

---

# Foreword: MASS at Penn State University

This book is part of a collection published jointly by the American Mathematical Society and the MASS (Mathematics Advanced Study Semesters) program as a part of the Student Mathematical Library series. The books in the collection are based on lecture notes for advanced undergraduate topics courses taught at the MASS (Mathematics Advanced Study Semesters) program at Penn State. Each book presents a self-contained exposition of a non-standard mathematical topic, often related to current research areas, which is accessible to undergraduate students familiar with an equivalent of two years of standard college mathematics, and is suitable as a text for an upper division undergraduate course.

Started in 1996, MASS is a semester-long program for advanced undergraduate students from across the USA. The program's curriculum amounts to sixteen credit hours. It includes three core courses from the general areas of algebra/number theory, geometry/topology, and analysis/dynamical systems, custom designed every year; an interdisciplinary seminar; and a special colloquium. In addition, every participant completes three research projects, one for each core course. The participants are fully immersed into mathematics, and this, as well as intensive interaction among the students, usually leads

to a dramatic increase in their mathematical enthusiasm and achievement. The program is unique for its kind in the United States.

Detailed information about the MASS program at Penn State can be found on the website [www.math.psu.edu/mass](http://www.math.psu.edu/mass).

---

## Chapter 2

# Infinite Ramsey theory

### 2.1. The infinite Ramsey theorem

In this chapter, we will look at Ramsey's theorem for colorings of infinite sets. We start with the simplest infinite Ramsey theorem. We carry over the notation from the finite case. Given any set  $Z$  and a natural number  $p \geq 1$ ,  $[Z]^p$  denotes the set of all  $p$ -element subsets of  $Z$ , or simply the  $p$ -sets of  $Z$ .

**Theorem 2.1** (Infinite Ramsey theorem). *Let  $Z$  be an infinite set. For any  $p \geq 1$  and  $r \geq 1$ , if  $[Z]^p$  is colored with  $r$  colors, then there exists an infinite set  $H \subseteq Z$  such that  $[H]^p$  is monochromatic.*

Compared with the finite versions of Ramsey's theorem in Chapter 1, the statement of the theorem seems rather elegant. This is due to a robustness of infinity when it comes to subsets: It is possible to remove infinitely many elements from an infinite set and still have an infinite set. It is customary to call a monochromatic set  $H$  as in Theorem 2.1 a **homogeneous** (for  $c$ ) subset, and from here on we will use *monochromatic* and *homogeneous* interchangeably.

**Proof.** Fix  $r \geq 1$ . We will proceed via induction on  $p$ . For  $p = 1$  the statement is the simplest version of an **infinite pigeonhole principle**:

If we distribute infinitely many objects into finitely many drawers, one drawer must contain infinitely many objects.

In our case, the drawers are the colors  $1, \dots, r$ , and the objects are the elements of  $Z$ .

Next assume  $p > 1$  and let  $c : [Z]^p \rightarrow \{1, \dots, r\}$  be an  $r$ -coloring of the  $p$ -element subsets of  $Z$ .

To use the induction hypothesis, we fix an arbitrary element  $z_0 \in Z$  and use  $c$  to define a coloring of  $(p-1)$ -sets: For  $\{b_1, \dots, b_{p-1}\} \in [Z \setminus \{z_0\}]^{p-1}$ , define

$$c_0(b_1, \dots, b_{p-1}) := c(z_0, b_1, \dots, b_{p-1}).$$

Note that  $Z \setminus \{z_0\}$  is still infinite. Hence, by the inductive hypothesis, there exists an infinite homogeneous  $Z_1 \subseteq Z \setminus \{z_0\}$  for  $c_0$ , which in turn means that all  $p$ -sets  $\{z_0, b_1, \dots, b_{p-1}\}$  with  $b_1, \dots, b_{p-1} \in Z_1$  have the same  $c$ -color.

Pick an element  $z_1$  of  $Z_1$ . Now define a coloring of the  $(p-1)$ -sets of  $Z_1 \setminus \{z_1\}$ : For  $b_1, \dots, b_{p-1} \in Z_1 \setminus \{z_1\}$ , put

$$c_1(b_1, \dots, b_{p-1}) := c(z_1, b_1, \dots, b_{p-1}).$$

Again, our inductive hypothesis tells us that there is an infinite homogeneous subset  $Z_2 \subseteq Z_1 \setminus \{z_1\}$  for  $c_1$ .

We can continue this construction inductively and obtain infinite sets  $Z \supset Z_1 \supset Z_2 \supset Z_3 \supset \dots$ , where  $Z_{i+1}$  is homogeneous for a coloring  $c_i$  of the  $(p-1)$ -sets of  $Z_i$  that is derived from  $c$  by fixing one element  $z_i$  of  $Z_i$ , and thus all  $p$ -sets of  $\{z_i\} \cup Z_{i+1}$  that contain  $z_i$  have the same  $c_i$ -color.

By virtue of our choice of the  $Z_i$  and the  $z_i$ , the sequence of the  $z_i$  has the crucial property that for any  $i \geq 0$ ,

$$\{z_{i+1}, z_{i+2}, \dots\}$$

is homogeneous for  $c_i$  (namely, it is a subset of  $Z_{i+1}$ ). Let  $k_i$  denote the color ( $\in \{1, \dots, r\}$ ) for which the homogeneous set  $Z_{i+1}$  is monochromatic.

Now use the infinite pigeonhole principle one more time: At least one color, say  $k^*$ , must occur infinitely often among the  $k_i$ . Collect

the corresponding  $z_i$ 's in a set  $H$ . We claim that  $H$  is homogeneous for  $c$ .

To verify the claim, let  $\{h_1, h_2, \dots, h_p\} \subset H$ . Every element of  $H$  is a  $z_i$ , i.e. there exist  $i_1, \dots, i_p$  such that

$$h_1 = z_{i_1}, \dots, h_p = z_{i_p}.$$

Without loss of generality, we can assume that  $i_1 < i_2 < \dots < i_p$  (otherwise reorder). Then  $\{h_1, h_2, \dots, h_p\} \subset Z_{i_1}$ , and hence the color of  $\{h_1, h_2, \dots, h_p\}$  is  $k_{i_1} = k^*$  (all the colors corresponding to a  $z_j$  in  $H$  are equal to  $k^*$ ). The choice of  $\{h_1, h_2, \dots, h_p\}$  was arbitrary, and thus  $H$  is homogeneous for  $c$ .  $\square$

Conceptually, this proof is not really different from the proofs of the finite Ramsey theorem, Theorem 1.31. We start with an arbitrary element of  $Z$  and “thin out” the set  $\mathbb{N}$  so that all possible completions of this element to a  $p$ -set have the same  $c$ -color. We pick one of the remaining elements and do the same for all other remaining elements, and so on. Then we apply the pigeonhole principle one more time to homogenize the colors. The difference is that in the finite case we argued that if we start with enough numbers (or vertices), the process will produce a large enough finite sequence of numbers (vertices). In the infinite case, the process *never stops*. This is the robustness of infinity mentioned above: It is possible to take out infinitely many elements infinitely many times from an infinite set and still end up with an infinite set. In some sense, the set we end up with ( $H$ ) is smaller than the set we started with ( $Z$ ). But in another sense, it is of the same size: it is still infinite.

This touches on the important concept of *infinite cardinalities*, to which we will return in Section 2.5.

## 2.2. König's lemma and compactness

As noted before, the infinite Ramsey theorem is quite elegant, in that its nature seems *more qualitative than quantitative*. We do not have to worry about keeping count of finite cardinalities. Instead, the robustness of infinity takes care of everything.

It is possible to exploit infinitary results to prove finite ones. This technique is usually referred to as **compactness**. The essential ingredient is a result about infinite trees known as **König's lemma**. This is a purely combinatorial statement, but we will see in the next section that it can in fact be seen as a result in *topology*, where compactness is originally rooted.

Using compactness relieves us of much of the counting and book-keeping we did in Chapter 1, but usually at the price of not being able to derive bounds on the finite Ramsey numbers. In fact, using compactness often introduces huge numbers. In Chapters 3 and 4, we will see how large these numbers actually get.

**Partially ordered sets.** In Section 1.2, we introduced trees as a special family of graphs (those without cycles). We also saw that every tree induces a *partial order* on its vertex set. Conversely, if a partial order satisfies certain additional requirements, it induces a tree structure on its elements, on which we will now elaborate.

Orders play a fundamental role not only in mathematics but in many other fields from science to finance. Many things we deal with in our life come with some characteristics that allow us to compare and order them: gas mileage or horse power in cars, interest rates for mortgages, temperatures in weather reports—the list of examples is endless. Likewise, the mathematical theory of orders studies sets that come equipped with a binary relation on the elements, the *order*.

Most mathematical orders you encounter early on are *linear*, and they are so natural that we often do not even realize there is an additional structure present. The integers, the rationals, and the reals are all linearly ordered: If we pick any two numbers from these sets one will be *smaller* than the other. But we can think of examples where this is not necessarily the case. For example, take the set  $\{1, 2, 3\}$  and consider all possible subsets:

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

We order these subsets by saying that  $A$  is smaller than  $B$  if  $A \subset B$ . Then  $\{1\}$  is smaller than  $\{1, 2\}$ , but what about  $\{1, 2\}$  and  $\{1, 3\}$ ? Neither is contained in the other, and so the two sets are *incomparable* with respect to our order—the order is *partial* (Figure 2.1).

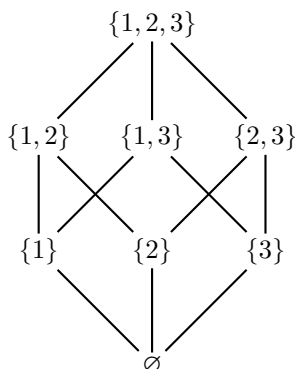


Figure 2.1. The subset partial order on  $\{1, 2, 3\}$

The notion of a **partially ordered set** captures the minimum requirements for a binary relation on a set to be meaningfully considered a partial order.

**Definition 2.2.** Let  $X$  be a set. A **partial order** on  $X$  is a binary relation  $<$  on  $X$  such that

- (P1) for all  $x \in X$ ,  $x \not< x$  (*irreflexive*);
- (P2) for all  $x, y, z \in X$ , if  $x < y$  and  $y < z$ , then  $x < z$  (*transitive*).

The pair  $(X, <)$  is often simply called a **poset**. A partial order  $<$  on  $X$  is **linear** (also called *total*) if additionally

- (L) for all  $x, y \in X$ ,  $x < y$  or  $x = y$  or  $y < x$ .

If  $(X, <)$  is a poset one writes, as usual,  $x \leq y$  to express that either  $x < y$  or  $x = y$ . As we saw above, the usual order on the integers, rationals, and reals is linear, while the subset-ordering of the subsets of  $\{1, 2, 3\}$  is a partial order but not linear. Another example of a partial order that is not linear is the following: Let  $X = \mathbb{R}^2$ , and for  $x = (x_1, x_2)$  and  $y = (y_1, y_2)$  in  $\mathbb{R}^2$  put

$$x < y \iff \|x\| < \|y\|,$$

that is, we order vectors by their length. This order is not linear since for each length  $l > 0$ , there are infinitely many vectors of length  $l$  (which therefore cannot be compared).

**Trees from partial orders.** Let  $(T, <)$  be a partially ordered set.  $(T, <)$  is called a **tree** (as a partial order) if

- (T1) there exists an  $r \in T$  such that for all  $x \in T$ ,  $r \leq x$   
( $r$  is the *root* of the tree);
- (T2) for any  $x \in T$ , the set of *predecessors* of  $x$ ,  $\{y \in T: y < x\}$ , is finite and linearly ordered by  $<$ .

Note that not every poset is a tree. For example, in the set of all subsets of  $\{1, 2, 3\}$ , the predecessors of  $\{1, 2, 3\}$  are not linearly ordered. Often a poset also lacks a root element. For example, the usual ordering of the integers  $\mathbb{Z}$ ,  $\dots < -2 < -1 < 0 < 1 < 2 < \dots$ , satisfies neither (T1) nor (T2).

Trees arising from partial orders can be interpreted as graph-theoretic trees, as introduced in Section 1.2. In fact, the elements of  $T$  are called *nodes*, and sets of the form  $\{y \in T: y \leq x\}$  are called *branches*.

**Exercise 2.3.** Let  $(T, <)$  be a tree (partial order). Define a graph by letting the node set be  $T$  and connect two nodes if one is an *immediate predecessor* of the other. (Node  $s$  is an immediate predecessor of  $t$  if  $s < t$  and if for all  $u \in T$ ,  $u < t$  implies  $u \leq s$ .) Show that the resulting graph is a tree in the graph-theoretic sense.

As an example, consider the set  $\{0, 1\}^*$  of all *binary strings*. A binary string  $\sigma$  is a finite sequence of 0s and 1s, for example,

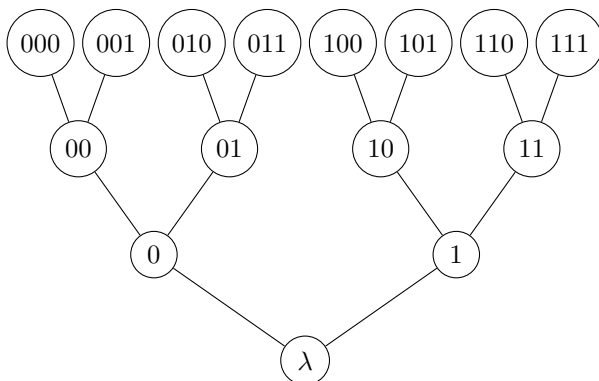
$$\sigma = 01100010101.$$

We order strings via the *initial segment relation*:  $\sigma < \tau$  if  $\sigma$  is shorter than  $\tau$  and the two strings agree on the bits of  $\sigma$ . For example, 011 is an initial segment of 01100, but 010 is not (the two strings disagree on the third bit). It is not hard to verify that (P1) and (P2) hold for this relation. Furthermore, the *empty string*  $\lambda$  is an initial segment of any other string and the initial segments of a string are linearly ordered by  $<$ ; for example, for  $\sigma = 010010$ ,

$$\lambda < 0 < 01 < 010 < 0100 < 01001 < \sigma.$$

Therefore,  $(\{0, 1\}^*, <)$  is a tree, the **full binary tree** (Figure 2.2).





**Figure 2.2.** The full binary tree up to strings of length 3

**Paths and König's lemma.** While we think of branches as something finite—due to (T2)—a sequence of them can give rise to an infinite branch, also called an *infinite path*.

**Definition 2.4.** An **infinite path** in a tree  $(T, \leq)$  is a sequence  $r = x_0 < x_1 < x_2 < \dots$  where all  $x_i \in T$  and for all  $i$ ,  $\{x \in T: x < x_i\} = \{x_0, x_1, \dots, x_{i-1}\}$ .

One can think of an infinite path as a sequence of elements on the tree where each element is a “one-step” extension of the previous one. In the full binary tree, an infinite path corresponds to an infinite sequence of zeros and ones.

It is clear that if a tree  $T$  has a path, then it must be infinite (i.e.  $T$  as a set is infinite). But does the converse hold? That is:

*If a tree is infinite, does it have a path?*

It is easy to give an example to show that this is not true. Consider the tree on  $\mathbb{N}$  where 0 is the root and every number  $n \geq 1$  is an immediate successor of 0 (an infinite fan of depth one).

**Definition 2.5.** A tree  $(T, \leq)$  is *finitely branching* if for every  $x \in T$ , there exist at most finitely many  $y_1, \dots, y_n \in T$  such that whenever  $z > x$  for some  $z \in T$ , we have  $z \geq y_i$  for some  $i$ . (That is, every  $x \in T$  has at most finitely many immediate successors.)

Our example of the full binary tree  $\{0, 1\}^*$  is finitely branching.

**Theorem 2.6** (König's lemma). *If an infinite tree  $(T, <)$  is finitely branching, then it has an infinite path.*

**Proof.** We construct an infinite sequence on  $T$  by induction. Let  $x_0 = r$ . Given any  $x \in T$ , let  $T_x$  denote the part of  $T$  “above”  $x$ , i.e.

$$T_x = \{y \in T : x \leq y\}.$$

$T_x$  inherits a tree structure from  $T$  by letting  $x$  be its root. Note that  $T = T_r$ . Let  $y_1^{(r)}, \dots, y_n^{(r)}$  be the immediate successors of  $r$  in  $T$ . Since we assume  $T$  to be infinite, by the infinite pigeonhole principle, one of the trees  $T_{y_1^{(r)}}, \dots, T_{y_n^{(r)}}$  must be infinite, say  $T_{y_i^{(r)}}$ . Put  $x_1 = y_i^{(r)}$ .

We can now iterate this construction, using the infinite pigeonhole principle on the finitely many disjoint trees above the immediate successors of  $x_1$ , and so on. Since we always maintain an infinite tree above our current  $x_k$ , the construction will carry on indefinitely and we obtain an infinite sequence  $x_0 < x_1 < x_2 < \dots$ , an infinite path through  $T$ .  $\square$

**Proving finite results from infinite ones.** We can use König's lemma and the infinite Ramsey theorem (Theorem 2.1) to prove the general finite Ramsey theorem (Theorem 1.31).

Assume, for the sake of contradiction, that for some  $k, p$ , and  $r$ , the statement of the finite Ramsey theorem does not hold. That is, for all  $n$  there exists at least one coloring  $c_n : [n]^p \rightarrow 1, \dots, r$  such that no monochromatic subsets of size  $k$  exist. Collect these counterexamples  $c_n$ , for all  $n$ , in a single set  $T$ , and order them by *extension*: Let  $c_m < c_n$  if and only if  $m < n$  and the restriction of  $c_n$  to  $[m]^p$  is equal to  $c_m$ , i.e.  $c_m$  extends  $c_n$  as a function.

We make three crucial observations:

- (1)  $(T, <)$  is a tree; the root  $r$  is the empty function and the predecessors of a coloring  $c_n \in T$  are the restricted colorings

$$\emptyset < c_n \upharpoonright_{[p]^p} < c_n \upharpoonright_{[p+1]^p} < \dots < c_n \upharpoonright_{[n-1]^p} < c_n.$$

- (2)  $T$  is finitely branching; this is clear since for every  $n$  there are only finitely many functions  $c : [n]^p \rightarrow \{1, \dots, r\}$  at all.

- (3)  $T$  is infinite; this is true because we are assuming that at least one such coloring exists for all  $n$ .

Therefore, we can apply König's lemma and obtain an infinite path

$$\emptyset < c_p < c_{p+1} < c_{p+2} < \dots$$

Since each  $c_n$  on our path is an extension of all the previous functions on the path, we can construct a well-defined function  $C : [\mathbb{N}]^p \rightarrow \{1, \dots, r\}$  which has the property that  $C|_{[n]^p} = c_n$ . That is, if  $X$  is a  $p$ -set of  $\mathbb{N}$  where  $N$  is the largest integer in  $X$ , then  $C(X) = c_N(X)$ .

Now we have a coloring on  $\mathbb{N}$  and we can apply the infinite Ramsey theorem to deduce that there exists an infinite subset  $H \subseteq \mathbb{N}$  such that  $[H]^p$  is monochromatic.

We write  $H = \{h_1 < h_2 < h_3 < \dots\}$  and let  $N = h_k$ . Since  $H$  is monochromatic for the coloring  $C$ , so is every subset of  $H$ . In particular,  $H_k := \{h_1 < h_2 < \dots < h_k\}$  is a monochromatic subset of size  $k$  for the coloring  $C$ . But we have  $C(H_k) = c_N(H_k)$ , and therefore  $c_N$  has a monochromatic subset of size  $k$ , which contradicts our initial assumption.

**A blueprint for compactness arguments.** We can use the previous proof as a prototype for future uses of compactness. Suppose we have a statement  $P(\vec{\pi})$  with a vector of parameters  $\vec{\pi}$  that asserts the existence of a certain object, and we want to show that for a sufficiently large finite set  $\{1, \dots, N\}$ ,  $P(\vec{\pi})$  is always true. Suppose further that we have shown  $P(\vec{\pi})$  is true for  $\mathbb{N}$ . Here is a blueprint for an argument using compactness:

- (1) Assume, for a contradiction, that the finite version of  $P$  fails.
- (2) Then we can find counterexamples for every set  $[n]$ .
- (3) Collect these counterexamples in a set  $T$ , order them by extension, and show that under this ordering  $T$  forms an infinite, finitely branching tree.
- (4) Apply König's lemma to obtain an infinite path in  $T$ , which corresponds to an instance of our statement  $P(\vec{\pi})$  for  $\mathbb{N}$ .
- (5) Since  $P(\vec{\pi})$  is true for  $\mathbb{N}$ , we can choose a witness example for this instance.

- (6) By restricting the witness to a sufficiently large subset, we obtain a contradiction to the fact that  $T$  contains only counterexamples to  $P$ .

We will later see that the introduction of infinitary methods opens a fascinating metamathematical door: There exist “finitary” statements (in the sense that all objects involved—sets, functions, numbers—are finite) for which *only infinitary proofs exist*. In a certain sense, the finite sets whose existence the infinitary methods establish are so huge that a “finitary accounting method” cannot keep track of them. We will investigate this phenomenon in Chapter 4.

### 2.3. Some topology

If you have learned about compactness in a topology or analysis class, you might be wondering why we are using this word. We will show that König’s lemma can be rephrased in terms of sequential compactness. While we will provide all the necessary definitions, we can of course not even scratch the surface of the theory of metric spaces and topology. For a reader who has no previous experience in this area, we recommend consulting on the side one of the numerous textbooks on analysis or topology, for example [51].

**Metric spaces.** The concept of a metric space is a generalization of distance. We use it to describe how close or far two elements in a set are from each other.

**Definition 2.7.** A **metric** on a set  $X$  is any function  $d : X^2 \rightarrow \mathbb{R}$  such that for all  $x, y, z \in X$ :

- (1)  $d$  is *non-negative*, that is,  $d(x, y) \geq 0$ , and moreover  $d(x, y) = 0$  if and only if  $x = y$ ;
- (2)  $d$  is *symmetric*, that is,  $d(x, y) = d(y, x)$ ; and
- (3)  $d$  satisfies the *triangle inequality*, that is,  $d(x, z) \leq d(x, y) + d(y, z)$ .

A **metric space**  $(X, d)$  is a set  $X$  together with a metric  $d$ .

The prototypical examples of metric spaces include  $\mathbb{R}$  with the standard distance  $d(x, y) = |x - y|$ , or  $\mathbb{R}^2$  with the distance function

which comes from the Pythagorean theorem,

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}.$$

These are both examples of the  $n$ -dimensional Euclidean metric: For  $x, y \in \mathbb{R}^n$ ,

$$d(x, y) = \left( \sum_{i=1}^n (x_i - y_i)^2 \right)^{1/2}.$$

This metric is the natural function with which we associate the idea of “distance” between two points. However, there are many other important metrics. For any non-empty set  $X$ , we can consider the *discrete metric*, defined by

$$d(x, y) = \begin{cases} 0 & \text{if } x = y \\ 1 & \text{if } x \neq y \end{cases}$$

This is a metric where every two distinct points are the same distance away—a rather crude measure of distance.<sup>1</sup> One can refine this idea to take the combinatorial structure into account. For any connected graph  $G$ , we can define a metric on  $G$  by letting

$$d(v, w) = \text{length of the shortest path between } v \text{ and } w.$$

**Exercise 2.8.** Show that  $d(v, w)$  defines a metric on a connected graph  $G$ .

**Neighborhoods and open sets.** Given a point  $x$  in a metric space  $(X, d)$  and a real number  $\epsilon$ , the  $\epsilon$ -**neighborhood of  $x$**  is the set

$$B_\epsilon(x) := \{y \in X : d(x, y) < \epsilon\},$$

i.e. all points which are less than  $\epsilon$  away from  $x$ .

For example, an open ball  $B_\epsilon(x)$  (with respect to the Euclidean metric) on the real line  $\mathbb{R}$  is just an open interval of the form  $(x - \epsilon, x + \epsilon)$ .

An **open set** is any set  $U \subseteq X$  where, for every  $x$  in  $U$ , there exists an  $\epsilon > 0$  such that  $B_\epsilon(x)$  is contained entirely in  $U$ . The union of open sets is also an open subset. The complement of an open set is called a **closed set**. Note that in any metric space  $(X, d)$ , the entire set  $X$  and the empty set  $\emptyset$  are both open and closed.

---

<sup>1</sup>The mathematician Stanislaw Ulam once wrote that Los Angeles is a discrete metric space, where the distance between any two points is an hour’s drive [67].

We can now define the topological notion of compactness using open coverings: A collection of open subsets  $\{U_i\}_{i \in I}$  is defined to be an *open cover* of  $Y$  if  $Y \subseteq \bigcup U_i$ .

**Definition 2.9.** A subset  $Y$  in a metric space  $(X, d)$  is **compact** if whenever  $\{U_i\}_{i \in I}$  is an open cover of  $Y$ , there is a finite subset  $J \subseteq I$  such that  $\{U_j\}_{j \in J}$  is also an open cover of  $Y$ , in other words, if every open cover has a finite subcover.

Suppose we cover  $\mathbb{R}$  with balls  $B_\varepsilon(x)$  for an arbitrarily small  $\varepsilon$ , so that every  $x \in \mathbb{R}$  contributes an interval  $(x - \varepsilon, x + \varepsilon)$ . Together these intervals clearly cover all of  $\mathbb{R}$ . It is easy to see that this covering has no finite subcover, as choosing finitely many of the  $B_\varepsilon(x)$  covers at most an interval of the form  $(-M, M)$ , where  $M < \infty$ . Therefore,  $\mathbb{R}$  with the Euclidean metric is not a compact space.

This example suggests that compact sets, although possibly infinite (as sets), should somehow be considered “finite”. In Euclidean space this is confirmed by the **Heine-Borel theorem**: A set  $X \subseteq \mathbb{R}^n$  is compact if and only if it is closed and bounded, that is,  $X$  is contained in some  $n$ -dimensional cube  $(-M, M)^n$ .

**Sequential compactness.** Given a sequence of points  $(x_i)$  in a metric space  $(X, d)$ , the sequence **converges** to a point  $x$  if

$$\lim_{i \rightarrow \infty} d(x_i, x) = 0.$$

A metric space  $(X, d)$  is **sequentially compact** if every sequence has a convergent subsequence. In metric spaces, the notion of compactness and sequential compactness are equivalent (see [51]), although this is not true for general topological spaces.

In  $\mathbb{R}^n$  (with the Euclidean metric) the equivalence of compactness and sequential compactness follows from the *Bolzano-Weierstrass theorem*: Every bounded sequence has a convergent subsequence.

**Exercise 2.10.** Use the infinite Ramsey theorem to prove that every sequence in  $\mathbb{R}$  has a monotone subsequence. As any bounded, monotone sequence converges in  $\mathbb{R}$ , this implies the Bolzano-Weierstrass theorem.

**Infinite trees as metric spaces.** Given an infinite, finitely branching tree  $T$ , König's lemma tells us that there will be at least one infinite path. We can collect all of the infinite paths into a set, denoted by  $[T]$ . It will be useful to visualize the elements of  $[T]$  not just as paths on a tree, but also as infinitely long sequences of nodes.

Suppose we have two elements of  $[T]$ ,  $\vec{s}$  and  $\vec{t}$ , and their sequences of nodes

$$\vec{s} = \{r = s_0 < s_1 < s_2 < \dots\} \text{ and } \vec{t} = \{r = t_0 < t_1 < t_2 < \dots\}.$$

To define a notion of distance, two paths will be regarded as "close" if their sequences agree for a long time. We put, for distinct  $\vec{s}$  and  $\vec{t}$ ,

$$D_{\vec{s}, \vec{t}} = \min\{i \geq 0: s_i \neq t_i\}$$

and then define our distance function as

$$d(\vec{s}, \vec{t}) = \begin{cases} 0 & \text{if } s = t, \\ 2^{-D_{\vec{s}, \vec{t}}} & \text{if } s \neq t. \end{cases}$$

We claim that  $d$  is a metric on  $[T]$  and will call it the **path metric** on  $[T]$ . Non-negativity and symmetry are clear from the definition of  $d$ . To verify the triangle inequality, suppose  $\vec{s}, \vec{t}$ , and  $\vec{u}$  are pairwise distinct paths in  $[T]$ . (If any two of the sequences are identical, the statement is easy to verify.) We distinguish two cases:

**Case 1:**  $D_{\vec{s}, \vec{u}} \leq D_{\vec{s}, \vec{t}}$ .

This means that  $\vec{s}$  agrees at least as long with  $\vec{t}$  as with  $\vec{u}$ .

But this implies that  $\vec{t}$  agrees with  $\vec{u}$  precisely as long as  $\vec{s}$  does, which in turn means  $D_{\vec{t}, \vec{u}} = D_{\vec{s}, \vec{u}}$ , and hence

$$d(\vec{s}, \vec{u}) = 2^{-D_{\vec{s}, \vec{u}}} = 2^{-D_{\vec{t}, \vec{u}}} \leq 2^{-D_{\vec{s}, \vec{t}}} + 2^{-D_{\vec{t}, \vec{u}}} = d(\vec{s}, \vec{t}) + d(\vec{t}, \vec{u}).$$

**Case 2:**  $D_{\vec{s}, \vec{u}} > D_{\vec{s}, \vec{t}}$ .

In this case  $\vec{s}$  agrees with  $\vec{u}$  longer than it agrees with  $\vec{t}$ . But this directly implies that

$$d(\vec{s}, \vec{u}) = 2^{-D_{\vec{s}, \vec{u}}} < 2^{-D_{\vec{s}, \vec{t}}} \leq 2^{-D_{\vec{s}, \vec{t}}} + 2^{-D_{\vec{t}, \vec{u}}} = d(\vec{s}, \vec{t}) + d(\vec{t}, \vec{u}).$$

What do the neighborhoods  $B_\varepsilon(s)$  look like for this metric? A sequence  $t$  is in  $B_\varepsilon(s)$  if and only if  $2^{-D_{\vec{s}, \vec{t}}} < \varepsilon$ , which means  $D_{\vec{s}, \vec{t}} > -\log_2 \varepsilon$ . Hence  $t$  is in the  $\varepsilon$ -neighborhood of  $\vec{s}$  if and only if it agrees with  $\vec{s}$  on the first  $\lceil -\log_2 \varepsilon \rceil$  bits.

**Exercise 2.11.** Draw a picture of  $B_{1/8}(10101010\dots)$ .

**König's lemma and compactness.** We can now interpret König's lemma as an instance of topological compactness.

**Theorem 2.12.** *If  $T$  is a finitely branching tree, then  $[T]$  with the path metric is a compact metric space.*

**Proof.** Assume that  $[T]$  is not empty. Let  $(\vec{s}_n)$  be a sequence in  $[T]$ . (We use the vector notation, after all.) This means that every  $\vec{s}_n$  is itself a sequence  $r = s_n^0 < s_n^1 < s_n^2 < \dots$  in  $T$ . We will construct a convergent subsequence  $(\vec{s}_{n_i})$ . It follows that  $[T]$  is sequentially compact and therefore compact.

Let  $T^*$  be the subtree of  $T$  defined as follows: Let  $\sigma \in T^*$  if and only if  $\sigma = r$  or  $\sigma = s_n^i$  for some  $i$  and  $n$ ; that is,  $\sigma$  is in  $T^*$  if and only if it occurs in one of the paths  $\vec{s}_n$ .

We observe that  $T^*$  is infinite, because each  $\vec{s}_n$  is an infinite path. It is also finitely branching as it is a subtree of the finitely branching tree  $T$ .

By König's lemma,  $T^*$  has an infinite path  $\vec{t}$  of the form  $r = t^0 < t^1 < t^2 < t^3 < \dots$ . We use this path to identify a subsequence of  $(\vec{s}_n)$  that converges to  $\vec{t} \in [T]$ .

The path  $\vec{t}$  is built from nodes that occur as a node in *some* path  $\vec{s}_n$ . This means that for every  $i$  there exists an  $n$  such that  $t^i = s_n^i$ . We use this to define a subsequence of  $(s_n)$  as follows: Let

$$n_i = \min\{n: s_n^i = t^i\}.$$

*Claim:*  $(\vec{s}_{n_i})$  converges to  $\vec{t}$ .

By the definition of the path metric  $d$ ,

$$d(\vec{s}_{n_i}, \vec{t}) \xrightarrow{i \rightarrow \infty} 0 \quad \text{iff} \quad \vec{s}_{n_i} \text{ and } \vec{t} \text{ agree} \\ \text{on longer and longer segments.}$$

But this is built into the definition of  $(\vec{s}_{n_i})$ : We have

$$s_{n_i}^i = t^i,$$

and since initial segments in trees are unique, this implies that

$$s_{n_i}^0 = t^0, s_{n_i}^1 = t^1, \dots, s_{n_i}^{i-1} = t^{i-1}.$$



Therefore,  $\vec{s}_{n_i}$  and  $\vec{t}$  have an agreement of length  $i$ , and thus

$$d(\vec{s}_{n_i}, \vec{t}) \xrightarrow{i \rightarrow \infty} 0.$$

□

**Exercise 2.13.** Every real number  $x \in [0, 1]$  has a *dyadic expansion*  $\vec{s}^{(x)} \in \{0, 1\}^\infty$  such that

$$x = \sum_i s_i^{(x)} 2^{-i}.$$

The expansion is unique except when  $x$  is of the form  $m/2^n$ , with  $m, n \leq 1$  integer. To make it unique, we require the dyadic expansion to eventually be constant  $\equiv 1$ .

Show that a sequence  $(x_i)$  of real numbers in  $[0, 1]$  converges with respect to the Euclidean metric if and only if  $(\vec{s}^{(x_i)})$  converges with respect to the path metric  $d$ .

## 2.4. Ordinals, well-orderings, and the axiom of choice

The natural numbers form the mathematical structure that we use to count things. In the process of counting, we bestow an order on the objects we are counting. We speak of the first, the second, the third element, and so on. The realm of the natural numbers is sufficient as long as we count only finite objects. But how can we count infinite sets? This is made possible by the theory of *ordinal numbers*.

**Properties of ordinals.** Ordinal numbers are formally defined using set theory, as transitive sets that are well-ordered by the element relation  $\in$ . We will not introduce ordinals formally here, but instead simply list some crucial properties of ordinal numbers that let us extend the counting process into the infinite realm. For a formal development of ordinals, see for example [35].

- (O1) Every natural number is an ordinal number.
- (O2) The ordinal numbers are *linearly ordered*, and 0 is the least ordinal.
- (O3) Every ordinal has a unique *successor* (the *next* number); that is, for every ordinal  $\alpha$  there exists an ordinal  $\beta > \alpha$  such that

$$(\forall \gamma) \alpha < \gamma \Rightarrow \beta \leq \gamma.$$

The successor of  $\alpha$  is denoted by  $\alpha + 1$ .

- (O4) For every set  $A$  of ordinals, there exists an ordinal  $\beta$  that is the *least upper bound* of  $A$ , that is,

for all  $\alpha \in A$ ,

$\alpha \leq \beta$  and  $\beta$  is the least number with this property.

If we combine (O1) and (O4), there must exist a least ordinal that is *greater than every natural number*. This number is called  $\omega$ . (O3) tells us that  $\omega$  has a successor,  $\omega + 1$ , which in turn has a successor itself,  $(\omega + 1) + 1$ , which we write as  $\omega + 2$ . We can continue this process and obtain

$$\omega, \omega + 1, \omega + 2, \omega + 3, \dots, \omega + n, \dots$$

But the ordinals do not stop here. Applying (O4) to the set  $\{\omega + n : n \in \mathbb{N}\}$ , we obtain a number that is greater than any of these, denoted by  $\omega + \omega$ . Here is a graphical representation of these first infinite ordinals:

$$\begin{array}{ll} \omega & \circ \circ \circ \dots \\ \omega + 1 & \circ \circ \circ \dots \bullet \\ \omega + 2 & \circ \circ \circ \dots \bullet \circ \\ \omega + \omega & \circ \circ \circ \dots \bullet \circ \circ \dots \end{array}$$

One can continue enumerating:

$$\omega + \omega + 1, \omega + \omega + 2, \dots, \omega + \omega + \omega, \dots, \omega + \omega + \omega + \omega, \dots$$

In this process we encounter two types of ordinals.

- *Successor ordinals*: Any ordinal  $\alpha$  for which there exists an ordinal  $\beta$  such that  $\alpha = \beta + 1$ . Examples include all natural numbers greater than 0,  $\omega + 1$ , and  $\omega + \omega + 3$ .

- *Limit ordinals*: Any ordinal that is not a successor ordinal, for example 0,  $\omega$ , and  $\omega + \omega + \omega$ .

Since there is always a successor, the process never stops. An attentive reader will remark that, on the other hand, we could apply (O4) to the set of *all* ordinals. Would this not yield a contradiction? This is known as the *Burali-Forti paradox*. We have to be careful which mathematical objects we consider a set and which not. And in our case we say:

*There is no set of all ordinals.*

There are simply *too many* to form a set. The ordinals form what is technically referred to as a *proper class*, which we will denote by Ord. Other examples of proper classes are the class of all sets and the class of all sets that do not contain themselves (this is *Russell's paradox*). The assumption that either of these is a set leads to a contradiction similar to assuming that a set of all ordinals exists. Classes behave in many ways like sets—for example, we can talk about the *elements* of a class. But these elements cannot be other classes; classes are *too large* to be an element of something.

**Ordinal arithmetic.** The counting process above indicates that we can define arithmetical operations on ordinals similar to the operations we have on the natural numbers. For the natural numbers, addition and multiplication are defined by *induction*, by means of the following identities.

- $m + (n + 1) = (m + n) + 1$ ,
- $m \cdot (n + 1) = m \cdot n + m$ .

For ordinals, we use *transfinite induction*. This is essentially the same as “ordinary” induction, except that we also have to account for limit ordinals in the induction step.

*Addition of ordinals.*

$$\begin{aligned} \alpha + 0 &= \alpha, \\ \alpha + (\beta + 1) &= (\alpha + \beta) + 1, \\ \alpha + \lambda &= \sup\{\alpha + \gamma : \gamma < \lambda\} \quad \text{if } \lambda \text{ is a limit ordinal.} \end{aligned}$$

There is an important aspect in which ordinal addition behaves very differently from addition for natural numbers. Take, for example,

$$1 + \omega = \sup\{1 + n: n \in \mathbb{N}\} = \sup\{m: m \in \mathbb{N}\} = \omega \neq \omega + 1.$$

So ordinal addition is **not commutative**; that is, it is not true in general that  $\alpha + \beta = \beta + \alpha$ .

*Multiplication of ordinals.*

$$\begin{aligned} \alpha \cdot 0 &= 0, \\ \alpha \cdot (\beta + 1) &= (\alpha \cdot \beta) + \alpha, \\ \alpha \cdot \lambda &= \sup\{\alpha \cdot \gamma: \gamma < \lambda\} \quad \text{if } \lambda \text{ is a limit ordinal.} \end{aligned}$$

Let us calculate a few examples.

$$\omega \cdot 2 = \omega(1 + 1) = \omega + \omega,$$

and similarly  $\omega \cdot 3 = \omega + \omega + \omega$ ,  $\omega \cdot 4 = \omega + \omega + \omega + \omega$ , and so on. On the other hand,

$$2 \cdot \omega = \sup\{2n: n \in \mathbb{N}\} = \sup\{m: m \in \mathbb{N}\} = \omega.$$

(Think of  $\alpha \cdot \beta$  as “ $\alpha$  repeated  $\beta$ -many times.”) Hence ordinal multiplication is not commutative either. Moreover, we have

$$\omega \cdot \omega = \sup\{\omega \cdot n: n \in \mathbb{N}\}.$$

Hence we can view  $\omega \cdot \omega$  as the *limit* of the sequence

$$\omega, \omega + \omega, \omega + \omega + \omega, \omega + \omega + \omega + \omega, \dots$$

Similarly, we can now form the sequence

$$\omega, \omega \cdot \omega, \omega \cdot \omega \cdot \omega, \dots$$

What should the limit of this sequence be? If we let ourselves be guided by the analogy of the finite world of natural numbers, it ought to be

$$\omega^\omega.$$

Just as multiplication is obtained by iterating addition, exponentiation is obtained by iterating multiplication. We can do this for ordinals, too.

*Exponentiation of ordinals.*

$$\begin{aligned}\alpha^0 &= 1, \\ \alpha^{\beta+1} &= \alpha^\beta \cdot \alpha, \\ \alpha^\lambda &= \sup\{\alpha^\gamma : \gamma < \lambda\} \quad \text{if } \lambda \text{ is a limit ordinal.}\end{aligned}$$

By this definition,  $\omega^\omega$  is the limit of  $\omega, \omega^2, \omega^3, \dots$  indeed.

Using exponentiation, we can form the sequence

$$\omega, \omega^\omega, \omega^{\omega^\omega}, \dots$$

The limit of this sequence is called  $\varepsilon_0$ . It is the least ordinal with the property that

$$\omega^{\varepsilon_0} = \varepsilon_0.$$

This property seems rather counterintuitive, since in the finite realm  $m^n$  is much larger than  $n$  as  $m$  and  $n$  grow larger and larger.

**Ordinals, well-orderings and the axiom of choice.** Let's look at the sequence of ordinals we have encountered so far:

$$0 < 1 < 2 < \dots < \omega < \omega + 1 < \dots < \omega + \omega < \dots < \omega\omega < \dots < \omega^\omega < \dots < \varepsilon_0.$$

Property (O2) requires that the ordinal numbers be linearly ordered, and our initial list above clearly reflects this property. It turns out that the ordinals are a linear ordering of a special kind, a *well-ordering*.

**Definition 2.14.** Assume that  $(S, <)$  is a linearly ordered set. We say that  $(S, <)$  is a **well-ordering** if every non-empty subset of  $S$  has a  $<$ -least element.

In particular, this means that  $S$  itself must have a  $<$ -minimal element. Therefore,  $\mathbb{Z}, \mathbb{Q}$ , and  $\mathbb{R}$  are not well-orderings. On the other hand, the natural numbers with their standard ordering are a well-ordering—in every non-empty subset of  $\mathbb{N}$  there is a least number. If we restrict the rationals (or reals) to  $[0, 1]$ , we do not get a well-ordering, since the subset  $\{1/n : n \geq 1\}$  does not have a minimal element *in the subset*.

The last example hints at an equivalent characterization of well-orderings.

**Proposition 2.15.** *A linear ordering  $(S, <)$  is a well-ordering if and only if there does not exist an infinite descending sequence*

$$s_0 > s_1 > s_2 > \cdots$$

in  $S$ .

**Proof.** It is clear that if such a sequence exists, then the ordering cannot be a well-ordering, since the set  $\{s_0, s_1, s_2, \dots\}$  does not have a minimal element.

On the other hand, suppose  $(S, <)$  is not a well-ordering. Then there exists a non-empty subset  $M \subseteq S$  that has no minimal element with respect to  $<$ . We use this fact to construct a descending sequence  $s_0 > s_1 > s_2 > \cdots$  in  $M$ .

Let  $s_0$  be any element of  $M$ ; then  $s_0$  cannot be a minimum of  $M$ , since  $M$  does not have a minimum. Hence we can find an element  $s_1 < s_0$  in  $M$ . But  $s_1$  cannot be a minimum of  $M$  either, and hence we can find  $s_2 < s_1$ , and so on.  $\square$

This characterization shows us that well-orderings have a strong *asymmetry*. While we can “count up” unboundedly through an infinite well-ordering, we cannot “count down” in the same way. No matter how we do it, after *finitely many steps* we reach the end, i.e. the minimal element.

We now verify that the ordinals are well-ordered by  $<$ .

**Proposition 2.16.** *Any set of ordinal numbers is well-ordered by  $<$ .*

**Proof.** Suppose  $S$  is a set of ordinals not well-ordered by  $<$ . Then there is an infinite descending sequence  $\alpha_0 > \alpha_1 > \cdots$  in  $S$ . Let  $M$  be the set of all ordinals smaller than every element of the sequence, i.e.

$$M = \{\beta: \beta < \alpha_i \text{ for all } i\}.$$

By (O4),  $M$  has a least upper bound  $\gamma$ . Now  $\gamma$  has to be below all  $\alpha_i$  as well, because otherwise every  $\alpha_i < \gamma$  would be a smaller upper bound for  $M$ . Therefore,  $\gamma \in M$ .

By (O3), there exists a smallest ordinal greater than  $\gamma$ , namely its successor  $\gamma + 1$ . But  $\gamma + 1$  cannot be in  $M$ , for otherwise  $\gamma$  would not be an upper bound for  $M$ . Hence there must exist an  $i$  such that

$\alpha_i < \gamma + 1$ . But since  $\gamma + 1$  is the smallest ordinal greater than  $\gamma$ , it follows that  $\alpha_{i+1} < \alpha_i \leq \gamma$ , contradicting that  $\gamma \in M$ .  $\square$

As we have mentioned above, the collection of all ordinals does not form a set. But if we fix any ordinal  $\beta$ , the **initial segment** of Ord up to  $\beta$ ,

$$\text{Ord} \upharpoonright_{\beta} = \{\alpha : \alpha \text{ is an ordinal and } \alpha < \beta\},$$

does form a set, and the proof as above shows that this initial segment is well-ordered.

Of course not every set that comes with a partial ordering is a well-ordering (or even a linear ordering). But if we are given the freedom to impose our own ordering, can every set be well-ordered?

This appears to be clear for finite sets. Suppose  $S$  is a non-empty finite set. We pick any element, declare it to be the minimal element, and pick another (different) element, which we declare to be the minimum of the remaining elements. We continue this process till we have covered the whole set. What we are really doing is constructing a bijection  $\pi : \{0, 1, \dots, n-1\} \rightarrow S$ , where  $n$  is a natural number. The well-ordering of  $S$  is then given by

$$(2.1) \quad s < t \text{ :} \Leftrightarrow \pi^{-1}(s) < \pi^{-1}(t).$$

Can we implement a similar process for infinite sets? We have introduced the ordinals as a transfinite analogue of the natural numbers, so what we could try to do is find a bijection  $\pi$  between our set and an initial segment of the ordinals of the form

$$\{\alpha : \alpha \text{ is an ordinal and } \alpha < \beta\},$$

where  $\beta$  is an ordinal. Then, as can be easily verified, (2.1) again defines a well-ordering on our set.

Another way to think of a well-ordering of a set  $S$  is as an *enumeration of  $S$  indexed by ordinals*: If we let  $s_{\xi} = \pi(\xi)$ , then

$$S = \{s_0 < s_1 < s_2 < \dots\} = \{s_{\xi} : \xi < \beta\}$$

for some ordinal  $\beta$ .

If a set is well-ordered, one can in turn show that the ordering is isomorphic to the well-ordering of an initial segment of Ord.

**Proposition 2.17.** *Suppose  $(A, <)$  is a well-ordering. Then there exists a unique ordinal  $\beta$  and a bijection  $\pi : A \rightarrow \{\alpha : \alpha < \beta\}$  such that for all  $a, b \in A$*

$$a < b \Leftrightarrow \pi(a) < \pi(b).$$

We call  $\beta$  the **order type** of the well-ordering  $(A, <)$ .

You can try to prove Proposition 2.17 yourself as an exercise, but you will need to establish some more (albeit easy) properties of ordinals and well-orderings along the way. You can also look up a proof, for example in [35].

One can furthermore show that the order type of  $\text{Ord} \upharpoonright_\beta$  is  $\beta$ . For this reason, ordinals are usually *identified with their initial segments in Ord*.

Returning to our question above, is it possible to well-order any set? The answer to this question is, maybe somewhat surprisingly, a hesitant “it depends”.

**The axiom of choice and the well-ordering principle.** Intuitively, an argument for the possibility of well-ordering an arbitrary set  $S$  might go like this:

Let  $S' = S$ . If  $S' \neq \emptyset$ , let  $\xi$  be the least ordinal to which we have not yet assigned an element of  $S$ . Pick any  $x \in S'$ , map  $\xi \mapsto x$ , put  $S' := S' \setminus \{x\}$ , and iterate.

The problem here is the “pick any  $x \in S'$ ”. It seems an innocent step; after all,  $S'$  is assumed to be non-empty. But we have to look at the fact that we *repeatedly apply* this step. In fact, what we seem to assert here is the existence of a special kind of **choice function**: There exists a function  $f$  whose domain is the *set of all non-empty subsets of  $S$* ,  $\mathcal{P}^0(S) = \{S' : \emptyset \neq S' \subseteq S\}$ , such that for all  $S' \in \mathcal{P}^0(S)$ ,

$$f(S') \in S'.$$

Indeed, equipped with such a function, we can formalize our argument above.

If  $S = \emptyset$ , we are done. So assume  $S \neq \emptyset$ . Put  $s_0 = f(S)$ .



Suppose now we have enumerated elements

$$\{s_\xi : \xi < \alpha\}$$

from  $S$ . If  $S \setminus \{s_\xi : \xi < \alpha\}$  is non-empty, put

$$s_\alpha = f(S \setminus \{s_\xi : \xi < \alpha\}).$$

Now iterate. This procedure has to stop at some ordinal, i.e. there exists an ordinal  $\beta$  such that

$$S = \{s_\xi : \xi < \beta\}.$$

If not, that is, if the procedure traversed *all* ordinals, we would have constructed an injection  $F : \text{Ord} \rightarrow S$ . Using some standard axioms about sets, this would imply that, since  $\text{Ord}$  is not a set,  $S$  cannot be a set (it would be as large as the ordinals, which form a proper class), which is a contradiction.

Does such a choice function exist? Most mathematicians are comfortable to assume this, or at least they do not feel that there is overwhelming evidence against it. It turns out, however, that the existence of general choice functions is a mathematical principle that cannot be reduced to or proved from other, more evident principles (this is the result of some seminal work on the foundations of mathematics first by Gödel [20] and then by Cohen [8, 9]). It is therefore usually stated as an *axiom*.

**Axiom of choice (AC):** Every family of non-empty sets has a choice function. That is, if  $\mathcal{S}$  is a family of sets and  $\emptyset \notin \mathcal{S}$ , then there exists a function  $f$  on  $\mathcal{S}$  such that  $f(S) \in S$  for all  $S \in \mathcal{S}$ .

The axiom of choice is equivalent to the following principle.

**Well-ordering principle (WO):** Every set can be well-ordered.

We showed above that (AC) implies (WO). It is a nice exercise to show the converse.

**Exercise 2.18.** Derive (AC) from (WO).

There are some consequences of the axiom of choice that are, to say the least, puzzling. Arguably the most famous of these is the

**Banach-Tarski paradox:** Assuming the axiom of choice, it is possible to partition a unit ball in  $\mathbb{R}^3$  into finitely many pieces and rearrange the pieces so that we get *two* unit balls in  $\mathbb{R}^3$ .

Why has the Banach-Tarski paradox not led to an outright rejection of the axiom of choice, as this consequence clearly seems to run counter to our geometric intuition?

The reason is that our intuitions about notions such as *length* and *volume* are not so easy to formalize mathematically. The pieces obtained in the Banach-Tarski decomposition of a ball are what is called *non-measurable*, meaning essentially that the concept of volume in Euclidean space as we usually think about it (the *Lebesgue measure*) is not applicable to these pieces.

For now, let us just put on record that the use of the axiom of choice may present some foundational issues. By using a choice function without specifying further the specific objects which are chosen, the axiom introduces a non-constructive aspect into proofs. For this reason, one often tries to clarify whether the axiom of choice is needed in its full strength, whether it can be replaced by weaker (and foundationally less critical) principles such as the axiom of countable choice ( $AC_\omega$ ) or the axiom of dependent choice (DC), or whether it can be avoided altogether (for example by giving an explicit, constructive proof).

The book by Jech [36] is an excellent source on many questions surrounding the axiom of choice.

## 2.5. Cardinality and cardinal numbers

We introduced ordinals as a continuation of the counting process through the transfinite. In the finite realm, one of the main purposes of counting is to establish cardinalities. We count a finite set by assigning its elements successive natural numbers. In other words,

to count a finite set  $S$  means to put the elements of  $S$  into a one-to-one correspondence with the set  $\{0, \dots, n-1\}$ , for some natural number  $n$ . In this case we say  $S$  has cardinality  $n$ .

How can this be generalized to infinite sets? The basic idea is that:

Two sets have the same cardinality if there is a bijection (a mapping that is one-to-one and onto) between them.

For example, the sets  $\{1, 2, 3, 4, 5\}$  and  $\{6, 7, 8, 9, 10\}$  have the same cardinality. In the finite realm, it is impossible for a set to be a proper subset of another set yet have the same cardinality as the other set. This is no longer the case for infinite sets. The set of integers has the same cardinality as the set of even integers, as witnessed by the bijection  $z \mapsto 2z$ .

A very interesting case is  $\mathbb{N}$  versus  $\mathbb{N} \times \mathbb{N}$ . While  $\mathbb{N}$  is not a subset of  $\mathbb{N} \times \mathbb{N}$ , we can embed it via the mapping  $n \mapsto (n, 0)$  as a proper subset of  $\mathbb{N} \times \mathbb{N}$ . But there is actually a bijection between the two sets, the **Cantor pairing function**

$$(x, y) \mapsto \langle x, y \rangle = \frac{(x+y)^2 + 3x + y}{2}.$$

**Exercise 2.19.** (a) Draw the points of  $\mathbb{N} \times \mathbb{N}$  in a two-dimensional grid. Start at  $(0, 0)$ , which maps to 0, and find the point which maps to 1. Connect the two with an arrow. Next find the point that maps to 2, and connect it by an arrow to the point that maps to 1. Continue in this way. What pattern emerges?

(b) We can rewrite the pairing function as

$$\frac{(x+y)^2 + 3x + y}{2} = x + \frac{(x+y+1)(x+y)}{2}.$$

Recall that the sum of all numbers from 1 to  $n$  is given by

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

How does this help to explain the pattern in part (a)?

It can be quite hard to find a bijection between two sets of the same cardinality. The **Cantor-Schröder-Bernstein theorem** can be very helpful in this regard.

**Theorem 2.20.** *If there is an injection  $f : X \rightarrow Y$  and an injection  $g : Y \rightarrow X$ , then  $X$  and  $Y$  have the same cardinality.*

You can find a proof in [35]. You can of course try proving it yourself, too.

**Exercise 2.21.** Use the Cantor-Schröder-Bernstein theorem to show that  $\mathbb{R}$  and  $[0, 1]$  have the same cardinality.

Being able to map a set bijectively to another set is another important example of an equivalence relation (see Section 1.2). Let us write

$$A \sim B: \Leftrightarrow \text{there exists a bijection } \pi : A \rightarrow B.$$

**Exercise 2.22.** Show that  $\sim$  is an equivalence relation, that is, it is reflexive, symmetric, and transitive.

We could define the cardinality of a set to be its equivalence class with respect to  $\sim$ . (This would indeed be a proper *class*, not a set.) While this is mathematically sound, it makes thinking about and working with cardinalities rather cumbersome.

One way to overcome this is to pick a canonical representative for each equivalence class and then study the system of representatives.

In the case of cardinalities, what should be our representatives? For finite sets, we use natural numbers. For infinite sets, we can try to use ordinals, as they continue the counting process beyond the finite realm. Counting, in this generalized sense, means establishing a bijection between the set we are counting and an ordinal. If we assume the axiom of choice, the well-ordering principle ensures that every set can be well-ordered, so every set would have a representative. The only problem is that *an infinite set can be well-ordered in more than one way*.

Consider for instance the set of integers,  $\mathbb{Z}$ . We can well-order  $\mathbb{Z}$  as follows:

$$0 < 1 < -1 < 2 < -2 < 3 < \dots$$

This gives a well-ordering of order type  $\omega$ . But we could also proceed like this:

$$1 < -1 < 2 < -2 < 3 < -3 < \dots < 0,$$

that is, we put 0 on top of all other numbers. This gives a well-ordering of order type  $\omega + 1$ . Or we could put all the negative numbers on top of the positive integers:

$$0 < 1 < 2 < 3 < \dots < -1 < -2 < -3 < \dots,$$

which gives a well-ordering of type  $\omega + \omega$ .

This implies, in particular, that  $\omega$ ,  $\omega + 1$ , and  $\omega + \omega$  all have the same cardinality. Recall that we identify ordinals with their initial segment, i.e. we put  $\beta = \{\alpha \in \text{Ord} : \alpha < \beta\}$ . Hence  $\omega + 1 = \{0, 1, 2, \dots, \omega\}$ , and we can map  $\omega + 1$  bijectively to  $\omega$  as follows

$$\omega \mapsto 0, \quad 0 \mapsto 1, \quad 1 \mapsto 2, \quad \dots$$

**Exercise 2.23.** Show that  $\omega^\omega$  has the same cardinality as  $\omega$ .

To obtain a unique representative for each cardinality, we pick the *least ordinal* in each equivalence class. (Here it comes in very handy that the ordinals are well-ordered.)

**Definition 2.24.** An ordinal  $\kappa$  is a **cardinal** if for all ordinals  $\beta < \kappa$ ,  $\beta \neq \kappa$ .

For example,  $\omega + 1$  is not a cardinal, while  $\omega$  is—every ordinal below  $\omega$  is finite and hence not of the same cardinality as  $\omega$ . Thus,  $\omega$  enjoys a special status in that it is the *first infinite cardinal*.

**Exercise 2.25.** Show that every cardinal greater than  $\omega$  is a limit ordinal.

To define the **cardinality** of a set  $S$ , denoted by  $|S|$ , we now simply pick out the one ordinal among all possible order types of  $S$  that is a cardinal

$$\begin{aligned} |S| &= \min\{\alpha : \text{there exists a well-ordering of } S \text{ of order type } \alpha\} \\ &= \text{the unique cardinal } \kappa \text{ such that } S \sim \kappa. \end{aligned}$$

Note that this definition uses the axiom of choice, since we have to ensure that each set has at least one well-ordering.

**Exercise 2.26.** Show that  $|A| \leq |B|$  if and only if there exists a one-to-one mapping  $A \rightarrow B$ .

How many cardinals are there? Infinitely many. This is Cantor's famous theorem.

**Theorem 2.27.** *For every set  $S$ , there exists a set of strictly larger cardinality.*

**Proof.** Consider  $\mathcal{P}(S) = \{X: X \subseteq S\}$ , the *power set* of  $S$ . The mapping  $S \rightarrow \mathcal{P}(S)$  given by  $s \mapsto \{s\}$  is clearly injective, so  $|S| \leq |\mathcal{P}(S)|$ .

We claim that there is no bijection  $f: S \rightarrow \mathcal{P}(S)$ . Suppose there were such a bijection  $f$ , that is, in particular,

$$\mathcal{P}(S) = \{f(x): x \in S\}.$$

Every subset of  $S$  is the image of an element of  $S$  under  $f$ . To get a contradiction, we exhibit a set  $X \subseteq S$  for which this is impossible, namely by letting

$$x \in X \iff x \notin f(x).$$

Now, if there were  $x_0 \in S$  such that  $f(x_0) = X$ , then, by the definition of  $X$ ,

$$x_0 \in X \iff x_0 \notin f(x_0) \iff x_0 \notin X,$$

a contradiction. This is a set-theoretic version of Cantor's *diagonal argument*.  $\square$

The power set operation always yields a set of higher cardinality. But by how much? Since the ordinals are well-ordered, so are the cardinals. We can therefore define, for any cardinal  $\kappa$ ,

$$\kappa^+ = \text{the least cardinal greater than } \kappa.$$

**Cardinal arithmetic.** We now define the basic arithmetic operations on cardinals. Given cardinals  $\kappa$  and  $\lambda$ , let  $A$  and  $B$  be sets such that  $|A| = \kappa$ ,  $|B| = \lambda$ , and  $A \cap B = \emptyset$ . Let

$$(2.2) \quad \kappa + \lambda = |A \cup B|,$$

$$(2.3) \quad \kappa \cdot \lambda = |A \times B|,$$

$$(2.4) \quad \kappa^\lambda = |A^B| = |\{f: f \text{ maps } B \text{ to } A\}|.$$

**Exercise 2.28.** Verify that the definitions above are independent of the choice of  $A$  and  $B$ .

The power operation  $2^\kappa$  is particularly important because it coincides with the cardinality of the power set of  $\kappa$ :

$$2^\kappa = \text{cardinality of } \mathcal{P}(\kappa).$$

In some ways, cardinal arithmetic behaves just like the familiar arithmetic of real numbers.

**Exercise 2.29.** Let  $\kappa, \lambda$ , and  $\mu$  be cardinals. Show that

$$(\kappa^\lambda)^\mu = \kappa^{\lambda \cdot \mu}.$$

But in other regards, cardinal arithmetic is very different.

**Proposition 2.30.** *Let  $\kappa$  and  $\lambda$  be infinite cardinals. Then*

$$\kappa + \lambda = \kappa \cdot \lambda = \max\{\kappa, \lambda\}.$$

**Exercise 2.31.** Prove Proposition 2.30.

Note that for many arguments involving cardinals and cardinal arithmetic, the axiom of choice is needed.

**Alephs and the continuum hypothesis.** Let us denote the cardinality of  $\mathbb{N}$  by  $\aleph_0$ , i.e. any countable, infinite set has cardinality  $\aleph_0$ . Pronounced “aleph”,  $\aleph$  is the first letter of the Hebrew alphabet. The cardinal  $\aleph_0$  is the smallest infinite cardinal. We know that the real numbers  $\mathbb{R}$  are uncountable, i.e.  $|\mathbb{R}| > \aleph_0$ , and it is not hard to show (identifying reals with their binary expansions, which in turn can be interpreted as characteristic functions of subsets of  $\mathbb{N}$ ) that  $|\mathbb{R}| = 2^{\aleph_0}$ . But is the cardinality of the reals actually the smallest uncountable cardinality? That is, is it true that

$$\aleph_0^+ = 2^{\aleph_0}?$$

This is the **continuum hypothesis (CH)**. Like the axiom of choice, the continuum hypothesis is independent over the most common axiom system for set theory, **ZF**. This means that the continuum hypothesis can be neither proved nor disproved in this axiom system. We will say more about independence in Chapter 4.

Since every cardinal has a successor cardinal (just like ordinals), we can use ordinals to index cardinals: We let

$$\aleph_1 = \aleph_0^+,$$

and more generally, for any ordinal  $\alpha$ ,

$$\aleph_{\alpha+1} = \aleph_{\alpha}^{+}.$$

We use  $\omega_{\alpha}$  instead of  $\aleph_{\alpha}$  to denote the order type of the cardinal  $\aleph_{\alpha}$ . If  $\lambda$  is a limit ordinal, we define

$$\aleph_{\lambda} = \sup\{\omega_{\alpha} : \alpha < \lambda\}.$$

**Exercise 2.32.** Show that  $\aleph_{\lambda}$  as defined above is indeed a cardinal. In other words, if  $S$  is a set of cardinals, so is the supremum of  $S$ .

**Exercise 2.33.** Show that every cardinal is an aleph, i.e. if  $\kappa$  is a cardinal, then there exists an ordinal  $\alpha$  such that  $\kappa = \aleph_{\alpha}$ .

### Generalized continuum hypothesis (GCH):

$$\text{For any } \alpha, \aleph_{\alpha+1} = 2^{\aleph_{\alpha}}.$$

If the GCH is true, it means that cardinalities are neatly aligned with the power set operation. The **beth function**  $\beth_{\alpha}$  is defined inductively as

$$\beth_0 = \aleph_0, \quad \beth_{\alpha+1} = 2^{\beth_{\alpha}}, \quad \beth_{\lambda} = \sup\{\beth_{\alpha} : \alpha < \lambda\} \text{ for } \lambda \text{ a limit ordinal.}$$

That is,  $\beth_{\alpha}$  enumerates the cardinalities obtained by iterating the power set operation, starting with  $\mathbb{N}$ . If the GCH holds, then  $\beth_{\alpha} = \aleph_{\alpha}$  for all  $\alpha$ .

## 2.6. Ramsey theorems for uncountable cardinals

Equipped with the notion of a cardinal, we can now attack the question of whether Ramsey's theorem holds for uncountable sets. It also makes sense now to consider colorings with *infinitely many colors*—the corresponding Ramsey statements are not trivially false anymore. We will also look at colorings of *sets of infinite tuples* over a set.

It is helpful to extend the arrow notation for these purposes. Recall that  $N \rightarrow (k)_r^p$  means that every  $r$ -coloring of  $[N]^p$  has a monochromatic subset of size  $k$ . This is really a statement about cardinalities, which we can extend from natural numbers to cardinals.

Let  $\kappa, \mu, \eta$ , and  $\lambda$  be cardinals, where  $\mu, \eta \leq \kappa$ .

$$\kappa \rightarrow (\eta)_{\lambda}^{\mu}$$



means:

If  $|X| \geq \kappa$  and  $c : [X]^\mu \rightarrow \lambda$ , then there exists  $H \subseteq X$  with  $|H| \geq \eta$  such that  $c|_{[H]^\mu}$  is constant.

Here  $[X]^\mu$  is the set of all subsets of  $X$  of cardinality  $\mu$ :

$$[X]^\mu = \{D : D \subseteq X \text{ and } |D| = \mu\}.$$

The following lemma keeps track of the cardinality of  $[X]^\mu$ .

**Lemma 2.34.** *If  $\kappa \geq \mu$  are infinite cardinals and  $|A| = \kappa$ , then*

$$[A]^\mu = \{D : D \subseteq A \text{ and } |D| = \mu\}$$

*has cardinality  $\kappa^\mu$ .*

**Proof.** As  $|A| = \kappa$ , any element of  $\kappa^\mu$  corresponds to a mapping  $f : \mu \rightarrow A$ , which is a subset of  $\mu \times A$ . Moreover, any such  $f$  satisfies  $|f| = \mu$ . Hence  $\kappa^\mu \leq |[\mu \times A]^\mu| = |[A]^\mu|$ , as  $|\mu \times A| = |A|$ .

On the other hand, we can define an injection  $[A]^\mu \rightarrow A^\mu$ : If  $D \subseteq A$  with  $|D| = \mu$ , we can choose a function  $f_D : \mu \rightarrow A$  whose range is  $D$ . Then the mapping  $D \mapsto f_D$  is one-to-one.  $\square$

As  $\aleph_0$  is the smallest infinite cardinal, we can now write the infinite Ramsey theorem, Theorem 2.1, as

$$\aleph_0 \longrightarrow (\aleph_0)_r^p \quad (\text{for any } p, r \in \mathbb{N}).$$

**Finite colorings of uncountable sets.** Does the infinite Ramsey theorem still hold if we pass to uncountable cardinalities?

Let us try to lift the proof from  $\mathbb{N}$  to  $\aleph_1$ . To keep things simple, let us assume we are coloring pairs of real numbers with two colors. In the proof of  $\aleph_0 \longrightarrow (\aleph_0)_2^2$ , one proceeds by constructing a sequence of natural numbers

$$z_0, z_1, z_2, \dots$$

along with a sequence of sets

$$\mathbb{N} = Z_0 \supseteq Z_1 \supseteq Z_2 \supseteq \dots$$

such that

- $z_i \in Z_i$ ;
- for each  $i$ , the color of  $\{z_i, z_j\}$  is the same for all  $j > i$ ; and
- each  $Z_i$  is infinite.

It was possible to find these sequences because of the simple infinite pigeonhole principle: *If we partition an infinite set into finitely many parts, one of them must be infinite.*

This principle still holds for uncountable sets: Any finite partition of an uncountable set must have an uncountable part. In fact, we have something stronger:

*In any partition of an uncountable set into countably many parts, one of the parts must be uncountable.*

Using the language of cardinals, we can state and prove a formal version of this principle.

**Proposition 2.35.** *If  $\kappa$  is an uncountable cardinal and  $f : \kappa \rightarrow \omega$ , then there exists an  $\alpha < \omega$  such that  $|f^{-1}(\{\alpha\})| = \aleph_1$ .*

**Proof.** Assume for a contradiction that  $\kappa_\alpha = |f^{-1}(\{\alpha\})|$  is countable for all  $\alpha < \omega$ . Then

$$\kappa = \bigcup_{\alpha < \omega} \kappa_\alpha$$

would be a countable union of countable sets, which is countable—a contradiction.  $\square$

Looking at the countable infinite case, one might conjecture that an even stronger pigeonhole principle should be true, namely that there is an  $\alpha$  such that  $|f^{-1}(\{\alpha\})| = \kappa$ . This is not quite so; it touches on an aspect of cardinals called *cofinality*. We will learn more about it when we look at large cardinals.

The pigeonhole principle is one instance where uncountable cardinals can behave rather differently from  $\aleph_0$ . We will see that this has consequences for Ramsey's theorem.

We return to Ramsey's theorem and try to prove  $\aleph_1 \longrightarrow (\aleph_1)_2^2$ .

We start with the usual setup. We choose  $z_0 \in \aleph_1$  and look at all  $z \in \aleph_1 \setminus \{z_0\}$  such that  $\{z_0, z\}$  is red. If the set of such  $z$  is uncountable,

then we put  $Z_1 = \{z: c(z_0, z) = \text{red}\}$ . Otherwise, by the uncountable pigeonhole principle,  $\{z: c(z_0, z) = \text{blue}\}$  is uncountable, and we let  $Z_1$  be this set. We can now continue as usual inductively and construct the sequences  $z_0, z_1, z_2, \dots$  and  $Z_1 \supseteq Z_2 \supseteq Z_3 \supseteq \dots$ , where each  $Z_i$  is uncountable. In the countable case, we were almost done, since it took only one more application of the (countable) pigeonhole principle to select an infinite homogeneous subsequence from the  $z_i$ . Now, however, we cannot do this, since we are looking for an *uncountable* homogeneous set. We therefore need to continue our sequence into the transfinite. How can this be done? We need to choose a “next” element of our sequence. We have learned in Section 2.4 that ordinals are made for exactly that purpose. Hence we would index the next element by  $z_\omega$ . But what should the corresponding set  $Z_\omega$  be? We required  $Z_1 \supseteq Z_2 \supseteq Z_3 \supseteq \dots$ , and hence we should have that

$$Z_1 \supseteq Z_2 \supseteq Z_3 \supseteq \dots \supseteq Z_\omega.$$

The only possible choice would therefore be

$$Z_\omega = \bigcap_i Z_i.$$

But this is a problem, because *the intersection of countably many uncountable nested sets is not necessarily uncountable*. Consider for instance the intersection of countably many open intervals

$$\bigcap_n (0, 1/n),$$

which is empty. Indeed, this obstruction is not a coincidence.

**Proposition 2.36.** *Ramsey’s theorem does not hold for the real numbers:*

$$|\mathbb{R}| = 2^{\aleph_0} \not\rightarrow (2^{\aleph_0})_2^2.$$

We will, in fact, show something slightly stronger:

$$2^{\aleph_0} \not\rightarrow (\aleph_1)_2^2.$$

Of course, if the continuum hypothesis holds, this is equivalent to the previous statement.

**Proof.** The proof is based on the fact that a well-ordering of  $\mathbb{R}$  must look *very different* from the usual ordering of the real line.

Using the axiom of choice, let  $<$  be any well-ordering of  $\mathbb{R}$ . Hence we can write

$$\mathbb{R} = \{x_0 < x_1 < \cdots < x_\alpha < x_{\alpha+1} < \cdots\} \quad \text{with all } \alpha < 2^{\aleph_0}.$$

We define a coloring  $c : [\mathbb{R}]^2 \rightarrow \{\text{red}, \text{blue}\}$ . Let  $y \neq z$  be real numbers and denote the usual ordering of  $\mathbb{R}$  by  $<_{\mathbb{R}}$ . Let

$$c(y, z) = \begin{cases} \text{red} & \text{if } (y <_{\mathbb{R}} z \text{ and } y < z) \text{ or } (z <_{\mathbb{R}} y \text{ and } z < y), \\ \text{blue} & \text{otherwise.} \end{cases}$$

Here  $<_{\mathbb{R}}$  denotes the usual order of  $\mathbb{R}$ . In other words, we color a set  $\{y, z\}$  red if the two orderings  $<_{\mathbb{R}}$  and  $<$  agree for  $y$  and  $z$ . If they differ, we color the pair blue.

Assume for a contradiction that  $H$  is a homogeneous subset for  $c$  of size  $\aleph_1$ . We can write  $H$  as

$$H = \{x_{\alpha_0} < x_{\alpha_1} < \cdots < x_{\alpha_\xi} < \cdots\} \quad \text{with } \xi < \aleph_1.$$

If  $c \upharpoonright_{[H]^2} \equiv \text{red}$ , then we have by definition of  $c$  that also

$$x_{\alpha_0} <_{\mathbb{R}} x_{\alpha_1} <_{\mathbb{R}} \cdots <_{\mathbb{R}} x_{\alpha_\xi} <_{\mathbb{R}} \cdots,$$

that is,  $H$  gives us a  $<_{\mathbb{R}}$ -increasing sequence of length  $\aleph_1$ . If  $c \upharpoonright_{[H]^2} \equiv \text{blue}$ , then we get a  $<_{\mathbb{R}}$ -decreasing sequence of length  $\aleph_1$ .

We claim that there cannot be such a sequence.

The rationals are dense in  $\mathbb{R}$  with respect to  $<_{\mathbb{R}}$ , i.e. between any two real numbers is a rational number (not equal to either of them). If there were a strictly  $<_{\mathbb{R}}$ -increasing or strictly  $<_{\mathbb{R}}$ -decreasing sequence of length  $\aleph_1$ , there would also have to be a strictly  $<_{\mathbb{R}}$ -increasing/decreasing sequence of rational numbers of length  $\aleph_1$ , but this is impossible, since the rationals are countable.  $\square$

The essence of the proof lies in the fact that a homogeneous set would “line-up” the well-ordering  $<$  with the standard ordering  $<_{\mathbb{R}}$  of  $\mathbb{R}$ . If this line-up is too long (uncountable), we get a contradiction due to the fact that  $\mathbb{R}$  contains  $\mathbb{Q}$  as a dense “backbone” (under  $<_{\mathbb{R}}$ ).

The proof also links back to the difficulties encountered earlier when trying to lift Ramsey’s theorem to  $2^{\aleph_0}$ .

We are dealing with two orderings here: a well ordering  $<$  of  $\mathbb{R}$  and the familiar linear ordering  $<_{\mathbb{R}}$  of the real line. Let us call the first one the “enumeration” ordering, since it determines the order in which we enumerate  $\mathbb{R}$  and which element we choose next during our attempted construction—the  $<$ -least available. We do not know much about this ordering other than that it is a well-ordering. (In fact, there are some metamathematical issues that prevent us from proving that any explicitly defined function from  $\mathbb{R}$  to an ordinal is a bijection.)

The standard ordering  $<_{\mathbb{R}}$ , on the other hand, is the “color” ordering, since going up or down along it determines whether we color red or blue.

Let us try to follow the construction of a homogeneous set in the proof of Theorem 2.1 and see where it fails for coloring  $c$ . Pick the  $<$ -first element of  $Z_0 = \mathbb{R}$ , say  $x_{\alpha_0}$ . Next check whether the set  $\{y: c(x_{\alpha_0}, y) = \text{red}\}$  is uncountable. This is the case, since there are uncountably many  $y$  “to the right” of  $x_{\alpha_0}$ , and also uncountably many  $y$  not yet enumerated (these appear *after*  $x_{\alpha_0}$  in the well-ordering). Hence we put  $Z_1 = \{y: c(x_{\alpha_0}, y) = \text{red}\}$  and repeat the argument for  $Z_1$ : Pick the  $<$ -least element of  $Z_1$  (which must exist since  $<$  is a well-ordering) and observe that  $\{y \in Z_1: c(x_{\alpha_0}, y) = \text{red}\}$  is again uncountable. Inductively, we construct an increasing sequence

$$x_{\alpha_0} <_{\mathbb{R}} x_{\alpha_1} <_{\mathbb{R}} x_{\alpha_2} <_{\mathbb{R}} \cdots$$

and a nested sequence of sets

$$Z_0 \supset Z_1 \supset Z_2 \supset \cdots$$

such that

$$(x_{\alpha_n}, \infty) \supseteq Z_{n+1}.$$

But if  $x_{\alpha_n} \rightarrow \infty$  (which might well be the case), this implies

$$\bigcap_n Z_n = \emptyset,$$

and hence after  $\omega$ -many steps we cannot continue our construction.

We could try to select the  $\alpha_n$  a little more carefully; in particular, we could, for instance, let  $x_{\alpha_{n+1}}$  be the  $<$ -least element of  $Z_{n+1}$  such that  $x_{\alpha_n} <_{\mathbb{R}} x_{\alpha_{n+1}} < x_{\alpha_n} + 1/2^n$ . This way we would guarantee that we

could continue our construction beyond stage  $\omega$  and into the transfinite. In fact, by choosing the  $x_\alpha$  carefully enough, we can ensure that the construction goes on for  $\beta$ -many stages for *any fixed countable ordinal*  $\beta$ . But the cardinality argument of the proof above tells us it is impossible to do this for  $\aleph_1$ -many stages.

**Exercise 2.37.** Show that Proposition 2.36 generalizes to

$$2^\kappa \not\rightarrow (\kappa^+)_2^2.$$

(*Hint:* Show that  $\{0, 1\}^\kappa$  has no increasing or decreasing sequence of length  $\kappa^+$ .)

An obvious question now arises: If we allow higher cardinalities  $\kappa$  beyond  $2^{\aleph_0}$ , does  $\kappa \rightarrow (\aleph_1)_2^2$  become true eventually? The *Erdős-Rado theorem* shows that we in fact only have to pass to the next higher cardinality.

**Theorem 2.38** (Erdős-Rado theorem).

$$(2^{\aleph_0})^+ \longrightarrow (\aleph_1)_2^2.$$

Compared to the counterexample in Proposition 2.36, the extra cardinal gives us some space to

*set aside an uncountable set such that whenever we extend our current homogeneous set, we leave this set untouched, i.e. we do not add elements from it.*

In this way, we can now guarantee that the sets  $Z_\alpha$  will have a non-empty, in fact uncountable, intersection.

This “setting aside” happens by virtue of the following lemma.

**Lemma 2.39.** *There exists a set  $R \subset (2^{\aleph_0})^+$  of cardinality  $|R| = 2^{\aleph_0}$  such that for every countable  $D \subseteq R$  and for every  $x \in (2^{\aleph_0})^+ \setminus D$ , there exists an  $r \in R \setminus D$  such that for all  $d \in D$ ,*

$$(2.5) \quad c(x, d) = c(r, d).$$

Informally, whenever we choose an  $x$  and a countable  $D \subset R$ , we can find a “replacement” for  $x$  in  $R$  that behaves in a color-identical manner with respect to  $D$ . This will enable us, in our construction

of a homogeneous set of size  $\aleph_1$ , to choose the  $x_\alpha$  from a set of size  $2^{\aleph_0}$ , leaving a “reservoir” of uncountable cardinality.

**Proof.** We construct the set  $R$  by extending it step by step, adding the witnesses required by (2.5).

We start by putting  $R_0 = 2^{\aleph_0}$ . We have to ensure that (2.5) holds for every countable subset  $D \subset R_0$  and every  $x \in (2^{\aleph_0})^+ \setminus D$ . To simplify notation, let us put

$$c_x(y) = c(x, y).$$

Hence every  $x$  fixes a function  $c_x : (2^{\aleph_0})^+ \setminus \{x\} \rightarrow \{0, 1\}$ . We are interested in the functions  $c_x \upharpoonright_D$  for countable  $D \subset R_0$ . Each such function maps a countable subset of  $R_0$  to  $\{0, 1\}$ .

We count the number of such functions. If we fix a countable  $D \subset R_0$ , there are at most  $2^{\aleph_0}$ -many ways to map  $D$  to  $\{0, 1\}$ . By Lemma 2.34, there are

$$(2^{\aleph_0})^{\aleph_0} = 2^{\aleph_0 \cdot \aleph_0} = 2^{\max\{\aleph_0, \aleph_0\}} = 2^{\aleph_0}$$

countable subsets of  $2^{\aleph_0}$ . Therefore, there are at most

$$2^{\aleph_0} \cdot 2^{\aleph_0} = 2^{\aleph_0}$$

possible functions  $c_x \upharpoonright_D$ . (Note that while each  $x \in (2^{\aleph_0})^+$  gives rise to such a function, many of them will actually be identical, by the pigeonhole principle.)

Therefore, we need to add at most  $2^{\aleph_0}$ -many witnesses to  $R_0$ , one  $r \in (2^{\aleph_0})^+$  for each function  $c_x \upharpoonright_D$  (of which there are at most  $2^{\aleph_0}$ -many). This gives us  $R_1$ , and  $R_1$  in turn gives rise to new countable subsets  $D$  which we have to witness by possibly adding new elements from  $(2^{\aleph_0})^+$  to  $R_1$ . But the crucial fact here is that the cardinality of  $R_1$  is still  $2^{\aleph_0}$ , since  $2^{\aleph_0} + 2^{\aleph_0} = 2^{\aleph_0}$ , and therefore we can resort to the same argument as before, adding at most  $2^{\aleph_0}$ -many witnesses, resulting in a set  $R_2$  of cardinality  $2^{\aleph_0}$ .

We have to run our construction into the transfinite. Let  $\alpha$  be a countable ordinal, and assume that we have defined sets

$$R_0 \subseteq R_1 \subseteq \dots \subseteq R_\beta \subseteq \dots$$

for all  $\beta < \alpha$ ,  $|R_\beta| = 2^{\aleph_0}$ . If  $\alpha$  is a successor ordinal,  $\alpha = \beta + 1$ , we define  $R_\alpha$  by the argument given above, adding at most  $2^{\aleph_0}$  new witnesses. If  $\alpha$  is a limit ordinal, we put

$$R_\alpha = \bigcup_{\beta < \alpha} R_\beta.$$

This is a countable union of sets of cardinality  $2^{\aleph_0}$ , and hence is also of cardinality  $2^{\aleph_0}$  (see Proposition 2.30).

Finally, put

$$R = \bigcup_{\alpha < \omega_1} R_\alpha.$$

We claim that this  $R$  has the desired property. First note that the cardinality of  $R$  is

$$\aleph_1 \cdot 2^{\aleph_0} = \max\{\aleph_1, 2^{\aleph_0}\} = 2^{\aleph_0}.$$

Let  $D \subset R$  be countable and let  $x \in (2^{\aleph_0})^+ \setminus D$ . The crucial observation is that

all elements of  $D$  must have been added by some stage  $\alpha < \omega_1$ ,  
i.e. there exists an  $\alpha < \omega_1$  such that  $D \subseteq R_\alpha$ .

If this were not the case, every stage would add a new element of  $D$ , which would mean  $D$  has at least  $\omega_1$ -many elements, in contradiction to  $D$  being countable. But then the necessary witness for  $c_x \upharpoonright_D$  is present in  $R_{\alpha+1}$ ; that is, there exists an  $r \in R_{\alpha+1} \subseteq R$  such that  $c_r \upharpoonright_D = c_x \upharpoonright_D$ .

This completes the proof of the lemma. □

We can now use the lemma to modify our construction of a homogeneous set  $H$ .

**Proof of the Erdős-Rado theorem.** Let  $x^*$  be an arbitrary element of  $(2^{\aleph_0})^+ \setminus R$ . This will be our “anchor point”. Choose  $x_0 \in R$  arbitrary.

Suppose now, given  $\alpha < \omega_1$ , we have chosen  $x_\beta$  for all  $\beta < \alpha$ .

Let

$$D = \{x_\beta : \beta < \alpha\}.$$

This is a countable set (since  $\alpha < \omega_1$ ). By Lemma 2.39, there exists an  $r \in R$  such that  $c_{x^*} \upharpoonright_D = c_r \upharpoonright_D$ . Put  $x_\alpha = r$ .



By the pigeonhole principle, there exists  $i \in \{0, 1\}$  such that

$$H = \{x_\alpha : \alpha < \omega_1, c(x^*, x_\alpha) = i\}$$

is uncountable. We claim that  $c \upharpoonright_{[H]^2} \equiv i$ , i.e.  $H$  is homogeneous for  $c$ . For suppose  $x_\zeta, x_\xi \in H$  and  $\zeta < \xi$ . Then, by definition of  $x_\xi$ ,

$$c(x_\xi, x_\zeta) = c_\xi(x_\zeta) = c_{x^*}(x_\zeta) = c(x^*, x_\alpha) = i.$$

□

Note that the proof becomes quite elegant once we have proved the lemma. After we ensured the existence of the set  $R$ , we can work in some sense “backwards”: We choose a single “anchor point”  $x^*$  from the reserved set  $(2^{\aleph_0})^+ \setminus R$ . You can think of  $x^*$  as always being the next point chosen in the sense of the standard construction, only then to be replaced by an element from  $R$  which behaves exactly like it in terms of color pairings with the already constructed  $x_\beta$ .

Note also that we do not need to construct a sequence of shrinking sets  $Z_\alpha$  anymore. In the standard construction, the  $Z_\alpha$  represent the reservoir from which the next potential elements of the homogeneous set are chosen. They are no longer needed since  $x^*$  is always available (as explained above).

We also do not need to keep track of the color choices we made along the way, as  $x^*$  does this job for us, too. For example, if  $c(x^*, x_0) = 0$ , it follows by construction that  $c(x_\beta, x_0) = 0$  for all  $\beta > 0$ , which in the previous constructions means that we restrict the  $Z$  to all elements which color 0 with  $x_0$ .

**Exercise 2.40.** Generalize the proof of the Erdős-Rado theorem (and Lemma 2.39) to show that

$$\beth_n^+ \longrightarrow (\aleph_1)_{\aleph_0}^{n+1}.$$

**Infinite colorings.** The Erdős-Rado theorem holds for countable colorings, too (see Exercise 2.40). What else can we say about infinite colorings? Clearly, the number of colors should be smaller than the set we are trying to color. For example,

$$\aleph_0 \not\rightarrow (2)_{\aleph_0}^1.$$

But even if we make the colored set larger than the number of colors, this does not mean we can find even a finite homogeneous set.

**Proposition 2.41.** *For any infinite cardinal  $\kappa$ ,*

$$2^\kappa \not\rightarrow (3)_\kappa^2.$$

**Proof.** We define  $c : [2^\kappa]^2 \rightarrow \kappa$  as follows. An element of  $2^\kappa$  corresponds to a  $\{0, 1\}$ -sequence  $(x_\beta : \beta < \kappa)$  of length  $\kappa$  (recall that  $2^\kappa$  is the cardinality of the power set of  $\kappa$ , and every element of the power set can be coded by its characteristic sequence;  $(x_\beta : \beta < \kappa)$  is such a characteristic sequence). Given two such sequences  $(x_\beta) \neq (y_\beta)$ , we let

$$c((x_\beta), (y_\beta)) = \text{the least } \alpha < \kappa \text{ such that } x_\alpha \neq y_\alpha.$$

Now assume that  $(x_\beta), (y_\beta), (z_\beta)$  are pairwise distinct. Let  $c((x_\beta), (y_\beta)) = \alpha$ . Without loss of generality,  $x_\alpha = 0$  and  $y_\alpha = 1$ . Now  $z_\alpha \in \{0, 1\}$ , so either  $z_\alpha = x_\alpha$  or  $z_\alpha = y_\alpha$ . In the first case  $c((x_\beta), (z_\beta)) \neq \alpha$  and in the second case  $c((y_\beta), (z_\beta)) \neq \alpha$ . Therefore, there cannot exist a  $c$ -homogeneous subset of size 3.  $\square$

## 2.7. Large cardinals and Ramsey cardinals

The results of the previous section make the set of natural numbers stand out among the infinite sets not only because  $\aleph_0$  is the first infinite cardinal but also for another reason: With respect to finite colorings of finite tuples, the natural numbers *admit a homogeneous subset of the same size*, or, in the Ramsey arrow notation,

$$\aleph_0 \rightarrow (\aleph_0)_r^p$$

for any positive integers  $p$  and  $r$ .

In the previous section, we saw that this is no longer true for  $\aleph_1$  (Proposition 2.36). In fact, for *any* infinite cardinal  $\kappa$ ,

$$2^\kappa \not\rightarrow (\kappa^+)_2^2,$$

so in particular

$$2^\kappa \not\rightarrow (2^\kappa)_2^2$$

for any infinite cardinal  $\kappa$ .

This in turn means that *any* cardinal  $\lambda$  that can be written as  $\lambda = 2^\kappa$  cannot satisfy  $\lambda \rightarrow (\lambda)_2^2$ . But are there any cardinals that *cannot*

be written this way?  $\aleph_0$ , the cardinality of  $\mathbb{N}$ , is such a cardinal—the power set of a finite set is still finite. But other than  $\mathbb{N}$ ?

**Definition 2.42.** A cardinal  $\lambda$  is a **limit cardinal** if  $\lambda = \aleph_\gamma$  for some limit ordinal  $\gamma$ .  $\lambda$  is a **strong limit cardinal** if for all cardinals  $\kappa < \lambda$ ,  $2^\kappa < \lambda$ .

Any strong limit cardinal is a limit cardinal. For if  $\lambda$  is a successor cardinal, then  $\lambda = \aleph_{\alpha+1} = \aleph_\alpha^+ \leq 2^{\aleph_\alpha}$ . Is being *strong limit* truly stronger than being limit? Well, it depends. If the GCH holds, then  $2^{\aleph_\alpha} = \aleph_\alpha^+ = \aleph_{\alpha+1}$  for all  $\alpha$ , and therefore every limit cardinal is actually a strong limit cardinal.

For now just let us assume that  $\kappa$  is a strong limit cardinal. Is this sufficient for

$$\kappa \longrightarrow (\kappa)_2^2?$$

Take for example  $\aleph_\omega$ . This is clearly a limit cardinal and, if the GCH holds, also a strong limit cardinal. Does it hold that

$$\aleph_\omega \longrightarrow (\aleph_\omega)_2^2?$$

The problem is that we can “reach”  $\aleph_\omega$  rather “quickly” from below, since

$$\aleph_\omega = \bigcup_{n < \omega} \aleph_n.$$

We can use this fact to devise a coloring of  $\aleph_\omega$  that cannot have a homogeneous subset of size  $\aleph_\omega$ . Namely, let us put, for each  $n < \omega$ ,

$$X_{n+1} = \aleph_{n+1} \setminus \aleph_n.$$

Then  $\aleph_\omega$  is the disjoint union of the  $X_n$ , and the cardinality of each  $X_n$  is strictly less than  $\aleph_\omega$ . Now define a coloring  $c : [\aleph_\omega]^2 \rightarrow \{0, 1\}$  by

$$c(x, y) = \begin{cases} 1 & \text{if } x \text{ and } y \text{ are in different } X_n, \\ 0 & \text{if } x \text{ and } y \text{ are in the same } X_n. \end{cases}$$

Let  $H \subset \aleph_\omega$  be a homogeneous subset for  $c$ . If  $c \upharpoonright [H]^2 \equiv 1$ , then no two elements of  $H$  can be in the same  $X_n$ , but there are only  $\aleph_0$ -many  $X_n$ , and hence  $H$  is countable. If  $c \upharpoonright [H]^2 \equiv 0$ , then all elements of  $H$  have to be in the same  $X_n$ , but as noted above,  $|X_n| < \aleph_\omega$  for each  $n$ .

The proof works in general for any cardinal  $\kappa$  that we can reach in fewer than  $\kappa$  steps. This brings us to the notion of *cofinality*.

**Definition 2.43.** The **cofinality** of a limit ordinal  $\alpha$ ,  $\text{cf}(\alpha)$ , is the least ordinal  $\beta$  such that there exists an increasing sequence  $(\alpha_\gamma)_{\gamma < \beta}$  of length  $\beta$  such that  $\alpha_\gamma < \alpha$  for all  $\alpha$  and

$$\alpha = \lim_{\gamma \rightarrow \beta} \alpha_\gamma = \sup\{\xi_\gamma : \gamma < \beta\}.$$

Obviously, we always have  $\text{cf}(\alpha) \leq \alpha$ . Here are some examples as an exercise.

**Exercise 2.44.** Prove the following cofinalities:

- (i)  $\text{cf}(\omega) = \omega$ ,
- (ii)  $\text{cf}(\omega + \omega) = \omega$ ,
- (iii)  $\text{cf}(\alpha) = \omega$  for every countable, infinite  $\alpha$ ,
- (iv)  $\text{cf}(\omega_1) = \omega_1$  (if we assume AC),
- (v)  $\text{cf}(\omega_\omega) = \omega$ .

The last statement can be generalized to

$$\text{cf}(\omega_\lambda) = \text{cf}(\lambda),$$

where  $\lambda$  is any limit ordinal.

If  $\kappa$  is an infinite cardinal and  $\text{cf}(\kappa) < \kappa$ , it means that  $\kappa$  can be reached from below by means of a “ladder” that has fewer steps than  $\kappa$ . Such a cardinal is called **singular**. If  $\text{cf}(\kappa) = \kappa$ ,  $\kappa$  is called **regular**. Hence  $\aleph_0$  and  $\aleph_1$  are regular cardinals, while  $\aleph_\omega$  is singular. Assuming the axiom of choice, one can show that every successor cardinal, that is, a cardinal of the form  $\aleph_{\alpha+1}$ , is a regular cardinal.

It seems much harder for a limit cardinal to be regular. For this to be true, the following must hold:

$$\aleph_\lambda = \text{cf}(\aleph_\lambda) = \text{cf}(\lambda) \leq \lambda.$$

But since clearly  $\aleph_\lambda \geq \lambda$ , this means that

$$\text{if } \aleph_\lambda \text{ is regular } (\lambda \text{ limit}), \text{ then } \aleph_\lambda = \lambda.$$

This seems rather strange. Going, for example, from  $\aleph_0$  to  $\aleph_1$ , we traverse  $\omega_1$ -many ordinals, but the jump “costs” only one step in terms of cardinals. That means we have to go a *long, long way* if we ever want to “catch up” with the alephs again.

Anyway, let us capture the notions of large cardinals we have found so far in a formal definition.

**Definition 2.45.** Let  $\kappa$  be an uncountable cardinal.

- (i)  $\kappa$  is **weakly inaccessible** if it is regular and a limit cardinal.
- (ii)  $\kappa$  is **(strongly) inaccessible** if it is regular and a strong limit cardinal.
- (iii)  $\kappa$  is **Ramsey** if  $\kappa \rightarrow (\kappa)_2^2$ .

Usually, strongly inaccessible cardinals are called simply *inaccessible*. Our investigation leading up to Definition 2.45 now yields the following.

**Theorem 2.46.** *Every Ramsey cardinal is inaccessible.*

Do Ramsey cardinals exist? We cannot answer this question here, nor will we in this book. In fact, in a certain sense we cannot answer this question *at all*. More precisely, the *existence of Ramsey cardinals cannot be proved in ZF*. As mentioned before in Section 2.5, ZF is a common axiomatic framework for set theory in which most of contemporary mathematics can be formalized. We will say more about axiom systems and formal proofs in Chapter 4.