} \cdot Z_{\mathcal{B}}.$$

5.9.4.3. *Substitution.* Let $\mathcal{A}$ and $\mathcal{B}$ be two classes of structures with disjoint relational signatures τ_1 and τ_2 such that $\mathcal{B}$ does not contain empty structures. The *composite* of $\mathcal{B}$ in $\mathcal{A}$, denoted by $\mathcal{A} \circ \mathcal{B}$, is the class of all $\tau_1 \cup \tau_2 \cup \{E\}$-structures $\underline{S}$, where E is a new binary relation symbol, such that

- $E^{\underline{S}}$ is an equivalence relation on S; the equivalence class of $s \in S$ with respect to $E^{\underline{S}}$ is denoted by $[s]$.
- there is a structure $\underline{A} \in \mathcal{A}$, called the *template structure*, and a bijection b between the equivalence classes of $E^{\underline{S}}$ and A such that for every $R \in \tau_1$ of arity k and for every $(s_1, \dots, s_k) \in S^k$ we have $(s_1, \dots, s_k) \in R^{\underline{S}}$ if and only if $(b([s_1]), \dots, b([s_k])) \in R^{\underline{A}}$.
- for each equivalence class $\{s_1, \dots, s_n\}$ of $E^{\underline{S}}$ there exists an isomorphism between the τ_2-reduct of the structure induced by $\underline{S}$ on $\{s_1, \dots, s_n\}$ and a structure $\underline{B} \in \mathcal{B}$ (a *component structure*).

Let σ be an automorphism of $\underline{S}$. Since σ preserves E, it induces a permutation τ of the equivalence classes of E, and τ must be an automorphism of $\underline{A}$ because σ preserves all the relations from τ_1. Now consider an equivalence class $B = \{s_1, \dots, s_n\}$ of $E^{\underline{S}}$ and let $\underline{B}$ the τ_2-reduct of the structure induced by $\underline{S}$ on B. Let k be the length of the cycle of τ that contains B. Then σ^k maps B to B, so the restriction of σ^k to B is an automorphism of $\underline{B}$.

Conversely, automorphisms of the template structure from $\mathcal{A}$ and the component structures from $\mathcal{B}$ give rise to automorphisms of the composite; see Figure 5.5 for an illustration.

PROPOSITION 5.9.9. *Let $\mathcal{A}$ and $\mathcal{B}$ be classes of finite relational structures with disjoint signatures. Then*

$$Z_{\mathcal{A}\circ\mathcal{B}} = Z_{\mathcal{A}} \circ Z_{\mathcal{B}} := Z_{\mathcal{A}}(Z_{\mathcal{B}}(s_1, s_2, \dots), Z_{\mathcal{B}}(s_2, s_4, \dots), Z_{\mathcal{B}}(s_3, s_6, \dots), \dots). \quad (71)$$

PROOF. TODO. □

REMARK 5.9.10. The composition of the formal power series $Z_{\mathcal{A}}$ and $Z_{\mathcal{B}}$ in (71) is called *plethystic substitution.* A treatment of the content of Proposition 5.9.9 in the context of combinatorial species can be found in [**5**].

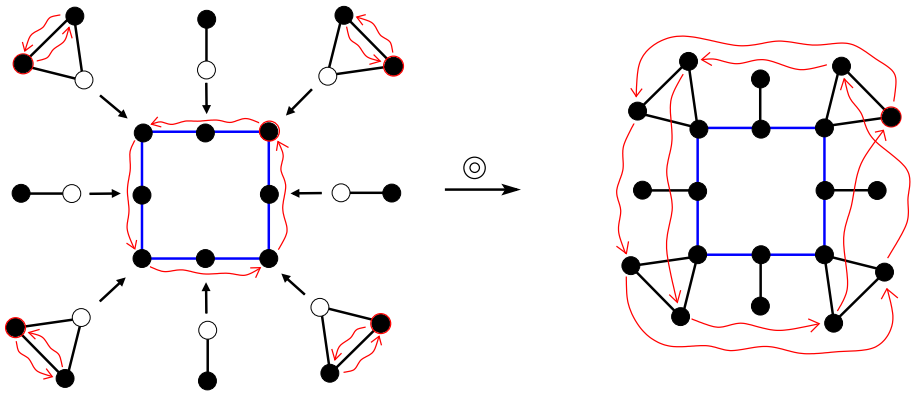

FIGURE 5.5. A substitution operation and some cycles of automorphisms.

5.9.4.4. *Pointing (Rooting).* When computing the number of labelled trees, it was useful to instead count *rooted* trees. Counting labelled trees and counting rooted labelled trees is essentially the same because every labelled structure of size n gives rise to exactly n rooted objects. Unfortunately, the same is false for unlabelled enumeration. Adding a root adds a *bias* in the sense that if we generate a rooted structure with domain $\{1, \dots, n\}$ uniformly at random and then drop the root, structures with a small automorphism group are more likely than they are in the uniform distribution.

Let $\mathcal{A}$ be a class of τ-structures and let $A(x)$ be the ordinary generating function for the number of unlabelled structures of size n in $\mathcal{A}$. Then $\mathcal{A}^\bullet$ is the class of all $(\tau \cup \{P\})$-structures $\underline{B}$ where $|P^{\underline{B}}| = 1$. The rooting operation has an effect on the corresponding ordinary generating function for the number of unlabelled structures in the class that cannot be described on the level of ordinary generating functions alone; we need to look at cycle index sums instead (which then provide the ordinary generating functions via Lemma 5.9.3).

PROPOSITION 5.9.11. $Z_{\mathcal{A}^\bullet}(s_1, s_2, s_3, \dots) = s_1(\frac{\partial}{\partial s_1} Z_{\mathcal{A}})(s_1, s_2, s_3, \dots)$

PROOF. TODO. □

5.9.4.5. *Cycle Pointing.* Instead of rooting a structure $\underline{A}$ at a vertex we may add a *cycle* which is the cycle of an automorphism of $\underline{A}$. This operation, which is called *cycle pointing*, is *unbiased* (see previous section), as we will see in Theorem 5.9.12. The advantage of working with cycle pointed objects is that the distinguished cycle may serve as the starting point in a recursive decomposition of the structures in the class, similarly as rooting the trees is useful when enumerating labelled trees.

Let $\mathcal{A}$ be a class of τ-structures. Then the *cycle-pointed class of* $\mathcal{A}$, denoted by $\mathcal{A}^\circ$, consists of all expansions of structures $\underline{A} \in \mathcal{A}$ by a new relation C which denotes the edge relation of a cycle $(a_0, \dots, a_{\ell-1})$ of elements of A which is the cycle of some automorphism α of $\underline{A}$.

Let $A(x) = \sum_{n \in \mathbb{N}} a_n x^n$ be the ordinary generating function for the number a_n of structures in $\mathbb{A}_n$ up to isomorphism, and let $A^\circ(x) = \sum_{n \in \mathbb{N}} a_n^\circ x^n$ be the ordinary generating function for the number a_n° of structures in $\mathcal{A}_n^\circ$ up to isomorphism.

THEOREM 5.9.12 (Unbiased pointing [**6**]). *Let $n \in \mathbb{N}$. For each unlabelled structure $\underline{S}$ of size n in $\mathcal{A}_n$ there are exactly n non-isomorphic structures $(\underline{S}, C)$ in $\mathcal{A}_n^\circ$. Hence,*

$$A^\circ(x) = xA'(x).$$

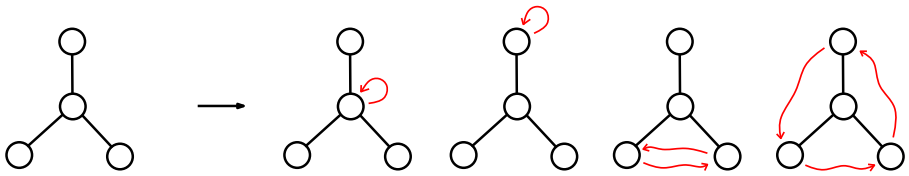

FIGURE 5.6. An unlabelled tree of size 4 yields 4 unlabelled cycle-pointed trees.

PROOF. TODO. □

See Figure 5.6.

REMARK 5.9.13. Theorem 5.9.12 can be translated into the language of permutation groups, where it is known as *Parker's lemma*.

5.9.4.6. *Cycle-pointed Substitution.* Let $\mathcal{A}$ and $\mathcal{B}$ be two classes of relational structures with disjoint signatures τ_1 and τ_2 and assume that $\mathcal{B}$ does not contain the empty structure. Let $\underline{P}$ be a structure from $(\mathcal{A} \circ \mathcal{B})^\circ$ and let $(c_1, \ldots, c_k)$ be the cycle marked by $C^{\underline{S}}$. Let $\underline{S} \in \mathcal{A} \circ \mathcal{B}$ be the $\tau_1 \cup \tau_2$-reduct of $\underline{P}$, and let $\underline{A} \in \mathcal{A}$ be the template structure for $\underline{S}$. Then $(c_1, \ldots, c_k)$ is the cycle of an automorphism of $\underline{S}$, and this automorphism induces a cycle on the equivalence classes of $E^{\underline{S}}$. Consider the $\{P\} \cup \tau_1$-expansion of $\underline{A}$ where P denotes this cycle; note that this expansion is from $\mathcal{A}^\circ$, and we call it the *template structure of* $\underline{S}$.

For $\mathcal{P} \subseteq \mathcal{A}^\circ$ we define the class $\mathcal{P} \circledcirc \mathcal{B}$ as the class of all structures $\underline{S}$ in $(\mathcal{A} \circ \mathcal{B})^\circ$ where the template structure is from $\mathcal{P}$. Note that

$$(\mathcal{A} \circ \mathcal{B})^\circ = \mathcal{A}^\circ \circledcirc \mathcal{B}.$$

The cycle index sum of $\mathcal{P} \circledcirc \mathcal{B}$ can be computed systematically, similarly as in Proposition 5.9.9 for $\mathcal{A} \circ \mathcal{B}$. However, the details are technical and we refer the interested reader to [**6**].

REMARK 5.9.14. To avoid clumsy expressions with many brackets, we make the convention that with respect to binding strength, the symbols are ordered as follows: $+$ (lowest binding strength), $\cdot$, the binary composition operations $\circ$ and $\circledcirc$, and the unary pointing operation $\circ$ (strongest binding strength).

5.9.5. Unlabelled rooted trees. In Section 5.9.6 we will reduce the task to counting unlabelled trees to counting *rooted* unlabelled trees $\mathcal{R}$, i.e., trees with a single distinguished vertex, considered up to isomorphisms that preserve the root. Rooted trees are easier to count. The reason is that rooted trees can be decomposed recursively at the root into several rooted trees, similarly as in Section 5.8.3.

Let $\mathcal{X}$ denote the class of structures over the empty signature that just contains all one-element structures. Note that $Z_{\mathcal{X}} = s_1$. The idea of the decomposition of $\mathcal{R}$ can then be expressed recursively as follows (such equations can be formalised, which goes beyond the scope of this course; we refer to [**6**]).

$$\mathcal{R} \equiv \mathcal{X} \cdot \mathcal{K} \circ \mathcal{R} \tag{72}$$

THEOREM 5.9.15. *If r_n is the number of rooted unlabelled trees with at least one vertex and let $R(x) = \sum_{n \in \mathbb{N}} r_n x^n$ be the corresponding ordinary generating function. Then*

$$R(x) = x \exp\left(\sum_{i \geq 1} \frac{R(x^i)}{i} \right). \tag{73}$$

PROOF. Recall from Example 28 that $Z_{\mathcal{K}} = \exp\left(\sum_{r\geq 1} \frac{s_r}{r}\right)$. Using the rule of computing the cycle index sum of substitutions (Proposition 5.9.9) we obtain

$$Z_{\mathcal{K}\circ\mathcal{R}} = \exp\left(\sum_{i\geq 1} \frac{Z_{\mathcal{R}}(s_i, s_{2i}, \dots)}{i}\right).$$

Using the rule of computing the cycle index sum of products (Section 5.9.4.2) we obtain that

$$Z_{\mathcal{X}\cdot\mathcal{K}\circ\mathcal{R}} = s_1 \cdot \exp\left(\sum_{i\geq 1} \frac{Z_{\mathcal{R}}(s_i, s_{2i}, \dots)}{i}\right).$$

Specialising via Lemma 5.9.3 yields

$$\begin{aligned} R(x) = Z_{\mathcal{X}\cdot\mathcal{K}\circ\mathcal{R}}(x, x^2, x^3, \dots) &= x \cdot \exp\left(\sum_{i\geq 1} \frac{Z_{\mathcal{R}}(x^i, x^{2i}, \dots)}{i}\right) \\ &= x \cdot \exp\left(\sum_{i\geq 1} \frac{R(x^i)}{i}\right). \end{aligned}$$

□

REMARK 5.9.16. Equation (73) allows for an efficient computation of r_n; this can for instance be done using the computer algebra system Maple.

5.9.6. Unlabelled trees. In this section we present a formula for the number t_n of unlabelled trees (i.e., connected acyclic graphs) with n vertices and the corresponding ordinary generating function $T(x) = \sum_{n\in\mathbb{N}} t_n x^n$. We use the results about rooted unlabelled trees from the previous section, and in particular we use the ordinary generating function $R(x)$.

THEOREM 5.9.17 (from [**6**]). $xT'(x) = R(x) + x^2R'(x^2) + R(x)\sum_{\ell\geq 2} x^\ell R'(x^\ell)$

REMARK 5.9.18. In combination with Theorem 5.9.15 and Remark 5.9.16 this theorem allows for an efficient computation of t_n.

To prove Theorem 5.9.17 we present a recursive decomposition for the class $\mathcal{T}$ of all finite trees structures. Our strategy is as follows.

(1) By Theorem 5.9.12, the enumeration task is equivalent to the task of enumerating unlabelled cycle-pointed trees, $\mathcal{T}^\circ$.
(2) We now distinguish whether the marked cycle in an element $\underline{A} \in \mathcal{T}^\circ$ has length one or length greater than one.
(3) If the cycle has length one, we view the element of the cycle as a root, and hence reduced the situation to the rooted case.
(4) If the cycle has length greater than one, then we call $\underline{A}$ *symmetric*. The class of all symmetric elements of $\mathcal{T}^\circ$ will be denoted by $\mathcal{S}$. A key observation for the description of $\mathcal{S}$ is that there exists either a unique vertex or a unique edge which we call the *center of symmetry*, and which may serve as a starting point for a decomposition (Proposition 5.9.19). See Figure 5.7.

The following proposition can be shown easily by induction, removing simultaneously all leaves of the tree until the tree is reduced to a single vertex or a single edge.

PROPOSITION 5.9.19. *Let $\underline{A}$ be a finite tree and let $(c_0c_1\cdots c_{k-1})$ be a cycle from an automorphism of $\underline{A}$. Then there exists either*

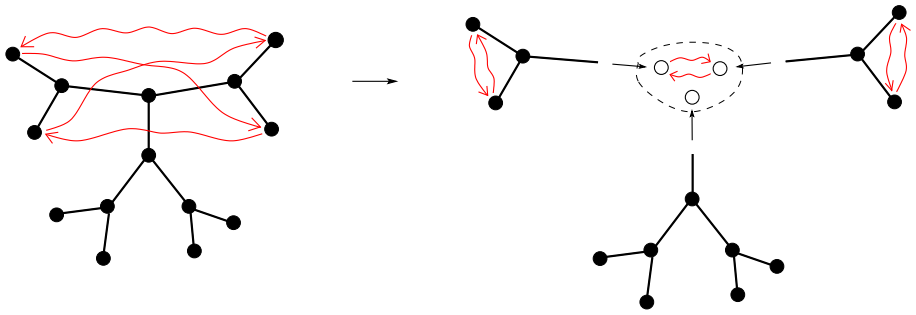

FIGURE 5.7. Decomposition of a tree at its center of symmetry (in the case that the center of symmetry is a vertex).

- *a unique vertex $v \in A$ such that for every $i \in \{0, \dots, k-1\}$ the (unique) path P from v_i to v_{i+1} (indices modulo k) has odd length and v is the middle vertex of P, or*
- *a unique edge e of $\underline{A}$ such that for every $i \in \{0, \dots, k-1\}$ the (unique) path P from v_i to v_{i+1} (indices modulo k) has even length and e is the middle edge on P.*

To formally specify the recursive definition of $\mathcal{T}^\circ$, we again introduce two basic classes of structures that serve as building blocks.

DEFINITION 5.9.20. The class

- $\mathcal{E}$ denotes the class of cycle pointed trees with two vertices where the marked cycle has size exactly two.
- $\mathcal{K}^{(\geq 2)}$ denotes the class of cycle-pointed complete digraphs with an arbitrary number of vertices such that the marked cycle has size at least two. Note that $\mathcal{K}^{(\geq 2)} \subseteq \mathcal{K}^\circ$ where $\mathcal{K}$ has been introduced in Example 28.

Similarly as in (73), the idea of the strategy to decompose $\mathcal{T}$ that we explained above may be phrased recursively as follows.

$$\mathcal{T}^\circ \equiv \mathcal{R} + \mathcal{S} \tag{74}$$

$$\mathcal{S} \equiv \mathcal{E} \circledcirc \mathcal{R} + \mathcal{X} \cdot (\mathcal{K}^{(\geq 2)} \circledcirc \mathcal{R}) \tag{75}$$

PROOF SKETCH OF THEOREM 5.9.17. We write $T^\circ(x)$ for the ordinary generating function of $\mathcal{T}^\circ$ and $S(x)$ for the ordinary generating function of $\mathcal{S}$. By Theorem 5.9.12 we have $xT'(x) = T^\circ(x)$. Clearly, $T^\circ(x) = R(x) + S(x)$ (see (74) and Section 5.9.4.1).

It can be shown that the ordinary generating function for $\mathcal{E} \circledcirc \mathcal{R}$ is $x^2 \cdot R'(x^2)$, and that the ordinary generating function for $\mathcal{X} \cdot (\mathcal{K}^{(\geq 2)} \circledcirc \mathcal{R})$ is

$$\left(\sum_{\ell \geq 2} x^\ell R'(x^\ell) \right) \cdot R(x).$$

Putting these together (see (75)) we obtain that

$$\begin{aligned} xT'(x) &= R(x) + S(x) \\ &= R(x) + x^2 R'(x^2) + R(x) \sum_{\ell \geq 2} x^\ell R'(x^\ell). \end{aligned}$$

□

REMARK 5.9.21. Otter [**33**] proved with a similar, but different approach that

$$T(x) = R(x) - \frac{R(x)^2 - R(x^2)}{2}.$$

REMARK 5.9.22. The description of $R(x)$ in Theorem 5.9.15 and of $T(x)$ in Theorem 5.9.17 combined with results from analytic combinatorics can be used to obtain the precise growth rates for r_n and t_n; details can be found in [**6**]. We just mention that

$$r_n \sim c^{-3/2}\rho^{-n}$$
$$t_n \sim (2\pi c^3)n^{-5/2}\rho^{-n}$$

for constants $c \approx 0.43922$ and $\rho \approx 0.33832$.

Bibliography

[1] S. Aaronson. P=?NP. *Electronic Colloquium on Computational Complexity (ECCC)*, 24:4, 2017.

[2] N. Alon, A. Kostochka, B. Reiniger, D. B. West, and X. Zhu. Coloring, sparseness, and girth, 2015. Preprint available at Arxiv:1412.8002.

[3] N. Alon and J. H. Spencer. *The Probabilistic Method*. Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons, Inc., New York, 2000.

[4] S. Arora and B. Barak. *Computational Complexity: A Modern Approach*. Cambridge University Press, 594 pages.

[5] F. Bergeron, G. Labelle, and P. Leroux. *Théorie des espèces et combinatoire des structures arborescentes*. LaCIM, Montréal, 1994. English version: Combinatorial Species and Tree-like Structures, Cambridge University Press (1998).

[6] M. Bodirsky, É. Fusy, M. Kang, and S. Vigerske. Boltzmann samplers, Pólya theory, and cycle pointing. *SIAM J. Comput.*, 40(3):721–769, 2011.

[7] D. A. Cohen, M. C. Cooper, P. Creed, P. G. Jeavons, and S. Zivny. An algebraic theory of complexity for discrete optimization. *SIAM J. Comput.*, 42(5):1915–1939, 2013.

[8] A. Coja-Oghlan and A. Taraz. Exact and approximative algorithms for coloring G(n,p). *Random Struct. Algorithms*, 24(3):259–278, 2004.

[9] A. Condon. The complexity of stochastic games. *Information and Computation*, 96(2):203 – 224, 1992.

[10] C. Daskalakis, P. W. Goldberg, and C. H. Papadimitriou. The complexity of computing a Nash equilibrium. *SIAM J. Comput.*, 39(1):195–259, 2009.

[11] F. d'Epenoux. A probabilistic production and inventory problem. *Management Science*, 10(1):98–108, 1963.

[12] R. Diestel. *Graph Theory*. Springer–Verlag, New York, 2005. Third edition.

[13] P. Erdős. On a problem in graph theory. *The Mathematical Gazette*, 47(361):220–223, 1963.

[14] T. Feder and M. Y. Vardi. The computational structure of monotone monadic SNP and constraint satisfaction: a study through Datalog and group theory. *SIAM Journal on Computing*, 28:57–104, 1999.

[15] J. Filar and K. Vrieze. *Competitive Markov Decision Processes*. Springer, New York, 1996.

[16] P. Flajolet and R. Sedgewick. *Analytic Combinatorics*. Cambridge University Press, 2009.

[17] E. Fusy. Random generation. MPRI Lecture Notes, 2011. available at http://www.lix.polytechnique.fr/Labo/Eric.Fusy/Teaching/notes.pdf.

[18] R. L. Graham, B. L. Rothschild, and J. H. Spencer. *Ramsey theory*. Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons, Inc., New York, 1990. Second edition.

[19] M. Grötschel, L. Lovász, and L. Schrijver. *Geometric Algorithms and Combinatorial Optimization*. Springer, Heidelberg, 1994. Second edition.

[20] A. W. Hales and R. I. Jewett. Regularity and positional games. *Transactions of the AMS*, 106:222–229, 1993.

[21] W. Hodges. *A shorter model theory*. Cambridge University Press, Cambridge, 1997.

[22] S. Janson, T. Luczak, and A. Rucinski. *Random Graphs*. John Wiley and Sons, New York, 2000.

[23] S. Jukna. *Extremal Combinatorics (With Applications in Computer Science)*. Springer-Verlag, 2001.

[24] A. Kechris, V. Pestov, and S. Todorčević. Fraïssé limits, Ramsey theory, and topological dynamics of automorphism groups. *Geometric and Functional Analysis*, 15(1):106–189, 2005.

[25] L. Khachiyan. A polynomial algorithm in linear programming. *Doklady Akademii Nauk SSSR*, 244:1093–1097, 1979.

[26] L. Libkin. *Elements of Finite Model Theory*. Springer, 2004.

[27] J. Matoušek and J. Nešetřil. *Invitation to Discrete Mathematics*. Oxford University Press, 1998.

[28] J. Matoušek and B. Gärtner. *Understanding and Using Linear Programming*. Springer, 2007.

[29] J. Nash. Non-cooperative games. *Annals of Mathematics*, 54:289–295, 1951.

[30] J. Nešetřil. Ramsey classes and homogeneous structures. *Combinatorics, Probability & Computing*, 14(1-2):171–189, 2005.
[31] J. Nešetřil and V. Rödl. A short proof of the existence of highly chromatic hypergraphs without short cycles. *J. Comb. Theory, Ser. B*, 27(2):225–227, 1979.
[32] J. Nešetřil and V. Rödl. Chromatically optimal rigid graphs. *J. Combin. Theory Ser. B*, 46:133–141, 1989.
[33] R. Otter. The number of trees. *Annals of Math.*, 49:583–599, 1948.
[34] G. Pólya. *Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen. Acta Mathematica*, 68(1):145–254, 1937.
[35] V. Rosta. Ramsey theory applications. *Electronic Journal of Combinatorics*, 2004. Dynamic Survey D13.
[36] J. M. Ruiz. *The Basic Theory of Power Series*. Advanced Lectures in Mathematics. Springer, 1993.
[37] R. Schilling. *Wahrscheinlichkeit*. De Gruyter Studium, 2017.
[38] A. Schrijver. *Theory of Linear and Integer Programming*. Wiley - Interscience Series in Discrete Mathematics and Optimization, 1998.
[39] S. Shapley. Stochastic games. *Proceedings of the national Academy of Sciences, USA*, 39:1095–1100, 1953.
[40] J. H. Spencer. *The strange logic of random graphs*. Springer, 2001.
[41] H. S. Wilf. *generatingfunctionology*. Academic Press, Inc., 1990. Free internet edition.

APPENDIX A

Basics from Calculus

We first recall the formal definition of convergence. Let $(P, \leq)$ be a partially ordered set. A *upper (lower) bound* of $S \subseteq P$ is an element $p \in P$ that is larger (smaller) than all elements of S. An upper bound b is called a *supremum (infimum) of* S if b larger than all other upper bounds of S in P, and written by $\sup(S)$. Suprema and infima do not necessarily exist. But every non-empty subset S of the real numbers $\mathbb{R}$ has a supremum and an infimum (by the definition of $\mathbb{R}$).

DEFINITION A.0.1. Let $(a_n)_{n\in\mathbb{N}}$ be a sequence of numbers in $\mathbb{R}$.

- Then the *limit superior of* $(a_n)_{n\in\mathbb{N}}$ is an element of $\mathbb{R} \cup \{-\infty, +\infty\}$ defined as
$$\limsup_{n\to\infty} a_n := \inf_{n\to\infty} (\sup_{m\geq n} x_m).$$
- The *limit inferior of* $(a_n)_{n\in\mathbb{N}}$ is an element of $\mathbb{R} \cup \{-\infty, +\infty\}$ defined as
$$\limsup_{n\to\infty} a_n := \sup_{n\to\infty} (\inf_{m\geq n} x_m).$$
- $(a_n)_{n\in\mathbb{N}}$ *converges* if
$$\limsup_{n\to\infty} a_n = \liminf_{n\to\infty} =: \lim_{n\to\infty} a_n \in \mathbb{R};$$
equivalently, $\lim(a_n)_{n\in\mathbb{N}}$ if and only if for every $\epsilon > 0$ there exists an $m > 0$ such that for all $n \geq m$ $|f(a_n) - f(a)| < \epsilon$.

Let $S \subseteq \mathbb{R}$, let $z \in S$, and $f\colon S \to \mathbb{R}$. We write $\lim_{a\to z} f(a) = b$ if for every $\epsilon > 0$ there exists $\delta > 0$ such that for all $a \in S$ with $|a - z| < \delta$ we have $|f(a) - f(z)| < \epsilon$. The definitions for complex-valued functions are analogous. It is a basic lemma in calculus that limits in $\mathbb{C}$ can be evaluated componentwise: we identify $\mathbb{C}$ with $\mathbb{R}^2$. Then $(a_i)_{n\in\mathbb{N}}$ converges against $a \in \mathbb{C}$ if the sequence of real parts converges against the real part of a and the sequence of imaginary parts converges against the imaginary part of a.

A.1. Divergence and Convergence Tests

We recall some of the basic divergence and convergence tests.

LEMMA A.1.1 (term divergence test). *If* $\lim_{n\to\infty} a_n$ *does not exist or* $\lim_{n\to\infty} \neq 0$ *then* $\sum_{n\in\mathbb{N}} a_n$ *diverges.*

PROOF. We show the contrapositive: if $\sum_{n\in\mathbb{N}} a_n$ converges, then $\lim_{n\to\infty} a_n = 0$. Writing s_n for $\sum_{i=0}^{n} a_i$ and ℓ for $\lim_{n\to\infty} s_n$, we have

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty} (s_n - s_{n-1}) = \lim_{n\to\infty} s_n - \lim_{n\to\infty} s_{n-1} = \ell - \ell = 0. \qquad \square$$

LEMMA A.1.2 (comparison test). *If* $\sum_{n\in\mathbb{N}} b_n$ *converges, and there exists an* $m \in \mathbb{N}$ *such that* $0 \leq a_n \leq b_n$ *for all* $n \geq m$*, then* $\sum_{n\in\mathbb{N}} a_n$ *converges, too.*

PROOF. Let $s_n := \sum_{i=0}^{n} a_n$. Then $s_{n-1} = a_{n+1} \geq 0$ for all $n \geq m$. So $s_{n+1} \geq s_n$ which means that s_n is non-decreasing. It is one of the fundamental properties of real numbers that a non-decreasing sequence converges if it has an upper bound. Such an upper bound exists since

$$\begin{aligned} s_n = a_0 + a_1 + \cdots + a_n &\leq b_0 + b_1 + \cdots + b_n \\ &\leq b_0 + b_1 + \cdots + b_n + b_{n+1} + \cdots \end{aligned}$$ □

A sequence $(a_n)_{n\in\mathbb{N}}$ of real or complex numbers is called a *Cauchy sequence* if and only if for every $\epsilon > 0$ there exists an $n_0 \in \mathbb{N}$ such that for all $n, m \geq n_0$ we have $|a_m - a_n| < \epsilon$. Since the real and complex numbers are complete a sequence converges if and only if it is a Cauchy sequence. This condition is useful since we do not need to know the limit of the sequence in order to prove convergence. It gives the following convergence test for series.

LEMMA A.1.3 (Cauchy's convergence test). *$\sum_{n\in\mathbb{N}} a_i$ is convergent if and only if for every $\epsilon > 0$ there exists an $n_0 \in \mathbb{N}$ such that $|a_{n+1} + a_{n+2} + \cdots + a_{n+p}| < \epsilon$ for all $n > n_0$ and $p \in \mathbb{N}$.*

PROOF. Let $s_n := \sum_{m=0}^{n} a_m$. By definition, the series $\sum_{n\in\mathbb{N}} a_i$ is convergent if and only if $(s_n)_{n\in\mathbb{N}}$ is convergent, which is the case if and only if $(s_n)_{n\in\mathbb{N}}$ is a Cauchy sequence. Writing out this last condition we arrive at the statement. □

And we have a further divergence test.

LEMMA A.1.4 (tail of convergent series tends to zero). *Let $\sum_{n\in\mathbb{N}} a_n$ be a convergent series. Then $b_m := \sum_{n=m}^{\infty} a_n$ converges and $\lim_{m\to\infty} b_m = 0$.*

PROOF. The convergence of $\sum_{n=m}^{\infty} a_n$ follows immediately from Cauchy's test. We have $\sum_{n\in\mathbb{N}} a_n = \sum_{n=0}^{m} a_m + b_m$. Since $\lim_{m\to\infty} \sum_{n=0}^{m} a_m = \sum_{n\in\mathbb{N}} a_n$ we have that $\lim_{m\to\infty} b_m = 0$. □

A.2. Inequalities

LEMMA A.2.1 (inequality of arithmetic and geometric means). *Let $x_1, \ldots, x_n$ be non-negative real numbers. Then*

$$\bar{x}_a := \frac{x_1 + x_2 + \cdots + x_n}{n} \geq \sqrt[n]{x_1 \cdot x_2 \cdots x_n} =: \bar{x}_g$$

and equality holds if and only if $x_1 = x_2 = \cdots = x_n$.

PROOF. Apply (30) to $x_i/\bar{x}_a - 1$, and we obtain

$$\exp(x_i/\bar{x}_a - 1) \geq x_i/\bar{x}_a.$$

Multiplying all of these inequalities for all $i \in \{1, \ldots, n\}$ we obtain

$$\exp\left(\sum_{i=1}^{n} x_i/\bar{x}_a - n\right) \geq \prod_i x_i/\bar{x}_a$$

and hence

$$1 = \exp(n - n) \geq \bar{x}_g^n/\bar{x}_a^n$$

and finally

$$\bar{x}_a^n \geq \bar{x}_g^n.$$ □

LEMMA A.2.2 (Cauchy-Schwarz). *Let $x, y \in \mathbb{R}^n$. Then*

$$\left(\sum_{i=1}^{n} x_i y_i\right)^2 \leq \left(\sum_{i=1}^{n} x_i^2\right)\left(\sum_{i=1}^{n} y_i^2\right). \tag{76}$$

PROOF. We use the usual notation $\langle x, y\rangle$ for the scalar product $\sum_{i=1}^{n} x_i y_i$ and $||x|| := \sqrt{\langle x, x\rangle}$ for the associated norm; then (76) can be written as $\langle x, y\rangle \leq ||x|| \cdot ||y||$. Note that for every $\lambda \in \mathbb{R}$ we have

$$0 \leq \langle \lambda y, \lambda y\rangle = \langle x, x\rangle - 2\lambda\langle x, y\rangle + \lambda^2\langle y, y\rangle.$$

Choosing $\lambda := \frac{\langle x,y\rangle}{\langle y,y\rangle}$ we obtain

$$0 \leq \langle x, x\rangle - 2\frac{\langle x, y\rangle^2}{\langle y, y\rangle} + \frac{\langle x, y\rangle^2}{\langle y, y\rangle} = \langle x, x\rangle - \frac{\langle x, y\rangle^2}{\langle y, y\rangle}$$

and hence $\langle x, y\rangle^2 \leq \langle x, x\rangle \cdot \langle y, y\rangle$. We obtain the statement by taking square roots. □

APPENDIX B

Some Basics from Complexity Theory

For a set A, we write A^* for the set of all words over the alphabet A. A *word over A* can be seen as a function from $\{1, \ldots, n\} \to A$, for some $n \in \mathbb{N}$. We write ϵ for the empty word (i.e., for the function with the empty domain).

The most classical setting of complexity theory is the study of the computational complexity of functions f from $\{0,1\}^* \to \{0,1\}$. Alternatively, we may view f as a set of words, namely that set of words w such that $f(w) = 1$; such sets are also called *formal languages*. There are several mathematically rigorous machine models to formalise the set of such functions that are *computable* or *efficiently computable*. The first insight is that most of these machine models lead to the same, or to closely related classes of functions. Complexity theory maps out the landscape of the resulting classes of functions. Typically the first machine model that is introduced in introductory courses are *Turing machines*.

B.1. Turing Machines

Turing machines strike a good balance between the following two (almost contradictory!) requirements that a theoretician has for these machine models:

- the model should be relatively simple, so that it is easy to show that it can be simulated by many other machine models.
- the model should be relatively powerful, so that it is easy to show that it can simulate many other machine models.

Turing machines are simple, but still the definition does not easily fit into a few lines. On the other hand, today academics are most likely to already have a very good idea of what a computer program can do (in polynomially many steps); and this coincides with what a Turing machine M can do (in polynomially many computational steps). In a nutshell, a Turing machine

- has an unboundedly large memory containing values from $\{-1, 0, 1\}$ (the symbol -1 will be called the *blank* symbol);
- has finitely many *states* Q;
- has a *read-* and *write* head;
- has a finite transition function $\delta\colon Q \times \{-1, 0, 1\} \to \Sigma \times Q \times \{l, r\}$;
- has a *accept* state $y \in Q$.
- has a *start* state $s \in Q$.

Initially, the memory just contains the word $w \in \{0,1\}^*$, i.e., in the first cell there is w_1, in the second cell there is w_2, etc, and in all further memory cells there is -1, and the machine *is in state s*. Depending on its state $u \in Q$ and the tape content c under the read-write head, let $(v, d, m) := \delta(u, c)$; then

(1) the machine changes to state v;
(2) the tape content under the read-write head is changed from c to d,
(3) the read-write tape moves one cell to the left if $m = l$, and one to the right if $m = r$.

If the machine reaches state y it accepts. Every Turing machine describes a formal language, namely the function $f\colon \{0,1\}^* \to \{0,1\}$ such that $f(w) = 1$ if and only if when running the machine on input w it eventually accepts. We also say that M *computes* f, and we then sometimes write $M(f)$ instead of $f(w)$. More generally, Turing machines can be used to describe functions f from $\{0,1\}^*$ to $\{0,1\}^*$ where $f(w)$, for a given word w, is the string that is written on the output tape when the Turing machine accepts (here we require that the machine terminates on every input after finitely many steps, and again we say that M *computes* f).

So we will pretend in the following that the reader already knows what Turing machines M are. It turns out that despite the simplicity of Turing machines, they can simulate most of the other machine models, and they can simulate any machine that humans ever constructed (even when neglecting the restriction that we one have some fixed finite maximal memory size in this universe).

B.2. Complexity

In complexity theory we are interested in the number of computation steps that M needs to perform to compute $f(w)$, which corresponds to computation time. For example, we say that a Turing machine runs *in polynomial time* if the number of computation steps is in $O(|w|^k)$ for some $k \in \mathbb{N}$. The class of such functions is denoted by P.

Coding. In the combinatorics course we have met computational complexity for example in the section about colorability. We mentioned that 2-colorability is in P and that k-colorability, for $k \geq 3$, is NP-hard. But these were problems about finite graphs, whereas in the above we only treated formal languages. But this is just a matter of coding. We first observe that we can simulate any alphabet by our alphabet $\{0,1\}$, by just grouping bits together to represent a richer alphabet. In particular, we will typically use the letter # to separate different numbers in the input. One way to represent a graph as a word is to first write the number n of vertices, followed by the symbol #, followed by a sequence of n^2 bits for the adjacency matrix.

The second most important complexity class is NP.

Definition B.2.1. NP (for *nondeterministic polynomial time*) stands for the class of all functions $f\colon \{0,1\}^* \to \{0,1\}$ such that there exists a polynomial-time Turing machine M and a $d \in \mathbb{N}$ such that for every $w \in \{0,1\}^*$ there exists a $a \in \{0,1\}^*$ with $|a| \in O(n^d)$ such that $f(w) = M(w\#a)$.

It is a famous open problem whether P = NP, and it is widely conjectured that P $\neq$ NP. To explain the significance of this conjecture, we need a couple of more concepts. Let $f_1, f_2\colon \{0,1\}^* \to \{0,1\}$. A *reduction* from f_1 to f_2 is a function $g\colon \{0,1\}^* \to \{0,1\}^*$ such that $f_1(w) = f_2(g(w))$. A reduction g is *polynomial-time* if g can be computed a Turing machine that runs in polynomial time.

Definition B.2.2. A function $f\colon \{0,1\}^* \to \{0,1\}$ is *NP-hard* if every function g in NP has a polynomial-time reduction to f. A function is called *NP-complete* if it is in NP and NP-hard.

The class *coNP* is dual to NP: it is the class of all functions f such that $1-f$ is in NP. There is an analogous definition for any complexity class K: a function is in co-K if $1 - f$ is in K. Clearly, every function in P is both in NP and in coNP. There are some problems that are simultaneously in NP and in coNP, but that are not known to be in P: we have seen some examples in Chapter 2.

B.3. A Logic Perspective

A class of finite graphs $\mathcal{C}$ is in NP if there exists a formal language in NP such that each word in the language codes a graph in $\mathcal{C}$ (say in the way we described above), and every graph in $\mathcal{C}$ is coded by some word in the language. Unlike the class P, it is possible to define the class of all graph classes in NP transparently and fully formally in a few lines (without any reference to Turing machines).

THEOREM B.3.1 (Fagin). *A class of finite graphs $\mathcal{C}$ is in NP if and only if there exists an existential second-order sentence Φ such that for every finite graph G we have*

$$G \in \mathcal{C} \text{ if and only if } G \models \Phi.$$

We do not define *existential second-order logic* here. The interested reader is referred to a textbook on finite model theory to learn more about such connections between logic and complexity theory, e.g. [**26**].

B.4. The P Versus NP Problem

We now return to the question why most researchers believe that P $\neq$ NP. In order to show that P=NP is suffices to provide for *any* of the known NP-complete problems a polynomial-time algorithm. There are many NP-complete problems that are of central importance in optimisation, scheduling, cryptography, bioinformatics, artificial intelligence and many more areas. If P=NP, then this would mean a simultaneous breakthrough in all of these areas. It is fair to say that every day, thousands of researchers are directly or indirectly working on proving that P=NP (since they work on things that are related to the better understanding of some NP-complete problem). The fact that nobody has succeeded (not even came close to) is one of the reasons why we believe that P cannot be equal to NP. A world where P = NP would probably be drastically different from the world we live in. On the other hand, we also have no clue on how to possibly prove that P $\neq$ NP. And quite a bit is known about approaches to proving P $\neq$ NP that must fail (see [**1**]).