Topology

Neil Strickland

Contents

1. Introduction

To do:

- Appendix on complex analysis.
- Puppe sequences
- Quasifibrations
- ENRs
- Exercise on operator norms?
- Exercise on continuity of the spectral radius?
- Functional calculus as an application of Stone-Weierstrass?
- Extended example: the complement of the Mandelbrot set is connected.

2. Basic Concepts

Definition 2.1. [`defn-topology`]
A *topology* on a set X is a set τ of subsets of X (called *τ-open sets*, or just *open sets*) such that:

T0: The empty set and the whole set X are both open.
T1: If we have a family $(U_i)_{i\in I}$ of open sets, then the union $U=\bigcup_{i\in I} U_i$ is also open.
T2: If we have two open sets U_0 and U_1, then $U_0\cap U_1$ is also open.

A *topological space* is a set X equipped with a specified topology.

Example 2.2. [`eg-R-topology`]
Consider the set $X=\mathbb{R}$. We say that a subset $U\subseteq X$ is *open* if for all $x\in U$ there exists $\epsilon>0$ such that $(x-\epsilon,x+\epsilon)\subseteq U$, and we let τ be the family of all subsets satisfying this condition. For example, the set (a,b) is open for any $a<b$, because if $x\in(a,b)$ then we can put $\epsilon=\min(x-a,b-x)>0$ and we find that $(x-\epsilon,x+\epsilon)\subseteq(a,b)$. However, the set $[a,b]$ is not open, because the defining condition is violated when $x=a$ or $x=b$. The empty set is open (because there is nothing to check) and $\mathbb{R}$ itself is open (because we can always take $\epsilon=1$). Now suppose we have a family of open sets U_i, and we put $U=\bigcup_i U_i$. Suppose that $x\in U$, so $x\in U_i$ for some i. As U_i is open we can find $\epsilon>0$ with $(x-\epsilon,x+\epsilon)\subseteq U_i\subseteq U$. It follows that U is open. Next, suppose we have open sets U_0 and U_1 and a point $x\in U_0\cap U_1$. As $x\in U_0$ and U_0 is open, there exists $\epsilon_0>0$ such that $(x-\epsilon_0,x+\epsilon_0)\subseteq U_0$. Similarly, there exists $\epsilon_1>0$ such that $(x-\epsilon_1,x+\epsilon_1)\subseteq U_1$. It follows that if we put $\epsilon=\min(\epsilon_0,\epsilon_1)$ then $(x-\epsilon,x+\epsilon)\subseteq U_0\cap U_1$. This means that $U_0\cap U_1$ is again open, so all the axioms are satisfied, and the collection τ is a topology on $\mathbb{R}$. We call it the *standard topology* on $\mathbb{R}$.

Example 2.3. [`eg-Rn-topology`]
Let X be any subset of $\mathbb{R}^n$ (possibly $\mathbb{R}^n$ itself). For any $x\in\mathbb{R}^n$ we write $\|x\|=\left(\sum_i x_i^2\right)^{1/2}$, and

$$\begin{aligned} OB_\epsilon(x) &= \{y\in\mathbb{R}^n \;:\; \|y-x\|<\epsilon\} \\ B_\epsilon(x) &= \{y\in\mathbb{R}^n \;:\; \|y-x\|\leq\epsilon\}. \end{aligned}$$

We say that a subset $U\subseteq X$ is *open* if for all $x\in U$ there exists $\epsilon>0$ such that $OB_\epsilon(x)\cap X\subseteq U$. This defines a topology on X, which we again call the *standard topology*. The proof that it is a topology is the same as for $X=\mathbb{R}$, but with the set $OB_\epsilon(x)\cap X$ replacing the interval $(x-\epsilon,x+\epsilon)$. Note that $\mathbb{C}^n$ can be identified with $\mathbb{R}^{2n}$, so this procedure gives a topology on any subset of $\mathbb{C}^n$ as well.

Example 2.4. [`eg-Sn`]
One example that we will use repeatedly is the n-sphere:

$$S^n=\{(x_0,\dots,x_n)\in\mathbb{R}^{n+1} \;:\; \sum_i x_i^2=1\}.$$

This is a subset of $\mathbb{R}^n$, so it has a topology as described in Example 2.3. We will identify S^1 with $\{z\in\mathbb{C} \;:\; |z|=1\}$ by the correspondence $(x_0,x_1)\mapsto x_0+ix_1$.

Example 2.5. [`eg-mandelbrot`]

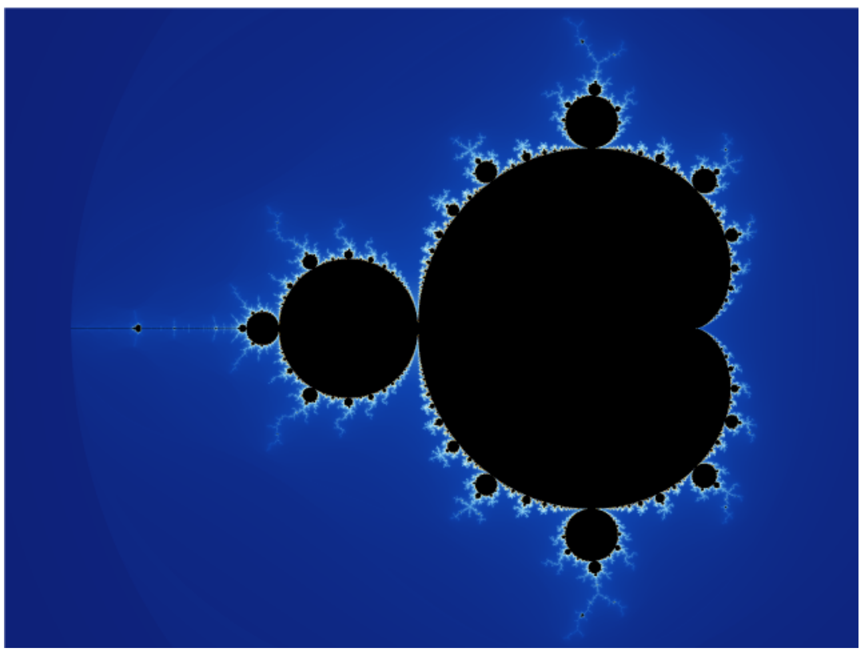
We now briefly introduce a space $M \subset \mathbb{C}$ that we will later revisit several times as an example of various different phenomena in topology. It is called the *Mandelbrot set*. For any $c \in \mathbb{C}$, we can define a function $q_c \colon \mathbb{C} \to \mathbb{C}$ by $q_c(z) = z^2 + c$. Using this we define a sequence of values

$$\begin{aligned}
&f_0(c) = 0 \\
&f_1(c) = q_c(f_0(c)) = q_c(0) = c \\
&f_2(c) = q_c(f_1(c)) = q_c(c) = c^2 + c \\
&f_3(c) = q_c(f_2(c)) = c^4 + 2c^3 + c^2 + c \\
&f_4(c) = q_c(f_3(c)) = c^8 + 4c^7 + 6c^6 + 6c^{+}5c^4 + 2c^3 + c^2 + c.
\end{aligned}$$

and so on. The explicit formulae rapidly become unmanageable, but we will not need them. We remark, however, that $f_n(c)$ is a polynomial of degree 2^{n-1} in c. We put

$$M = \{c \in \mathbb{C} \; : \; |f_n(c)| \leq 2 \text{ for all } n\}.$$

This is a subset of $\mathbb{C} \simeq \mathbb{R}^2$ so it has a topology as described in Example 2.3. The structure of M is extremely intricate. The black region in the picture below is an initial approximation, but much finer structure is revealed if you blow up a small region on the boundary of the set.



Example 2.6. [eg-binary-seq]

A *binary sequence* is a sequence $x = (x_0, x_1, \dots)$ with $x_i \in \{0, 1\}$ for all i. Let X be the set of all binary sequences. For $x \in X$ and $n \in \mathbb{N}$ put

$$C_n(x) = \{y \; : \; y_i = x_i \text{ for all } i < n\}.$$

Say that $U \subseteq X$ is *open* if for each $x \in U$ there exists $n \in \mathbb{N}$ such that $C_n(x) \subseteq U$. For example, consider the sets

$$\begin{aligned}
U &= \{x \; : \; x_0 = x_2 = x_4 = x_6 = 0\} \\
V &= \{x \; : \; x_{2i} = 0 \text{ for all } i\} \\
W &= \{x \; : \; x_i = 1 \text{ for at least five indices } i\}.
\end{aligned}$$

If $x \in U$ then $C_7(x) \subseteq U$, so U is open. However, V is not open. To see this, let e_m be the sequence given by $(e_m)_m = 1$ and $(e_m)_i = 0$ for $i \neq m$. The zero sequence is in V, but for any n we have $e_{2n} \in C_n(0) \setminus V$, so

$C_n(0) \not\subseteq V$. On the other hand, if $x \in W$ then we can find $i_1 < i_2 < \dots < i_5$ such that $x_{i_1} = \dots = x_{i_5} = 1$, and then $C_{i_5+1}(x) \subseteq W$; so W is open.

We next claim that the above definition gives a topology on X. Indeed, the empty set is open (because there is nothing to check) and the whole space is open (because we can take $n = 0$). Now suppose we have a family of open sets U_i, and we put $U = \bigcup_i U_i$. Suppose that $x \in U$, so $x \in U_i$ for some i. As U_i is open we can find n with $C_n(x) \subseteq U_i \subseteq U$. It follows that U is open. Next, suppose we have open sets U_0 and U_1 and a point $x \in U_0 \cap U_1$. As $x \in U_0$ and U_0 is open, there exists n_0 such that $C_{n_0}(x) \subseteq U_0$. Similarly, there exists n_1 such that $C_{n_1}(x) \subseteq U_1$. It follows that if we put $n = \max(n_0, n_1)$ then $C_n(x) = C_{n_0}(x) \cap C_{n_1}(x) \subseteq U_0 \cap U_1$. This means that $U_0 \cap U_1$ is again open, so all the axioms are satisfied.

Example 2.7. [`eg-sierpinski`]

Take $X = \{0, 1\}$. We declare that the sets $\emptyset$, $\{1\}$ and X are open, but that $\{0\}$ is not. This gives a topology on X called the *Sierpinski topology*; the set X equipped with this topology is called the *Sierpinski space*.

Remark 2.8. [`rem-finite-topologies`]

There is a close link between the theory of topologies on finite sets and the algebraic theory of partially ordered sets, and this has some applications in theoretical computer science. However, we will not emphasise such examples in these notes.

Maybe we should have a section on frames and locales. If so, edit this remark.

Example 2.9. [`eg-discrete`]

Let X be any set. We can define a topology on X by declaring that *all* subsets are open. This is called the *discrete topology*. We can define a different topology by declaring that the only open sets are $\emptyset$ and X itself. This is called the *indiscrete topology*.

Remark 2.10. [`rem-de-morgan`]

Arguments in topology very often involve manipulations with intersections and unions of infinite families of subsets. For these it is important to recall the following identities, known as *De Morgan's laws*:

(a) If A_0 and A_1 are subsets of X, then $(A_0 \cup A_1)^c = A_0^c \cap A_1^c$.
(b) More generally, if we have any family of subsets $A_i \subseteq X$ (for $i \in I$, say) then $\left(\bigcup_{i \in I} A_i\right)^c = \bigcap_{i \in I} A_i^c$.
(c) If A_0 and A_1 are subsets of X, then $(A_0 \cap A_1)^c = A_0^c \cup A_1^c$.
(d) More generally, if we have any family of subsets $A_i \subseteq X$ (for $i \in I$, say) then $\left(\bigcap_{i \in I} A_i\right)^c = \bigcup_{i \in I} A_i^c$.

All these are just logical reformulations of the relevant definitions. For example, consider (d). We have

$$\begin{aligned} x \in \textstyle\bigcap_{i \in I} A_i &\Leftrightarrow \text{ For all } i \in I \text{ we have } x \in A_i \\ x \in \left(\textstyle\bigcap_{i \in I} A_i\right)^c &\Leftrightarrow x \notin \textstyle\bigcap_{i \in I} A_i \\ &\Leftrightarrow \text{ For some } i \in I \text{ we have } x \notin A_i \\ &\Leftrightarrow \text{ For some } i \in I \text{ we have } x \in A_i^c \\ &\Leftrightarrow x \in \textstyle\bigcup_{i \in I} A_i^c. \end{aligned}$$

Definition 2.11. [`defn-top-omni`]

Let X be a topological space.

(a) An *open neighbourhood* of a point $x \in X$ is an open set $U \subseteq X$ such that $x \in U$. More generally, a *neighbourhood* of x is a set Y that contains an open neighbourhood of x.
(b) An *interior point* of a set $Y \subseteq X$ is a point $x \in Y$ such that Y is a neighbourhood of x. Equivalently, a point $x \in Y$ counts as an interior point if there is an open set U such that $x \in U \subseteq Y$.
(c) We write $\overset{\circ}{Y}$ or $\operatorname{int}(Y)$ for the set of interior points of Y, otherwise known as the *interior* of Y. Note that this depends on the ambient space X, so we will sometimes write $\operatorname{int}_X(Y)$ to avoid ambiguity.
(d) A *closure point* of a set $Y \subseteq X$ is a point $x \in X$ such that every open neighbourhood of x meets Y (or equivalently, every neighbourhood of x meets Y).
(e) We write $\overline{Y}$ or $\operatorname{cl}(Y)$ or $\operatorname{cl}_X(Y)$ for the set of closure points of Y, otherwise known as the *closure* of Y.
(f) A set $Y \subseteq X$ is *closed* if and only if its complement $Y^c = X \setminus Y$ is open.

(g) The *boundary* of a set $Y \subseteq X$ is the set $\mathrm{bdy}_X(Y) = \mathrm{cl}_X(Y) \cap \mathrm{cl}_X(Y^c)$.

REMARK 2.12. [`rem-clopen`]
One point that may cause confusion is that it is possible for a set to be both open and closed, and it is also possible for a set to be neither open nor closed. More specifically:

(a) In the space $\mathbb{R}$, the set $[0,1) = \{x : 0 \leq x < 1\}$ is neither open nor closed. Similarly, $\mathbb{Q}$ is neither open nor closed in $\mathbb{R}$.
(b) In any topological space X, the sets $\emptyset$ and X are both open and closed.
(c) If X is any set with the discrete topology, then every subset of X is both open and closed.
(d) Let X be the space of binary sequences as in Example 2.6, and put $Y = \{x : x_0 = 0\}$ and $Z = \{x : x_0 = 1\}$. Then Y and Z are both open, so $X \setminus Y$ and $X \setminus Z$ are closed. However $X \setminus Y = Z$ and $X \setminus Z = Y$, so Y and Z are closed as well as open.

PROPOSITION 2.13. *Let X be a topological space, and let Y be a subset of X.*

(a) *The set* $\mathrm{int}(Y)$ *is open; it is the union of the collection of all open sets U such that $U \subseteq Y$.*
(b) *The set Y itself is open if and only if* $\mathrm{int}(Y) = Y$.
(c) *The set* $\mathrm{cl}(Y)$ *is closed; it is the intersection of the collection of all closed sets F such that $Y \subseteq F$.*
(d) *The set Y itself is closed if and only if* $\mathrm{cl}(Y) = Y$.

PROOF. Just by translating the definition, we see that $\mathrm{int}(Y)$ is the union of the collection of all open sets U such that $U \subseteq Y$. The union of any family of open sets is again open, so $\mathrm{int}(Y)$ is open in X. In particular, if $\mathrm{int}(Y) = Y$ then Y is open. Conversely, if Y is open then it is a neighbourhood of each of its points, so every element of Y is an interior point of y, so $Y = \mathrm{int}(Y)$.

Next, note that x is not a closure point of Y if and only if there is some open neighbourhood N of x such that $N \cap Y = \emptyset$, or equivalently $N \subseteq Y^c$. This means that $\mathrm{cl}(Y)^c = \mathrm{int}(Y^c)$, or equivalently $\mathrm{cl}(Y) = \mathrm{int}(Y^c)^c$. We can thus prove (c) and (d) by applying (a) and (b) to Y^c. □

REMARK 2.14. [`rem-closed-axioms`]
Because the closed sets are just the complements of the open sets and *vice versa*, we can specify a topology completely by describing the closed sets instead of describing the open sets. The closed sets must have the following properties:

Z0: The empty set and the whole set X are both closed.
Z1: If we have a family $(F_i)_{i \in I}$ of closed sets, then the intersection $F = \bigcap_{i \in I} F_i$ is also closed.
Z2: If we have two closed sets F_0 and F_1, then $F_0 \cup F_1$ is also closed.

These are equivalent to the axioms in Definition 2.1, because $\left(\bigcup_i U_i\right)^c = \bigcap_i U_i^c$ and $(U_0 \cap U_1)^c = U_0^c \cup U_1^c$.

REMARK 2.15. [`rem-bdy-closed`]
We now see that $\mathrm{cl}_X(Y)$ and $\mathrm{cl}_X(Y^c)$ are closed, so the boundary $\mathrm{bdy}(Y) = \mathrm{cl}_X(Y) \cap \mathrm{cl}_X(Y^c)$ is also closed.

EXAMPLE 2.16. [`eg-cofinite`]
Let X be any set, and declare that a subset $F \subseteq X$ is closed if and only if it is either finite or all of X. It is easy to see that the above axioms are satisfied, so this defines a new topology on X, called the *cofinite topology*. If X itself is finite then the cofinite topology is the same as the discrete topology, but not in general.

EXAMPLE 2.17. [`eg-three-points`]
Now consider the set $X = \{0,1,2\}$. One can check that this admits 29 different topologies. To describe these, note that the sets of size 0 or 3 are automatically open, and that a set of size 2 is open if and only if the complementary set of size one is closed. We will abuse language slightly and say that a point $x \in X$ is open (or closed) if the singleton set $\{x\}$ is open (or closed). Thus, we can describe a topology on X by listing the open points and the closed points. If x and y are two distinct points and the remaining point is z, then $\{z\} = (\{x\} \cup \{y\})^c$. It follows that if x and y are both open then z is closed, and if x and y are both closed then z is open. By a systematic check of cases, we obtain the following list of possibilities:

(a) The indiscrete topology has no open or closed points.

(b) For the discrete topology, every point is both open and closed.
(c) For each point $x \in X$, there is a topology where x is the only open point and there are no closed points.
(d) For each point $x \in X$, there is a topology where x is the only closed point and there are no open points.
(e) For each point $x \in X$, there is a topology where x is open but not closed, and the other two points are closed but not open.
(f) For each point $x \in X$, there is a topology where x is closed but not open, and the other two points are open but not closed.
(g) For each point $x \in X$, there is a topology where x is both open and closed, and the other two points are neither open nor closed.
(h) For each pair of points $x \neq y$, there is a topology where x is the only open point, and y is the only closed point.
(i) For each pair of points $x \neq y$, there is a topology where x is the only open point, and y is the only point that is not closed.

Note that in cases (c) to (g) we have three choices for x, and in cases (h) and (i) we have six choices for the pair (x, y); this gives $2 + 5 \times 3 + 2 \times 6 = 29$ topologies altogether.

Example 2.18. [`eg-zariski`]
Readers who are familiar with commutative algebra can consider the following example. Let $P = \mathbb{R}[x_1, \dots, x_n]$ denote the ring of all polynomial functions on $\mathbb{R}^n$. Given an ideal $I \leq P$, we put

$$V(I) = \{x \in \mathbb{R}^n \ : \ f(x) = 0 \text{ for all } f \in I\}.$$

One can check that

$$\begin{aligned} V(P) &= \emptyset \\ V(0) &= \mathbb{R}^n \\ V(\sum_i I_i) &= \bigcap_i V(I_i) \\ V(I_0 I_1) &= V(I_0 \cap I_1) = V(I_0) \cup V(I_1). \end{aligned}$$

It follows that the sets of the form $V(I)$ are the closed sets for a new topology on $\mathbb{R}^n$, which is called the *Zariski topology*. Nothing here really depends on the fact that we work over $\mathbb{R}$; there is a Zariski topology on K^n for any field K. This is of central importance in algebraic geometry. In the case $n = 1$, the Zariski topology is the same as the cofinite topology.

Definition 2.19. [`defn-dense`]
A subset $Y \subseteq X$ is *dense* if $\overline{Y} = X$.

This can be reformulated as follows:

Lemma 2.20. [`lem-dense`]
Y is dense in X if and only if Y meets every nonempty open subset of X.

Proof. First suppose that Y is dense. Let U be a nonempty open subset of X. As $U \neq \emptyset$ we can choose a point $x \in U$, so that U is an open neighbourhood of x. As Y is dense we have $\overline{Y} = X$, so $x \in \overline{Y}$, which means that every neighbourhood of x meets Y. In particular U meets Y, as required.

Suppose instead that Y meets every nonempty open set. Consider a point $x \in X$. Then any open neighbourhood of x is a nonempty open set, so it meets Y. This means that x is a closure point of Y, so $x \in \overline{Y}$. As x was arbitrary this means that $\overline{Y} = X$ as required. □

Example 2.21. [`eg-dense`]
The set $\mathbb{Q}$ is dense in $\mathbb{R}$. Indeed, any nonempty open subset $U \subseteq \mathbb{R}$ contains an interval (a, b) for some $a < b$. We can choose an integer $n > 0$ such that $1/n < b - a$, and then let k be the largest integer such that $k/n \leq a$. We then find that $(k + 1)/n \in \mathbb{Q} \cap U$, so $\mathbb{Q}$ meets U as required. We also see that for sufficiently large m we have $(k + 1)/n + \sqrt{2}/m \in U$, and using this that $\mathbb{R} \setminus \mathbb{Q}$ is also dense.

We next explore a convenient way of describing and analysing topologies.

Definition 2.22. [defn-basis]
A *topological basis* on X is a collection β of subsets of X such that:

B0: For every $x \in X$ there exists some $U \in \beta$ such that $x \in U$. (Equivalently, the union of all the sets in β is X.)

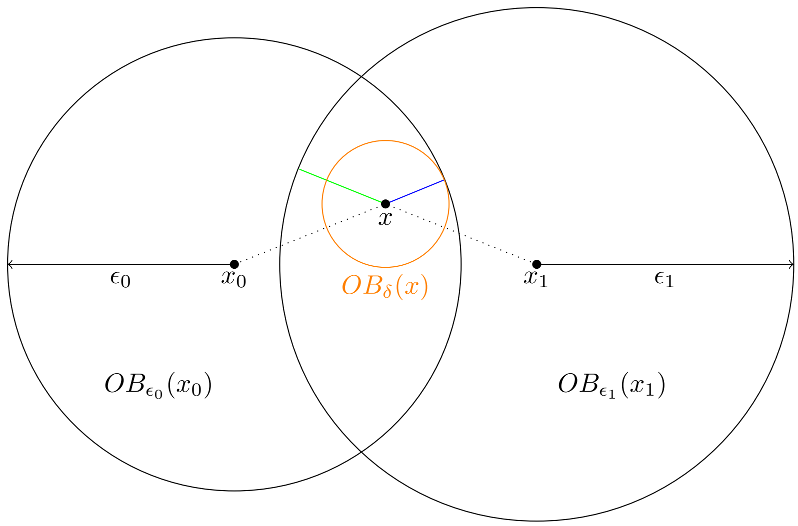
B1: If U_0 and U_1 are in β and $x \in U_0 \cap U_1$ then there exists $V \in \beta$ with $x \in V \subseteq U_0 \cap U_1$.

If β is such a collection, we write $\tau(\beta)$ for the larger collection of all sets $V \subseteq X$ with the following property: for all $x \in V$ there exists $U \in \beta$ with $x \in U \subseteq V$.

Example 2.23. [eg-basis]

(a) If we let β be the collection of all intervals $(a, b) \subseteq \mathbb{R}$, then β is a topological basis on $\mathbb{R}$ and $\tau(\beta)$ is just the standard topology on $\mathbb{R}$.

(b) Now instead let X be a subset of $\mathbb{R}^n$, and let β be the collection of all sets of the form $OB_\epsilon(x) \cap X$ for $x \in X$ and $\epsilon > 0$, as in Example 2.3. It is clear that axiom B0 is satisfied. For B1, suppose we have $x \in OB_{\epsilon_0}(x_0) \cap OB_{\epsilon_1}(x_1)$. Put $\delta_i = \epsilon_i - \|x_i - x\| > 0$ and $\delta = \min(\delta_0, \delta_1)$. Using the triangle inequality $\|x_i - y\| \leq \|x_i - x\| + \|x - y\|$ we see that $x \in OB_\delta(x) \subseteq OB_{\epsilon_0}(x_0) \cap OB_{\epsilon_1}(x_1)$, which proves B1. We again see from the definitions that $\tau(\beta)$ is the standard topology on X.



(c) In Example 2.6, the set $\beta = \{C_n(x) \; : \; x \in X, \; n \in \mathbb{N}\}$ is a topological basis, and $\tau(\beta)$ is just the topology considered previously.

Part (b) above used the triangle inequality for $\mathbb{R}^n$. For completeness, we record a proof.

Lemma 2.24. [lem-triangle]
For vectors $u, v \in \mathbb{R}^n$ we have $|\langle u, v\rangle| \leq \|u\|\|v\|$ (the Cauchy-Schwartz inequality*) and $\|u+v\| \leq \|u\|+\|v\|$ (the* Triangle Inequality*).*

Proof. The most direct way to prove the first inequality is to check that

$$\langle u, v\rangle^2 + \sum_{1\leq i<j\leq n} (u_i v_j - u_j v_i)^2 = \|u\|^2\|v\|^2.$$

Indeed, we have

$$\langle u, v\rangle^2 = (\sum_i u_i v_i)^2 = \sum_i u_i^2 v_i^2 + 2\sum_{i<j} u_i u_j v_i v_j.$$

On the other hand, we have

$$\sum_{i<j} (u_i v_j - u_j v_i)^2 = \sum_{i<j} u_i^2 v_j^2 + \sum_{i<j} u_j^2 v_i^2 - 2\sum_{i<j} u_i u_j v_i v_j.$$

The first two sums here can be combined and reindexed as $\sum_{i\neq j} u_i^2 v_j^2$, and the last sum cancels with the last term in our previous equation, leaving

$$\langle u,v\rangle^2 + \sum_{i<j}(u_iv_j - u_jv_i)^2 = \sum_i u_i^2 v_i^2 + \sum_{i\neq j} u_i^2 v_j^2 = \sum_{i,j} u_i^2 v_j^2 = \|u\|^2\|v\|^2$$

as claimed. This means that $\langle u,v\rangle^2 \leq \|u\|^2\|v\|^2$, and by taking square roots we obtain $\langle u,v\rangle \leq |\langle u,v\rangle| \leq \|u\|\|v\|$. This in turn gives

$$\begin{aligned}\|u+v\|^2 = \langle u+v, u+v\rangle = \langle u,u\rangle + \langle v,v\rangle + 2\langle u,v\rangle &= \|u\|^2 + \|v\|^2 + 2\langle u,v\rangle \\ \leq \|u\|^2 + \|v\|^2 + 2\|u\|\|v\| = (\|u\| + \|v\|)^2,\end{aligned}$$

so $\|u+v\| \leq \|u\| + \|v\|$. □

The following result should now come as no surprise.

Proposition 2.25. [prop-basis]
Let β be a topological basis on a a set X. Then:
- (a) *$\tau(\beta)$ is a topology on X.*
- (b) *$\beta \subseteq \tau(\beta)$.*
- (c) *A set U lies in $\tau(\beta)$ if and only if it is the union of some family of elements of β.*
- (d) *If τ' is any topology on X such that $\beta \subseteq \tau'$, then $\tau(\beta) \subseteq \tau'$.*

Proof.
- (a) The empty set is open (because there is nothing to check). We next claim that X is open. Indeed, if $x \in X$ then axiom B0 gives us a set $U \in \beta$ with $x \in U \subseteq X$, as required. Now suppose we have a family of open sets U_i, and we put $U = \bigcup_i U_i$. Suppose that $x \in U$, so $x \in U_i$ for some i. As U_i is open we can find $V \in \beta$ with $x \in V \subseteq U_i \subseteq U$. It follows that U is open. Next, suppose we have open sets U_0 and U_1 and a point $x \in U_0 \cap U_1$. As $x \in U_0$ and U_0 is open, there exists $V_0 \in \beta$ such that $x \in V_0 \subseteq U_0$. Similarly, there exists $V_1 \in \beta$ such that $x \in V_1 \subseteq U_1$. Now axiom B1 gives us a set $V \in \beta$ with $x \in V \subseteq V_0 \cap V_1 \subseteq U_0 \cap U_1$. This means that $U_0 \cap U_1$ is again open, so all the axioms are satisfied.
- (b) Suppose that $U \in \beta$. For each $x \in U$ we must find $V \in \beta$ with $x \in V \subseteq U$; but we can just take $V = U$.
- (c) This is just a translation of the definition.
- (d) If τ' is a topology containing β then it contains the union of any family of elements of β, so it contains all of $\tau(\beta)$ by part (c).

□

Example 2.26. [eg-padic-basis]
Fix a prime number p. Take $X = \mathbb{Z}$ and

$$\beta = \{n + p^i\mathbb{Z} \;:\; n \in \mathbb{Z},\ i \in \mathbb{N}\}.$$

As $0 + p^0\mathbb{Z} = \mathbb{Z}$, we see that B0 is satisfied. Next, note that $x \in n + p^i\mathbb{Z}$ if and only if $x = n \pmod{p^i}$, in which case $n + p^i\mathbb{Z} = x + p^i\mathbb{Z}$. Thus, if $x \in (n_0 + p^{i_0}\mathbb{Z}) \cap (n_1 + p^{i_1}\mathbb{Z})$ we can put $i = \max(i_0, i_1)$ and we find that $x + p^i\mathbb{Z} \in \beta$ and $(x + p^i\mathbb{Z}) \subseteq (n_0 + p^{i_0}\mathbb{Z}) \cap (n_1 + p^{i_1}\mathbb{Z})$. This means that B1 is also satisfied, so we have a topological basis. The corresponding topology $\tau(\beta)$ is called the *p-adic topology* on $\mathbb{Z}$; it is important in algebraic number theory.

Definition 2.27. [defn-basis-for]
Suppose we are given a topology θ on a set X, and also a topological basis β on the same set. We say that β *is a basis for* θ if $\theta = \tau(\beta)$. In this context we refer to the elements of β as *basic open sets*.

Proposition 2.28. [prop-basis-for]
Let X be a topological space, and let β be a collection of open subsets. Then the following are equivalent:
- (a) *β is a basis for the topology*
- (b) *β contains a basis for the topology*

(c) *For every open set U and every point $x \in U$ there exists $V \in \beta$ such that $x \in V \subseteq U$.*

PROOF. We will write θ for the originally given topology, so $\beta \subseteq \theta$. It is clear that (a) implies (b). Now suppose that β contains a subset β' that is a basis for θ. Let U be a set that is open with respect to $\theta = \tau(\beta')$. By the definition of $\tau(\beta')$, we see that for every $x \in U$ there exists $V \in \beta' \subseteq \beta$ such that $x \in V \subseteq U$. This shows that (b) implies (c). Now suppose that (c) holds. By taking $U = X$ we see that the sets in β cover X, so axiom B0 is satisfied. Similarly, given sets $U_0, U_1 \in \beta \subseteq \theta$ we can take $U = U_0 \cap U_1$ to see that axiom B1 is satisfied. This shows that β is a topological basis. This means that we have a topology $\tau(\beta)$, and after examining the definition we see that (c) says that $\theta \subseteq \tau(\beta)$. On the other hand, we have $\beta \subseteq \theta$ by hypothesis, and every set in $\tau(\beta)$ is the union of some family of sets in β, so $\tau(\beta) \subseteq \theta$. We thus have $\theta = \tau(\beta)$, so (a) holds. □

It is sometimes convenient to go one step further, as follows.

DEFINITION 2.29. [`defn-subbasis`]
A *topological subbasis* on X is just a collection σ of subsets of X. Given such a collection, we put

$$\beta(\sigma) = \{U_1 \cap \ldots \cap U_n \;:\; U_1, \ldots, U_n \in \sigma\}.$$

We allow the case $n = 0$, in which case the intersection is interpreted as X, so $X \in \beta(\sigma)$. Given this, it is easy to see that $\beta(\sigma)$ is a topological basis on X. We also abbreviate $\tau(\beta(\sigma))$ as $\tau(\sigma)$; this is the smallest topology on X that contains σ. If we are given a topology θ and it works out that $\tau(\sigma) = \theta$, we say that σ is a *subbasis for θ*.

EXAMPLE 2.30. [`eg-subbasis`]

(a) Put
$$\sigma = \{(-\infty, a) \;:\; a \in \mathbb{R}\} \cup \{(a, \infty) \;:\; a \in \mathbb{R}\}.$$
Note that for $x \in \mathbb{R}$ and $\epsilon > 0$ we have
$$(x - \epsilon, x + \epsilon) = (-\infty, x + \epsilon) \cap (x - \epsilon, \infty) \in \beta(\sigma).$$
Using this, we see that σ is a subbasis for the standard topology on $\mathbb{R}$.

(b) Let X be the space of binary sequences as in Example 2.6, and let τ be the topology discussed there. For $n \in \mathbb{N}$ and $b \in \{0, 1\}$ put $U_{nb} = \{x \in X \;:\; x_n = b\}$. If $x \in U_{nb}$ then $C_{n+1}(x) \subseteq U_{nb}$, so we see that U_{nb} is open. Now put
$$\sigma = \{U_{nb} \;:\; n \in \mathbb{N},\ b \in \{0, 1\}\} \subseteq \tau.$$
Recall that the sets $C_m(y)$ form a basis for τ, and note that
$$C_m(y) = U_{0,y_0} \cap \cdots \cap U_{m,y_m} \in \beta(\sigma).$$
Using this, we see that σ is a subbasis for τ.

(c) Let X be any set, and let τ be the cofinite topology on X. Put $\sigma = \{\{x\}^c \;:\; x \in X\}$. If U is a nonempty open set then $X \setminus U$ is finite, say $X \setminus U = \{x_1, \ldots, x_n\}$. We then have
$$U = \{x_1\}^c \cap \cdots \cap \{x_n\}^c \in \beta(\sigma).$$
Using this, we see that σ is a subbasis for τ.

We can now use the terminology of subbases to introduce a topology on an arbitrary real vector space V. This is not very useful unless V has finite dimension, and in that case we can identify V with $\mathbb{R}^n$ for some n, so Example 2.3 gives us a topology on V. The real point here, however, is to describe the topology in a way that is independent of the choice of isomorphism $V \simeq \mathbb{R}^n$.

DEFINITION 2.31. [`defn-linear-topology`]
Let V be any vector space over $\mathbb{R}$. For any linear map $\phi \colon V \to \mathbb{R}$ and any $a, b \in \mathbb{R}$ we put

$$U(\phi, a, b) = \{x \in V \;:\; a < \phi(x) < b\}.$$

The collection of all sets of this form is a subbasis for a topology on V, which we call the *linear topology*.

Proposition 2.32. [prop-linear-topology]
The linear topology on $\mathbb{R}^n$ is the same as the standard topology described in Example 2.3.

Proof. Let τ_S be the standard topology, which has basis

$$\beta_S = \{OB_\epsilon(x) \;:\; x \in \mathbb{R}^n,\ \epsilon > 0\}.$$

Let σ_L be the family of all sets of the form $U(\phi, a, b)$, which is a subbasis for the linear topology τ_L.

Let V be a set in τ_S, and consider a point $x \in V$. As $V \in \tau_S$, we have $OB_\epsilon(x) \subseteq V$ for some $\epsilon > 0$. We can define linear maps $\pi_i \colon \mathbb{R}^n \to \mathbb{R}$ by $\pi_i(u) = u_i$, and using these we can define a set

$$N = \bigcap_{i=1}^{n} U(\pi_i, x_i - \epsilon/\sqrt{n}, x_i + \epsilon/\sqrt{n}) \in \tau_L.$$

Note that $y \in N$ iff $|x_i - y_i| < \epsilon/\sqrt{n}$ for all n, and if so, we have

$$\|x - y\| = \sqrt{\textstyle\sum_{i=1}^n |x_i - y_i|^2} < \sqrt{\textstyle\sum_{i=1}^n \epsilon^2/n} = \epsilon,$$

so $y \in OB_\epsilon(x) \subseteq V$. We thus have $N \subseteq V$, so x is in the τ_L-interior of V. As x was arbitrary, we have $V \in \tau_L$. This means that $\tau_S \subseteq \tau_L$.

In the other direction, suppose we have a set $W = U(\phi, a, b) \in \sigma_L$. Here ϕ is a linear map $\mathbb{R}^n \to \mathbb{R}$, so it have the form $\phi(x) = \langle x, u\rangle$ for some $u \in \mathbb{R}^n$. If $u = 0$ then W is either $\mathbb{R}^n$ (if $a < 0 < b$) or $\emptyset$ (otherwise), so $W \in \tau_S$. We may therefore assume that $u \neq 0$. Suppose we have a point $x \in W$, so $\langle x, u\rangle \in (a, b)$. Put

$$\delta = \min(\langle x, u\rangle - a,\ b - \langle x, u\rangle),$$

and $\epsilon = \delta/\|u\|$. If $\|y - x\| < \epsilon$ then the Cauchy-Schwartz inequality gives $|\phi(y) - \phi(x)| = |\langle y - x, u\rangle| \leq \delta$ and so $\phi(y) \in (a, b)$. We thus have $OB_\epsilon(x) \subseteq W$, so x is in the τ_S-interior of W. As x was arbitrary, we have $V \in \tau_S$. This means that $\sigma_L \subseteq \tau_S$, so $\tau_L \subseteq \tau_S$. □

If we have a suitable notion of distance between points of a set X, we can use it to define a topology by a straightforward generalisation of Example 2.3. We next explain this in more detail.

Definition 2.33. [defn-metric]
A *metric* on a space X is a function $d \colon X \times X \to [0, \infty]$ satisfying the following axioms:

M0: For all $x \in X$ we have $d(x, x) = 0$.
M1: For all $x, y \in X$ we have $d(x, y) = d(y, x)$.
M2: For all $x, y, z \in X$ we have $d(x, z) \leq d(x, y) + d(y, z)$ (the *triangle inequality*).
M3: For all $x, y \in X$, if $d(x, y) = 0$ then $x = y$.

In axiom M2 we use the convention $\infty + x = \infty = x + \infty$ for all $x \in [0, \infty]$ if necessary. A *semimetric* is a function that satisfies M0 to M2, but not necessarily M3. A *(semi)metric space* is a set equipped with a specified (semi)metric.

Remark 2.34. [rem-semimetric-quotient]
Suppose that X is a semimetric space. Let E be the relation on X given by xEy if and only if $d(x, y) = 0$. It follows easily from the semimetric axioms that this is an equivalence relation. Let $\overline{X}$ be the quotient set. If x_0Ex_1 and y_0Ey_1 it follows from the triangle inequality that $d(x_0, y_0) = d(x_1, y_1)$. We therefore have an induced map $\overline{d} \colon \overline{X} \times \overline{X} \to [0, \infty)$, which is easily seen to give a metric on $\overline{X}$. Most questions about the semimetric space X reduce easily to questions about the metric space $\overline{X}$. However, it is occasionally useful to be able to work in the more general semimetric context.

Definition 2.35. [defn-metric-topology]
If d is a (semi)metric on X then we put

$$\begin{aligned} OB_\epsilon(x) &= \{y \in X \;:\; d(x, y) < \epsilon\} \\ B_\epsilon(x) &= \{y \in X \;:\; d(x, y) \leq \epsilon\} \\ \beta_d &= \{OB_\epsilon(x) \;:\; x \in X, \epsilon > 0\}. \end{aligned}$$

The set $OB_\epsilon(x)$ is called the *open ball of radius ϵ around x*, and $B_\epsilon(x)$ is called the *closed ball.* The set β_d is a basis for a topology τ_d on X, called the *(semi)metric topology.*

Remark 2.36. [`rem-metric-topology`]
The proof that β_d is a topological basis is essentially the same as Example 2.23(b). Axiom $B0$ is satisfied because $x \in OB_1(x) \in \beta_d$ for all $x \in X$. For B1, suppose we have $x \in OB_{\epsilon_0}(x_0) \cap OB_{\epsilon_1}(x_1)$. Put $\delta_i = \epsilon_i - d(x_i, x) > 0$ and $\delta = \min(\delta_0, \delta_1)$. Using axiom M2 for d we see that

$$d(x_i, y) \leq d(x_i, x) + d(x, y) = \epsilon_i - \delta_i + d(x, y).$$

If $d(x, y) < \delta$ this gives $d(x_i, y) < \epsilon_i$, so we have $x \in OB_\delta(x) \subseteq OB_{\epsilon_0}(x_0) \cap OB_{\epsilon_1}(x_1)$, which proves B1.

Remark 2.37. Open balls are open by the definition of the semimetric topology. We claim that closed balls are closed (as one would expect from the terminology). To see this, suppose that $y \in B_\epsilon(x)^c$. This means that the number $\delta = d(x, y) - \epsilon$ is strictly positive. If there were a point $z \in B_\epsilon(x) \cap OB_\delta(y)$ we would have

$$d(x, y) \leq d(x, z) + d(z, y) < \epsilon + \delta = d(x, y),$$

which is impossible. It follows that $B_\epsilon(x) \cap OB_\delta(y) = \emptyset$, so $OB_\delta(y) \subseteq B_\epsilon(x)^c$, so y is in the interior of $B_\epsilon(x)^c$. As y was arbitrary this means that $B_\epsilon(x)^c$ is open, so $B_\epsilon(x)$ is closed, as claimed. It is also clear that when $\epsilon < \delta$ we have

$$OB_\epsilon(x) \subseteq B_\epsilon(x) \subseteq OB_\delta(x) \subseteq B_\delta(x).$$

It is often the case that $OB_\epsilon(x)$ is the interior of $B_\epsilon(x)$, and that $B_\epsilon(x)$ is the closure of $OB_\epsilon(x)$. However, both of these statements can fail, for example when $X = \mathbb{Z}$ (with the metric $d(n, m) = |n - m|$) and $x = 0$ and $\epsilon = 1$.

Example 2.38. [`eg-Rn-metric`]
We can now define a metric on $\mathbb{R}^n$ by

$$d(x, y) = \|x - y\| = \sqrt{\sum_i (x_i - y_i)^2},$$

and the associated topology is the standard one. Here axioms M0, M1 and M3 are clear, and M2 follows from Lemma 2.24.

Example 2.39. [`eg-lanes`]
We can define a different metric on $\mathbb{R}^2$ (called the *lane metric*) by

$$d((x, y), (x', y')) = \begin{cases} |y - y'| & \text{if } x = x' \\ |y| + |x - x'| + |y'| & \text{if } x \neq x'. \end{cases}$$

The heuristic picture here is that there is a main road along the x-axis, and vertical lanes covering the whole plane. If $x = x'$ then you can travel a distance $|y - y'|$ along a vertical lane from (x, y) to (x', y'); otherwise, you need to travel a distance $|y|$ from (x, y) to the main road, then $|x - x'|$ along the main road, then $|y'|$ along another lane to reach (x', y').

Axioms M0, M1 and M3 are clear for this metric. For M2, suppose we have points $a = (x, y)$ and $a' = (x', y')$ and $a'' = (x'', y'')$. Put $u = d(a, a') + d(a', a'') - d(a, a'')$, so we must show that $u \geq 0$.

(a) If $x = x' = x''$ then $u = |x'' - x| - |x'' - x'| - |x' - x|$, which is nonnegative by the triangle inequality.

(b) Suppose instead that $x = x' \neq x''$. Note that

$$|y| = |(y - y') + y'| \leq |y - y'| + |y'|.$$

We have

$$\begin{aligned} u &= |y - y'| + (|y'| + |x' - x''| + |y''|) - (|y| + |x - x''| + |y''|) \\ &= |y - y'| + |y'| - |y| \geq 0. \end{aligned}$$

(c) The case where $x \neq x' = x''$ is essentially the same as (b).

(d) Suppose that $x \neq x' \neq x''$ but $x'' = x$. We then have

$$\begin{aligned} u &= (|y| + |x - x'| + |y'|) + (|y'| + |x' - x''| + |y''|) - |y - y''| \\ &\geq |y| + |y''| - |y - y''| \geq 0. \end{aligned}$$

(e) Finally, suppose that x, x' and x'' are all different. We then have

$$\begin{aligned} u &= (|y| + |x - x'| + |y'|) + (|y'| + |x' - x''| + |y''|) - (|y| + |x - x''| + |y''|) \\ &= 2|y'| + (|x - x'| + |x' - x''| - |x - x''|) \geq 0. \end{aligned}$$

Example 2.40. [eg-matrix-metric]

Let $M_n(\mathbb{R})$ denote the set of $n \times n$ matrices over $\mathbb{R}$. We can identify this with $\mathbb{R}^{n^2}$, which gives a metric and thus a topology. The metric can be related nicely to the algebra of matrices, as we now describe. Recall that the *trace* of a matrix $A \in M_n(\mathbb{R})$ is defined by $\text{trace}(A) = \sum_{i=1}^n A_{ii}$, and the *transpose* is the matrix A^T with entries $(A^T)_{ij} = A_{ji}$. One checks that $\text{trace}(A^T A) = \sum_{i=1}^n \sum_{j=1}^n A_{ij}^2$, and thus that

$$d(A, B) = \sqrt{\text{trace}((A - B)^T (A - B))}.$$

We can restrict this definition to give a metric and thus a topology on any subset of $M_n(\mathbb{R})$. Some interesting examples include the following:

$$\begin{aligned} O(n) &= \{A \in M_n(\mathbb{R}) \;:\; A^T A = 1\} \\ \mathfrak{o}(n) &= \{A \in M_n(\mathbb{R}) \;:\; A^T + A = 0\} \\ SL_n(\mathbb{R}) &= \{A \in M_n(\mathbb{R}) \;:\; \det(A) = 1\} \\ SO(n) &= O_n \cap SL_n(\mathbb{R}) \\ GL_n^+(\mathbb{R}) &= \{A \in M_n(\mathbb{R}) \;:\; \det(A) > 0\} \\ GL_n(\mathbb{R}) &= \{A \in M_n(\mathbb{R}) \;:\; \det(A) \neq 0\} \\ \mathbb{R}P^n &= \{A \in M_{n+1}(\mathbb{R}) \;:\; A^T = A = A^2, \text{ trace}(A) = 1\}. \end{aligned}$$

Various topological properties of these spaces will be described in Example 5.21. Note that $\mathbb{R}P^n$ is often defined in a rather different way, as a quotient of the sphere S^n. This will be reconciled with our definition in Examples 5.24 and 5.69.

We will also consider complex analogues of these spaces. In this context we need to use the hermitian transpose $A^\dagger$, given by $(A^\dagger)_{ij} = \overline{A_{ji}}$. We put

$$\begin{aligned} U(n) &= \{A \in M_n(\mathbb{C}) \;:\; A^\dagger A = 1\} \\ \mathfrak{u}(n) &= \{A \in M_n(\mathbb{C}) \;:\; A^\dagger + A = 0\} \\ SL_n(\mathbb{C}) &= \{A \in M_n(\mathbb{C}) \;:\; \det(A) = 1\} \\ SU(n) &= U_n \cap SL_n(\mathbb{C}) \\ GL_n(\mathbb{C}) &= \{A \in M_n(\mathbb{C}) \;:\; \det(A) \neq 0\} \\ \mathbb{C}P^n &= \{A \in M_{n+1}(\mathbb{C}) \;:\; A^\dagger = A = A^2, \text{ trace}(A) = 1\}. \end{aligned}$$

Example 2.41. [eg-trivial-metrics]

Let X be any set, and define $d(x, y) = 0$ when $x = y$, and $d(x, y) = 1$ when $x \neq y$. This defines a metric on X, for which the corresponding topology is the discrete topology (i.e. every subset of X is open). Alternatively, we can define $d'(x, y) = 0$ for all x and y. This is a semimetric on X, for which the associated topology is the indiscrete topology (only $\emptyset$ and X are open).

Example 2.42. [eg-binary-metric]

Let X be the set of binary sequences, as in Example 2.6. For $x, y \in X$ we define $d(x, y)$ as follows. If $x = y$ we put $d(x, y) = 0$. Otherwise, we let i be the least integer such that $x_i \neq y_i$, and put $d(x, y) = 2^{-i}$. It is clear that $d(x, y) = d(y, x)$ and that $d(x, y) \geq 0$, with equality if and only if $x = y$. Now suppose we have a third sequence z. If $x_n = y_n$ for $n < i$ and $y_n = z_n$ for $n < j$, it is clear that $x_n = z_n$ for $n < \min(i, j)$. Using this we see that

$$d(x, z) \leq \max(d(x, y), d(y, z)) \leq d(x, y) + d(y, z).$$

It follows that d is a metric on X. Moreover, if $2^{-n} < \epsilon \leq 2^{1-n}$ then $OB_\epsilon(x) = C_n(x)$. It follows from this that the metric topology is the same as the topology considered earlier.

EXAMPLE 2.43. [eg-padic-metric]

Fix a prime number p. For $n,m \in \mathbb{Z}$ we define $d(n,m)$ as follows. If $n = m$ we put $d(n,m) = 0$. Otherwise there is a largest integer v such that $n-m$ is divisible by p^v, and we put $d(n,m) = p^{-v}$. It is clear that $d(n,m) = d(m,n)$ and that $d(m,n) \geq 0$, with equality if and only if $m = n$. Now suppose we have a third integer k. If $n-m$ is divisible by p^v and $m-k$ is divisible by p^w, we see that the integer $n-k = (n-m)+(m-k)$ is divisible by $p^{\min(v,w)}$. Using this we see that

$$d(n,k) \leq \max(d(n,m), d(m,k)) \leq d(n,m) + d(m,k).$$

It follows that d is a metric on $\mathbb{Z}$. Moreover, if $p^{-v} < \epsilon \leq p^{1-v}$ then $OB_\epsilon(n) = n + p^v\mathbb{Z}$. It follows from this that the metric topology is the same as the p-adic topology considered earlier described in Example 2.26.

Although metrics can be very convenient, they are less canonical than the associated topologies. In particular, it often happens that there are many different metrics that define the same topology. We next investigate this phenomenon.

PROPOSITION 2.44. [prop-truncated-metric]

Let d be a semimetric on a set X, let c be a positive constant, and put $d'(x,y) = \min(d(x,y), c)$. Then d' is another metric on X, which defines the same topology.

Note that if we have a metric that sometimes takes the value ∞, then we can use this proposition to replace it by one that is always finite. The intuition is as follows. We can think of $d(x,y)$ as the time it takes to walk from x to y. Suppose we have a teleportation device that takes a time c to warm up, and we walk or teleport depending only on which is faster; then $d'(x,y)$ is the travel time from x to y.

PROOF. It is clear that d' satisfies M0 and M1, and it satisfies M3 if and only if d does. The only issue is the triangle inequality. Consider three points $x,y,z \in X$. If $d(x,y) \geq c$ or $d(y,z) \geq c$ then $d'(x,y) = c$ or $d'(y,z) = c$ so $d'(x,y) + d'(y,z) \geq c$, but visibly $d'(x,z) \leq c$ so the triangle inequality holds. This leaves only the case where $d(x,y) < c$ and $d(y,z) < c$, so $d'(x,y) = d(x,y)$ and $d'(y,z) = d(y,z)$. Here the triangle inequality for d gives $d(x,z) \leq d'(x,y) + d'(y,z)$ and from the definitions $d'(x,z) \leq d(x,z)$, so the triangle inequality for d' is again valid.

We have now shown that d' is a semimetric, but we still need to understand the topology $\tau_{d'}$ that it defines. Put $OB_\epsilon(x) = \{y \,:\, d(x,y) < \epsilon\}$ and $OB'_\epsilon(x) = \{y \,:\, d'(x,y) < \epsilon\}$. If $\epsilon \leq c$ then $OB'_\epsilon(x) = OB_\epsilon(x)$, and if $\epsilon \geq c$ then $OB'_\epsilon(x) = X$. It follows that $OB'_\epsilon(x)$ is always open with respect to τ_d. Conversely, if $y \in OB_\epsilon(x)$ and $\delta = \min(c, \epsilon - d(x,y))$ then the set $OB'_\delta(y) = OB_\delta(y)$ is contained in $OB_\epsilon(x)$. This shows that $OB_\epsilon(x)$ is open in $\tau_{d'}$, and from this we conclude that $\tau_d = \tau_{d'}$ as claimed. $\square$

DEFINITION 2.45. [defn-strong-equiv]

Let d and d' be two semimetrics on the same set X. We say that d and d' are *weakly equivalent* if the corresponding topologies τ_d and $\tau_{d'}$ are the same. We say that they are *strongly equivalent* if there are constants $A, A' > 0$ such that $d(x,y) \leq Ad'(x,y)$ and $d'(x,y) \leq A'd(x,y)$ for all $x,y \in X$.

LEMMA 2.46. [lem-weak-equiv]

Let d and d' be two semimetrics on the same set X, and write OB and OB' for open balls defined using d and d' respectively. Then d and d' are weakly equivalent iff

(a) *For each $x \in X$ and $\epsilon > 0$ there exists $\delta > 0$ such that $OB_\delta(x) \subseteq OB'_\epsilon(x)$; and*
(b) *For each $x \in X$ and $\epsilon > 0$ there exists $\delta > 0$ such that $OB'_\delta(x) \subseteq OB_\epsilon(x)$.*

PROOF. Suppose that (a) holds. Consider a set $U \in \tau_{d'}$. Then for each $x \in U$ there exists $\epsilon > 0$ such that $OB'_\epsilon(x) \subseteq U$, but then (a) means that there exists $\delta > 0$ with $OB_\delta(x) \subseteq OB'_\epsilon(x) \subseteq U$, so x is in the τ_d-interior of U. Using this we see that $\tau_{d'} \subseteq \tau_d$. Similarly, if (b) holds then $\tau_d \subseteq \tau_{d'}$. Conversely, suppose that $\tau_d = \tau_{d'}$. Now $OB'_\epsilon(x)$ is open with respect to d', so it must be open with respect to d, and it contains x, so there must exist δ as in (a). A symmmetrical argument shows that (b) also holds. $\square$

LEMMA 2.47. [lem-strong-equiv]

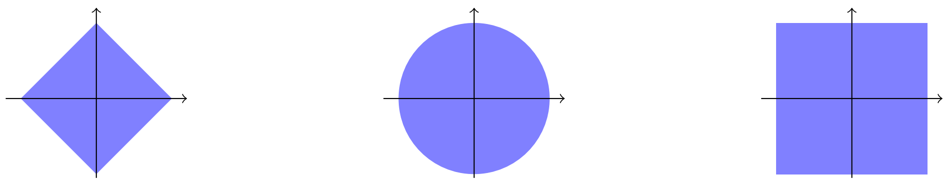
If d and d' are strongly equivalent, then they are weakly equivalent (so they give the same topology).

PROOF. Let A and A' be as in the definition. Then in part (a) of Lemma 2.46 we can take $\delta = \epsilon/A'$, and in part (b) we can take $\delta = \epsilon/A$. $\square$

Definition 2.48. [defn-standard-norms]
We can define norms and metrics on $\mathbb{R}^n$ as follows:

$$\begin{aligned} \|x\|_1 &= \sum_i |x_i| & d_1(x,y) &= \|x-y\|_1 \\ \|x\|_2 &= \left(\sum_i x_i^2\right)^{1/2} & d_2(x,y) &= \|x-y\|_2 \\ \|x\|_\infty &= \max\{|x_i| \;:\; 1 \le i \le n\} & d_\infty(x,y) &= \|x-y\|_\infty. \end{aligned}$$

Thus, $\|x\|_2$ is the norm that we previously denoted by the undecorated symbol $\|x\|$.



$\{x \in \mathbb{R}^2 \;:\; \|x\|_1 \le 1\}$ $\{x \in \mathbb{R}^2 \;:\; \|x\|_2 \le 1\}$ $\{x \in \mathbb{R}^2 \;:\; \|x\|_\infty \le 1\}$

Lemma 2.49. [lem-Rn-strong-equiv]
The metrics d_1, d_2 and d_∞ on $\mathbb{R}^n$ are all strongly equivalent to each other, and so define the same topology.

Proof. Suppose $x, y \in \mathbb{R}^n$. Put $z_i = |x_i - y_i|$ (for $i = 1, \dots, n$) and $r_k = \|x-y\|_k = \|z\|_k$ (for $k = 1, 2, \infty$). It will be enough to show that

$$r_2 \le \sqrt{n} r_\infty \le \sqrt{n} r_1 \le n r_2.$$

Clearly $z_i \le \max\{z_1, \dots, z_n\} = r_\infty$ for all i, so

$$r_2^2 = \sum_i z_i^2 \le \sum_i r_\infty^2 = n r_\infty^2,$$

so $r_2 \le \sqrt{n} r_\infty$.

Next, note that $r_\infty = z_j$ for some j, so r_∞ is one of the terms in the sum that defines r_1, and all these terms are nonnegative, so $r_\infty \le r_1$, so $\sqrt{n} r_\infty \le \sqrt{n} r_1$.

Finally, put $w_i = z_i - r_1/n$, so $w_i^2 = z_i^2 - 2 z_i r_1/n + r_1^2/n^2$. We then have

$$\begin{aligned} \sum_{i=1}^n w_i^2 &= \sum_{i=1}^n z_i^2 - 2 r_1 n^{-1} \sum_{i=1}^n z_i + r_1^2 n^{-2} \sum_{i=1}^n 1 \\ &= r_2^2 - 2 r_1 n^{-1} r_1 + r_1^2 n^{-2}.n \\ &= r_2^2 - r_1^2/n. \end{aligned}$$

As $\sum_i w_i^2$ is clearly nonnegative, we conclude that $r_2^2 - r_1^2/n \ge 0$ so $r_1^2/n \le r_2^2$ so $r_1 \le \sqrt{n} r_2$ so $\sqrt{n} r_1 \le n r_2$, as claimed. □

We can generalise the above as follows.

Definition 2.50. [defn-product-metric]
Let X and Y be metric spaces, and write d_X and d_Y for the associated metrics. We can define three different metrics on $X \times Y$ as follows:

$$\begin{aligned} d_1((x,y),(x',y')) &= d_X(x,x') + d_Y(y,y') \\ d_2((x,y),(x',y')) &= \sqrt{d_X(x,x')^2 + d_Y(y,y')^2} \\ d_\infty((x,y),(x',y')) &= \max(d_X(x,x'), d_Y(y,y')). \end{aligned}$$

In other words, if we put

$$u = (d_X(x,x'), d_Y(y,y')) \in \mathbb{R}^2$$

then $d_k((x,y),(x',y')) = \|u\|_k$. By essentially the same argument as in Lemma 2.49 we see that these are all strongly equivalent and so determine the same topology on $X \times Y$. Unless otherwise specified, we will generally use d_∞ by default.

We next discuss some questions about sequences.

Definition 2.51. [defn-converge]
Let X be a topological space, let $\underline{x} = (x_n)_{n\in\mathbb{N}}$ be a sequence of points in X, and let a be another point in X. We say that the sequence *converges* to a if for every open neighbourhood U of a, there exists $N \in \mathbb{N}$ such that $x_n \in U$ whenever $n \geq N$. We say that $\underline{x}$ is *convergent* if there is some point in X to which it converges.

Remark 2.52. [rem-R-convergence]
It is easy to see that this reduces to the usual notion of convergence if $X = \mathbb{R}$.

Lemma 2.53. [lem-metric-convergence]
Let X be a metric space, let $\underline{x} = (x_n)_{n\in\mathbb{N}}$ be a sequence of points in X, and let a be another point in X. Then the following are equivalent.

(a) *The sequence $(x_n)_{n\in\mathbb{N}}$ converges to a in X.*
(b) *For every $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that $d(x_n, a) < \epsilon$ whenever $n \geq N$.*
(c) *The sequence $(d(x_n, a))_{n\in\mathbb{N}}$ converges to 0 in $\mathbb{R}$.*

Proof. If (a) holds then (b) also holds, just by taking $U = OB_\epsilon(a)$ in Definition 2.51. Conversely, if (b) holds and we are given an open neighbourhood U of a then (by the definition of the metric topology) we can find $\epsilon > 0$ such that $OB_\epsilon(a) \subseteq U$. We can then take N as in (b) and we find that for $n \geq N$ we have $x_n \in OB_\epsilon(a) \subseteq U$; so (a) holds.

Finally, it is immediate from the definition of convergence in $\mathbb{R}$ (and the fact that d is nonnegative) that (b) and (c) are equivalent. □

Corollary 2.54. [cor-unique-limits]
Let $\underline{x}$ be a sequence in a metric space X. Then $\underline{x}$ converges to at most one point in X.

Proof. Suppose that $\underline{x}$ converges to a and also to b. For any $\epsilon > 0$ we can find N such that $d(a, x_n) < \epsilon/2$ when $n \geq N$, and we can also find M such that $d(x_m, b) < \epsilon/2$ when $n \geq M$. now put $L = \max(N, M)$ and observe that $d(a,b) \leq d(a, x_L) + d(x_L, b) < \epsilon/2 + \epsilon/2 = \epsilon$. As this holds for all $\epsilon > 0$ we must have $d(a,b) = 0$ and thus $a = b$. □

Definition 2.55. [defn-subsequence]
Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in X. By a *subsequence* we mean a sequence of the form $(x_{n_k})_{k\in\mathbb{N}}$, where $n_k \in \mathbb{N}$ for all k with $n_0 < n_1 < n_2 < \cdots$ (and so $n_k \geq k$).

Lemma 2.56. [lem-subsequence-limit]
If $\underline{x}$ converges to a and $\underline{y}$ is a subsequence of $\underline{x}$, then $\underline{y}$ also converges to a.

Proof. We have $y_k = x_{n_k}$ for some strictly increasing sequence of integers n_k. Let U be an open neighbourhood of a. By assumption, there exists N such that $x_n \in U$ for all $n \geq N$. Now for $k \geq N$ we have $n_k \geq k \geq N$ and so $y_k = x_{n_k} \in U$. Thus, $\underline{y}$ also converges to a. □

Definition 2.57. [defn-sequentially-closed]
Let X be a topological space, and let Y be a subset of X. We say that Y is *sequentially closed* if for every sequence $\underline{y}$ in Y and every limit $a \in X$ for $\underline{y}$, we actually have $a \in Y$.

Proposition 2.58. [prop-sequentially-closed]
In any topological space, every closed set is sequentially closed. In any metric space, every sequentially closed set is closed.

Proof. First let X be arbitrary, and suppose that Y is closed. Let $\underline{y}$ be a sequence in Y that converges to $a \in X$. Then for any open neighbourhood U of a there exists N such that $y_n \in U$ whenever $n \geq N$. In particular we have $y_N \in U \cap Y$, so $U \cap Y \neq \emptyset$. This means that x is a closure point of Y, but Y is closed, so $x \in Y$ as required.

Now suppose that X is a metric space, and consider the converse. Let Y be a sequentially closed subset of X. Let x be a closure point of Y. Then for each $n \in \mathbb{N}$ we have an open neighbourhood $OB_{2^{-n}}(x)$ of x which must meet Y, so we can choose $y_n \in Y$ with $d(y_n, x) < 2^{-n}$. This gives a sequence $\underline{y} = (y_n)_{n\in\mathbb{N}}$, which clearly converges to x. By hypothesis, we must have $x \in Y$. As x was an arbitrary closure point of Y, we deduce that Y is closed. $\square$

2.1. Countability properties. We next consider various countability properties that a topological space may or may not enjoy. See Section 35.1 for a review of basic facts about countability.

DEFINITION 2.59. [`defn-nbhd-basis`]
Let X be a topological space, and let x be a point of X. A *neighbourhood basis* at x is set β of neighbourhoods of x such that for any neighbourhood U of x there is a set $V \in \beta$ such that $V \subseteq U$.

REMARK 2.60. [`rem-nbhd-basis`]
We have not insisted that the neighbourhoods in β must be open, because that would be inconvenient for certain applications. However, if β is a neighbourhood basis at x then it is easy to see that the set $\beta' = \{\mathrm{int}(V) \;:\; V \in \beta\}$ is a neighbourhood basis consisting of open neighbourhoods.

DEFINITION 2.61.

(a) We say that X is *separable* if it has a countable dense subset.
(b) We say that X is C_1 or *first countable* if each point $x \in X$ has a countable neighbourhood basis.
(c) We say that X is C_2 or *second countable* if there is a countable basis for the topology on X.

PROPOSITION 2.62. [`prop-metric-first-countable`]
Any metric space is first countable. Moreover, any separable metric space is second countable.

PROOF. Let X be a metric space. For each point $x \in X$ the balls $OB_{2^{-n}}(x)$ (for $n \in \mathbb{N}$) form a countable neighbourhood basis for x, so X is first countable. Now suppose we have a countable dense subset $Y \subseteq X$, and put $\beta = \{OB_{2^{-n}}(y) \;:\; y \in Y\}$, which is again countable. Consider an open set $U \subseteq X$ and a point $x \in X$. As U is open, there exists $n \in \mathbb{N}$ such that $OB_{2^{-n}}(x) \subseteq U$. As Y is dense, there exists $y \in Y$ with $y \in OB_{2^{-n-1}}(x)$, or equivalently $x \in OB_{2^{-n-1}}(y)$. Moreover, for $z \in OB_{2^{-n-1}}(y)$ we have

$$d(x,z) \leq d(x,y) + d(y,z) < 2^{-n-1} + 2^{-n-1} = 2^{-n},$$

so $z \in U$. Thus, $OB_{2^{-n-1}}(y)$ is a set in β that contains x and is contained in U. It follows that β is a countable basis for X, as required. $\square$

PROPOSITION 2.63. [`prop-second-countable`]
Any second countable space is both separable and first countable.

PROOF. Let X be a second countable space, and let β be a countable basis for the topology. For each nonempty set $U \in \beta$, choose a point $y_U \in U$, then put $Y = \{y_U \;:\; U \in \beta,\ U \neq \emptyset\}$, so Y is countable. Consider an arbitrary point $x \in X$. For any open neighbourhood V of x we can find a basic open set U with $x \in U \subseteq V$, and then $y_U \in U \cap Y$, so $U \cap Y \neq \emptyset$. This means that x is a closure point of Y, but x was arbitrary, so $\overline{Y} = X$, or in other words Y is dense. Thus, X is separable.

It is also clear that $\{U \in \beta \;:\; x \in U\}$ is a countable basis of neighbourhoods for x, so X is first countable. $\square$

PROPOSITION 2.64. [`prop-subspace-second-countable`]
Any subspace of a second countable space is second countable.

PROOF. If β is a countable basis for the topology on X and $Y \subseteq X$ then $\{U \cap Y \;:\; U \in \beta\}$ is a countable basis for the subspace topology on Y. $\square$

EXAMPLE 2.65. [`prop-binary-countable`]
In Example 2.6 we defined the space X of binary sequences and the subsets

$$C_n(x) = \{y \in X \;:\; x_i = y_i \text{ for all } i < n\},$$

which form a basis β for a topology on X.

Say that $x \in X$ is *eventually zero* if there exists n such that $x_i = 0$ for all $i \geq n$. Let Y be the subset of sequences that are eventually zero. This is countable; in fact, we can define an explicit bijection $f\colon Y \to \mathbb{N}$ by $f(x) = \sum_i x_i 2^i$, with the inverse being given by binary expansion. For any $x \in X$ and $n \in \mathbb{N}$ we can define $\zeta_n(x) \in Y$ by

$$\zeta_n(x)_i = \begin{cases} x_i & \text{if } i < n \\ 0 & \text{if } i \geq n. \end{cases}$$

Note that $C_n(x) = C_n(\zeta_n(x))$, so $\beta = \{C_n(y) \,:\, n \in \mathbb{N}, y \in Y\}$, so β is countable. This means that X is second countable, and therefore also first countable and separable. In fact, the set Y meets every basic open set, so it is dense as well as countable.

Example 2.66. [`eg-padic-countable`]
In Example 2.26 we exhibited a basis

$$\beta = \{n + p^i\mathbb{Z} \,:\, n \in \mathbb{Z},\ i \in \mathbb{N}\}$$

for the p-adic topology on $\mathbb{Z}$. It is clear that β is countable, so with this topology $\mathbb{Z}$ is second countable, and therefore also first countable and separable. Of course separability is trivial here, because $\mathbb{Z}$ itself is countable.

Example 2.67. [`eg-Rn-countable`]
Given vectors $a, b \in \mathbb{Q}^n$, we put

$$U(a,b) = \{x \in \mathbb{R}^n \,:\, a_i < x_i < b_i \text{ for all } i\} = \prod_{i=1}^{n} (a_i, b_i).$$

We then put $\beta = \{U(a,b) \,:\, a, b \in \mathbb{Q}^n\}$. This is a countable family of open subsets of $\mathbb{R}^n$, and it is not hard to check that it is a basis for the standard topology. This means that $\mathbb{R}^n$ is second countable. The set $\mathbb{Q}^n$ is countable and dense.

Example 2.68. [`eg-lanes-countable`]
Let X denote $\mathbb{R}^2$ equipped with the lane metric described in Example 2.39, so

$$d((x,y),(x',y')) = \begin{cases} |y - y'| & \text{if } x = x' \\ |y| + |x - x'| + |y'| & \text{if } x \neq x'. \end{cases}$$

We claim that X is not separable. Indeed, let A be any countable subset of X, and let B denote the image of A under the vertical projection $(x,y) \mapsto x$. Then B will also be countable, so it cannot be all of $\mathbb{R}$, so we can choose $u \in \mathbb{R} \setminus B$. We then find that the point $(u,1)$ has distance at least one from all points in A, so it is not a closure point of A, so A is not dense. As X is not separable, it cannot be second countable. However, it is first countable by Proposition 2.62.

For the next example, we will need the notion of the supremum of a subset of $\mathbb{R}$. We pause to recall some basic facts about this.

Definition 2.69. [`defn-sup`]
Consider a subset $A \subseteq \mathbb{R}$.

(a) An *upper bound* for A is a number $u \in \mathbb{R}$ such that $a \leq u$ for all $a \in A$. We say that A is *bounded above* if there exists an upper bound.
(b) A *lower bound* for A is a number $v \in \mathbb{R}$ such that $v \leq a$ for all $a \in A$. We say that A is *bounded below* if there exists a lower bound.
(c) Suppose that A is nonempty and bounded above. Then there exists a unique upper bound u for A with the property that $u \leq u'$ for any other upper bound u'. We write $\sup(A)$ for u, and call it the *least upper bound* (or *supremum*) of A.
(d) Suppose that A is nonempty and bounded below. Then there exists a unique lower bound v for A with the property that $v \geq v'$ for any other lower bound v'. We write $\inf(A)$ for v, and call it the *greatest lower bound* (or *infimum*) of A.

(e) If A is not bounded above, we will sometimes write $\sup(A) = \infty$. Similarly, if A is not bounded below, we sometimes write $\inf(A) = -\infty$. By default we also take $\sup(\emptyset) = -\infty$ and $\inf(\emptyset) = +\infty$. However, in any argument that only involves subsets of $[a, b]$, we may make a temporary convention that $\inf(\emptyset) = b$ and $\sup(\emptyset) = a$.

It is not obvious that numbers u and v as in (c) and (d) exist, but this is one of the key properties of $\mathbb{R}$. We have included a proof in Appendix 34.

Example 2.70. [`eg-not-separable`]
Consider the set X of all sequences $x = (x_0, x_1, x_2, \dots)$ with $x_i \in [0, 1]$ for all i. We will make this a metric space using the metric

$$d(x, y) = \sup\{|x_i - y_i| \,:\, i \in \mathbb{N}\}.$$

We claim that this is not separable. To see this, let P denote the set of all subsets of $\mathbb{N}$, and recall that this is uncountable. Define a function $\sigma\colon X \to P$ by $\sigma(x) = \{i \in \mathbb{N} \,:\, x_i > 1/2\}$. (We do not claim that this is continuous in any sense.) For any countable subset $Y \subseteq X$, note that $\sigma(Y)$ is a countable subset of P, so it cannot be all of P, so we can choose $A \subseteq \mathbb{N}$ such that $A \neq \sigma(y)$ for all $y \in Y$. Define $x \in X$ by

$$x_i = \begin{cases} 1 & \text{if } i \in A \\ 0 & \text{if } i \notin A. \end{cases}$$

Consider a sequence $y \in Y$. As $\sigma(y) \neq A$, one of the following must hold:

(a) There exists i with $i \in \sigma(y)$ but $i \notin A$, so $y_i > 1/2$ and $x_i = 0$, so $|x_i - y_i| > 1/2$; or
(b) There exists i with $i \in A$ but $i \notin \sigma(y)$, so $y_i \leq 1/2$ and $x_i = 1$, so $|x_i - y_i| \geq 1/2$.

Either way we find that $d(x, y) \geq 1/2$. Thus, the ball $OB_{1/3}(x)$ is an open neighbourhood of x not meeting Y, so Y is not dense in X.

Lemma 2.71. *Let X be a first countable space, and let x be a point of X. Then there is a basis of neighbourhoods for x of the form $(U_n)_{n\in\mathbb{N}}$ with $U_{i+1} \subseteq U_i$ for all i.*

Proof. By the definition of first countability we can choose a countable basis of open neighbourhoods for x, and as it is countable we can index it as $(V_n)_{n\in\mathbb{N}}$ say. We then put $U_n = V_0 \cap \dots \cap V_n$. This certainly gives a sequence of open neighbourhoods of x with $U_{i+1} \subseteq U_i$ for all i. If W is an arbitrary neighbourhood of x then by assumption we have $x \in V_n \subseteq W$ for some n, but this implies that $x \in U_n \subseteq W$. It follows that $(U_n)_{n\in\mathbb{N}}$ is again a neighbourhood basis. □

Recall from Definition 2.57 that a subset $Y \subseteq X$ is *sequentially closed* if Y contains every limit in X of every sequence in Y. Closed sets are always sequentially closed.

Proposition 2.72. *Let X be a first countable space. Then any sequentially closed subset of X is closed.*

Proof. Let $Y \subseteq X$ be a sequentially closed subset, and let x be a closure point of Y; we must show that $x \in Y$. Choose a nested neighbourhood basis $(U_n)_{n\in\mathbb{N}}$ for x as in Lemma 2.71. As x is a closure point of Y, each set U_n must meet Y, so we can choose $y_n \in U_n \cap Y$ for all n. As the sets U_n are nested, we have $y_m \in U_n$ for all $m \geq n$. As the sets U_n form a neighbourhood basis, this means that the sequence $(y_n)_{n\in\mathbb{N}}$ converges to x. As $y_n \in Y$ and Y is sequentially closed, we see that $x \in Y$ as required. □

Exercise 2.1. [`ex-open`]
For which of the following pairs of sets $Y_i \subseteq X_i$ is Y_i open in X_i?

$$\begin{aligned}
X_0 &= \mathbb{R} & Y_0 &= \mathbb{R} \setminus \mathbb{Z} = \{x \in \mathbb{R} \,:\, x \notin \mathbb{Z}\} \\
X_1 &= [-1, 1] & Y_1 &= [0, 1] \\
X_2 &= \mathbb{Q} & Y_2 &= \mathbb{Q} \cap [-\sqrt{2}, \sqrt{2}] \\
X_3 &= \mathbb{R} & Y_3 &= \{x \in \mathbb{R} \,:\, x \neq 1/n \text{ for any } n \in \mathbb{N}^+\} \\
X_4 &= \{1/n \,:\, n \in \mathbb{N}^+\} & Y_4 &= \{1/(n+1) \,:\, n \in \mathbb{N}^+\} \\
X_5 &= [0, 1] \cup [2, 3] & Y_5 &= [0, 1].
\end{aligned}$$

Solution: The set Y_i is open in X_i for $i \in \{0, 2, 4, 5\}$. For $i = 2$ one should note that Y_2 can also be described as $\mathbb{Q} \cap (-\sqrt{2}, \sqrt{2})$ (because $\pm\sqrt{2} \notin \mathbb{Q}$). For $i = 1$ and $i = 3$, one should note that 0 is in Y_i but no interval $(-\epsilon, \epsilon)$ is contained in Y_i.

Exercise 2.2. [ex-findex]
Find examples of the following situations:

(a) A set $X \subset \mathbb{R}$ which is equal to its boundary.
(b) A set $X \subset \mathbb{R}$ which is not the closure of its interior.
(c) A set $X \subset \mathbb{R}$ which is the interior of its closure.
(d) A set $X \subset \mathbb{Q}$ which is both open and closed in $\mathbb{Q}$.
(e) An infinite, bounded, closed set $X \subset \mathbb{R}$ with empty interior.
(f) Subsets $X, Y \subset \mathbb{R}$ with $\overline{X \cap Y} \neq \overline{X} \cap \overline{Y}$
(g) A sequence of open sets $U_n \subset \mathbb{R}$ for $n \in \mathbb{N}$ whose intersection is not open.

Solution: Many solutions are possible. Some examples are as follows.

(a) We need X to be closed with empty interior, so we can take $X = \mathbb{Z}$ or $X = \{0\}$ or $X = \emptyset$.
(b) We can take $X = (0, 1)$, so $\mathrm{int}(X) = X = (0, 1) \neq [0, 1] = \mathrm{cl}(\mathrm{int}(X))$. Alternatively, we can take $X = \mathbb{Z}$ so $\mathrm{int}(X) = \emptyset$ so $\mathrm{cl}(\mathrm{int}(X)) = \emptyset \neq X$.
(c) We can take $X = (0, 1)$, so $\mathrm{cl}(X) = [0, 1]$ and $\mathrm{int}(\mathrm{cl}(X)) = (0, 1) = X$. Alternatively, we can just take $X = \emptyset$.
(d) We can take $X = (-\infty, \pi) \cap \mathbb{Q} = (-\infty, \pi] \cap \mathbb{Q}$. The first description shows that this is closed in $\mathbb{Q}$, and the second shows that it is open. Alternatively, we can take $X = \{x \in \mathbb{Q} \;:\; x^2 < 2\}$.
(e) We can take $X = \{1/n \;:\; n \in \mathbb{Z}_+\} \cup \{0\}$. The Cantor set (which will be discussed in Example 3.25) is another commonly used example.
(f) We can take $X = (-\infty, 0)$ and $Y = (0, \infty)$, so $\overline{X \cap Y} = X \cap Y = \emptyset$ but $\overline{X} \cap \overline{Y} = \{0\}$.
(g) $U_n = (-2^{-n}, 2^{-n})$.

Exercise 2.3. [ex-topology]
Let τ denote the usual topology on $\mathbb{R}$. Which of the following collections of subsets of $\mathbb{R}$ form topologies ?

(a) $\sigma_0 = \{U \subseteq \mathbb{R} \;:\; U \subseteq \mathbb{Q}\}$
(b) $\sigma_1 = \{U \subseteq \mathbb{R} \;:\; U \cap \mathbb{Q} = V \cap \mathbb{Q} \text{ for some } V \in \tau\}$
(c) $\sigma_2 = \{[a, \infty) \;:\; a \in \mathbb{R}\} \cup \{\emptyset, \mathbb{R}\}$
(d) $\sigma_3 = \{(a, \infty) \;:\; a \in \mathbb{R}\} \cup \{\emptyset, \mathbb{R}\}$
(e) $\sigma_4 = \{U \subseteq \mathbb{R} \;:\; 1 \in U\}$
(e) $\sigma_5 = \{U \subseteq \mathbb{R} \;:\; 0 \notin U \text{ or } 1 \in U\}$
(f) $\sigma_6 = \{U \subseteq \mathbb{R} \;:\; x \in U \iff x + 1 \in U\}$

Solution:

(a) The collection σ_0 is not a topology, because $\mathbb{R} \notin \sigma_0$, so axiom T0 fails.
(b) The collection σ_1 is a topology.
(c) The collection σ_2 is not a topology. The sets $[\epsilon, \infty)$ for $\epsilon > 0$ lie in σ_2, but their union does not:
$$\bigcup_{\epsilon > 0} [\epsilon, \infty) = (0, \infty) \notin \sigma$$
This contradicts axiom T1.
(d) The collection σ_3 is a topology. Axiom T0 holds by definition. Axiom T1 holds essentially because
$$\bigcup_I (a_i, \infty) = (\inf_I a_i, \infty)$$
Suppose we have a family $(U_i)_{i \in I}$ of sets in σ_3. We want to know that $U = \bigcup_I U_i \in \sigma_3$. For some i we may have $U_i = \emptyset$; these terms can be discarded without affecting the union. For other i we may have $U_i = X$; if so then $U = X \in \sigma_3$. If there are no i for which $U_i = X$ then the question reduces to the equation above. Similarly, T2 holds because of the equation
$$(a, \infty) \cap (b, \infty) = (\max(a, b), \infty)$$

apart from a little fiddling with exceptional cases.

(e) The collection σ_4 does not contain the empty set, so it is not a topology.
(f) The collection σ_5 is a topology.
(g) The collection σ_6 is a topology.

EXERCISE 2.4. [ex-metric]
Which of the following pairs (X, d) is a metric space ?

(a) $X = \mathbb{R}^n$, $\quad d(\underline{x}, \underline{y}) = \sum_{k=1}^n k|x_k - y_k|$
(b) $X = \mathbb{R}$, $\quad d(x, y) = (x - y)^2$
(c) $X = \mathbb{R}$, $\quad d(x, y) = \begin{cases} \min(|x-y|, 1) & \text{if } x - y \in \mathbb{Q} \\ 1 & \text{if } x - y \notin \mathbb{Q} \end{cases}$
(d) $X = \mathbb{R}^n$, $\quad d(\underline{x}, \underline{y}) = \min_k |x_k - y_k|$
(e) $X = \mathbb{Q}$. If $x = y$ we take $d(x, y) = 0$, otherwise we can write $x - y$ as $2^n a/b$ where a and b are odd integers and n is also an integer. In this case we take $d(x, y) = 2^{-n}$.
(f) $X = \mathbb{Z}$, $\quad d(x, x) = 0$. If $x \neq y$ write $x - y = 3^n a$ where a is an integer not divisible by 3, and take $d(x, y) = 3^n$.

Solution:

(a) Yes.
(b) No. The triangle inequality M2 fails for $x = -1$, $y = 0$, $z = 1$ for example.
(c) Yes. To prove this, it helps to show first that
$$\overline{d}(x, y) = \min(|x - y|, 1)$$
gives a metric on $\mathbb{R}$ (in fact, it induces the same topology as the usual metric). This is theorem 9.1 in the book.
(d) This is certainly not a metric space, as we have
$$d((1, 0), (0, 1)) = 0 \text{ but } (1, 0) \neq (0, 1)$$
contrary to axiom M3. It is not even a semimetric space, as the triangle inequality fails for $x = (0, 0)$, $y = (1, 0)$ and $z = (1, 1)$.
(e) Yes.
(f) No. The triangle inequality fails for $x = 0$, $y = 1$ and $z = 3$.

EXERCISE 2.5. [ex-fourteen]
For this question we use the notation
$$iA = \text{ interior of } A$$
$$kA = \text{ closure of } A$$
$$cA = \text{ complement of } A$$
It is interesting to ask what sets we can get by starting with a given set A and repeatedly applying the operators i, k, and c.

(a) "Simplify" the following expressions:
$$ccA \quad kkA \quad iiA \quad cicA \quad ckcA$$
(b) Prove that if A is the closure of some open set U, then $A = kiA$.
(c) Prove that $kikiA = kiA$ for any A, and hence that $ikikB = ikB$ for any B.
(d) Prove that for any set A, at most fourteen different sets (including A itself) can be obtained from A by repeatedly applying the operations i,k, and c. Seven of these are "roughly the same as A" and the other seven "roughly the same as cA".
(e) Find a subset $A \subset \mathbb{R}$ such that all fourteen of these sets are different. Hints: If you take care of the first seven, the other seven will probably take care of themselves. You will want to build A out of several different chunks spaced out along the real line.

Solution:

(a)

$$\begin{aligned} ccA &= A \\ kkA &= kA \\ iiA &= iA \\ cicA &= kA \\ ckcA &= iA. \end{aligned}$$

(b) As $U \subseteq A$ and U is open, we have $U \subseteq iA$. This implies that $A = kU \subseteq kiA$. On the other hand, we have $iA \subseteq A$ so $kiA \subseteq kA = kkU = kU = A$. Thus $kiA = A$.

(c) Applying the above to the case $U = iB$, we find that $kikiB = kiB$ for any B. Applying this in turn with $B = cC$ we get $cikikC = kikicC = kicC = cikC$ and thus $ikikC = ikC$.

(d) A typical set obtained from A by applying the operations i, k and c is something like $kkciccikkkciA$. We use the equations $ci = kc$ and $ck = ic$ to sweep the c's to the right, and then cancel them using $c^2 =$ identity. This leaves $kkkkiiiiA$. We then use $k^2 = k$ and $i^2 = i$ to eliminate repetitions, giving kiA. In the general case, we are left with a string of alternating i's and k's, followed either by A or by cA. If the string of i's and k's has length > 3, then we can use $kiki = ki$ or $ikik = ik$ to shorten it. This leaves us with 14 possibilities:

$$\begin{array}{rl} A & cA \\ iA & icA \\ kA & kcA \\ ikA & ikcA \\ kiA & kicA \\ kikA & kikcA \\ ikiA & ikicA \end{array}$$

The sets on the left are in some sense roughly the same size as A; they are at least bounded if A is bounded, for example. The ones on the right are roughly the same size as cA.

(e)

$$\begin{aligned} A &= A_0 \cup A_1 \cup A_2 \cup A_3 \\ A_0 &= \mathbb{Q} \cap (0,1) \\ A_1 &= [2,5] \setminus (\mathbb{Q} \cap (3,4)) \\ A_2 &= \{6 + 1/n \ : \ n \in \mathbb{Z}_+\} \\ A_3 &= [8,10] \setminus \{9 + 1/n \ : \ n \in \mathbb{Z}_+\}. \end{aligned}$$

Exercise 2.6. [`ex-boundary`]
Now write

$$bA = \text{ boundary of } A = kA \cap kcA$$

(a) "Simplify" the following expressions:

$$k(A \cup B) \quad i(A \cap B) \quad c(A \cap B) \quad A \cap (B \cup C) \quad bcA \quad kbA$$

(b) Prove that if A is closed, then $ibA = \emptyset$, and thus that $b^2A = bA$.
(c) Prove that $b^3A = b^2A$ for any set A.
(d) Find a set $A \subset \mathbb{R}$ with $A \neq bA \neq b^2A$.
(e) Show that $bA = \emptyset$ if and only if A is both open and closed.

Solution:

(a)

$$\begin{aligned} k(A \cup B) &= kA \cup kB \\ i(A \cap B) &= iA \cap iB \\ c(A \cap B) &= cA \cup cB \\ A \cap (B \cup C) &= (A \cap B) \cup (A \cap C) \\ bcA &= bA \\ kbA &= bA \end{aligned}$$

(b) Suppose A is closed. Then $bA = A \cap kcA \subseteq A$, so $ibA \subseteq iA$. On the other hand,

$$ibA \subseteq bA = A \cap kcA \subseteq kcA = ciA$$

Thus $ibA \subseteq iA \cap ciA = \emptyset$, so $ibA = \emptyset$. This implies that $bbA = kbA \cap kcbA = bA \cap cibA = bA$.

(c) For general A, we know that $B = bA$ is closed so by part (b) $b^2B = bB$, in other words $b^3A = b^2A$.

(d)

$$A = \mathbb{Q} \qquad bA = \mathbb{R} \qquad b^2A = \emptyset$$

(e) For any A we have $bA = kA \cap ciA = kA \setminus iA$, and it is always the case that $iA \subseteq A \subseteq kA$. From this we deduce that $bA = \emptyset$ iff $iA = A = kA$ iff A is both open and closed.

3. Continuous Maps

Definition 3.1. [`defn-preimage`]

Let $f\colon X \to Y$ be any function. Given a subset $A \subseteq X$ we write $f(A) = \{f(a) \;:\; a \in A\} \subseteq Y$, and call this the *image* of A under f. Given a subset $B \subseteq Y$ we write $f^{-1}(B) = \{a \in X \;:\; f(a) \in B\}$, and call this the *preimage* of B under f.

Proposition 3.2. [`prop-preimage`]

Let $f\colon X \to Y$ be a function. Then for any sets $A, B \subseteq Y$ we have

(a) $f^{-1}(A \cap B) = f^{-1}(A) \cap f^{-1}(B)$

(b) $f^{-1}(A \cup B) = f^{-1}(A) \cup f^{-1}(B)$

(c) $f^{-1}(Y \setminus A) = X \setminus f^{-1}(A)$

(d) $f(f^{-1}(B)) \subseteq B$, *with equality if f is surjective.*

(e) $A \subseteq f^{-1}(f(A))$, *with equality if f is injective.*

Proof. This is all trivial once one has untangled the notation. For (a), we note that

$$\begin{aligned} x \in f^{-1}(A \cap B) &\Leftrightarrow f(x) \in A \cap B \\ &\Leftrightarrow (f(x) \in A) \text{ and } (f(x) \in B) \\ &\Leftrightarrow (x \in f^{-1}(A)) \text{ and } (x \in f^{-1}(B)) \\ &\Leftrightarrow x \in f^{-1}(A) \cap f^{-1}(B). \end{aligned}$$

The next two parts are similar. For part (d), if $y \in f(f^{-1}(B))$ then $y = f(a)$ for some $a \in f^{-1}(B)$, and the fact that $a \in f^{-1}(B)$ means precisely that $f(a) \in B$, so $y \in B$ as required. Now suppose that f is surjective, and consider the reverse inclusion. For $b \in B$ we certainly have $b = f(a)$ for some $a \in X$. As $f(a) = b \in B$ we have $a \in f^{-1}(B)$, and therefore $b = f(a) \in f(f^{-1}(B))$ as required. Part (e) is similar to (d). □

Definition 3.3. [`defn-cts`]

Let X and Y be topological spaces, and let f be a function from X to Y. We say that f is *continuous* if for each open set $V \subseteq Y$, the preimage $f^{-1}(V)$ is open in X.

Remark 3.4. [`rem-cts-closed`]

Using Proposition 3.2(c) we see that continuity can also be defined in terms of closed sets: a function $f\colon X \to Y$ is continuous if and only if for every closed subset $G \subseteq Y$, the preimage $f^{-1}(G)$ is closed in X.

Proposition 3.5. [`prop-comp-cts`]

If $f\colon X \to Y$ and $g\colon Y \to Z$ are continuous then so is $g \circ f\colon X \to Z$.

Proof. For any open set $W \subseteq Z$ we see from the definitions that $(g \circ f)^{-1}(W) = f^{-1}(g^{-1}(W))$. Here $g^{-1}(W)$ is open (because g is continuous), and thus $f^{-1}(g^{-1}(W))$ is open (because f is continuous). We conclude that $g \circ f$ is continuous as claimed. $\square$

Remark 3.6. [rem-cat-top]
Throughout this book we will use the language of category theory, which is reviewed in Appendix 36. The above proposition (together with the trivial fact that identity maps are continuous) means that we have a category **Spaces**, whose objects are topological spaces, and whose morphisms are the continuous functions.

Remark 3.7. [rem-cat-metric]
Similarly one can define a category **MetricSpaces**, whose objects are metric spaces, and whose morphisms from X to Y are the functions $f\colon X \to Y$ that are continuous with respect to the corresponding metric topologies. There is a functor $U\colon \mathbf{MetricSpaces} \to \mathbf{Spaces}$ that sends each metric space X to the same set equipped with the metric topology. Here $\mathbf{MetricSpaces}(X,Y)$ is exactly the same set as $\mathbf{Spaces}(UX,UY)$, and the map

$$U\colon \mathbf{MetricSpaces}(X,Y) \to \mathbf{Spaces}(UX,UY)$$

is just the identity. Thus, the functor U is full and faithful. It is not hard to check that the topology on the Sierpinski space (Example 2.7) does not arise from any metric (or even any semimetric). It follows that the functor U is not essentially surjective, so it is not an equivalence.

Proposition 3.8. [prop-cts-subbasis]
Let $f\colon X \to Y$ be a function between topological spaces. Suppose that σ is a subbasis for the topology on Y, and that for all $V \in \sigma$ the preimage $f^{-1}(V)$ is open in X. Then f is continuous.

Proof. Recall that $\beta(\sigma)$ is the set of subsets of the form $V = V_1 \cap \dots \cap V_r$, with $V_1, \dots, V_r \in \sigma$. For such V we have $f^{-1}(V) = \bigcap_{i=1}^r f^{-1}(V_i)$, and the sets $f^{-1}(V_i)$ are open by assumption, so $f^{-1}(V)$ is open. Next, $\tau(\beta(\sigma))$ is the set of subsets $W \subseteq Y$ that can be written as the union of some family $(W_j)_{j\in J}$, with $W_j \in \beta(\sigma)$ for all j. This means that $f^{-1}(W)$ is the union of the sets $f^{-1}(W_j)$, each of which is open by our first step, so $f^{-1}(W)$ is open in X. We have assumed that $\tau(\beta(\sigma))$ is the given topology on Y, and the claim now follows. $\square$

Proposition 3.9. [prop-cts-metric]
Let $f\colon X \to Y$ be a function between metric spaces. Then f is continuous (with respect to the metric topologies) if and only if for every $x \in X$ and $\epsilon > 0$ there exists $\delta > 0$ such that $d(x,x') < \delta$ implies $d(f(x), f(x')) < \epsilon$.

Proof. First suppose that f is continuous. Given $x \in X$ and $\epsilon > 0$ we observe that $OB_\epsilon(f(x))$ is an open set in Y containing $f(x)$, so the set $U = f^{-1}(OB_\epsilon(x))$ is an open subset of X containing x. This means that U must contain $OB_\delta(x)$ for some $\delta > 0$, so for $d(x,x') < \delta$ we have $x' \in U$ and therefore $f(x') \in OB_\epsilon(f(x))$, or equivalently $d(f(x), f(x')) < \epsilon$ as required.

Conversely, suppose that f satisfies the ϵ-δ condition. Suppose that V is open in Y, and that $x \in f^{-1}(V)$. This means that $f(x)$ is a point in the open set V, so there exists $\epsilon > 0$ such that $OB_\epsilon(f(x)) \subseteq V$. By hypothesis, we can find $\delta > 0$ such that $d(x,x') < \delta$ implies $d(f(x), f(x')) < \epsilon$, and thus $f(x') \in OB_\epsilon(f(x)) \subseteq V$, and thus $x' \in f^{-1}(V)$. This means that $OB_\delta(x) \subseteq f^{-1}(V)$, so x is in the interior of $f^{-1}(V)$. As x was an arbitrary element of $f^{-1}(V)$, we conclude that $f^{-1}(V)$ is open, as required. $\square$

Proposition 3.10. [prop-cts-misc]
The following maps are continuous (metric topologies used everywhere):

$$\begin{aligned}
\sigma\colon \mathbb{R}^2 &\to \mathbb{R} & \sigma(x,y) &= x+y \\
\mu\colon \mathbb{R}^2 &\to \mathbb{R} & \mu(x,y) &= xy \\
\nu\colon \mathbb{R}\setminus\{0\} &\to \mathbb{R}\setminus\{0\} & \nu(x) &= 1/x \\
\rho\colon [0,\infty) &\to [0,\infty) & \rho(x) &= \sqrt{x} \\
\max,\min\colon \mathbb{R}^2 &\to \mathbb{R}
\end{aligned}$$

Proof. For definiteness, we use the metric $d = d_\infty$ on $\mathbb{R}^2$, so $d((x,y),(x',y')) = \max(|x-x'|,|y-y'|)$. It is easy to see that

$$d(\sigma(x,y),\sigma(x',y')) = |(x-x')+(y-y')| \leq |x-x'|+|y-y'| \leq 2d((x,y),(x',y')),$$

and thus that σ is continuous. Next, we claim that

$$d(\max(x,y),\max(x',y')) \leq d((x,y),(x',y')).$$

To see this, put

$$\begin{aligned} f(x,y,x',y') &= d((x,y),(x',y')) - d(\max(x,y),\max(x',y')) \\ &= \max(|x-x'|,|y-y'|) - |\max(x,y)-\max(x',y')|, \end{aligned}$$

so the claim is that $f(x,y,x',y') \geq 0$ for all $(x,y,x',y') \in \mathbb{R}^4$. It is clear that $f(x,y,x',y') = f(x',y',x,y) = f(y,x,y',x')$, and using these symmetries we reduce to the case where y, x' and y' are all less than or equal to x. In that case we have

$$\begin{aligned} \max(|x-x'|,|y-y'|) &\geq |x-x'| = x-x' \\ |\max(x,y)-\max(x',y')| &= x - \max(x',y') \leq x-x' \\ f(x,y,x',y') &\geq (x-x')-(x-x') \geq 0 \end{aligned}$$

as claimed. It follows that the map $\max\colon \mathbb{R}^2 \to \mathbb{R}$ is continuous (because we can take $\delta = \epsilon$ in the criterion of Proposition 3.9). We can also see that the map $\min\colon \mathbb{R}^2 \to \mathbb{R}$ is continuous, either by a parallel argument, or by using the identity $\min(x,y) = -\max(-x,-y)$.

Next, suppose we have a point $(x,y) \in \mathbb{R}^2$ and a constant $\epsilon > 0$. Put $\delta = \min(\epsilon/(|x|+|y|+1),1)$. If $d((x,y),(x',y')) < \delta$ we find that

$$\begin{aligned} d(\mu(x,y),\mu(x',y')) &= |x'y'-xy| = |(x'-x)y + (y'-y)x + (x'-x)(y'-y)| \\ &\leq |x'-x||y| + |y'-y||x| + |x'-x||y'-y| < \delta|y| + \delta|x| + \delta^2 \\ &\leq \delta(|x|+|y|+1) \leq \epsilon. \end{aligned}$$

It follows from this that μ is continuous. Now consider ν. Suppose we have $x \in \mathbb{R}\setminus\{0\}$ and $\epsilon > 0$. Put $\delta = \min(|x|/2, \epsilon|x|^2/2)$, and suppose that $d(x',x) < \delta$. We have the triangle inequality

$$|x| = |(x-x')+x'| \leq |x-x'| + |x'| < |x|/2 + |x'|,$$

which can be rearranged to give $|x'| > |x|/2$. We also have

$$d(\nu(x),\nu(x')) = \left|\frac{x'-x}{xx'}\right| \leq \frac{2}{|x|^2}|x-x'| < \frac{2}{|x|^2}\delta \leq \epsilon.$$

It follows that ν is also continuous. Finally, consider ρ. Suppose we have $x \geq 0$ and $\epsilon > 0$. If $x = 0$ we just take $\delta = \epsilon^2 > 0$; then for $x' \in [0,\infty)$ with $|x-x'| < \delta$ we have $0 \leq x' < \epsilon^2$ and so $0 \leq \sqrt{x'} < \epsilon$, so $|\rho(x') - \rho(x)| < \epsilon$ as required. Suppose instead that $x > 0$. Put $\delta = \epsilon\sqrt{x}$. For all $x' \geq 0$ we note that

$$(\sqrt{x'}-\sqrt{x})(\sqrt{x'}+\sqrt{x}) = x'-x,$$

so

$$|\sqrt{x'}-\sqrt{x}| = \frac{|x'-x|}{\sqrt{x'}+\sqrt{x}} \leq \frac{|x'-x|}{\sqrt{x}}.$$

Thus, if $|x'-x| < \delta$ we have $|\rho(x')-\rho(x)| < \epsilon$, as required. ☐

Corollary 3.11. [cor-CX-ops]
Let X be any space, and let $f,g\colon X \to \mathbb{R}$ be continuous maps. Then the maps

$$fg,\ f+g,\ \max(f,g),\ \min(f,g)\colon X \to \mathbb{R}$$

are all continuous. Moreover, if $f(x) \neq 0$ for all $x \in X$ then $1/f\colon X \to \mathbb{R}$ is also continuous, and if $f(x) \geq 0$ for all x then $\sqrt{f}$ is continuous.

Proof. Consider the map $h\colon X \to \mathbb{R}^2$ given by $h(x) = (f(x), g(x))$. Consider a basic open set $V = OB_\epsilon(a,b) \subset \mathbb{R}^2$ (where $\epsilon > 0$ and $a, b \in \mathbb{R}$). We have $h(x) \in V$ iff $\max(|f(x) - a|, |g(x) - b|) < \epsilon$ iff $f(x) \in OB_\epsilon(a)$ and $g(x) \in OB_\epsilon(b)$, so $h^{-1}OB_\epsilon(a,b) = (f^{-1}OB_\epsilon(a)) \cap (g^{-1}OB_\epsilon(b))$. As f and g are continuous, the sets $f^{-1}OB_\epsilon(a)$ and $g^{-1}OB_\epsilon(b)$ are open, so the same is true of $h^{-1}OB_\epsilon(a,b)$. Thus, the map h is continuous by Proposition 3.8. It follows by Proposition 3.5 that the maps $f+g = \sigma\circ h$, $fg = \mu \circ h$, $\max(f,g) = \max \circ h$ and $\min(f,g) = \min \circ h$ are also continuous. Similarly, if f is everywhere nonzero then we can regard f as a continuous map $X \to \mathbb{R} \setminus \{0\}$ and $\nu\colon \mathbb{R}\setminus\{0\} \to \mathbb{R}$ is continuous so $1/f = \nu \circ f$ is also continuous. The proof for $\sqrt{f}$ is essentially the same. □

Proposition 3.12. [eg-poly-cts]
Let $f\colon \mathbb{R} \to \mathbb{R}$ be a polynomial function; then f is continuous.

Proof. One approach is to start with the fact that constant maps and the identity map are continuous, and then prove that polynomials of degree d are continuous by induction on d using Corollary 3.11. Here, however, we will explain a more direct proof, using the criterion in Proposition 3.9.

As f is polynomial, we can write it in the form $f(x) = \sum_{i=0}^n a_i x^i$ say. Suppose we are given $x \in \mathbb{R}$ and $\epsilon > 0$. We have

$$f(x+u) = \sum_{i=0}^n a_i (x+u)^i = \sum_{i=0}^n \sum_{j=0}^i \left(\begin{array}{c} i \\ j \end{array} \right) a_i x^{i-j} u^j,$$

which can be rewritten as $\sum_{j=0}^n b_j u^j$ for certain numbers b_j. Note in particular that $f(x) = f(x+0) = b_0$. Put $\delta = \min(1, \epsilon / \sum_{j=1}^n |b_j|)$. If $|u| < \delta$ then $|u| < 1$ so $|u^j| \leq |u|$ for all $j > 0$, so

$$|f(x+u) - f(x)| = \left| \left(\sum_{j=0}^n b_j u^j \right) - b_0 \right| = \left| \sum_{j=1}^n b_j u^j \right| \leq \sum_{j=1}^n |b_j||u| < \epsilon$$

as required. □

Proposition 3.13. [prop-lin-cts]
Let V and W be vector spaces over $\mathbb{R}$, and let $f\colon V \to W$ be a linear map. Then f is continuous with respect to the linear topologies on V and W.

Proof. We first recall the basic definitions: for each linear map $\psi\colon W \to \mathbb{R}$ and each pair of real numbers a, b with $a < b$, we put

$$U(\psi, a, b) = \{w \in W \ : \ a < \psi(w) < b\}.$$

We then let σ_W denote the family of all sets of this form, and we let τ_W denote the topology with σ_W as a subbasis. Now it is clear that $f^{-1}(U(\psi,a,b)) = U(\psi\circ f, a, b) \in \sigma_V$, so the preimage under f of any subbasic open set is a subbasic open set in V. It follows by Proposition 3.8 that f is continuous. □

Definition 3.14. [defn-seq-cts]
Let X and Y be topological spaces. We say that a map $f\colon X \to Y$ is *sequentially continuous* if it has the following property: for every sequence $\underline{x}$ in X and every point $a \in X$ such that $\underline{x}$ converges to a, the sequence $(f(x_n))_{n\in\mathbb{N}}$ converges to $f(a)$.

Proposition 3.15. [prop-seq-cts]
Any continuous map is sequentially continuous. Conversely, if X is first countable, then any sequentially continuous map from X to Y is continuous. (In particular, this holds if X is a metric space.)

Proof. First let X and Y be arbitrary, and let $f\colon X \to Y$ be continuous. Consider a sequence $\underline{x}$ in X converging to a point $a \in X$. Consider a neighbourhood V of $f(a)$. Then the set $U = f^{-1}(V) \subseteq X$ is open (because f is continuous) and contains a. It follows that there is some $N \in \mathbb{N}$ such that $x_n \in U$ whenever $n \geq N$. By the definition of U, this means that $f(x_n) \in V$ whenever $n \geq N$. This means that the sequence $(f(x_n))_{n\in\mathbb{N}}$ converges to $f(a)$, as required.

Now suppose that X is first countable, and consider the converse. Let F be a closed subset of Y. By Remark 3.4, it will be enough to show that $f^{-1}(F)$ is closed. Thus, by Proposition 2.72, it will be enough to show that $f^{-1}(F)$ is sequentially closed. Consider a sequence $(x_n)_{n\in\mathbb{N}}$ in $f^{-1}(F)$ converging to a point

$a \in X$. As f is sequentially continuous, we see that $f(x_n)$ converges to $f(a)$ in Y. As $x_n \in f^{-1}(F)$, we see that the points $f(x_n)$ lie in the closed set F, so the limit $f(a)$ must also lie in F, so $a \in f^{-1}(F)$ as required. □

Definition 3.16. [`defn-lipschitz`]
Let $f\colon X \to Y$ be a function between metric spaces. We say that f is *Lipschitz* if there is a constant $K > 0$ (called a *Lipschitz constant*) such that $d(f(x), f(x')) \leq K d(x, x')$ for all $x, x' \in X$.

Proposition 3.17. [`prop-lipschitz`]
If $f\colon X \to Y$ is Lipschitz, then it is continuous.

Proof. Let K be a Lipschitz constant. Then in the criterion of Proposition 3.9, we can just take $\delta = \epsilon/K$. □

Example 3.18. [`eg-not-lipschitz`]
To see that this argument is not reversible, consider the map $f\colon \mathbb{R}^+ \to \mathbb{R}^+$ given by $f(x) = x^2$. Then for $x \neq y$ we have

$$\frac{d(f(x), f(y)}{d(x,y)} = \frac{|x^2 - y^2|}{|x-y|} = |x+y|,$$

which can be arbitrarily large. This means that f is not Lipschitz. It is, however, continuous by Example 3.12.

Remark 3.19. [`rem-cat-lipschitz`]
Let X, Y and Z be metric spaces. We write $\mathbf{Lipschitz}(X,Y)$ for the set of Lipschitz maps from X to Y, and $\mathbf{Lipschitz}_1(X,Y)$ for the subset of maps where $d(f(x), f(x')) \leq d(x, x')$, so that 1 is a Lipschitz constant for f. It is then clear that

$$\mathbf{Lipschitz}_1(X,Y) \subseteq \mathbf{Lipschitz}(X,Y) \subseteq \mathbf{MetricSpaces}(X,Y).$$

Moreover, if $f\colon X \to Y$ has Lipschitz constant K and $g\colon Y \to Z$ has Lipschitz constant L, then LK is a Lipschitz constant for $g \circ f\colon X \to Z$; and we also have $1_X \in \mathbf{Lipschitz}_1(X,X)$. This means that we have wide subcategories

$$\mathbf{Lipschitz}_1 \subseteq \mathbf{Lipschitz} \subseteq \mathbf{MetricSpaces}\,.$$

Definition 3.20. [`defn-homeo`]
A map $f\colon X \to Y$ is a *homeomorphism* if and only if it is bijective and both f and f^{-1} are continuous. Two spaces X and Y are *homeomorphic* if and only if there is a homeomorphism from one to the other.

Proposition 3.21. [`prop-interval-homeo`]
There are homeomorphisms

$$(0,1) \xrightarrow{f} (0,\infty) \xrightarrow{g} \mathbb{R}$$

given by

$$\begin{aligned} f(x) &= \frac{x}{1-x} & f^{-1}(y) &= \frac{y}{1+y} \\ g(y) &= \frac{y - y^{-1}}{2} & g^{-1}(z) &= \sqrt{1+z^2} + z \\ gf(x) &= \frac{2x-1}{2x(1-x)} & (gf)^{-1}(z) &= \frac{z + \sqrt{1+z^2}}{1 + z + \sqrt{1+z^2}}. \end{aligned}$$

Proof. First note that the definition of $f(x)$ is meaningful, because the denominator $1-x$ does not vanish for $x \in (0,1)$. Similarly, the definition of $g(y)$ is meaningful, and the expression for $gf(x)$ can be checked by straightforward algebra. We also have a well-defined function $f_1\colon (0,\infty) \to \mathbb{R}$ given by $f_1(y) = y/(1+y)$, and for $y > 0$ we can divide the inequalities $0 < y < 1+y$ by the positive number $1+y$ to obtain $0 < f_1(y) < 1$, so we can regard f_1 as a map $(0,\infty) \to (0,1)$. It is again straightforward algebra to check that $f_1(f(x)) = x$ and $f(f_1(y)) = y$, so $f\colon (0,1) \to (0,\infty)$ is a bijection with inverse f_1. Next, for any $z \in \mathbb{R}$ we know that $1 + z^2 > z^2 \geq 0$ so $\sqrt{1+z^2}$ is defined and $\sqrt{1+z^2} > \sqrt{z^2} = |z| \geq 0$, so $\sqrt{1+z^2} + z > |z| + z = 2\max(0,z) \geq 0$. We can thus define a function $g_1\colon \mathbb{R} \to (0,\infty)$ by $g_1(z) = \sqrt{1+z^2} + z$. Note that $(\sqrt{1+z^2}+z)(\sqrt{1+z^2}-z) = 1$, and from this it follows that $g(g_1(z)) = z$. It is also easy to see that when $y \in (0, \infty)$ we have $g(y)^2 + 1 = ((y + y^{-1})/2)^2$ and $(y + y^{-1})/2 > 0$ so $\sqrt{g(y)^2 + 1} = (y + y^{-1})/2$; from this it follows that $g_1(g(y)) = y$. Thus, the function $g\colon (0, \infty) \to \mathbb{R}$ is a bijection with inverse g_1. It follows that gf is a bijection with inverse

$$(gf)^{-1}(z) = f_1(g_1(z)) = \frac{z + \sqrt{1 + z^2}}{1 + z + \sqrt{1 + z^2}}.$$

By repeated use of Corollary 3.11, we see that all the functions considered are continuous. As f and $f^{-1} = f_1$ are both continuous, we see that f is a homeomorphism. Similarly, g and gf are both homeomorphisms, by the same argument. □

Remark 3.22. There are also homeomorphisms

$$\begin{aligned} \exp\colon \mathbb{R} &\to (0, \infty) \\ \log\colon (0, \infty) &\to \mathbb{R} \\ \tan\colon (-\pi/2, \pi/2) &\to \mathbb{R} \\ \tanh\colon \mathbb{R} &\to (-1, 1). \end{aligned}$$

However, one needs various digressions in real analysis to justify these.

Example 3.23. [`eg-three-homeo`]
In Example 2.17 we discussed the 29 possible topologies on the set $X = \{0, 1, 2\}$. For example, under case (c) we had the three topologies

$$\begin{aligned} \gamma_0 &= \{\emptyset, \{0\}, \{0, 1, 2\}\} \\ \gamma_1 &= \{\emptyset, \{1\}, \{0, 1, 2\}\} \\ \gamma_2 &= \{\emptyset, \{2\}, \{0, 1, 2\}\}. \end{aligned}$$

We can define a homeomorphism $f\colon (X, \gamma_0) \to (X, \gamma_1)$ by $f(0) = 1$ and $f(1) = 2$ and $f(2) = 0$. The same function f also gives a homeomorphism $(X, \gamma_1) \to (X, \gamma_2)$, so all the spaces (X, γ_i) are homeomorphic to each other. By a straightforward extension of this, we see that the 29 different topologies fall into nine homeomorphism types, corresponding to the nine cases (a) to (i) in Example 2.17.

Example 3.24. [`eg-padic-cts`]
Give $\mathbb{Z}$ the p-adic topology as in Examples 2.26 and 2.43. Consider the maps

$$\begin{aligned} \sigma\colon \mathbb{Z}^2 &\to \mathbb{Z} & \sigma(x, y) &= x + y \\ \mu\colon \mathbb{Z}^2 &\to \mathbb{Z} & \mu(x, y) &= xy \\ \max, \min\colon \mathbb{Z}^2 &\to \mathbb{Z} \end{aligned}$$

We claim that σ and μ are continuous, whereas max and min are not. Indeed, it is clear that if $x = x' \pmod{p^v}$ and $y = y' \pmod{p^v}$ then $x + y = x' + y' \pmod{p^v}$ and also $xy = x'y' \pmod{p^v}$. Now use the p-adic metric d on $\mathbb{Z}$, and the metric

$$d((x, y), (x', y')) = \max(d(x, x'), d(y, y'))$$

on $\mathbb{Z} \times \mathbb{Z}$. The above congruences then tell us that

$$\begin{aligned} d(\sigma(x, y), \sigma(x', y')) &\leq d((x, y), (x', y')) \\ d(\mu(x, y), \mu(x', y')) &\leq d((x, y), (x', y')), \end{aligned}$$

so σ and μ are Lipschitz and therefore continuous.

Now put $V = p\mathbb{Z} = \{n \in \mathbb{Z} \,:\, n = 0 \pmod{p}\}$, which is open in $\mathbb{Z}$, and put

$$U = \max^{-1}(V) = \{(n, m) \in \mathbb{Z}^2 \,:\, \max(n, m) = 0 \pmod{p}\}.$$

Now $(1, p) \in U$, but every basic open neighbourhood of $(1, p)$ contains points of the form $(1 + p^n, p)$ for large n, and $\max(1 + p^n, p) = 1 + p^n \neq 0 \pmod{p}$ so $(1 + p^n, p) \notin U$. This shows that U is not open, so max is not continuous. A similar argument works for min.

Example 3.25. [eg-shift-maps]

Let X be the space of binary sequences, as in Examples 2.6 and 2.42. Define left and right shift maps $L, R\colon X \to X$ by

$$L(x_0, x_1, x_2, x_3, \dots) = (x_1, x_2, x_3, x_4 \dots)$$
$$R(x_0, x_1, x_2, x_3, \dots) = (0, x_0, x_1, x_2, \dots).$$

We claim that these are continuous. Indeed, if we use the metric defined in Example 2.42, we find that $d(R(x), R(y)) = d(x, y)/2$, whereas $d(L(x), L(y)) \leq 2d(x, y)$, so both L and R are Lipschitz. Now define $F\colon X \to \mathbb{R}$ by

$$F(x) = \sum_{i=0}^{\infty} 2x_i/3^i.$$

We claim that this is also continuous. To see this, consider an open set $V \subseteq \mathbb{R}$ and a point $x \in F^{-1}(V)$. As $F(x) \in V$ and V is open we can find $\epsilon > 0$ such that $OB_\epsilon(F(x)) \subseteq V$. We can then find $n \in \mathbb{N}$ such that $3^{1-n} < \epsilon$. Suppose that $y \in C_n(x)$; then

$$\begin{aligned} d(F(y), F(x)) = |F(y) - F(x)| &= \left| \sum_{i=n}^{\infty} 2(x_i - y_i)/3^i \right| \leq \sum_{i=n}^{\infty} 2|x_i - y_i|/3^i \leq \frac{2}{3^n} \sum_{j=0}^{\infty} 3^{-j} \\ &= \frac{2}{3^n} \frac{1}{1 - 1/3} = \frac{1}{3^{n-1}} < \epsilon, \end{aligned}$$

so $F(y) \in V$. This shows that $C_n(x) \subseteq F^{-1}(V)$, and we conclude that $F^{-1}(V)$ is open, as required. The image of F is known as the *Cantor set*.

Proposition 3.26. [prop-linear-iso]

Let V be a vector space over $\mathbb{R}$ of dimension $n < \infty$. Then V (with the linear topology) is homeomorphic to $\mathbb{R}^n$ (with the linear topology, which is the same as the standard topology).

Proof. By standard linear algebra, there is a linear isomorphism $f\colon \mathbb{R}^n \to V$. Now f^{-1} is also linear, so f and f^{-1} are both continuous with respect to the linear topologies, so they are homeomorphisms. We also know from Proposition 2.32 that the linear topology on $\mathbb{R}^n$ is the same as the standard one. □

Remark 3.27. [rem-not-homeo]

It is surprisingly difficult to prove that two spaces are not homeomorphic, even in cases where this seems visually obvious. To address such questions in a systematic way, we need methods from algebraic topology, most of which are outside the scope of this book. However, we will prove some non-homeomorphism results by *ad hoc* methods, for example in Section 8.1.

Exercise 3.1. [ex-dense]

(a) Give an example of a countable dense subset of $\mathbb{R}$.
(b) Prove that two continuous functions f and g from $\mathbb{R}$ to $\mathbb{R}$ which agree on a dense set are equal.
(c) Suppose that $f\colon \mathbb{R} \to \mathbb{R}$ is continuous and that $f(x) + f(y) = f(x + y)$ for all x and y in $\mathbb{R}$. Prove that $f(x) = f(1)x$ for all x.

Solution:

(a) $\mathbb{Q}$ is a countable dense subset of $\mathbb{R}$.
(b) Suppose that $X \subseteq \mathbb{R}$ is dense, and that we are given two continuous functions f and g from $\mathbb{R}$ to $\mathbb{R}$ which agree on X. Put $h(x) = f(x) - g(x)$, which gives a continuous function by Corollary 3.11. It follows that the set $Y = h^{-1}(\{0\})$ is closed in $\mathbb{R}$. For $x \in X$ we have $f(x) = g(x)$ so $h(x) = 0$, which means that $X \subseteq Y$. As Y is closed, it follows that $\overline{X} \subseteq Y$, but X is dense, so $\overline{X} = \mathbb{R}$, so $Y = \mathbb{R}$. This means that $h(\mathbb{R}) = 0$, or equivalently $f = g$.
(c) Suppose that $f\colon \mathbb{R} \to \mathbb{R}$ is continuous and that $f(x) + f(y) = f(x + y)$ for all x and y in $\mathbb{R}$. Put $g(x) = f(x) - f(1)\,x$. Corollary 3.11 tells us that g is continuous, and it is easy to see that $g(x + y) = g(x) + g(y)$. In particular, we have $g(0) = g(0 + 0) = g(0) + g(0)$, which means that $g(0) = 0$. We can use this to start an induction to show that $g(nx) = n\,g(x)$ for all $n \in \mathbb{N}$.

By construction we also have $g(1) = f(1) - f(1) = 0$. This means that for $n > 0$ we have $n\,g(1/n) = g(1) = 0$, so $g(1/n) = 0$. It follows in turn that $g(m/n) = 0$ for all $m \in \mathbb{N}$, or in other words $g(q) = 0$ for all nonnegative rationals q. This gives $g(-q) = g(q) + g(-q) = g(0) = 0$, so $g(\mathbb{Q}) = 0$. As g is continuous and $\mathbb{Q}$ is dense, we deduce that $g = 0$, so $f(x) = f(1)x$ for all x as claimed.

3.1. Spaces of continuous functions. It is an important idea that various sets of continuous functions can themselves be regarded as topological spaces. A very general and powerful version of this idea will be discussed in Section 23. For the moment we will consider some more restricted versions.

DEFINITION 3.28. [`defn-CXY-metric`]
Let X be a topological space, and let Y be a metric space. Let $C(X,Y)$ be the set of continuous functions from X to Y. For $f, g \in C(X,Y)$ we put

$$d(f,g) = \sup\{d(f(x), g(x)) \; : \; x \in X\}.$$

Note that we are taking the supremum of a set that may be unbounded, so the result may be infinite. We will often write $C(X)$ for $C(X,\mathbb{R})$.

PROPOSITION 3.29. [`prop-CXY-metric`]
The above definition gives a metric on $C(X,Y)$. (The corresponding topology is called the topology of uniform convergence.*)*

PROOF. Axioms M0 and M1 are clear. For the triangle inequality, suppose we have elements $f, g, h \in C(X,Y)$. For each $x \in X$ we can apply the triangle inequality in Y together with the inequalities $d(f(x), g(x)) \leq d(f,g)$ and $d(g(x), h(x)) \leq d(g,h)$ to get

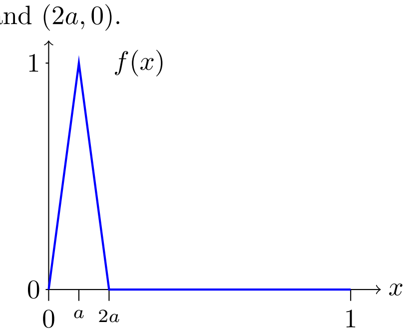
$$d(f(x), h(x)) \leq d(f(x), g(x)) + d(g(x), h(x)) \leq d(f,g) + d(g,h).$$

This means that $d(f,g) + d(g,h)$ is an upper bound for the values $d(f(x), h(x))$, so it is at least as large as the least upper bound, so $d(f,h) \leq d(f,g) + d(g,h)$. This proves M2. Finally, if $d(f,g) = 0$ then we must have $d(f(x), g(x)) = 0$ for all x, so $f(x) = g(x)$ by axiom M3 for Y, so $f = g$. □

EXAMPLE 3.30. [`eg-function-metrics`]
Now consider the set $X = C([0,1],\mathbb{R})$. We can define metrics d_1, d_2 and d_∞ on X by $d_p(f,g) = \|f - g\|_p$, where

$$\|f\|_1 = \int_{x=0}^{1} |f(x)|\,dx$$
$$\|f\|_2 = \left(\int_{x=0}^{1} f(x)^2\,dx\right)^{1/2}$$
$$\|f\|_\infty = \sup\{|f(x)| \; : \; 0 \leq x \leq 1\}.$$

Thus d_∞ is as in Definition 3.28, but the other two are different. In Corollary 10.33 we will prove the well-known fact that $\|f\|_\infty$ is finite, and it follows that the other two are finite as well. These metrics are not strongly equivalent, and in fact they define different topologies (all of which are used for different purposes in functional analysis). To see this, suppose that $a \in [0, 1/2]$, and let f_a be the function whose graph is a triangle with vertices $(0,0)$, $(a,1)$ and $(2a,0)$.



We find that

$$\|f_a\|_1 = a$$
$$\|f_a\|_2 = \sqrt{2a/3}$$
$$\|f_a\|_\infty = 1.$$

Using this we see that $\|f_a\|_\infty/\|f_a\|_2$ and $\|f_a\|_2/\|f_a\|_1$ can both be arbitrarily large.

We can use similar formulae to define semimetrics on various spaces of functions that are integrable but not continuous. Often these are not metrics. For example, if f is the function given by $f(0) = 1$ and $f(x) = 0$ for $0 < x \leq 1$ then $f \neq 0$ but $d_2(f, 0) = 0$. In this context it is usual to use Remark 2.34 to pass to a quotient set where there is a genuine metric.

We now pause briefly to discuss a common theme in Definition 2.48 and Example 3.30.

Definition 3.31. [`defn-norm`]
Let $\mathbb{K}$ be $\mathbb{R}$ or $\mathbb{C}$, and let V be a vector space over $\mathbb{K}$. A *norm* on V is a map $\phi\colon V \to [0, \infty)$ such that

N0: $\phi(tx) = |t|\,\phi(x)$ for all $t \in \mathbb{K}$ and $x \in V$.
N1: $\phi(x + y) \leq \phi(x) + \phi(y)$ for all $x, y \in V$.
N2: $\phi(x) = 0$ if and only if $x = 0$.

A *seminorm* is a function that satisfies N0 and N1 but not necessarily N2. We write

$$B(V, \phi) = \{x \in V \;:\; \phi(x) \leq 1\}$$
$$S(V, \phi) = \{x \in V \;:\; \phi(x) = 1\}$$
$$OB(V, \phi) = \{x \in V \;:\; \phi(x) < 1\}.$$

We also write $d_\phi(x, y) = \phi(x - y)$. If ϕ is clear from the context, we just write $B(V)$, $S(V)$ and $OB(V)$ for the spaces described above, and we may also write $\|x\|$ rather than $\phi(x)$.

It is clear that Definition 2.48 and Example 3.30 are instances of this construction. More generally, it is clear that for any seminorm ϕ the function d_ϕ is a semimetric, and that d_ϕ is a metric iff ϕ is a norm. In this context there is a useful criterion for continuity of linear maps, as follows:

Proposition 3.32. [`prop-bounded-continuous`]
Let V and W be vector spaces equipped with norms, and let $f\colon V \to W$ be a linear map. Then f is continuous iff there is a constant K such that $\|f(x)\| \leq K$ for all $x \in B(V)$.

Proof. Suppose there exists K as described. For any $x \in V \setminus \{0\}$ we put $u = x/\|x\|$ so $\|u\| = 1$, so $\|f(u)\| \leq K$. We also have $\|f(u)\| = \|f(x)/\|x\|\| = \|f(x)\|/\|x\|$, so we see that $\|f(x)\| \leq K\|x\|$. This also holds trivially when $x = 0$. We thus have

$$d(f(x), f(y)) = \|f(x) - f(y)\| = \|f(x - y)\| \leq K\|x - y\| = K\,d(x, y),$$

so f is Lipschitz and therefore continuous.

Conversely, suppose that f is continuous. In the standard metric criterion for continuity (Proposition 3.9) we can take $x = 0$ and $\epsilon = 1$; we conclude that there exists $\delta > 0$ such that $\|f(x)\| < 1$ whenever $\|x\| < \delta$. It follows that the number $K = 1/\delta$ has the required property. □

We will see in Corollary 10.37 that if V is finite-dimensional, then any linear map $f\colon V \to W$ is continuous.

Definition 3.33. [`defn-operator-norm`]
Let $\mathbb{K}$ be $\mathbb{R}$ or $\mathbb{C}$, and let V and W be normed vector spaces over $\mathbb{K}$. We write $\mathrm{Hom}^c(V, W)$ or $\mathrm{Hom}^c_{\mathbb{K}}(V, W)$ for the set of continuous linear maps from V to W. For $f \in \mathrm{Hom}^c(V, W)$ we put

$$\|f\|_{\mathrm{op}} = \sup\{\|f(v)\| \;:\; v \in V, \|v\| \leq 1\},$$

noting that this is finite by Proposition 3.32. We call this the *operator norm* of f. It is clear that $\|f(v)\| \leq \|f_{\mathrm{op}}\|\|v\|$ for all $v \in V$. Where there is no danger of confusion, we will write $\|f\|$ rather than $\|f\|_{\mathrm{op}}$. We will also write $V^* = \mathrm{Hom}^c(V, \mathbb{K})$, and call this the *continuous dual* of V.

Proposition 3.34. [prop-operator-norm]
$\mathrm{Hom}^c(V,W)$ is a vector space, and the operator norm is a norm. Moreover, for continuous linear maps $U \xrightarrow{g} V \xrightarrow{f} W$ we have $\|fg\|_{op} \leq \|f\|_{op}\|g\|_{op}$.

Proof. Suppose that $f,g \in \mathrm{Hom}^c(V,W)$. It is clear that the map $h = f+g$ is linear, and for $v \in V$ we have

$$\|h(v)\| = \|f(v)+g(v)\| \leq \|f(v)\| + \|g(v)\| \leq \|f\|_{\mathrm{op}}\|v\| + \|g\|_{\mathrm{op}}\|v\| = (\|f\|_{\mathrm{op}} + \|g\|_{\mathrm{op}})\|v\|.$$

This proves that $f+g \in \mathrm{Hom}^c(V,W)$ with $\|f+g\|_{\mathrm{op}} \leq \|f\|_{\mathrm{op}} + \|g\|_{\mathrm{op}}$. Similarly, for $t \in \mathbb{K}$ we have $tf \in \mathrm{Hom}^c(V,W)$ with $\|tf\|_{\mathrm{op}} = |t|\|f\|_{\mathrm{op}}$. This proves that $\mathrm{Hom}^c(V,W)$ is a vector space and that the operator norm is at least a seminorm. It is also clear that $\|f\|_{\mathrm{op}}$ can only vanish if $f=0$, so in fact we have a norm.

Now suppose we have continuous linear maps $U \xrightarrow{g} V \xrightarrow{f} W$. For $u \in U$ we have

$$\|fg(u)\| \leq \|f\|_{\mathrm{op}}\|g(u)\| \leq \|f\|_{\mathrm{op}}\|g\|_{\mathrm{op}}\|u\|,$$

so $\|fg\|_{\mathrm{op}} \leq \|f\|_{\mathrm{op}}\|g\|_{\mathrm{op}}$ as claimed. □

It is sometimes useful to define a topology using not just a single metric, but a whole family of them.

Definition 3.35. [defn-metrics-topology]
Let $D = (d_i)_{i\in I}$ be a family of semimetrics on X. Put

$$U(i,\epsilon,x) = \{y \in X \;:\; d_i(x,y) < \epsilon\}.$$

More generally, for any finite subset $J \subseteq I$ we put

$$U(J,\epsilon,x) = \{y \in X \;:\; d_j(x,y) < \epsilon \text{ for all } j \in J\} = \bigcap_{j\in J} U(j,\epsilon,x).$$

Note that $U(\{i\},\epsilon,x) = U(i,\epsilon,x)$ and

$$x \in U(J\cup K, \min(\epsilon,\delta), x) \subseteq U(J,\epsilon,x) \cap U(K,\delta,x).$$

Now put

$$\begin{aligned}
\sigma_D &= \{U(i,\epsilon,x) \;:\; i \in I,\ \epsilon > 0,\ x \in X\} \\
\beta_D &= \{U(J,\epsilon,x) \;:\; J \subseteq I \text{ finite },\ \epsilon > 0,\ x \in X\}.
\end{aligned}$$

We let τ_D be the topology with subbasis σ_D, and note that β_D is a basis for this topology.

Proposition 3.36. [prop-metrics-topology]
Let $D = (d_i)_{i\in I}$ be a family of semimetrics as above, and suppose that the index set I is countable. Then there is a single semimetric d on X that defines the same topology as D.

Proof. We will assume that I is infinite; the finite case is similar but easier and is left to the reader. We can then choose a bijection from $\mathbb{N}$ to I, so it will be harmless to assume that $I = \mathbb{N}$. We then put $d(x,y) = \sum_n \min(2^{-n}, d_n(x,y))$, and note that this is a convergent sum, with $d(x,y) \leq 2$. It is clear that $d(x,y) = d(y,x) \geq 0$ and $d(x,x) = 0$. For any three points x,y,z Proposition 2.44 tells us that

$$\min(2^{-n}, d_n(x,z)) \leq \min(2^{-n}, d_n(x,y)) + \min(2^{-n}, d_n(y,z))$$

and by adding these inequalities we see that $d(x,z) \leq d(x,y) + d(y,z)$. Thus, d is a semimetric. (It is also clear that $d(x,y) = 0$ iff $d_n(x,y) = 0$ for all n, so quite often d will be a metric even if the d_n individually are only semimetrics.)

Suppose that V is d-open, and that $x \in V$. We can then find $\delta > 0$ such that $y \in V$ whenever $d(x,y) < \delta$. Choose $m > 0$ such that $2^{-m} < \delta$. Put $J = \{0,1,\dots,m\}$ and $\epsilon = (\delta - 2^{-m})/(m+1)$. If $d_j(x,y) < \epsilon$ for all

$j \in J$ we have

$$
\begin{aligned}
d(x,y) &= \sum_{j=0}^{\infty} \min(2^{-j}, d_j(x,y)) \\
&\leq \sum_{j=0}^{m} d_j(x,y) + \sum_{j=m+1}^{\infty} 2^{-j} \\
&< \left(\sum_{j=0}^{m} \frac{\delta - 2^{-m}}{m+1} \right) + 2^{-m} = \delta,
\end{aligned}
$$

so $y \in V$. This proves that $U(J,\epsilon,x) \subseteq V$, and we conclude that V is D-open.

Conversely, suppose that V is D-open, and that $x \in U$. We can thus find a finite set $J \subset \mathbb{N}$ and a number $\delta > 0$ such that $U(J,\epsilon,x) \subseteq V$. Let m be the largest element of J and put $\epsilon = \min(\delta, 2^{-m})$. Suppose that $d(x,y) < \epsilon$. For each $j \in J$ we have $\min(2^{-j}, d_j(x,y)) \leq d(x,y) < \epsilon$ but $2^{-j} \geq 2^{-m} \geq \epsilon$ so we must have $d_j(x,y) < \epsilon$. As this holds for all $j \in J$, we have $y \in V$. This proves that $OB_\epsilon(x) \subseteq V$, and we conclude that V is d-open, as required. □

The definition of τ_D can be simplified if the semimetrics in D are all comparable with each other, as shown by the following result.

Lemma 3.37. [`lem-comparable-semimetrics`]
Suppose we have a sequence of semimetrics $d_0, d_1, \dots$ on X such that for all x and y we have

$$d_0(x,y) \leq d_1(x,y) \leq d_2(x,y) \leq \cdots .$$

Then the subbasis $\sigma_D = \{U(n,\epsilon,x) \;:\; n \in \mathbb{N},\ \epsilon > 0,\ x \in X\}$ for τ_D is actually a basis.

Proof. The inequalities $d_{i+1} \geq d_i$ imply that $U(J,\epsilon,x) = U(\max(J),\epsilon,x)$. □

Example 3.38. [`eg-locally-uniform`]
For each $n \in \mathbb{N}$ we can define a semimetric d_n on $C(\mathbb{R})$ by

$$d_n(f,g) = \sup\{|f(x) - g(x)| \;:\; x \in [-n,n]\}.$$

The topology determined by this family of semimetrics is called the *topology of locally uniform convergence* on $C(\mathbb{R})$. It is clear here that $d_i(x,y) \leq d_{i+1}(x,y)$, so Lemma 3.37 is applicable.

Example 3.39. [`eg-smooth-space`]
Recall that a function $f\colon \mathbb{R} \to \mathbb{R}$ is said to be *smooth* (or *infinitely differentiable*) if there are continuous functions $f^{(k)}(x)$ for all $k \geq 0$ such that $f^{(0)} = f$ and each $f^{(k)}$ is differentiable with derivative $f^{(k+1)}$. We write $C^\infty(\mathbb{R})$ for the set of all smooth functions. For $n, m \in \mathbb{N}$ we put

$$\|f\|_{n,m} = \sup\{|f^{(n)}(x)| \;:\; x \in [-m,m]\},$$

and $d_{n,m}(f,g) = \|f - g\|_{n,m}$. This family of semimetrics gives a topology τ on $C^\infty(\mathbb{R})$, called the *topology of locally uniform convergence of all derivatives.*

Example 3.40. [`eg-holomorphic-metric`]
Put $U = \{z \in \mathbb{C} \;:\; |z| < 1\}$, and let $H(U)$ denote the space of holomorphic functions on U. For any r with $0 < r < 1$ we put

$$\|f\|_r = \sup\{|f(z)| \;:\; |z| = r\} = \sup\{|f(z)| \;:\; |z| \leq r\}.$$

(The two definitions are the same by the Maximum Modulus Principle from complex analysis.) The second description shows that if $\|f\|_r = 0$ then $f = 0$ on the disc $\{z \;:\; |z| < r\}$ and so all derivatives of f vanish on that disc, so the Taylor series of f is zero, so $f = 0$. Given this, it is not hard to see that the function $d_r(f,g) = \|f - g\|_r$ gives a metric on $H(U)$, and that $d_r(f,g) \leq d_s(f,g)$ whenever $0 < r < s < 1$. By considering the functions $f_n(z) = z^n$, we see that $\|f\|_s / \|f\|_r$ can be arbitrarily large, so we have a family of inequivalent metrics. The most useful topology on $H(U)$ is the one determined by the whole family.

3.2. Operators. Let A and B be spaces of functions of some kind. Continuous maps $F\colon A \to B$ are often called *operators*, and continuous functions $G\colon A \to \mathbb{R}$ are often called *functionals*. We now discuss some examples.

EXAMPLE 3.41. Let $C([0,1])$ be the space of continuous functions $f\colon [0,1] \to \mathbb{R}$, with the metric d_∞ as in Example 3.30. We can define maps $I, R_n, T_n\colon C([0,1]) \to \mathbb{R}$ (for $n > 0$) by

$$I(f) = \int_0^1 f(x)\,dx$$
$$R_n(f) = \frac{1}{n}\sum_{k=1}^{n} f\left(\frac{2k-1}{2n}\right)$$
$$T_n(f) = \frac{1}{n}\left(\frac{1}{2}f(0) + \sum_{k=1}^{n-1} f(k/n) + \frac{1}{2}f(1)\right)$$

(so $R_n(f)$ and $T_n(f)$ are the approximations to $I(f)$ given by the rectangle rule and the trapezium rule, respectively). We claim that these maps are all Lipschitz (with constant one) and thus continuous. To see this, consider a pair of functions $f, g \in C([0,1])$, and put $h = f - g$ and , so $r = \|h\|_\infty = d(f,g) \in \mathbb{R}$. In the case of I, we have

$$|I(f) - I(g)| = |I(h)| = \left|\int_0^1 h(x)\,dx\right| \le \int_0^1 |h(x)|\,dx \le \int_0^1 r\,dx = r = d(f,g),$$

as required. In the case of R_n, the real point is again to prove that $|R_n(h)| \le r$. For this we note that $|h((2k-1)/(2n))| \le r$ for all k by the definition of r, so

$$|R_n(h)| = n^{-1}\left|\sum_{k=1}^{n} h((2k-1)/(2n))\right| \le n^{-1}\sum_{k=1}^{n} |h((2k-1)/(2n))| \le n^{-1}\sum_{k=1}^{n} r = r,$$

as required. The proof for T_n is similar.

EXAMPLE 3.42. [eg-integral-continuous]
We next define $J\colon C([0,r]) \to C([0,r])$ by

$$J(f)(x) = \int_{t=0}^{x} f(t)\,dt.$$

If $d(f,g) \le \epsilon$ we have

$$|J(f)(x) - J(g)(x)| = \left|\int_{t=0}^{x} f(t) - g(t)\,dt\right| \le \int_{t=0}^{x} |f(t) - g(t)|\,dt \le \int_{t=0}^{x} \epsilon = x\epsilon \le r\epsilon.$$

It follows that J is Lipschitz (with Lipschitz constant r) and therefore continuous. This is the simplest possible example of an *integral operator*; these are important in the general theory of differential equations.

EXAMPLE 3.43. [eg-integral-continuous-ii]
Now consider instead the space $C(\mathbb{R})$, and define $J\colon C(\mathbb{R}) \to C(\mathbb{R})$ by the same formula as above. If we give $C(\mathbb{R})$ the topology of uniform convergence (corresponding to the metric d_∞), then J is not continuous. To see this, note that the sequence of constant functions $1/n$ converges to zero, but $J(1/n)(x) = x/n$ (which is unbounded) so $d(J(1/n), J(0)) = \infty$, so $J(1/n) \not\to J(0)$.

However, we claim that J is continuous with respect to the topology of locally uniform convergence, as described in Example 3.38. Recall that this is determined by the increasing family of semimetrics d_n, where

$$d_n(f,g) = \sup\{|f(x) - g(x)| \;:\; x \in [-n,n]\}.$$

As in the previous example, we have $d_n(J(f), J(g)) \le n\,d_n(f,g)$.

Consider an open set $U \subseteq C(\mathbb{R})$, and an element $f \in J^{-1}(U)$. We then have $J(f) \in U$ so there exists $n > 0$ and $\epsilon > 0$ such that $p \in U$ whenever $d_n(J(f), p) < \epsilon$. If $d_n(f,g) < \epsilon/n$ we find that $d_n(J(f), J(g)) < \epsilon$ so $J(g) \in U$ so $g \in J^{-1}(U)$. This shows that $J^{-1}(U)$ is open, as required.

Example 3.44. [eg-shift-operator]

Fix a constant $a \in \mathbb{R}$. We can define a shift operator $S_a \colon C(\mathbb{R}) \to C(\mathbb{R})$ by $(S_a(f))(x) = f(x-a)$. For example, we have $S_{\pi/2}(\cos) = \sin$ and $S_\pi(\cos) = -\cos$ and $S_1(\exp) = e^{-1}.\exp$. It is clear that S_a is bijective, with inverse S_{-a}. We also have $d_\infty(S_a(f), S_a(g)) = d_\infty(f,g)$, so S_a is an isometric isomorphism from $C(\mathbb{R})$ to itself. This means that it is a homeomorphism with respect to the topology of uniform convergence. One can also check that it is a homeomorphism with respect to the topology of locally uniform convergence, but we will leave this as an exercise for the reader.

Example 3.45. [eg-evaluation]

Consider a space X and a point $a \in X$. We can define an *evaluation functional* $\mathrm{ev}_a \colon C(X) \to \mathbb{R}$ by $\mathrm{ev}_a(f) = f(a)$. For example, consider the space $X = [0, 2\pi]$ and the function $f(x) = x^2/6$. We then have

$$\begin{aligned} \mathrm{ev}_\pi(\sin) &= 0 \\ \mathrm{ev}_\pi(\cos) &= -1 \\ \mathrm{ev}_\pi(f) &= \pi^2/6. \end{aligned}$$

It is clear that

$$|\,\mathrm{ev}_a(f) - \mathrm{ev}_a(g)| = |f(a) - g(a)| \leq \sup\{|f(x) - g(x)| \;:\; x \in X\} = d(f,g).$$

This means that 1 is a Lipschitz constant for ev_a, so ev_a is continuous.

Example 3.46. [eg-diff-discts]

Let $C^1([0,1])$ denote the set of differentiable functions from $[0,1]$ to $\mathbb{R}$. We can use the metric $d_\infty(f,g) = \sup\{|f(x) - g(x)| \;:\; x \in [0,1]\}$ to make this a topological space, just as with $C([0,1])$. We can then define $D \colon C^1([0,1]) \to C([0,1])$ by $D(f) = f'$. However, this is not continuous. To see this, put $f_n(x) = \sin(2\pi n^2 x)/n$, so $f_n'(x) = 2\pi n \cos(2\pi n^2 x)$. We find that $d(f_n, 0) = 1/n$ but $d(D(f_n), 0) = 2\pi n$, so $f_n \to 0$ but $D(f_n) \not\to D(0)$.

Example 3.47. [eg-diff-cts]

Now consider instead the space $C^\infty(\mathbb{R})$ with the topology discussed in Example 3.39. We can again define $D \colon C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R})$ by $D(f) = f'$. We claim that in this context, D is continuous. Indeed, by Proposition 3.8 it will suffice to check that the preimage under D of the subbasic open set

$$U((n,m), \epsilon, f) = \{g \;:\; d_{n,m}(f,g) < \epsilon\}$$

is open. Suppose that $p \in D^{-1}(U((n,m),\epsilon,f))$, so the number $\delta = \epsilon - d_{n,m}(p', f)$ is strictly positive. Note that by definition we have $d_{n,m}(q', p') = d_{n+1,m}(q,p)$. Thus, if $d_{n+1,m}(q,p) < \delta$ then

$$d_{n,m}(q', f) \leq d_{n,m}(q', p') + d_{n,m}(p', f) = d_{n+1,m}(q,p) + (\epsilon - \delta) < \delta + (\epsilon - \delta) = \epsilon.$$

This proves that $U((n+1,m), p, \delta) \subseteq D^{-1}(U((n,m),\epsilon,f))$, and we conclude that $D^{-1}(U((n,m),\epsilon,f))$ is open as required.

We now discuss another context in which differentiation is continuous.

Example 3.48. [eg-holomorphic-diff]

Let $H(U)$ be the set of functions that are holomorphic on the open unit disc in $\mathbb{C}$, topologised using the family of metrics

$$d_r(f,g) = \|f - g\|_r = \sup\{|f(z) - g(z)| \;:\; |z| \leq r\}$$

(for $0 < r < 1$) as in Example 3.40. We claim that $D \colon H(U) \to H(U)$ is continuous. The key point is as follows: suppose we have $0 < r < s < 1$ and $|z| \leq r$. Standard complex analysis then gives the integral formula

$$f'(z) = \frac{1}{2\pi i} \oint_{|w|=s} \frac{f(w)}{(z-w)^2} dw.$$

The length of the contour where $|w| = s$ is $2\pi s$. For z on this contour we have $|f(w)| \leq \|f\|_s$ and $|w - z| \geq s - r$, so $1/|w-z|^2 \leq 1/(s-r)^2$. From this we deduce that

$$|f'(z)| \leq \frac{1}{2\pi} 2\pi s \frac{\|f\|_s}{(s-r)^2} = \frac{s}{(s-r)^2} \|f\|_s.$$

As z was an arbitrary point with $|z| \leq r$, this means that $\|f'\|_r \leq s(s-r)^{-2}\|f\|_s$.

Now suppose we have an open set $V \subseteq H(U)$, and an element $f \in D^{-1}(V)$. As V is open there exists r and ϵ such that $p \in V$ whenever $\|p - f'\|_r < \epsilon$. Choose any s with $r < s < 1$, and put $\delta = (s-r)^2\epsilon/s$. If $\|g - f\|_s < \delta$ we deduce that $\|g' - f'\|_r < \epsilon$ and so $g' \in V$, so $g \in D^{-1}(V)$. This proves that $D^{-1}(V)$ is open, as required.

4. Other Properties of Maps

Definition 4.1. [`defn-more-maps`]

Let $f\colon X \to Y$ be a map between topological spaces.

(a) We recall for convenience of comparison that f is said to be *continuous* if for every open set $V \subseteq Y$, the preimage $f^{-1}(V)$ is open in X.
(b) We say that f is an *embedding* if it is injective and continuous, and every open set $U \subseteq X$ has the form $U = f^{-1}(V)$ for some open subset $V \subseteq Y$.
(c) We say that f is a *quotient map* if it is surjective and continuous, and whenever $V \subseteq Y$ and $f^{-1}(V)$ is open in X, the set V is itself open in Y.
(d) We say that f is *proper* if it is continuous, and for any compact subset $K \subseteq Y$, the preimage $f^{-1}(K)$ is compact. (Compactness will not be defined until Section 10, but we record the definition of properness here to keep it together with other similar properties of maps.)
(e) We say that f is *open* if for every open set $U \subseteq X$, the image $f(U)$ is open in Y.
(f) We say that f is *closed* if for every closed set $F \subseteq X$, the image $f(F)$ is closed in Y.

Remark 4.2. [`rem-more-maps-complement`]

As open sets are precisely the complements of closed sets, and preimages of complements are complements of preimages, we see that (a) to (c) can be reformulated as follows:

(a') f is continuous if and only if for every closed set $G \subseteq Y$ the preimage $f^{-1}(G)$ is closed in X. (This was already discussed in Remark 3.4.)
(b') f is an embedding if it is injective and continuous, and every closed set $F \subseteq X$ has the form $F = f^{-1}(G)$ for some closed subset $G \subseteq Y$.
(c') f is a quotient map if and only if it is surjective and continuous, and whenever $G \subseteq Y$ and $f^{-1}(G)$ is closed in X, the set G is itself closed in Y.

We cannot reformulate (e) and (f) in the same way, because the image of the complement need not be the same as the complement of the image.

Proposition 4.3. [`prop-more-maps`]

Every homeomorphism has all the above properties. Moreover, if we have maps $X \xrightarrow{f} Y \xrightarrow{g} Z$ which both have one of the above properties, then $gf\colon X \to Z$ also has the same property.

Remark 4.4. This means that we have a wide subcategory of **Spaces** whose morphisms are all the embeddings, and similarly for all the other classes of maps.

Proof. This is clear from the definitions. □

We can generalise this slightly as follows.

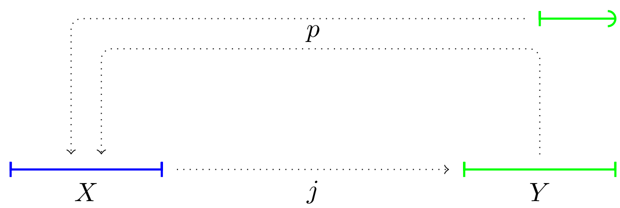
Proposition 4.5. [`prop-retraction`]

Suppose we have continuous maps $X \xrightarrow{j} Y \xrightarrow{p} X$ with $pj = 1_X$ (but not necessarily $jp = 1_Y$). Then j is an embedding and p is a quotient map.

Proof. First, if $j(x) = j(x')$ then we can apply p to both sides and use $pj = 1_X$ to see that $x = x'$. Thus, j is injective. Similarly, if $x \in X$ then $x = p(y)$ for some y, namely $y = j(x)$. Thus p is surjective. Now suppose we have an open set $U \subseteq X$. As p is continuous, the set $V = p^{-1}(U)$ is open in Y. We also have $j^{-1}(V) = j^{-1}(p^{-1}(U)) = (pj)^{-1}(U) = 1_X^{-1}(U) = U$, so we see that $U = j^{-1}(V)$ for some open set V. This means that j is an embedding. Finally, suppose we have subset $U' \subseteq X$ such that the set $V' = p^{-1}(U')$ is open in Y. As j is continuous we deduce that $j^{-1}(V')$ is open in X. However, just as before we have $j^{-1}(V') = U'$, so U' is open in X. This proves that p is a quotient map as claimed. □

Example 4.6. [eg-retraction-not-open]
Consider the following example:

$$
\begin{aligned}
X &= [0,1] \times \{0\} \subset \mathbb{R}^2 & \qquad & j(t,0) = (t,0) \\
T &= [1/2,1) \times \{1\} \subset \mathbb{R}^2 & & \\
Y &= X \cup T & & p(t,0) = p(t,1) = (t,0).
\end{aligned}
$$



The conditions of Proposition 4.5 are satisfied, so j is an embedding and p is a retraction. Moreover, T is both open and closed in Y, but $p(T)$ is neither open nor closed in X. This shows that a quotient map need not be open or closed.

Proposition 4.7. [prop-open-embedding]
Let $f\colon X \to Y$ be an injective continuous map. If f is either open or closed, then f is an embedding.

Proof. First suppose that f is open. If $U \subseteq X$ is open then the set $V = f(U)$ is open in Y (because f is open) and $U = f^{-1}(V)$ (because f is injective). Thus f is an embedding. If we suppose instead that f is closed, then we can argue in the same way, using Remark 4.2(b') instead of Definition 4.1(b). □

We will show in Proposition 8.5 that any injective continuous map $f\colon \mathbb{R} \to \mathbb{R}$ is open, and thus is an embedding.

Proposition 4.8. [prop-quotient-map]
Let $f\colon X \to Y$ be a surjective continuous map. If f is either open or closed, then f is a quotient map.

Proof. First suppose that f is open. Let V be a subset of Y such that the preimage $U = f^{-1}(V)$ is open in X. As f is an open map, it follows that $f(U)$ is open in Y. As f is surjective, we have $f(U) = f(f^{-1}(V)) = V$, so V is open in Y. This proves that f is a quotient map. If we suppose instead that f is closed, then we can argue in the same way, using Remark 4.2(c') instead of Definition 4.1(c). □

Proposition 4.9. [prop-cts-bij]
Let $f\colon X \to Y$ be a bijective continuous map. If f is open, or closed, or an embedding, or a quotient map, then f is a homeomorphism (and therefore f^{-1} is continuous).

Proof. As f is bijective, there is an inverse function $g\colon Y \to X$, and we need to show that g is continuous. First suppose that f is an embedding. Let U be open in X. As f is an embedding, we have $U = f^{-1}(V)$ for some open set $V \subseteq Y$. As f and g are mutually inverse bijections, this can be rewritten as $V = f(U)$ or $V = g^{-1}(U)$, so $g^{-1}(U)$ is open in Y. As U was an arbitrary open subset of X, this shows that g is continuous. This completes the argument when f is an embedding. If f is open or closed then it is an embedding by Proposition 4.7, so that case is covered as well. Suppose instead that f is a quotient map. Consider again an open set $U \subseteq X$ and put $V = g^{-1}(U) = f(U) \subseteq Y$. Note that $f^{-1}(V) = U$, which is open. As f is a quotient map, this means that V is open. In other words, $g^{-1}(U)$ is open for all open $U \subseteq X$, so $g\colon Y \to X$ is continuous. □

Proposition 4.10. [prop-embedding-first]
Let $X \xrightarrow{f} Y \xrightarrow{g} Z$ be continuous maps, and suppose that $gf\colon X \to Z$ is an embedding. Then f is also an embedding.

Proof. The map f is continuous by hypothesis. If $f(x) = f(x')$ then certainly $gf(x) = gf(x')$, but gf is assumed to be injective, so $x = x'$. This shows that f is injective. Now suppose that $U \subseteq X$ is open. As gf is an embedding, there is an open set $W \subseteq Z$ such that $U = (gf)^{-1}(W) = f^{-1}(g^{-1}(W))$. Here the set $V = g^{-1}(W) \subseteq Y$ is open (because g is assumed continuous) and $U = f^{-1}(V)$. This means that f is an embedding as claimed. □

Remark 4.11. [`rem-hol-open`]
Let U be an open disc in $\mathbb{C}$, and let $f\colon U \to \mathbb{C}$ be a nonconstant holomorphic map; then f is open. A detailed proof can be found in almost any textbook on complex analysis; here we will just give a sketch. Let γ be a simple closed curve in U. For any $a \in \mathbb{C}$ such that $f \circ \gamma$ does not pass through a, let $n(\gamma, a)$ be the number of preimages of a inside γ counted with appropriate multiplicity. It is then a standard fact (known as the Argument Principle) that

$$n(\gamma, a) = \frac{1}{2\pi i} \oint_\gamma \frac{f'(z)}{f(z) - a}\, dz.$$

If we move a slightly then the integral will only change slightly, but the integral is constrained to be an integer so it cannot change at all. In particular, if $n(\gamma, a) > 0$ then $n(\gamma, b) > 0$ for b sufficiently close to a; this proves that f is open as claimed.

Remark 4.12. Further results showing that certain classes of maps are open include Corollary 5.30, Lemma 5.72, Theorem 15.8 and Remark 28.15.

5. Constructs

5.1. Subspaces.

Definition 5.1. [`defn-subspace-topology`]
Let τ be a topology on a set X, and let Y be a subset of X. We declare that a subset $V \subseteq Y$ is open in Y if and only if it has the form $V = U \cap Y$ for some set $U \subseteq X$ that is open in X. This is easily seen to give a topology on Y, which we call the *subspace topology*, and denote by $\tau|_Y$. By a *subspace* of X we mean a subset considered as a topological space using the subspace topology.

Lemma 5.2. [`lem-subspace-closed`]
A subset $G \subseteq Y$ is closed (for the subspace topology) if and only if it has the form $G = F \cap Y$ for some set $F \subseteq X$ that is closed in X.

Proof. If G is closed, it must have the form $G = Y \setminus V$, where V is open in Y. This means that $V = U \cap Y$ for some open set $U \subseteq X$. Put $F = X \setminus U$; this is closed in X, and we find that $G = F \cap Y$ as required. We leave it to the reader to check that this argument is reversible. □

Remark 5.3. [`rem-embedding`]
Let Y be a subset of X, and consider it as a topological space using the subspace topology. Write j (or j_Y or j_{YX}) for the inclusion map $Y \to X$. For any subset $A \subseteq X$, the preimage $j^{-1}(A)$ is just $A \cap Y$. It follows that j is an embedding (as in Definition 4.1(c)). We can also reformulate that definition as follows. Suppose we have a map $f\colon X \to Y$ of topological spaces, and we put $X' = f(X) \subseteq Y$, considered as a topological space using the subspace topology. We can then regard f as a function $f\colon X \to X'$, and $f\colon X \to Y$ is an embedding if and only if $f\colon X \to X'$ is a homeomorphism. In more pernickety notation, we could introduce a new symbol f' for the restricted map $X \to X'$, and then we would have $f = j \circ f'$, and we would say that f is an embedding if and only if f' is a homeomorphism.

Proposition 5.4. [`prop-maps-to-subspace`]
Let X be a topological space, and let Y be a subspace of X. Let Z be another topological space, and let f be a map from Z to Y. Then f is continuous (with respect to the subspace topology on Y) if and only if the map $j \circ f\colon Z \to X$ is continuous. Thus, continuous maps from Z to Y biject with continuous maps from Z to X with image contained in Y.

Proof. By definition, f is continuous if and only if $f^{-1}(V)$ is open in Z for every set $V \subseteq Y$ that is open in the subspace topology. The relevant sets V are precisely those of the form $U \cap Y = j^{-1}(U)$ for some set U that is open in X, and we have $f^{-1}(j^{-1}(U)) = (j \circ f)^{-1}(U)$. Thus, f is continuous if and only if $(j \circ f)^{-1}(U)$ is open for all open sets $U \subseteq X$, and this is equivalent to the continuity of $j \circ f$. The rest of the statement follows directly from this. $\square$

PROPOSITION 5.5. [prop-subspace-basis]
If β is a basis for θ then the set

$$\beta|_Y = \{V \cap Y \,:\, V \in \beta\}$$

is a basis for $\theta|_Y$. Moreover, the corresponding result also holds for subbases.

PROOF. First suppose that $U \in \tau(\beta|_Y)$. This means that for each $y \in U$ we can choose $U_y \in \beta|_Y$ with $y \in U_y \subseteq U$. As $U_y \in \beta|_Y$ we can choose $V_y \in \beta$ with $U_y = V_y \cap Y$. Let V be the union of all the sets V_y, so $V \in \tau(\beta) = \theta$. We also find that $V \cap Y$ is the union of all the sets $V_y \cap Y = U_y$, or in other words $V \cap Y = U$, which means that $U \in \theta|_Y$. As U was an arbitrary set in $\tau(\beta|_Y)$, we conclude that $\tau(\beta|_Y) \subseteq \theta|_Y$.

In the opposite direction, suppose that $U \in \theta|_Y$. This means that we can choose $V \in \tau$ with $U = V \cap Y$. If $y \in U$ then certainly $y \in V$, and $V \in \theta = \tau(\beta)$, so we can find $V_y \in \beta$ with $y \in V_y \subseteq V$. We now put $U_y = V_y \cap Y \in \beta|_Y$, and observe that $y \in U_y \subseteq V \cap Y = U$. As y was an arbitrary element of U, it follows that $U \in \tau(\beta|_Y)$. We conclude that $\theta|_Y \subseteq \tau(\beta|_Y)$, which completes the proof of the statement for bases. The statement for subbases is similar, and is left to the reader. $\square$

PROPOSITION 5.6. [prop-subspace-metric]
Let X be a set equipped with a metric d and thus an associated topology τ_d. Let Y be a subset of X. We can then restrict the metric to get a metric d' on Y, and form the associated topology $\tau_{d'}$ on Y; or we can use the subspace topology $\tau_d|_Y$. These are the same topology.

PROOF. We will write OB and OB' for open balls with respect to d and d', so $OB'_\epsilon(y) = OB_\epsilon(y) \cap Y$ for all $y \in Y$ and $\epsilon > 0$.

Suppose that $V \in \tau_{d'}$. For each $y \in V$ there exists $\epsilon_y > 0$ such that $OB'_{\epsilon_y}(y) \subseteq V$. If we put $U = \bigcup_{y \in Y} OB_{\epsilon_y}(y)$ we find that $U \in \tau_d$ and $U \cap Y = V$ so $V \in \tau_d|_Y$.

Conversely, suppose that $V \in \tau_d|_Y$, so there is a set $U \in \tau_d$ such that $V = U \cap Y$. Now for any $y \in Y$ there exists $\epsilon > 0$ such that $OB_\epsilon(y) \subseteq U$, but this implies that $OB'_\epsilon(y) \subseteq V$. It follows that $V \in \tau_{d'}$. $\square$

PROPOSITION 5.7. [prop-rel-open]

(a) *Suppose that Y is open in X. Then the open subsets for the subspace topology are precisely the sets that are open in X and contained in Y.*
(b) *Suppose that Y is closed in X. Then the closed subsets for the subspace topology are precisely the sets that are closed in X and contained in Y.*

PROOF.

(a) Suppose that Y is open in X. The open sets for the subspace topology are precisely the sets $Y \cap U$, where U is open in X. In this case it is clear that $Y \cap U$ is open in X and contained in Y. Conversely, suppose that V is open in X and contained in Y. From the definition of the subspace topology we see that $V \cap Y$ is open in Y, but $V \cap Y$ is just the same as V.
(b) This is essentially the same as (a).

$\square$

PROPOSITION 5.8. [prop-subspace-closure]
If $Z \subseteq Y$ then we have $\mathrm{cl}_Y(Z) = \mathrm{cl}_X(Z) \cap Y$.

PROOF. As $\mathrm{cl}_X(Z)$ is closed in X, we see from Lemma 5.2 that $\mathrm{cl}_X(Z) \cap Y$ is closed in Y. This set also contains Z, and $\mathrm{cl}_Y(Z)$ is by definition the smallest set that is closed in Y and contains Z, so we see that $\mathrm{cl}_Y(Z) \subseteq \mathrm{cl}_X(Z) \cap Y$. On the other hand, $\mathrm{cl}_Y(Z)$ is closed in the subspace topology on Y, so it has the form $\mathrm{cl}_Y(Z) = F \cap Y$ for some set F that is closed in X. This means that $Z \subseteq \mathrm{cl}_Y(Z) \subseteq F$, and $\mathrm{cl}_X(Z)$ is the smallest set that is closed in X and contains Z. It follows that $\mathrm{cl}_X(Z) \subseteq F$ and so $\mathrm{cl}_X(Z) \cap Y \subseteq F \cap Y = \mathrm{cl}_Y(Z)$. This now proves the claim. $\square$

Proposition 5.9 (Patching). [prop-patching]
Let $f\colon X \to Y$ be a function between topological spaces. Suppose that X is the union of some family of subsets $(X_i)_{i\in I}$, and that the restriction $f_i = f|_{X_i}\colon X_i \to Y$ is continuous (with respect to the subspace topology on X_i and the original topology on Y) for all i.

(a) *If the sets X_i are all open in X, then f is continuous.*
(b) *If the sets X_i are all closed, and the index set I is finite, then f is again continuous.*

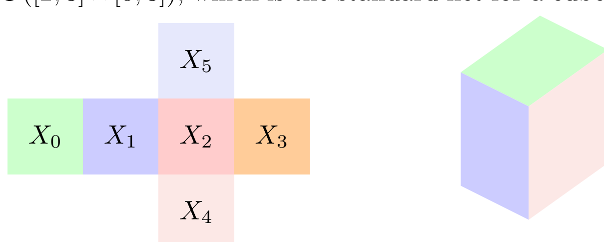
This result is used extensively to define continuous functions in homotopy theory, as we will start to see in Section 27. It is also often used in complex algebraic geometry, where many functions are initially given by different formulae over different open subsets of projective space.

Proof.

(a) Suppose that the sets X_i are open in X. Consider an open set $V \subseteq Y$. As f_i is continuous, we see that $f_i^{-1}(V)$ is open in X_i with respect to the subspace topology. By Proposition 5.7(a), this means that $f_i^{-1}(V)$ is open in X. On the other hand, it is clear that $f_i^{-1}(V)$ is just the same as $f^{-1}(V)\cap X_i$ and $X = \bigcup_i X_i$ so $f^{-1}(V) = \bigcup_i f_i^{-1}(V)$. This means that $f^{-1}(V)$ is the union of a family of sets that are open in X, so $f^{-1}(V)$ is open in X. As V was an arbitrary open subset of Y, we conclude that f is continuous.
(b) Suppose that the sets X_i are closed in X, and that the index set I is finite. Consider a closed set $G \subseteq Y$. As f_i is continuous, we see (via Remark 3.4) that $f_i^{-1}(G)$ is closed in X_i with respect to the subspace topology. By Proposition 5.7(b), this means that $f_i^{-1}(G)$ is closed in X. On the other hand, it is clear that $f_i^{-1}(G)$ is just the same as $f^{-1}(G)\cap X_i$ and $X = \bigcup_i X_i$ so $f^{-1}(G) = \bigcup_i f_i^{-1}(G)$. This means that $f^{-1}(G)$ is the union of a finite family of sets that are closed in X, so $f^{-1}(G)$ is closed in X. As G was an arbitrary closed subset of Y, we conclude that f is continuous.

□

Example 5.10. [eg-closed-patching]
Put $X = ([0,4]\times[1,2]) \cup ([2,3]\times[0,3])$, which is the standard net for a cube, as shown on the left below.



We can write this as a union of six squares as follows:

$$X_0 = [0,1]\times[1,2] \qquad X_1 = [1,2]\times[1,2] \qquad X_2 = [2,3]\times[1,2]$$
$$X_3 = [3,4]\times[1,2] \qquad X_4 = [2,3]\times[0,1] \qquad X_5 = [2,3]\times[2,3].$$

We define maps $f_i\colon X_i \to \mathbb{R}^3$ by

$$\begin{aligned}
f_0(x,y) &= (x, 2-y, 1)\\
f_1(x,y) &= (1, 2-y, 2-x)\\
f_2(x,y) &= (3-x, 2-y, 0)\\
f_3(x,y) &= (0, 2-y, x-3)\\
f_4(x,y) &= (3-x, 1, 1-y)\\
f_5(x,y) &= (3-x, 0, y-2).
\end{aligned}$$

One can check that f_i and f_j agree on $X_i \cap X_j$ for all i and j. For example, we have $X_0 \cap X_1 = \{(1,y) : 1 \le y \le 2\}$ and $f_0(1,y) = (1, 2-y, 1) = f_1(1,y)$. Using part (b) of the proposition we see that the maps f_i can be patched together to give a continuous map $f\colon X \to \mathbb{R}^3$. The image of f is the surface of a cube, as on the right above.

5.2. Products.

Definition 5.11. [defn-product-set]
Suppose we have a family of sets $(X_i)_{i\in I}$. The *product set* is the set $X = \prod_{i\in I} X_i$ of all indexed families $(x_i)_{i\in I}$ where $x_i \in X_i$ for all i. We write π_i for the map $X \to X_i$ that sends a family $(x_i)_{i\in I}$ to the i'th entry x_i. We call this the *i'th projection map.*

Remark 5.12. [rem-maps-to-product]
Consider a map $f\colon W \to \prod_{i\in I} X_i$. We can compose with the projections $\pi_i\colon \prod_{i\in I} X_i \to X_i$ to get a family of maps $f_i = \pi_i \circ f\colon W \to X_i$. Conversely, given such a family of maps $f_i\colon W \to X_i$, we can define $f\colon W \to \prod_{i\in I} X_i$ by $f(w) = (f_i(w))_{i\in I}$. These constructions are visibly inverse to each other.

Example 5.13. [eg-product-sets]

(a) If $I = \{0, 1\}$ then we just have two sets X_0 and X_1, and $X = X_0 \times X_1$ is just the set of ordered pairs (x_0, x_1) with $x_0 \in X_0$ and $x_1 \in X_1$. The projection maps are $\pi_0(x_0, x_1) = x_0$ and $\pi_1(x_0, x_1) = x_1$. Any map $f\colon W \to X$ has the form $f(x) = (f_0(x), f_1(x))$ for some $f_0\colon W \to X_0$ and $f_1\colon W \to X_1$. We will use the abbreviated notation $f = (f_0, f_1)$.
(b) Similarly, if $I = \{0, 1, 2\}$ then X is the set of triples (x_0, x_1, x_2) with $x_0 \in X_0$ and $x_1 \in X_1$ and $x_2 \in X_2$.
(c) If all the sets X_i are the same (equal to Y, say) then $\prod_i X_i$ can be identified with the set of all functions from I to Y.
(d) If $I = \mathbb{N}$ and $X_n = [-n, n]$ then $\prod_n X_n$ is the set of all sequences $x = (x_0, x_1, \dots)$ for which $|x_n| \leq n$ for all n.

Definition 5.14. [defn-product-space]
Now suppose we have a family of topological spaces $(X_i)_{i\in I}$. We write τ_i for the topology on X_i, and then we put

$$\sigma = \{\pi_i^{-1}(U) \ : \ i \in I, U \in \tau_i\}.$$

This is a subbasis for a topology on $\prod_i X_i$, which we call the *product topology.* By the *product space* we mean the product set equipped with the product topology.

Remark 5.15. [rem-proj-cts]
It is immediate from the definitions that the projection maps π_i are continuous.

Proposition 5.16. [prop-maps-to-product]
Consider a topological space W and a map $f\colon W \to \prod_{i\in I} X_i$. Then f is continuous (with respect to the product topology) if and only if the maps $f_i = \pi_i \circ f\colon W \to X_i$ are all continuous. Thus, continuous maps $f\colon W \to \prod_{i\in I} X_i$ biject with families of continuous maps $f_i\colon W \to X_i$.

Proof. Proposition 3.8 tells us that f is continuous if and only if the preimages of all the subbasic open sets $\pi_i^{-1}(U)$ (for $i \in I$ and U open in X_i) are open in W. Now $f^{-1}(\pi_i^{-1}(U)) = f_i^{-1}(U)$, and f_i is continuous if and only if $f_i^{-1}(U)$ is open in W whenever U is open in X_i. It follows that f is continuous if and only if all the maps f_i are continuous. In combination with Remark 5.12 this gives a bijection between continuous maps $f\colon W \to \prod_{i\in I} X_i$ and families of maps $f_i\colon W \to X_i$, as claimed $\square$

Remark 5.17. [rem-product-categorical]
This proposition means that the product space is a product in the sense of category theory, as discussed in Appendix 36.

Corollary 5.18. [cor-product-map]
Suppose we have a family of continuous maps $f_i\colon X_i \to Y_i$, and we define

$$f = \prod_I f_i\colon \prod_I X_i \to \prod_I Y_i$$

by $f((x_i)_{i\in I}) = (f_i(x_i))_{i\in I}$. Then f is also continuous.

Proof. By the proposition, it will suffice to show that $\pi_j \circ f\colon \prod_i X_i \to Y_j$ is continuous for all j. However, $\pi_j \circ f$ is just the same as $f_j \circ \pi_j$, and f_j and π_j are continuous, so $f_j \circ \pi_j$ is continuous by Proposition 3.5. The maps considered are conveniently displayed in the following diagram:

$$\begin{tikzcd}
\prod_I X_i \arrow[r, "f"] \arrow[d, "\pi_j"'] & \prod_I Y_i \arrow[d, "\pi_j"] \\
X_j \arrow[r, "f_j"'] & Y_j.
\end{tikzcd}$$

□

Proposition 5.19. [prop-product-metric]

Let X_0 and X_1 be metric spaces. We can then form a product metric on $X_0 \times X_1$ as in Definition 2.50 and use the associated metric topology on $X_0 \times X_1$, or we can take the metric topologies on X_0 and X_1 and form the product topology on $X_0 \times X_1$. In fact, these two topologies are the same.

Proof. Let τ_m be the metric topology, and let τ_p be the product topology. For definiteness, we use the product metric given by

$$d((x_0, x_1), (y_0, y_1)) = \max(d(x_0, y_0), d(x_1, y_1)).$$

(As remarked in Definition 2.50, either of the other two variants would give the same topology.) With this convention we have

$$OB_\epsilon(x_0, x_1) = OB_\epsilon(x_0) \times OB_\epsilon(x_1) = \pi_0^{-1}(OB_\epsilon(x_0)) \cap \pi_1^{-1}(OB_\epsilon(x_1)),$$

which is open in τ_p. As these balls form a basis for τ_m, we see that $\tau_m \subseteq \tau_p$. On the other hand, consider an open set $U_0 \subseteq X_0$, which gives a subbasic open set $\pi_0^{-1}(U_0) = U_0 \times X_1$ for τ_p. If $(x_0, x_1) \in U_0 \times X_1$ then x_0 lies in the open set U_0 so we can find $\epsilon > 0$ such that $OB_\epsilon(x_0) \subseteq U_0$, and it follows easily that $OB_\epsilon(x_0, x_1) \subseteq U_0 \times X_1$. This means that the sets $U_0 \times X_1$ are open with respect to τ_m, as are all sets of the form $X_0 \times U_1$ (for U_1 open in X_1) by a symmetrical argument. These two families of sets give a subbasis for τ_p, so we conclude that $\tau_p \subseteq \tau_m$. □

Example 5.20. [eg-CX-ops]

In Corollary 3.11 we showed that if f and g are continuous maps from X to $\mathbb{R}$, then the maps $f+g$, fg, $\max(f,g)$ and $\min(f,g)$ are also continuous. The first step in the proof was to introduce the function $h = (f,g)\colon X \to \mathbb{R} \times \mathbb{R}$, and to check that this is continuous. This can now be seen as a special case of Proposition 5.16.

Example 5.21. [eg-matrix-maps]

Consider the map $f\colon \mathbb{R}^3 \to \mathbb{R}^3$ given by

$$f(x,y,z) = (x+y+z, xy+yz+zx, xyz).$$

We claim that this is continuous. Indeed, we have $f = (f_0, f_1, f_2)$, where

$$f_0(x,y,z) = x+y+z \qquad f_1(x,y,z) = xy+yz+zx \qquad f_2(x,y,z) = xyz,$$

so it will suffice to show that each map f_i is continuous. There are projection maps

$$\pi_0(x,y,z) = x \qquad \pi_1(x,y,z) = y \qquad \pi_2(x,y,z) = z$$

which are continuous by construction. We can describe f_0 as $\pi_0 + \pi_1 + \pi_2$, and Corollary 3.11 tells us that the sum of any two continuous real-valued functions is again continuous, so we see that f_0 is continuous. The same Corollary also tells us that the product of any two continuous real-valued functions is again continuous, and using this repeatedly we see that the maps $f_1 = \pi_0\pi_1 + \pi_1\pi_2 + \pi_2\pi_0$ and $f_2 = \pi_0\pi_1\pi_2$ are also continuous, as claimed.

By a more general argument along the same lines, we see that any multivariate polynomial map $\mathbb{R}^n \to \mathbb{R}^m$ is continuous.

We can get many more interesting examples by identifying $M_n(\mathbb{R})$ with $\mathbb{R}^{n^2}$. In particular, we find that the map $\det\colon M_n(\mathbb{R}) \to \mathbb{R}$ is polynomial and therefore continuous. The set $GL_n(\mathbb{R})$ of invertible matrices is the preimage under det of the open set $\mathbb{R}\setminus\{0\}$, so $GL_n(\mathbb{R})$ is open in $M_n(\mathbb{R})$. The set $SL_n(\mathbb{R})$ is the preimage under det of the closed set $\{1\}$, so $SL_n(\mathbb{R})$ is closed in $M_n(\mathbb{R})$. Similarly, we have a continuous map $f\colon M_n(\mathbb{R})\to M_n(\mathbb{R})$ given by $f(A)=A^TA$. The set $O(n)$ of orthogonal matrices is the preimage of the closed set $\{I\}$ under f, so $O(n)$ is closed in $M_n(\mathbb{R})$, as is the intersection $SO(n)=SL_n(\mathbb{R})\cap O(n)$. Similarly, we can define continuous maps as follows:

$$\begin{aligned} &g\colon M_n(\mathbb{R})\to M_n(\mathbb{R}) && g(A)=A^T-A\\ &h\colon M_n(\mathbb{R})\to M_n(\mathbb{R}) && h(A)=A^2-A\\ &k\colon M_n(\mathbb{R})\to \mathbb{R} && k(A)=\mathrm{trace}(A)-1. \end{aligned}$$

The space $\mathbb{R}P^{n-1}$ is defined as $g^{-1}\{0\}\cap h^{-1}\{0\}\cap k^{-1}\{0\}$; this is the intersection of three closed sets and so is again closed in $M_n(\mathbb{R})$.

Example 5.22. [`eg-matrix-inversion`]
Now consider the inversion map $\chi\colon GL_n(\mathbb{R})\to GL_n(\mathbb{R})$, given by $\chi(A)=A^{-1}$; we will show that this is continuous. A standard formula says that $A^{-1}=\mathrm{adj}(A)/\det(A)$, where $\mathrm{adj}(A)_{ij}$ is $(-1)^{ij}$ times the determinant of the matrix obtained by deleting the i'th column and j'th row from A. In other words, we can write χ as the composite

$$GL_n(\mathbb{R})\xrightarrow{(\det,\mathrm{adj})}\mathbb{R}^\times\times GL_n(\mathbb{R})\xrightarrow{\nu\times 1}\mathbb{R}^\times\times GL_n(\mathbb{R})\xrightarrow{\mathrm{mult}}GL_n(\mathbb{R}).$$

(Here $\mathbb{R}^\times$ denotes $\mathbb{R}\setminus\{0\}$, and ν is given by $x\mapsto 1/x$ as in Proposition 3.10.) We use the obvious subspace topologies on $\mathbb{R}^\times$ and $GL_n(\mathbb{R})$. Proposition 5.4 tells us that a map to $\mathbb{R}^\times$ is continuous if and only if it is continuous when regarded as a map to $\mathbb{R}$. Similarly, a map to $GL_n(\mathbb{R})$ is continuous if and only if it is continuous when regarded as a map to $M_n(\mathbb{R})\simeq\mathbb{R}^{n^2}$. Using this we see that the maps $\det\colon GL_n(\mathbb{R})\to\mathbb{R}^\times$ and $\mathrm{adj}\colon GL_n(\mathbb{R})\to GL_n(\mathbb{R})$ and $\mathrm{mult}\colon\mathbb{R}^\times\times M_n(\mathbb{R})\to M_n(\mathbb{R})$ are continuous. We also saw in Proposition 3.10 that ν is continuous. It follows using Proposition 5.16 and Corollary 5.18 that $(\det,\mathrm{adj})$ and $\nu\times 1$ are continuous, and so we can compose to see that χ is continuous as claimed.

Example 5.23. [`eg-GL-two`]
Consider the space

$$X=\mathbb{R}^+\times\mathbb{R}\times\mathbb{R}^\times\times S^1=\{(u,v,w,x,y)\in\mathbb{R}^5\ :\ u>0,\ w\neq 0,\ x^2+y^2=1\}.$$

We can define a map $f\colon X\to M_2(\mathbb{R})$ by

$$f(u,v,w,x,y)=u\begin{bmatrix}1&v\\0&1\end{bmatrix}\begin{bmatrix}w&0\\0&1\end{bmatrix}\begin{bmatrix}x&-y\\y&x\end{bmatrix}=\begin{bmatrix}uwx+uvy&uvx-uwy\\uy&ux\end{bmatrix}.$$

One can check that this is actually a homeomorphism $X\to GL_2(\mathbb{R})$, with inverse given by

$$f^{-1}\begin{bmatrix}a&b\\c&d\end{bmatrix}=\left(\sqrt{c^2+d^2},\frac{ac+bd}{c^2+d^2},\frac{ad-bc}{c^2+d^2},\frac{d}{\sqrt{c^2+d^2}},\frac{c}{\sqrt{c^2+d^2}}\right).$$

By restricting attention to the subspace where $u=w^2=1$ and $v=0$, we obtain a homeomorphism $g\colon S^0\times S^1=\{1,-1\}\times S^1\to O(2)$, given by

$$g(w,x,y)=\begin{bmatrix}wx&-wy\\y&x\end{bmatrix}\qquad\qquad g^{-1}\begin{bmatrix}a&b\\c&d\end{bmatrix}=(ad-bc,d,c).$$

We can also put

$$Y=\mathbb{R}^+\times\mathbb{R}^+\times\mathbb{R}^2\times S^1\times S^3=\{y\in\mathbb{R}^{10}\ :\ y_0,y_1>0,\ y_4^2+y_5^2=y_6^2+y_7^2+y_8^2+y_9^2=1\},$$

and we can define $h\colon Y\to M_2(\mathbb{C})$ by

$$h(y)=\begin{bmatrix}y_0&y_2+iy_3\\0&y_1\end{bmatrix}\begin{bmatrix}y_4+iy_5&0\\0&1\end{bmatrix}\begin{bmatrix}y_6+iy_7&y_8+iy_9\\-y_8+iy_9&y_6-iy_7\end{bmatrix}.$$

It can be shown that this is again a homeomorphism from Y to $GL_2(\mathbb{C})$, although in this case it is less easy to give a tidy formula for the inverse.

Example 5.24. [eg-Sn-RPn]
Recall that

$$S^n = \{x \in \mathbb{R}^{n+1} \;:\; \|x\| = 1\}$$
$$\mathbb{R}P^n = \{A \in M_{n+1}(\mathbb{R}) \;:\; A^T = A = A^2,\ \text{trace}(A) = 1\}.$$

Here we topologise $M_{n+1}(\mathbb{R})$ using the metric $d_2(P,Q) = \sqrt{\text{trace}((P-Q)(P-Q)^T)}$ as in Example 2.40; as explained there, this is the same metric as we get by identifying $M_{n+1}(\mathbb{R})$ with $\mathbb{R}^{(n+1)^2}$ in the obvious way.

We define a map $f\colon S^n \to M_{n+1}(\mathbb{R})$ by $f(x)_{ij} = x_i x_j$; this is now easily seen to be continuous. We claim that $f(x) \in \mathbb{R}P^n$. To prove this, it is convenient to regard elements of $\mathbb{R}^{n+1}$ as column vectors; we then have $\langle x, y\rangle = x^T y$ and $f(x) = xx^T$. From this we get $f(x)^T = x^{TT}x^T = f(x)$ and also

$$f(x)^2 = x(x^T x)x^T = x.\langle x, x\rangle.x^T = \|x\|^2 f(x) = f(x).$$

Moreover, we have $\text{trace}(f(x)) = \sum_i x_i^2 = \|x\|^2 = 1$. We have thus defined a continuous map $f\colon S^n \to \mathbb{R}P^n$. If y is another point in $\mathbb{R}^{n+1}$ we have

$$f(x)y = x(x^T y) = \langle x, y\rangle x.$$

Next, we claim that $f\colon S^n \to \mathbb{R}P^n$ is surjective. Indeed, suppose we have $B \in \mathbb{R}P^n$. As $\text{trace}(B) = 1$ we have $B \neq 0$, so we can find $w \in \mathbb{R}^{n+1}$ such that $Bw \neq 0$. Put $x = (Bw)/\|Bw\| \in S^n$ and $A = f(x) \in \mathbb{R}P^n$. It will be enough to show that $A = B$, or equivalently that $d_2(A,B) = 0$, or equivalently that the matrix $C = A - B$ satisfies $\text{trace}(CC^T) = 0$. For this, we first note that $Bx = (B^2 w)/\|Bw\| = (Bw)/\|Bw\| = x$. We saw in the previous paragraph that $Ay = \langle x, y\rangle x$, so $BAy = \langle x, y\rangle Bx = \langle x, y\rangle x = Ay$. This holds for all y, so $BA = A$. Taking transposes gives $A^T B^T = (BA)^T = A^T$, but $A^T = A$ and $B^T = B$ so we see that $AB = A$ also. It follows that

$$CC^T = (A-B)(A-B) = A^2 - AB - BA + B^2 = A - A - A + B = B - A,$$

so $\text{trace}(CC^T) = \text{trace}(B) - \text{trace}(A) = 1 - 1 = 0$ as required. (This argument is efficient, but may seem a little miraculous. As an alternative one can check that $\mathbb{R}^{n+1}$ splits orthogonally as $\text{img}(B) \oplus \ker(B)$, choose orthonormal bases adapted to this splitting, and proceed from there.)

We next observe that $f(-x) = f(x)$, so f is not injective. We claim that this is the only source of non-injectivity. Indeed, for any $x, y \in S^n$ we have

$$\text{trace}(f(x)f(y)) = \sum_{i,j} f(x)_{ij} f(y)_{ij} = \sum_{i,j} x_i y_i x_j y_j = \langle x, y\rangle^2,$$

and thus that

$$d_2(f(x), f(y)) = \text{trace}((f(x) - f(y))^2)^{1/2} = \text{trace}(f(x) - f(x)f(y) - f(y)f(x) + f(y))^{1/2} = \sqrt{2(1 - \langle x, y\rangle^2)}.$$

As x and y are unit vectors, we see from the Cauchy-Schwartz inequality that $1 - \langle x, y\rangle^2$ can only vanish if $y = \pm x$. It follows that we have $f(x) = f(y)$ iff $x = \pm y$.

Example 5.25. [eg-rref]
Now define a map $\phi\colon M_n(\mathbb{R}) \to M_n(\mathbb{R})$ by sending each matrix A to its row reduced echelon form. In other words, $\phi(A)$ is characterised by the following properties:

(a) The linear span of the rows of $\phi(A)$ is the same as the linear span of the rows of A.
(b) For each nonzero row in $\phi(A)$, the first nonzero entry (called the *pivot*) is equal to one, and lies to the right of any pivots in previous rows.
(c) All entries above a pivot are zero.
(d) Any zero rows come after all the nonzero rows.

We claim that the map ϕ is not continuous. Indeed, we have $\phi(0) = 0$, but $\phi(\epsilon I) = I$ for all $\epsilon \neq 0$. Consider the set $U = M_n(\mathbb{R}) \setminus \{I\}$, which is open. The preimage $\phi^{-1}(U)$ contains 0, but it does not contain $OB_\epsilon(0)$ for any $\epsilon > 0$. This means that $\phi^{-1}(U)$ is not open, so ϕ is not continuous.

We have defined the product topology using a subbasis, but for some purposes it is more convenient to have a basis. For this, we need a preliminary definition.

Definition 5.26. Suppose we have a family of spaces X_i indexed by a set I, and a subset $J \subseteq I$. We then write π^I_J or π_J for the map $\prod_I X_i \to \prod_J X_i$ that sends each family $(x_i)_{i\in I}$ to the subfamily $(x_j)_{j\in J}$. For example, if $I = \{0,1,2,3,4,5\}$ and $J = \{0,2,5\}$ we just have

$$\pi_{\{0,2,5\}}(x_0, x_1, x_2, x_3, x_4, x_5) = (x_0, x_2, x_5).$$

Lemma 5.27. [lem-proj-cts]
The maps π_J are continuous.

Proof. This is immediate from Proposition 5.16, as $\pi_j \circ \pi_J = \pi_j \colon \prod_I X_i \to X_j$ for all $j \in J$. □

Proposition 5.28. [prop-product-basis]
The following sets are bases for the product topology:

$$\beta = \{\pi_J^{-1}(U) \ : \ J \subseteq I \textit{ is finite and } U \textit{ is open in } \prod_J X_i\}$$

$$\beta' = \{\pi_J^{-1}\left(\prod_{j\in J} U_j\right) \ : \ J \subseteq I \textit{ is finite and } U_j \textit{ is open in } X_j \textit{ for all } j \in J\}.$$

Remark 5.29. [rem-bt-prime]
We can redescribe β' as the collection of all products $\prod_{i\in I} U_i$, where U_i is open in X_i for all i, and $U_i = X_i$ for all but finitely many indices i.

Proof. We write τ for the product topology on $\prod_I X_i$, and σ for the standard subbasis that generates τ. It is clear (by considering sets J of size one) that $\sigma \subseteq \beta'$. Next, suppose we have a finite subset $J \subseteq I$ and open sets $U_j \subseteq X_j$ for all $j \in J$. The subset $\prod_J U_j$ can then be expressed as the intersection of the finite family of open sets $\pi_j^{-1}(U_j)$, so $\prod_J U_j$ is open in $\prod_J X_j$. Given this, we see that $\beta' \subseteq \beta$. On the other hand, as π_J is continuous we see that $\beta \subseteq \tau$, so we now know that $\sigma \subseteq \beta' \subseteq \beta \subseteq \tau$. As τ is the smallest topology that contains σ, it is therefore also the smallest topology that contains β or β'. All that is now left is to check that β and β' are actually topological bases. From Remark 5.29 it is clear that the intersection of any two sets in β' is again in β', which implies that β' is a topological basis. Now suppose we have sets $\pi_J^{-1}(U)$ and $\pi_K^{-1}(V)$ in β. The set $L = J \cup K$ is then finite, and the set

$$W = (\pi^L_J)^{-1}(U) \cap (\pi^L_K)^{-1}(V)$$

is open in $\prod_K X_i$. One checks that $\pi_J^{-1}(U) \cap \pi_K^{-1}(V) = \pi_L^{-1}(W) \in \beta$. This means that the intersection of any two sets in β is again in β, so β is also a topological basis. □

Corollary 5.30. [cor-projections-open]
The projection maps $\pi_j \colon \prod_i X_i \to X_j$ are open (but not closed in general). They are also quotient maps (except in the trivial case where $X_j \neq \emptyset$ but $X_i = \emptyset$ for some $i \neq j$.)

Proof. If we just have a product $X_0 \times X_1$ of two factors, we note that open sets of the form $U_0 \times U_1$ form a basis for the topology, and $\pi_i(U_0 \times U_1)$ is either U_i or empty (in a trivial case). Either way, it is open, and it follows that π_i is an open map.

We now treat the general case, which is essentially the same idea, but needs more notation. Write $X = \prod_i X_i$. Consider an open set $V \subseteq X$, and a point $x_j \in \pi_j(V)$. This means that there exists a point $x \in V$ whose j'th coordinate is x_j. As V is open, there is a set $U \in \beta'$ with $x \in U \subseteq V$. As in Remark 5.29, we can describe U as $\prod_i U_i$ for some family of open sets $U_i \subseteq X_i$, with $U_i = X_i$ for all but finitely many indices i. As $x \in U$ we have $x_i \in U_i$ for all i. Now define $f \colon X_j \to X$ by

$$f(y) = \begin{cases} x_i & \text{if } i \neq j \\ y & \text{if } i = j. \end{cases}$$

We find that $f(U_j) \subseteq U \subseteq V$, so $\pi_j(f(U_j)) \subseteq \pi(V)$, but $\pi_j \circ f = 1$ so $x_j \in U_j \subseteq \pi(V)$. This proves that $\pi(V)$ is a neighbourhood of x_j, as required.

For an example showing that π_j need not be closed, consider the set $F = \{(x, y) \in \mathbb{R}^2 \,:\, xy = 1\}$. This is closed in $\mathbb{R}^2$, but

$$\pi_0(F) = \pi_1(F) = \mathbb{R} \setminus \{0\},$$

which is not closed in $\mathbb{R}$.

If each set X_i (for $i \neq j$) is nonempty, we can choose $x_i \in X_i$ for all i, and then define $f\colon X_j \to X$ as above. We again have $\pi_j \circ f = 1_{X_j}$, and it follows from Proposition 4.8 (or from Proposition 4.5) that π_j is a quotient map. □

Example 5.31. [eg-binary-seq-product]

Suppose we give the set $\{0, 1\}$ the discrete topology, and let τ' denote the resulting product topology on the set $X = \prod_{k=0}^{\infty}\{0, 1\}$ of binary sequences. We claim that this is the same as the topology τ defined in Example 2.6. Indeed, the basic open set $C_n(x)$ for τ can be described as $\pi_J^{-1}(U)$, where $J = \{0, 1, \dots, n-1\}$ and U is the single point $(x_0, \dots, x_{n-1})$, which is open in the discrete space $\prod_{i=0}^{n-1}\{0, 1\}$. This means that the basic open sets for τ are τ'-open. Conversely, suppose we have a basic open set $V = \pi_J^{-1}(U)$ for τ' (so J is a finite subset of $\mathbb{N}$). Put $n = \max(J) + 1$, and note that whenever $x \in V$ we have $C_n(x) \subseteq V$. Using this we see that V is τ-open, as required.

Proposition 5.32. [prop-sub-product]

If $Y_i \subseteq X_i$ for each $i \in I$ then the two obvious topologies on $\prod_I Y_i$ (as a product of subspaces of the X_i or as a subspace of the product $\prod_I X_i$) are the same.

Proof. Put $X = \prod_I X_i$ and $Y = \prod_I Y_i$. We write τ for the product topology on X, and σ for the standard subbasis that generates τ. Similarly, we write τ' for the product topology on Y, and σ' for the standard subbasis that generates τ'. Proposition 5.5 tells us that $\sigma|_Y$ is a subbasis for $\tau|_Y$, so it will suffice to show that $\sigma|_Y = \sigma'$. Let π_i be the projection $X \to X_i$, and let π'_i be the projection $Y \to Y_i$. The subbasis σ consists of the sets $\pi_i^{-1}(U)$, where $i \in I$ and U is open in X_i. This means that $\sigma|_Y$ consists of the sets $Y \cap \pi_i^{-1}(U)$, which are the same as the sets $(\pi'_i)^{-1}(Y_i \cap U)$. Moreover, the sets $Y_i \cap U$ are precisely the subsets of Y_i that are open with respect to the subspace topology. It follows that $\sigma|_Y = \sigma'$ as required. □

Remark 5.33. [rem-embedding-product]

In the light of Remark 5.3, the above proposition can be restated as follows: If we have a family of embeddings $f_i\colon Y_i \to X_i$, then the product map $\prod_I f_i\colon \prod_i Y_i \to \prod_i X_i$ is also an embedding.

Proposition 5.34. [prop-product-closure]

Suppose we have a family of spaces X_i and subsets $Y_i \subseteq X_i$, giving a subset $Y = \prod_I Y_i$ in $X = \prod_I X_i$.

(a) *If each set Y_i is closed in X_i, then Y is closed in X.*
(b) *More generally, we always have $\mathrm{cl}_X(Y) = \prod_I \mathrm{cl}_{X_i}(Y_i)$.*

Proof. (a) Suppose that Y_i is closed for all i. As π_i is continuous, it follows that $\pi_i^{-1}(Y_i)$ is closed in X. The intersection of any family of closed sets is again closed, so $\bigcap_I \pi_i^{-1}(Y_i)$ is closed, but that intersection is visibly equal to Y.

(b) Now let the sets Y_i be arbitrary. Part (a) tells us that the set $Z = \prod_I \mathrm{cl}_{X_i}(Y_i)$ is closed and it clearly contains Y, so it must contain $\mathrm{cl}_X(Y)$. Conversely, suppose that $z \in Z$. Let U be any neighbourhood of z. By the standard relationship between topologies and their bases, we can find a set V in the basis β' (from Proposition 5.28) such that $z \in V \subseteq U$. By the definition of β', there is a finite set J and open sets $V_j \subseteq X_j$ (for $j \in J$) such that $V = \bigcap_{j \in J} \pi_j^{-1}(V_j)$. For $i \notin J$ we put $V_i = X_i$, so now $V = \prod_I V_i$ and V_i is a neighbourhood of z_i for all i. As $z \in Z$ we must also have $z_i \in \mathrm{cl}_{X_i}(Y_i)$, which means that $V_i \cap Y_i$ must be nonempty, containing some point y_i say. This gives a point $y = (y_i)_{i \in I}$ lying in $Y \cap V \subseteq Y \cap U$. In particular, we see that $Y \cap U$ is always nonempty, so $z \in \mathrm{cl}_X(Y)$. This gives the required reverse inclusion.

□

Proposition 5.35. [prop-product-seq-gen]

Let $(X_i)_{i \in I}$ be a family of topological spaces, and let $\underline{x} = (x_n)_{n \in \mathbb{N}}$ be a sequence in the product space $X = \prod_I X_i$. Then $\underline{x}$ converges to a point $a \in X$ if and only if $(\pi_i(x_n))_{n \in \mathbb{N}}$ converges to $\pi_i(a)$ for all $i \in I$.

Proof. First, we know that the maps π_i are continuous and therefore (by Proposition 3.15) sequentially continuous. Thus, if $(x_n)_{n\in\mathbb{N}}$ converges to a then $(\pi_i(x_n))_{n\in\mathbb{N}}$ converges to $\pi_i(a)$.

Conversely, suppose that $(\pi_i(x_n))_{n\in\mathbb{N}}$ converges to $\pi_i(a)$ for all $i \in I$. Consider a neighbourhood U of a in X. Using Proposition 5.28 we see that there is a finite set $J \subseteq I$ and a family of open sets $U_j \subseteq X_j$ for $j \in J$ such that $a \in \pi_J^{-1}(\prod_j U_j) \subseteq U$. This means that for each $j \in J$, the set U_j is a neighbourhood of $\pi_j(a)$ in X_j. By hypothesis the sequence $(\pi_j(x_n))_{n\in\mathbb{N}}$ converges to $\pi_j(a)$, so there exists $N_j \in \mathbb{N}$ such that $\pi_j(x_n) \in U_j$ whenever $n \geq N_j$. As J is finite we have a well-defined integer $N = \max\{N_j \ : \ j \in J\}$. For $n \geq N$ we have

$$x_n \in \bigcap_{j\in J} \pi_j^{-1}(U_j) = \pi_J^{-1}(\prod_J U_j) \subseteq U.$$

As U was arbitrary this means that (x_n) converges to a, as required. □

5.3. Disjoint Unions.

Definition 5.36. [defn-coproduct-set]

Suppose we have a family of sets $(X_i)_{i\in I}$. The *disjoint union* or *coproduct set* is the set of all pairs (i, x), where $i \in I$ and $x \in X_i$. We write $\coprod_{i\in I} X_i$ for this set. For each $i \in I$ we have a map $\iota_i \colon X_i \to \coprod_{i\in I} X_i$ given by $\iota_i(x) = (i, x)$.

Remark 5.37. [rem-maps-from-coproduct]

Consider a map $f \colon \coprod_{i\in I} X_i \to Y$. We can compose with the inclusions $\iota_i \colon X_i \to \coprod_{i\in I} X_i$ to get a family of maps $f_i = f \circ \iota_i \colon X_i \to Y$. Conversely, given such a family of maps $f_i \colon X_i \to Y$, we can define $f \colon \coprod_{i\in I} X_i \to Y$ by $f(i, x) = f_i(x)$. These constructions are visibly inverse to each other.

Remark 5.38. [rem-union-disjoint]

Suppose that the sets X_i are all subsets of some other set X. We can then define a surjective map $m \colon \coprod_{i\in I} X_i \to \bigcup_{i\in I} X_i$ by $m(i, x) = x$. If the sets X_i are disjoint, then m is a bijection. In this situation it is common to silently identify $\coprod_{i\in I} X_i$ with $\bigcup_{i\in I} X_i$.

Remark 5.39. For any subset $U \subseteq \coprod_{i\in I} X_i$ we have subsets $U_i = \iota_i^{-1}(U_i) \subseteq X_i$, and it is easy to see that $U = \coprod_{i\in I} U_i$. Thus, subsets of $\coprod_{i\in I} X_i$ biject with families $(U_i)_{i\in I}$ with $U_i \subseteq X_i$ for all i.

Definition 5.40. [defn-coproduct-space]

Now suppose we have a family of topological spaces $(X_i)_{i\in I}$. We declare that a subset $U \subseteq \coprod_{i\in I} X_i$ is open if and only if the set $U_i = \iota_i^{-1}(U)$ is open in X_i for all i. This defines a topology on $\coprod_{i\in I} X_i$, which we call the *coproduct topology*. By the *coproduct space* we mean the coproduct set equipped with the coproduct topology.

Remark 5.41. [rem-coproduct-closed]

Consider a family of subsets $F_i \subseteq X_i$ and the corresponding subset $F = \coprod_i F_i \subseteq \coprod_i X_i$, so $F_i = \iota_i^{-1}(F)$. It is easy to see that F is closed in the coproduct topology iff F_i is closed in X_i for all i. Note here that the index set I may be infinite, in which case we have a situation where certain infinite unions of closed sets are automatically closed.

Remark 5.42. [rem-coproduct-misc]

It is clear that the maps ι_i are continuous. Moreover, the subset $\iota_i(X_i)$ (with the subspace topology) is homeomorphic to X_i and is both open and closed in $\coprod_{i\in I} X_i$.

Proposition 5.43. [prop-maps-from-coproduct]

Consider a topological space Y and a map $f \colon \coprod_{i\in I} X_i \to Y$. Then f is continuous (with respect to the coproduct topology) if and only if the maps $f_i = f \circ \iota_i \colon X_i \to Y$ are all continuous. Thus, continuous maps $f \colon \coprod_{i\in I} X_i \to Y$ biject with families of continuous maps $f_i \colon X_i \to Y$.

Proof. The map f is continuous if and only if for every open set $V \subseteq Y$, the preimage $f^{-1}(V)$ is open in the coproduct topology. This means by definition that the sets $\iota_i^{-1}(f^{-1}(V))$ must be open in X_i, or in other words $f_i^{-1}(V)$ must be open. Thus, f is continuous if and only if all the f_i are continuous. □

Remark 5.44. [rem-coproduct-categorical]
This proposition means that the coproduct space is a coproduct in the sense of category theory, as in Definition 36.80.

Corollary 5.45. [cor-coproduct-map]
Suppose we have a family of continuous maps $f_i\colon X_i \to Y_i$, and we define

$$f = \coprod_I f_i\colon \coprod_I X_i \to \coprod_I Y_i$$

by $f(i,x) = (i, f_i(x)))$. Then f is also continuous.

Proof. By the proposition, it will suffice to show that $f \circ \iota_j\colon X_j \to \coprod_i Y_i$ is continuous for all j. However, $f \circ \iota_j$ is just the same as $\iota_j \circ f_j$, and ι_j and f_j are continuous, so $\iota_j \circ f_j$ is continuous by Proposition 3.5. The maps considered are conveniently displayed in the following diagram:

$$\begin{tikzcd} X_j \arrow[r, "f_j"] \arrow[d, "\iota_j"'] & Y_j \arrow[d, "\iota_j"] \\ \coprod_I X_i \arrow[r, "f"'] & \coprod_I Y_i. \end{tikzcd}$$

□

Proposition 5.46. [prop-union-disjoint]
Suppose that the sets X_i are open disjoint subsets of some larger space X, equipped with their subspace topologies. Then the map $m\colon (i,x) \mapsto x$ gives a homeomorphism

$$\coprod_I X_i \to \bigcup_I X_i \subseteq X.$$

Remark 5.47. It is necessary to assume here that the subsets X_i are open, as we see by considering the example where $X = \mathbb{R}$ and $X_0 = (-\infty, 0)$ and $X_1 = [0, \infty)$. Here m gives a continuous bijection $X_0 \amalg X_1 \to X$, but m^{-1} is not continuous, so m is not a homeomorphism.

Proof. Put $A = \coprod_I X_i$ and $B = \bigcup_i X_i$. Note that B is the union of a family of open sets in X, so B is open in X. The map m certainly gives a bijection $A \to B$. The composite $m \circ \iota_i$ is just the inclusion of X_i in B, which is continuous. It follows that m is continuous. Now suppose we have a open subset $U \subseteq A$. This must have the form $U = \coprod_I U_i$, where U_i is open in X_i. As X_i is open in X, this means that U_i is open in X, and thus that the set $m(U) = \bigcup_I U_i$ is also open in X (and therefore in B). Using this we see that m^{-1} is also continuous, and thus that m is a homeomorphism. □

5.4. Quotients. We start by recalling the basic facts about equivalence relations and quotient sets; then we will discuss topologies on such sets.

Definition 5.48. [defn-relation]
A *relation* on a set X is a subset $R \subseteq X \times X$. In this context, we write xRy instead of $(x, y) \in R$.

Definition 5.49. [defn-equivalence]
A relation E on X is an *equivalence relation* if it satisfies the following axioms:

E0: (reflexivity) For all $x \in X$ we have xEx.
E1: (symmetry) For all $x, y \in X$ with xEy we also have yEx.
E2: (transitivity) For all $x, y, z \in X$ with xEy and yEz we also have xEz.

Remark 5.50. [rem-generated-equivalence]
Let R be any relation on X. Define a new relation R' by

$$xR'y \Leftrightarrow (x = y \text{ or } xRy \text{ or } yRx).$$

Then define $\overline{R}$ so that xRy iff there exists a chain $x = u_0, u_1, \ldots, u_r = y$ with $u_i R' u_{i+1}$ for $0 \leq i < r$. We find that $\overline{R}$ is an equivalence relation, and that it is the smallest equivalence relation containing R.

DEFINITION 5.51. [defn-quotient-set]
Let X be a set, and let E be an equivalence relation on X. For any point $x \in X$, we define

$$q_E(x) = [x]_E = \{x' \in X \ : \ xEx'\}.$$

(The subscript E will be dropped if there is no danger of confusion.) We call the set $[x]_E$ the *equivalence class* of x. We write X/E for the set of all equivalence classes, and call this the *quotient set* of X by E. Note that q_E gives a surjective map $X \to X/E$, which we call the *quotient map*.

EXAMPLE 5.52. [eg-mod-five]
Take $X = \mathbb{Z}$, and put $E = \{(n, n+5m) \ : \ n, m \in \mathbb{Z}\} \subseteq \mathbb{Z}^2$, so that pEq if and only if $p - q$ is divisible by 5. This is easily seen to be an equivalence relation. There are equivalence classes $[0], [1], [2], [3]$ and $[4]$, and every other equivalence class is equal to precisely one of these. For example, we have $[555] = [5] = [0]$ and $[6] = [1]$ and so on. Thus $\mathbb{Z}/E$ is a finite set of size five.

EXAMPLE 5.53. [eg-trivial-equiv]
The sets $E = X^2$ and $\Delta = \{(x, x) \ : \ x \in X\}$ are both equivalence relations on X. The relation E has only one equivalence class, namely the set X itself, so X/E is a single point. The equivalence classes for Δ are just the sets of the form $\{x\}$ for some $x \in X$, so X/Δ can be identified with X itself.

EXAMPLE 5.54. [eg-glue-ends]
We can define an equivalence relation E on $[0, 1]$ by

$$E = \{(t, t) \ : \ 0 \leq t \leq 1\} \cup \{(0, 1), (1, 0)\}.$$

The equivalence classes are the singletons $[t] = \{t\}$ for $0 < t < 1$, together with the set $[0] = [1] = \{0, 1\}$. Intuitively, the quotient set $[0, 1]/E$ can be thought of as the result of taking the interval $[0, 1]$ and gluing the two ends together. Later we will introduce a topology that reflects this idea.

EXAMPLE 5.55. [eg-collapse]
Let X be a set, and let Y be a nonempty subset. The set $E = \Delta \cup Y^2 \subseteq X^2$ is then an equivalence relation. Each point $x \in X \setminus Y$ gives an equivalence class $\{x\}$, and there is one more equivalence class, namely the set Y. It is conventional to write X/Y for X/E, and ∞ for Y considered as a point of X/Y, so as sets we just have $X/Y \simeq (X \setminus Y) \amalg \{\infty\}$. We say that X/Y is obtained from X by *collapsing* Y *to a point*. Example 5.54 is the special case where $X = [0, 1]$ and $Y = \{0, 1\}$.

For completeness, we should mention a technical issue. It is common to use this construction in situations where Y is not known in advance and might be empty. To ensure that this case is covered in a useful way, it is usual to modify the definition as follows. We take ∞ to be an arbitrary point disjoint from X, and define a relation $E = \{\infty\} \times Y$ on $X \amalg \{\infty\}$. The equivalence relation $\overline{E}$ generated by E is

$$\begin{aligned}
\overline{E} &= \Delta_{X \amalg \{\infty\}} \cup (Y \amalg \{\infty\})^2 \\
&= (\Delta_X \cup Y^2) \amalg (\{\infty\} \times Y) \amalg (Y \times \{\infty\}) \amalg \{(\infty, \infty)\} \\
&\subseteq X^2 \amalg (\{\infty\} \times Y) \amalg (Y \times \{\infty\}) \amalg \{(\infty, \infty)\} = (X \amalg \{\infty\})^2.
\end{aligned}$$

We then define $X/Y = (X \amalg \{\infty\})/E$. This is the same as before when Y is nonempty, but gives $X/\emptyset = X \amalg \{\infty\}$. There are compelling reasons from category theory to consider this as the right definition.

EXAMPLE 5.56. [eg-equaliser-relation]
Let $f \colon X \to Y$ be any map of sets, and put

$$E = \mathrm{eq}(f) = \{(x, x') \in X^2 \ : \ f(x) = f(x')\} \subseteq X^2.$$

This is easily seen to be an equivalence relation on X. For each point $y \in \mathrm{img}(f) \subseteq Y$ we find that $f^{-1}\{y\}$ is an equivalence class, and that this construction gives a bijection $\mathrm{img}(f) \to X/E$.

PROPOSITION 5.57. [prop-equivalence]
Let E be an equivalence relation on a set X. Note that every element of X/E is an equivalence class and so is a subset of X.

(a) *If $x \in X$ and $y \in X/E$ then x lies in y if and only if $y = [x]$.*
(b) *If $y, y' \in X/E$ then either $y = y'$ or $y \cap y' = \emptyset$.*

(c) *Each element $x \in X$ lies in precisely one equivalence class, namely $[x]$.*

PROOF. (a) Let y be an equivalence class. If $y = [x] = \{x' \;:\; xEx'\}$ then $x \in y$ by axiom E0. Conversely, let x be any element of y. As y is an equivalence class, we must have $y = [u]$ for some u. By the definition of $[u]$ we must have uEx and so also xEu (by E1). If xEv then we can use uEx and E2 to see that also uEv; this shows that $[x] \subseteq [u]$. Essentially the same argument shows that $[u] \subseteq [x]$, so $y = [x]$ as required.

(b) Now let y and y' be two equivalence classes. If $y \cap y' \neq \emptyset$ then we can choose $x \in y \cap y'$ and part (a) shows that $y = [x] = y'$.

(c) This is just a paraphrase of (a).

$\square$

COROLLARY 5.58. [`cor-quotient-induced`]
Let E be an equivalence relation on a set X, and let f be a function from X to another set Y. This gives another equivalence relation $\operatorname{eq}(f)$ as in Example 5.56.

(a) *If $f(x) = f(x')$ whenever xEx' (or equivalently, $E \subseteq \operatorname{eq}(f)$), then there is a unique map $\overline{f}\colon X/E \to Y$ such that $f = \overline{f} \circ q_E$.*
(b) *Otherwise, there is no map $\overline{f}\colon X/E \to Y$ such that $f = \overline{f} \circ q_E$.*
(c) *Suppose that $\overline{f}$ exists. Then $\overline{f}$ is surjective if and only if f is surjective.*
(d) *Similarly, $\overline{f}$ is injective if and only if $E = \operatorname{eq}(f)$.*
(e) *If f is surjective and $\operatorname{eq}(f) = E$, then $\overline{f}$ is bijective and the inverse is just*

$$\overline{f}^{-1}(y) = f^{-1}\{y\} = \{x \in X \;:\; f(x) = y\} \in X/E.$$

PROOF.

(a) Suppose that $f(x) = f(x')$ whenever xEx', or equivalently whenever $[x] = [x']$. Given $y \in X/E$ we can choose $x \in X$ such that $y = [x]$, and then define $\overline{f}(y) = f(x)$; the hypothesis shows that this is well-defined. We have $(\overline{f} \circ q_E)(x) = \overline{f}([x]) = f(x)$ for all x, so $\overline{f} \circ q_E = f$. As q_E is surjective, it is clear that $\overline{f}$ is uniquely determined by this property.

(b) Conversely, suppose that f factors as $\overline{f} \circ q_E$ for some map $\overline{f}$. Whenever xEx' we then have $q_E(x) = q_E(x')$ and so $f(x) = \overline{f}(q_E(x)) = \overline{f}(q_E(x')) = f(x')$. By taking the contrapositive we obtain statement (b) above.

(c) Suppose that f is surjective. Then for any $y \in Y$ we can find $x \in X$ with $f(x) = y$, which means that $\overline{f}([x]) = y$, so y is in the image of $\overline{f}$. This means that $\overline{f}$ is surjective. Conversely, suppose that $\overline{f}$ is surjective. As q_E is always surjective and $f = \overline{f} \circ q_E$, we see that f is surjective.

(d) Suppose instead that $\operatorname{eq}(f) = E$. If $a, a' \in X/E$ and $\overline{f}(a) = \overline{f}(a')$, we can choose x, x' such that $a = [x]$ and $a' = [x']$, and we see that $f(x) = f(x')$. This means that $(x, x') \in \operatorname{eq}(f) = E$, so xEx', so $[x] = [x']$, so $a = a'$. It follows that $\overline{f}$ is injective. Conversely, suppose that $\overline{f}$ is injective. This means that for $a, a' \in X/E$ we have ($\overline{f}(a) = \overline{f}(a')$ iff $a = a'$). It follows that for $x, x' \in X$ we have ($\overline{f}([x]) = \overline{f}([x'])$ iff $[x] = [x']$), or in other words ($f(x) = f(x')$ iff xEx'), so $eq(f) = E$.

(e) Now suppose that f is surjective and $\operatorname{eq}(f) = E$, so $\overline{f}$ is bijective by (c). Consider a point $y \in Y$, and put $a = f^{-1}\{y\} \subseteq X$. As f is surjective, this is nonempty, so we can choose $x \in a$. If x' is another point in X we have $f(x') = y$ iff $(x, x') \in \operatorname{eq}(f) = E$ iff $x' \in [x]$. It follows that $a = [x]$, so $a \in X/E$. It is clear that $\overline{f}(a) = y$, so $a = \overline{f}^{-1}(y)$ as claimed.

$\square$

PROPOSITION 5.59. [`prop-quotient-top`]
Let X be a set with a topology τ, and let E be an equivalence relation on X. The family

$$\tau/E = \{V \subseteq X/E \;:\; q_E^{-1}(V) \in \tau\}$$

is then a topology on X/E, with respect to which the map q_E is continuous. (We will call τ the quotient topology *on X/E.)*

Proof. Firstly, we have $q^{-1}(X/E) = X$ and $q^{-1}(\emptyset) = \emptyset$, and both of these are open in X. It follows that X/E and $\emptyset$ are open in X/E. Next, consider a family $(V_i)_{i \in I}$ of sets in τ/E. This means that the sets $q^{-1}(V_i)$ are open in X, so the union $\bigcup_I q^{-1}(V_i)$ is open in X. However, this union is just the same as $q^{-1}(\bigcup_I V_i)$, so $\bigcup_I V_i \in \tau/E$. Now suppose that the index set I is finite. It follows that $\bigcap_I q^{-1}(V_i)$ is then open in X, but this is just the same as $q^{-1}(\bigcap_I V_i)$, which proves that $\bigcap_I V_i \in \tau/E$. Thus, all the axioms for a topology are satisfied. It is tautological that this makes q_E continuous. □

Proposition 5.60. [`prop-quotient-test`]
Let X be a topological space, and let E be an equivalence relation on X. Let g be a function from X/E to another space Y. Then g is continuous (with respect to the quotient topology) if and only if $g \circ q \colon X \to Y$ is continuous.

Proof. The map g is continuous if and only if for every open set $V \subseteq Y$, the preimage $g^{-1}(V)$ is open in the quotient topology, or equivalently $q^{-1}(g^{-1}(V))$ is open in X. Here $q^{-1}(g^{-1}(V)) = (g \circ q)^{-1}(V)$, so the stated condition is equivalent to continuity of $g \circ q$, as required. □

Proposition 5.61. [`prop-maps-from-quotient`]
Let X be a topological space, let E be an equivalence relation on X, and let f be a continuous map from X to another space Y. Suppose that $f(x) = f(x')$ whenever xEx', or equivalently that $E \subseteq \mathrm{eq}(f)$.

(a) *There is a unique map $\overline{f} \colon X/E \to Y$ such that $f = \overline{f} \circ q_E$, and this map is continuous.*
(b) *The map $\overline{f}$ is injective if and only if $E = \mathrm{eq}(f)$.*
(c) *The map $\overline{f}$ is surjective if and only if f is surjective.*
(d) *The map $\overline{f}$ is a quotient map if and only if f is a quotient map.*
(e) *The map $\overline{f}$ is a homeomorphism if and only if f is a quotient map and $E = \mathrm{eq}(f)$.*

Proof.

(a) The map $\overline{f}$ exists and is unique by Corollary 5.58, and is continuous by Proposition 5.60.
(b),(c) These are already part of Corollary 5.58, and are just repeated for ease of reference.
(d) In view of (c) we can assume that both f and $\overline{f}$ are surjective here. Suppose that f is a quotient map. Let W be a subset of Y for which $\overline{f}^{-1}(W)$ is open in X/E; we must show that W is open. As q is continuous and $\overline{f}^{-1}(W)$ is open, we see that $q^{-1}(\overline{f}^{-1}(W))$ is open in X. As $f = \overline{f} \circ q$, this set is just $f^{-1}(W)$. As f is a quotient map and $f^{-1}(W)$ is open, we see that W is open as required. This means that $\overline{f}$ is a quotient map.

Conversely, suppose we start with the assumption that $\overline{f}$ is a quotient map. Let W be a subset of Y for which $f^{-1}(W)$ is open. This set is the same as $q^{-1}(\overline{f}^{-1}(W))$; by the definition of the quotient topology we deduce that $\overline{f}^{-1}(W)$ is open in X/E. As $\overline{f}$ is a quotient map this means that W is open in Y. We conclude that f is a quotient map, as required.
(e) This now follows from the previous parts together with Proposition 4.9.

□

Example 5.62. [`eg-complex-exp`]
We can introduce an equivalence relation E on $\mathbb{R}$ by

$$xEy \Leftrightarrow (x - y \in \mathbb{Z}),$$

so $\mathbb{R}/E$ is just the quotient group $\mathbb{R}/\mathbb{Z}$. We then define a map $f \colon \mathbb{R} \to S^1 = \{z \in \mathbb{C} \ : \ |z| = 1\}$ by $f(x) = \exp(2\pi i x)$. It is a standard fact from complex analysis that this is continuous, and that $f(x+n) = f(x)f(n) = f(x)$ for all $n \in \mathbb{Z}$. This means that whenever xEy we have $f(x) = f(y)$, so there is an induced continuous map $\overline{f} \colon \mathbb{R}/E \to S^1$. Further standard facts say that every $z \in S^1$ can be written as $\exp(2\pi i x)$ for some x (so f is surjective), and that $\exp(2\pi i x) = \exp(2\pi i y)$ if and only if $x - y \in \mathbb{Z}$ (so $\mathrm{eq}(f) = E$). All these facts are reviewed in Section 34.3, where we also prove that f is an open map. Proposition 4.8 therefore tells us that f is a quotient map, and it follows using Proposition 5.61(e) that $\overline{f} \colon \mathbb{R}/\mathbb{Z} \to S^1$ is a homeomorphism. We can also define $f \colon \mathbb{C} \to \mathbb{C} \setminus \{0\}$ by the same formula $f(x) = \exp(2\pi i x)$, and prove in the same way that $\mathbb{C}/\mathbb{Z}$ is homeomorphic to $\mathbb{C} \setminus \{0\}$.

Example 5.63. [`eg-R-mod-Q`]
Now consider instead the relation E on $\mathbb{R}$ given by

$$xEy \Leftrightarrow (x-y\in\mathbb{Q}).$$

We claim that the quotient topology on $\mathbb{R}/E$ is the indiscrete topology. Indeed, let V be a nonempty open set in $\mathbb{R}/E$. As $q\colon\mathbb{R}\to\mathbb{R}/E$ is continuous and surjective we see that the set $U=q^{-1}(V)\subseteq\mathbb{R}$ is nonempty and open an satisfies $q(U)=V$. Choose a point $x\in U$, and then choose $\epsilon>0$ such that $(x-\epsilon,x+\epsilon)\subseteq U$. Consider an arbitrary number $y\in\mathbb{R}$. As $\mathbb{Q}$ is dense in $\mathbb{R}$ (by Example 2.21) we can find a rational number a in the interval $(x-y-\epsilon,x-y+\epsilon)$, so $a+y\in(x-\epsilon,x+\epsilon)\subseteq U$, so $q(a+y)\in q(U)=V$. On the other hand, from the definition of E we have $q(y)=q(a+y)$, so $q(y)\in V$. As y was arbitrary this means that $V=\mathbb{R}/E$. Thus, the only nonempty open set in $\mathbb{R}/E$ is $\mathbb{R}/E$ itself, so the topology is indiscrete as claimed.

Group actions are an important source of equivalence relations and quotient spaces, as we now explain. We first recall some basic definitions, to fix the terminology and notation.

Definition 5.64. [`defn-G-set`]
Let G be a group. A *G-set* is a set X with a given action of G, so for all $g\in G$ and $x\in X$ we have an element $gx\in X$, defined in such a way that

- When g is the identity element $1\in G$, we have $1x=x$ for all x.
- For all $g,h\in G$ and $x\in X$ we have $(gh)x=g(hx)$.

For any G-set X and $x\in X$ we put

$$\begin{aligned}\mathrm{stab}_G(x)&=\{g\in G\ :\ gx=x\}\\ \mathrm{orb}_G(x)&=\{gx\ :\ g\in G\}\end{aligned}$$

We call these the *stabiliser* and the *orbit* of x, and we note that $\mathrm{stab}_G(x)$ is a subgroup of G. We say that the action is *free* if $\mathrm{stab}_G(x)=\{1\}$ for all x. Next, we introduce an equivalence relation E_G on X by

$$E_G=\{(x,gx)\ :\ x\in X,\ g\in G\}\subseteq X^2,$$

so

$$xE_Gy \text{ iff } x\in\mathrm{orb}_G(y) \text{ iff } y\in\mathrm{orb}_G(x) \text{ iff } \mathrm{orb}_G(x)=\mathrm{orb}_G(y),$$

so the equivalence classes are precisely the orbits. We write X/G for $X/\sim_G$ and call this the *orbit set*. We say that the action is *transitive* if X/G is a singleton, or equivalently $X\neq\emptyset$ and for all $x,y\in X$ there exists $g\in G$ with $gx=y$.

Definition 5.65. [`defn-G-space`]
A *G-space* is a topological space with an action of G such that for each $g\in G$, the map $\alpha_g\colon X\to X$ given by $\alpha_g(x)=gx$ is continuous. In this context we will equip the set X/G with the quotient topology, and call it the *orbit space*.

Remark 5.66. [`rem-action-homeo`]
If the maps α_g are all continuous then in fact they are all homeomorphisms, because $\alpha_{g^{-1}}$ is an inverse for α_g.

Remark 5.67. [`rem-G-space-functors`]
As explained in Example 36.90, G-sets can be identified with functors from the one-object category bG to the category of sets, and similarly G-spaces are functors from bG to the category of spaces.

Remark 5.68. [`rem-topological action`]
For the moment we are treating G as a discrete set. In many cases G will also have a natural topology, and in that context it is natural to ask for the action map $G\times X\to X$ (given by $(g,x)\mapsto gx$) to be continuous with respect to the product topology. We will postpone detailed consideration of this situation, however.

Postpone or omit?

Example 5.69. [`eg-Sn-RPn-quot`]
In Example 5.24 we defined a surjective map $f\colon S^n\to\mathbb{R}P^n$ and showed that $f(x)=f(y)$ iff $y=\pm x$. We can let the group $C_2=\{1,-1\}$ act on S^n by multiplication, and we see that f induces a continuous

bijection $\overline{f}\colon S^n/C_2 \to \mathbb{R}P^n$ with $\overline{f}q = f$. In fact, this is a homeomorphism. One way to prove this is as follows. Let $e_0, \dots, e_n$ be the standard basis for $\mathbb{R}^{n+1}$, and put

$$U_k = \{A \in \mathbb{R}P^n \;:\; Ae_k \neq 0\}.$$

It is easy to see that these are open in $\mathbb{R}P^n$, and that $\mathbb{R}P^n = U_0 \cup \cdots \cup U_n$. We can define a continuous map $s_k\colon U_k \to S^n$ by $s_k(A) = (Ae_k)/\|Ae_k\|$. By examining the proof of Example 5.24 we see that $f(s_k(A)) = A$ for all $A \in U_k$, so $q \circ s_k$ is the restriction of $\overline{f}^{-1}$ to U_k. In particular, these restrictions are all continuous. As the sets U_k are open and cover $\mathbb{R}P^n$, it follows that $\overline{f}^{-1}$ itself is continuous, so $\overline{f}$ is a homeomorphism.

Example 5.70. [eg-lens-space]

Put $C_d = \{z \in \mathbb{C} \;:\; z^d = 1\}$, which is a group under multiplication. We can let this act on $S^{2n-1} = S(\mathbb{C}^n)$ by multiplication, and this action is free. More generally, given integers $p_1, \dots, p_n$ we can define a less obvious action on S^{2n-1} by

$$z.(x_1, \dots, x_n) = (z^{p_1}x_1, \dots, z^{p_n}x_n).$$

We claim that this is free iff the integers p_k are all coprime to d. Indeed, if $m > 1$ is a common factor of d and p_k and e_k is the k'th basis vector and $z = e^{2\pi i/m} \in C_d$ then $z.e_k = e_k$, proving that the action is not free. Conversely, suppose that all the p_k are coprime to d and that $z \in C_d$ and $x \in S^{2n-1}$ with $z.x = x$. We can then find k with $x_k \neq 0$ and the equation $z.x = x$ gives $(z^{p_k} - 1)x_k = 0$ so $z^{p_k} = 1$. We also have $z^d = 1$. Moreover, as $(d, p_k) = 1$ there exist integers s, t with $sd + tp_k = 1$, which gives

$$z = z^{sd+tp_k} = (z^d)^s\,(z_k^p)^t = 1,$$

proving that the action is free.

In this case where the action is free, we write $L(p_1, \dots, p_n)$ for the orbit space. Spaces of this type are called *Lens spaces*, and they have been studied extensively.

Example 5.71. [eg-fta-orbits]

We can let the permutation group Σ_n act on $\mathbb{C}^n$ by the rule

$$\sigma.(z_1, \dots, z_n) = (z_{\sigma^{-1}(1)}, \dots, z_{\sigma^{-1}(n)}).$$

In a slightly different notation, if we write $z(i)$ instead of z_i then z becomes a function from the set $N = \{1, \dots, n\}$ to $\mathbb{C}$, and the rule is $\sigma.z = z \circ \sigma^{-1}$. We then have

$$\tau.(\sigma.z) = \tau.(z \circ \sigma^{-1}) = z \circ \sigma^{-1} \circ \tau^{-1} = z \circ (\tau\sigma)^{-1} = (\tau\sigma).z$$

as required. This formulation makes it easy to see why inversion is necessary. Note that the action is not free (unless $n \leq 1$) because the stabiliser of the zero vector is all of Σ_n. We let q be the quotient map $\mathbb{C}^n \to \mathbb{C}^n/\Sigma_n$

Now let P_n denote the set of polynomials of the form

$$f(t) = a_0 + a_1t + \cdots + a_{n-1}t^{n-1} + t^n$$

with $a_i \in \mathbb{C}$. We give this the obvious topology so that it is homeomorphic to $\mathbb{C}^n$. Now define $\phi\colon \mathbb{C}^n \to P_n$ by

$$\phi(z)(t) = \prod_{i=1}^{n}(t - z_i).$$

This is clearly continuous. It is also clear that this product is unchanged if we reorder the numbers z_i, so $\phi(\sigma.z) = \phi(z)$ for all σ and z, so there is an induced continuous map $\overline{\phi}\colon \mathbb{C}^n/\Sigma_n \to P_n$ with $\overline{\phi}q = \phi$. It turns out that this is a homeomorphism.

Indeed, the Fundamental Theorem of Algebra says that every polynomial $f(t) \in P_n$ has a root, say z_n, so $f(t) = (t - z_n)g(t)$ for some $g(t) \in P_{n-1}$. By an obvious induction based on this we can write $f(t) = \prod_{i=1}^{n}(t - z_n)$ for some list $z = (z_1, \dots, z_n)$, so $f = \phi(z)$. This proves that ϕ is surjective, and so $\overline{\phi}$ is surjective.

Next, suppose that $\phi(z) = f$ say. For any complex number u, the number of times that u occurs in the list z is the order of vanishing of $f(t)$ at $t = u$. If we also have $\phi(w) = f$ we see that w must contain the same entries as z, repeated the same number of times. It follows that there exists a permutation σ with $\sigma.z = w$, so $q(z) = q(w)$. It follows from this that $\overline{\phi}$ is also injective, so it is a continuous bijection. It is also true that the inverse is continuous, but we do not yet have an efficient method to prove this. **Refer forward to a proof.**

We saw in Example 4.6 that quotient maps need not be open in general. However, the situation is better for quotient maps arising from group actions, as we now show.

LEMMA 5.72. [`lem-quotient-open`]
For any G-space X, the quotient map $q\colon X \to X/G$ is open. If G is finite then q is also closed.

In fact, it is common for q to be closed even when G is infinite, but more subtle criteria are needed for this. We will not discuss details here.

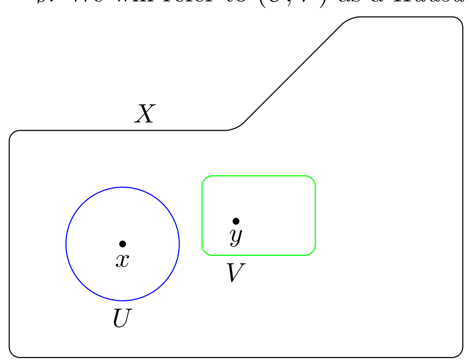
PROOF. Let U be an open subset of X, and put $V = q(U) \subseteq X/G$. We must show that this is open in the quotient topology on X/G, or equivalently that $q^{-1}(V)$ is open in X. Now we have

$$\begin{aligned} x \in q^{-1}(V) &\Leftrightarrow q(x) \in V = q(U) \\ &\Leftrightarrow q(x) = q(y) \text{ for some } y \in U \\ &\Leftrightarrow x = gy \text{ for some } y \in U \text{ and } g \in G \\ &\Leftrightarrow x \in \bigcup_{g\in G} gU, \end{aligned}$$

so $q^{-1}(V) = \bigcup_{g\in G} gU$. As the maps $x \mapsto gx$ are homeomorphisms, we see that gU is open, so $q^{-1}(V)$ is open as required. Now suppose that G is finite. For any closed set $F \subseteq X$ we again have $q^{-1}(q(F)) = \bigcup_{g\in G} gF$, which is a finite union of closed sets and thus is again closed. As q is a quotient map, we deduce that $q(F)$ is closed in X/G, as claimed. □

6. The Hausdorff Property

DEFINITION 6.1. [`defn-hausdorff`]
A space X is *Hausdorff* if and only if for every pair $x, y \in X$ with $x \neq y$ there are open sets U, V such that $x \in U$ and $y \in V$ and $U \cap V = \emptyset$. We will refer to (U, V) as a *Hausdorff pair* for (x, y).



PROPOSITION 6.2. [`prop-metric-hausdorff`]
A semimetric space is Hausdorff if and only if it is a metric space; in particular $\mathbb{R}$ is Hausdorff.

PROOF. Let X be a set equipped with a semimetric d and the corresponding topology. First suppose that d is a metric. Given distinct points $x, y \in X$ we put $\epsilon = d(x,y)/2 > 0$ and $U = OB_\epsilon(x)$ and $V = OB_\epsilon(y)$, so U and V are open and $x \in U$ and $y \in V$. If $z \in U \cap V$ then we have $d(x,z) < \epsilon$ and $d(z,y) < \epsilon$ so $d(x,y) \leq d(x,z) + d(z,y) < 2\epsilon = d(x,y)$, which is a contradiction. It follows that $U \cap V = \emptyset$, and thus that X is Hausdorff.

Now suppose instead that d is not a metric, so there is some pair x, y with $x \neq y$ but $d(x,y) = 0$. If U is any neighbourhood of x then U must contain $OB_\epsilon(x)$ for some $\epsilon > 0$, but that implies that $y \in U$, so there cannot be any neighbourhood V of y that is disjoint from U. Thus, X is not Hausdorff. □

EXAMPLE 6.3. [`eg-hausdorff`]

(a) If X is an infinite set with the cofinite topology (Example 2.16) then any two nonempty open sets have nonempty intersection, and it follows from this that X is not Hausdorff.
(b) Consider $\mathbb{R}^n$ with the Zariski topology as in Example 2.18. We will assume the standard fact that a polynomial $f \in \mathbb{R}[x_1, \dots, x_n]$ vanishes everywhere on $\mathbb{R}^n$ if and only if all the coefficients of f are zero. If $V(I)$ and $V(J)$ are proper closed subsets of $\mathbb{R}^n$ then the ideals I and J must be nonzero. If we choose $f \in I \setminus \{0\}$ and $g \in J \setminus \{0\}$ we see that fg is a nonzero polynomial that vanishes everywhere on $V(I) \cup V(J)$, so $V(I) \cup V(J)$ is again a proper subset of $\mathbb{R}^n$. By taking complements, we see that the intersection of any two nonempty open sets is nonempty. This clearly means that the Zariski topology is not Hausdorff. Examples related to this one are probably the most important application of non-Hausdorff spaces in other areas of mathematics.
(c) More obviously, any indiscrete space with at least two points is not Hausdorff.
(d) If E is an equivalence relation on a Hausdorff space X, it can easily happen that X/E is not Hausdorff. This holds for the quotient group $\mathbb{R}/\mathbb{Q}$ discussed in Example 5.63, for example. In many applications, the appropriate response is to find a larger equivalence relation $\overline{E}$ such that $X/\overline{E}$ is Hausdorff, and work with $X/\overline{E}$ instead.

PROPOSITION 6.4. [`prop-hausdorff-constructs`]
Subspaces, products and coproducts of Hausdorff spaces are Hausdorff.

PROOF.
(a) Let X be a Hausdorff space, and let Y be a subset with the subspace topology. Suppose that y_0 and y_1 are distinct points in Y. By assumption there is a Hausdorff pair (U_0, U_1) for (y_0, y_1) in X, and it follows that $(U_0 \cap Y, U_1 \cap Y)$ is a Hausdorff pair for (y_0, y_1) in Y.
(b) Let $(X_i)_{i \in I}$ be a family of Hausdorff spaces, and put $X = \prod_I X_i$. Let x and y be distinct points in X. As $x \neq y$ we must have $x_i \neq y_i$ for some i. As X_i is Hausdorff there is a Hausdorff pair (U, V) for (x_i, y_i) in X_i. It follows that $(\pi_i^{-1}(U), \pi_i^{-1}(V))$ is a Hausdorff pair for (x, y) in X.
(c) Now instead consider the coproduct $X' = \coprod_I X_i$. Let x' and y' be distinct points in X', say $x' = (i, u)$ and $y' = (j, v)$. If $i \neq j$ then $(\iota_i(X_i), \iota_j(X_j))$ is a Hausdorff pair for (x', y'). Otherwise, u and v are distinct points in the Hausdorff space X_i, so we can find a corresponding Hausdorff pair (U, V), and we find that $(\iota_i(U), \iota_i(V))$ is the required Hausdorff pair for (x', y').

$\square$

PROPOSITION 6.5. [`prop-finite-subspace`]
Let X be a Hausdorff space, and let Y be a finite subset. Then Y is closed in X, and the subspace topology on Y is discrete.

PROOF. Consider a point $y \in Y$. For any $x \in \{y\}^c$ we can choose a Hausdorff pair (U, V) for (x, y), and we see that U is an open neighbourhood of x contained in $\{y\}^c$. It follows that $\{y\}^c$ is open, so $\{y\}$ is closed. If $Y = \{y_1, \dots, y_n\}$ then $Y = \bigcup_{i=1}^n \{y_i\}$ which is a finite union of closed sets and thus is closed. The same argument shows that every subset $Z \subseteq Y$ is closed, which means that the subspace topology is discrete. $\square$

PROPOSITION 6.6. [`prop-closed-diagonal`]
A space X is Hausdorff if and only if the diagonal $\Delta = \{(x, x) \,:\, x \in X\}$ is a closed subset of $X \times X$ (or equivalently Δ^c is open).

PROOF. Recall that sets of the form $U \times V$ (with U and V open in X) form a basis for the product topology. Moreover, we have $(U \times V) \cap \Delta = \{(x, x) \,:\, x \in U \cap V\}$, so $U \times V$ is contained in Δ^c if and only if $U \cap V = \emptyset$. Thus, Hausdorff pairs for (x, y) biject with basic neighbourhoods of (x, y) that are contained in Δ^c. The claim is clear from this. $\square$

LEMMA 6.7. [`lem-unique-limits`]
If X is Hausdorff (and in particular, if X is a metric space) then any sequence in X converges to at most one point.

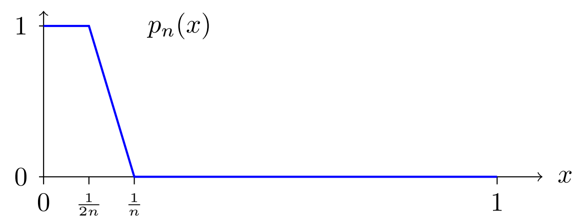
PROOF. Suppose that $\underline{x}$ converges to both a and b, where $a \neq b$. As X is Hausdorff we can choose disjoint open sets A and B with $a \in A$ and $b \in B$. As $\underline{x}$ converges to a we can find $N \in \mathbb{N}$ such that

$x_i \in A$ for all $i \geq N$. As $\underline{x}$ converges to b we can find $M \in \mathbb{N}$ such that $x_i \in B$ for all $i \geq M$. Now $x_{\max(N,M)} \in A \cap B = \emptyset$, which is a contradiction. □

Example 6.8. [eg-germs]
Consider the set $X = C([0,1])$ of continuous functions $u\colon [0,1] \to \mathbb{R}$. We introduce an equivalence relation E on X by

$$uEv \Leftrightarrow (\text{ There exists } \epsilon > 0 \text{ such that } u(x) = v(x) \text{ for all } x \in [0, \epsilon] \text{ }).$$

The equivalence classes are called *germs*. The set of germs is very useful in differential geometry, but it is not obvious how to give it a nontrivial topology. Here we will just show that some naive approaches do not work well. Suppose we use the metric $d_1(u,v) = \int_0^1 |u(x) - v(x)|\, dx$ to give a topology on X, and then consider the quotient topology on X/E. We claim that this is indiscrete. The proof will use the map $p_n\colon [0,1] \to \mathbb{R}$ as shown below:



Consider a nonempty closed set $G \subseteq X/E$, and put $F = q^{-1}(G) \subseteq X$. As $G \neq \emptyset$, there is at least one point $u \in F$. Let v be any other point in X, and put $w_n = p_n u + (1 - p_n)v$. As $p_n(x) = 1$ for $x < 1/(2n)$, we see that $q(w_n) = q(u) \in G$, so $w_n \in F$ for all n. We also have $(w_n - v)(t) = p_n(t)(u(t) - v(t))$, which is zero for $t > 1/n$ and bounded by $\|u - v\|_\infty$ for $t \leq 1/n$, so $d_1(w_n, v) \leq \|u - v\|_\infty/n \to 0$. As F is closed we deduce that $v \in F$. As v was arbitrary this means that $F = X$ and so $G = X/E$. It follows that X/E is indiscrete, as claimed.

Now suppose that we use instead the topology on X defined by the metric $d_\infty(u,v) = \max\{|u(x) - v(x)| \,:\, x \in [0,1]\}$. We can define $f\colon X \to \mathbb{R}$ by $f(u) = u(0)$, and we find that this is continuous with respect to d_∞ (but not with respect to d_1). It is clear that when uEv we have $f(u) = f(v)$, so there is an induced map $\overline{f}\colon X/E \to \mathbb{R}$ which is continuous with respect to the new quotient topology on X/E. One can check that the open sets in X/E are precisely the sets $\overline{f}^{-1}(U)$ where U is open in $\mathbb{R}$. This means that the topology on X/E is still very coarse, and does not really distinguish between X/E and the much smaller set $\mathbb{R}$.

Definition 6.9. [dfn-graph]
The *graph* of a function $f\colon X \to Y$ is the set

$$\Gamma(f) = \{(x, f(x)) \,:\, x \in X\} \subseteq X \times Y.$$

Proposition 6.10. [prop-closed-graph]
Suppose that $f\colon X \to Y$ is continuous and that Y is Hausdorff. Then $\Gamma(f)$ is closed in $X \times Y$.

Proof. The map $f \times 1\colon X \times Y \to Y \times Y$ is continuous, and the diagonal $\Delta \subseteq Y \times Y$ is closed by Proposition 6.6, so $(f \times 1)^{-1}(\Delta)$ is closed in $X \times Y$. On the other hand, we have

$$(f \times 1)^{-1}(\Delta) = \{(x,y) \in X \times Y \,:\, (f(x), y) \in \Delta\} = \{(x,y) \,:\, y = f(x)\} = \Gamma(f).$$

□

There are some important results giving converses to the above proposition under various additional conditions. However, the converse is not valid in general, as the example below will show.

Example 6.11. [eg-closed-graph]
Define $f\colon \mathbb{R} \to \mathbb{R}$ by

$$f(x) = \begin{cases} 1/x & \text{if } x \neq 0 \\ 0 & \text{if } x = 0. \end{cases}$$

This is clearly not continuous. The graph can also be described as $\Gamma(f) = Y \cup \{(0,0)\}$, where $Y = \{(x,y) : xy = 1\}$. The multiplication map $\mu\colon \mathbb{R}^2 \to \mathbb{R}$ is continuous, so the set $Y = \mu^{-1}\{1\}$ is closed, so $\Gamma(f)$ is closed.

7. Connectedness

Definition 7.1. [defn-separation]
A *separation* of a space X is a pair of open subsets A,B such that $X = A \cup B$ and $A \cap B = \emptyset$. A separation is *trivial* if $A = \emptyset$ or $B = \emptyset$. The space X is *connected* if and only if it is nonempty and has no nontrivial separation.

Remark 7.2. [rem-separation]
If $X = A \cup B$ with $A \cap B = \emptyset$ then $A = B^c$ and $B = A^c$. Thus, A is open iff B is closed, and A is closed iff B is open. In particular, if A and B give a separation then they are both open and so they are both closed as well.

Example 7.3. [eg-separation]

(a) The sets $(-\infty, 0)$ and $(0, \infty)$ give a nontrivial separation of $\mathbb{R} \setminus \{0\}$.
(b) Recall that $GL_n(\mathbb{R})$ is the space of invertible $n \times n$ matrices over $\mathbb{R}$. Put

$$GL_n^+(\mathbb{R}) = \{A \in GL_n(\mathbb{R}) \;:\; \det(A) > 0\} \qquad\qquad GL_n^-(\mathbb{R}) = \{A \in GL_n(\mathbb{R}) \;:\; \det(A) < 0\}.$$

As the determinant map is continuous, these sets are open. They give a separation of $GL_n(\mathbb{R})$, which is nontrivial (for $n > 0$). In Example 8.10 we will show that $GL_n^+(\mathbb{R})$ and $GL_n^-(\mathbb{R})$ are connected, so this separation cannot be refined any further.
(c) Consider $\mathbb{Q}$ with its topology as a subspace of $\mathbb{R}$. For any irrational number α, the sets $(-\infty, \alpha) \cap \mathbb{Q}$ and $(\alpha, \infty) \cap \mathbb{Q}$ give a nontrivial separation.
(d) Let X be the space of binary sequences as in Example 2.6. For any $n \in \mathbb{N}$, the sets $A_n = \{x : x_n = 0\}$ and $B_n = \{x \;:\; x_n = 1\}$ give a nontrivial separation of X.
(e) We shall show later (in Corollary 7.19) that $\mathbb{R}^n$ is connected.

Proposition 7.4. [prop-interval-connected]
The space $[0, 1]$ *is connected.*

Proof. Let A and B give a separation of $[0, 1]$. After exchanging A and B if necessary, we may assume that $0 \in A$. Put $P = \{x \in (0, 1] \;:\; [0, x) \subseteq A\}$. As A is open it must contain some neighbourhood of 0, so $P \neq \emptyset$. It is also clear that P is bounded above by 1. As explained in Appendix 34, it follows that P has a least upper bound, which we denote by $p = \sup(P) \in (0, 1]$. We claim that $[0, p) \subseteq A$ (so $p \in P$). Indeed, if $0 \leq x < p$ then x cannot be an upper bound for P, so there exists $a \in P$ with $x < a$, so $x \in [0, a) \subseteq A$ as required. Next, recall that B is open and $A = B^c$ so A is closed. As $[0, p) \subseteq A$ we see that p is a closure point for A and so $p \in A$. If $p = 1$ then this means that $A = [0, 1]$ and the separation is trivial, as required. Suppose instead that $p < 1$. As $p \in A$ and A is open in $[0, 1]$ we see that there must be some neighbourhood $(p - \epsilon, p + \epsilon)$ of p contained in A. It then follows that $p + \epsilon \in P$, which contradicts the fact that $p = \sup(P)$. We must therefore have $p = 1$ after all. □

Corollary 7.5 (The Intermediate Value Theorem). [cor-IVT]
Suppose we have real numbers $a < b$ *and a continuous function* $f\colon [a, b] \to \mathbb{R}$. *Suppose that there are numbers* $c, d \in [a, b]$ *and* $y \in \mathbb{R}$ *with* $f(c) \leq y \leq f(d)$. *Then there exists* $x \in [a, b]$ *with* $f(x) = y$.

Proof. If not, the open sets $f^{-1}((-\infty, y))$ and $f^{-1}((y, \infty))$ would give a nontrivial separation of the interval $[a, b]$. This is impossible because $[a, b]$ is homeomorphic to $[0, 1]$ and so is connected. □

Remark 7.6. One can prove along similar lines that sets such as $\mathbb{R}$ and $[a, b)$ are connected. However, we prefer to leave this until we have a slightly more efficient method available.

It will now be convenient to introduce some slightly more flexible terminology.

Definition 7.7. [defn-relative-separation]
Suppose we have a topological space X and a nonempty subset Y. A *relative separation* is a pair of open sets $A, B \subseteq X$ such that $Y \subseteq A \cup B$ and $A \cap B \cap Y = \emptyset$. Such a relative separation is *trivial* if $A \cap Y = \emptyset$ or $B \cap Y = \emptyset$. (Equivalently, the separation is trivial if $Y \subseteq B$ or $Y \subseteq A$.)

Proposition 7.8. [prop-relative-separation]
The set Y (considered with the subspace topology) is connected if and only if every relative separation is trivial.

Proof. This is a straightforward translation of the definitions. □

Proposition 7.9. [prop-connected-closure]
Let X be a topological space, and let Z be a connected subspace. Then any subspace Y with $Z \subseteq Y \subseteq \overline{Z}$ is also connected. In particular, $\overline{Z}$ is connected.

Proof. Let A and B be a relative separation for Y. We then have $Z \subseteq Y \subseteq A \cup B$ and $A \cap B \cap Z \subseteq A \cap B \cap Y = \emptyset$, so A and B also give a relative separation for Z. As Z is connected this must be trivial. We may therefore assume without loss of generality that $A \cap Z = \emptyset$. This means that $Z \subseteq A^c$, but A^c is closed, so $Y \subseteq \overline{Z} \subseteq A^c$, so $A \cap Y = \emptyset$. Thus, our original relative separation of Y is trivial, as required. □

Proposition 7.10. [prop-connected-union]
Let X be a space, and let Y and Z be connected subspaces of X such that $Y \cap Z \neq \emptyset$. Then $Y \cup Z$ is again connected.

Proof. Let A and B give a relative separation of $Y \cup Z$. It then follows that they also give a relative separation of Y, which must be trivial, so either $Y \subseteq A$ or $Y \subseteq B$. Similarly we have $Z \subseteq A$ or $Z \subseteq B$. If $Y \subseteq A$ and $Z \subseteq A$ then $Y \cup Z \subseteq A$ and so the original relative separation is trivial. Similarly, if $Y \subseteq B$ and $Z \subseteq B$ then the original relative separation is again trivial. Otherwise we must have $Y \subseteq A$ and $Z \subseteq B$ or *vice versa*, but either possibility would give $Y \cap Z \subseteq A \cap B = \emptyset$, contrary to assumption. Thus the original relative separation is trivial as required. □

Proposition 7.11. [prop-image-connected]
Let $f\colon X \to Y$ be a continuous map, and let X' be a connected subset of X. Then the image $Y' = f(X')$ is again connected.

Proof. Let C and D give a relative separation of Y'. As f is continuous, it follows that the sets $A = f^{-1}(C)$ and $B = f^{-1}(D)$ are open in X. We also have $X' \subseteq f^{-1}(f(X')) = f^{-1}(Y') \subseteq f^{-1}(C \cup D) = A \cup B$ and $X' \cap A \cap B \subseteq f^{-1}(Y') \cap f^{-1}(C) \cap f^{-1}(D) = f^{-1}(Y' \cap C \cap D) = f^{-1}(\emptyset) = \emptyset$. This means that A and B give a relative separation of the connected set X', which must therefore be trivial. We may assume without loss of generality that $X' \subseteq A = f^{-1}(C)$, which gives $Y' = f(X') \subseteq f(f^{-1}(C)) \subseteq C$. This proves that our original relative separation of Y' is trivial, as required. □

Corollary 7.12. [cor-quotient-connected]
Let X be a connected space, and let E be an equivalence relation on X. Then the quotient X/E is again connected.

Proof. Apply the proposition with $X' = X$ and $Y = X/E$ and $f = q_E$. □

Proposition 7.13. [prop-product-connected]
Let $(X_i)_{i \in I}$ be a family of connected spaces. Then the product $X = \prod_I X_i$ is again connected.

The proof for the general case (where the index set I may be infinite) looks somewhat involved, so we will start with a simpler case for motivation.

Lemma 7.14. [lem-product-connected]
Let X and Y be connected spaces; then $X \times Y$ is also connected.

Proof. As X and Y are connected they must be nonempty, so we can choose $u \in X$ and $v \in Y$ say. Let A and B give a separation of $X \times Y$. The point (u, v) must lie in either A or B, and we may assume without loss of generality that $(u, v) \in A$. Now consider the sets $C_u = \{y \,:\, (u, y) \in A\}$ and $D_u = \{y \,:\, (u, y) \in B\}$.

These are easily seen to give a separation of the connected space Y, and $v \in C_u$ so we must have $C_u = Y$ and $D_u = \emptyset$. Now fix $y \in Y$ and put $E_y = \{x \in X \ : \ (x,y) \in A\}$ and $F_y = \{x \in X \ : \ (x,y) \in B\}$. These are easily seen to give a separation of the connected space Y. Moreover, we have $y \in Y = C_u$ so $(u,y) \in A$ so $u \in E_y$. We must therefore have $E_y = X$ and $F_y = \emptyset$. This means that $X \times \{y\} \subseteq A$ but y was arbitrary so $A = X \times Y$. □

Proof of Proposition 7.13. Let A and B give a separation of X. For any finite subset $J \subseteq I$, we put

$$A_J = \{x \in X \ : \ \text{there exists } a \in A \text{ such that } x_i = a_i \text{ for all } i \notin J\}.$$

We will prove by induction on $|J|$ that $A_J = A$. It is clear that $A \subseteq A_J$ for all J, and that $A_\emptyset = A$, so the induction starts. Now consider the case $J = \{j\}$. If $x \in A_{\{j\}}$ we can choose $a \in A$ such that $a_i = x_i$ for all $i \neq j$. We can then define $f \colon X_j \to X$ by $f(u)_j = u$ and $f(u)_i = a_i = x_i$ for all $i \neq j$. This is clearly continuous, so the sets $f^{-1}(A)$ and $f^{-1}(B)$ form a separation of the connected space X_j, which must therefore be trivial. We have $f(a_j) = a \in A$, so $a_j \in f^{-1}(A) \neq \emptyset$, so we must have $f^{-1}(A) = X_j$. In particular we have $x_j \in f^{-1}(A)$, so the point $x = f(x_j)$ lies in A. This proves that $A_{\{j\}} = A$ as required. Now suppose that $A_J = A$, and consider a point $y \in A_{J\cup\{k\}}$ for some $k \notin J$. Choose $a \in A$ such that $x_i = a_i$ for $i \notin J \cup \{k\}$. Define $y \in X$ by $y_i = x_i$ for $i \neq k$, and $y_k = a_k$. We find that $y \in A_J = A$, and x only differs from y in the k'th place, so $x \in A_{\{k\}} = A$. It follows that $A_J = A$ for all finite J, as claimed. Similarly, we can define B_J in the same way, and we find that $B_J = B$.

The sets X_i are assumed to be connected, so they are nonempty, so X is nonempty. It follows that at least one of A and B is nonempty, so without loss of generality we can assume that we have a point $x \in A$. As A is open, it must therefore contain a basic open neighbourhood of x. Thus, there is a finite set $J \subseteq I$ and a family of open sets $U_j \subseteq X_j$ with $x_j \in U_j$ such that the set $U = \bigcap_{j\in J} U_j$ is contained in A. Now let y be an arbitrary point in X. Define $z \in X$ by $z_i = x_i$ for $i \in J$, and $z_i = y_i$ for $i \notin J$. We then have $z \in U$ so $z \in A$. It follows from this that $y \in A_J = A$. Thus, we find that $A = X$, so the original separation is trivial as required. □

Proposition 7.15. [prop-component-equivalence]
Let X be a topological space. Define a relation E on X by xEy iff (there is a connected set $Y \subseteq X$ with $\{x,y\} \subseteq Y$). Then E is an equivalence relation.

Proof. First, for any point $x \in X$ the singleton $\{x\}$ is connected, so xEx. Next, it is immediate that xEy iff yEx. Now suppose that xEy and yEz. This means that there are connected sets Y and Z with $\{x,y\} \subseteq Y$ and $\{y,z\} \subseteq Z$. Note that $Y \cap Z$ contains y and so is nonempty. We therefore see from Proposition 7.10 that $Y \cup Z$ is connected. As $\{x,z\} \subseteq Y \cup Z$ we deduce that xEz as required. □

Definition 7.16. [defn-components]
The equivalence classes for the above relation are called the *connected components* (or just *components*) of X.

Proposition 7.17. [prop-components-closed]
The components of X are closed connected sets. Moreover, they are maximal in the following sense: if C is a component and D is a connected set containing C, then $D = C$.

Proof. First, let C be a component. Then C must be nonempty, so we can choose $x \in C$. Let A and B give a relative separation of C. After exchanging A and B if necessary, we may assume that $x \in A$. Now let y be any other point in C. By the definition of the equivalence relation, we can find a connected set Y with $\{x,y\} \subseteq Y$. The definition also implies that all points in Y are equivalent to x, so $Y \subseteq [x] = C$. It follows that A and B give a relative separation of the connected set Y, which must therefore be trivial. As $x \in A \cap Y \neq \emptyset$, it follows that $Y \subseteq A$, so in particular $y \in A$. As y was an arbitrary element of C, this means that $C \subseteq A$, so the relative separation (A,B) of C is trivial. This means that C is connected, as claimed.

Now let D be any other connected set containing C. If $z \in D$ then D is a connected set containing x and z, which means that xEz. However, C is the full equivalence class of x, so we conclude that $z \in C$. This shows that $D = C$.

In particular, Proposition 7.9 shows that $\overline{C}$ is connected, so we can take $D = \overline{C}$ and conclude that $C = \overline{C}$, so C is closed. □

Corollary 7.18. [cor-open-components]
If X has only finitely many components, then the components are open as well as closed, and X is homeomorphic to the coproduct of the components.

Proof. Let the components be $C_1, \dots, C_n$, so X is the disjoint union of these sets. If we fix k then the set $C_k^c = \bigcup_{i \neq k} C_i$ is a finite union of closed sets and so is still closed, and it follows that C_k is open. The last statement now follows from Proposition 5.46. □

Corollary 7.19. [cor-Rn-connected]
The space $\mathbb{R}^n$ is connected.

Proof. Using Proposition 7.13 we reduce to the case $n = 1$. If $a, b \in \mathbb{R}$ with $a < b$ then the interval $[a, b]$ is homeomorphic to $[0, 1]$ and so is connected, so a and b lie in the same component. This means that there is only one component, namely $\mathbb{R}$ itself, so $\mathbb{R}$ is connected. □

Remark 7.20. [eg-QZ-components]
Now consider the space $\mathbb{Q}$. For any $x, y \in \mathbb{Q}$ with $x < y$, the number $u = x + (y - x)/\sqrt{2}$ is irrational with $x < u < y$. It follows that the sets $U = (-\infty, u) \cap \mathbb{Q}$ and $V = (u, \infty) \cap \mathbb{Q}$ give a separation with $x \in U$ and $y \in V$. It follows that x and y cannot lie in the same component, so the component of x is just $\{x\}$. Thus, $\mathbb{Q}$ has infinitely many components, and they are closed but not open. In the space $\mathbb{Z}$ it is again true that the components are singletons and that there are infinitely many of them, but in this case the singletons are open as well as closed.

The above example is quite typical of countable Hausdorff spaces, which might lead one to think that a countable Hausdorff space with more than one point cannot be connected. This is not true, as shown by the next example.

Example 7.21. [eg-countable-conected]
Consider the set
$$X_0 = \mathbb{Q}(\sqrt{2}) = \{a + b\sqrt{2} \ : \ a, b \in \mathbb{Q}\}.$$
This inherits a topology from $\mathbb{R}$, which we call τ. We say that a subset $U \subseteq X_0$ is σ_0-open if there is a τ-open set U^* with $\mathbb{Q} \cap U^* \subseteq U \subseteq U^*$. This gives a new topology σ_0 on X_0. Note that any τ-open set is σ_0-open, which implies that σ_0 is Hausdorff. Also, if $\mathbb{Q} \subseteq U \subseteq X_0$ then we can take $U^* = \mathbb{R}$ to see that U is σ_0-open. In particular, if T is any subset of $X_0 \setminus \mathbb{Q}$ then we find that the set $T \cup \mathbb{Q}$ is σ_0-open, so T is open with respect to the subspace topology $\sigma_0|_{X_0 \setminus \mathbb{Q}}$. This means that this subspace topology is just the discrete topology.

Next, for $u = a + b\sqrt{2} \in X_0$ we define $u^* = a - b\sqrt{2}$, and we let X be the quotient set in which u is identified with u^* for all u. The operation $u \mapsto u^*$ is not continuous with respect to σ_0, but that does not prevent us from constructing a quotient topology on X, which we call σ. We claim that X is countable, Hausdorff and connected. Countability is clear.

For the rest, we put
$$U_\epsilon(u) = \{u\} \cup (\mathbb{Q} \cap (u - \epsilon, u + \epsilon)).$$
These sets form a basis of neighbourhoods of u in X_0. Note that $U_\epsilon(u^*)$ is usually different from $U_\epsilon(u)^*$, but nonetheless the set
$$V_\epsilon(u) = U_\epsilon(u) \cup U_\epsilon(u^*)$$
is the same as $U_\epsilon(u) \cup U_\epsilon(u)^*$ and so is invariant under $t \mapsto t^*$. These sets form a basis of $*$-invariant neighbourhoods of $\{u, u^*\}$. Moreover, if $\pi(u) \neq \pi(v)$ in X then $\{u, u^*\}$ and $\{v, v^*\}$ will be disjoint, so $V_\epsilon(u)$ and $V_\delta(v)$ will be disjoint when ϵ and δ are sufficiently small. Using this, we see that X is Hausdorff.

To prove that X is connected, we need a more precise version of the above argument. Let $\eta(u, v)$ denote the minimum possible distance between a member of $\{u, u^*\}$ and a member of $\{v, v^*\}$. It is not hard to see that $V_\epsilon(u) \cap V_\delta(v) \neq \emptyset$ iff $\epsilon + \delta < \eta(u, v)$. By letting δ tend to zero, we see that $\pi(v)$ lies in the closure of $\pi(V_\epsilon(u))$ iff $\eta(u, v) \leq \epsilon$. Thus, the preimage of the closure of $\pi(V_\epsilon(u))$ is the set
$$F_\epsilon(u) = \{v \ : \ \eta(u, v) \leq \epsilon\}.$$

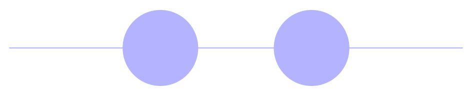
We next claim that for all points $u = a + b\sqrt{2}$ and $v = c + d\sqrt{2}$ in X_0, and all $\epsilon > 0$, the set $F_\epsilon(u) \cap F_\epsilon(v)$ is nonempty.

To see this, choose rational numbers p, q with $|p - (u+v)/2| < \epsilon/2$ and $|q\sqrt{2} - (u-v)/2| < \epsilon/2$. We then put $z = p + q\sqrt{2} \in X_0$, and we find that $|z - u| < \epsilon$ and $|z^* - v| < \epsilon$, so $z \in F_\epsilon(u) \cap F_\delta(v)$ as required.

Now suppose we have a separation (A, B) of X, so the sets $A_0 = \pi^{-1}(A)$ and $B_0 = \pi^{-1}(B)$ give a $*$-invariant separation of X_0. Suppose for a contradiction that both A_0 and B_0 are nonempty, say with $u \in A_0$ and $v \in B_0$. As A_0 and B_0 are open and invariant, we can choose $\epsilon > 0$ such that $V_\epsilon(u) \subseteq A_0$ and $V_\epsilon(v) \subseteq B_0$. As A_0 and B_0 are also closed, this gives $F_\epsilon(u) \subseteq A_0$ and $F_\epsilon(v) \subseteq B_0$, which means that $F_\epsilon(u) \cap F_\epsilon(v) = \emptyset$, contrary to what we proved above. Thus, the separation must actually be trivial, as required.

Exercise 7.1. [`ex-clconn`]
Find a connected subset $X \subseteq \mathbb{R}^n$ (for some n) such that $\mathrm{int}(X)$ is not connected.

Solution: Let $X \subseteq \mathbb{R}^2$ be the union of the x-axis with two open discs of radius $1/2$ centred at $(-1, 0)$ and $(1, 0)$.



Then X is connected but the interior of X is just the union of the two discs, which is not connected.

Exercise 7.2. [`ex-puncture`]
Suppose that $X \subseteq \mathbb{R}^2$ is connected and $x \in \mathrm{int}(X)$. Prove that $X \setminus \{x\}$ is connected.

Solution: As $x \in \mathrm{int}(X)$ we have a small open disc D centred at x and contained in X. Put $X' = X \setminus \{x\}$ and $D' = D \setminus \{x\}$, and note that D' is connected. Suppose that U and V are open in $\mathbb{R}^2$ and have $X' \subseteq U \cup V$ and $U \cap V \cap X' = \emptyset$. It follows that $D' \subseteq U \cup V$ and $D' \cap U \cap V = \emptyset$. As D' is connected we must have $D' \subseteq U$ or $D' \subseteq V$, and we may assume without loss of generality that $D' \subseteq U$. It follows that the set $U^* = U \cup \{x\}$ is open. Next, as $D' \cap U \cap V = \emptyset$ we deduce that $D' \cap V = \emptyset$, so V cannot contain any open disc centred at x. As V is open it follows that $x \notin V$. We now have $X \subseteq U^* \cup V$ with $U^* \cap V \cap X = \emptyset$ and $U^* \cap X \neq \emptyset$. As X is connected it follows that $X \subseteq U^*$, and so $X' \subseteq U$. This implies that X' is connected, as claimed.

Exercise 7.3. [`ex-quasi`]
Let X be a topological space. Define a relation F on X by xFy iff (there is no separation $X = A \cup B$ into disjoint open sets such that $x \in A$ and $y \in B$). Prove that this is an equivalence relation. We will call the equivalence classes *quasicomponents*. Show that each quasicomponent is closed. Show that each component is contained in a quasicomponent.

Solution: We need to prove that the relation F is reflexive, symmetric and transitive. The first two are immediate. For transitivity, suppose that xFy and yFz. Suppose that $X = A \cup B$ is a separation into disjoint open sets. By assumption either $x, y \in A$ or $x, y \in B$, and also either $y, z \in A$ or $y, z \in B$. On the other hand, A and B are disjoint so it cannot happen that $y \in A$ and $y \in B$; thus the only possibilities are $x, y, z \in A$ or $x, y, z \in B$. In either case, x and z lie in the same half of the partition. Thus xFz as required.

The quasicomponent C containing x is the set of points y such that for every open and closed set A containing x, we also have $y \in A$. In other words,

$$C = \bigcap \{A \ : \ x \in A \text{ and } A \text{ is open and closed}\}$$

This is the intersection of a family of closed sets, hence is closed.

Now write xEy if there is a connected set containing x and y, so the E-equivalence classes are by definition the components. Suppose that xFy, say $x, y \in Z$ with $Z \subseteq X$ connected. Consider a separation $X = A \cup B$ as before. Then the separation $Z = (Z \cap A) \cup (Z \cap B)$ is trivial, so wlog $Z \cap B = \emptyset$ and so $Z \subseteq A$. Thus $x, y \in A$. As this happens for every separation $X = A \cup B$, we see that xFy. It follows that the component $D = \{y \ : \ yEx\}$ containing x is a subset of the quasicomponent $C = \{y \ : \ yFx\}$. Thus every component is contained in a quasicomponent, as claimed.

Exercise 7.4. [ex-totdis]
A space X is said to be *totally disconnected* if the only connected subsets are single points.

(a) Prove that $\mathbb{Q}$ (considered as a subspace of $\mathbb{R}$) is totally disconnected.
(b) Fix a prime p, and consider $\mathbb{Z}$ with the p-adic topology, as in Example 2.26. Prove that this is also totally disconnected.

Solution:

(a) Let $Y \subseteq \mathbb{Q}$ be connected. By definition this means that Y is nonempty, so we can choose $y \in Y$. We claim that $Y = \{y\}$. If not, then there is some $z \in Y$ with $z \neq y$ and then the number $x = y + (z-y)/\sqrt{2}$ is irrational and lies strictly between y and z. This means that $(-\infty, x) \cap Y$ and $(x, \infty) \cap Y$ form a nontrivial separation of Y, contrary to the assumption. Thus $\mathbb{Q}$ is totally disconnected as claimed.
(b) Now consider $\mathbb{Z}$ with the p-adic topology. For any $k \geq 0$ and any $a \in \{0, 1, \dots, p^k - 1\}$ the set $U_{k,a} = \{m : m = a \pmod{p^k}\}$ is then a (basic) open set. It is easy to see that $U_{k,a}^c$ is the union of the open sets $U_{k,b}$ for $b \in \{0, \dots, p^k - 1\}$ with $b \neq a$. This means that $U_{k,a}^c$ is open, so $U_{k,a}$ is closed as well as being open. Now suppose we have a connected set $Y \subseteq X$, so we can choose a point $y \in Y$. We claim that $Y = \{y\}$. If not, there is some point $z \in Y$ with $0 \neq z - y \in \mathbb{Z}$, so for sufficiently large k we see that $z - y$ is not divisible by p^k. This means that there is some set $U_{k,a}$ with $y \in U_{k,a}$ and $z \in U_{k,a}^c$. This means that $(U_{k,a}, U_{k,a}^c)$ gives a nontrivial separation of Y, contrary to hypothesis. The claim follows.

8. Path Connectedness

Definition 8.1. [defn-pi-zero]

(a) A *path from x_0 to x_1* in a space X is a continuous map $u \colon [0,1] \to X$ such that $u(0) = x_0$ and $u(1) = x_1$.
(b) For any point $x \in X$, we write c_x for the constant map $[0,1] \to X$ with value x, so c_x is a path from x to x.
(c) If u is a path from x_0 to x_1 then we define $\overline{u}(t) = u(1-t)$, which gives a path from x_1 to x_0, called the *reverse* of u.
(d) If v is another path that runs from x_1 to x_2, we define $v * u \colon [0,1] \to X$ by
$$(v * u)(t) = \begin{cases} u(2t) & \text{if } 0 \leq t \leq 1/2 \\ v(2t-1) & \text{if } 1/2 \leq t \leq 1. \end{cases}$$
This is continuous by Proposition 5.9 applied to the sets $[0, 1/2]$ and $[1/2, 1]$, so it gives a path from x_0 to x_2.
(e) Write xEy iff there exists a path from x to y in X. Parts (b) to (d) show that this is an equivalence relation. The equivalence classes are called *path components*. We write $\pi_0(X)$ for the set X/E of path components.
(f) We say that X is *path connected* if it is nonempty and any two points can be joined by a path, so there i precisely one path component, namely X itself.

Note that in the definition of $v * u$, we use u first, then v. This is consistent with the idea that a path from x_0 to x_1 is like a map from x_0 to x_1, and that joining paths is analogous to composition. This idea will be explored further in Section 28.

Proposition 8.2. [prop-path-connected]
Any path connected space is connected.

Proof. Let X be a path connected space, and let (U, V) be a separation of X. By definition, a path connected space must be nonempty, so we can choose a point $a \in X$, and we may assume that $a \in U$. For any other point $x \in X$, we can choose a path u from a to x. Now the sets $u^{-1}(U)$ and $u^{-1}(V)$ give a separation of the connected space $[0,1]$ with $0 \in u^{-1}(U)$, so we must have $u^{-1}(U) = [0,1]$ and $u^{-1}(V) = \emptyset$. In particular we have $1 \in u^{-1}(U)$ so $x = u(1) \in U$. As x was arbitrary this gives $U = X$, so the separation is trivial as required.

□

Example 8.3. [eg-convex-connected]
We say that a subset $X \subseteq \mathbb{R}^n$ is *convex* if whenever $a, b \in X$ and $0 \leq t \leq 1$ the point $(1-t)a + tb$ also lies in X. If so, we have a path from a to b given by $u(t) = (1-t)a + tb$, showing that X is path connected, and thus connected (provided that $X \neq \emptyset$).

Lemma 8.4. [lem-convex-classify]
The convex subsets of $\mathbb{R}$ are the sets of the following types:

(a) *The empty set, single points and $\mathbb{R}$ itself.*
(b) *Semi-infinite intervals $(-\infty, a)$, $(-\infty, a]$, $[a, \infty)$ and (a, ∞) for $a \in \mathbb{R}$.*
(c) *Finite intervals (a, b), $[a, b)$, $(a, b]$ and $[a, b]$ where $a, b \in \mathbb{R}$ and $a < b$.*

Proof. Let $X \subseteq \mathbb{R}$ be convex. First note that if $p, q \in X$ and $p < x < q$ we can put $t = (x-p)/(q-p)$ and we find that $0 < t < 1$ and $x = (1-t)p + tq$, so $x \in X$ by convexity.

Now suppose for the moment that X is nonempty and bounded above and below. We then have numbers $a = \inf(X)$ and $b = \sup(X)$ with $a \leq b$, so $X \subseteq [a, b]$. If $a = b$ then we must have $X = \{a\}$. Suppose instead that $a < b$. Consider a point $x \in (a, b)$. Then x is strictly less than $b = \sup(X)$, so x is not an upper bound for X, so there exists $q \in X$ with $x < q \leq b$. Similarly, there exists $p \in X$ with $a \leq p < x$. As $p < x < q$ with $p, q \in X$ we see that $x \in X$. This proves that $(a, b) \subseteq X \subseteq [a, b]$, so X has one of the types in (c).

Now suppose instead that X has neither an upper bound nor a lower bound. For any $x \in \mathbb{R}$, we note that x is not an upper bound for X, so there exists $q \in X$ with $x < q$. Similarly there exists $p \in X$ with $p < x$. As $p < x < q$ with $p, q \in X$ we see that $x \in X$. As x was arbitrary we see that $X = \mathbb{R}$, which is covered by case (a).

Finally, suppose that X has an upper bound but not a lower bound, or vice-versa. By a mixture of the methods above, we see that X is covered by case (b). □

Proposition 8.5. [prop-convex-open]
Let $U \subseteq \mathbb{R}$ be a nonempty convex open set, and let $f\colon U \to \mathbb{R}$ be a continuous injective map. Then $f(U)$ is also convex and open, and $f\colon U \to f(U)$ is a homeomorphism. Moreover, f is either strictly increasing or strictly decreasing.

Proof. As U is nonempty and open, we can choose $a, b \in U$ with $a < b$. As f is injective we have $f(a) \neq f(b)$. As $-f$ is also continuous and injective, we can reduce to the case where $f(a) < f(b)$. Now suppose we have another pair of points $x, y \in U$ with $x < y$. Put $u(t) = (1-t)a + tx$ and $v(t) = (1-t)b + ty$ for $0 \leq t \leq 1$. By convexity, we have $u(t), v(t) \in U$, and it is also clear that $u(t) < v(t)$ for all t. Now put $w(t) = f(v(t)) - f(u(t))$, and note that $w(0) = f(b) - f(a) > 0$. As $u(t) < v(t)$ and f is injective we see that $w(t) \neq 0$ for all t, and it follows by the Intermediate Value Theorem that $w(1) > 0$, so $f(x) < f(y)$. Thus, the map f is strictly increasing.

Now suppose again that we have $x, y \in U$ with $x < y$. If $p \in \mathbb{R}$ with $f(x) \leq p \leq f(y)$ then the Intermediate Value Theorem tells us that there exists $z \in [x, y]$ such that $f(z) = p$, so $f([x, y]) \supseteq [f(x), f(y)]$. On the other hand, as f is strictly increasing we have $f([x, y]) \subseteq [f(x), f(y)]$, so $f([x, y]) = [f(x), f(y)]$. As all points of the form $(1-t)f(x) + tf(y)$ (for $t \in [0, 1]$) lie in $[f(x), f(y)]$, we deduce that $f(U)$ is convex. As f is injective it follows that $f((x, y)) = (f(x), f(y))$. Now, every open subset $V \subseteq U$ (including U itself) can be written as a union of such open intervals (x, y), so $f(V)$ is the union of the corresponding intervals $(f(x), f(y))$, so $f(V)$ is open. It follows that $f(U)$ is open and that $f\colon U \to f(U)$ is a continuous bijection and an open map, and therefore a homeomorphism. □

Proposition 8.6. [prop-product-path-connected]
Let $(X_i)_{i \in I}$ be a family of path connected spaces. Then the product $X = \prod_I X_i$ is also path connected.

Proof. Let x and y be points in X. As X_i is path connected, we can choose a path u_i from x_i to y_i in X_i. We then define $u(t) = (u_i(t))_{i \in I}$. This gives a map $u\colon [0, 1] \to X$, which is continuous by Proposition 5.16, giving a path from x to y. □

Proposition 8.7. [prop-image-path-connected]

Let $f\colon X \to Y$ be a continuous map, and let X' be a path connected subset of X. Then the image $Y' = f(X')$ is also path connected.

Proof. Suppose we have points $y_0, y_1 \in Y'$. As $Y' = f(X')$ we can find points $x_0, x_1 \in X'$ with $f(x_i) = y_i$. As X' is path connected we can find a map $u\colon [0,1] \to X'$ with $u(0) = x_0$ and $u(1) = x_1$. Now put $v = f \circ u\colon [0,1] \to Y'$; we find that this gives the required path from y_0 to y_1. $\square$

Example 8.8. [eg-Sn-connected]

For $n > 0$ the sphere $S^n = \{x \in \mathbb{R}^n \;:\; \|x\| = 1\}$ is path connected (and thus connected). To see this, consider points $x, y \in S^n$ with $y \neq -x$. Define $u\colon [0,1] \to \mathbb{R}^{n+1}$ by $u(t) = (1-t)x + ty$. As $y \neq -x$, we claim that this does not pass through zero. This should be clear geometrically. For an algebraic proof, one can check by expanding everything out that

$$4\|u(t)\|^2 = (2t-1)^2\|x-y\|^2 + \|x+y\|^2 \geq \|x+y\|^2,$$

which implies the claim. We can thus define a path v from x to y in S^n by $v(t) = u(t)/\|u(t)\|$. For the exceptional case where $y = -x$, we simply choose a point $z \in S^n \setminus \{x, -x\}$ (which is possible because $n > 0$); then the general case gives paths from x to z and from z to $-x$, which we can join to give a path from x to $-x$.

Example 8.9. [eg-RPn-connected]

In Example 5.24 we exhibited a continuous surjection from S^n to $\mathbb{R}P^n$. Using this and Proposition 8.7, we see that $\mathbb{R}P^n$ is also path connected.

Example 8.10. [eg-GLnR-connected]

Consider the space $GL_n(\mathbb{R})$ and the open subspaces

$$GL_n^{\pm}(\mathbb{R}) = \{A \in GL_n(\mathbb{R}) \;:\; \pm\det(A) > 0\}.$$

We claim that these are both path connected. As a first step, let U be the path component containing the identity matrix, so $U \subseteq GL_n^+(\mathbb{R})$. If u is a path from I to A and v is a path fom I to B then the map $t \mapsto u(t)^{-1}$ gives a path from I to A^{-1} and the map $t \mapsto u(t)v(t)$ gives a path from I to AB. It follows that U is a subgroup of $GL_n^+(\mathbb{R})$.

Now let $e_1, \ldots, e_n$ be the standard basis vectors for $\mathbb{R}^n$, so we can define a matrix A by describing the columns Ae_i. Consider the following matrices:

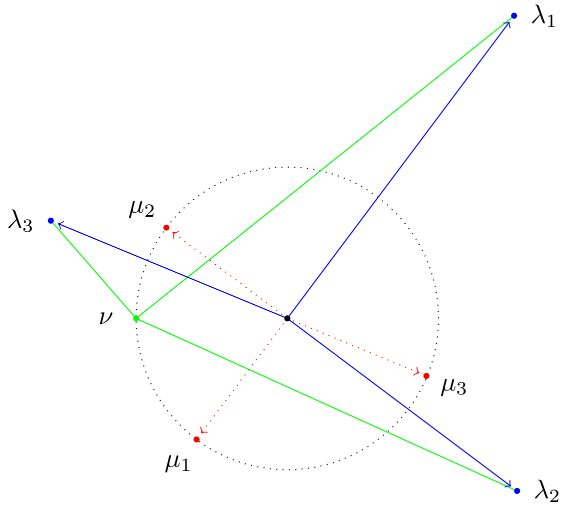
(a) For $1 \leq i \leq n$ and $t > 0$ we define $D_i(t)e_i = te_i$, and $D_i(t)e_k = e_k$ for $k \neq i$.
(b) For $1 \leq i, j \leq n$ and $t \in \mathbb{R}$ we define $E_{ij}(t)e_i = e_i + te_j$ and $E_{ij}(t)e_k = e_k$ for $k \notin \{i, j\}$.
(c) For $1 \leq i, j \leq n$ and $0 \leq t \leq \pi$ we define $R_{ij}(t)e_i = \cos(t)e_i + \sin(t)e_j$ and $R_{ij}(t)e_j = -\sin(t)e_i + \cos(t)e_j$ and $R_{ij}(t)e_k = e_k$ for $k \notin \{i, j\}$.

It is visible from the definitions that these lie in U. Now let A be an arbitrary matrix in $GL_n^+(\mathbb{R})$. We find that:

(p) $E_i(t)A$ is the result of multiplying the i'th row by t.
(q) $E_{ij}(t)A$ is the result of adding t times the i'th row to the j'th row.
(r) $R_{ij}(\pi/2)A$ is the result of exchanging the i'th and j'th rows and multiplying one of them by minus one.
(s) $R_{ij}(\pi)A$ is the result of multiplying the i'th and j'th rows by minus one.

By a minor adjustment of the standard row-reduction algorithm, we can perform a sequence of operations of types (p), (q) and (r) to convert A to a diagonal matrix A' in which all the diagonal entries are ± 1. None of these operations change the sign of the determinant, and $\det(A) > 0$ so we must have $\det(A') > 0$, so the number of -1's on the diagonal must be even. We can thus perform operations of type (s) to convert A' to I. Because all of these operations are given by left-multiplying by matrices in U, we find that there exists $B \in U$ with $BA = I$, so $A = B^{-1}$. As U is a subgroup we deduce that $A \in U$. As A was an arbitrary element of $GL_n^+(\mathbb{R})$ we deduce that $GL_n^+(\mathbb{R}) = U$, so $GL_n^+(\mathbb{R})$ is path connected as claimed. Now let J be the matrix with $Je_1 = -e_1$ and $Je_k = e_k$ for $k > 1$. It is then clear that the map $A \mapsto JA$ gives a homeomorphism $GL_n^+(\mathbb{R}) \to GL_n^-(\mathbb{R})$ (with inverse again given by $A \mapsto JA$) so $GL_n^-(\mathbb{R})$ is also path connected.

EXAMPLE 8.11. [eg-GLnC-connected]
Now consider the space $GL_n(\mathbb{C}) = \{A \in M_n(\mathbb{C}) \;:\; \det(A) \neq 0\}$. We claim that this is path connected (and thus connected). This could be proved by a similar method to that used for $GL_n^+(\mathbb{R})$, but we will instead explain a rather different approach. Consider a point $A \in GL_n(\mathbb{C})$, and let $\lambda_1, \dots, \lambda_r$ be the eigenvalues of A. As A is invertible, these are all nonzero.



Put $\mu_i = -\lambda_i/|\lambda_i| \in S^1$, and then choose any point $\nu \in S^1$ different from $\mu_1, \dots, \mu_r$. Put $u(t) = t\nu I + (1-t)A \in M_n(\mathbb{C})$ (for $0 \leq t \leq 1$). The eigenvalues of $u(t)$ are the numbers $t\nu + (1-t)\lambda_i$, which lie on the line segments joining λ_i to ν. Our choice of ν ensures that these line segments do not pass through the origin, so $u(t)$ is invertible, so we have defined a path $u\colon [0,1] \to GL_n(\mathbb{C})$ joining A to νI. Now choose θ such that $\nu = \exp(i\theta)$, and define $v(t) = \exp(i(1-t)\theta)I$; this gives a path joining νI to I. We conclude that any point $A \in GL_n(\mathbb{C})$ can be joined to the identity, and it follows that $GL_n(\mathbb{C})$ is path connected as claimed.

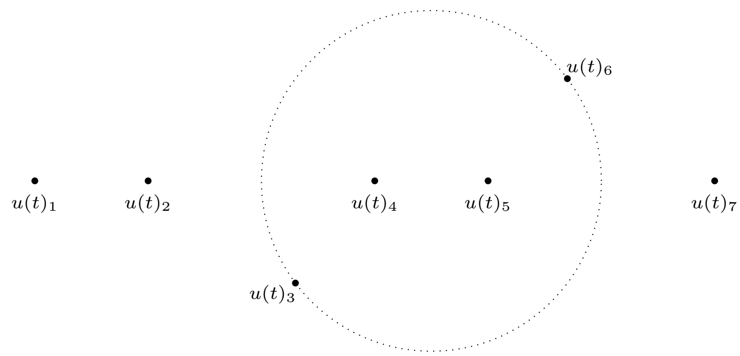
EXAMPLE 8.12. [eg-FnC-connected]
Consider the space

$$F_n(\mathbb{C}) = \{z \in \mathbb{C}^n \;:\; z_i \neq z_j \text{ for all } i \neq j\}.$$

Put $e_1 = (1, 2, 3, \dots, n) \in F_n(\mathbb{C})$, and let E_1 be the path component of e_1. We will show that E_1 is all of $F_n(\mathbb{C})$, so $F_n(\mathbb{C})$ is path connected. Indeed, for any permutation σ of $\{1, \dots, n\}$ we can put $e_\sigma = (\sigma(1), \dots, \sigma(n))$, so $e_\sigma \in F_n(\mathbb{C})$. Consider a transposition $\tau = (p\ q)$. Put $a = (\sigma(p) + \sigma(q))/2$ and $b = (\sigma(p) - \sigma(q))/2$, and define $u\colon [0,1] \to \mathbb{C}^n$ by

$$u(t)_k = \begin{cases} \sigma(k) & \text{if } k \notin \{p, q\} \\ a + e^{\pi i t} b & \text{if } k = p \\ a - e^{\pi i t} b & \text{if } k = q. \end{cases}$$

The following picture illustrates the case where $n = 7$, σ is the identity permutation, $p = 6$, $q = 3$ and $t = 0.2$.



We find that when $j \neq k$ we have $u(t)_j \neq u(t)_k$, so $u(t) \in F_n(\mathbb{C})$ for all t. We also have $u(0) = e_\sigma$ and $u(1) = e_{\sigma\tau}$, so e_σ and $e_{\sigma\tau}$ lie in the same path component. As every permutation can be written as a product of transpositions, we see that all the points e_σ lie in E_1.

Next, put

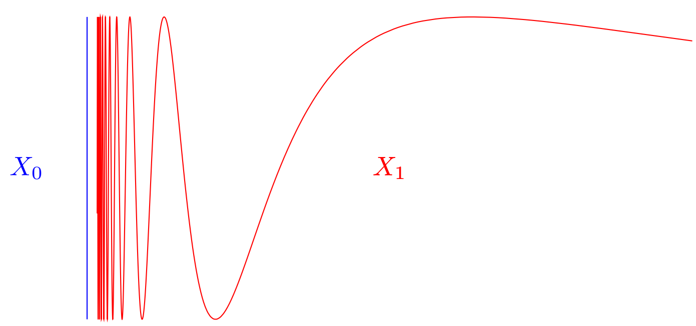
$$U_\sigma = \{z \in \mathbb{C}^n \; : \; \mathrm{Re}(z_{\sigma(1)}) < \mathrm{Re}(z_{\sigma(2)}) < \cdots < \mathrm{Re}(z_{\sigma(n)})\}.$$

We find that U_σ is open and convex and contains the point e_σ, so every point in U_σ can be connected by a linear path to e_σ and so lies in E_1.

Now consider a general point $z \in F_n(\mathbb{C})$. This gives us a finite collection of points $i(z_j - z_k)/|z_j - z_k| \in S^1$ for $1 \leq j, k \leq n$ with $j \neq k$. Choose θ such that $e^{i\theta}$ is not one of these points. This means that none of the numbers $e^{-i\theta} z_j - e^{-i\theta} z_k$ is purely imaginary, so the numbers $e^{-i\theta} z_1, \ldots, e^{-i\theta} z_n$ have distinct real parts, so $e^{-i\theta} z \in U_\sigma$ for some σ, so $e^{-i\theta} z \in E_1$. The path $v(t) = e^{-it\theta} z$ connects z to $e^{-i\theta} z$ in $F_n(\mathbb{C})$, so $z \in E_1$ as claimed.

Example 8.13. [`eg-wild-sin`]
Consider the spaces

$$X_0 = \{(0, y) \; : \; -1 \leq y \leq 1\}$$
$$X_1 = \{(x, \sin(1/x)) \; : \; x > 0\}$$
$$X = X_0 \cup X_1 \subset \mathbb{R}^2.$$



We claim that X is connected but not path connected. Indeed, we have a continuous surjection $(0, \infty) \to X_1$ given by $x \mapsto (x, \sin(1/x))$, so X_1 is connected, and X is the closure of X_1 in $\mathbb{R}^2$, so X is connected. Now consider a path $u \colon [0,1] \to X$, given by $u(t) = (v(t), w(t))$ say. Put $A = v^{-1}\{0\}$, which is closed in $[0,1]$. We claim that it is also open. Indeed, suppose $v(t) = 0$, and suppose for the moment that $w(t) < 1$. As w is continuous there is some $\epsilon > 0$ such that $w(s) < 1$ for s in the set $N = (t - \epsilon, t + \epsilon) \cap [0,1]$. If $v(s) = 2/((4n+1)\pi)$ for some $n \geq 0$ we see that $w(s) = \sin(1/v(s)) = 1$, which cannot happen for $s \in N$. It follows that $v(N)$ is contained in the set

$$V = [0, \infty) \setminus \{2/((4n+1)\pi) \; : \; n \in \mathbb{N}\}.$$

Moreover, N is connected and contains t, so $v(N)$ is connected and contains $v(t) = 0$. It is clear that the only connected subset of V containing 0 is $\{0\}$, so $v(N) = 0$, so $N \subseteq A$, so t is an interior point of A. This argument does not work if $w(t) = 1$, but in that case we can use a very similar argument based on the inequality $w(t) > -1$. We therefore see that A is both open and closed, as claimed. This means that A and A^c give a separation of the connected space $[0,1]$, so either $A = \emptyset$ or $A = [0,1]$. This means that no path can cross between X_0 and X_1, so X is not path connected.

We next explain how to consider π_0 as a functor. All the relevant concepts are reviewed in Appendix 36.

PROPOSITION 8.14. [`prop-pi-zero-functor`]
*The construction π_0 can be made into a functor **Spaces** $\to$ **Sets**, in such a way that $\pi_0(f)([x]) = [f(x)]$ for all continuous maps $f\colon X \to Y$ and all points $x \in X$. Moreover, this functor preserves all products and coproducts.*

In this context we will typically write f_* rather than $\pi_0(f)$ (as in Remark 36.36).

PROOF. For each object $X \in \mathbf{Spaces}$, we have already defined $\pi_0(X) \in \mathbf{Sets}$. Now suppose we have a continuous map $f\colon X \to Y$. If $[x_0] = [x_1]$ in $\pi_0(X)$, then there exists a path $u\colon [0,1] \to X$ joining x_0 to x_1, and then the path $f \circ u\colon [0,1] \to Y$ joins $f(x_0)$ to $f(x_1)$, so $[f(x_0)] = [f(x_1)]$ in $\pi_0(Y)$. We thus have a well-defined map $f_*\colon \pi_0(X) \to \pi_0(Y)$ given by $f_*([x]) = [f(x)]$. If $f = 1_X$ it is clear from this that $f_* = 1_{\pi_0(X)}$. Similarly, if we have a second map $g\colon Y \to Z$ then

$$(gf)_*([x]) = [g(f(x))] = g_*[f(x)] = g_*(f_*([x]))$$

so $(gf)_* = g_*f_*\colon \pi_0(X) \to \pi_0(Z)$. We have thus defined a functor, as claimed.

Now consider a family of spaces $(X_i)_{i\in I}$. Put $P = \prod_i X_i$, and let $p_i\colon P \to X_i$ be the i'th projection. (This would normally be denoted by π_i, but we wish to avoid confusion with the notation π_0 for the set of path components.) Note that p_i induces a map $(p_i)_*\colon \pi_0(P) \to \pi_0(X_i)$. We can put these together to define a map $\phi\colon \pi_0(P) \to \prod_{i\in I} \pi_0(X_i)$ by $\phi(a) = ((p_i)_*(a))_{i\in I}$. We claim that ϕ is a bijection. Indeed, suppose we have an element $b \in \prod_i \pi_0(X_i)$, or equivalently a family of elements $b_i \in \pi_0(X_i)$ for $i \in I$. We can then choose elements $x_i \in X_i$ such that $b_i = [x_i]$, and put $x = (x_i)_{i\in I} \in P$ and $a = [x] \in \pi_0(P)$. We then have $(p_i)_*(a) = (p_i)_*([x]) = [p_i(x)] = [x_i] = b_i$ for all i, so $\phi(a) = (b_i)_{i\in I} = b$. This shows that ϕ is surjective. Now suppose we have two elements $a, a' \in \pi_0(P)$ with $\phi(a) = \phi(a')$. We can then choose $x, x' \in X$ with $a = [x]$ and $a' = [x']$. Here x is a family $(x_i)_{i\in I}$ of points $x_i \in X_i$, and x' is a family $(x'_i)_{i\in I}$ of points $x'_i \in X_i$. We have $\phi(a) = ([x_i])_{i\in I}$, and this is the same as $\phi(a') = ([x'_i])_{i\in I}$. We must therefore have $[x_i] = [x'_i]$ for all i, so we can choose a path $u_i\colon [0,1] \to X_i$ with $u_i(0) = x_i$ and $u_i(1) = x'_i$. We can then use these to define a path $u\colon [0,1] \to P$ by $u(t) = (u_i(t))_{i\in I}$; this is continuous by Proposition 5.16. We have $u(0) = x$ and $u(1) = x'$, so we have $a = [x] = [x'] = a'$. This proves that ϕ is also injective, so it is a bijection. This means that the functor π_0 preserves products as claimed.

Now put $Q = \coprod_i X_i$, and consider instead the set $\pi_0(Q)$. If $x \in X_i$ we have a point $[x] \in \pi_0(X_i)$ and thus a point $(i,[x]) \in \coprod_i \pi_0(X_i)$. On the other hand, we also have $(i,x) \in Q$ and thus $[(i,x)] \in \pi_0(Q)$. If $(i,[x]) = (i,[y])$ then we can choose a path $u\colon [0,1] \to X_i$ joining x to y, and this gives a path $t \mapsto (i,u(t))$ joining (i,x) to (i,y), so $[(i,x)] = [(i,y)]$. Conversely, suppose we have a path $v\colon [0,1] \to Q$ joining (i,x) to (j,y). The set $\{i\} \times X_i$ is both open and closed in Q, so the preimage $A = u^{-1}(\{i\} \times X_i)$ is both open and closed in $[0,1]$. It also contains zero and so is nonempty. As $[0,1]$ is connected we must have $A = [0,1]$, so $v(t) \in \{i\} \times X_i$ for all t. This means that $i = j$ and $y \in X_i$ and that there is a continuous map $u\colon [0,1] \to X_i$ such that $v(t) = (i,u(t))$ for all t. In particular, u joins x to y, so $(i,[x]) = (i,[y]) = (j,[y])$. We thus have a bijection $\psi\colon \coprod_i \pi_0(X_i) \to \pi_0(Q)$ given by $\psi(i,[x]) = [(i,x)]$, which means that π_0 also preserves coproducts. $\square$

8.1. Non-homeomorphism results. [`subsec-non-homeo`]

As mentioned previously, it is often hard to prove that two spaces are not homeomorphic, even in cases where this seems to be clear. However, we can use the functor π_0 to prove some results of this type. The most obvious approach is this: if $f\colon X \to Y$ is a homeomorphism, then functoriality means that $f_*\colon \pi_0(X) \to \pi_0(Y)$ is a bijection, so $|\pi_0(X)| = |\pi_0(Y)|$. Thus, if we have spaces X and Y with $|\pi_0(X)| \neq |\pi_0(Y)|$, then they cannot be homeomorphic.

EXAMPLE 8.15. If $X = \mathbb{R} \setminus \mathbb{Z} = \bigcup_{n\in\mathbb{Z}} (n, n+1)$ and $Y = \mathbb{R} \setminus \{0\}$ then $\pi_0 X$ is infinite but $|\pi_0 Y| = 2$ so X is not homeomorphic to Y.

Example 8.16. We saw in Example 8.10 that $GL_n(\mathbb{R})$ is not path-connected, but $M_n(\mathbb{R})$ is a vector space and thus is path-connected, so $GL_n(\mathbb{R})$ is not homeomorphic to $M_n(\mathbb{R})$.

This method is unfortunately inadequate to prove many other visually obvious facts such as that S^1 is not homeomorphic to $\mathbb{R}$ or that $\mathbb{R}^2$ is not homeomorphic to $\mathbb{R}^3$. At least the first of these can be proved by a small adaptation of the method, however.

Proposition 8.17. *S^1 is not homeomorphic to $\mathbb{R}$.*

Proof. Suppose for a contradiction that $f\colon \mathbb{R} \to S^1$ is a homeomorphism. Put $a = f(0) \in S^1$. Because f is injective we have $f(t) \neq a$ when $t \neq 0$ so f gives a continuous map $f\colon \mathbb{R} \setminus \{0\} \to S^1 \setminus \{a\}$. Similarly, f^{-1} gives a continuous map $f^{-1}\colon S^1 \setminus \{a\} \to \mathbb{R} \setminus \{0\}$. These maps are clearly inverse to each other, so $\mathbb{R} \setminus \{0\}$ is homeomorphic to $S^1 \setminus \{a\}$. This is a contradiction, because it is easy to see that $S^1 \setminus \{a\}$ is path-connected but $\mathbb{R} \setminus \{0\}$ is not. □

More generally, if X is homeomorphic to Y and $a_1, \ldots, a_n$ are n distinct points in X then there exist n distinct points $b_1, \ldots, b_n$ in Y such that $X \setminus \{a_1, \ldots, a_n\}$ is homeomorphic to $Y \setminus \{b_1, \ldots, b_n\}$. Indeed, if $f\colon X \to Y$ is a homeomorphism, we can just take $b_i = f(a_i)$. Using this, we can prove a number of other non-homeomorphism results.

Proposition 8.18. *$[0, 1]$ is not homeomorphic to S^1.*

Proof. If $[0, 1]$ were homeomorphic to S^1, then $(0, 1) = [0, 1] \setminus \{0, 1\}$ would be homeomorphic to $S^1 \setminus \{b_1, b_2\}$ for some b_1, b_2. This is a contradiction, because $(0, 1)$ is path-connected whereas $S^1 \setminus \{b_1, b_2\}$ is always disconnected for any pair $\{b_1, b_2\}$ of distinct points. □

Proposition 8.19. *$\mathbb{R}$ is not homeomorphic to $\mathbb{R}^2$, because $\mathbb{R}$ is disconnected by the removal of any point, whereas there is no point in $\mathbb{R}^2$ whose removal disconnects the space. Similarly, there are precisely two points whose removal fails to disconnect $[0, 1]$, but every point in $[0, 1]^2$ has that property, so $[0, 1]$ is not homeomorphic to $[0, 1]^2$.* □

Proposition 8.20. *$(0, 1)$ is not homeomorphic to $[0, 1)$, because $[0, 1) \setminus \{0\}$ is connected and thus not homeomorphic to $(0, 1) \setminus \{b\}$ for any b.* □

Proposition 8.21. *Put $X = \{(x, y) \in \mathbb{R}^2 \;:\; xy = 0\} = (\mathbb{R} \times \{0\}) \cup (\{0\} \times \mathbb{R})$. Then X is not homeomorphic to $\mathbb{R}$, because $X \setminus \{(0, 0)\}$ has four path components and thus is not homeomorphic to $\mathbb{R} \setminus \{b\}$ for any b.* □

These ideas can be extended to give some numerical invariants of spaces.

Definition 8.22. Let X be a metric space. We write

$$\begin{aligned} a(X) &= \max\{|Y| \;:\; Y \subseteq X \text{ and } X \setminus Y \text{ is path-connected }\} \\ &= \text{ the greatest number of points that can be removed without disconnecting } X \\ b(X) &= \min\{|Y| \;:\; Y \subseteq X \text{ and } X \setminus Y \text{ is disconnected }\} \\ &= \text{ the least number of points that need to be removed to disconnect the space } X. \end{aligned}$$

These invariants can be infinite: for example $a(\mathbb{R}^2) = \infty$, because the plane remains connected after the removal of any finite set of points. However, we will principally be interested in cases in which they are finite.

Proposition 8.23. *If X is homeomorphic to X' then $a(X) = a(X')$ and $b(X) = b(X')$.*

Proof. Let $f\colon X \to X'$ be a homeomorphism. If $Y \subseteq X$ and $X \setminus Y$ is path-connected then we put $Y' = f(Y) \subseteq X'$. As f is a bijection we have $|Y'| = |Y|$. We also note that f gives a homeomorphism $X \setminus Y \to X' \setminus Y'$, so $X' \setminus Y'$ is path-connected. We must therefore have $a(X') \geq |Y'| = |Y|$. By taking the maximum over all possible Y's, we see that $a(X') \geq a(X)$. By applying this argument to $f^{-1}\colon X' \to X$ instead of $f\colon X \to X'$, we also see that $a(X) \geq a(X')$. This means that $a(X) = a(X')$. The argument for b is similar. □

Now consider the letters of the alphabet (drawn with infinitely thin lines) as subsets of $\mathbb{R}^2$. We can try to use the invariants a and b to classify them. If we just consider the letters A to F, we have the following table:

	A	B	C	D	E	F
$a(R)$	3	2	2	1	3	3
$b(R)$	1	2	1	2	1	1

We see that the letters A, E and F have the same invariants, but otherwise there are no coincidences. It is possible to remove a single point from E to get a space with three components (and similarly for F) but this is not possible for A, so A is not homeomorphic to E or F. It is not hard to exhibit a homeomorphism between E and F. To be horribly explicit, we could use the following formal definitions:

$$E = (\{0\} \times [0,2]) \cup ([0,1] \times \{0,1,2\}) \subset \mathbb{R}^2$$
$$F = (\{0\} \times [0,2]) \cup ([0,1] \times \{1,2\}) \subset \mathbb{R}^2.$$

Let X be the union of the top two horizontal lines and the top half of the vertical line in E, so $X = (\{0\} \times [1,2]) \cup ([0,1] \times \{1,2\})$. We can then define a homeomorphism $f\colon E \to F$ by

$$f(x,y) = \begin{cases} (x,y) & \text{if } (x,y) \in X \\ (0,(y+1/2)) & \text{if } (x,y) \in \{0\} \times [0,1] \\ ((1-x)/2, 0) & \text{if } (x,y) \in [0,1] \times \{0\}. \end{cases}$$

The conclusion is that E is homeomorphic to F, and no other pair of different letters in $\{A,B,C,D,E,F\}$ are homeomorphic to each other.

9. Local Connectedness

Definition 9.1. [`defn-locally-connected`]
Let X be a topological space. We say that X is *locally connected* if it satisfies the following conditions:

(a) The family of all connected open sets forms a basis for the topology.
(b) Some family of connected open sets forms a basis for the topology.
(c) For every point $x \in X$ and every open neighbourhood U of x, there is a connected open neighbourhood V of x with $V \subseteq U$.

(Proposition 2.28 shows that these three conditions are equivalent.) Similarly, we say that X is *locally path connected* if it satisfies the following equivalent conditions:

(d) The family of all path connected open sets forms a basis for the topology.
(e) Some family of path connected open sets forms a basis for the topology.
(f) For every point $x \in X$ and every open neighbourhood U of x, there is a path connected open neighbourhood V of x with $V \subseteq U$.

Remark 9.2. As every path connected open set is connected, it is clear that every locally path connected space is locally connected.

Proposition 9.3. [`prop-Rn-locally-connected`]
The space $\mathbb{R}^n$ is locally path connected, and thus locally connected.

Proof. The open balls $OB_\epsilon(x)$ form a basis for the topology, and they are convex and therefore connected by Example 8.3. $\square$

REMARK 9.4. [rem-locally-connected]
In $\mathbb{Q}$ the only connected subsets are the singletons (by Exercise 7.4), and these have empty interior. It follows that $\mathbb{Q}$ is not locally connected. On the other hand, in a discrete space (such as $\mathbb{Z}$), every point is a connected neighbourhood of itself; so discrete spaces are locally connected.

PROPOSITION 9.5. [prop-subspace-locally-connected]
Let X be a topological space, and let U be an open subspace.

(a) *If X is locally connected, then so is U.*
(b) *If X is locally path connected, then so is U.*

PROOF. This is clear, because the open subsets of U are just the open subsets of X that are contained in U. □

PROPOSITION 9.6. [prop-product-locally-connected]
Let I be a finite set, and let $(X_i)_{i\in I}$ be a family of spaces indexed by I. Put $X = \prod_{i\in I} X_i$.

(a) *If each X_i is locally connected, then so is X.*
(b) *If each X_i is locally path connected, then so is X.*

PROOF. For part (a), let β_i denote the family of connected open sets in X_i, and suppose that this gives a basis for all i. Let β denote the family of all products $\prod_I U_i$, where $U_i \in \beta_i$ for all i. Proposition 5.28 tells us that β is a basis for the product topology, and all the sets in β are connected by Proposition 7.13, so X is locally connected as claimed. Essentially the same argument works for (b). □

PROPOSITION 9.7. [prop-local-components]

(a) *If X is locally connected, then the components of X are both open and closed in X.*
(b) *If X is locally path connected, then the path components of X are the same as the components, and they are both open and closed in X.*

PROOF.

(a) Suppose that X is locally connected. Consider a component C, and a point $x \in C$. As X is locally connected we can find a connected open set V with $x \in V$, and from the definition of components we see that $V \subseteq C$, so x is in the interior of C. As x was an arbitrary point of C, it follows that C is open. We also know from Proposition 7.17 that C is closed.
(b) Suppose instead that X is locally path connected. Consider a path component C. By essentially the same argument as above, we see that C is open. Now C^c is the union of all the other path components, each of which is open for the same reason, so C^c is open, so C is closed. Now pick a point $x \in C$, and let D be the connected component containing x. As C is connected and contains x we see that $C \subseteq D$ (so in particular D does not really depend on the choice of x). Now C and C^c give a relative separation of the connected set D, which must therefore be trivial. As $x \in C\cap D \neq \emptyset$, we must have $C \subseteq D$ and thus $C = D$. This shows that the path components are the same as the components, and they are both open and closed.

□

COROLLARY 9.8. *If X is connected and locally path connected, then it is path connected.*

PROOF. As X is connected there is only one component, but the proposition tells us that the components are the same as the path components, so there is only one path component. □

EXERCISE 9.1. [ex-lconn]
Suppose that $f\colon X \to Y$ is continuous and surjective, and that X is locally connected. Need Y be locally connected?

Solution: No. Let X be the discrete space $\mathbb{N}$, which is locally connected ($\{n\}$ is a connected neighbourhood of n contained in every neighbourhood of n). Let Y be $\{1/n \;:\; n \in \mathbb{N}, n > 0\} \cup \{0\}$. The map $f\colon X \to Y$

defined by

$$f(n) = \begin{cases} 0 & \text{if } n = 0 \\ 1/n & \text{if } n > 0 \end{cases}$$

is surjective and continuous (trivially, because $\mathbb{N}$ is discrete). The point $0 \in Y$ has no connected neighbourhoods, so Y is not locally connected.

10. Compactness

Definition 10.1. [`defn-open-cover`]
An *open cover* of a space X is a family $(U_i)_{i\in I}$ such that $\bigcup_{i\in I} U_i = X$. A *finite subcover* is a subfamily $(U_i)_{i\in J}$ for some finite subset J of I, such that $\bigcup_{j\in J} U_j = X$.

Example 10.2. [`eg-cover-Rn`]
The intervals $U_n = (n-1, n+1)$ give an open cover of $\mathbb{R}$. Any finite subfamily of these can only cover a bounded part of $\mathbb{R}$, so there is no finite subcover.

Example 10.3. [`eg-cover-Sn`]
Put

$$\begin{aligned} S^n &= \{x \in \mathbb{R}^{n+1} \ :\ \sum_{i=0}^{n} x_i^2 = 1\} \\ U_i^+ &= \{x \in S^n \ :\ x_i > 0\} \\ U_i^- &= \{x \in S^n \ :\ x_i < 0\}. \end{aligned}$$

Thus S^n is the unit sphere in $(n+1)$-dimensional euclidean space, and the sets $U_i^\pm$ are open hemispheres. The collection $U_0^+, \ldots, U_n^+, U_0^-, \ldots, U_n^-$ is then a finite open cover of S^n.

Definition 10.4. [`defn-compact`]
A space X is *compact* if every open cover of X has a finite subcover.

For proofs about compactness, it is convenient to have some slightly more flexible terminology.

Definition 10.5. [`defn-cover-misc`]
Let X be a space, and let Y be a subset of X. Given a family of subsets $(U_i)_{i\in I}$ (not necessarily open) we say that the family *covers* Y if $Y \subseteq \bigcup_{i\in I} U_i$. Now suppose that the family $(U_i)_{i\in I}$ is fixed. For any subset $J \subseteq I$ we say that J *covers* Y if the subfamily $(U_j)_{j\in J}$ covers Y, or equivalently $Y \subseteq \bigcup_{j\in J} U_j$. We say that Y is *finitely covered* if there is some finite subset J that covers Y.

Remark 10.6. [`rem-cover-misc`]
If Y is covered by J and Z is covered by K then $Y \cup Z$ is covered by $J \cup K$. It follows that if Y and Z are both finitely covered then so is $Y \cup Z$. Thus, if X is covered by the whole family, then every finite subset is finitely covered.

Example 10.7. [`eg-finite-compact`]
It follows directly from the above remark that every finite space is compact.

Example 10.8. [`eg-cofinite-compact`]
Let X be any nonempty set with the cofinite topology (as in Example 2.16). We claim that X is compact. Indeed, let $(U_i)_{i\in I}$ be an open cover. Choose a point $a \in X$, and an index i_a such that $a \in U_{i_a}$. Now U_{i_a} is a nonempty open set, so the complement $Y = X \setminus U_{i_a}$ must be finite, and thus covered by some finite set $J \subseteq I$. Now X is covered by the finite set $J \cup \{i_a\}$, as required.

Proposition 10.9. [`prop-interval-compact`]
The space $[0,1]$ *is compact.*

Proof. Let $(U_i)_{i\in I}$ be an open cover. Let P be the set of all t such that $[0,t]$ is finitely covered. We can choose i such that $0 \in U_i$, and U_i is open so there exists $\epsilon > 0$ with $[0,\epsilon) \subseteq U_i$, which means that $\epsilon/2 \in P$. This means that P is nonempty and bounded above by 1, so we can put $p = \sup(P)$, and we find that $0 < \epsilon/2 \leq p \leq 1$. Now choose j such that $p \in U_j$. As U_j is open in $[0,1]$ we see that U_j contains $(p-\delta, p+\delta) \cap [0,1]$ for some $\delta > 0$. Now $p - \delta$ cannot be an upper bound for P, so we can find $q \in P$ with $p - \delta < q$, so $[0, q]$ is covered by some finite set J. Now put $K = J \cup \{j\}$ and observe that this covers $[0, p+\delta) \cap [0,1]$, so $[0, p+\delta) \cap [0,1] \subseteq P$. This can only be consistent with the fact that $p = \sup(P)$ if $p = 1$. We conclude that K covers all of $[0,1]$, as required. □

PROPOSITION 10.10. [`prop-compact-subspace`]
A subspace $Y \subseteq X$ is compact in the subspace topology if and only if the following holds: for every family of open subsets of X that covers Y, some finite subfamily covers Y.

PROOF. First suppose that Y is compact in the subspace topology. Let $(U_i)_{i \in I}$ be a family of open subsets of X that covers Y. Then the sets $V_i = U_i \cap Y$ form an open cover of Y. As Y is compact there must exist a finite subset $J \subseteq I$ such that $Y = \bigcup_{j \in J}(U_j \cap Y)$, but this means that $Y \subseteq \bigcup_{j \in J} U_j$ as required.

Conversely, suppose that Y satisfies the condition in the proposition. Let $(V_i)_{i \in I}$ be an open cover of Y. By the definition of the subspace topology, there must be open sets $U_i \subseteq X$ such that $V_i = U_i \cap Y$. As the sets V_i give a cover of Y, we must have $Y \subseteq \bigcup_{i \in I} U_i$. By hypothesis, there must exist a finite set $J \subseteq I$ such that $Y \subseteq \bigcup_{j \in J} U_j$, or equivalently $Y = \bigcup_{j \in J} V_j$. This shows that Y is compact. □

It is sometimes convenient to have a formulation of compactness in terms of closed sets rather than open sets.

DEFINITION 10.11. [`defn-fip`]
A collection $\mathcal{F} = (F_i)_{i \in I}$ of subsets of a space X has the *finite intersection property* (FIP) if for every finite set of indices $J \subseteq I$, the intersection $\bigcap_{i \in J} F_i$ is nonempty.

PROPOSITION 10.12. [`prop-fip`]
Let X be a topological space. Then X is compact if and only if every family $(F_i)_{i \in I}$ of closed sets with FIP has $\bigcap_{i \in I} F_i \neq \emptyset$.

PROOF. Suppose that X is compact. Let $(F_i)_{i \in I}$ be a family of closed subsets with FIP. Put $U_i = X \setminus F_i$, which is open. Given any finite subset $J \subseteq I$ we have

$$X \setminus \bigcup_{j \in J} U_j = \bigcap_{j \in J}(X \setminus U_j) = \bigcap_{j \in J} F_j \neq \emptyset,$$

so $\bigcup_{j \in J} U_j \neq X$. In other words, no finite subfamily of the open sets U_i can cover X. As X is compact, this means that the whole family cannot cover X either. This means that $X \setminus \bigcup_{i \in I} U_i$ is nonempty, or in other words $\bigcap_{i \in I} F_i \neq \emptyset$, as required. We leave it to the reader to check that the whole argument is straightforwardly reversible. □

PROPOSITION 10.13. [`prop-closed-compact`]
Suppose that X is compact, and that Y is closed in X. Then Y is also compact.

PROOF. Let $(U_i)_{i \in I}$ be a family of open subsets of X that cover Y. Let I_0 consist of I together with an extra point 0, and put $U_0 = X \setminus Y$. This gives a larger family of open sets that covers all of X. As X is compact, it follows that there is some finite subset J_0 with $\bigcup_{j \in J_0} U_j = X$. Clearly U_0 does not contribute to covering Y, so we can put $J = J_0 \setminus \{0\} \subseteq I$ and conclude that $Y \subseteq \bigcup_{j \in J} U_j$ as required. □

DEFINITION 10.14. [`defn-precompact`]
Let X be a topological space, and let Y be a subspace. We say that Y is *precompact* in X if $\overline{Y}$ is compact in the subspace topology.

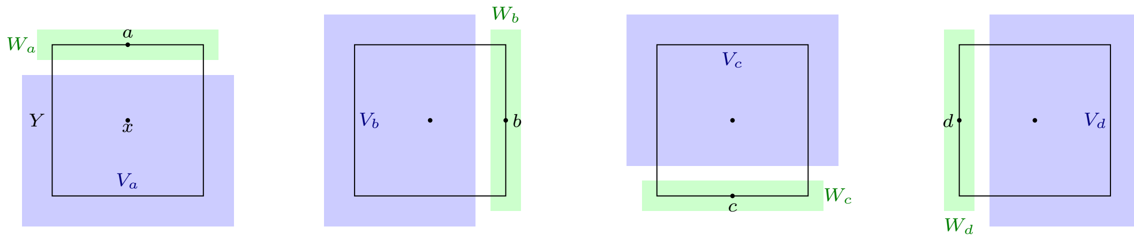
REMARK 10.15. [`rem-precompact`]
Suppose that $Z \subseteq Y \subseteq X$ and Y is precompact. We claim that Z is also precompact. Indeed, $\overline{Z}$ is a closed subset of the compact set $\overline{Y}$, so it is again compact as required.

There is a partial converse to Proposition 10.13 as follows.

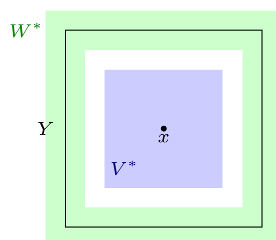
PROPOSITION 10.16. [`prop-compact-closed`]
Let X be a Hausdorff space, and let Y be a compact subspace. Then Y is closed in X.

PROOF. Put $U = X \setminus Y$; it will suffice to show that this is open. Consider a point $x \in U$. For each $y \in Y$ we have $x \neq y$, so (by the Hausdorff property) we can find disjoint open sets V_y, W_y with $x \in V_y$ and $y \in W_y$. The family $(W_y)_{y\in Y}$ covers Y, so (by compactness) there exists a finite subset $J \subseteq Y$ such that Y is contained in the set $W^* = \bigcup_{y\in J} W_y$. Now put $V^* = \bigcap_{y\in J} V_y$. As $V_y \cap W_y = \emptyset$, we find that $V^* \cap Y \subseteq V^* \cap W^* = \emptyset$, so $V^* \subseteq U$. Moreover, as V^* is the intersection of a finite family of open sets containing x, we see that V^* is an open set containing x. This shows that x is an interior point of U. This holds for any $x \in U$, so U is open as required.

For example, consider the case where Y is the boundary of a square in $\mathbb{R}^2$, and x is the centre of the square. Let a, b, c and d be the midpoints of the edges. For $y \in \{a, b, c, d\}$ we could choose V_y and W_y as shown below.



We could then take $J = \{a, b, c, d\}$ and V^* and W^* would be as follows:



□

REMARK 10.17. [rem-compact-intrinsic]

Suppose we have a space X and subsets $Z \subseteq Y \subseteq X$ (which we consider with their subspace topologies). If we ask whether Z is closed, we need to specify whether we mean closed in Y or closed in X. For example, if $X = \mathbb{R}$ and $Y = (0, \infty)$ and $Z = (0, 1]$ then Z is closed in Y but not in X. However, compactness of Z is an intrinsic property of the space Z. We can use the last two propositions to deduce compactness from closedness or *vice-versa* in various different settings, and thus relate closedness in various different ambient spaces. This theme will crop up repeatedly in later sections.

PROPOSITION 10.18. [rem-union-compact]

Let X be any space, and let $Y_1, \dots, Y_n$ be compact subspaces of X. Then $\bigcup_{i=1}^n Y_i$ is also compact.

PROOF. This is clear from Remark 10.6. □

PROPOSITION 10.19. [prop-image-compact]

Let $f: X \to Y$ be a continuous map, and let X' be a compact subspace of X. Then the image $Y' = f(X')$ is a compact subspace of Y.

PROOF. Let $(V_i)_{i\in I}$ be a family of open subsets of Y that covers Y'. Put $U_i = f^{-1}(V_i)$, which is open because f is continuous. If $x \in X'$ then $f(x) \in Y' \subseteq \bigcup_I V_i$, so $f(x) \in V_i$ for some i, so $x \in U_i$ for some i. This shows that the family $(U_i)_{i\in I}$ covers X', but X' is compact, so some finite subfamily $(U_j)_{j\in J}$ must cover X'. Now consider a point $y \in Y' = f(X')$. We can choose $x \in X'$ with $f(x) = y$, then we can choose $j \in J$ with $x \in U_j = f^{-1}(V_j)$. This means that $f(x) \in V_j$, or in other words $y \in V_j$. This works for any $y \in Y'$, so we see that $Y' \subseteq \bigcup_{j\in J} V_j$, so we have the required finite subcover for Y'. □

Corollary 10.20. [cor-image-compact]
Let Y be a topological space, and suppose that there exists a compact space X and a continuous surjective map $f\colon X \to Y$. Then Y is also compact.

Proof. Just take $X' = X$ in the proposition. $\square$

Corollary 10.21. [cor-quotient-compact]
Let X be a compact space, and let E be an equivalence relation on X. Then the quotient space X/E is compact.

Proof. Apply the previous corollary to the quotient map $X \to X/E$. $\square$

We can now prove the following very convenient result. Part (d) is often used to check that various explicitly constructed maps are homeomorphisms.

Proposition 10.22. [prop-comp-to-haus]
Let X and Y be spaces such that X is compact and Y is Hausdorff, and let $f\colon X \to Y$ be a continuous map. Then:

(a) *f is always a closed map.*
(b) *If f is injective then it is an embedding.*
(c) *If f is surjective then it is a quotient map.*
(d) *If f is bijective then it is a homeomorphism.*

Proof. Let F be a closed subset of X. Proposition 10.13 tells us that F is compact, so Proposition 10.19 tells us that $f(F)$ is compact, so Proposition 10.16 tells us that $f(F)$ is closed in Y. This proves (a). Parts (b), (c) and (d) then follow by Propositions 4.7 to 4.9. $\square$

Lemma 10.23. [lem-compact-basis]
Let X be a space, and let β be a basis for the topology. Suppose that every cover of X by basic open sets has a finite subcover. Then X is compact.

Proof. Let $(U_i)_{i\in I}$ be a cover of X by arbitrary open sets. For each $x \in X$ we can then choose an index i_x such that $x \in U_{i_x}$. As β is a basis we can then choose $V_x \in \beta$ with $x \in V_x \subseteq U_{i_x}$. The family $(V_x)_{x\in X}$ is then a cover of X by basic open sets. By hypothesis, we can find a finite subset $J \subseteq X$ such that $X = \bigcup_{y\in J} V_y$. Now put $J' = \{i_y \;:\; y \in J\}$, which is a finite subset of I. As $U_{i_y} \supseteq V_y$ we see that $\bigcup_{j\in J'} U_j \supseteq \bigcup_{y\in J} V_y = X$, which gives the required finite subcover. $\square$

Proposition 10.24. [prop-tychonov-binary]
Let X and Y be compact spaces. Then $X \times Y$ is also compact.

Proof. Let β be the family of sets of the form $U \times V$, where U is open in X and V is open in Y. This is a basis for the product topology on $X \times Y$. Consider a family of basic open sets $(U_i \times V_i)_{i\in I}$ that covers all of $X \times Y$. For each $x \in X$ we put $I(x) = \{i \in I \;:\; x \in U_i\}$. Now for any $y \in Y$ we must have $(x, y) \in U_i \times V_i$ for some i, so $i \in I(x)$ and $y \in V_i$. This means that $Y = \bigcup_{i\in I(x)} V_i$, but Y is compact, so there is a finite subset $J(x) \subseteq I(x)$ with $Y = \bigcup_{j\in J(x)} V_j$. Now put $W_x = \bigcap_{j\in J(x)} U_j$. This is the intersection of a finite family of open sets containing x, so it is an open set containing x. Thus, the family $(W_x)_{x\in X}$ is an open cover of the compact space X, so there a finite subset $K \subseteq X$ with $X = \bigcup_{t\in K} W_t$. Put $J = \bigcup_{t\in K} J(t)$, which is a finite subset of I. We claim that $X \times Y = \bigcup_{j\in J}(U_j \times V_j)$. To see this, consider an arbitrary point $(x, y) \in X \times Y$. As the sets $(W_t)_{t\in K}$ cover X, we can choose $t \in K$ with $x \in W_t$. As the sets $(V_j)_{j\in J(t)}$ cover Y, we can choose $j \in J(t)$ such that $y \in V_j$. From the definition of W_t we have $W_t \subseteq U_j$, so $x \in U_j$, so $(x, y) \in U_j \times V_j$ as required.

This now shows that every cover of $X \times Y$ by basic open sets has a finite subcover, so $X \times Y$ is compact by Lemma 10.23. $\square$

Corollary 10.25. [cor-tychonov-finite]
Let $(X_i)_{i\in I}$ be a finite family of compact spaces. Then the product $\prod_{i\in I} X_i$ is also compact.

Proof. This follows easily by induction from the proposition. $\square$

It is an important fact that the restriction to finite families is not really necessary here: the product of *any* family of compact spaces is compact. This is called Tychonov's Theorem, and it will be proved as Theorem 21.25 using the theory of ultrafilters. For the moment we will explain how to prove one special case by more elementary means.

PROPOSITION 10.26. [prop-tychonov-profinite]
Suppose that for each $n \in \mathbb{N}$ we have a finite set X_n with the discrete topology. Then the product $X = \prod_{n=0}^{\infty} X_n$ is compact.

PROOF. We will suppose that $(U_i)_{i \in I}$ is an open cover with no finite subcover, and derive a contradiction. We will say that a subset $Y \subseteq X$ is *bad* if it is not finitely covered, so X itself is bad by assumption. Note that if $Y \cup Z$ is bad, then either Y or Z must be bad.

Next, for any finite sequence $x = (x_0, \dots, x_r)$, we put

$$C(x) = \{y \in X \ : \ y_i = x_i \text{ for all } i \leq r\}.$$

Note that $C(0) \cup C(1) = X$ and this is bad, so either $C(0)$ or $C(1)$ must be bad. We can thus choose $x_0 \in \{0, 1\}$ such that $C(x_0)$ is bad. Next, we have $C(x_0, 0) \cup C(x_0, 1) = C(x_0)$ and this is bad, so we can choose x_1 such that $C(x_0, x_1)$ is bad. Similarly, we can choose x_2 such that $C(x_0, x_1, x_2)$ is bad. Continuing in this way, we obtain an infinite sequence x such that all the sets $C(x_0, \dots, x_r)$ are bad. Now the sets U_i cover X, so we have $x \in U_i$ for some i. Moreover, the set U_i is open, so we have $C_n(x) \subseteq U_i$ for some i, so certainly $C_n(x)$ is finitely covered. However $C_n(x)$ is just the same as $C(x_0, \dots, x_{n-1})$, which is bad by construction; this is the required contradiction. $\square$

COROLLARY 10.27. [cor-binary-compact]
It follows using Example 5.31 that the space of binary sequences is compact. $\square$

We can now characterise the compact subspaces of $\mathbb{R}^n$.

PROPOSITION 10.28. [prop-Rn-compact]
Let Y be a subset of $\mathbb{R}^n$. Then Y is compact if and only if it is bounded and closed.

PROOF. First suppose that Y is compact. Put $U_k = (-k, k)^n$, so $(U_k)_{k \in \mathbb{N}}$ is an open cover of $\mathbb{R}^n$. As Y is compact there must exist a finite subset $J \subseteq \mathbb{N}$ with $Y \subseteq \bigcup_{j \in J} U_j$. If we put $r = \max(J)$ this means that $Y \subseteq (-r, r)^n$, so Y is bounded. We also know from Proposition 10.16 that Y is closed.

Conversely, suppose that Y is bounded and closed. As Y is bounded, it is contained in $[-r, r]^n$ for some $r > 0$. As $[-r, r]$ is homeomorphic to $[0, 1]$, it is compact (by Proposition 10.9). It follows by Corollary 10.25 that $[-r, r]^n$ is compact. Now Y is a closed subset of a compact space, so it is compact by Proposition 10.13. $\square$

EXAMPLE 10.29. [eg-Sn-compact]
The sets $B^n = \{x \in \mathbb{R}^n \ : \ \|x\| \leq 1\}$ and $S^{n-1} = \{x \in \mathbb{R}^n \ : \ \|x\| = 1\}$ are both bounded and closed, so they are compact.

EXAMPLE 10.30. [eg-mandelbrot-compact]
Consider the Mandelbrot set $M \subseteq \mathbb{C} \simeq \mathbb{R}^2$, as defined in Example 2.5. We will show that this is bounded and closed, and thus is compact. Recall that

$$M = \{c \in \mathbb{C} \ : \ |f_n(c)| \leq 2 \text{ for all } n\},$$

where $f_0(c) = 0$ and $f_{k+1}(c) = f_k(c)^2 + c$. Here $f_k \colon \mathbb{C} \to \mathbb{C}$ is a polynomial function, so it is continuous. The disc $D = \{c \in \mathbb{C} \ : \ |c| \leq 2\}$ is closed, so $f_n^{-1}(D)$ is also closed. The set M is the intersection of the closed sets $f_n^{-1}(D)$, so it is also closed. Moreover, we have $f_1(c) = c$ so $M \subseteq f_1^{-1}(D) = D$, so M is bounded.

EXAMPLE 10.31. [eg-On-compact]
Recall the spaces

$$O(n) = \{A \in M_n(\mathbb{R}) \ : \ A^T A = I\}$$
$$\mathbb{R}P^{n-1} = \{A \in M_n(\mathbb{R}) \ : \ A^T = A = A^2, \ \text{trace}(A) = 1\}.$$

Here as usual we identify $M_n(\mathbb{R})$ with $\mathbb{R}^{n^2}$, giving the metric $d_2(A,B)=\sqrt{\text{trace}((A-B)^T(A-B))}$ explained in Example 2.40. We explained in Example 5.21 that $O(n)$ and $\mathbb{R}P^{n-1}$ are closed in $M_n(\mathbb{R})$. For $A\in O(n)$ we have $d_2(A,0)=\text{trace}(A^TA)^{1/2}=\text{trace}(I)^{1/2}=\sqrt{n}$, so $O(n)$ is also bounded. Similarly, for $A\in\mathbb{R}P^{n-1}$ we have $A^TA=A^2=A$ and so $d_2(A,0)=\sqrt{\text{trace}(A)}=1$; it follows that $\mathbb{R}P^n$ is bounded. Thus, both $O(n)$ and $\mathbb{R}P^n$ are compact.

On the other hand, by considering matrices of the form $\left[\begin{smallmatrix}1&a\\0&1\end{smallmatrix}\right]$ we see that $SL_2(\mathbb{R})$ is unbounded and so is not compact.

Corollary 10.32. [`cor-R-compact`]
Let X be a nonempty compact subset of $\mathbb{R}$. Then the numbers $a=\inf(X)$ and $b=\sup(X)$ are finite and $\{a,b\}\subseteq X\subseteq[a,b]$.

Proof. It follows from the proposition that X is bounded, so a and b are finite. It is clear from the definition of sup and inf that $X\subseteq[a,b]$. Next, as a is the greatest lower bound for X we see that $a+\epsilon$ (for $\epsilon>0$) is not a lower bound, so the set $[a,a+\epsilon)\cap X$ is nonempty. This means that a is a closure point of X, but X is compact and therefore closed, so $a\in X$. Essentially the same argument shows that $b\in X$ as well. □

Corollary 10.33. [`cor-CX-bounded`]
Let X be a nonempty compact space, and let $f\colon X\to\mathbb{R}$ be continuous. Then there exist real numbers a and b such that

- *$a\leq f(z)\leq b$ for all $z\in X$.*
- *$a=f(x)$ for some $x\in X$*
- *$b=f(y)$ for some $y\in X$*

In other words, the function f is bounded and attains its bounds.

Proof. Proposition 10.19 tells us that $f(X)$ is a nonempty compact subset of $\mathbb{R}$. We can thus apply Corollary 10.32 to the set $f(X)$, and the claim follows directly. □

We next discuss a generalisation of Lemma 2.49, which showed that the metrics d_1, d_2 and d_∞ on $\mathbb{R}^n$ are all equivalent.

Proposition 10.34. [`prop-norm`]
Let $\phi\colon\mathbb{R}^n\to[0,\infty)$ be a norm, as in Definition 3.31. Then

(a) *There are constants $k,K>0$ such that $k\|x\|_2\leq\phi(x)\leq K\|x\|_2$ for all $x\in\mathbb{R}^n$.*
(b) *The metric $d_\phi(x,y)=\phi(x-y)$ on $\mathbb{R}^n$ is strongly equivalent to the usual metric d_2.*
(c) *The map ϕ is continuous with respect to the usual topology on $\mathbb{R}^n$.*
(d) *There is a homeomorphism $f\colon B^n\to B(\mathbb{R}^n,\phi)$ given by $f(x)=\phi(x)x/\|x\|_2$ for $x\neq 0$, and $f(0)=0$. This satisfies $\phi(f(x))=\|x\|_2$, and it restricts to give homeomorphisms $S^{n-1}\to S(\mathbb{R}^n,\phi)$ and $OB^n\to OB(\mathbb{R}^n,\phi)$.*

Proof. Let e_i be the i'th basis vector $(0,\dots,1,\dots 0)$. Write $a_i=\phi(e_i)$ and $K=\sum_i a_i$. For $x\in\mathbb{R}^n$ it is clear that $|x_i|=\sqrt{x_i^2}\leq\sqrt{\sum_j x_j^2}=\|x\|_2$. Using the axioms for a norm we deduce that

$$\phi(x)=\phi(\sum_i x_ie_i)\leq\sum_i|x_i|\phi(e_i)\leq\sum_i\|x\|_2a_i=K\|x\|_2.$$

This proves half of (a).

Now suppose we have two points $x,y\in\mathbb{R}^n$. We can apply axiom N1 to the pair $y,x-y$ to see that

$$\phi(x)\leq\phi(y)+\phi(x-y)\leq\phi(y)+K\|x-y\|_2,$$

so $\phi(x)-\phi(y)\leq K\|x-y\|_2$. A symmetrical argument shows that we also have $\phi(y)-\phi(x)\leq K\|x-y\|_2$, so $|\phi(x)-\phi(y)|\leq K\|x-y\|_2$. This means that the map $\phi\colon\mathbb{R}^n\to\mathbb{R}$ is Lipschitz, and therefore continuous, so (c) holds. Now consider ϕ as a continuous function from the compact space S^{n-1} to $\mathbb{R}$, and apply Corollary 10.33 to see that there are numbers $a,b\in\mathbb{R}$ with $\{a,b\}\subseteq\phi(S^{n-1})\subseteq[a,b]$. Axiom N2 tells us that $0\not\in\phi(S^{n-1})$ so we must have $a>0$. For any $x\in\mathbb{R}^n$ we can write $x=\|x\|_2u$ with $u\in S^{n-1}$ and then $\phi(u)\geq a$ and so

$\phi(x) = \|x\|_2\phi(u) \geq a\|x\|_2$. We can thus complete the proof of (a) by taking $k = a$. By applying (a) to $x - y$ we deduce (b).

Now define $m\colon \mathbb{R}^n \setminus \{0\} \to (0, \infty)$ by $m(x) = \phi(x)/\|x\|_2$. This is continuous and strictly positive, so $1/m$ is also continuous. Note also that $m(tx) = m(x)$ for all $t \neq 0$, and that $k \leq m(x) \leq K$ for all x. We can thus define continuous maps

$$\mathbb{R}^n \setminus \{0\} \xrightarrow{f_1} \mathbb{R}^n \setminus \{0\} \xrightarrow{g_1} \mathbb{R}^n \setminus \{0\}$$

by $f_1(x) = x/m(x)$ and $g_1(x) = x\,m(x)$. We find that these are inverse to each other, so they are homeomorphisms. Note also that $\phi(f_1(x)) = \phi(x)/m(x) = \|x\|_2$ and similarly $\|g_1(x)\|_2 = \phi(x)$. Note also that $\|f_1(x)\| \leq \|x\|/k$ and $\|g_1(x)\| \leq K\|x\|$.

Now define $f\colon \mathbb{R}^n \to \mathbb{R}^n$ by $f(x) = f_1(x)$ for $x \neq 0$, and $f(0) = 0$. Define $g\colon \mathbb{R}^n \to \mathbb{R}^n$ in the same way. We claim that f is continuous. To see this, consider a sequence $(x_j)_{j\in\mathbb{N}}$ in $\mathbb{R}^n$ converging to $a \in \mathbb{R}^n$ say. If $a \neq 0$ then we must $x_j \neq 0$ for all sufficiently large j, and it follows from the continuity of f_1 that $f(x_k) \to f(a)$. If $a = 0$ then we must have $\|x_j\| \to 0$ and it follows from the inequality $\|f(x_j)\| \leq \|x_j\|/k$ that $f(x_j) \to 0 = f(0)$. This shows that that f is sequentially continuous and therefore continuous. The map g is also continuous by the same argument, and it is inverse to f, so it is a homeomorphism. As $\phi(f(x)) = \|x\|_2$ and $\|g(x)\|_2 = \phi(x)$, we see that f restricts to give homeomorphisms $S^{n-1} \to S(\mathbb{R}^n, \phi)$ and $B^n \to B(\mathbb{R}^n, \phi)$ and $OB^n \to OB(\mathbb{R}^n, \phi)$. □

Example 10.35. We can apply Proposition 10.34 to the norm $\phi(x, y) = \max(\|x\|_2, \|y\|_2)$ on $\mathbb{R}^{p+q} = \mathbb{R}^p \times \mathbb{R}^q$; it gives us a homeomorphism $B^p \times B^q \simeq B^{p+q}$.

Corollary 10.36. [`cor-norm`]
Let V be a finite-dimensional vector space over $\mathbb{R}$, and let ϕ be a norm on V, giving a metric d_ϕ and thus a topology τ_ϕ on V. Then τ_ϕ is the same as the linear topology (as in Definition 2.31).

Proof. Put $n = \dim(V)$. The proposition shows that any norm on $\mathbb{R}^n$ gives the standard topology, which is the same as the linear topology by Proposition 2.32. After choosing a basis we can identify V with $\mathbb{R}^n$; this identifies norms on V with norms on $\mathbb{R}^n$, and identifies the linear topology on V with the linear topology on $\mathbb{R}^n$, so the claim follows. □

Corollary 10.37. [`cor-linear-finite`]
Let $f\colon V \to W$ be a linear map between normed vector spaces, and suppose that V has finite dimension. Then f is continuous.

Proof. If W is also finite-dimensional, this follows easily from the previous corollary. For the general case, let $i\colon f(W) \to W$ be the inclusion, and let f_0 denote f considered as a map $V \to f(W)$. The special case shows that f_0 is continuous with respect to the norm topologies, and it is clear that i is continuous, so the same is true of $f = if_0$. □

We now discuss a lemma that will be useful in various places below.

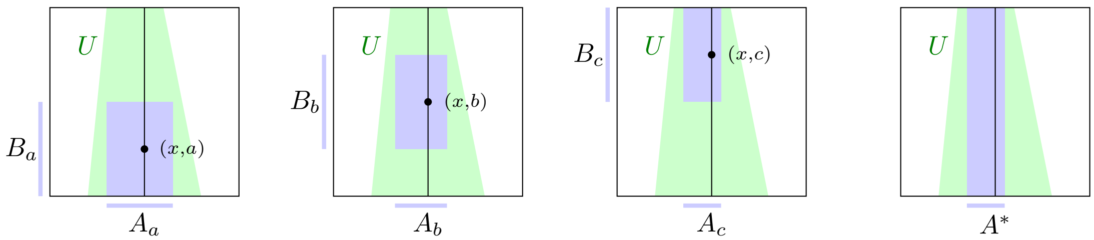
Lemma 10.38 (Tube Lemma). [`lem-tube`]
If $U \subseteq X \times Y$ is open and $Z \subseteq Y$ is compact then the set

$$V = \{x \in X \;:\; \{x\} \times Z \subseteq U\}.$$

is open in X.

Proof. Suppose that $x \in V$, so $\{x\} \times Z \subseteq U$. For each $z \in Z$ we have $(x, z) \in U$ so we can choose an open neighbourhood A_z of x and an open neighbourhood B_z of z such that $A_z \times B_z \subseteq U$. Now $(B_z)_{z\in Z}$ is a family of open sets in Y that covers the compact set Z, so we can find a finite subset $J \subseteq Z$ such that $Z \subseteq \bigcup_{z\in J} B_z$. Put $A^* = \bigcap_{z\in J} A_z$, which is an open neighbourhood of x in X. If $x' \in A^*$ and $z' \in Z$ then we must have $z' \in B_z$ for some $z \in J$, and then $x' \in A^* \subseteq A_z$ so $(x', z') \in A_z \times B_z \subseteq U$. This means that $A^* \times Z \subseteq U$, so $A^* \subseteq V$, showing that x is in the interior of V. This holds for any point $x \in V$, so V is open.

In the following illustration we have $X = Y = Z = [0, 1]$ and $J = \{a, b, c\}$.



□

Exercise 10.1. Let X be a compact Hausdorff space, and $C(X)$ the set of continuous functions $u\colon X \to \mathbb{R}$. Define two maps b and t ("bottom" and "top") from $C(X)$ to $\mathbb{R}$ by

$$b(u) = \min\{u(x) \ : \ x \in X\}$$
$$t(u) = \max\{u(x) \ : \ x \in X\}.$$

Prove that t and b are continuous.

Solution: Suppose we have two functions $u, v \in C(X)$. For all x we have

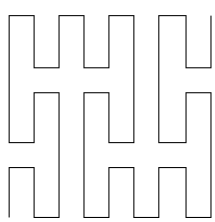
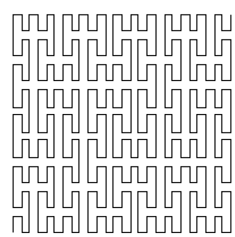
$$u(x) = v(x) + (u(x) - v(x)) \leq v(x) + |u(x) - v(x)| \leq v(x) + \|u - v\| \leq t(v) + \|u - v\|.$$

This holds for all x, so we must have $t(u) \leq t(v) + \|u - v\|$. By symmetry, we also have $t(v) \leq t(u) + \|v - u\| = t(u) + \|u - v\|$. It follows that $|t(u) - t(v)| \leq \|u - v\| = d(u, v)$. It follows that t is a Lipschitz map, with Lipschitz constant 1, so it is continuous. It also follows (using the identity $b(u) = -t(-u)$) that b is continuous.

11. Space-filling-curves

[`sec-peano`]

We now turn briefly to the theory of space-filling curves. This is useful as an extended example illustrating many of the ideas that we have studied so far. It is also important as a warning that continuous functions can be more wild than one might naively think. In particular, one might imagine that the image of a continuous map $k\colon [0,1] \to [0,1]^2$ is necessarily one-dimensional, in some sense. However, we will construct an example (due to Peano) where k is actually surjective. (Note, however, that k cannot be a homeomorphism, by Proposition 8.19.) The general idea is not too hard to understand. The pictures below show two different maps $k_4, k_6\colon [0,1] \to [0,1]^2$:

Close inspection shows that k_6 follows roughly the same route as k_4, but with more squiggles. It also passes close to every point in the square. By adding even more squiggles we can define maps k_n for $n > 6$ that pass even closer to all the points in the square. If we arrange the squiggling in the right way then the functions k_n will converge to a continuous function $k\colon [0,1] \to [0,1]^2$ whose image is dense. As $[0,1]$ is compact, the image will also be closed, so it will be all of $[0,1]^2$. To carry out this programme, we need a good way to organise the combinatorial structure of the squiggles. To do this, it turns out to be convenient to work everywhere in base three.

Definition 11.1. [`defn-ternary`]

We put $T = \{0,1,2\}$, and call this the set of *ternary digits*. We then put $X = \prod_{k=1}^{\infty} T$. We give T the discrete topology and X the product topology. (This has many features in common with the space of binary sequences, which we have discussed in various places.) We also define a map $f\colon X \to [0,1]$ by $f(x) = \sum_{k=1}^{\infty} x_k/3^k$.

We claim that the map f is continuous and surjective. This just means that every number in $[0,1]$ has a ternary expansion, analogous to the usual decimal representation. Just as decimal representations are unique apart from the usual ambiguity about infinite strings of nines, we will show that ternary expansions are unique apart from a similar ambiguity about infinite strings of twos. To formalise this, we need some further definitions.

Definition 11.2. [`defn-ternary-relation`]
We put

$$X_0 = \{x \in X \;:\; x \neq 0 \text{ but } x_k = 0 \text{ for } k \gg 0\}.$$

Every element $x \in X_0$ then has a unique representation $x = (u, t, 0^\infty)$ where u is a finite (possibly empty) sequence of ternary digits and $t \in \{1, 2\}$. We define a map $g \colon X_0 \to X$ by

$$g(u, t, 0^\infty) = (u, t - 1, 2^\infty).$$

We then define

$$R = \{(x, y) \in X^2 \;:\; x = y \text{ or } (x \in X_0 \text{ and } y = g(x)) \text{ or } (y \in X_0 \text{ and } x = g(y))\} \subseteq X^2.$$

Proposition 11.3. [`prop-ternary`]
The set R is an equivalence relation on X and a closed subset of X^2. The map f is a quotient map with ($f(x) = f(y)$ iff $(x, y) \in R$), so it induces a homeomorphism $X/R \to [0,1]$.

Proof. Suppose we $x \in X$ and $\epsilon > 0$. Choose $m \in \mathbb{N}$ such that $3^{-m} < \epsilon$, and put

$$U = \{y \in X \;:\; y_i = x_i \text{ for all } i \leq m\}.$$

As T is discrete, we see that U is an open neighbourhood of x in the product topology. If $y \in U$ we have $y_k - x_k = 0$ for $k \leq m$ and $|y_k - x_k| \leq 2$ for $k > m$ so

$$|f(y) - f(x)| = \left| \sum_{k=m+1}^{\infty} (y_k - x_k)/3^k \right| \leq \sum_{k=m+1}^{\infty} 2/3^k = 3^{-m} < \epsilon.$$

From this it is clear that f is continuous. Next, suppose we have a rational number q of the form $q = a/3^m$ for some integer a with $0 < a < 3^m$. We can then write a in base three, say as $a = \sum_{j=0}^{m-1} a_j 3^j$ with $a_j \in T$. If we put

$$x = (a_{m-1}, a_{m-2}, \ldots, a_0, 0^\infty) \in X_0$$

we find that $f(x) = q$. It follows that $f(X_0)$ is dense in $[0,1]$. On the other hand, we know from Proposition 10.26 that X is compact, so $f(X)$ is compact and therefore closed in $[0,1]$; it follows that f is surjective. Using Proposition 10.22 we deduce that f is a quotient map. Now put

$$\operatorname{eq}(f) = \{(x, y) \in X^2 \;:\; f(x) = f(y)\},$$

which is a closed subset of X^2 and an equivalence relation. In view of Proposition 5.61, it will suffice to check that $R = \operatorname{eq}(f)$. Using the fact that $\sum_{k=m+1}^{\infty} 2/3^k = 1/3^m$, we see that $fg = f \colon X_0 \to [0,1]$, and it follows that $R \subseteq \operatorname{eq}(f)$. For the converse, suppose that $(x, y) \in \operatorname{eq}(f)$. If $x = y$ then clearly $(x, y) \in R$. Suppose instead that $x \neq y$, and let m be the smallest index where $x_m \neq y_m$. After exchanging x and y if necessary, we may assume that $x_m > y_m$, so $x_m - y_m - 1 \geq 0$. After rearranging the equation $f(x) = f(y)$ using $\sum_{k=m+1}^{\infty} 2/3^k = 1/3^m$, we obtain

$$\frac{x_m - y_m - 1}{3^m} + \sum_{k=m+1}^{\infty} \frac{x_k}{3^k} + \sum_{k=m+1}^{\infty} \frac{2 - y_k}{3^k} = 0.$$

Here all the terms on the left hand side are nonnegative, so they must vanish individually. We therefore have $x_m = y_m + 1$, and for $k > m$ we have $x_k = 0$ and $y_k = 2$. This means that $x \in X_0$ and $y = g(x)$, so $(x, y) \in R$ as required. $\square$

Definition 11.4. [defn-peano-curve]

We define $\chi\colon T \to T$ by $\chi(t) = 2 - t$ (so $\chi^2 = 1$). We then define a map $k\colon X \to X^2$ by $k(x) = (k_0(x), k_1(x))$, where

$$\begin{aligned} k_0(x) &= (x_1, \chi^{x_2}(x_3), \chi^{x_2+x_4}(x_5), \chi^{x_2+x_4+x_6}(x_7), \dots) \\ k_1(x) &= (\chi^{x_1}(x_2), \chi^{x_1+x_3}(x_4), \chi^{x_1+x_3+x_5}(x_6), \dots) \end{aligned}$$

or equivalently

$$\begin{aligned} k_0(x)_i &= \chi^{\sum_{j<i} x_{2j}}(x_{2i-1}) \\ k_1(x)_i &= \chi^{\sum_{j\leq i} x_{2j-1}}(x_{2i}). \end{aligned}$$

Proposition 11.5. [prop-peano-curve]

The map $k\colon X \to X^2$ is a homeomorphism. Moreover, there is a unique map $\overline{k}\colon [0,1] \to [0,1]^2$ making the following diagram commute:

$$\begin{tikzcd} X \arrow[r, "k"] \arrow[d, two heads, "f"'] & X^2 \arrow[d, two heads, "f^2"] \\ {[0,1]} \arrow[r, "\overline{k}"'] & {[0,1]^2}. \end{tikzcd}$$

This map $\overline{k}$ is a quotient map. In particular, it is continuous and surjective.

Proof. First note that $k_0(x)_i$ depends only on $x_1, \dots, x_{2i-1}$, so there is a commutative diagram of the form

$$\begin{tikzcd} X \arrow[r, "k_0"] \arrow[d, "\pi"'] & X \arrow[d, "\pi_i"] \\ T^{2i-1} \arrow[r, "k_{0,i}"'] & T. \end{tikzcd}$$

Here the map π is continuous (by an easy argument recorded as Lemma 5.27) and $k_{0,i}$ is automatically continuous because T is discrete. This means that $\pi_i \circ k_0$ is continuous for all i, so k_0 is continuous by Proposition 5.16. The map k_1 is also continuous by essentially the same argument, and it follows that the combined map $k\colon X \to X^2$ is continuous. Next, note that as $\chi^2 = 1$ we have $\chi^{\chi(t)} = \chi^t$ for all $t \in T$. Using this, it is straightforward to check that k is a bijection with inverse

$$k^{-1}(y,z) = (y_1, \chi^{y_1}(z_1), \chi^{z_1}(y_2), \chi^{y_1+y_2}(z_2), \chi^{z_1+z_2}(y_3), \cdots)$$

or equivalently

$$\begin{aligned} k^{-1}(y,z)_{2k-1} &= \chi^{\sum_{j<k} z_j}(y_k) \\ k^{-1}(y,z)_{2k} &= \chi^{\sum_{j\leq k} y_j}(z_k). \end{aligned}$$

The same line of argument used for k also shows that k^{-1} is continuous, so k is a homeomorphism.

Now suppose we have $x, y \in X$ with $f(x) = f(y)$. We claim that $fk_0(x) = fk_0(y)$. Using Proposition 11.3, we reduce easily to the case where $x = (u, t, 0^\infty) \in X_0$ and $y = g(x) = (u, t-1, 2^\infty)$. Suppose for the moment that t occurs in an odd-numbered position, say as x_{2m-1}. Put $v_i = \chi^{\sum_{j<i} u_{2j}}(u_{2i-1})$ (for $1 \leq i < m$) and $r = \sum_{i=1}^{m-1} u_{2i}$. We find that

$$\begin{aligned} k_0(x) &= (v, \chi^r(t), \chi^r(0)^\infty) \\ k_0(y) &= (v, \chi^r(t-1), \chi^r(2)^\infty). \end{aligned}$$

If r is even we find that $k_0(x) = (v, t, 0^\infty) \in X_0$ and $k_0(y) = (v, t-1, 2^\infty) = g(k_0(x))$. If r is odd we find instead that $k_0(y) = (v, 3-t, 0^\infty) \in X_0$ and $k_0(x) = (v, 2-t, 2^\infty) = g(k_0(y))$. Either way, we have

$f(k_0(x)) = f(k_0(y))$, as required. Suppose instead that t occurs as x_{2m}. Then v_i is defined for $1 \leq i \leq m$ and we have

$$\begin{aligned} k_0(x) &= (v, \chi^{r+t}(0)^\infty) \\ k_0(y) &= (v, \chi^{r+t-1}(2)^\infty). \end{aligned}$$

These are the same, because $\chi(2) = 0$. We therefore again have $f(k_0(x)) = f(k_0(y))$. A similar argument shows that k_1 has the same property, so we see that $(f^2 \circ k)(x) = (f^2 \circ k)(y)$ whenever $f(x) = f(y)$. As f is surjective, it follows that there is a unique map $\overline{k}\colon [0,1] \to [0,1]^2$ making the diagram

$$\begin{tikzcd} X \arrow[r, "k"] \arrow[d, two heads, "f"'] & X^2 \arrow[d, two heads, "f^2"] \\ {[0,1]} \arrow[r, "\overline{k}"'] & {[0,1]^2} \end{tikzcd}$$

commute. Here f is a quotient map, and $\overline{k} \circ f$ is continuous (because it is the same as $f^2 \circ k$) so $\overline{k}$ is continuous. As k is a homeomorphism and f is surjective, it follows easily that $\overline{k}$ is surjective. As $[0,1]$ is compact and Hausdorff, it follows that $\overline{k}$ is a quotient map. □

COROLLARY 11.6. [`cor-peano`]
For every $n \geq 0$ there is a continuous surjective map $k_n\colon [0,1] \to [0,1]^n$, and also a continuous surjective map $m_n\colon \mathbb{R} \to \mathbb{R}^n$.

PROOF. The cases $n = 0$ and $n = 1$ are trivial, and the proposition gives a map k_2. Next, as $\mathbb{Z}^2$ is countable, we can choose a bijection $p\colon \mathbb{Z} \to \mathbb{Z}^2$. We can then define

$$m_{2,d}\colon [2d, 2d+1] \to p(d) + [0,1]^2 \subset \mathbb{R}^2$$

by $m_{2,d}(2d+t) = p(d) + k_2(t)$. By combining these we get a continuous surjection $m_2\colon \coprod_d [2d, 2d+1] \to \mathbb{R}^2$. We can then extend this over $\mathbb{R}$ by putting $m_2(2d+1+s) = (1-s)m_2(2d+1) + s m_2(2d+2)$ for $d \in \mathbb{Z}$ and $s \in [0,1]$. Finally, we can define k_n and m_n for $n > 2$ by the recursive rules

$$\begin{aligned} k_n &= ([0,1] \xrightarrow{k_2} [0,1] \times [0,1] \xrightarrow{1 \times k_{n-1}} [0,1] \times [0,1]^{n-1} = [0,1]^n) \\ m_n &= (\mathbb{R} \xrightarrow{m_2} \mathbb{R} \times \mathbb{R} \xrightarrow{1 \times m_{n-1}} \mathbb{R} \times \mathbb{R}^{n-1} = \mathbb{R}^n). \end{aligned}$$

□

12. Compactness and Completeness in Metric Spaces

We now study various special features of metric spaces.

DEFINITION 12.1. [`defn-bounded`]
Let X be a metric space, and let Y be a subset of X. We say that Y is *bounded* if it satisfies the equivalent conditions (a) and (b) below:

(a) There exists $R \geq 0$ such that $d(y, y') < R$ for all $y, y' \in Y$.
(b) For all $x \in X$ there exists $R \geq 0$ such that $Y \subseteq B_R(x)$, so $d(x, y) \leq R$ for all $y \in Y$.
(c) There exists $x \in X$ and $R \geq 0$ such that $Y \subseteq B_R(x)$, so $d(x, y) \leq R$ for all $y \in Y$.

It is an easy exercise with the triangle inequality to see that (a) and (b) are equivalent to each other, and also to (c) except in the trivial case where $X = \emptyset$. We say that a sequence $(x_n)_{n \in \mathbb{N}}$ is bounded if and only if the corresponding set $\{x_n : n \in \mathbb{N}\}$ is bounded.

DEFINITION 12.2. [`defn-cauchy`]
Now let X be a metric space. A sequence $\underline{x} = (x_n)_{n \in \mathbb{N}}$ is *Cauchy* if for all $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that $d(x_i, x_j) < \epsilon$ whenever $i, j \geq N$.

LEMMA 12.3. [`lem-cauchy`]
Every convergent sequence is Cauchy.

Proof. Let $\underline{x}$ be a convergent sequence in a metric space X, converging to a point a say. The ball $OB_{\epsilon/2}(a)$ is an open neighbourhood of a, so there exists N such that $x_i \in OB_{\epsilon/2}(a)$ for all $i \geq N$. Now if $i, j \geq N$ we have $d(x_i, a) < \epsilon/2$ and $d(a, x_j) < \epsilon/2$ so

$$d(x_i, x_j) \leq d(x_i, a) + d(a, x_j) < \epsilon/2 + \epsilon/2 = \epsilon$$

as required. □

Lemma 12.4. [lem-cauchy-bounded]
Every Cauchy sequence is bounded.

Proof. Let $\underline{x}$ be a Cauchy sequence. By taking $\epsilon = 1$ in the definition, we see that there exists $N \in \mathbb{N}$ such that $d(x_i, x_j) < 1$ for all $i, j \geq N$. Now put

$$K = \max(d(x_0, x_N), \ldots, d(x_{N-1}, x_N), 1).$$

We find that $d(x_i, x_N) \leq K$ for all i, so the sequence is bounded. □

Lemma 12.5. [lem-cauchy-bound]
Let $\underline{x}$ be a sequence that converges to a, and suppose we have $d(x_n, x_m) < \epsilon$ for all $n, m \geq N$. Then $d(x_n, a) \leq \epsilon$ for all $n \geq N$.

Proof. Consider a number $\delta > 0$. As $x_n \to a$, there exists M such that $d(x_m, a) < \delta$ for all $m \geq M$. After replacing M by $\max(M, N)$ if necessary, we may assume that $M \geq N$. For $n \geq N$ we then have $d(x_n, x_M) < \epsilon$ and $d(x_M, a) < \delta$ so $d(x_n, a) < \epsilon + \delta$. As this holds for all $\delta > 0$, we deduce that $d(x_n, a) \leq \epsilon$. □

Lemma 12.6. [lem-cauchy-subsequence]
Let $\underline{x}$ be a Cauchy sequence, and let $\underline{y}$ be a subsequence of $\underline{x}$. Then $\underline{y}$ is also Cauchy, and $\underline{x}$ converges to a point a if and only if $\underline{y}$ converges to a.

Proof. We have $y_k = x_{n_k}$ for some strictly increasing sequence of atural numbers n_k (so $n_k \geq k$). Suppose we are given $\epsilon > 0$. By the Cauchy property of $\underline{x}$, there exists $N \in \mathbb{N}$ with $d(x_i, x_j) < \epsilon$ whenever $i, j \geq N$. Now for $i, j \geq N$ we have $n_i, n_j \geq N$ so $d(y_i, y_j) = d(x_{n_i}, x_{n_j}) < \epsilon$. This shows that $\underline{y}$ is Cauchy.

We saw in Lemma 2.56 that if $\underline{x}$ converges to a, then so does $\underline{y}$. Conversely, suppose that $\underline{y}$ converges to a. Given $\epsilon > 0$ we can then choose N such that $d(x_i, x_j) < \epsilon/2$ for $i, j \geq N$, and then we can choose $N \geq M$ such that $d(y_i, a) < \epsilon/2$ for $i \geq N$. Suppose that $i \geq n_N$. Now i and n_N are both at least N, so $d(x_i, y_N) = d(x_i, x_{n_N}) < \epsilon/2$. We also have $d(y_N, a) < \epsilon/2$, so $d(x_i, a) < \epsilon$. It follows that $\underline{x}$ converges to a. □

Definition 12.7. [defn-complete]
A metric space X is *complete* if every Cauchy sequence in X is convergent.

Lemma 12.8. [lem-nondecreasing]
Let $\underline{x}$ be a sequence of real numbers that is bounded above (so there exists $K \in \mathbb{R}$ with $x_i \leq K$ for all i) and nondecreasing (so $x_i \leq x_{i+1}$ for all i). Then $\underline{x}$ converges to $\sup(\{x_i \;:\; i \in \mathbb{N}\})$.

Proof. Put $Y = \{x_i \;:\; i \in \mathbb{N}\}$ and $y = \sup(Y)$ (which is well-defined because Y is nonempty and bounded above). Suppose we are given $\epsilon > 0$. As y is by definition the least upper bound of Y, we see that $y - \epsilon$ is not an upper bound, so there is some number $y_N \in Y$ with $y_N > y - \epsilon$. Now for $i \geq N$ we have $y - \epsilon < y_N \leq y_i \leq y$, so $|y_i - y| < \epsilon$ as required. □

Proposition 12.9. [prop-R-complete]
The metric space $\mathbb{R}$ is complete.

Proof. Let $\underline{x}$ be a Cauchy sequence in $\mathbb{R}$. By Lemma 12.4 we can find K such that $x_i \in [-K, K]$ for all i, so every subset of the x_i's has a sup and an inf. Now put

$$X_i = \{x_j \;:\; j \geq i\} \qquad y_i = \sup(X_i)$$
$$Y_i = \{y_j \;:\; j \geq i\} \qquad z_i = \inf(Y_i).$$

Note that $X_{k+1} \subseteq X_k$, so $y_{k+1} \leq y_k$, so

$$y_0 \geq y_1 \geq \cdots \geq y_{i-1} \geq y_i.$$

This means that the lower bounds for Y_i are the same as the lower bounds for Y_0, and so $z_i = z_0$ for all i. We now claim that the sequence $\underline{x}$ converges to z_0. Indeed, suppose we are given a number $\epsilon > 0$. Let M be such that $|x_i - x_j| < \epsilon/2$ whenever $i, j \geq M$. For $M \leq i \leq j$ we then have $X_j \subseteq [x_i - \epsilon/2, x_i + \epsilon/2]$ so $y_j \in [x_i - \epsilon/2, x_i + \epsilon/2]$. This shows that $Y_i \subseteq [x_i - \epsilon/2, x_i + \epsilon/2]$, so $z_0 = z_i \in [x_i - \epsilon/2, x_i + \epsilon/2]$, so $|x_i - z_0| \leq \epsilon/2 < \epsilon$, as required. □

Remark 12.10. [`rem-not-complete`]
The space $X = (0, \infty)$ (with the standard metric $d(x, y) = |x-y|$) is not complete, because the sequence $(2^{-n})_{n\in\mathbb{N}}$ is a Cauchy sequence in X that has no limit in X. We saw in Proposition 3.21 that the map $g(y) = (y - y^{-1})/2$ gives a homeomorphism $(0, \infty) \to \mathbb{R}$, so we see that it is possible for an incomplete metric space to be homeomorphic to a complete one. For a different perspective on the same phenomenon, we can define a new metric d' on $\mathbb{R}$ by $d'(x, y) = |g^{-1}(x) - g^{-1}(y)|$. As g is a homeomorphism, we see that d' is weakly equivalent to the standard metric d on $\mathbb{R}$. However, the sequence $(g(2^{-n}))_{n\in\mathbb{N}}$ is Cauchy but not convergent with respect to d'. Thus, $\mathbb{R}$ is complete with respect to d, but incomplete with respect to d'. On the other hand, it is straightforward to check that if d and d' are *strongly* equivalent metrics on a set X, then X is complete with respect to d iff it is complete with respect to d'.

Proposition 12.11. [`prop-binary-complete`]
Let X be the space of binary sequences as in Example 2.6, with the metric describe in Example 2.42. Then X is complete.

Proof. Let $(x_n)_{n\geq 0}$ be a Cauchy sequence in X. Note that each x_n is itself a sequence, say $x_n = (x_{n0}, x_{n1}, x_{n2}, \dots)$, with $x_{ni} \in \{0, 1\}$ for all n and i. For each m the Cauchy condition gives us an integer N_m such that $d(x_p, x_q) < 2^{-m}$ whenever $p, q \geq N_m$. After inspecting the definition of the metric, we deduce that $x_{pm} = x_{qm}$ whenever $p, q \geq N_m$, so in particular $x_{pm} = x_{N_m,m}$. We write a_m for this value, so $a_m \in \{0, 1\}$. This gives us a point $a = (a_0, a_1, \dots) \in X$. If $p \geq \max(N_0, \dots, N_m)$ we find that $d(x_p, a) < 2^{-m}$, which means that $x_p \to a$. Thus, every Cauchy sequence is convergent, as claimed. □

Proposition 12.12. [`prop-padic-incomplete`]
Let X denote the set $\mathbb{Z}$ with the p-adic metric as in Example 2.43. Then X is not complete.

Proof. Suppose for the moment that $p > 2$. Put

$$x_n = (1 - p^n)/(1 - p) = \sum_{i=0}^{n-1} p^i \in \mathbb{Z}.$$

Then $x_n - x_m$ is divisible by $p^{\min(n,m)}$, and it follows easily that the sequence $(x_n)_{n\geq 0}$ is Cauchy. We claim, however, that it has no limit in X. Indeed, if x_n converged to a then the sequence of terms $1-p^n = (1-p)x_n$ would converge to $(1 - p)a$. On the other hand, it is clear that $1 - p^n \to 1$, so we would have $(1 - p)a = 1$. As $p > 2$, this is clearly impossible for $a \in X = \mathbb{Z}$. To cover the case $p = 2$, just use the sequence $x_n = (1 - 4^n)/(1 - 4)$ instead. □

We next discuss some questions about products of metric spaces. For definiteness, we will always use the metric on $X \times Y$ given by

$$d((x, y), (x', y')) = \max(d(x, x'), d(y, y')).$$

The other metrics discussed in Definition 2.50 are strongly equivalent to this one, which means that none of the questions that we consider will depend on the choice of metric.

Proposition 12.13. [`prop-product-seq`]
Suppose we have a sequence $(z_n)_{n\in\mathbb{N}}$ in $X \times Y$ given by $z_n = (x_n, y_n)$, and a point $c = (a, b) \in X \times Y$.

(a) *The sequence $\underline{z}$ converges to c if and only if $\underline{x}$ converges to a and $\underline{y}$ converges to b.*
(b) *The sequence $\underline{z}$ is Cauchy if and only if $\underline{x}$ and $\underline{y}$ are both Cauchy.*

Proof.

(a) This is a special case of Proposition 5.35, but we will give a direct argument. Suppose that $\underline{x}$ converges to a and $\underline{y}$ converges to b. Given $\epsilon > 0$ there exists N such that $d(x_n, a) < \epsilon$ whenever $n \geq N$, and there exists M such that $d(y_n, b) < \epsilon$ whenever $n \geq M$. Now if $n \geq \max(N, M)$ we find that
$$d(z_n, c) = \max(d(x_n, a), d(y_n, b)) < \max(\epsilon, \epsilon) = \epsilon.$$
It follows that $\underline{z}$ converges to c. We leave the converse to the reader.

(b) Suppose that $\underline{z}$ is Cauchy. Given $\epsilon > 0$ there exists N such that $d(z_i, z_j) < \epsilon$ for all $i, j \geq N$. For such i and j we then have
$$d(x_i, x_j) \leq \max(d(x_i, x_j), d(y_i, y_j)) = d(z_i, z_j) < \epsilon.$$
This shows that $\underline{x}$ is Cauchy, and similarly $\underline{y}$ is Cauchy. We leave the converse to the reader.

□

Corollary 12.14. [`cor-product-complete`]
If X and Y are complete metric spaces, then so is $X \times Y$. □

Corollary 12.15. [`cor-Rn-complete`]
The space $\mathbb{R}^n$ is complete.

Proof. Induction on n, starting with Proposition 12.9 and using Corollary 12.14. □

Proposition 12.16. [`prop-CXY-complete`]
Let X be a topological space, and let Y be a complete metric space. Let $C(X, Y)$ be the set of continuous functions from X to Y, with the metric $d(f, g) = \sup\{d(f(x), g(x)) : x \in X\}$ as in Definition 3.28. Then $C(X, Y)$ is complete.

Proof. Let $\underline{f} = (f_n)_{n\in\mathbb{N}}$ be a Cauchy sequence in $C(X, Y)$. For each $\epsilon > 0$ we therefore have $N(\epsilon) \in \mathbb{N}$ such that $d(f_i, f_j) < \epsilon$ whenever $i, j \geq N(\epsilon)$. For each $x \in X$ we note that $d(f_i(x), f_j(x)) \leq d(f_i, f_j)$, and thus that $(f_n(x))_{n\in\mathbb{N}}$ is a Cauchy sequence in Y. As Y is complete this converges to a point $g(x) \in Y$. More precisely, if $N(\epsilon) \leq i \leq j$ we have $d(f_i(x), f_j(x)) < \epsilon$ so
$$d(f_i(x), g(x)) \leq d(f_i(x), f_j(x)) + d(f_j(x), g(x)) < \epsilon + d(f_j(x), g(x)).$$
We now let j tend to infinity in this inequality to see that $d(f_i(x), g(x)) \leq \epsilon$ whenever $i \geq N(\epsilon)$. If we knew that g was continuous, this would give us a point $g \in C(X, Y)$ with $d(f_n, g) < \epsilon$ whenever $n \geq N(\epsilon/2)$, proving that $\underline{f}$ is convergent as required. Thus, all that is left is to verify that g is indeed continuous.

Suppose we are given $x \in X$ and $\epsilon > 0$. Put $n = N(\epsilon/4)$, so $d(f_n(u), g(u)) < \epsilon/4$ for all u. As f_n is continuous, we can choose $\delta > 0$ such that $d(f_n(u), f_n(x)) < \epsilon/4$ whenever $d(u, x) < \delta$. For such u we then have
$$d(g(u), g(x)) \leq d(g(u), f_n(u)) + d(f_n(u), f_n(x)) + d(f_n(x), g(x)) \leq \epsilon/4 + \epsilon/4 + \epsilon/4 < \epsilon,$$
as required. □

Corollary 12.17. [`cor-CX-complete`]
In particular, the metric space $C(X, \mathbb{R})$ is complete. □

Definition 12.18. [`defn-l-two`]
Let $l^2(\mathbb{N})$ denote the set of sequences $x = (x_0, x_1, \dots)$ of real numbers for which $\sum_i x_i^2 < \infty$. For $x \in l^2(\mathbb{N})$ we put $\|x\| = \sqrt{\sum_i x_i^2}$.

Proposition 12.19. [`prop-l-two`]
The set $l^2(\mathbb{N})$ is a vector space, and the function $x \mapsto \|x\|$ is a norm (in the sense of Definition 3.31). Moreover, $l^2(\mathbb{N})$ is complete with respect to the corresponding metric.

Proof. In this proof we will use the norm $\|x\| = \sqrt{\sum_{i=0}^{n-1} x_i^2}$ on $\mathbb{R}^n$, and recall that the metric topology is the same as the product topology. (The corresponding fact will not hold for $l^2(\mathbb{N})$, however.) We define $\tau_n : l^2(\mathbb{N}) \to \mathbb{R}^n$ by $\tau_n(x) = (x_0, \dots, x_{n-1})$, and note that the numbers $\|\tau_n(x)\|$ form a nondecreasing sequence converging to $\|x\|$.

Suppose we have $x, y \in l^2(\mathbb{N})$. By straightforward expansion we have

$$\|\tau_n(x-y)\|^2 + \|\tau_n(x+y)\|^2 = \sum_{i=0}^{n-1}(x_i - y_i)^2 + \sum_{i=0}^{n}(x_i + y_i)^2 = 2\sum_{i=0}^{n} x_i^2 + 2\sum_{i=0}^{n} y_i^2 \leq 2\|x\|^2 + 2\|y\|^2 < \infty.$$

By passing to the limit as $n \to \infty$ we deduce that $x+y, x-y \in l^2(\mathbb{N})$ with $\|x+y\|^2 + \|x-y\|^2 \leq 2\|x\|^2 + 2\|y\|^2$. It is also clear that for $t \in \mathbb{R}$ we have $tx \in l^2(\mathbb{N})$ with $\|tx\| = |t|\,\|x\|$. It follows that $l^2(\mathbb{N})$ is a vector space under the obvious operations of addition and scalar multiplication. We next show that the map $x \mapsto \|x\|$ is a norm on $l^2(\mathbb{N})$. All of the axioms are clear except for the triangle inequality, which says that $\|x\| + \|y\| - \|x+y\| \geq 0$. The usual triangle inequality for $\mathbb{R}^n$ (Lemma 2.24) says that $\|\tau_n(x)\| + \|\tau_n(y)\| - \|\tau_n(x+y)\| \geq 0$ for all n, and we can recover the inequality for $l^2(\mathbb{N})$ by letting n tend to infinity. We thus have a norm as claimed, and a metric $d(x,y) = \|x-y\|$. It is clear that for all z and i we have $|z_i| \leq \|z\|$, so $|x_i - y_i| \leq d(x,y)$. It follows that the projection $\pi_i \colon l^2(\mathbb{N}) \to \mathbb{R}$ is Lipschitz (with constant 1) and therefore continuous. Similarly, the truncation maps τ_n also satisfy $d(\tau_n(x), \tau_n(y)) \leq d(x,y)$, so they are continuous.

Finally, we must show that $l^2(\mathbb{N})$ is complete. Let $(x_k)_{k\in\mathbb{N}}$ be a Cauchy sequence in $l^2(\mathbb{N})$ (so each x_k is itself a sequence, say $x_k = (x_{k0}, x_{k1}, \cdots)$). For any n, we recall that π_n is Lipschitz so the sequence $(\pi_n(x_k))_{k\in\mathbb{N}}$ is a Cauchy sequence in $\mathbb{R}$, so it converges to some point $y_n \in \mathbb{R}$. This gives a sequence $y = (y_n)_{n\in\mathbb{N}}$; we claim that this lies in $l^2(\mathbb{N})$, and that it is a limit for the the original sequence $(x_k)_{k\in\mathbb{N}}$. We will first observe a weaker fact. If we fix m, we have a sequence $(\tau_m(x_k))_{k\in\mathbb{N}}$ in $\mathbb{R}^m$, and we claim that this converges to $\tau_m(y)$. Indeed, by Proposition 5.35 it will suffice to check that for all $n < m$ the numbers $\pi_n(\tau_m(x_k)) = \pi_n(x_k)$ converge to $\pi_n(\tau_m(y)) = \pi_n(y) = y_n$, and this is true by the definition of y_n. Next, as our original sequence is Cauchy, it must be bounded, so there is a constant K with $\|x_k\| \leq K$ for all k. It follows that $\|\tau_m(x_k)\| \leq K$, and by letting k tend to infinity we deduce that $\|\tau_m(y)\| \leq K$. As this holds for all m we deduce that $\|y\| \leq K$, so $y \in l^2(\mathbb{N})$.

Now put $z_k = x_k - y$, so $\tau_m(z_k) \to 0$ for all m. Suppose we are given $\epsilon > 0$. By the Cauchy property of $(x_k)_{k\in\mathbb{N}}$, we can find N such that $d(z_j, z_k) = d(x_j, x_k) < \epsilon/3$ for all $j, k \geq N$. Consider an index $j \geq N$. As $\|z_j\|$ is the nondecreasing limit of the numbers $\|\tau_m(z_j)\|$, we can choose m with $\|z_j\| < \|\tau_m(z_j)\| + \epsilon/3$. As $\|\tau_m(z_i)\| \to 0$, we can choose $k \geq j$ with $\|\tau_m(z_k)\| < \epsilon/3$. As $j, k \geq N$ we have $\|z_j - z_k\| < \epsilon/3$. Putting this together, we have

$$\begin{aligned}\|z_j\| &< \epsilon/3 + \|\tau_m(z_j)\| \\ &\leq \epsilon/3 + \|\tau_m(z_k)\| + \|\tau_m(z_j - z_k)\| \\ &\leq \epsilon/3 + \|\tau_m(z_k)\| + \|z_j - z_k\| \\ &< \epsilon/3 + \epsilon/3 + \epsilon/3 = \epsilon.\end{aligned}$$

This shows that $\|z_i\| \to 0$, so $x_i \to y$ in $l^2(\mathbb{N})$ as claimed. □

Remark 12.20. [rem-l-two-I]

It is sometimes useful to generalise this as follows. Let I be an arbitrary set. For any function $x \colon I \to \mathbb{R}$ and any finite set $J \subseteq I$ we put $\|x\|_J = \sqrt{\sum_{j\in J} x_j^2}$. We then put

$$\|x\| = \sup\{\|x\|_J \ : \ J \subseteq I \text{ finite }\} \in [0, \infty],$$

and $l^2(I) = \{x \ : \ \|x\| < \infty\}$. By essentially the same argument as above we see that $l^2(I)$ is a vector space, the map $x \mapsto \|x\|$ is a norm, and $l^2(I)$ is complete with respect to the corresponding metric.

The last example can be placed in a broader framework as follows:

Definition 12.21. [defn-banach-space]

A *Banach space* is a vector space (over $\mathbb{R}$ or $\mathbb{C}$) equipped with a norm such that it is complete under the resulting metric.

Example 12.22. [eg-banach-space]

Any finite-dimensional vector space is a Banach space under any norm, as we see from Corollary 10.36. If X is any compact space, then we see from Corollary 12.17 that $C(X, \mathbb{R})$ is a Banach space under the norm

$\|f\| = \sup\{|f(x)| \,:\, x \in X\}$. (We need X to be compact here to ensure that $\|f\|$ is finite.) Moreover, $l^2(I)$ is a Banach space under the norm introduced in Definition 12.18.

PROPOSITION 12.23. [prop-hom-banach]
Let V and W be normed vector spaces over $\mathbb{K}$ (where $\mathbb{K}$ is $\mathbb{R}$ or $\mathbb{C}$). Consider the space $\mathrm{Hom}^c(V,W)$ of continuous linear maps from V to W, with the operator norm as in Definition 3.33. If W is a Banach space, then $\mathrm{Hom}^c(V,W)$ is also a Banach space. In particular, the continuous dual $V^* = \mathrm{Hom}^c(V,\mathbb{K})$ is always a Banach space.

PROOF. Let $(f_n)_{n\in\mathbb{N}}$ be a Cauchy sequence in $\mathrm{Hom}^c(V,W)$. For any point $v \in V$ we then have a sequence $(f_n(v))_{n\in\mathbb{N}}$ in W. Using the inequality

$$d(f_n(v), f_m(v)) = \|(f_n - f_m)(v)\| \leq \|f_n - f_m\|_{\mathrm{op}}\|v\| = d(f_n, f_m)\|v\|$$

we see that this is again Cauchy. As W is assumed to be complete, there is a unique limit point for the sequence, which we denote by $f(v)$. It is straightforward to check that the resulting map $f\colon V \to W$ is linear. Moreover, as the sequence $(f_n)_{n\in\mathbb{N}}$ is Cauchy, it must be bounded, so there is a constant K such that $\|f_n\|_{\mathrm{op}} \leq K$ for all n. We thus have $\|f_n(v)\| \leq K\|v\|$ for all v, and it follows that $\|f(v)\| \leq K\|v\|$ as well. This proves that f is continuous. All that is left is to prove that $f_n \to f$ in $\mathrm{Hom}^c(V,W)$. Suppose we are given $\epsilon > 0$. As the original sequence is Cauchy, we can choose N such that $\|f_n - f_m\|_{\mathrm{op}} \leq \epsilon$ for all $n, m \geq N$. This means that for such n and m we have $\|f_n(v) - f_m(v)\| \leq \epsilon\|v\|$ for all $v \in V$. By a straightforward argument that we gave as Lemma 12.5, it follows that $\|f_n(v) - f(v)\| \leq \epsilon\|v\|$ for all $n \geq N$. As this holds for all v we have $d(f_n, f) \leq \epsilon$ for all $n \geq N$, as required. □

PROPOSITION 12.24. [prop-complete-closed]
Let X be a complete metric space, and let Y be a subset of X (considered as a metric space in the obvious way). Then Y is complete if and only if it is closed in X.

PROOF. First suppose that Y is complete. Let $\underline{y}$ be a sequence in Y that converges to $x \in X$. As the sequence is convergent in X, it must be Cauchy. As Y is complete and $\underline{y}$ is Cauchy, it must converge to a point $x' \in Y$. As limits in metric spaces are unique we must have $x = x'$, so $x \in Y$. Thus Y is closed (by the criterion in Proposition 2.58).

Conversely, suppose that Y is closed. Let $\underline{y}$ be a Cauchy sequence in Y. Then $\underline{y}$ can also be regarded as a Cauchy sequence in the complete metric space X, so it must converge to some point $x \in X$. As Y is closed we see from Proposition 2.58 that we actually have $x \in Y$, so $\underline{y}$ is convergent in Y. This proves that Y is complete. □

We next consider compactness and related properties in metric spaces.

DEFINITION 12.25. [defn-totally-bounded]
Let X be a metric space, and let ϵ be a positive real number. An *ϵ-net* for X is a finite subset $F \subseteq X$ such that

$$X = \bigcup_{x\in F} OB_\epsilon(x).$$

We say that X is *totally bounded* if it has an ϵ-net for every $\epsilon > 0$.

EXAMPLE 12.26. [eg-totally-bounded]
(a) The set $\{k/n \,:\, 0 \leq k \leq n\}$ is a $1/n$-net for $[0,1]$.
(b) Let X be the space of binary sequences as in Example 2.6, with the metric describe in Example 2.42. Put

$$F_n = \{x \in X \,:\, x_i = 0 \text{ for all } i > n\}.$$

Then F_n is a 2^{-n}-net for X.
(c) Fix a prime p, and consider $\mathbb{Z}$ with the p-adic metric as in Example 2.43. Then $\{0, 1, \ldots, p^{n+1} - 1\}$ is a p^{-n}-net for $\mathbb{Z}$.

It follows that all three of these spaces are totally bounded.

Definition 12.27. [defn-lebesgue-number]

Let X be a metric space, and let $(U_i)_{i\in I}$ be an open covering of X. A *Lebesgue number* for the covering is a number $\epsilon > 0$ such that for each $x \in X$ there exists $i \in I$ such that $OB_\epsilon(x) \subseteq U_i$.

Theorem 12.28. [thm-compact-metric]

Let X be a metric space, and consider the following conditions:

(a) *X is compact.*
(b) *X is totally bounded and complete.*
(c) *Every sequence in X has a convergent subsequence.*
(d) *Every open cover of X has a Lebesgue number.*

Then (a), (b) and (c) are equivalent, and they imply (d).

Perhaps do something with neighbourhoods of the diagonal?

This summarises a number of smaller results, which we will prove separately. We will gather the threads together to prove the theorem at the end of this section.

First, however, we give a sample application.

Example 12.29. [eg-binary-compact]

Let X be the space of binary sequences as in Example 2.6. Using Example 12.26 and Proposition 12.11 we see that X is totally bounded and complete, so it is compact by the above theorem. This was already proved by a different method in Corollary 10.27.

Lemma 12.30. [lem-compact-metric]

Let X be a compact metric space. Then X is totally bounded and complete.

Proof. First suppose that X is compact. Consider a number $\epsilon > 0$. The family $\{OB_\epsilon(x) \;:\; x \in X\}$ is an open cover, so there is a finite subcover, say $\{OB_\epsilon(x) \;:\; x \in F\}$. This means that F is an ϵ-net for X. It follows that X is totally bounded. Now let $\underline{x}$ be a Cauchy sequence in X. Put $X'_n = \{x_i \;:\; i \geq n\}$ and let X_n be the closure of X'_n. If we have a finite list of these sets, say $X_{n_1}, \dots, X_{n_p}$, then we can put $m = \max(n_1, \dots, n_p)$ and we find that $x_m \in \bigcap_{t=1}^p X_{n_t}$. This means that the family $(X_n)_{n\in\mathbb{N}}$ has the Finite Intersection Property. As X is compact, Proposition 10.12 tells us that $\bigcap_n X_n \neq \emptyset$, so we can choose $a \in \bigcap_n X_n$. We claim that the sequence $\underline{x}$ converges to a. To see this, consider a number $\epsilon > 0$. Choose N such that $d(x_i, x_j) < \epsilon/2$ whenever $i, j \geq N$. We also have $a \in X_N$, so the ball $OB_{\epsilon/2}(a)$ must meet X'_N, so we can choose $n \geq N$ with $d(a, x_n) < \epsilon/2$. Now for $i \geq N$ we have $d(x_i, x_n) < \epsilon/2$ and $d(x_n, a) < \epsilon/2$ so $d(x_i, a) < \epsilon$, as required. We now see that X is complete as claimed. □

It is now convenient to introduce more flexible terminology.

Definition 12.31. [defn-totally-bounded-aux]

Let X be a metric space, let Y be a subset of X, and let ϵ be a positive real number. An *external ϵ-net* for Y in X is a finite subset $F \subseteq X$ such that

$$Y \subseteq \bigcup_{x\in F} OB_\epsilon(x).$$

We say that Y is *externally totally bounded in X* if it has an external ϵ-net for every $\epsilon > 0$.

Lemma 12.32. [lem-totally-bounded-subset]

Let X be a metric space, and let Y be a subset of X. Then Y is totally bounded if and only if it is externally totally bounded in X.

Proof. It is clear that an ϵ-net gives an external ϵ-net, so if Y is totally bounded then it is externally totally bounded.

Conversely, suppose that Y is externally totally bounded. For each $\epsilon > 0$, let F be an external $\epsilon/2$-net for Y. We can harmlessly discard from F any points x such that $OB_{\epsilon/2}(x) \cap Y = \emptyset$, so we may as well assume that there are no points of this type. Thus, for each $x \in F$ we can choose $\alpha(x) \in Y \cap OB_{\epsilon/2}(x)$. Put $G = \{\alpha(x) \;:\; x \in F\}$, which is a finite subset of Y. We claim that it is an ϵ-net. Indeed, if $y \in Y$ then we can find $x \in F$ such that $d(y, x) < \epsilon/2$ (because F is an $\epsilon/2$-net). We also have $d(x, \alpha(x)) < \epsilon/2$ and so

$d(y,\alpha(x)) < \epsilon$ by the triangle inequality. As $\alpha(x) \in G$ this means that $y \in OB_\epsilon(\alpha(x)) \subseteq \bigcup_{u\in G} OB_\epsilon(u)$. As y was an arbitrary point of Y, this shows that G is an ϵ-net as claimed. □

COROLLARY 12.33. [`cor-totally-bounded-subset`]
Any subset of a totally bounded set is again totally bounded.

PROOF. Any external ϵ-net for the larger set is also an external ϵ-net for the smaller set. □

PROPOSITION 12.34. [`prop-closure-totally-bounded`]
Let X be a metric space, and let Y be a totally bounded subset of X. Then $\overline{Y}$ is also totally bounded.

PROOF. Let F be an $\epsilon/2$-net for Y. We claim that this is an ϵ-net for $\overline{Y}$, which will prove the proposition. Indeed, if $x \in \overline{Y}$ then $OB_{\epsilon/2}(x)$ must meet Y. We choose $y \in OB_{\epsilon/2}(x) \cap Y$, and note that there must exist $z \in F$ with $y \in OB_{\epsilon/2}(z)$. Now $d(x,y) < \epsilon/2$ and $d(y,z) < \epsilon/2$ so $d(x,z) < \epsilon$ as required. □

PROPOSITION 12.35. [`prop-cauchy-subseq`]
Let X be a totally bounded metric space. Then every sequence in X has a subsequence that is Cauchy.

PROOF. Let $\underline{x}$ be a sequence in X. For each $n \in \mathbb{N}$, choose a 2^{-n}-net F_n for X. For each $u \in F_0$, put $I_0(u) = \{i \;:\; x_i \in OB_1(u)\}$. As F_0 is finite and $\bigcup_{u\in F_0} I_0(u) = \mathbb{N}$ we see that there must exist $u_0 \in F_0$ such that $I_0(u_0)$ is infinite. We put $\phi(0) = \min(I_0(u_0))$ and $\Phi(1) = I_0(u_0) \setminus \{\phi(0)\}$, so $\Phi(1)$ is an infinite set of integers, all of them larger than $\phi(0)$. Next, for $u \in F_1$ we put $I_1(u) = \{i \in \Phi(1) \;:\; x_i \in OB_{1/2}(u)\}$. As F_1 is finite and $\bigcup_{u\in F_1} I_1(u) = \Phi(1)$ we see that there must exist $u_1 \in F_1$ such that $I_1(u_1)$ is infinite. We put $\phi(1) = \min(I_1(u_1))$ and $\Phi(2) = I_1(u_1) \setminus \{\phi(1)\}$. Continuing in the same way, we define elements $u_k \in F_k$ and integers $\phi(k)$ and subsets $\Phi(k) \subseteq \mathbb{N}$ such that

(a) The set $I_k(u_k) = \{i \in \Phi(k) \;:\; x_i \in OB_{2^{-k}}(u_k)\}$ is infinite
(b) $\phi(k) = \min(I_k(u_k))$
(c) $\Phi(k+1) = I_k(u_k) \setminus \{\phi(k)\} \subset \Phi(k)$.

It follows that the function $\phi\colon \mathbb{N} \to \mathbb{N}$ is strictly increasing, so we have a subsequence $(x_{\phi(n)})_{n\in\mathbb{N}}$. We claim that this is Cauchy. Indeed, for $i,j > n$ we have $\phi(i), \phi(j) \in \Phi(n+1) \subseteq I_n(u_n)$ so $x_{\phi(i)}, x_{\phi(j)} \in OB_{2^{-n}}(u_n)$, so

$$d(x_{\phi(i)}, x_{\phi(j)}) \leq d(x_{\phi(i)}, u_n) + d(u_n, x_{\phi(j)}) < 2^{1-n},$$

and the Cauchy property follows easily from this. □

COROLLARY 12.36. [`cor-convergent-subseq`]
If X is totally bounded and complete then every sequence in X has a convergent subsequence. □

LEMMA 12.37. [`lem-sc-tb`]
Let X be a metric space in which every sequence has a convergent subsequence. Then X is totally bounded.

PROOF. If not, we can choose $\epsilon > 0$ such that there is no ϵ-net. We then choose a sequence recursively as follows. We start with an arbitrary point x_0. If $x_0, \dots, x_{n-1}$ have been chosen, we observe that they cannot give an ϵ-net by hypothesis, so there must exist a point x_n that does not lie in $\bigcup_{i=0}^{n-1} OB_\epsilon(x_i)$, so $d(x_i, x_n) \geq \epsilon$ for all $i < n$. This gives us a sequence with the property that $d(x_i, x_j) \geq \epsilon$ for all $i \neq j$, and it is clear from this that no subsequence can be Cauchy, so no subsequence can be convergent. This contradicts our assumption on X. □

LEMMA 12.38. [`lem-sc-lebesgue`]
Let X be a metric space in which every sequence has a convergent subsequence. Then every open cover of X has a Lebesgue number.

PROOF. Let $(U_i)_{i\in I}$ be an open cover of X, and suppose that there is no Lebesgue number. Then, for each n we can find a point x_n such that the ball $OB_{2^{-n}}(x_n)$ is not contained in any of the sets U_i. The sequence $(x_n)_{n\in\mathbb{N}}$ must have a subsequence $(x_{\phi(n)})_{n\in\mathbb{N}}$ converging to some point $x \in X$ say. As the sets U_i cover X, we can choose i such that $x \in U_i$. As U_i is open, we have $OB_\epsilon(x) \subseteq U_i$ for some $\epsilon > 0$. Now for m sufficiently large we have $2^{-\phi(m)} < \epsilon/2$ and also $d(x_{\phi(m)}, x) < \epsilon/2$, so $OB_{2^{-\phi(m)}}(x_{\phi(m)}) \subseteq OB_\epsilon(x) \subseteq U_i$, which contradicts the choice of $x_{\phi(m)}$. □

Lemma 12.39. [`lem-tbl-compact`]
Let X be a totally bounded metric space in which every open cover has a Lebesgue number. Then X is compact.

Proof. Let $(U_i)_{i\in I}$ be an open cover of X. Let ϵ be a Lebesgue number for the cover, and let F be an ϵ-net for X. For each $x \in F$ choose $i_x \in I$ such that $OB_\epsilon(x) \subseteq U_{i_x}$ (which is possible because ϵ is a Lebesgue number). Now put $J = \{i_x \;:\; x \in F\}$, which is a finite subset of I. We have $X = \bigcup_{x\in F} OB_\epsilon(x) \subseteq \bigcup_{x\in F} U_{i_x} = \bigcup_{j\in J} U_j$, so we have a finite subcover as required. □

Proof of Theorem 12.28.

- Lemma 12.30 shows that (a) implies (b).
- Corollary 12.36 shows that (b) implies (c).
- Lemma 12.38 shows that (c) implies (d).
- Lemmas 12.38, 12.37 and 12.39 together show that (c) implies (a).

□

Example 12.40. We will consider $\mathbb{R}^2$ with the lane metric as defined in Example 2.39:

$$d((x,y),(x',y')) = \begin{cases} |y-y'| & \text{if } x = x' \\ |y| + |x-x'| + |y'| & \text{if } x \neq x'. \end{cases}.$$

We will investigate which sets $K \subseteq \mathbb{R}^2$ are compact with respect to the metric topology. First, define $\pi\colon \mathbb{R}^2 \to \mathbb{R}$ by $\pi(x,y) = x$. For $a \in \mathbb{R}$ we also define $\iota_a\colon \mathbb{R} \to \mathbb{R}^2$ by $\iota_a(y) = (a,y)$. Given $K \subseteq \mathbb{R}^2$, we put

$$\begin{aligned} K(a) = \iota_a^{-1}(K) &= \{y \in \mathbb{R} \;:\; (a,y) \in K\} \\ A = \pi(K) &= \{a \;:\; K(a) \neq \emptyset\} \\ B = \pi(K \setminus (\mathbb{R}\times\{0\})) &= \{a \;:\; K(a) \not\subseteq \{0\}\} \subseteq A. \end{aligned}$$

These are all subsets of $\mathbb{R}$, and we consider $\mathbb{R}$ with the usual metric and topology. We claim that if K is compact in $\mathbb{R}^2$, then A is compact in $\mathbb{R}$, and B is finite, and $K(b)$ is compact in $\mathbb{R}$ for all $b \in B$. First, it is easy to see that π decreases distances and so is continuous, so the set $A = \pi(K)$ is compact. Similarly, ι_a is a closed embedding, so $K(a)$ is essentially a closed subset of a compact set and so is also compact. (Of course, this statement only has any force for $a \in B$.) Next, for $b \in B$ we put $r(b) = \sup\{|t| \;:\; t \in K(b)\} > 0$, and $U(b) = (-r(b), r(b))$. Note that this is open and contains 0 but does not contain $K(b)$. For $x \notin B$ we just put $U(x) = \mathbb{R}$, and then we put $U = \{(x,y) \in \mathbb{R}^2 \;:\; y \in U(x)\}$. Next, for $b \in B$ we put $V(b) = \{(b,y) \;:\; y \neq 0\}$. The sets U and $V(b)$ together give an open cover of K, so there must be a finite subcover. However, because $K(b) \not\subseteq U(b)$, this finite subcover must contain all of the sets $V(b)$. It follows that B is finite, as claimed. It is also not hard to see that these conditions are sufficient as well as necessary for the compactness of K.

12.1. Contraction mappings. We next explain a very useful principle called the Contraction Mapping Theorem.

Definition 12.41. [`defn-contraction`]
Let X be a metric space, and let f be a function from X to itself. If f has a Lipschitz constant α with $\alpha < 1$ (so $d(f(x), f(y)) \leq \alpha f(x,y)$ for all $x, y \in X$), we say that f is a *contraction mapping of ratio α*. A *fixed point* for f is a point $x \in X$ with $f(x) = x$.

Theorem 12.42 (The Contraction Mapping Theorem). [`thm-contraction`]
Let X be a nonempty complete metric space, and let $f\colon X \to X$ be a contraction mapping of ratio α. Then f has a unique fixed point. Moreover, if a is an arbitary point and b is the fixed point, then

$$d(a,b) \leq d(a, f(a))/(1-\alpha).$$

Proof. Choose $a = a_0 \in X$ and write $r = d(a, f(a))$. Then put $a_n = f^{(n)}(a_0)$ for all $n \geq 0$, so $a_{n+1} = f(a_n)$. We claim that the sequence (a_n) is Cauchy. Indeed, by induction we see that

$$d(a_n, a_{n+1}) = d(f(a_{n-1}), f(a_n)) \leq \alpha^n r$$

for all n. Thus, if $m \leq n$ we have

$$\begin{aligned} d(a_m, a_n) &\leq d(a_m, a_{m+1}) + \dots d(a_{n-1}, a_n) \\ &\leq r(\alpha^m + \dots \alpha^{n-1}) \\ &\leq r \sum_{k=m}^{\infty} \alpha^k = \alpha^m r/(1-\alpha), \end{aligned}$$

which easily implies the claim. Thus, as X is complete, the sequence converges to a limit b. Moreover

$$f(b) = f(\lim_{n\to\infty} a_n) = \lim_{n\to\infty} f(a_n) = \lim_{n\to\infty} a_{n+1} = b,$$

so b is a fixed point. Suppose that c is another fixed point. Then

$$d(b, c) = d(f(b), f(c)) \leq \alpha d(b, c)$$

As $\alpha < 1$, this implies $d(b, c) = 0$ and thus $b = c$. Thus b is the unique fixed point.

Finally,

$$d(a, b) = \lim_{n\to\infty} d(a_0, a_n) \leq \lim_{n\to\infty} \frac{1-\alpha^n}{1-\alpha} r = \frac{d(a, f(a))}{1-\alpha}.$$

□

The typical strategy for using the Contraction Mapping Theorem is as follows. Suppose we want to construct an object b with a particular property. We first find a space X of "potential solutions", in which we expect b to lie. We then try to define an "improvement function" $f\colon X \to X$, such that $f(x)$ is a better potential solution than x is. If we do this correctly then the fixed points of f will be precisely the actual solutions to our problem. If we are lucky then there will be a metric on X such that X is complete and f is a contraction mapping. If so, the theorem will tell us that there is a unique point $b \in X$ with the required property.

For example, one can use this approach to prove quite general existence theorems for solutions to ordinary differential equations. Here we will just illustrate the method by treating a simple special case.

Proposition 12.43. [prop-ode]

Let $f\colon \mathbb{R} \to \mathbb{R}$ be a continuously differentiable function, and suppose that there exists $\alpha < 1$ such that $|f'(x)| \leq \alpha$ for all $x \in \mathbb{R}$. Then there is a unique continuously differentiable map $u\colon [0,1] \to \mathbb{R}$ that satisfies the differential equation $u'(t) = f(u(t))$ with boundary condition $u(0) = 0$.

Proof. First, for any $x < y$ the Mean Value Theorem tells us that there exists $a \in [x, y]$ such that $f(y) - f(x) = f'(a)(y - x)$. It follows that $|f(y) - f(x)| \leq \alpha|y - x|$, so f is a contraction mapping of ratio α. However, we will not apply the Contraction Mapping Theorem to f itself, but to a rather more complicated operator defined in terms of f.

Put $X = C([0,1], \mathbb{R})$ (which is a complete metric space by Proposition 12.16). For $u \in X$ and $s \in [0,1]$ put $Fu(s) = \int_{t=0}^{s} f(u(t))\, dt$ and $K = \|f \circ u\|_\infty$. Note that

$$|Fu(r) - Fu(s)| = \left| \int_s^r f(u(t))\, dt \right| \leq \int_s^r |f(u(t))|\, dt \leq K|r - s|.$$

This means that $Fu\colon [0,1] \to \mathbb{R}$ is Lipschitz, and therefore continuous. We have therefore defined an operator $F\colon X \to X$.

Suppose that $u, v \in X$. We then have

$$\begin{aligned} d(Fu, Fv) &= \sup\{|Fu(s) - Fv(s)| \;:\; s \in [0,1]\} \\ &\leq \int_0^1 |f(u(t)) - f(v(t))|\, dt \\ &\leq \sup\{|f(u(t)) - f(v(t))| \;:\; t \in [0,1]\} \\ &\leq \alpha \sup\{|u(t) - v(t)| \;:\; t \in [0,1]\} = \alpha d(u, v). \end{aligned}$$

This means that $F\colon X \to X$ is a contraction mapping of ratio α. It follows that there is a unique point $u \in X$ with $Fu = u$, so

$$u(s) = \int_{t=0}^{s} f(u(t))\,dt.$$

We can put $s = 0$ here to see that $u(0) = 0$. We also see that

$$\frac{u(s+h)-u(s)}{h} = \frac{1}{h}\int_{t=s}^{s+h} f(u(t))\,dt,$$

which is easily seen to converge to $f(u(s))$ as $h \to 0$. This means that u is continuously differentiable, with $u'(s) = f(u(s))$ as required. $\square$

The Contraction Mapping Theorem can also be used to prove various versions of the Inverse Function Theorem and the Implicit Function Theorem. Again, we will treat only a simple case that illustrates the method.

Proposition 12.44. [`prop-implicit-function`]
Let $f\colon \mathbb{R}^k \times \mathbb{R}^n \to \mathbb{R}^k$ be a continuously differentiable map such that $f(u,0) = u + O(\|u\|^2)$ for small u in $\mathbb{R}^k$. Then there is a ball $Y = B_\epsilon(0) \subset \mathbb{R}^n$ for some $\epsilon > 0$ and a map $g\colon Y \to \mathbb{R}^k$ such that $f(g(y), y) = 0$ for all $y \in Y$.

Proof. It will be convenient to work with the function $h(x,y) = x - f(x,y)$ rather than f. This is again continuously differentiable, so there are continuous maps $\alpha\colon \mathbb{R}^k \times \mathbb{R}^n \to M_{k,k}(\mathbb{R})$ and $\beta\colon \mathbb{R}^k \times \mathbb{R}^n \to M_{k,n}(\mathbb{R})$ such that we have

$$h(x+u, y+v) = h(x,y) + \alpha(x,y)u + \beta(x,y)v + O(\|u\|^2 + \|v\|^2)$$

(for $x, u \in \mathbb{R}^k$ and $y, v \in \mathbb{R}^n$). The condition $f(u,0) = u + O(\|u\|^2)$ means that $\alpha(0,0) = 0$. As α is continuous we can choose $\delta > 0$ such that $\|\alpha(x,y)\| < 1/3$ whenever $\|x\| \leq \delta$ and $\|y\| \leq \delta$. As β is also continuous, and the set of such pairs (x,y) is compact, there is a constant K such that $\|\beta(x,y)\| \leq K$ for all (x,y) with $\|x\| \leq \delta$ and $\|y\| \leq \delta$. For such x and y we therefore have

$$\|h(x,y)\| = \left|\int_{t=0}^{1} \frac{d}{dt} h(tx,ty)\,dt\right| = \left\|\int_{t=0}^{1} \alpha(tx,ty)x + \beta(tx,ty)y\,dt\right\| \leq \|x\|/3 + K\|y\|.$$

Now put $\epsilon = \delta/\max(1, 3K)$ and $X = \{x \in \mathbb{R}^k \;:\; \|x\| \leq \delta\}$ and $Y = \{y \in \mathbb{R}^n \;:\; \|y\| \leq \epsilon\}$. Then define $M = C(Y,X)$, which is a complete metric space. For $g \in M$ we define $Tg\colon Y \to \mathbb{R}^k$ by $Tg(y) = h(g(y),y)$. As $y \in Y$ and $g \in M$ we have $\|y\| \leq \epsilon \leq \delta$ and $\|g(y)\| \leq \delta$ so we can use the estimate $\|h(g(y),y)\| \leq \|g(y)\|/3 + K\|y\| \leq \delta/3 + K\epsilon \leq 2\delta/3$. It follows that $Tg \in M$, so T gives a map $M \to M$.

Next, note that when $\|x_0\|, \|x_1\|, \|y\| \leq \delta$ we have

$$\|h(x_1,y) - h(x_0,y)\| = \left|\int_{t=0}^{1} \alpha(tx_1 + (1-t)x_0, y).(x_1 - x_0)\,dt\right| \leq \int_{t=0}^{1} \|\alpha(tx_1 + (1-t)x_0, y)\|.\|x_1 - x_0\|\,dt \leq \|x_1 - x_0\|/3.$$

It follows that $\|Tg_0 - Tg_1\| \leq \|g_0 - g_0\|/3$ for all $g_0, g_1 \in M$, so T is a contraction mapping on M. If we let g denote the unique fixed point of T, we find that $g(y) = h(g(y),y) = g(y) - f(g(y),y)$, so $f(g(y),y) = 0$ as required. $\square$

It is sometimes useful to know that the fixed point of f depends continuously on X. This can be formalised as follows.

Proposition 12.45. [`prop-fp-cts`]
Let X be a complete metric space, and let α be a number in $(0,1)$. Let $CM_\alpha(X)$ be the set of contraction mappings of ratio α on X, and define $\phi\colon CM_\alpha(X) \to X$ by

$$\phi(f) = \text{ the unique fixed point of } f.$$

Then $CM_\alpha(X)$ is closed in $C(X,X)$, and ϕ is continuous.

Proof. First suppose that $f \in C(X,X) \setminus CM_\alpha(X)$. This means that there exist points $x, y \in X$ with $d(f(x), f(y)) > \alpha d(x,y)$. Put $\epsilon = (d(f(x), f(y)) - \alpha d(x,y))/2 > 0$. If $d(f,g) < \epsilon$ we find that

$$d(g(x), g(y)) \geq d(f(x), f(y)) - d(f(x), g(x)) - d(f(y), g(y)) > d(f(x), f(y)) - 2\epsilon > \alpha d(x,y),$$

so $g \notin CM_\alpha(X)$. This means that the complement of $CM_\alpha(X)$ is open, so $CM_\alpha(X)$ is closed. Next, suppose we have $f, g \in CM_\alpha(X)$. Let x be the fixed point of f, and let y be the fixed point of g. Now the last part of Theorem 12.42 tells us that $d(y,x) \leq d(y, f(y))/(1-\alpha)$. As $g(y) = y$, we can rewrite this as

$$d(y,x) \leq \frac{d(g(y), f(y))}{1-\alpha} \leq \frac{d(f,g)}{1-\alpha},$$

or in other words $d(\phi(f), \phi(g)) \leq d(f,g)/(1-\alpha)$. This means that ϕ is Lipschitz and therefore continuous. □

12.2. Uniform continuity.

Definition 12.46. [`defn-uniformly-cts`]
Let $f\colon X \to Y$ be a map of metric spaces. We say that f is *uniformly continuous* if for all $\epsilon > 0$ there exists $\delta > 0$ such that $d(f(x), f(x')) < \epsilon$ whenever $d(x, x') < \delta$.

Remark 12.47. [`rem-uniformly-cts`]
Note here that δ is not allowed to depend on x (or x'). If δ were allowed to depend on x, this would just reduce to the ordinary notion of continuity. Thus, any uniformly continuous map is continuous.

Remark 12.48. [`rem-lipschitz-uniform`]
If $f\colon X \to Y$ is Lipschitz, with Lipschitz constant A say, then we can just take $\delta = \epsilon/A$ to see that f is uniformly continuous.

Proposition 12.49. [`prop-uniformly-cts`]
Let $f\colon X \to Y$ be a continuous map of metric spaces, and suppose that X is compact. Then f is uniformly continuous.

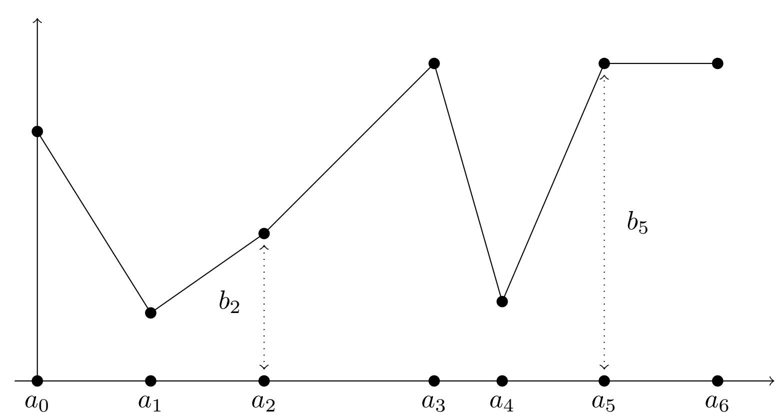
Proof. Suppose we are given $\epsilon > 0$. For each $y \in Y$, put $U_y = f^{-1}(OB_{\epsilon/2}(y))$, which is open in X. As $x \in U_{f(x)}$ for all x, we see that this gives an open covering of X. Let δ be a Lebesgue number for this covering. Now suppose we have points x and x' in X with $d(x,x') < \delta$, so $x' \in OB_\delta(x)$. As δ is a Lebesgue number we have $OB_\delta(x) \subseteq U_y$ for some y. This means that $f(x)$ and $f(x')$ lie in $OB_{\epsilon/2}(y)$, so $d(f(x), f(x')) \leq d(f(x), y) + d(y, f(x')) < \epsilon$ as required. □

As an application, we can now exhibit a countable dense subset of the space $C([0,1])$.

Definition 12.50. [`defn-piecewise-linear`]
Consider a function $f\colon [0,1] \to \mathbb{R}$. We say that f is *piecewise-linear* if there exist numbers $0 = a_0 < a_1 < \cdots < a_r = 1$ such that $f|_{[a_i, a_{i+1}]}$ is linear for all i. Equivalently, if we put $b_i = f(a_i)$, we should have

$$f(x) = b_i + \frac{(x-a_i)(b_{i+1}-b_i)}{(a_{i+1}-a_i)}$$

for all $x \in [a_i, a_{i+1}]$.



We will say that f is *rationally piecewise-linear* if in addition, the numbers a_i and b_i are rational. We write RPL for the set of all rationally piecewise-linear functions (which is easily seen to be countable).

PROPOSITION 12.51. [`prop-rpl-dense`]
The set RPL is dense in $C([0,1])$ (so $C([0,1])$ is separable).

PROOF. Suppose we are given $f \in C([0,1])$ and $\epsilon > 0$. Proposition 12.49 tells us that f is uniformly continuous, so there exists $n > 0$ such that $|f(x)-f(y)| < \epsilon/5$ whenever $|x-y| \leq 1/n$. For each $k \in \{0,\dots,n\}$, let b_k be a rational number with $|b_k - f(k/n)| < \epsilon/5$. Let $g\colon [0,1] \to \mathbb{R}$ be the function that satisfies $g(k/n) = b_k$ for $k = 0,1,\dots,n$ and is linear on each interval $[k/n,(k+1)/n]$, so $g \in RPL$. Consider a point $x \in [0,1]$. Let k be the integer such that $k/n \leq x < (k+1)/n$. If $k = n$ then we must have $x = 1$ and $|f(x) - g(x)| = |f(1) - b_n| < \epsilon/5 < 4\epsilon/5$. Suppose instead that $k < n$. As g is linear on $[k/n,(k+1)/n]$ we have

$$|g(x) - b_k| \leq |b_{k+1} - b_k| \leq |b_{k+1} - f((k+1)/n)| + |b_k - f(k/n)| + |f((k+1)/n) - f(k/n)| < 3\epsilon/5,$$

so

$$|g(x) - f(x)| \leq |g(x) - b_k| + |f(x) - b_k| < 3\epsilon/5 + \epsilon/5 = 4\epsilon/5.$$

As x was arbitrary, it follows that $d(f,g) \leq 4\epsilon/5 < \epsilon$ as required. □

12.3. Spaces of subsets. We next discuss ways to measure the distance between two subsets of a given metric space X. This will allow us to consider the set of closed subsets of X as a metric space in its own right. Later (in Definition 21.45) we will give a non-metric version by defining a topology on the space of closed subsets of an arbitrary compact Hausdorff space.

DEFINITION 12.52. [`defn-hausdorff-metric`]
Let X be a metric space.

(a) Given a point $x \in X$ and a subset $Y \subseteq X$, we put $d(x,Y) = \inf\{d(x,y) \;:\; y \in Y\}$. This is interpreted as ∞ if $Y = \emptyset$.
(b) Given subsets $Y, Z \subseteq X$ we put

$$d'(Y,Z) = \sup\{d(y,Z) \;:\; y \in Y\}$$
$$\overline{d}(Y,Z) = \max(d'(Y,Z), d'(Z,Y)).$$

This is interpreted as ∞ if $Y = \emptyset$ or $Z = \emptyset$.

LEMMA 12.53. [`lem-dbar-zero`]
We have $d(x,Z) = 0$ if and only if $x \in \overline{Z}$. Thus, we have $d'(Y,Z) = 0$ iff $Y \subseteq \overline{Z}$.

PROOF. Suppose that $d(x,Z) = 0$. This means that for all $\epsilon > 0$, the number ϵ is not a lower bound for $\{d(x,z) \;:\; z \in Z\}$, so we can choose $z \in Z$ with $d(x,z) < \epsilon$, so $OB_\epsilon(x)$ meets Z. This means that $x \in \overline{Z}$. All steps in this argument can be reversed, so $d(x,Z) = 0$ iff $x \in \overline{Z}$. It follows that $d'(Y,Z) = 0$ iff ($d(y,Z) = 0$ for all $y \in Y$) iff $Y \subseteq \overline{Z}$. □

LEMMA 12.54. [`lem-dbar-lipschitz`]
Suppose that $d(x,y) < \infty$ for all $x,y \in X$, and that $Y \neq \emptyset$. Then for all $x,x' \in X$ we have $|d(x,Y) - d(x',Y)| \leq d(x,x')$. The map $x \mapsto d(x,Y)$ is therefore Lipschitz and so continuous.

PROOF. The auxiliary conditions ensure that $d(x,Y) < \infty$ for all x, so the expression $d(x,Y) - d(x',Y)$ is meaningful. For all $y \in Y$ we have

$$d(x,Y) \leq d(x,y) \leq d(x,x') + d(x',y).$$

This means that $d(x,Y) - d(x,x')$ is a lower bound for the numbers $d(x',y)$, and $d(x',Y)$ is by definition the greatest lower bound, so we have $d(x,Y) - d(x,x') \leq d(x',Y)$, or equivalently $d(x,Y) - d(x',Y) \leq d(x,x')$. By reversing the roles of x and x' we see that also $d(x',Y) - d(x,Y) \leq d(x,x')$, so $|d(x,Y) - d(x',Y)| \leq d(x,x')$. □

Lemma 12.55. [lem-hausdorff-metric]
Let P denote the set of subsets of X. Then the function $\overline{d}$ defines a semimetric on P, with $\overline{d}(A,B)=0$ iff $\overline{A}=\overline{B}$.

Proof. It is clear that $\overline{d}(A,B)=\overline{d}(B,A)\geq 0$, and that $\overline{d}(A,A)=0$. Using Lemma 12.53 we also see that $\overline{d}(Y,Z)=0$ iff ($Y\subseteq\overline{Z}$ and $Z\subseteq\overline{Y}$), which is equivalent to $\overline{Y}=\overline{Z}$.

This just leaves the triangle inequality. Suppose that $d(A,B)=r$ and $d(B,C)=s$ and $\epsilon>0$. If $a\in A$ then $d(a,B)\leq r$, so there exists $b\in B$ with $d(a,b)<r+\epsilon$. Similarly, we then have $d(b,C)\leq s$, so there exists $c\in C$ with $d(b,c)<s+\epsilon$. We have now found $c\in C$ with $d(a,c)<r+s+2\epsilon$, so $d(a,C)\leq r+s+2\epsilon$. As $\epsilon>0$ was arbitrary, we deduce that $d(a,C)\leq r+s$. A similar argument shows that for all $c\in C$ we have $d(c,A)\leq r+r$, so $\overline{d}(A,B)\leq r+s$ as required. □

Proposition 12.56. [prop-hausdorff-metric]
Let X be a compact metric space, and let K be the set of closed subsets of X. Then $\overline{d}$ gives a metric on K, and it is compact with respect to the metric topology.

Proof. It is immediate from Lemma 12.55 that $\overline{d}$ gives a metric on K. By Theorem 12.28, it will suffice to show that K is totally bounded and complete with respect to $\overline{d}$. Consider a number $\epsilon>0$. As X is compact, there exists an $\epsilon/2$-net F for X. Let G be the set of subsets of F, which is a finite subset of K. For any $A\in K$, put $B=\{x\in F\;:\;d(x,A)\leq\epsilon/2\}$, so $B\in G$. By definition we have $d'(B,A)\leq\epsilon/2$. If $y\in A$ then (because F is an $\epsilon/2$-net) there exists $x\in F$ with $d(x,y)<\epsilon/2$. This means that $d(x,A)<\epsilon/2$, so $x\in B$. As $x\in B$ and $d(x,y)<\epsilon/2$ we have $d(y,B)<\epsilon/2$. As y was an arbitrary element of A, this gives $d'(A,B)\leq\epsilon/2<\epsilon$. Thus, G is an ϵ-net in K. It follows that K is totally bounded.

Now consider a Cauchy sequence $(A_n)_{n=0}^\infty$ in K. By the Cauchy property, we can choose $n_0<n_1<n_2<\cdots$ such that $\overline{d}(A_i,A_j)<2^{-k}$ whenever $i,j\geq n_k$. Put $B_i=A_{n_i}$, so $\overline{d}(B_i,B_j)<2^{-\min(i,j)}$. Then put

$$C=\{x\in X\;:\;d(x,B_i)\leq 2^{-i}\text{ for all }i\}.$$

By construction we have $d'(C,B_i)\leq 2^{-i}$. Consider a point $x\in B_i$. Put

$$D_j=\{y\in X\;:\;d(y,B_j)\leq 2^{-j},\ d(x,y)\leq 2^{-i}\},$$

and note that this is closed in X. We claim that for all k, we have $\bigcap_{j<k}D_j\neq\emptyset$. Indeed, it will be harmless to assume that $k\geq i$. We then have $d(B_i,B_k)<2^{-i}$ and $x\in B_i$, so we can choose $y\in B_k$ with $d(x,y)<2^{-i}$. Now for $j<k$ we have

$$d(y,B_j)\leq d'(B_k,B_j)\leq\overline{d}(B_k,B_j)\leq 2^{-\min(k,j)}=2^{-j},$$

so $y\in D_j$ as required. This shows that the family $(D_j)_{j\in\mathbb{N}}$ has the finite intersection property, but X is compact, so $\bigcap_j D_j\neq\emptyset$. From the definitions we have $\bigcap_j D_j=\{y\in C\;:\;d(x,y)\leq 2^{-i}\}$; as this set is nonempty, we have $d(x,C)\leq 2^{-i}$. As x was an arbitrary point in B_i, we deduce that $d'(B_i,C)\leq 2^{-i}$. We also saw previously that $d'(C,B_i)\leq 2^{-i}$, so $\overline{d}(B_i,C)\leq 2^{-i}$, so the sequence $(B_i)_{i\geq 0}$ converges to C. It follows by Lemma 12.6 that the original sequence $(A_i)_{i\geq 0}$ also converges to C. □

Remark 12.57. Note that $d(\emptyset,A)=\infty$ for all $A\neq\emptyset$, so the empty set gives an isolated point in K. It follows that the set $K'=K\setminus\{\emptyset\}$ is closed in K and so is again compact.

There are interesting applications of the above theory to the study of fractals and iterated function systems. Fractals are subspaces $T\subseteq\mathbb{R}^n$ that are usually self-similar in some sense. One possible sense is that $T=T_1\cup\cdots\cup T_r$, where each T_i is the image of some continuous map $f_i\colon T\to T$. To construct examples, we can start with a cube $X=[-R,R]^n$ and a list of maps $f_i\colon X\to X$ (for $1\leq i\leq r$ say). Let K' be the space of nonempty compact subsets of X as before, and define $F\colon K\to K$ by $F(A)=\bigcup_{i=1}^r f_i(A)$. If we can show that F is a contraction mapping, then the Contraction Mapping Theorem (Theorem 12.42) will tell us that there is a unique point $T\in K$ with $T=F(T)$, or in other words a unique compact subset

$T \subseteq X$ with the self-similarity property $T = \bigcup_i f_i(T)$. For example, we can define maps $f_i : \mathbb{R}^2 \to \mathbb{R}^2$ by

$$f_0 \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.16 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$
$$f_1 \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0.85 & 0.04 \\ -0.04 & 0.85 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} 0.00 \\ 1.60 \end{bmatrix}$$
$$f_2 \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -0.15 & 0.28 \\ 0.26 & 0.24 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} 0.00 \\ 0.44 \end{bmatrix}$$
$$f_3 \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0.20 & -0.26 \\ 0.23 & 0.22 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} 0.00 \\ 1.60 \end{bmatrix}$$

One can check that these restrict to give contraction mappings $f_i : X \to X$, where $X = [-20, 20]^2$. Moreover, the resulting map $F : K' \to K'$ is again a contraction mapping, so there is a unique fixed set T as discussed. This set T is called *Barnsley's Fern*.

To verify that F is a contraction mapping, we need the following lemmas:

Lemma 12.58. [`lem-matrix-lipschitz`]
Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be a map of the form $f(x) = Ax + a$, for some matrix $A \in M_n(\mathbb{R})$ and some vector $a \in \mathbb{R}^n$. Let k be the maximum absolute value of the eigenvalues of $A^T A$. Then $\sqrt{k}$ is a Lipschitz constant for f (with respect to the metric $d_2(x, y) = \|x - y\|_2$ on $\mathbb{R}^n$). In particular, if $k < 1$ then f is a contraction mapping.

Proof. First, we note that the matrix $B = A^T A$ is symmetric, so by standard linear algebra, there is an orthonormal basis $u_1, \dots, u_n$ for $\mathbb{R}^n$ and real numbers $t_1, \dots, t_n$ with $Bu_i = t_i u_i$ for all i. After renumbering the basis elements if necessary, we may assume that $t_1 \leq \cdots \leq t_n$. Note that for any vector v we have

$$\langle v, Bv \rangle = \langle v, A^T Av \rangle = \langle Av, Av \rangle = \|Av\|^2 \geq 0.$$

Taking $v = e_i$, we deduce that $t_i \geq 0$. It follows that the number k in the statement is just t_n. Moreover, we can write v as $\sum_{i=1}^n s_i u_i$ say, and we find that

$$\|Av\|^2 = \langle v, Bv \rangle = \langle \sum_i s_i u_i, \sum_i s_i t_i u_i \rangle = \sum_i s_i^2 t_i \leq k \sum_i s_i^2 = k\|v\|^2,$$

so $\|Av\| \leq \sqrt{k}\|v\|$. Taking $v = x - y$, we deduce that

$$d_2(f(x), f(y)) = \|A(x - y)\| \leq \sqrt{k}\|x - y\| = \sqrt{k}\, d_2(x, y)$$

as required. $\square$

LEMMA 12.59. [lem-IFS-contraction]
Let X be a metric space, and let $f_1, \dots, f_r$ be contraction mappings on X. Let K' be the space of nonempty compact subsets of X, and define $F\colon K' \to K'$ by $F(A) = \bigcup_i f_i(A)$. Then F is also a contraction mapping.

PROOF. As f_i is a contraction mapping, there exists $m_i \in (0,1)$ such that $d(f_i(x), f_i(y)) \leq m_i d(x,y)$ for all x and y. Put $m = \max(m_1, \dots, m_r) \in (0,1)$. Suppose we have distinct sets $A, B \in K'$ and a number $t > d(A,B)$. This means precisely that for all $a \in A$ there exists $b \in B$ with $d(a,b) < t$, and similarly for all $b \in B$ there exists $a \in A$ with $d(a,b) < t$. Now suppose we have a point $a' \in F(A)$. This means that $a' = f_i(a)$ for some i and some $a \in A$. There exists $b \in B$ with $d(a,b) < t$, and we put $b' = f_i(b) \in F(B)$. We find that

$$d(a', b') = d(f_i(a), f_i(b)) \leq m_i d(a,b) < mt.$$

Similarly, if $b' \in F(B)$ there exists $a' \in F(A)$ with $d(a', b') < mt$. It follows that $d(F(A), F(B)) \leq mt$. We can now let t approach $d(A,B)$ to see that $d(F(A), F(B)) \leq m\, d(A,B)$, so m is a Lipschitz constant for F, as required. $\square$

EXERCISE 12.1. [ex-dbar]
Let X be a metric space, and let Y be a closed subset. Define $e\colon X^2 \to [0,\infty)$ by

$$e(a,b) = \min(d(a,b), d(a,Y) + d(b,Y)).$$

Prove that this gives a semimetric on X.

REMARK 12.60. The intuition for e is as follows. We imagine that there is a new hyperspace travel facility covering the region Y. To move from a to b, we either use the old route (of length $d(a,b)$) or we travel a distance $d(a,Y)$ to the nearest available point in Y, then jump instantaneously to a point that is as close as possible to b, then travel a distance $d(b,Y)$ to b.

Solution: It is immediate that $e(a,a) = 0$ and $e(a,b) = e(b,a)$. Thus we need only show that

$$e(a,c) \leq e(a,b) + e(b,c)$$

We need to separate four cases. For brevity we write $P(a,b)$ to mean that $d(a,b) \leq \overline{d}(a,Y) + \overline{d}(b,Y)$ and $Q(a,b)$ to mean that $d(a,b) \geq \overline{d}(a,Y) + \overline{d}(b,Y)$. Note that $P(a,b)$ implies that $e(a,b) = d(a,b)$, and so on.

(a) Suppose that $P(a,b)$ and $P(b,c)$ hold. Then

$$e(a,c) \leq d(a,c) \leq d(a,b) + d(b,c) = e(a,b) + e(b,c)$$

(b) Suppose $P(a,b)$ and $Q(b,c)$. Using

$$d(a,Y) \leq d(a,b) + d(b,Y)$$

we get

$$\begin{aligned} e(a,c) &\leq \overline{d}(a,Y) + \overline{d}(c,Y) \\ &\leq d(a,b) + \overline{d}(b,Y) + \overline{d}(c,Y) \\ &= e(a,b) + e(b,c) \end{aligned}$$

(c) The case when $Q(a,b)$ and $P(b,c)$ hold is similar.
(d) Suppose $Q(a,b)$ and $Q(b,c)$. Then

$$\begin{aligned} e(a,c) &\leq \overline{d}(a,Y) + \overline{d}(c,Y) \\ &\leq \overline{d}(a,Y) + \overline{d}(b,Y) + \overline{d}(b,Y) + \overline{d}(c,Y) \\ &= e(a,b) + e(a,c) \end{aligned}$$

13. Completion

Definition 13.1. [defn-completion]

An *isometry* (or *isometric embedding*) is a map $f\colon X \to Y$ of metric spaces such that $d(f(x_0), f(x_1)) = d(x_0, x_1)$ for all $x_0, x_1 \in X$. An *isometric isomorphism* is a bijective isometry. A *completion* of X is a complete metric space Y together with an isometry $i\colon X \to Y$ such that $i(X)$ is dense in Y.

Note that an isometry is automatically continuous and injective. A bijective isometry is a homeomorphism and the inverse is an isometry. If $f\colon X \to Y$ is an isometry, then the induced map $f\colon X \to f(X)$ (where $f(X) \subseteq Y$ is given the subspace topology) is a homeomorphism, hence the use of the term "embedding".

Construction 13.2. [cons-completion]

Let X be a metric space. Write

$$CS(X) = \{ \text{ Cauchy sequences } \underline{x} = (x_n)_{n\in\mathbb{N}} \text{ in } X \ \}.$$

For $\underline{x}, \underline{y} \in CS(X)$ write

$$d(\underline{x}, \underline{y}) = \lim_{n\to\infty} d(x_n, y_n).$$

We will check later that this limit really exists, and that the resulting function d is a semimetric on $CS(X)$. We can thus introduce an equivalence relation on $CS(X)$ by ($\underline{x}E\underline{y}$ iff $d(\underline{x}, \underline{y}) = 0$), and we find (as in Remark 2.34) that the quotient set $\widetilde{X} = CS(X)/E$ is a metric space. We define $i\colon X \to \widetilde{X}$ by sending x to the equivalence class of the constant sequence $(x, x, x, \dots)$.

Proposition 13.3. [prop-completion]

Everything above makes sense, and $i\colon X \to \widetilde{X}$ becomes a completion of X. Moreover, for any Cauchy sequence $\underline{x}$ in X, the resulting sequence $(i(x_n))_{n\in\mathbb{N}}$ converges to $[\underline{x}]$ in $\widetilde{X}$.

Proof. First, consider a pair of Cauchy sequences $\underline{x}$ and $\underline{y}$. We must check that the sequence

$$(d(x_n, y_n))_{n\in\mathbb{N}}$$

is convergent, so that the definition of $d(\underline{x}, \underline{y})$ is meaningful. As $\mathbb{R}$ is complete it will suffice to show that the sequence is Cauchy. Suppose we are given $\epsilon > 0$. As $\underline{x}$ is Cauchy we can choose N such that $d(x_i, x_j) < \epsilon/2$ whenever $i, j \geq N$. Similarly, we can choose M such that $d(y_i, y_j) < \epsilon/2$ whenever $i, j \geq M$. Now when $i, j \geq \max(N, M)$ we have

$$d(x_i, y_i) \leq d(x_i, x_j) + d(x_j, y_j) + d(y_j, y_i) < \epsilon/2 + d(x_j, y_j) + \epsilon/2 = d(x_j, y_j) + \epsilon.$$

By a symmetrical argument we also have $d(x_j, y_j) < d(x_i, y_i) + \epsilon$, so $|d(x_i, y_i) - d(x_j, y_j)| < \epsilon$ as required.

We now see that d is a well-defined function from $CS(X) \times CS(X)$ to $[0, \infty)$. We claim that it is in fact a semimetric. Indeed, axioms M0 and M1 are clear. For the triangle inequality M2, suppose we have a third Cauchy sequence $\underline{z}$. The triangle inequality for X means that we have $d(x_i, z_i) \leq d(x_i, y_i) + d(y_i, z_i)$ in $\mathbb{R}$. We can then pass to the limit as $i \to \infty$ to see that $d(\underline{x}, \underline{z}) \leq d(\underline{x}, \underline{y}) + d(\underline{y}, \underline{z})$ as required. This means that there is a well-defined metric $\overline{d}$ on the quotient set $\widetilde{X}$ satisfying $\overline{d}([\underline{x}], [\underline{y}]) = d(\underline{x}, \underline{y})$ for all $\underline{x}, \underline{y} \in CS(X)$. Now let $i'(x)$ be the constant sequence with value x, so that $i(x) = [i'(x)]$. It is clear that $d(i'(x), i'(y)) = d(x, y)$, so $\overline{d}(i(x), i(y)) = d(x, y)$, so the map $i\colon X \to \widetilde{X}$ is an isometric embedding.

We next claim that the image of i is dense. To see this, consider a point $[\underline{x}] \in \widetilde{X}$, and a number $\epsilon > 0$. We must show that $OB_\epsilon([\underline{x}])$ meets $i(X)$. As $\underline{x}$ is Cauchy, we can choose $N \in \mathbb{N}$ such that $d(x_i, x_j) < \epsilon/2$ for $i, j \geq N$. In particular we see that $d(x_i, x_N) \leq \epsilon/2$ for all $i \geq N$, and by passing to the limit we see that $\overline{d}([\underline{x}], i(x_N)) \leq \epsilon/2 < \epsilon$ as required. This also proves that $(i(x_n))_{n\in\mathbb{N}}$ converges to $[\underline{x}]$ as claimed in the proposition.

All that is now left is to show that the space $\widetilde{X}$ is actually complete. Consider a Cauchy sequence $(u_n)_{n\in\mathbb{N}}$ in $\widetilde{X}$. By the previous paragraph, we can choose $x_n \in X$ such that $\overline{d}(i(x_n), u_n) < 2^{-n}$. It follows that

$$d(x_n, x_m) = d(i(x_n), i(x_m)) \leq d(i(x_n), u_n) + d(u_n, u_m) + d(u_m, i(x_m)) < 2^{-n} + 2^{-m} + d(u_n, u_m).$$

From this we can easily see that the sequence $\underline{x}$ is Cauchy, so it determines a point $a = [\underline{x}] \in \widetilde{X}$. We claim that $u_n \to a$ in $\widetilde{X}$. To see this, suppose we are given $\epsilon > 0$. As $\underline{x}$ is Cauchy we can find N such that

$d(x_i, x_j) < \epsilon/2$ for $i, j \geq N$. As in the proof of density, we see that $d(i(x_j), a) \leq \epsilon/2$ for all $j \geq N$. We also have $d(i(x_j), u_j) < 2^{-j}$, and by taking j large enough we can arrange for this to be less than $\epsilon/2$, which gives $d(a, u_j) < \epsilon$. This shows that $u_j \to a$ so the sequence $\underline{u}$ is convergent, as required. □

Proposition 13.4. [`prop-completion-adjoint`]
If $f\colon X \to Y$ is an isometry and Y is a complete metric space then there is a unique isometry $\tilde{f}\colon \widetilde{X} \to Y$ such that $\tilde{f} \circ i = f$. It is given by

$$\tilde{f}([\underline{x}]) = \lim_{n\to\infty} f(x_n).$$

Proof. Let $\underline{x}$ be a Cauchy sequence in X. Then we have $d(f(x_n), f(x_m)) = d(x_n, x_m)$ (because f is an isometry) and it follows directly that $(f(x_n))_{n\in\mathbb{N}}$ is also a Cauchy sequence in Y. As Y is complete, it follows that this sequence converges to some point in Y. We can thus define $\widehat{f}\colon CS(X) \to Y$ by $\widehat{f}(\underline{x}) = \lim_{n\to\infty} f(x_n)$.

Now consider another Cauchy sequence $(x'_n)_{n\in\mathbb{N}}$ in X. We claim that

$$d(\widehat{f}(\underline{x}), \widehat{f}(\underline{x}')) = d(\underline{x}, \underline{x}') = \lim_{n\to\infty} d(x_n, x'_n).$$

To prove this, it will be convenient to write $b = \widehat{f}(\underline{x})$ and $b' = \widehat{f}(\underline{x}')$. Given $\epsilon > 0$ we can find an integer N such that when $n \geq N$ we have $d(f(x_n), b) < \epsilon/2$ and also $d(f(x'_n), b') < \epsilon/2$. It then follows that

$$d(f(x_n), f(x'_n)) \leq d(f(x_n), b) + d(b, b') + d(b', f(x'_n)) < d(b, b') + \epsilon.$$

By essentially the same argument we also have $d(b, b') < d(f(x_n), f(x'_n)) + \epsilon$, so $|d(f(x_n), f(x'_n)) - d(b, b')| < \epsilon$. We also have $d(f(x_n), f(x'_n)) = d(x_n, x'_n)$ because f is an isometry, so $|d(x_n, x'_n) - d(b, b')| < \epsilon$ as required.

In particular, we now see that if $[\underline{x}] = [\underline{x}']$ then $d(\underline{x}, \underline{x}') = 0$ and so $d(\widehat{f}(\underline{x}), \widehat{f}(\underline{x}')) = 0$. As Y is assumed to be a metric space (and not just a semimetric space) we conclude that $\widehat{f}(\underline{x}) = \widehat{f}(\underline{x}')$. We therefore have a well-defined map $\tilde{f}\colon \widetilde{X} \to Y$ given by $\tilde{f}([\underline{x}]) = \widehat{f}(\underline{x}) = \lim_{n\to\infty} f(x_n)$, and we see from the above that $\tilde{f}$ is an isometry. Note that $\tilde{f}(i(x))$ is the limit of the constant sequence $f(x)$, which is just $f(x)$; so $\tilde{f} \circ i = f$.

Finally, suppose that $g\colon \widetilde{X} \to Y$ is another isometry with $g \circ i = f$. For any Cauchy sequence $\underline{x}$ we have seen that $(i(x_n))_{n\in\mathbb{N}}$ converges to $[\underline{x}]$, so $g([\underline{x}])$ must be the limit of the points $g(i(x_n)) = f(x_n)$; this shows that $g = \tilde{f}$ as claimed. □

Remark 13.5. [`rem-completion-adjoint`]
The proposition can be reformulated in categorical terms as follows. Let **MSI** denote the category whose objects are metric spaces, and whose morphisms are isometric embeddings. Let **CMSI** be the full subcategory of complete metric spaces, and let $J\colon \mathbf{CMSI} \to \mathbf{MSI}$ be the inclusion functor. For $X \in \mathbf{MSI}$, put $CX = \widetilde{X}$, so $C\colon \mathrm{obj}(\mathbf{MSI}) \to \mathrm{obj}(\mathbf{CMSI})$. The proposition gives a canonical bijection $i^*\colon \mathbf{CMSI}(CX, Y) \to \mathbf{MSI}(X, JY)$ for all $X \in \mathbf{MSI}$ and $Y \in \mathbf{CMSI}$. By Proposition 36.129, there is a unique way to define C on morphisms such that C becomes a functor left adjoint to J. This means that **CMSI** is a reflective subcategory of **MSI**.

Proposition 13.6. [`prop-compiscl`]
Let X be a complete metric space, and let Y be a subspace of X. Then the completion $\widetilde{Y}$ is isometrically isomorphic to the closure $\overline{Y} = \mathrm{cl}_X(Y)$.

Proof. Write i for the isometric embedding of Y in its canonical completion $\widetilde{Y}$. The inclusion $j\colon Y \to X$ is an isometric embedding and X is complete so there is a unique isometric embedding $\tilde{j}\colon \widetilde{Y} \to X$ with $\tilde{j} \circ i = j$.

As $\tilde{j}$ is continuous, the set $\tilde{j}^{-1}(\overline{Y}) \subseteq \widetilde{Y}$ is closed and it contains $i(Y)$. However, $i(Y)$ is dense in $\widetilde{Y}$ so $\tilde{j}^{-1}(\overline{Y}) = \widetilde{Y}$ and so $\tilde{j}(\widetilde{Y}) \subseteq \overline{Y}$.

On the other hand, $\tilde{j}(\widetilde{Y})$ is isometrically isomorphic to the complete metric space $\widetilde{Y}$, so it is complete. However, a complete subspace of a metric space is closed and $\tilde{j}(\widetilde{Y}) \supseteq Y$ so $\tilde{j}(\widetilde{Y}) = \overline{Y}$. Thus $\tilde{j}\colon \widetilde{Y} \to \overline{Y}$ is an isometric isomorphism (and thus a homeomorphism). □

Proposition 13.7. [prop-banach-completion]
Let V be a vector space over $\mathbb{R}$ equipped with a norm, and let $\widetilde{V}$ be the completion of V with respect to the corresponding metric. Then $\widetilde{V}$ has a canonical structure as a Banach space, such that the map $i\colon V \to \widetilde{V}$ is linear and preserves norms.

Proof. Let $\underline{x}$ and $\underline{y}$ be Cauchy sequences in V, and let s and t be real numbers. Put $z_i = sx_i + ty_i$ for all i. We then have

$$d(z_i, z_j) = \|s(x_i - x_j) + t(y_i - y_j)\| \leq s\|x_i - x_j\| + t\|y_i - y_j\| = sd(x_i, x_j) + td(y_i, y_j).$$

From this we see that $\underline{z}$ is a Cauchy sequence. It follows that we can make $CS(V)$ into a vector space by defining $s\underline{x} + t\underline{y} = \underline{z}$. If we put $\|\underline{x}\| = d(\underline{x}, 0) = \lim_{n\to\infty} \|x_n\|$ we find that this is a seminorm on $CS(V)$ and that $d(\underline{x}, \underline{y}) = \|\underline{x} - \underline{y}\|$. Moreover, if we let $i'(x)$ denote the constant sequence $(x, x, x, \dots)$ then $\|i'(x)\| = \|x\|$. Now put $CS_0(V) = \{\underline{x} \;:\; \|\underline{x}\| = 0\}$. We find that this is a vector subspace and that $\widetilde{V} = CS(V)/CS_0(V)$. Everything else follows easily from this. $\square$

Example 13.8. [eg-l-two-functions]

We can complete the space $C([0,1])$ with respect to the norm $\|f\|_2 = \sqrt{\int_0^1 f(t)^2\,dt}$ to obtain a space known as $L^2([0,1])$. This is more usually defined in a different way, relying on the theory of Lebesgue integration, which we will not cover in this book. With that theory in hand, we can define $L^2_0([0,1])$ to be the set of Lebesgue-measurable functions $f\colon [0,1] \to \mathbb{R}$ for which $\int_0^1 f(t)^2\,dt < \infty$. For such f we put $\|f\|_2 = \sqrt{\int_0^1 f(t)^2\,dt}$; we find that this defines a seminorm. We will say that f is *null* if the Lebesgue measure of $f^{-1}\{0\}$ is one. It turns out that the set N of null functions is a vector subspace of $L^2_0([0,1])$ and that $\|f\|_2 = 0$ if and only if $f \in N$. We can thus define a norm on $L^2_0([0,1])/N$ by $\|f + N\|_2 = \|f\|_2$, and it can be shown that this is complete. One can also show that $C([0,1])$ is dense, so $L^2_0([0,1])/N$ can be identified with the completion $L^2([0,1])$.

14. Further separation axioms

Definition 14.1. [defn-sep-axioms]
Let X be a topological space.

(a) X is said to be T_0 if for any pair of distinct points y and z, there is either a neighbourhood of y that does not contain z, or a neighbourhood of z that does not contain y.
(b) X is said to be T_1 if for any pair of distinct points y and z, there is a neighbourhood of z that does not contain y.
(c) X is said to be T_2 (or Hausdorff) if for any pair of distinct points y and z, there is a neighbourhood U of y and a neighbourhood V of z such that $U \cap V = \emptyset$. (This is Definition 6.1, and is just repeated for ease of comparison.)
(d) X is said to be *regular* if it is T_1, and for any closed set $Y \subset X$ and any point $z \notin Y$ there exist disjoint open sets U and V such that $Y \subseteq U$ and $z \in V$.
(e) X is said to be *normal* if it is T_1, and for any disjoint closed sets $Y, Z \subset X$ there exist disjoint open sets U and V such that $Y \subseteq U$ and $Z \subseteq V$.

Remark 14.2. It is easy to see that T_2 implies T_1, and T_1 implies T_0.

Lemma 14.3. [lem-T-one]
X is T_1 if and only if each singleton set $\{y\}$ is closed in X.

Proof. Put $U = X \setminus \{y\}$. Note that z is distinct from y if and only if $z \in U$, and z has a neighbourhood that does not contain y if and only if z has a neighbourhood that is contained in U. Thus, the definition of T_1 is just that every point in U is interior, or equivalently that U is open, or equivalently that $\{y\}$ is closed. $\square$

Lemma 14.4. [lem-T-zero]
X is T_0 if and only if the following condition is satisfied: whenever y and z are distinct points of X, we have $\overline{\{y\}} \neq \overline{\{z\}}$.

Proof. Let y and z be distinct points of X. We first claim that $\overline{\{y\}} \subseteq \overline{\{z\}}$ if and only if $y \in \overline{\{z\}}$. Indeed, if $y \in \overline{\{z\}}$ then $\overline{\{z\}}$ is a closed set containing $\{y\}$, and $\overline{\{y\}}$ is the smallest closed set containing $\{y\}$, so $\overline{\{y\}} \subseteq \overline{\{z\}}$. The converse implication is clear.

We now negate both sides to see that $\overline{\{y\}} \not\subseteq \overline{\{z\}}$ if and only if $y \not\in \overline{\{z\}}$, and this holds if and only if y is not a closure point of $\{z\}$, or equivalently there is a neighbourhood of y that does not contain z.

After this translation the definition of T_0 reads as follows: for any two distinct points y and z, we either have $\overline{\{y\}} \not\subseteq \overline{\{z\}}$ or $\overline{\{z\}} \not\subseteq \overline{\{y\}}$. In other words, for any two distinct points y and z, we have $\overline{\{y\}} \neq \overline{\{z\}}$, as claimed. □

Proposition 14.5. [`prop-norm-reg-haus`]
Any normal space is regular, and any regular space is Hausdorff.

Proof. Any normal, regular or Hausdorff space is T_1, so we may assume that all singleton sets are closed. If we specialise the definition of regularity to the case where Y is a singleton, we get Hausdorff condition. If we specialise the definition of normality to the case where Z is a singleton, we get the regularity condition. □

Proposition 14.6. [`prop-reg-contra`]
Let X be a T_1 space.

(a) *X is regular if and only if the following holds: for each point y and each open neighbourhood U of y, there exists an open set V with $y \in V \subseteq \overline{V} \subseteq U$.*
(b) *X is normal if and only if the following holds: for each closed subset Y and each open set W containing Y, there exists an open set U with $Y \subseteq U \subseteq \overline{U} \subseteq W$.*

Proof. We will prove (b). The proof for (a) is essentially the same and is left to the reader. We can rewrite the condition in (b) in terms of the closed set $Z = W^c$ as follows: for each pair of disjoint closed sets Y and Z, there exists an open set U such that $Y \subseteq U$ and $\overline{U} \subseteq Z^c$. Here the condition $\overline{U} \subseteq Z^c$ is equivalent to $\overline{U} \cap Z = \emptyset$ or $Z \subseteq \overline{U}^c$. If this holds we can take $V = \overline{U}^c$ and we then have disjoint open sets U and V with $Y \subseteq U$ and $Z \subseteq V$, as in the definition of normality.

Conversely, suppose that X is normal. We then have disjoint open sets U and V with $Y \subseteq U$ and $Z \subseteq V$. This means that U is contained in the closed set V^c, so $\overline{U}$ must also be contained in V^c. We also have $Z \subseteq V$ and so $V^c \subseteq Z^c = W$, so $Y \subseteq U \subseteq \overline{U} \subseteq W$ as in (b). □

Lemma 14.7. [`lem-normal-open`]
Let X be a normal space, and let $U_1, \dots, U_n$ be a finite open cover of X. Then there exists another open cover $V_1, \dots, V_n$ such that $\overline{V_i} \subseteq U_i$ for all i.

Proof. Suppose we have found open sets $V_1, \dots, V_{m-1}$ (for some $m \leq n$) such that $\overline{V_i} \subseteq U_i$ for all $i < m$, and

$$V_1 \cup \cdots \cup V_{m-1} \cup U_m \cup \cdots \cup U_n = X.$$

Put

$$U'_m = V_1 \cup \cdots \cup V_{m-1} \cup U_{m+1} \cup \cdots \cup U_n,$$

so by assumption we have $U'_m \cup U_m = X$. This means that the sets $F'_m = X \setminus U'_m$ is closed and contained in the open set U_m. By Proposition 14.6(b), we can choose an open set V_m with $F'_m \subseteq V_m \subseteq \overline{V_m} \subseteq U_m$. As $F'_m \subseteq V_m$, we see that

$$V_1 \cup \cdots \cup V_m \cup U_{m+1} \cup \cdots \cup U_n = X.$$

Note that this construction works perfectly well even when $m = 1$, and in that case we do not need to have chosen any sets V_i. We can thus start an induction, and after n steps we reach the stated conclusion. □

Proposition 14.8. [`prop-metric-normal`]
Any metric space is normal.

Proof. Let X be a metric space, and let Y and Z be disjoint closed subsets of X. We must find disjoint open sets U and V with $Y \subseteq U$ and $Z \subseteq V$. If $Y = \emptyset$ we just take $U = \emptyset$ and $V = X$. Similarly, if $Z = \emptyset$ we just take $U = X$ and $V = \emptyset$. We may therefore assume that Y and Z are nonempty. We can also change the metric as in Proposition 2.44 if necessary, and thus assume that $d(x,x')<\infty$ for all $x,x'\in X$. We can now define $d(x,Y)$ and $d(x,Z)$ as in Definition 12.52, and Lemma 12.54 tells us that these are continuous. It follows that the expression $f(x)=d(x,Z)-d(x,Y)$ gives a continuous map $f\colon X\to\mathbb{R}$. We thus have disjoint open sets

$$U=\{x\;:\;f(x)>0\}=f^{-1}(0,\infty)$$
$$V=\{x\;:\;f(x)<0\}=f^{-1}(-\infty,0).$$

Moreover, Lemma 12.53 tells us that $d(x,Y)=0$ iff $x\in Y$, and $d(x,Z)=0$ iff $x\in Z$. Using this we see that $Y\subseteq U$ and $Z\subseteq V$, as required. □

Proposition 14.9. [`prop-comp-haus-normal`]
Any compact Hausdorff space is normal.

We will deduce this from the following lemma:

Lemma 14.10. [`lem-comp-haus-normal`]
If X is Hausdorff, $Y\subseteq X$ is compact and $z\not\in Y$ then are disjoint open sets U and V such that $y\in U$ and $Z\subseteq V$. It follows that $y\not\in\overline{V}$ and $\overline{U}\cap Z=\emptyset$.

Proof. For each $y\in Y$ we have $y\neq z$, so (by the Hausdorff condition) we can choose disjoint open sets U_y and V_y with $y\in U_y$ and $z\in V_y$. The family $(U_y)_{y\in Y}$ is then an open cover of the compact set Y, so there is a finite subset $J\subseteq Y$ such that $Y\subseteq\bigcup_{y\in J}U_y$. Now put $U=\bigcup_{y\in J}U_y$ and $V=\bigcap_{y\in J}V_y$. As J is finite these are again open. By construction we have $Y\subseteq U$. As $z\in V_y$ for all y, we also have $z\in V$. If $u\in U$ then $u\in U_y$ for some $y\in J$, so $u\not\in V_y$ (because U_y and V_y are disjoint), so $u\not\in V$. This shows that $U\cap V=\emptyset$, as required. Now V is contained in the closed se U^c, so $\overline{V}\subseteq U^c$ but $y\in U$ so $y\not\in\overline{V}$. Similarly, U is contained in the closed set V^c, so $\overline{U}\subseteq V^c$ but $Z\subseteq V$ so $\overline{U}\cap Z=\emptyset$. □

Proof of Proposition 14.9. Let X be a compact Hausdorff space, and let Y and Z be disjoint closed subsets of X. Proposition 10.13 tells us that Y and Z are also compact. For each $z\in Z$ we can therefore apply Lemma 14.10 to see that there exist disjoint open sets U_z and V_z, with $Y\subseteq U_z$ and $z\in V_z$. The family $(V_z)_{z\in Z}$ is then an open cover of the compact set Z, so there exists a finite subset $K\subseteq Z$ such that $Z\subseteq\bigcup_{z\in K}V_z$. We now put $U=\bigcap_{z\in K}U_z$ and $V=\bigcup_{z\in K}V_z$. As $Y\subseteq U_z$ for all z we have $Y\subseteq U$. As K is finite we see that U is open. By construction we have $Z\subseteq V$, and V is clearly open. If $v\in V$ then $v\in V_z$ for some $z\in K$, so $v\not\in U_z$ (because U_z and V_z are disjoint), so $v\not\in U$. It follows that $U\cap V=\emptyset$, as required. □

Proposition 14.11. [`prop-hausdorff-quotient`]
Let X be a compact Hausdorff space, and let E be an equivalence relation that is closed as a subset of X^2. Then the space X/E is also compact Hausdorff.

Proof. By a straightforward argument that we gave as Corollary 10.21, any quotient of a compact space is compact. The real point is to show that X/E is Hausdorff.

First, let $q\colon X\to X/E$ be the quotient map. Define $i_x\colon X\to X^2$ by $i_x(y)=(x,y)$. The equivalence class $[x]$ can be described as $i_x^{-1}(E)$, so it is closed. We also have $[x]=q^{-1}\{q(x)\}$, so the singleton $\{q(x)\}$ is closed in the quotient topology, so X/E is T_1.

Now let U be an open subset of X, and put $U^*=\{x\;:\;[x]\subseteq U\}$. We claim that U^* is open in X. To see this, note that the set $F=(X\times(X\setminus U))\cap E$ is closed and therefore compact. Let G be the image of F under the projection $(x,y)\mapsto x$, so G is again compact and therefore closed in X. We have $x\in G$ iff there exists y with $y\not\in U$ and xEy (so $y\in[x]$). Equivalently, we have $x\in G$ iff $[x]\not\subseteq U$ iff $x\not\in U^*$, so $U^*=X\setminus G$, which is open as claimed. Note also that when xEy we have $[x]=[y]$ so $x\in U^*$ iff $y\in U^*$. Using this, we see that $U^*=q^{-1}(q(U^*))$. As U^* is open in X, we deduce that $q(U^*)$ is open in X/E.

Now suppose we have distinct points $y_0,y_1\in X/E$. Choose $x_0,x_1\in X$ with $y_i=q(x_i)$. As $y_0\neq y_1$ we see that the sets $[x_0]$ and $[x_1]$ are disjoint, and we saw above that they are also closed. As X is normal, we can choose disjoint open sets U_0 and U_1 with $[x_i]\subseteq U_i$ (so $x_i\in U_i^*$). Now put $V_i=q(U_i^*)$. These are disjoint open subsets of X/E with $y_i\in V_i$, as required. □

Proposition 14.12. [prop-quasi-components]

Suppose that X is compact Hausdorff, and $x \in X$. Let C be the connected component containing X. Then C is also the intersection of the family

$$\mathcal{F} = \{ \text{ clopen sets } F \ : \ x \in F\}.$$

Proof. For any $F \in \mathcal{F}$ we see that (F, F^c) gives a relative separation for the connected set C, with $x \in C \cap F$, so we must have $C \subseteq F$. Thus, if we let D denote the intersection of $\mathcal{F}$, then $C \subseteq D$. We next claim that D is connected. To see this, let A be a set that is clopen in D and contains x; we must show that $A = D$. As A is closed in D, it is closed in X. As A is open in D, we see that the set $D \setminus A$ is closed in D and therefore in X, so the set $A \cup D^c = X \setminus (D \setminus A)$ is open in X. As A is closed and $A \cup D^c$ is open, Proposition 14.6(b) gives us an open set U with $A \subseteq U \subseteq A \cup D^c$. Now consider the boundary $E = \overline{U} \setminus U$. This is compact, and it is covered by the family $\mathcal{V} = \{F^c \ : \ F \in \mathcal{F}\}$, so it is covered by some finite subfamily. As $\mathcal{F}$ is closed under finite intersections, this subfamily can be reduced to a singleton, so we have $E \subseteq F^c$ for some $F \in \mathcal{F}$. The relation $E \subseteq F^c$ gives $F \subseteq U \cup \overline{U}^c$, which gives $F \cap \overline{U} \subseteq U$, so $F \cap \overline{U} = F \cap U$. Call this set G. Forom the description $G = F \cap \overline{U}$ we see that G is closed, and from the description $G = F \cap U$ we see that it is also open, so $G \in \mathcal{F}$. From the definition of D we see that $D \subseteq G$, so $D \cap G = D$. On the other hand, we have $U \subseteq A \cup D^c$ by construction, which gives $D \cap G \subseteq A$. This gives $A = D$ as required. $\square$

Corollary 14.13. [cor-component-relation]

Let X be a compact Hausdorff space, and put $E = \bigcup_C (C \times C)$, where C runs over the connected components of X. Then E is closed in $X \times X$ and is an equivalence relation, so that the quotient X/E is again a compact Hausdorff space.

Proof. Suppose that $(x, y) \notin E$, so x and y lie in different components. By Proposition 14.12, there is a clopen set F such that $x \in F$ and $y \notin F$. This means that the set $F \times F^c$ is a neighbourhood of (x, y) disjoint from E. We now see that E is closed, and it is clearly an equivalence relation. The quotient is compact Hausdorff by Proposition 14.11. $\square$

We now explain another useful feature of compact Hausdorff spaces. For any spaces X and Y we define an evaluation map

$$\mathrm{ev}\colon C(X, Y) \times X \to Y$$

by $\mathrm{ev}(f)(x) = f(x)$. Now suppose that X is compact Hausdorff, and that Y is a metric space. We give $C(X, Y)$ the topology corresponding to the metric $d(f, g) = \max\{d(f(x), g(x)) \ : \ x \in X\}$, and then we give $C(X, Y) \times X \to Y$ the product topology.

Proposition 14.14. [prop-eval-cts]

If X is compact Hausdorff and Y is a metric space, then the evaluation map $\mathrm{ev}\colon C(X, Y) \times X \to Y$ is continuous.

Proof. Let $V \subseteq Y$ be open, and consider a point $(f, x) \in \mathrm{ev}^{-1}(V)$. This means that $f(x) = \mathrm{ev}(f, x) \in V$, so $x \in f^{-1}(V)$. Now f is continuous, so $f^{-1}(V)$ is open. Moreover, X is normal (and therefore regular) by Proposition 14.9. It follows that there exists an open set $U \subseteq X$ with $x \in U \subseteq \overline{U} \subseteq f^{-1}(V)$. Here $\overline{U}$ is closed in the compact space X, so it is itself compact, so the set $L = f(\overline{U})$ is also compact. Now put $G = Y \setminus V$. This is a closed set that does not meet L, so the function $r(y) = d(y, G)$ is continuous and strictly positive on L. As L is compact we see that $1/r$ must be bounded on L, say by $1/(2\epsilon)$; this means that for all $u \in \overline{U}$ and $y \notin V$ we have $d(f(u), y) \leq \epsilon/2 < \epsilon$. It follows in turn that if $(g, u) \in OB_\epsilon(f) \times U$ we have $\mathrm{ev}(g, u) = g(u) \in V$, so $(g, u) \in \mathrm{ev}^{-1}(V)$. This proves that $\mathrm{ev}^{-1}(V)$ is a neighbourhood of (f, x), as required. $\square$

15. The Baire category theorem

Definition 15.1. [defn-baire]

Let X be a topological space. We say that a subset $Y \subseteq X$ is *nowhere dense* if $\overline{Y}$ has empty interior. We say that X is *Baire* if every countable union of nowhere dense sets has empty interior.

Note that we do not expect that a countable union of nowhere dense sets will itself be nowhere dense. For example, we can consider the countable family $\{\{q\} : q \in \mathbb{Q}\}$ of subsets of $\mathbb{R}$. Each individual set $\{q\}$ is certainly nowhere dense. The union is $\mathbb{Q}$, which has empty interior, reflecting the fact (to be proved below) that $\mathbb{R}$ is a Baire space. However, we have $\overline{\mathbb{Q}} = \mathbb{R}$, which has nonempty interior; so $\mathbb{Q}$ is not nowhere dense.

Proposition 15.2. [`prop-baire-equiv`]
A space X is Baire if and only if every countable intersection of dense open sets of X is dense.

Proof. Suppose that X is Baire. Let $(U_n)_{n\in\mathbb{N}}$ be a countable family of dense open sets, so $\overline{U_n} = X$ for all n. Put $F_n = U_n^c$, and note that F_n is closed and $\mathrm{int}(F_n) = \mathrm{cl}(U_n)^c = X^c = \emptyset$, so F_n is nowhere dense. Now put $A = \bigcup_n F_n$, so $A^c = \bigcap_n U_n$; we must show that A^c is dense. The Baire condition gives $\mathrm{int}(A) = \emptyset$, and thus $\mathrm{cl}(A^c) = \mathrm{int}(A)^c = X$ as required.

We leave it to the reader to check that the whole argument is reversible. $\square$

Theorem 15.3 (The Baire Category Theorem). [`thm-metric-baire`]
Every complete metric space is Baire.

The proof will be given after two preliminary results.

Remark 15.4. The name of this result refers to a traditional meaning of the word "category" which is now rarely used. It has nothing to do with the theory of categories and functors.

Lemma 15.5. [`lem-dense-open`]
Let X be an arbitrary space, and let $U_1, \dots, U_n$ be a finite list of open dense sets. Then $U_1 \cap \dots \cap U_n$ is also open and dense.

Proof. By an evident induction we can reduce to the case $n = 2$. Let x be an arbitrary point of X, and consider a number $\epsilon > 0$. As U_1 is dense, the set $V = OB_\epsilon(x) \cap U_1$ must be nonempty. Choose $y \in V$. Note that U_2 is dense, so y must be a closure point, so the open neighbourhood V of y must meet U_2. This means that the set $OB_\epsilon(x) \cap (U_1 \cap U_2) = V \cap U_2$ must be nonempty. As x and ϵ were arbitrary, this means that $U_1 \cap U_2$ is dense as claimed. $\square$

Corollary 15.6. [`cor-baire-nested`]
Let X be a space, and suppose that for every nested chain of dense open sets

$$V_0 \supseteq V_1 \supseteq V_2 \supseteq V_3 \supseteq \cdots$$

the intersection $V_\infty = \bigcap_n V_n$ is dense. Then X is Baire.

Proof. Let $(U_n)_{n\in\mathbb{N}}$ be a countable family of dense open sets. Put $V_n = U_0 \cap \dots \cap U_n$. These sets are open and dense by Lemma 15.5, and they are nested as above, so the intersection V_∞ is dense. However, it is clear that V_∞ is just the same as $\bigcap_n U_n$. $\square$

Proof of Theorem 15.3. Let X be a complete metric space, let $(V_n)_{n\in\mathbb{N}}$ be a nested chain of dense open sets, and let W be an arbitrary nonempty open set. We must show that $V_\infty \cap W \neq \emptyset$.

We first choose $x_0 \in V_0 \cap W$ (which is possible because V_0 is dense and W is nonempty and open). As $V_0 \cap W$ is open, we can now choose $\epsilon_0 < 1$ such that $OB_{2\epsilon_0}(x_0) \subseteq V_0 \cap W$. We put $A_0 = OB_{\epsilon_0}(x_0)$.

Next, we note that A_0 is nonempty and open, so it must meet the dense open set V_1. We choose $x_1 \in V_1 \cap A_0$, then we choose ϵ_1 with $0 < \epsilon_1 < \epsilon_0/2$ and $OB_{2\epsilon_1}(x_1) \subseteq V_1 \cap A_0$, then we put $A_1 = OB_{\epsilon_1}(x_1)$. Continuing in the same way, we choose points x_k, numbers ϵ_k and open balls A_k such that

(a) $B_{2\epsilon_k}(x_k) \subseteq V_k \cap A_{k-1}$
(b) $0 < \epsilon_k < \epsilon_{k-1}/2$
(c) $A_k = OB_{\epsilon_k}(x_k)$.

Suppose we have $j, k \geq n$ for some n. We then find that $x_j, x_k \in A_n$ and so $d(x_j, x_k) < \epsilon_n < 2^{-n}$. It follows that the sequence $(x_n)_{n\in\mathbb{N}}$ is Cauchy, but X is assumed to be complete, so the sequence converges to some point $x_\infty \in X$. As $d(x_n, x_i) < \epsilon_n$ for all $i \geq n$, we see that $d(x_n, x_\infty) \leq \epsilon_n < 2\epsilon_n$, but $B_{2\epsilon_n}(x_n) \subseteq V_n$ by (a), so $x_\infty \in V_n$. This holds for all n, so $x \in V_\infty$. Similarly, we have $d(x_0, x_i) < \epsilon_0$ for all i so $d(x_0, x_\infty) \leq \epsilon_0 < 2\epsilon_0$, but $B_{2\epsilon_0}(x_0) \subseteq W$ so $x_\infty \in W$. It follows that $V_\infty \cap W \neq \emptyset$, as claimed. $\square$

We now present a simple application, which should be compared with Proposition 8.19 (which says that $\mathbb{R} \not\simeq \mathbb{R}^2$ and $[0,1] \not\simeq [0,1]^2$) and Proposition 11.5 (which shows that Peano's construction gives a continuous surjection $[0,1] \to [0,1]^2$).

Proposition 15.7. [`prop-R-R-two-baire`]
For any continuous injective map $f\colon \mathbb{R} \to \mathbb{R}^2$, the image has empty interior. In particular, f cannot be surjective.

Proof. Put $F_n = f([-n,n])$. This gives a countable family of sets that are compact and therefore closed, and their union is $f(\mathbb{R})$. We will show shortly that F_n has empty interior. Because $\mathbb{R}^2$ is complete, we can use Theorem 15.3 to deduce that $f(\mathbb{R})$ has empty interior, as required.

To check that $\mathrm{int}(F_n) = \emptyset$, we first note that $f\colon [-n,n] \to F_n$ is a continuous bijection from a compact space to a Hausdorff space, so it is a homeomorphism. It follows there are precisely two points in F_n whose removal fails to disconnect the space (because the same property clearly holds for $[-n,n]$). However, if F_n has any interior points, then it contains an open disc, so it has infinitely many interior points, and Exercise 7.2 tells us that we can remove any one of them without disconnecting the space. We must therefore have $\mathrm{int}(F_n) = \emptyset$ as claimed. □

There are also important applications in functional analysis. The most basic one is as follows:

Theorem 15.8 (The Open Mapping Theorem). [`thm-banach-open`]
Let X and Y be Banach spaces (as in Definition 12.21), and let $f\colon X \to Y$ be a surjective continuous linear map. Then f is an open map.

The proof will be given after a lemma.

Lemma 15.9. [`lem-convex-closure`]
Let C be a convex subset of Y such that $-C = C$. Then the closure $\overline{C}$ is again convex, and satisfies $-\overline{C} = \overline{C}$. Moreover, if $\overline{C}$ has any interior points, then zero is an interior point.

Proof. For any $t \in [0,1]$ we can define a continuous map $m_t\colon Y^2 \to Y$ by $m_t(y,y') = (1-t)y + ty'$. As C is convex, we have $C \times C \subseteq m_t^{-1}(C) \subseteq m_t^{-1}(\overline{C})$. Moreover, the set $m_t^{-1}(\overline{C})$ is closed, so

$$\overline{C} \times \overline{C} = \overline{C \times C} \subseteq m_t^{-1}(\overline{C}).$$

As this holds for all $t \in [0,1]$ we deduce that $\overline{C}$ is convex. Next, as multiplication by -1 is a self-homeomorphism of Y, we have $-\overline{C} = \overline{-C} = \overline{C}$.

Now suppose that y is an interior point of $\overline{C}$, so there exists $\epsilon > 0$ such that $y + v \in \overline{C}$ whenever $\|v\| < \epsilon$. As $\|-v\| = \|v\| < \epsilon$ we also have $y - v \in \overline{C}$, but $\overline{C} = -\overline{C}$ so $v - y \in \overline{C}$ as well. It follows by convexity that $(v+y)/2 + (v-y)/2 \in \overline{C}$, or in other words $v \in \overline{C}$ whenever $\|v\| < \epsilon$. This proves that 0 is an interior point of $\overline{C}$. □

Proof of Theorem 15.8. Put $U_n = \{x \in X \;:\; \|x\| < n\}$ for all $n > 0$. These sets cover X and f is surjective, so $Y = \bigcup_n f(U_n)$. In particular, this union has nonempty interior, so by the Baire theorem, there must exist n such that the set $\overline{f(U_n)}$ has nonempty interior. As f is linear we see that $f(U_n)$ is convex and satisfies $-f(U_n) = f(U_n)$, so we see from the lemma that 0 must be in the interior of $\overline{f(U_n)}$. As multiplication by n is a self-homeomorphism of Y we deduce that 0 is also in the interior of $\overline{f(U_1)}$. In other words, there exists $\epsilon > 0$ such that whenever $\|y\| < \epsilon$ we have $y \in \overline{f(U_1)}$. This means in particular that we can choose $x \in X$ with $\|x\| < 1$ and $\|y - f(x)\| < \epsilon/2$. By rescaling we obtain the following statement: if $y \in Y$ with $\|y\| < t\epsilon$ then there exists $x \in X$ with $\|x\| < t$ and $\|y - f(x)\| < t\epsilon/2$.

Now revert to the case $t = 1$, and suppose that we have $y_0 \in Y$ with $\|y_0\| < \epsilon$. We can then choose $x_0 \in X$ with $\|x_0\| < 1$ and such that the vector $y_1 = y_0 - f(x_0)$ has $\|y_1\| < \epsilon/2$. We can then use the case $t = 1/2$ to see that there exists $x_1 \in X$ with $\|x_1\| < 1/2$ such that the vector $y_2 = y_1 - f(x_1) = y_0 - f(x_0 + x_1)$ has $\|y_2\| < \epsilon/4$. Continuing in this way, we obtain a sequence $(x_n)_{n\in\mathbb{N}}$ with $\|x_i\| < 2^{-i}$ such that the partial sums $w_n = \sum_{j<n} x_j$ satisfy $\|y_0 - f(w_n)\| < \epsilon/2^n$. In particular, we find that $f(w_n) \to y_0$. Moreover, as $\|x_i\| < 2^{-i}$ we see that the sequence $(w_n)_{n\in\mathbb{N}}$ is Cauchy. We assumed that X is a Banach space, so it is complete, so the sequence converges to some $w_\infty \in X$. As f is continuous, we have $f(w_\infty) = y_0$. Moreover, as $\|x_i\| < 2^{-i}$ for all i we have $\|w_\infty\| \leq \sum_i \|x_i\| < 2$. This proves that the open ball of radius ϵ in Y is contained in $f(U_2)$.

Now consider an arbitrary open set $A \subseteq X$, and a point $b \in f(A)$. We can then choose $a \in A$ with $f(a) = b$, and as A is open, there exists $r > 0$ such that $OB_r(a) \subseteq A$. It follows that $OB_{r\epsilon/2}(0) \subseteq f(OB_r(0))$ and so $OB_{r\epsilon/2}(b) \subseteq f(OB_r(a)) \subseteq f(A)$. This proves that $f(A)$ is open, as claimed. □

Corollary 15.10. [`cor-banach-iso`]
Let X and Y be Banach spaces, and let $f\colon X \to Y$ be a linear map that is continuous and bijective. Then $f^{-1}\colon Y \to X$ is also linear, continuous and bijective.

Proof. It is elementary that f^{-1} is linear and bijective. We also know from Theorem 15.8 that f is open, which is equivalent to f^{-1} being continuous. □

Corollary 15.11. [`cor-banach-closed-graph`]
Let $f\colon X \to Y$ be a linear map between Banach spaces, and suppose that the graph

$$G = \{(x, f(x)) \ : \ x \in X\} \subseteq X \times Y$$

is a closed subspace of $X \times Y$. Then f is continuous.

Proof. It is easy to see that the rule $\|(x, y)\| = max(\|x\|, \|y\|)$ defines a norm on $X \times Y$, which gives rise to the product metric and the product topology. As the product of two complete metric spaces is again complete, we see that this makes $X \times Y$ into a Banach space. As f is linear, we see that G is a vector subspace of $X \times Y$, so it inherits a norm. As G is closed, it is complete under the resulting metric, so it is again a Banach space. We have projection maps $X \xleftarrow{p} G \xrightarrow{q} Y$, which are continuous and linear. One can check directly that p is a bijection, with $f = qp^{-1}$. Corollary 15.10 tells us that p^{-1} is continuous, so f is also continuous. □

Another interesting class of applications of Baire's theorem is to prove that in some sense, almost all continuous functions are very wild. As an example, we have the following theorem of Banach and Mazurkiewicz.

Theorem 15.12. [`thm-mazurkiewicz`]
Put

$$X = \{f \in C([0, 1]) \ : \ f \textit{ is differentiable at some point } x \in [0, 1)\}.$$

Then X is nowhere dense in $C([0, 1])$.

Sketch proof. Let us say that f is n-tame at x if for all y with $x \leq y \leq 1$ we have $|f(y) - f(x)| \leq n|y - x|$. We claim that if f is differentiable at x, then it is n-tame at x for sufficiently large n. Indeed, differentiability will give an inequality $|f(y) - f(x)| \leq n|y - x|$ provided that n is large and $|y - x|$ is small, say $x - \epsilon < y < x + \epsilon$. Moreover, the function $y \mapsto |f(y) - f(x)|/|y - x|$ is continuous and therefore bounded on the compact interval $[x + \epsilon, 1]$, with upper bound m say. It then follows that f is $\max(n, m)$-tame at x, as required.

Now put

$$T_n = \{f \in C([0, 1]) \ : \ f \text{ is } n\text{-tame at some point } x \in [0, 1 - 1/n] \ \},$$

and $T_\infty = \bigcup_{n>0} T_n$. If $f \in X$ then we have seen that f is n-tame at x for some $x \in [0, 1)$ and n, and after increasing n if necessary we may assume that $x \in [0, 1 - 1/n]$, so $f \in T_\infty$. It will thus suffice to show that T_∞ is nowhere dense. This will follow from Baire's theorem if we can show that each set T_n is closed and nowhere dense.

For this, we first put

$$A_n = \{(x, y) \ : \ 0 \leq x \leq 1 - 1/n, \ x \leq y \leq 1\}.$$

We then define $\alpha_n(f)\colon A_n \to \mathbb{R}$ and $\beta_n(f)\colon [0, 1/n] \to \mathbb{R}$ and $\gamma_n(f) \in \mathbb{R}$ by

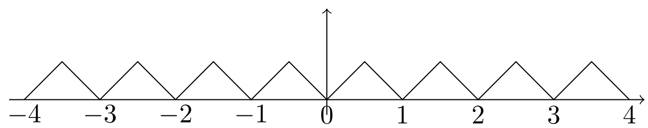
$$\begin{aligned} \alpha_n(f)(x, y) &= \max(0, |f(y) - f(x)| - n|y - x|) \\ \beta_n(f)(x) &= \max\{\alpha_n(x, y) \ : \ x \leq y \leq 1\} \\ \gamma_n(f) &= \min\{\beta_n(f)(x) \ : \ 0 \leq x \leq 1 - 1/n\}. \end{aligned}$$

It is then straightforward to check that

$$|\gamma_n(f) - \gamma_n(g)| \leq \|\beta_n(f) - \beta_n(g)\| \leq \|\alpha_n(f) - \alpha_n(g)\| \leq 2\|f - g\|.$$

Thus, the map $\gamma_n\colon C([0,1]) \to \mathbb{R}$ is continuous. Note also that $\gamma_n(f) = 0$ iff $\beta_n(f)(x) = 0$ for some $x \in [0, 1 - 1/n]$, or equivalently $\alpha_n(f)(x, y) = 0$ for all $y \in [x, 1]$, which means that f is n-tame at x. It follows that $T_n = \gamma_n^{-1}\{0\}$, so this set is closed.

Now let U_n be the interior of T_n. We must show that U_n is empty. If not, we recall from Proposition 12.51 that the set of piecewise-linear functions is dense in $C([0,1])$, so we can choose such a function g lying in U_n. As U_n is open, it will then contain $OB_{2\epsilon}(g)$ for some $\epsilon > 0$. Now define a "sawtooth" function $v\colon \mathbb{R} \to \mathbb{R}$ by $v(x) = \min\{|x - n| \;:\; n \in \mathbb{Z}\}$.



Let s be the maximum absolute value of the line segments in the graph of g, let m be large compared with ns/ϵ, and put $h(x) = g(x) + \epsilon v(mx)$. If $x < y \leq 1$ with $y - x$ small then v will contribute $\pm\epsilon m|y - x|$ to $h(y) - h(x)$, and this will dominate the contribution from g. Using this we see that h is not n-tame at x for any x, so $h \notin T_n$. However, we also have $d(g, h) = \epsilon < 2\epsilon$, so $h \in OB_{2\epsilon}(g) \subseteq U_n$, which is a contradiction. □

Proposition 15.13. [prop-compact-baire]
Every compact Hausdorff space is Baire.

Proof. Let X be a compact Hausdorff space, let $(V_n)_{n\in\mathbb{N}}$ be a nested chain of dense open sets, and let W be an arbitrary nonempty open set. We must show that $V_\infty \cap W \neq \emptyset$.

We first choose $x_0 \in V_0 \cap W$ (which is possible because V_0 is dense and W is nonempty and open). As $V_0 \cap W$ is open and X is regular we can choose an open neighbourhood A_0 of x_0 such that $x_0 \in A_0 \subseteq \overline{A_0} \subseteq V_0 \cap W$.

Next, we note that A_0 is nonempty and open, so it must meet the dense open set V_1. We choose $x_1 \in V_1 \cap A_0$, then we choose an open neighbourhood A_1 of x_1 such that $x_1 \in A_1 \subseteq \overline{A_1} \subseteq V_1 \cap A_0$. Continuing in the same way, we choose points x_k and open sets A_k such that $x_k \in A_k \subseteq \overline{A_k} \subseteq V_k \cap A_{k-1}$ for all k. As the sets $\overline{A_k}$ are nonempty and $\overline{A_k} \subseteq \overline{A_{k-1}}$ we see that the family $(\overline{A_k})_{k\in\mathbb{N}}$ has the finite intersection property. It follows by Proposition 10.12 that the intersection $\bigcap_k \overline{A_k}$ is nonempty. Choose $x_\infty \in \bigcap_k A_k$. For all k we have $x_\infty \in \overline{A_k} \subseteq V_k$, so $x_\infty \in V_\infty$. We also have $x_\infty \in \overline{A_0} \subseteq W$, so $V_\infty \cap W$ is nonempty, as required. □

16. Differentiation

Normed spaces give the most natural context for the theory of differentiation. We will need some of these ideas later, so we give a brief treatment here. Most of the discussion below applies equally well over $\mathbb{R}$ or $\mathbb{C}$. If we need to mention the scalar field we will call it $\mathbb{K}$ (so $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$).

Definition 16.1. [defn-derivative]
Let X and Y be normed spaces, and let U be an open subset of X. Consider a function $f\colon U \to Y$, a point $a \in U$, and a continuous linear map $\lambda\colon X \to Y$. We say that λ is a *derivative* for f at a if for all $\epsilon > 0$ there exists $\delta > 0$ such that $\|f(a + x) - f(a) - \lambda(x)\| \leq \epsilon\|x\|$ whenever $\|x\| \leq \delta$. We say that f is *differentiable at a* if such a λ exists. We say that f is *differentiable* if it is differentiable at every point in U.

Lemma 16.2. [lem-derivative-unique]
Let X, Y, U, f and a be as above. Then there is at most one derivative for f at a. (Thus, it is legitimate to denote that derivative by $f'(a)$, if it exists.)

Proof. Let λ_0 and λ_1 be derivatives, and put $\mu = \lambda_0 - \lambda_1$. For any $\epsilon > 0$ we can choose $\delta(\epsilon) > 0$ such that when $\|x\| \leq \delta(\epsilon)$ we have

$$\|f(a + x) - f(a) - \lambda_i(x)\| \leq \epsilon\|x\|/2$$

for $i = 0, 1$. It follows that for such x we have

$$\begin{aligned}\|\mu(x)\| &= \|(f(a+x) - f(a) - \lambda_1(x)) - (f(a+x) - f(a) - \lambda_0(x))\| \\ &\leq \|f(a+x) - f(a) - \lambda_1(x)\| + \|f(a+x) - f(a) - \lambda_0(x)\| \\ &\leq \epsilon\|x\|/2 + \epsilon\|x\|/2 = \epsilon\|x\|.\end{aligned}$$

Now let $x \in X$ be arbitrary. For sufficiently small $t > 0$ we will have $\|tx\| \leq \delta(\epsilon)$ so $\|\mu(tx)\| \leq \epsilon\|tx\|$ and we can divide by t to deduce that $\|\mu(x)\| \leq \epsilon\|x\|$. As this holds for all $\epsilon > 0$, we must have $\mu(x) = 0$ as claimed. □

PROPOSITION 16.3. [`prop-diff-cts`]
Any differentiable map is continuous.

PROOF. Let $f\colon U \to Y$ be differentiable. Suppose we are given $a \in X$ and $\epsilon > 0$. Put $\lambda = f'(a)$. We can take $\epsilon = 1$ in Definition 16.1 to see that there exists $\delta_0 > 0$ such that $\|f(a+x) - f(a) - \lambda(x)\| \leq \|x\|$ whenever $\|x\| \leq \delta_0$. For such x we therefore have

$$\|f(a+x) - f(a)\| \leq \|f(a+x) - f(a) - \lambda(x)\| + \|\lambda(x)\| \leq \|x\| + \|\lambda\|\|x\| = (1 + \|\lambda\|)\|x\|.$$

Now put $\delta = \min(\delta_0, \epsilon/(1 + \|\lambda\|))$. We find that for $\|x\| < \delta$ we have $\|f(a+x) - f(a)\| \leq (1 + \|\lambda\|)\|x\| < \epsilon$. By the usual metric criterion (Proposition 3.9) this proves that f is continuous. □

DEFINITION 16.4. [`defn-C-one`]
We say that $f\colon U \to Y$ is *continuously differentiable* if it is differentiable, and the map $f'\colon U \to \mathrm{Hom}^c(X, Y)$ is continuous.

REMARK 16.5. [`rem-affine-differentiable`]
Suppose that $f\colon X \to Y$ has the form $f(x) = c + \lambda(x)$ for some $c \in Y$ and some $\lambda \in \mathrm{Hom}^c(X, Y)$. It is then clear that f is continuously differentiable with $f'(a) = \lambda$ for all a.

PROPOSITION 16.6 (The Chain Rule). [`prop-chain-rule`]
Let X, Y and Z be normed spaces, and let $U \subseteq X$ and $V \subseteq Y$ be open sets. Suppose we have maps $U \xrightarrow{f} V \xrightarrow{g} Z$. Suppose we also have points $a \in U$ and $b \in V$ with $f(a) = b$, and that f is differentiable at a and that g is differentiable at b. Then $g \circ f$ is differentiable at a, with

$$(g \circ f)'(a) = g'(b) \circ f'(a)\colon X \to Z.$$

PROOF. First assume for simplicity that $a = 0$ and $b = 0$, and put $\lambda = f'(0)$ and $\mu = g'(0)$, so the claim is that $\mu \circ \lambda$ is a derivative for $g \circ f$ at 0. Suppose we are given $\epsilon > 0$. As μ is a derivative for g at 0, we can find $\delta > 0$ such that $\|g(y) - \mu(y)\| \leq \frac{\epsilon}{2(\|\lambda\|+1)}\|y\|$ whenever $\|y\| \leq \delta$. Next, we can choose $\eta > 0$ such that $\|f(x) - \lambda(x)\| \leq \min(1, \frac{\epsilon}{2(\|\mu\|+1)})\|x\|$ whenever $\|x\| \leq \eta$. After reducing η if necessary, we can assume that $\eta \leq \delta/(1 + \|\lambda\|)$. Now suppose that $\|x\| \leq \eta$. Note that $\|f(x) - \lambda(x)\| \leq \|x\|$ and so

$$\|f(x)\| \leq \|f(x) - \lambda(x)\| + \|\lambda(x)\| \leq (1 + \|\lambda\|)\|x\| \leq (1 + \|\lambda\|)\eta \leq \delta.$$

It follows that

$$\|g(f(x)) - \mu(f(x))\| \leq \frac{\epsilon}{2(\|\lambda\| + 1)}\|f(x)\| \leq \frac{\epsilon}{2}\|x\|.$$

On the other hand, we have

$$\|\mu(f(x)) - \mu(\lambda(x))\| \leq \|\mu\|\|f(x) - \lambda(x)\| \leq \|\mu\|\frac{\epsilon}{2(\|\mu\| + 1)}\|x\| \leq \frac{\epsilon}{2}\|x\|.$$

By combining these, we have

$$\|g(f(x)) - \mu(\lambda(x))\| \leq \|g(f(x)) - \mu(f(x))\| + \|\mu(f(x)) - \mu(\lambda(x))\| \leq \epsilon\|x\|$$

as required. The more general case (where a and b may be nonzero) can be reduced to the special case by considering the maps $f_0(x) = f(x + a) - b$ and $g_0(y) = g(y + b)$. □

COROLLARY 16.7. [`cor-chain-rule`]
For f and g as above, if f and g are both continuously differentiable then so is $g \circ f$.

Proof. Put $h = g \circ f$. The proposition tells us that h is differentiable, and that h' is the composite

$$U \xrightarrow{(f,1)} V \times U \xrightarrow{g' \times f'} \mathrm{Hom}^c(Y,Z) \times \mathrm{Hom}^c(X,Y) \xrightarrow{\text{compose}} \mathrm{Hom}^c(X,Z).$$

The map $(f,1)$ is continuous by Proposition 16.3, and $f' \times g'$ is continuous by assumption. Using the estimate $\|\mu\lambda\| \leq \|\mu\|\|\lambda\|$ it is not hard to see that the composition map is continuous, and it follows that h' is also continuous. $\square$

Proposition 16.8. [prop-diff-lipschitz]
Suppose that $U \subseteq X$ is convex, and that $f\colon U \to Y$ is differentiable with $\|f'(a)\| \leq K$ for all $a \in U$. Then $\|f(b) - f(a)\| \leq K\|b - a\|$ for all $a, b \in U$ (so f is Lipschitz).

Proof. Fix $a, b \in U$ and put $L = K\|b - a\|$. As U is open and convex there exist numbers $\alpha < 0$ and $\beta > 1$ such that $(1-t)a + tb \in U$ for $\alpha < t < \beta$. Define $g\colon (\alpha, \beta) \to Y$ by $g(t) = f(a + t(b-a)) - f(a)$. We find using the chain rule that g is differentiable. If we identify $\mathrm{Hom}^c(\mathbb{R}, Y)$ with Y in the obvious way, the derivative is $g'(t) = f'(a + t(b-a))(b-a)$ and thus $\|g'(t)\| \leq L$. The claim is that $\|g(1)\| \leq L$. To prove this, fix $\epsilon > 0$ and put

$$P = \{p \in [0,1] \ : \ \|g(t)\| \leq (L+\epsilon)t \text{ for all } t \in [0,p]\}.$$

It is easy to see that this is closed and contains 0. It follows that $P = [0,p]$ for some $p \in [0,1]$, so $\|g(p)\| \leq (L+\epsilon)p$. We claim that $p = 1$. To see this, note that there exists $\delta > 0$ such that when $p + x \in (\alpha, \beta)$ with $|x| \leq \delta$ we have $\|g(p+x) - g(p) - g'(p)x\| \leq \epsilon|x|$. Put $q = \min(p+\delta, 1)$. For $0 \leq x \leq q - p$ we then have

$$\|g(p+x)\| \leq \|g(p)\| + \|g'(p)x\| + \epsilon x \leq (L+\epsilon)p + Lx + \epsilon x = (L+\epsilon)(p+x).$$

This means that $q \in P$, which would contradict the definition of p unless $p = q = 1$. $\square$

We now want to consider iterated derivatives.

Definition 16.9. [defn-smooth]
Suppose we have normed vector spaces X and Y, an open subset $U \subseteq X$, and a map $f\colon U \to Y$. We can define normed spaces T_k recursively by $T_0 = Y$ and $T_{k+1} = \mathrm{Hom}^c(X, T_k)$ for all $k \geq 0$. We write $f^{(0)} = f\colon U \to T_0 = Y$. If f is differentiable we have a map $f'\colon U \to \mathrm{Hom}^c(X,Y) = T_1$, which we also denote by $f^{(1)}$. If $f^{(1)}$ is differentiable then the derivative is a map $U \to \mathrm{Hom}^c(X, \mathrm{Hom}^c(X,Y)) = T_2$, which we denote by $f^{(2)}$. More generally, if $f^{(k)}\colon U \to T_k$ is defined and is differentiable, we write $f^{(k+1)}\colon U \to T_{k+1}$ for the derivative. We say that f is *smooth* if $f^{(k)}$ is defined for all k.

It is somewhat awkward to work with the spaces T_k as defined above. The theory can instead be formulated in terms of multilinear maps, as we now explain.

Definition 16.10. [defn-multilinear]
Let $V_0, \dots, V_{r-1}$ and W be vector spaces, and put $V = \prod_t V_t$. We say that a map $f\colon V \to W$ is *multilinear* if it is linear in each variable separately. More precisely, suppose we have $p \in \{0, \dots, r-1\}$ and $a \in V$. We can then define $i_{p,a}\colon V_p \to \prod_t V_t$ by

$$i_{p,a}(x)_t = \begin{cases} x & \text{if } t = p \\ a_t & \text{if } t \neq p. \end{cases}$$

Then f is multilinear if and only if $f \circ i_{p,a}\colon V_p \to W$ is linear for all p and a. In the case $r = 2$ we say *bilinear* rather than multilinear. We write $\mathrm{Hom}_{(r)}(V, W)$ or $\mathrm{Hom}_{(r)}(V_0, \dots, V_{r-1}; W)$ for the set of all multilinear maps. This is itself a vector space in an evident way.

Example 16.11. [eg-multilinear]

(a) The standard inner product gives a bilinear map $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$. The standard Hermitian product does not give a bilinear map $\mathbb{C}^n \times \mathbb{C}^n \to \mathbb{C}$, because $\langle u, tv\rangle$ is $\overline{t}\langle u, v\rangle$ rather than $t\langle u, v\rangle$.
(b) The usual cross product of three-dimensional vectors gives a bilinear map $\mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^3$.
(c) Multiplication of matrices gives a bilinear map $M_{pq}(\mathbb{R}) \times M_{qr}(\mathbb{R}) \to M_{pr}(\mathbb{R})$.

(d) For any $n \geq 0$, we can define a multilinear map $\delta\colon (\mathbb{R}^n)^n \to \mathbb{R}$ as follows. Given vectors $v_1, \dots, v_n \in \mathbb{R}^n$, we let A denote the square matrix whose columns are $v_1, \dots, v_n$, then we put $\delta(v_1, \dots, v_n) = \det(A)$.

Definition 16.12. [`defn-multilinear-norm`]
Now suppose that $V_0, \dots, V_{r-1}$ and W have specified norms. We define a norm on the space $V = \prod_t V_t$ by

$$\|x\| = \max(\|x_0\|, \dots, \|x_{r-1}\|),$$

and recall that the corresponding topology is the same as the product topology. We also define $\nu\colon V \to \mathbb{R}_+$ by $\nu(x) = \prod_t \|x_t\|$.

Now consider a multilinear map $f\colon V \to W$. We define $\|f\| \in [0, \infty]$ by

$$\|f\| = \sup\{\|f(x)\| \;:\; x \in V, \|x\| \leq 1\}.$$

We say that f is *bounded* if $\|f\| < \infty$, and we write $\operatorname{Hom}^c_{(r)}(V, W)$ for the set of bounded multilinear maps.

Remark 16.13. [`rem-multilinear-norm`]
Multilinearity implies that

$$f(t_0x_0, \dots, t_{r-1}x_{r-1}) = \left(\prod_i t_i\right) f(x)$$

for all $x \in V$ and $t \in \mathbb{K}^r$. Using this we see that for all $x \in V$ we have $\|f(x)\| \leq \|f\|\nu(x)$.

Proposition 16.14. [`prop-multilinear-norm`]
The set $\operatorname{Hom}^c_{(r)}(V, W)$ is a vector subspace of $\operatorname{Hom}_{(r)}(V, W)$, and the map $f \mapsto \|f\|$ is a norm on $\operatorname{Hom}^c_{(r)}(V, W)$. Moreover, if W is a Banach space, then so is $\operatorname{Hom}^c_{(r)}(V, W)$.

Proof. Suppose that $f \in \operatorname{Hom}^c_{(r)}(V, W)$ and $t \in \mathbb{K}$. It is then clear from the definitions that $tf \in \operatorname{Hom}^c_{(r)}(V, W)$ with $\|tf\| = |t|\|f\|$. Now suppose we have another element $g \in \operatorname{Hom}^c_{(r)}(V, W)$. When $\|x\| \leq 1$ we then have

$$\|(f + g)(x)\| = \|f(x) + g(x)\| \leq \|f(x)\| + \|g(x)\| \leq \|f\| + \|g\|,$$

so $\|f + g\| \leq \|f\| + \|g\| < \infty$, so $f + g \in \operatorname{Hom}^c_{(r)}(V, W)$. It is clear from this that $\operatorname{Hom}^c_{(r)}(V, W)$ is a vector space, and that the map $f \mapsto \|f\|$ is at least a seminorm. It is also clear from the inequality $\|f(x)\| \leq \|f\|\nu(x)$ that ($\|f\| = 0$ iff $f = 0$), so we actually have a norm. All that is left is to prove that $\operatorname{Hom}^c_{(r)}(V, W)$ is complete if W is complete. The case $r = 1$ was proved as Proposition 12.23, and the argument for the general case is essentially the same. □

Proposition 16.15. [`prop-multilinear-curry`]
Suppose we have normed vector spaces $U_0, \dots, U_{q-1}, V_0, \dots, V_{r-1}$ and W. Then there is a linear isomorphism

$$\phi\colon \operatorname{Hom}_{(q)}(U, \operatorname{Hom}_{(r)}(V, W)) \to \operatorname{Hom}_{(q+r)}(U \times V, W)$$

given by

$$\phi(f)(u, v) = f(u)(v).$$

Moreover, this restricts to give an isometric linear isomorphism

$$\phi\colon \operatorname{Hom}^c_{(q)}(U, \operatorname{Hom}^c_{(r)}(V, W)) \to \operatorname{Hom}^c_{(q+r)}(U \times V, W)$$

Proof. Let $D(V, W)$ denote the space of all functions $V \to W$ and so on. Then the above prescription certainly defines a bijection $D(U \times V, W) \to D(U, D(V, W))$. We also see that $\phi(f) \in D(U, \operatorname{Hom}_{(r)}(V, W))$ iff $f(u, v)$ is linear in each of the variables v_k, and then that $\phi(f) \in \operatorname{Hom}_{(q)}(U, \operatorname{Hom}_{(r)}(V, W))$ iff f is also linear in each of the variables u_j, or equivalently $f \in \operatorname{Hom}_{(p+q)}(U \times V, W)$. We have thus defined a linear isomorphism

$$\phi\colon \operatorname{Hom}_{(q)}(U, \operatorname{Hom}_{(r)}(V, W)) \to \operatorname{Hom}_{(q+r)}(U \times V, W).$$

We also have

$$\|\phi(f)\| = \sup\{\|f(u)(v)\| \;:\; \|(u, v)\| \leq 1\}.$$

Here $\|(u,v)\| \leq 1$ iff $\|u\| \leq 1$ and $\|v\| \leq 1$ and by definition we have $\|f(u)\| = \sup\{\|f(u)(v)\| : \|v\| \leq 1\}$ so the above can be rewritten as

$$\|\phi(f)\| = \sup\{\|f(u)\| : \|u\| \leq 1\} = \|f\|.$$

It follows that ϕ restricts to give an isometric isomorphism

$$\phi\colon \operatorname{Hom}^c_{(q)}(U, \operatorname{Hom}^c_{(r)}(V,W)) \to \operatorname{Hom}^c_{(q+r)}(U \times V, W).$$

as claimed. □

Corollary 16.16. [`cor-multilinear-finite`]
If all the spaces V_t are finite-dimensional, then $\operatorname{Hom}_{(r)}(V,W) = \operatorname{Hom}^c_{(r)}(V,W)$.

Proof. This holds for $r = 1$ by Corollary 10.37, and the general case follows by induction on r using Proposition 16.15. □

Remark 16.17. [`rem-higher-derivatives`]
Suppose we have normed vector spaces X and Y. Definition 16.9 involves the spaces T_r defined by $T_0 = Y$ and $T_{k+1} = \operatorname{Hom}^c(X, T_k)$. Using Proposition 16.15 we see that T_k can be identified with $\operatorname{Hom}^c_{(k)}(X^k, Y)$.

Proposition 16.18. [`prop-derivatives-commute`]
Let $f\colon U \to Y$ be as in Definition 16.9, and suppose that $f^{(r)}\colon U \to \operatorname{Hom}^c_{(r)}(X^r, Y)$ is defined and continuous. Then for all $a \in U$, the multilinear map $f^{(r)}(a)\colon X^r \to Y$ is invariant under permutation of the arguments.

Before proving this in general, we will treat a special case.

Lemma 16.19. *Suppose we have $\alpha > 0$ and a continuous map $g\colon (-\alpha,\alpha)^2 \to Y$ such that*

(a) *$g(s,0) = g(0,t) = 0$ for all $s,t \in (-\alpha,\alpha)$.*
(b) *$g^{(2)}$ is defined and continuous and satisfies $g^{(2)}(0,0)(e_1,e_0) = 0$.*

Then for all $\epsilon > 0$ there exists $\beta > 0$ such that $\|g(s,t)\| \leq |s||t|\epsilon$ for all $(s,t) \in (-\beta,\beta)^2$.

Proof. As $g^{(2)}$ is continuous, we can choose $\beta > 0$ such that $\|g^{(2)}(s,t)(e_0,e_1)\| < \epsilon$ for all $(s,t) \in (-\beta,\beta)^2$. Now fix $t \in (-\beta,\beta)$ and define $p\colon (-\beta,\beta) \to Y$ by $p(s) = g'(s,t)(e_1)$. We can differentiate the relation $g(0,t) = 0$ with respect to t to see that $p(0) = 0$. On the other hand, the chain rule shows that p is differentiable with $p'(s) = g^{(2)}(s,t)(e_1,e_0)$, so $\|p'(s)\| < \epsilon$ for all $s \in (-\beta,\beta)$, so $\|p(s)\| = \|p(s) - p(0)\| \leq \epsilon|s|$ by Proposition 16.8. In other words, we have $\|g'(s,t)(e_1)\| \leq \epsilon|s|$ for all $(s,t) \in (-\beta,\beta)^2$.

Now fix $s \in (-\beta,\beta)$ and define $q\colon (-\beta,\beta) \to Y$ by $q(t) = g(s,t)$, so $q(0) = g(s,0) = 0$. The chain rule shows that this is differentiable, with $q'(t) = g'(s,t)(e_1)$, so $\|q'(t)\| \leq \epsilon|s|$. It follows that $\|g(s,t)\| = \|q(t) - q(0)\| \leq \epsilon|s||t|$ as claimed. □

Corollary 16.20. [`cor-derivatives-commute`]
Suppose we have $\alpha > 0$ and a map $f\colon (-\alpha,\alpha)^2 \to Y$ such that $f^{(2)}$ is defined and continuous. Then $f^{(2)}(0,0)(u,v) = f^{(2)}(0,0)(v,u)$ for all $u,v \in \mathbb{R}^2$.

Proof. Put $\mu = f^{(2)}(0,0)\colon (\mathbb{R}^2)^2 \to Y$. This is a bilinear map, so it must have the form

$$\mu((u_0,u_1),(v_0,v_1)) = \sum_{i=0}^{1}\sum_{j=0}^{1} u_i v_j y_{ij}$$

for some elements $y_{ij} = \mu(e_i,e_j) \in Y$. We must show that $\mu(u,v) = \mu(v,u)$ for all u and v, and this is easily equivalent to the claim that $y_{01} = y_{10}$. Define $g, h\colon (-\alpha,\alpha)^2 \to Y$ by

$$g(s,t) = f(s,t) - f(s,0) - f(0,t) + f(0,0) - sty_{01}$$
$$h(s,t) = f(s,t) - f(s,0) - f(0,t) + f(0,0) - sty_{10}.$$

It is clear that $g(s,0) = g(0,t) = h(s,0) = h(0,t) = 0$. We also have

$$\begin{aligned} g'(s,t)(e_0) &= f'(s,t)(e_0) - f'(s,0)(e_0) - ty_{01} \\ g(2)(s,t)(e_0,e_1) &= f^{(2)}(s,t)(e_0,e_1) - y_{01} \\ g(2)(0,0)(e_0,e_1) &= \mu(e_0,e_1) - y_{01} = 0. \end{aligned}$$

By the lemma, we deduce that for all $\epsilon > 0$ there exists $\beta > 0$ such that $\|g(s,t)\| \leq \epsilon|s||t|$ whenever $|s|,|t| \leq \beta$. By a symmetrical argument, there exists $\gamma > 0$ such that $\|h(s,t)\| \leq \epsilon|s||t|$ whenever $|s|,|t| \leq \gamma$. Now for $|s|,|t| \leq \min(\beta,\gamma)$ we have

$$|s||t|\|y_{10} - y_{01}\| = \|g(s,t) - h(s,t)\| \leq \|g(s,t)\| + \|h(s,t)\| \leq 2\epsilon|s||t|.$$

We can take $s = t = \min(\beta,\gamma) > 0$ and then divide by st to see that $\|y_{10} - y_{01}\| \leq 2\epsilon$. As this holds for all $\epsilon > 0$, we must have $y_{01} = y_{10}$ as claimed. □

Proof of Proposition 16.18. First consider the case $r = 2$. For any $a \in U$ and $u,v \in X$ we can choose $\alpha > 0$ such that $a + su + tv \in U$ whenever $|s|,|t| < \alpha$, then we can define $g\colon (-\alpha,\alpha)^2 \to Y$ by $g(s,t) = f(a + su + tv)$. Using the chain rule we see that $g^{(2)}$ is defined and continuous, with

$$\begin{aligned} g^{(2)}(0,0)(e_0,e_1) &= f^{(2)}(a)(u,v) \\ g^{(2)}(0,0)(e_1,e_0) &= f^{(2)}(a)(v,u). \end{aligned}$$

It therefore follows from Corollary 16.20 that $f^{(2)}(a)\colon X^2 \to Y$ is symmetric.

Now consider the case where $r > 2$. We must show that the function $f^{(r)}(a)(u_1,\dots,u_r)$ is invariant under the permutation group Σ_r. Inside Σ_r we have the subgroup H of all permutations that send r to itself, and also the subgroup K consisting of the identity together with the transposition that exchanges $r-1$ and r. We may assume by induction that $f^{(r-1)}$ is invariant under Σ_{r-1}, and by differentiating we conclude that $f^{(r)}$ is invariant under H. On the other hand, we can apply the $r = 2$ case to the function $f^{(r-2)}$ to see that $f^{(r)}$ is also invariant under K. It is standard and easy that H and K together generate Σ_r, so $f^{(r)}$ is invariant under Σ_r as claimed. □

Proposition 16.21. [`prop-multilinear-continuous`]
Let $f\colon V = \prod_{t=0}^{r-1} V_t \to W$ be multilinear. Then f is continuous if and only if it is bounded. (In particular, if the spaces V_t are all finite-dimensional, then f is continuous.)

Proof. First suppose that f is continuous. As $f(0) = 0$, there exists $\delta > 0$ such that $\|f(y)\| \leq 1$ whenever $\|y\| \leq \delta$. If $\|x\| \leq 1$ we can take $y = \delta x$ to deduce that $\|f(\delta x)\| \leq 1$, but multilinearity implies that $f(\delta x) = \delta^r f(x)$, so $\|f(x)\| \leq 1/\delta^r$. It follows that f is bounded with $\|f\| \leq 1/\delta^r$.

For the converse, suppose that f is bounded. Fix a point $a \in V$. For $P \subseteq \{0,\dots,r\}$ and $x \in V$ we define $j_P(x) \in V$ by

$$j_P(x)_t = \begin{cases} x_t & \text{if } t \in P \\ a_t & \text{if } t \notin P. \end{cases}$$

We also put $L = \|f\| \sum_{|P|>0} \prod_{t\notin P} \|a_t\|$. By expanding everything out we see that

$$f(a + x) - f(a) = \sum_{|P|>0} f(j_P(x)).$$

Now $\|j_P(x)\| \leq \|x\|^{|P|} \prod_{t\notin P} \|a_t\|$. If $\|x\| \leq 1$ we also have $\|x\|^{|P|} \leq \|x\|$ for all P with $|P| > 0$, and it follows easily that

$$\|f(a + x) - f(a)\| \leq L\|x\|$$

for all x with $\|x\| \leq 1$. It follows easily from this that f is continuous. □

This can be sharpened as follows.

Proposition 16.22. [prop-multilinear-continuous]
Let $f\colon V=\prod_{t=0}^{r-1}V_t\to W$ *be multilinear and bounded. Then* f *is smooth. The first derivative is given by*

$$f'(a)(x)=\sum_{p=0}^{r-1}f(i_{p,a}(x_p))$$

(where $i_{p,a}$ *is as in Definition 16.10).*

Proof. Fix $a\in V$, and put $g(x)=\sum_{p=0}^{r-1}f(i_{p,a}(x_p))$. To show that this is $f'(a)$, we need an upper bound for the norm of the error term $h(x)=f(a+x)-f(a)-g(x)$. In the notation in the proof of Proposition 16.22, we have $f(a)=\sum_{|P|=0}f(j_P(x))$ and $g(x)=\sum_{|P|=1}f(j_P(x))$ so $h(x)=\sum_{|P|>1}f(j_P(x))$. Now put $M=\|f\|\sum_{|P|>1}\prod_{t\not\in P}\|a_t\|$. Just as in the previous proof, if $\|x\|\leq 1$ we have $\|j_P(x)\|\leq\|x\|^2\prod_{t\not\in P}\|a_t\|$ whenever $|P|>1$, and thus $\|h(x)\|\leq M\|x\|^2$. Now given $\epsilon>0$ we can put $\delta=\min(1,\epsilon/M)$ and we find that $\|h(x)\|\leq\epsilon\|x\|$ whenever $\|x\|\leq\delta$, as required.

This proves that f is differentiable, with the claimed derivative. We still need to show that it is smooth. We may assume by induction that any continuous s-linear map with $s<r$ is smooth. In particular, for $0\leq p<r$ we can put $V'_p=\prod_{t\neq p}V_t$ and define $f_p\colon V'_p\to\mathrm{Hom}^c(V_p,W)$ by the obvious prescription

$$f_p(v)(x)=f(v_0,\dots,v_{p-1},x,v_{p+1},\dots,v_{r-1}).$$

As in Proposition 16.15 we see that f_p is $(r-1)$-linear and bounded, so it is smooth by the induction hypothesis. Now let $\pi_p\colon V\to V_p$ and $\pi'_p\colon V\to V'_p$ be the projections. The map π_p induces a map $\pi_p^*\colon\mathrm{Hom}^c(V_p,W)\to\mathrm{Hom}^c(V,W)$ by $\pi_p^*(u)=u\circ\pi_p$. This is easily seen to be bounded, with norm one. Now let k_p be the composite

$$V\xrightarrow{\pi'_p}V'_p\xrightarrow{f_p}\mathrm{Hom}(V_p,W)\xrightarrow{\pi_p^*}\mathrm{Hom}(V,W).$$

As f_p is smooth and the maps π'_p and π_p^* are linear and bounded, we see that k_p is also smooth. Moreover, our previous description of f' can be rewritten as $f'=\sum_p k_p$, so f' is smooth, so f is smooth. $\square$

17. Real valued functions

In this section, X is a topological space and $C(X)$ is the set of continuous real-valued functions $u\colon X\to\mathbb{R}$. We consider this as a metric space in the usual way.

17.1. The theorems of Urysohn and Tietze. The following theorem refers to normal spaces, which were introduced in Definition 14.1: a space is normal if any two disjoint closed sets have disjoint open neighbourhoods.

Theorem 17.1 (Urysohn's lemma). [thm-urysohn]
X *is normal iff for every pair* F_0,F_1 *of disjoint closed subsets there exists a continuous map* $f\colon X\to[0,1]$ *with* $f(F_0)\subseteq\{0\}$ *and* $f(F_1)\subseteq\{1\}$.

The proof will follow after some preliminaries. Note that one direction is trivial: if we have a function f as above, then the sets $U_0=\{x\;:\;f(x)<1/2\}$ and $U_1=\{x\;:\;f(x)>1/2\}$ are disjoint and open with $F_i\subseteq U_i$, as required for normality.

Definition 17.2. [defn-ufilt]
Write $A=[0,1]\cap\mathbb{Q}$. A *Urysohn filtration* on a space X is a family $U=(U_a)_{a\in A}$ of open subsets of X such that $\overline{U_a}\subseteq U_b$ whenever $a<b$. Such a filtration is *regular* if for all a we have $U_a=\bigcup_{b<a}U_b$. Note that this implies $U_0=\emptyset$.

Construction 17.3. Given a Urysohn filtration U define $f_U\colon X\to[0,1]$ by

$$f_U(x)=\inf\{a\in A\;:\;x\in U_a\}.$$

This is to be interpreted as 1 if $x\not\in U_a$ for any a.

Define also $\rho(U)_a=\bigcup_{b<a}U_b$, so U is regular if and only if $\rho(U)=U$. Finally, given $g\in C(X,[0,1])$ write

$$V(g)_a=\{x\;:\;g(x)<a\}=g^{-1}(-\infty,a).$$

THEOREM 17.4. [thm-ufilt]

(a) *$\rho(U)$ is a regular Urysohn filtration.*
(b) *$\rho^2(U) = U$.*
(c) *$V(g)$ is a regular Urysohn filtration.*
(d) *f_U is continuous.*
(e) *$V(f_U) = \rho(U)$*
(f) *$f_{V(g)} = g$*
(g) *There is a bijective correspondence between regular Urysohn filtrations on X and continuous functions $X \to [0,1]$, given by $U \mapsto f_U$ and $g \mapsto V(g)$.*

PROOF. In the following a, b and c are implicitly supposed to be elements of A, and x is supposed to be a point of X, and s to be an element of $[0,1]$. We shall repeatedly use without comment the fact that between any two distinct real numbers there lies a rational, and hence that $s = \inf\{a \,:\, s < a\}$ (where $\inf(\emptyset)$ is interpreted as 1).

(a) It is clear that $\rho(U)_a \subseteq U_a$. Suppose $a < b$, and write $c = (a+b)/2$ so $a < c < b$. As $\underline{U}$ is a Urysohn filtration, we have $\overline{U_a} \subseteq U_c$. Thus
$$\overline{\rho(U)_a} \subseteq \overline{U_a} \subseteq U_c \subseteq \rho(U)_b$$
This shows that $\rho(\underline{U})$ is a Urysohn filtration. Moreover,
$$\bigcup_{b<a} \rho(U)_b = \bigcup_{c<b<a} U_c = \bigcup_{c<a} U_c = \rho(U)_a$$
which shows that $\rho(U)$ is regular.
(b) This is immediate from the above.
(c) Suppose $g\colon X \to [0,1]$ is continuous and $a \in A$. Write
$$F(g)_a = \{x \,:\, g(x) \leq a\} = g^{-1}(-\infty, a]$$
This is a closed subset of X, and $V(g)_a \subseteq F(g)_a$ so $\overline{V(g)_a} \subseteq F_a(g)$. Moreover, if $a < b$ then $F(g)_a \subseteq V(g)_b$ and thus $\overline{V(g)_a} \subseteq V(g)_b$. Thus $V(g)$ is a Urysohn filtration.

Next note that if $x \in V(g)_a$ then $g(x) < a$ so there is a rational number $b \in A$ with $g(x) < b < a$ so $x \in V(g)_b$. Thus $V(g)_a = \bigcup_{b<a} V(g)_b$ and $V(g)$ is regular.
(d) To show that $f = f_U$ is continuous, we need only check that the preimages of the subbasic open sets $(-\infty, s)$ and (s, ∞) (for $s \in \mathbb{R}$) are open. For the first case, note that $f(x)$ is the greatest lower bound of the set $F = \{a \in A \,:\, x \in U_a\}$. Thus, we have $f(x) < s$ iff s is not a lower bound for that set, iff there exists $a \in F$ with $a < s$, iff $x \in \bigcup_{a<s} U_a$. We therefore see that $f^{-1}((-\infty, s)) = \bigcup_{a<s} U_a$, which is open as required. We next claim that $f^{-1}((s,\infty))$ is also open, or equivalently that $f^{-1}((-\infty, s])$ is closed. It will suffice to show that $f^{-1}((-\infty, s]) = \bigcap_{b>s} \overline{U_b}$. Suppose that $f(x) = \inf(F) \leq s$. Consider a point $b \in A$ with $b > s$. Then b is not a lower bound for F, so we can choose $a \in F$ with $a < b$. This means that $x \in U_a$, and by the Urysohn filtration condition we have $\overline{U_a} \subseteq U_b \subseteq \overline{U_b}$. It follows that $x \in \bigcap_{b>s} \overline{U_b}$ as required. Conversely, suppose that $x \in \bigcap_{b>s} \overline{U_b}$. Consider an element $c \in A$ with $c > s$. We can then find $b \in A$ with $c > b > s$. By assumption we have $x \in \overline{U_b}$, but $\overline{U_b} \subseteq U_c$ by the Urysohn filtration condition, so $x \in U_c$, so $c \in F$. Thus F contains $\{c \in A \,:\, c > s\}$, and it follows that $f(x) = \inf(F) \leq s$ as required.
(e) We have just showed that
$$V(f_U)_a = f_U^{-1}(-\infty, a) = \bigcup_{b<a} U_b = \rho(U)_a$$
as required.
(f)
$$f_{V(g)}(x) = \inf\{a \,:\, x \in V_a(g)\} = \inf\{a \,:\, g(x) < a\} = g(x).$$
(g) This is clear from the previous parts of the theorem.

□

Proof of Theorem 17.1. First suppose that X has the property in the statement. Then, given disjoint closed sets F_0 and F_1, we can choose f as above and put $U_0 = \{x : f(x) < 1/2\}$ and $U_1 = \{x : f(x) > 1/2\}$. This gives disjoint open sets U_i containing F_i, as required.

Conversely, suppose that X is normal, and let F_0 and F_1 be disjoint closed sets. As the set $A = \mathbb{Q} \cap [0,1]$ is countable, we can choose a bijection $n \mapsto a_n$ from $\mathbb{N}$ to A. We can arrange it such that $a_0 = 0$ and $a_1 = 1$. We shall choose recursively open sets $U_k \subseteq X$ such that

(a) $F_0 \subseteq U_k \subseteq X \setminus F_1$

(b) If $a_j < a_k$ then $\overline{U_j} \subseteq U_k$.

By normality, we can choose an open set U_0 with

$$F_0 \subseteq U_0 \subseteq \overline{U_0} \subseteq X \setminus F_1.$$

Similarly, we can choose U_1 with

$$\overline{U_0} \subseteq U_1 \subseteq \overline{U_1} \subseteq X \setminus F_1.$$

Suppose that $n > 0$ and $U_0, \dots U_n$ have been chosen. Some of the numbers $a_0, \dots, a_n$ (including $a_0 = 0$) will be less than a_{n+1}, and the rest (including $a_1 = 1$) will be greater than a_{n+1}. Let a_k be the largest of the values a_j of the first type, and let a_m be the smallest of the values a_j of the second type, so that $a_k < a_{n+1} < a_m$. By assumption we have $\overline{U_k} \subseteq U_m$ so by normality we can find U_{n+1} with

$$\overline{U_k} \subseteq U_{n+1} \subseteq \overline{U_{n+1}} \subseteq U_m$$

It is easy to check that this satisfies the requirements. Finally, we define a Urysohn filtration by

$$V_a = U_k \text{ where } a = a_k$$

and a continuous function $f = f_V$. It is easy to see that $f = 0$ on F_0 and $f = 1$ on F_1. □

Theorem 17.5 (The Tietze Extension Theorem). [thm-tietze]

Let X be a normal space, let Y be a closed subset, and let $f\colon Y \to \mathbb{R}$ be a continuous map. Then there exists a continuous function $g\colon X \to \mathbb{R}$ with $g|_Y = f$. Moreover, if $f(y) \in [a,b]$ for all y, then g can be chosen so that $g(x) \in [a,b]$ for all x.

The proof will follow after a preparatory lemma and a special case.

Lemma 17.6. [lem-tietze-approx]

Suppose X is normal and $Y \subseteq X$ is closed. Suppose $f\colon Y \to [-r,r]$ is continuous. Then there is a continuous function $g\colon X \to [-r/3, r/3]$ with

$$\|f - g\|_Y = \sup_{y \in Y} |f(y) - g(y)| \leq 2r/3.$$

Proof. Write

$$\begin{aligned} Y_- &= f^{-1}[-r, -r/3] \subseteq Y \\ Y_0 &= f^{-1}[-r/3, +r/3] \subseteq Y \\ Y_+ &= f^{-1}[+r/3, +r] \subseteq Y. \end{aligned}$$

These sets are clearly closed in Y, and Y is closed in X, so $Y_\pm$ are closed in X. As $Y_+ \cap Y_- = \emptyset$, Urysohn's theorem gives us a function $g\colon X \to [-r/3, r/3]$ with $g = -r/3$ on Y_- and $g = r/3$ on Y_+. By considering the cases $y \in Y_-$, $y \in Y_0$, and $y \in Y_+$ separately we see that $|f(y) - g(y)| \leq 2r/3$ for all $y \in Y$ and thus that $\|f - g\|_Y \leq 2r/3$. □

Lemma 17.7. [lem-tietze]

Let X be a normal space, let Y be a closed subset, and let $f\colon Y \to [-1,1]$ be a continuous map. Then there exists a continuous function $g\colon X \to [-1,1]$ with $g|_Y = f$.

Proof. We shall choose recursively continuous functions $g_k\colon X \to \mathbb{R}$ (for $k \geq 1$) with $\|g_k\| \leq (2/3)^k/2$ such that

$$\|f - \sum_{k=1}^{n} g_k\|_Y \leq (2/3)^n$$

Indeed, this works for $n = 0$ as $\|f\|_Y \leq 1$. Given g_k for $k \leq n$ we get g_{n+1} by applying Lemma 17.6 to $f - \sum_1^n g_k$ with $r = (2/3)^n$. Finally, we set $g = \sum_1^\infty g_k$. The sum is uniformly convergent so g is continuous and $\|f - g\|_Y = 0$ so $g|_Y = f$ as required. □

This lemma easily implies the special case of Theorem 17.5 where f is bounded (which is automatic if Y is compact). However, we will need an auxiliary construction to cover the general case.

Proof of Theorem 17.5. Let $\phi\colon \mathbb{R} \to (-1, 1)$ be a homeomorphism, for example $\phi(x) = x/\sqrt{1 + x^2}$ with $\phi^{-1}(y) = y/\sqrt{1 - y^2}$. Apply Lemma 17.7 to the map $\phi \circ f\colon Y \to (-1, 1) \subset [1, 1]$ to get a map $h\colon X \to [-1, 1]$ with $h|_Y = \phi \circ f$. Put $Z = \{x \in X \;:\; |h(x)| = 1\}$, so Z is closed and disjoint from Y. By Theorem 17.1 there exists $u\colon X \to [0, 1]$ with $u|_Y = 1$ and $u|_Z = 0$. We put $k = uh$ and observe that $|k(x)| < 1$ for all x, so $k\colon X \to (-1, 1)$. Put $g = \phi^{-1} \circ k$, which is a continuous map $g\colon X \to \mathbb{R}$. If $y \in Y$ then $u(y) = 1$ so $k(y) = h(y) = \phi(f(y))$, so $g(y) = f(y)$ as required. If we are given that $a \leq f(y) \leq b$ for all y, we simply replace g by the function $g'(x) = \min(b, \max(a, g(x)))$. □

17.2. The Stone-Weierstrass Theorem. The Stone-Weierstrass Theorem is a powerful tool for approximating real-valued functions by simpler functions such as polynomials. In order to state it, we need some preliminary definitions.

Definition 17.8. [`defn-subalg`]

A subset A of $C(X) = C(X, \mathbb{R})$ or $C(X, \mathbb{C})$ is a *subalgebra* if it contains all constant functions (real or complex, respectively) and is closed under addition and multiplication. In the complex case, we say that A is a $*$-subalgebra if it is also closed under conjugation.

Next, given $u, v \in C(X, \mathbb{R})$ define $u \vee v$ and $u \wedge v$ in $C(X, \mathbb{R})$ by

$$(u \vee v)(x) = \max(u(x), v(x))$$
$$(u \wedge v)(x) = \min(u(x), v(x)).$$

A subset $A \subseteq C(X, \mathbb{R})$ is a *sublattice* if it is closed under these two operations.

Definition 17.9. [`defn-separating`]

A subset A of $C(X, \mathbb{R})$ or $C(X, \mathbb{C})$ is *separating* if for all $x, y \in X$ with $x \neq y$, there exists $u \in A$ such that $u(x) \neq u(y)$.

Note that if X is normal then $C(X, \mathbb{R})$ itself is separating by Urysohn's theorem.

Theorem 17.10 (Stone-Weierstrass). [`thm-stone`]

If X is compact then any separating subalgebra of $C(X, \mathbb{R})$ is dense, as is any separating $*$-subalgebra of $C(X, \mathbb{C})$.

The proof will be given at the end of this subsection. Most of the time we will just discuss the real version. The complex case can be deduced easily from the real one, as we discuss at the end.

Example 17.11. [`eg-polynomial`]

Take $X = [0, 1]$ and $A = \mathbb{R}[x]$. This is easily seen to be a separating subalgebra of $C(X)$, so it is dense. In other words, for any continuous function $f\colon [0, 1] \to \mathbb{R}$ and any $\epsilon > 0$ we can find a polynomial p such that $|f(x) - p(x)| < \epsilon$ for all $x \in [0, 1]$.

Example 17.12. [`eg-fourier`]

Take $X = \mathbb{R}/(2\pi\mathbb{Z})$, and let A be the set of functions that have a finite Fourier series, say

$$f(x) = \sum_{k=0}^{N} a_k \cos(kx) + \sum_{k=1}^{N} b_k \sin(kx)$$

for some $a_k, b_k \in \mathbb{R}$. Equivalently, we have

$$f(x) = \sum_{k=-N}^{N} u_k e^{ikx}$$

for some complex numbers u_k with $u_{-k} = \overline{u_k}$. One can check that this is a separating subalgebra of $C(X)$, so it is dense. This is one of the first steps in a rigorous analytic treatment of Fourier series.

Example 17.13. [eg-hol-not-dense]
Put $X = \{z \in \mathbb{C} \;:\; |z| \leq 1\}$, and let A be the set of continuous functions $f\colon X \to \mathbb{C}$ that are holomorphic in the interior of X. This is a separating subalgebra of $C(X, \mathbb{C})$ that is not a $*$-subalgebra. We can define a continuous map $F\colon C(X, \mathbb{C}) \to \mathbb{C}$ by

$$F(f) = f(0) - \frac{1}{2\pi i}\int_{|z|=1} \frac{f(z)}{z}\, dz = f(0) - \frac{1}{2\pi}\int_{t=0}^{2\pi} f(e^{it})\, dt.$$

Cauchy's Theorem tells us that $F(f) = 0$ for all $f \in A$. However, if we put $g(z) = |z|$ then $g \in C(X, \mathbb{C})$ and $F(g) = -1$. It follows easily that A is not dense in $C(X, \mathbb{C})$.

Given a finite set $Y = \{y_1, \dots y_n\} \subseteq X$ write $F(Y)$ for the set of all (possibly discontinuous) functions $u\colon Y \to \mathbb{R}$. Note that $F(Y) \simeq \mathbb{R}^n$. In most cases of interest X is Hausdorff so Y is discrete and $F(Y) = C(Y)$.

Definition 17.14. [defn-interpolating]
A subset $A \subseteq C(X)$ is *interpolating* if for each finite $Y \subseteq X$ the restriction map $u \mapsto u|_Y$ is a surjection $A \to F(Y)$. Equivalently, given distinct points $y_1, \dots y_n$ and real numbers $a_1, \dots a_n$ there must exist a function $u \in A$ such that $u(y_k) = a_k$ for all k.

Proposition 17.15. [prop-subalgebra-closure]
The closure of a subalgebra is a subalgebra.

Proof. Suppose $A \subseteq C(X)$ is a subalgebra. Clearly all constant functions lie in $\overline{A}$. One checks easily that the following functions are continuous:

$$\begin{aligned} \sigma\colon C(X) \times C(X) &\to C(X) & \qquad \sigma(u, v) &= u + v \\ \mu\colon C(X) \times C(X) &\to C(X) & \qquad \mu(u, v) &= uv \end{aligned}$$

As A is an algebra, we have $\mu(A \times A) \subseteq A$ so

$$A \times A \subseteq \mu^{-1}(A) \subseteq \mu^{-1}(\overline{A})$$

As this last set is closed, we have

$$\overline{A} \times \overline{A} = \overline{A \times A} \subseteq \mu^{-1}(\overline{A})$$

so

$$\mu(\overline{A} \times \overline{A}) \subseteq \overline{A}$$

Similarly, $\sigma(\overline{A} \times \overline{A}) \subseteq \overline{A}$. It follows that $\overline{A}$ is a subalgebra as claimed. $\square$

Lemma 17.16. [lem-abs-approx]
There is a sequence $(p_n(x))_{n\in\mathbb{N}}$ of polynomial functions such that $p_n(x) \to |x|$ uniformly on the interval $[-1, 1]$.

Proof. We will use some standard theory of Taylor series. In particular, consider the function $f(x) = 1 - \sqrt{1 - x}$, and note that this is strictly increasing on the interval $[0, 1]$. We put $a_0 = 0$, and for $n > 0$ we put

$$a_n = (-1)^{n+1}\begin{pmatrix} 1/2 \\ n \end{pmatrix} = \frac{(-1)^{n+1}}{n!}\frac{1}{2}\frac{-1}{2}\frac{-3}{2}\dots\frac{3-2n}{2} = \frac{1}{2}\prod_{k=2}^{n}\frac{2k-3}{2k} > 0.$$

By a straightforward induction, we have $f^{(n)}(x) = n!a_n(1 - x)^{1/2-n}$ for $0 \leq x < 1$, and this function is strictly positive with $f^{(n)}(0) = n!a_n$. Now put $q_n(x) = \sum_{k=1}^{n-1} a_k x^k$, which is a Taylor approximation to $f(x)$. Taylor's theorem says that for $x \in (0, 1]$ there exists $w \in (0, x)$ with

$$f(x) = q_n(x) + f^{(n)}(w)/n! = q_n(x) + a_n w^n,$$

so $q_n(x) < f(x) < q_n(x) + a_n$. Next, as f and q_n are continuous on $[0, 1]$ and $q_n(x) < f(x)$ for $x \in [0, 1)$ we must also have $q_n(1) \leq f(1)$ for all n. Here $q_n(1) = \sum_{k=1}^{n-1} a_k$ and $f(1) = 1 - \sqrt{1-1} = 1$. As the numbers a_k are positive and all the partial sums are bounded by one, we deduce that $a_k \to 0$. The inequality $q_n \leq f \leq q_n + a_n$ therefore shows that $q_n \to f$ in $C([0, 1])$. Now put $p_n(x) = 1 - q_n(1 - x^2)$, and note that this is polynomial in x. It follows easily that these functions converge uniformly to $1 - f(1 - x^2) = \sqrt{x^2} = |x|$, as claimed. $\square$

Remark 17.17. The functions $p_n(x)$ arising in the above proof are not actually very good approximations to $|x|$. Consider instead the function

$$q_n(x) = \sum_{p,q=0}^{n} \frac{(-1)^{p+q}(2n+2p+1)!(2n+2q+1)!x^{2p}}{2^{4n}(2p+2q+1)(2q+2)(n+p)!(n+q)!(n-p)!(n-q)!(2p)!(2q)!}.$$

Using the theory of Hilbert matrices and Cauchy determinants, one can show that the L^2 distance from $q_n(x)$ to $|x|$ is minimal for polynomials of degree $2n$. For our present applications, we need to know about the L^∞ distance rather than the L^2 distance. Numerical experiments show that the L^∞ distance is approximately $0.3/n$, which is much better than the distance of approximately $1/\sqrt{\pi n}$ for $p_n(x)$. However, we do not know a formal proof of this fact.

Corollary 17.18. [cor-alg-lattice]
Any closed subalgebra is a sublattice.

Proof. Suppose A is a closed subalgebra, and that $w \in A$ has $\|w\| \leq 1$. As A is a subalgebra and p_n is a polynomial, we have $p_n(w) \in A$. Moreover, for all x we have $w(x) \in [-1, 1]$ so

$$|p_n(w(x)) - |w(x)|| \leq \sup_{[-1,1]} \{|p_n(t) - |t||\} \to 0$$

so $p_n(w) \to |w|$ in $C(X)$. As A is closed, this implies $|w| \in A$. More generally, let us not assume that $\|w\| \leq 1$. Write $\alpha = \|w\|$. By the previous argument, $|w/\alpha| \in A$ but A is an algebra so $|w| = \alpha|w/\alpha| \in A$.

Finally, suppose $u, v \in A$. Write $w = u - v \in A$ so $|w| = |u - v| \in A$. One checks easily that

$$\begin{aligned} u \vee v &= \tfrac{1}{2}(u + v + |u - v|) \in A \\ u \wedge v &= \tfrac{1}{2}(u + v - |u - v|) \in A, \end{aligned}$$

so A is a sublattice. □

Proposition 17.19. [prop-interpolating]
Any separating subalgebra is interpolating.

Proof. Let $A \subseteq C(X)$ be a separating subalgebra, and Y a finite subset of X. Suppose y, z are two distinct points in Y. As A is separating, there is a function $u_{yz} \in A$ such that the values $p = u_{yz}(y)$ and $q = u_{yz}(z)$ are distinct. Write

$$v_{yz}(x) = \frac{u_{yz}(x) - q}{p - q}.$$

As A is a subalgebra, we see that $v_{yz} \in A$. Clearly $v_{yz}(y) = 1$ and $v_{yz}(z) = 0$. Now write

$$v_y(x) = \prod_{z \in Y, z \neq y} v_{yz}(x)$$

Again, $v_y \in A$ because A is a subalgebra. Clearly $v_y(y) = 1$ and $v_y(z) = 0$ if z is any other point in Y. Finally, let $f \colon Y \to \mathbb{R}$ be an arbitary function. Define

$$v = \sum_{y \in Y} f(y) v_y$$

This is an element of A and for $z \in Y$ we have

$$v(z) = \sum_{y \in Y} f(y) v_y(z) = f(z)$$

(the terms in the sum for which $y \neq z$ are zero). Thus $v|_Y = f$. This shows that A is interpolating. □

Proposition 17.20. [prop-stone-lattice]
If X is compact then any interpolating sublattice is dense in $C(X)$.

Proof. Suppose $A \subseteq C(X)$ is an interpolating sublattice. Suppose $f \in C(X)$ and $\epsilon > 0$. We are required to find $h \in A$ such that $\|f - h\| < \epsilon$, or equivalently $f - \epsilon < h < f + \epsilon$.

For each $g \in A$ write

$$U(g) = \{x \ : \ g(x) > f(x) - \epsilon\}$$
$$V(g) = \{x \ : \ g(x) < f(x) + \epsilon\}.$$

These sets are clearly open. Fix $x \in X$ and write

$$\mathcal{U}(x) = \{U(g) \ : \ g \in A \text{ and } g(x) = f(x)\}$$

We claim that this is an open cover of X. Indeed, suppose that $y \in X$. By the interpolation property there is a function $g \in A$ with $g|_{\{x,y\}} = f|_{\{x,y\}}$, in other words $g(x) = f(x)$ and $g(y) = f(y) > f(y) - \epsilon$ so $y \in U(g) \in \mathcal{U}(x)$. This proves the claim.

There is thus a finite subcover $X = U(g_1) \cup \ldots U(g_n)$ with $g_k \in A$ and $g_k(x) = f(x)$. Write

$$g = g_1 \vee \ldots \vee g_n$$

As A is a sublattice we have $g \in A$. It is easy to see that $g > f - \epsilon$ and $g(x) = f(x)$ (so $x \in V(g)$).

Now write

$$\mathcal{V} = \{V(g) \ : \ g \in A \text{ and } g > f - \epsilon\}$$

If g is constructed as above then $x \in V(g) \in \mathcal{V}$, so $\mathcal{V}$ is an open cover of X. We thus have a finite subcover $X = V(g_1) \cup \cdots \cup V(g_m)$ with $g_k \in A$ and $g_k > f - \epsilon$. Write $h = \bigvee_{k=1}^m g_k$. This again lies in A and $f - \epsilon < h < f + \epsilon$ as required. $\square$

Proof of Theorem 17.10. We first consider the real case. Let A be a separating subalgebra of $C(X, \mathbb{R})$. Then $\overline{A}$ is a subalgebra by Proposition 17.15. It is thus a sublattice by Corollary 17.18, and it is interpolating by Proposition 17.19. It is thus dense by Proposition 17.20. As it is closed and dense, we see that $\overline{A} = C(X)$. In other words, A itself is dense.

Now instead let B be a separating $*$-subalgebra of $C(X, \mathbb{C})$. Put $A = B \cap C(X, \mathbb{R})$. For any $f \in B$ the functions $g = (f + f^*)/2 = \mathrm{Re}(f)$ and $h = (f - f^*)/(2i) = \mathrm{Im}(f)$ then lie in A and we have $f = g + ih$. It follows that $B = A + iA$ and that A is a separating subalgebra of $C(X; \mathbb{R})$. From the real case we see that A is dense in $C(X, \mathbb{R})$, and it follow that B is dense in $C(X, \mathbb{C})$. $\square$

As an application, we can deduce a classification of closed subalgebras that need not be separating.

Proposition 17.21. [`prop-closed-subalgebras`]
Let X be a compact Hausdorff space, and $A \subseteq C(X)$ a closed subalgebra. Define a relation E on X by

$$E = \bigcap_{f \in A} \mathrm{eq}(f) = \{(x, x') \in X^2 \ : \ f(x) = f(x') \textit{ for all } f \in A\}.$$

Then the space $Y = X/E$ is compact Hausdorff, and A is isometrically isomorphic to $C(Y)$.

Proof. Write q for the quotient map $X \to Y$. If $f \in A$ then it is tautological that $f(x) = f(x')$ whenever $(x, x') \in E$, so Proposition 5.61 tells us that there is a unique map $\phi(f) \colon Y \to \mathbb{R}$ with $f = \phi(f) \circ q$, and that this map is continuous. We write A' for the set of functions of the form $\phi(f)$ for some $f \in A$. Note that when $f, g \in A$ the function $phi(f) + \phi(g)$ is continuous with $(\phi(f) + \phi(g)) \circ q = f + g$, so we must have $\phi(f + g) = \phi(f) + \phi(g)$. By similar arguments we see that $\phi(fg) = \phi(f)\phi(g)$ and that ϕ sends constants to constants, so A' is a subalgebra of $C(Y)$. Moreover, as q is surjective we see that the values of the function $\phi(f)$ are the same as the values of f, so $\|\phi(f)\| = \|f\|$. In other words, ϕ gives an isometric isomorphism $A \to A'$.

Now note that $q \colon X \to Y$ is surjective, and X is compact, so Y is compact. Next, suppose y and y' are distinct points of Y. Then $y = q(x)$ and $y' = q(x')$ say, where $(x, x') \notin E$. This means that there is a function $f \in A$ with $f(x) \neq f(x')$, say $a = f(x) < a' = f(x')$. It follows that the function $g = \phi(f) \colon Y \to \mathbb{R}$ has $a = g(y) < a' = g(y')$. Now put $b = (a + a')/2$ and $U = \{z \ : \ g(z) < b\}$ and $U' = \{z \ : \ g(z) > b\}$. These give a Hausdorff pair for (y, y'). It follows that Y is Hausdorff as well as compact.

Now, A is closed in the complete space $C(X)$ so A is complete. Moreover, A' is isometrically isomorphic to A and hence complete, and hence closed in $C(Y)$. Thus, by the Stone-Weierstrass Theorem, we have $A' = C(Y)$. We conclude that $C(Y)$ is isometrically isomorphic to A as claimed. □

EXERCISE 17.1. [ex-osc]
Let $f\colon X \to Y$ be a continuous map of compact Hausdorff spaces. Show that the map

$$f^*\colon C(Y) \to C(X) \qquad\qquad f^*(u) = u \circ f$$

is continuous.

Now suppose X is a metric space (we shall use the same symbol d for all metrics). Define the ϵ-oscillation of u as

$$\mathrm{osc}_\epsilon(u) = \sup\{|u(x) - u(y)| \;:\; d(x,y) < \epsilon\}$$

Give a clean proof that $\mathrm{osc}_\epsilon\colon C(X) \to \mathbb{R}$ is continuous.

This shows that the set

$$U(\epsilon,\delta) = \{u \;:\; \mathrm{osc}_\epsilon(u) < \delta\}$$

is open. Prove that for fixed δ, we have

$$\bigcup_{\epsilon>0} U(\epsilon,\delta) = C(X)$$

These are the first steps in the proof of the following uniform Fourier approximation theorem. Let P be the space of functions $u\colon [-\pi,\pi] \to \mathbb{R}$ given by a finite Fourier series:

$$u(x) = \sum_{k=0}^{n} a_k \cos(kx) + b_k \sin(kx)$$

Then, given $\delta > 0$ there is a continuous map

$$F\colon C[-\pi,\pi] \to P$$

such that $d(u, F(u)) < \delta$ for all u. Of course, F is something like the Fourier transform, but we have to work out how to fix it up so that we take different but finite numbers of terms for different functions u, and have the result depending continuously on u.

Solution: The metric is derived from the norm

$$\|u\| = \|u\|_\infty = \sup\{|u(x)| \;:\; x \in X\}$$

Thus, if $u \in C(Y)$ then

$$\begin{aligned}
\|f^*(u)\| &= \sup\{|f^*(u)(x)| \;:\; x \in X\} \\
&= \sup\{|u(f(x))| \;:\; x \in X\} \\
&\leq \sup\{|u(y)| \;:\; y \in Y\} = \|u\|
\end{aligned}$$

Noting also that $f^*(u - v) = f^*(u) - f^*(v)$, we find that $d(f^*(u), f^*(v)) \leq d(u,v)$. This implies that f^* is continuous.

It is also easy to see that the norm function

$$n\colon C(X) \to \mathbb{R} \qquad\qquad n(u) = \|u\|$$

is continuous. This follows from the reversed triangle inequality:

$$|n(u) - n(v)| \leq d(u,v)$$

Now consider $Y = \{(x,x') \in X^2 \;:\; d(x,x') < \epsilon\}$. There are two continuous projection maps $\pi_0, \pi_1\colon Y \to X$. We have

$$\mathrm{osc}_\epsilon(u) = n(\pi_0^*(u) - \pi_1^*(u))$$

which shows that osc_ϵ is continuous.

We next want to show that $\bigcup_{\epsilon>0} U(\epsilon,\delta) = C(X)$. Consider $u \in C(X)$; we need to find $\epsilon > 0$ such that $\mathrm{osc}_\epsilon(u) < \delta$. This just means that u is uniformly continuous. A proof in the spirit of this problem is as follows. Write

$$K = \{(x,x') \in X^2 \;:\; |u(x) - u(x')| \geq \delta\}$$

The image $d(K)$ under the distance map $d\colon X^2 \to \mathbb{R}$ is compact and does not contain 0, so $d(K) \cap [0, \epsilon) = \emptyset$ for some $\epsilon > 0$. Thus $d(x, x') < \epsilon$ implies $|u(x) - u(x')| < \delta$ as required.

Exercise 17.2. [ex-qalg]
Let X be a space, and A a subset of $C(X)$. We identify a real number with the corresponding constant function, so $\mathbb{R} \subseteq C(X)$. We shall say that A is a $\mathbb{Q}$-*subalgebra* of $C(X)$ if it contains $\mathbb{Q}$ and is closed under addition and multiplication.

(a) Let B be a countable subset of $C(X)$. Prove that there is a countable $\mathbb{Q}$-subalgebra A such that $B \subseteq A \subseteq C(X)$. **Hint:** You may want to consider the construction which assigns to a set $C \subseteq C(X)$ the set
$$C' = C \cup \{f + g \;:\; f, g \in C\} \cup \{fg \;:\; f, g \in C\}.$$
(b) Let X be a compact metric space which has a countable dense subset. Prove that $C(X)$ has a countable dense subset.

Solution: We use the construction $C \mapsto C'$ as above. Note in particular that C' is countable if C is.

Suppose that $B \subseteq C(X)$ is countable. Define recursively
$$C_0 = \mathbb{Q} \cup B$$
$$C_{n+1} = C_n' \supseteq C_n$$
$$A = \bigcup_{n=0}^{\infty} C_n$$
I claim that A is a $\mathbb{Q}$-algebra. Indeed, $\mathbb{Q} \subseteq C_0 \subseteq A$. Moreover, if $f, g \in A$ then $f, g \in C_n$ for some n and so $f + g, fg \in C_{n+1} \subseteq A$. Also, each C_n is countable (by induction) so A is countable. Thus A is a countable $\mathbb{Q}$-algebra containing B, as required.

Now let X be a compact metric space which has a countable dense subset Y. Write $d_y(x) = d(y, x)$, so $d_y \in C(X)$. Write
$$B = \{d_y \;:\; y \in Y\}$$
(so B is a countable subset of $C(X)$).

I claim that $B \subseteq C(X)$ is separating. Indeed, suppose $u, v \in X$ and $u \neq v$, so $\epsilon = d(u, v)/2 > 0$. As Y is dense, there is a point $y \in Y \cap B(u, \epsilon)$. Then
$$d_y(u) = d(y, u) < \epsilon$$
$$d_y(v) = d(v, y) \geq d(v, u) - d(u, y) = 2\epsilon - d(u, y) > \epsilon$$
so $d_y(u) \neq d_y(v)$ as required.

Let A be a countable $\mathbb{Q}$-algebra containing B. Then $\overline{A}$ is a ring (see the proof that the closure of a $\mathbb{R}$-algebra is a $\mathbb{R}$-algebra) and contains $\overline{\mathbb{Q}} = \mathbb{R}$. Thus $\overline{A}$ is a closed separating $\mathbb{R}$-algebra. By Stone-Weierstrass, it is all of $C(X)$. Thus A is a countable dense subset of $C(X)$.

A popular error is to suppose that X need not be a metric space. One chooses a countable dense subset Y, uses Urysohn's lemma to choose a countable set B of functions separating any pair of points in Y and then argues as above. However, B need not separate the points of X. For example, take $X = [0, 1]$ and $B = \{f \in C[0, 1] \;:\; f(0) = f(1)\}$. Then B separates the points of the dense subset $(0, 1)$, but does not separate 0 from 1. This shows that we need to use functions of the special form indicated above. The result is false for non-metric spaces, the simplest example being $X = \beta\mathbb{N}$, the Stone-Čech compactification of the discrete space $\mathbb{N}$.

17.3. The Arzela-Ascoli Theorem. We next want to study compact subsets of $C(X, Y)$, particularly in the case where X is compact Hausdorff and $Y = \mathbb{R}$ or $Y = \mathbb{C}$. The key property that we need Y to satisfy is as follows:

Definition 17.22. [defn-arzela-space]
An *Arzela space* is a complete metric space Y such that every bounded closed subset is compact.

Remark 17.23. Proposition 10.28 tells us that $\mathbb{R}^n$ is an Arzela space. It is straightforward to check that any finite product of Arzela spaces is an Arzela space.

Remark 17.24. [`rem-arzela-bounded`]

Let Y be an Arzela space, and let Z be a bounded subset of Y; we claim that Z is totally bounded. Indeed, $\overline{Z}$ is both bounded and closed, so it is compact by the Arzela property, so it is totally bounded by Theorem 12.28. It follows by Corollary 12.33 that Z is also totally bounded.

Definition 17.25. [`defn-equicts`]

Let X be a topological space, and let Y be a metric space. Let A be a set of maps from X to Y. We say that A is *equicontinuous* if given $x \in X$ and $\epsilon > 0$ there is a neighbourhood U of x such that for any $f \in A$ and $x' \in U$ we have $d(f(x), f(x')) < \epsilon$.

Note that this is the same as the standard metric criterion for continuity, except that the same U is required to work simultaneously for all $f \in A$.

Theorem 17.26 (Arzela-Ascoli). [`thm-arzela`]

Let X be a compact Hausdorff space, let Y be an Arzela space, and let A be a subset of $C(X,Y)$. Then $\overline{A}$ is compact if and only if A is bounded and equicontinuous. Thus, A itself is compact if and only if A is closed, bounded and equicontinuous. In particular, this applies when $Y = \mathbb{R}$ or $Y = \mathbb{C}$.

Proof. First, suppose A is compact. Theorem 12.28 tells us that A is totally bounded (and therefore bounded) and closed. Thus, we need only show that A is equicontinuous. Suppose $x \in X$ and $\epsilon > 0$. We define maps

$$\psi\colon A \times X \to Y^2 \qquad \psi(a, x') = (u(x'), u(x)) = (\mathrm{ev}(u, x'), \mathrm{ev}(u, x))$$
$$\phi\colon A \times X \to \mathbb{R} \qquad \phi(a, x') = d(u(x), u(x')).$$

Note that ψ is continuous by Proposition 14.14 and $d\colon Y^2 \to \mathbb{R}$ is continuous by the triangle inequality so $\phi = d \circ \psi$ is also continuous. We define $U \subseteq A \times X$ by

$$U = \{(u, x') \in A \times X \ : \ d(u(x'), u(x)) < \epsilon\} = \phi^{-1}(-\epsilon, \epsilon).$$

This is open and contains $A \times \{x\}$. By the Tube Lemma (Lemma 10.38), there is a neighbourhood V of x such that $A \times V \subseteq U$, so for $u \in A$ and $x' \in V$ we have $d(u(x'), u(x)) < \epsilon$. This shows that A is equicontinuous.

Conversely, suppose A is closed, bounded and equicontinuous. We know from Proposition 12.16 that $C(X,Y)$ is a complete metric space, and from Theorem 12.28 that a subspace of a complete metric space is compact iff totally bounded and closed, so we need only show that A is totally bounded. Suppose $\epsilon > 0$. As A is equicontinuous we can find a neighbourhood U_x for each point $x \in X$ such that $d(f(x), f(x')) < \epsilon/4$ for all $x' \in U_x$ and $f \in A$. As X is compact we can choose a finite subset $T \subseteq X$ such that $\{U_x \ : \ x \in T\}$ covers X. Next, consider the set $F(T,Y)$ of all functions from T to Y. Note that if T has n points then $F(T,Y) \simeq Y^n$, so $F(T,Y)$ is an Arzela space. Now write

$$A|_Y = \{u|_Y \ : \ u \in A\} \subseteq F(T,Y)$$

As A is bounded we see that $A|_Y$ is bounded, and it follows by Remark 17.24 that $A|_Y$ is totally bounded. We can therefore find a finite set $B \subseteq A$ such that $B|_Y$ is an $\epsilon/4$-net in $A|_Y$. The claim is that B is an ϵ-net in A. Indeed, suppose $u \in A$. As $B|_Y$ is an $\epsilon/4$-net in $A|_Y$, there is an element $v \in B$ with $d(u|_Y, v|_Y) < \epsilon/4$. Suppose $x' \in X$. Then $x' \in U_x$ for some $x \in T$. Thus

$$d(u(x'), v(x')) < d(u(x'), u(x)) + d(u(x), v(x)) + d(v(x'), v(x)).$$

As $u, v \in A$ and $x' \in U_x$ we have

$$d(u(x'), u(x)), d(v(x'), v(x)) < \epsilon/4.$$

As $d(u|_Y, v|_Y) < \epsilon/4$, we have

$$d(u(x), v(x)) < \epsilon/4.$$

Putting this together gives $d(u(x'), v(x')) \leq 3\epsilon/4$. As x' was arbitrary we deduce that $d(u, v) \leq 3\epsilon/4 < \epsilon$, as required. $\square$

We can give a simple sufficient condition for equicontinuity as follows:

Definition 17.27. [`defn-equilipschitz`]
Let X and Y be metric spaces, and let A be a set of maps from X to Y. We say that A is *equilipschitz* if there is a constant $K > 0$ such that $d(f(x), f(x')) \leq Kd(x, x')$ for all $f \in A$ and $x, x' \in X$.

If A is equilipschitz then it is clear that the maps in A are continuous, and that the whole family is equicontinuous.

Example 17.28. [`eg-contractions-compact`]
Let X be a compact metric space, and let α be a number in $(0, 1)$. Let $CM_\alpha(X)$ be the set of contraction mappings of ratio α on X. We saw in Proposition 12.45 that $CM_\alpha(X)$ is closed in $C(X, X)$, and it is visibly equilipschitz (with constant α) and therefore equicontinuous. Moreover, as X is compact it must be bounded, and it follows that $CM_\alpha(X)$ is also bounded. We now see from the theorem that $CM_\alpha(X)$ is compact.

Example 17.29. [`eg-hol-equicts`]
Let X be the closed unit disc in $\mathbb{C}$. Fix some $r > 1$, let U be the open disc of radius r, and let A be the set of continuous maps $f : X \to X$ that can be extended to a holomorphic map $U \to X$. We claim that A bounded and equicontinuous, so the theorem will tell us that $\overline{A}$ is compact. Boundedness is clear. Next, fix s with $1 < s < r$. For $z \in X$ we have the Cauchy integral formula

$$f'(z) = \frac{1}{2\pi i} \int_{|w|=s} \frac{f(w)}{(w - z)^2},$$

in which $|f(w)| \leq 1$ and $|w - z| \geq s - 1$, so $|f'(z)| \leq (1 - s)^{-2}$. Thus, the Mean Value Theorem gives $|f(z) - f(y)| \leq (1 - s)^{-2}|z - y|$ for all $y, z \in X$, and the upper bound here is independent of f, which implies equicontinuity.

Example 17.30. [`eg-intop`]
Write $I = [0, 1]$, and let $k : I^2 \to \mathbb{R}$ be continuous. Define $K : C(I) \to C(I)$ by

$$K(f)(x) = \int_{t=0}^{1} k(x, y)f(y)\, dt.$$

Put $A = \{K(f) \;:\; f \in C(I),\ \|f\| \leq 1\}$. It is a a basic fact in the general theory of differential equations and integral operators that A is bounded and equicontinuous; we can prove this as follows. For boundedness, it is straightforward to check that $\|K(f)\| \leq \|K\|\|f\|$ so $\|g\| \leq \|K\|$ for all $g \in X$. Now suppose we are given $x \in I$ and $\epsilon > 0$. Consider the function $m(x', y) = k(x', y) - k(x, y)$ and the set $V = \{(x', y) \;:\; |m(x', y)| < \epsilon/2\}$. This is an open neighbourhood of $\{x\} \times I$ and I is compact, so V contains $U \times I$ for some open neighbourhood U of x (by the Tube Lemma 10.38). Now for $f \in C(I)$ with $\|f\| \leq 1$ and $x' \in U$ we have

$$|K(f)(x') - K(f)(x)| = \left| \int_{y=0}^{1} m(x', y)f(y)\, dy \right| \leq \epsilon\|f\|/2 < \epsilon.$$

This shows that A is equicontinuous, as claimed.

Exercise 17.3. [`ex-finite`]
Let X be a compact Hausdorff space such that the space $C(X, [0, 1])$ is compact. Prove that X is finite.

Solution: The Arzela-Ascoli theorem tells us that $C(X, [0, 1])$ must be equicontinuous. Consider a point $x \in X$. By equicontinuity, there is a neighbourhood U of x such that $|f(x) - f(y)| < 1$ for all $y \in U$ and all $f \in A$. We claim that $U = \{x\}$. Indeed, suppose not. Then there would be a point $y \neq x$ with $y \in U$. By Urysohn's Lemma (Theorem 17.1), we could choose a continuous function $f : X \to [0, 1]$ with $f(x) = 0$ and $f(y) = 1$, violating the equicontinuity estimate. Thus $U = \{x\}$ is open for each $x \in X$, so the sets $\{x\}$ form an open cover of X. By compactness there is a finite subcover $\{\{x_1\}, \dots \{x_n\}\}$, so $X = \{x_1, \dots x_n\}$ is finite.

We next discuss some aspects of equicontinuity that are important for understanding the Mandlebrot set and related ideas about fractals.

Let X be a topological space, let Y be a metric space, and let A be a subset of $C(X, Y)$. For any open subset $U \subseteq X$ we can consider the family

$$A|_U = \{f|_U : U \to Y\} \subseteq C(U, Y).$$

Even if the family A is not equicontinuous, it can happen that there are open sets U for which $A|_U$ is equicontinuous. We will show that there is a largest open set with this property (which may be empty).

PROPOSITION 17.31. [prop-fatou-exists]
Let X, Y and A be as above. Then there is an open set $U \subseteq X$ such that $A|_U$ is equicontinuous, and for other open sets V the family $A|_V$ is equicontinuous iff $V \subseteq U$.

PROOF. Put

$$U = \bigcup\{\text{open } V \subseteq X \;:\; A|_V \text{ is equicontinuous }\},$$

and note that this is open. It will suffice to show that $A|_U$ is equicontinuous. Suppose we are given $x \in U$ and $\epsilon > 0$. Then $x \in V$ for some V such that $A|_V$ is equicontinuous. This means that there is a set W containing x that is open in V (and so also in U or X) such that $d(f(y), f(x)) < \epsilon$ for all $f \in A$ and $y \in W$. This is precisely what is required for equicontinuity on U. □

An illustrative example is as follows.

PROPOSITION 17.32. [prop-powers-equicontinuous]
Define $f_n \colon \mathbb{R} \to \mathbb{R}$ by $f_n(x) = x^n$, and put $A = \{f_n \;:\; n \geq 2\} \subset C(\mathbb{R}, \mathbb{R})$. Then the maximal equicontinuity set U is the interval $(-1, 1)$.

It would not affect the conclusion if we included f_0 and f_1 in A, but the proof would need a few more words.

PROOF. The first step is to prove that when $0 < r < 1$, the family $A|_{(-r,r)}$ is equicontinuous. Using Lemma 17.33 below, we have the following chain of inequalities:

$$\begin{aligned}
|f_n(x) - f_n(y)| &= |x - y||x^{n-1} + x^{n-2}y + \dots y^{n-1}| \\
&\leq |x - y|(|x^{n-1}| + \dots |y^{n-1}|) \\
&\leq |x - y| n r^{n-1} \\
&\leq 2|x - y|/(r^{-1} - 1)
\end{aligned}$$

This estimate is independent of n, showing that $A|_{(-r,r)}$ is equilipschitz and hence equicontinuous, so $(-r, r) \subseteq U$. As this holds for all $r \in (0, 1)$ we deduce that $(-1, 1) \subseteq U$.

Now suppose that $x \geq 1$. We claim that there is no neighbourhood V of x such that $A|_V$ is equicontinuous. To see this, suppose $y = x + u \geq x$. Then (by the binomial expansion)

$$y^n - x^n = (x + u)^n - x^n \geq n x^{n-1} u$$

Note that $nx^{n-1} \to \infty$ as $n \to \infty$. Thus, we can only have $y^n - x^n < \epsilon$ for all n if $u = 0$, i.e. if $y = x$. Thus, there is no neighbourhood W of x such that $y \in W$ implies $|f_n(x) - f_n(y)| < \epsilon$ for all n. In other words, A is not equicontinuous in any neighbourhood of x. A similar argument works if $x \leq -1$. □

LEMMA 17.33. [lem-pow-diff]
For $0 < r < 1$ and $n \geq 2$ we have

$$n r^{n-1} \leq 2/(r^{-1} - 1)$$

PROOF. Write $\epsilon = r^{-1} - 1$ so $r = (1 + \epsilon)^{-1}$. Then, by the binomial expansion, we have

$$\begin{aligned}
(1 + \epsilon)^{n-1} &\geq (n - 1)\epsilon \\
r^{n-1} = \frac{1}{(1 + \epsilon)^{n-1}} &\leq \frac{1}{(n - 1)\epsilon}
\end{aligned}$$

Also, we assume $n \geq 2$ so $n/(n - 1) \leq 2$. Thus

$$n r^{n-1} \leq \frac{n}{(n - 1)\epsilon} \leq \frac{2}{\epsilon}$$

as claimed. □

18. Local Compactness

Definition 18.1. [defn-locally-compact]
A space X is said to be *locally compact* if every point has a precompact neighbourhood, or equivalently, the precompact open sets cover X.

Remark 18.2. Let X be a metric space. It is then easy to see that X is locally compact iff for all $x \in X$ thee exists $\epsilon > 0$ such that $B(x, \epsilon)$ is compact. From this it is clear that $\mathbb{R}^n$ (with the standard metric) is locally compact. Consider instead the space $\mathbb{R}^2$ with the lane metric. The compact sets were analysed in Example 12.40, and it is clear from that analysis that no neighbourhood of $(x, 0)$ is precompact, so this space is not locally compact.

Example 18.3. [eg-Q-not-locally-compact]
We claim that $\mathbb{Q}$ (topologised as a subspace of $\mathbb{R}$) is not locally compact. To see this, note that if $\epsilon > 0$ is irrational then in $\mathbb{Q}$ we have

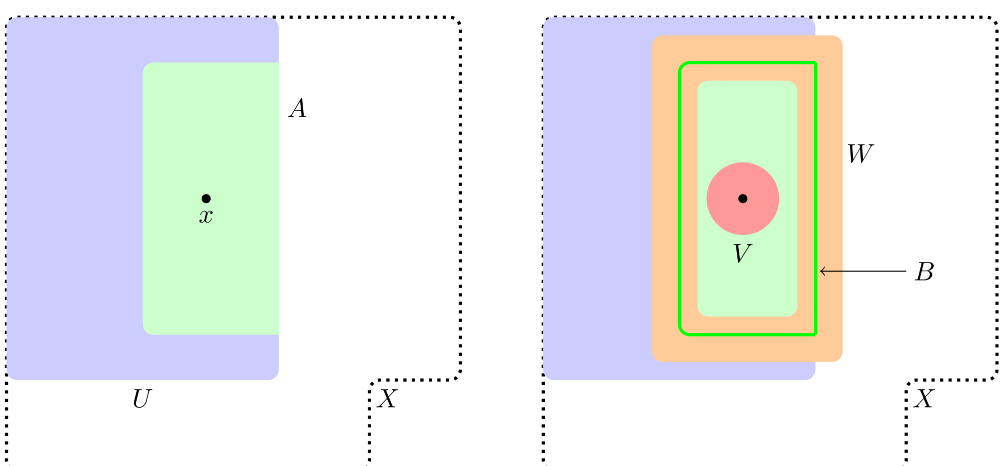
$$B(x, \epsilon) = [x - \epsilon, x + \epsilon] \cap \mathbb{Q} = (x - \epsilon, x + \epsilon) \cap \mathbb{Q} = OB(x, \epsilon).$$

The sets $B(x, \delta)$ (with $0 < \delta < \epsilon$) form an open cover of $B(x, \epsilon)$ with no finite subcover, so $B(x, \epsilon)$ is not compact. Moreover, if ϵ is rational we can still choose a smaller irrational number η, and we find that $B(x, \eta)$ is closed in $B(x, \epsilon)$ and is not compact, so $B(x, \epsilon)$ cannot be precompact.

Proposition 18.4. [prop-locally-compact]
Let X be a locally compact Hausdorff space.

(a) *For every $x \in X$ and every neighbourhood U of x, there is a precompact neighbourhood V of x such that $\overline{V} \subseteq U$.*
(b) *The precompact open sets form a basis for the topology on X.*
(c) *X is regular.*

Proof.
Let x be a point of X, and let U be an open neighbourhood of x. As X is locally compact we can choose a precompact open set A such that $x \in A$. Note that $A \cap U$ is also a precompact open neighbourhood of x, so we can replace A by $U \cap A$ and thus assume that $A \subseteq U$. Now put $B = \overline{A} \setminus A$ (the boundary of A); this is closed in the compact set $\overline{A}$, and so is compact. It follows from Lemma 14.10 that there exist disjoint open sets V and W such that $x \in V$ and $B \subseteq W$. After replacing V by $V \cap A$ if necessary, we may also assume that $V \subseteq A \subseteq U$. After this adjustment we see that V is contained in the precompact set A, so V is also precompact. Next, we note that $V \subseteq A$ and $V \cap W = \emptyset$, so V is contained in the closed set $C = \overline{A} \setminus W$, so $\overline{V} \subseteq C$. An the other hand, as $\overline{A} \setminus A = B \subseteq W$ we see that $C = \overline{A} \setminus W \subseteq A \subseteq U$. It follows that $\overline{V} \subseteq U$ as required.



(b) This is immediate from (a) and Proposition 2.28.

(c) This is immediate from (a) and Proposition 14.6.

□

We now consider when a subset Y of a locally compact Hausdorff space X is itself locally compact. It is not hard to see that this is true if Y is open or if Y is closed, but this does not exhaust the possibilities.

Proposition 18.5. [`prop-locally-closed`]
Let X be a locally compact Hausdorff space, and let Y be a subspace of X. Then the following are equivalent:

(a) *$Y = U \cap F$ for some open set U and some closed set F.*
(b) *Y is open in $\overline{Y}$.*
(c) *$Y \cup \overline{Y}^c$ is open in X.*
(d) *For each $y \in Y$ there exists an open neighbourhood U of y in X such that $Y \cap U$ is closed in U.*
(e) *Y is locally compact.*

(If these conditions hold, we say that Y is locally closed *in X.)*

The proof will be given after a number of lemmas.

Lemma 18.6. [`lem-locally-closed-a-b`]
Suppose that $Y = U \cap F$ for some open set U and some closed set F. Then $Y = U \cap \overline{Y}$, and so Y is open in $\overline{Y}$.

Proof. We have $Y \subseteq F$ and F is closed so $\overline{Y} \subseteq F$ so $U \cap \overline{Y} \subseteq U \cap F = Y$. In the other direction, we have $Y = U \cap F \subseteq U$ and certainly $Y \subseteq \overline{Y}$ so $Y \subseteq U \cap \overline{Y}$. □

Lemma 18.7. [`lem-locally-closed-b-c`]
Suppose that $Y = U \cap \overline{Y}$ for some open set U. Then $Y \cup \overline{Y}^c = U \cup \overline{Y}^c$, and this set is open in X.

Proof. As $Y = U \cap \overline{Y}$ we have $Y^c = U^c \cup \overline{Y}^c$. We can intersect both sides with $\overline{Y}$ to get $Y^c \cap \overline{Y} = U^c \cap \overline{Y}$, and then take complements again to get $Y \cup \overline{Y}^c = U \cup \overline{Y}^c$. As U and $\overline{Y}^c$ are open it follows that $Y \cup \overline{Y}^c$ is open. □

Lemma 18.8. [`lem-locally-closed-c-abd`]
Suppose that the set $U = Y \cup \overline{Y}^c$ is open in X. Then $Y = U \cap \overline{Y}$, so Y is open in $\overline{Y}$. Moreover, for each $y \in Y$ the set U is an open neighbourhood of y such that $Y \cap U$ is closed in U.

Proof. First, we have

$$U \cap \overline{Y} = (Y \cup \overline{Y}^c) \cap \overline{Y} = (Y \cap \overline{Y}) \cup (\overline{Y}^c \cap \overline{Y}) = Y \cup \emptyset = Y.$$

As U is open in X, this implies that Y is open in $\overline{Y}$. The identity $Y = U \cap \overline{Y}$ also shows that Y is closed in U, and of course $Y \cap U = Y$, so $Y \cap U$ is closed in U as claimed. □

Lemma 18.9. *Suppose that U is open in X and that $Y \cap U$ is closed in U; then $Y \cap U$ is open in $\overline{Y}$.*

Proof. By assumption there is a closed set F such that $Y \cap U = F \cap U$. We claim that $\overline{Y} \cap U$ is also equal to $F \cap U$. Indeed, as $Y \subseteq \overline{Y}$ we have $F \cap U = Y \cap U \subseteq \overline{Y} \cap U$. In the other direction, it is easy to see that $Y = (Y \cap U) \cup (Y \cap U^c)$ and $Y \cap U = F \cap U \subseteq F$ and $Y \cap U^c \subseteq U^c$ so $Y \subseteq F \cup U^c$. Here $F \cup U^c$ is closed, so $\overline{Y} \subseteq F \cup U^c$. We now intersect this with U to get $\overline{Y} \cap U \subseteq F \cap U$ as claimed. As $Y \cap U = \overline{Y} \cap U$ with U open in X, we see that $Y \cap U$ is open in $\overline{Y}$. □

Corollary 18.10. [`cor-locally-closed-d-b`]
Suppose that for each $y \in Y$ there exists an open neighbourhood U of y in X such that $Y \cap U$ is closed in U. Then Y is open in $\overline{Y}$.

Proof. We can consider Y as the union of the sets $Y \cap U$, and these are open in $\overline{Y}$ be the lemma. □

Corollary 18.11. [`cor-locally-closed-abcd`]
In Proposition 18.5, conditions (a) to (d) are equivalent.

Proof. Lemmas 18.6 and 18.7 show that (a) $\implies$ (b) $\implies$ (c). Lemma 18.8 shows that (c) implies (a), (b) and (d), whereas Corollary 18.10 shows that (d) implies (b). The claim follows by combining these results. □

Lemma 18.12. [lem-locally-closed-subspace]
Let X be a locally compact Hausdorff space, and let Y be a subspace of X.

(p) *If Y is closed in X, then it is locally compact.*
(q) *If Y is open in X, then it is locally compact.*
(r) *Suppose that $Y = U \cap F$, where U is open in X and F is closed in X; then Y is locally compact.*

Proof.

(p) Suppose that Y is closed, and consider a point $y \in Y$. As X is locally compact, there is a set U that is open in X such that $y \in U$ and $\overline{U}$ is compact. Now the set $V = U \cap Y$ is open in Y, and contains y. Moreover, as Y is closed in X we see that the closure of V in Y is the same as the closure in X. This is a closed subset of the compact set $\overline{U}$, so it is again compact. This proves that Y is locally compact.
(q) Suppose instead that Y is open. Now Proposition 18.4(a) tells us that for every $y \in Y$ there is an open set V such taht $\overline{Y}$ is compact with $y \in V \subseteq \overline{V} \subseteq Y$. This means that Y is locally compact.
(r) Now suppose that $Y = U \cap F$, where U is open in X and F is closed in X. Now U is locally compact by part (q), and Y is closed in U so it is locally compact by part (p).

□

Lemma 18.13. [lem-locally-closed-e-d]
Let X be a locally compact Hausdorff space, and let Y be a locally compact subspace of X. Then for each $y \in Y$ there exists an open neighbourhood U of y in X such that $Y \cap U$ is closed in U.

Proof. Let y be a point in Y. We can then find a set U_0 that is open and precompact in Y and has $y \in U_0$. As this is open in Y we have $U_0 = U \cap Y$ for some set U that is open in X. Now let $\overline{U_0}$ denote the closure of U_0 in X, so the closure of U_0 in Y is the set $F = \overline{U_0} \cap Y$. This is by assumption compact, and so is closed in X by Proposition 10.16. It will therefore suffice to show that $Y \cap U = F \cap U$. From the definition of F we have $F \subseteq Y$, and so $F \cap U \subseteq Y \cap U$. In the other direction, we have $Y \cap U = U_0 \subseteq \overline{U_0}$ and also $Y \cap U \subseteq Y$ so $Y \cap U \subseteq \overline{U_0} \cap Y = F$. It is also clear that $Y \cap U \subseteq U$, and we can combine these two inclusions to get $Y \cap U \subseteq F \cap U$ as required. □

Proof of Proposition 18.5. We have seen that conditions (a) to (d) are equivalent. Lemma 18.12(r) shows that (a) implies (e), and Lemma 18.13 shows that (e) implies (d). □

Remark 18.14. [rem-locally-closed-union]
If Y and Z are locally closed in X, it follows easily from criterion (a) that $Y \cap Z$ is also locally closed. However, the union $Y \cup Z$ need not be locally closed, as we see from the example where $X = \mathbb{R}^2$ and $Y = \{(0,0)\}$ and $Z = \{(x,y) \;:\; x > 0\}$. Here one can also check that the set $T = (Y \cup Z)^c = Y^c \cap Z^c$ is locally closed but T^c is not. Thus, the complement of a locally closed set need not be locally closed.

It is clear from the above discussion that any open subset of a compact Hausdorff space is a locally compact Hausdorff space. Conversely, given a locally compact Hausdorff space we can try to embed it as an open subset of a compact Hausdorff space. There are several different ways to do this. Our next task will be to discuss the simplest one.

Definition 18.15. [defn-one-point]
Let X be a locally compact Hausdorff. Let ∞ be some object not in X and write $X_\infty = X \cup \{\infty\}$. We declare a subset $U \subseteq X_\infty$ to be open if either

(a) U is an open subset of X; or
(b) $U = X_\infty \setminus K$ for some compact subset $K \subseteq X$.

We call this structure the *one-point compactification* of X.

Remark 18.16. [rem-one-point]
We claim that $U \subseteq X_\infty$ is open if and only if

(c) $U \cap X$ is open in X; and
(d) If $\infty \in U$ then $X \setminus U$ is compact.

Indeed, if (a) holds then (c) is immediate and (d) is vacuously satisfied. If (b) holds then $U \cap X = X \setminus K$ for some subset K which is compact and therefore closed (by Proposition 10.16); so (c) holds. We also have $X \setminus U = K$ so (d) holds. We leave it to the reader to check the converse, that (c) and (d) together imply ((a) or (b)).

Proposition 18.17. [prop-one-point]
The above definition gives a topology on X_∞, making it a compact Hausdorff space. Moreover, the subset $X \subset X_\infty$ is open, and the subspace topology on X is the same as the originally given topology.

Proof. It is clear that the sets $\emptyset$ and X_∞ are both open. Suppose that U and V are both open. Then $U \cap X$ and $V \cap X$ are both open in X, so the set $(U \cap V) \cap X = (U \cap X) \cap (V \cap X)$ is also open in X. If $\infty \in U \cap V$ then $\infty \in U$ and $\infty \in V$, so the sets $K = X \setminus U$ and $L = X \setminus V$ are both compact, so the set $X \setminus (U \cap V) = K \cup L$ is also compact. This shows that $U \cap V$ is again open in X_∞.

Now suppose we have a family $(U_i)_{i\in I}$ of open subsets of X_∞, and we put $U = \bigcup_I U_i$. The sets $U_i \cap X$ are then open in X, so the set $U \cap X = \bigcup_I (U_i \cap X)$ is also open. Suppose that $\infty \in U$. Then $\infty \in U_i$ for some i, so the set $K_i = X \setminus U_i$ must be compact. The set $K = X \setminus U$ is then closed in X and contained in the compact set K_i, so it must again be compact. This shows that U is open in X_∞, so we do indeed have a topology.

It is clear from the definitions that the subsets of X that are open in X_∞ are precisely the same as the subsets of X that are open in the original topology on X. In particular, X itself is open in X_∞, and we deduce that the subspace topology on X is the same as the original one.

We next claim that X_∞ is Hausdorff. Indeed, suppose we have two distinct points $x, y \in X_\infty$. If both of them lie in X, then (by the Hausdorff property of X) we can find a corresponding Hausdorff pair (U, V) in X, and this will still count as a Hausdorff pair in X_∞. Suppose instead that $x \in X$ but $y = \infty$. As X is locally compact, we can choose a precompact open neighbourhood U of x in X. We then put $V = X_\infty \setminus \overline{U}$, and observe that this is an open neighbourhood of $y = \infty$ that is disjoint from U. The case where $x = \infty$ and $y \in X$ can be treated in a symmetrical way.

Finally, we must show that X_∞ is compact. Suppose we have an open cover $(U_i)_{i\in I}$ of X_∞. We must then have $\infty \in U_i$ for some i, and thus the set $K_i = X_\infty \setminus U_i$ is a compact subset of X. It must therefore be finitely covered, so there is a finite subset $J \subseteq I$ such that $K_i \subseteq \bigcup_{j\in J} U_j$. It follows that if we put $J^* = J \cup \{i\}$ then $X_\infty = \bigcup_{j\in J^*} U_j$ as required. □

Remark 18.18. [rem-compact-cpfn]
If X itself is compact, we find that the set $\{\infty\} = X_\infty \setminus X$ is open in X_∞, and thus that X_∞ is just the disjoint union of X with an extra point.

Lemma 18.19. [lem-unpuncture]
Let X be a compact Hausdorff space, and let x be a point of X. Then $(X \setminus \{x\})_\infty = X$.

Proof. Put $Y = X \setminus \{x\}$, which is locally compact and Hausdorff. We have an obvious bijection $f\colon Y_\infty \to X$ given by $f(\infty) = x$ and $f(y) = y$ for $y \in Y$. We claim that this is a homeomorphism.

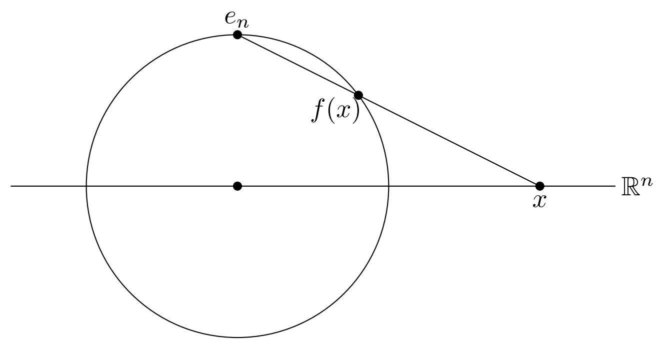
Let U be an open subset of X. If $x \notin U$ then U is an open subset of Y. If $x \in U$ then the set $K = X \setminus U$ is contained in Y, and it is closed in X and therefore compact, and we have $U = X \setminus K$. This is completely parallel to the description of the open sets in Definition 18.15, and the claim follows. □

Proposition 18.20. [prop-stereo]
The space $\mathbb{R}^n_\infty$ is homeomorphic to the space $S^n = \{x \in \mathbb{R}^{n+1} \,:\, \sum_i x_i^2 = 1\}$.

Proof. Put $e_n = (0, 1) \in \mathbb{R}^n \times \mathbb{R} = \mathbb{R}^{n+1}$. Define $f_0\colon \mathbb{R}^n \to S^n \setminus \{e_n\}$ by

$$f_0(x) = (2x, \|x\|^2 - 1)/(\|x\|^2 + 1).$$

Define $g_0 \colon S^n \setminus \{e_n\} \to \mathbb{R}^n$ by $g_0(u,t) = u/(1-t)$. One can check by direct calculation that f_0 and g_0 are continuous and that $f_0 g_0$ and $g_0 f_0$ are the respective identity maps, so f_0 and g_0 are mutually inverse homeomorphisms. Geometrically, the point $f(x)$ is just the unique point where the line from e_n to $(x,0)$ passes through S^n, so the picture is as follows:



We now see that $\mathbb{R}^n_\infty$ is homeomorphic to $(S^n \setminus \{e_0\})_\infty$, which in turn is homeomorphic to S^n by Lemma 18.19. More specifically, we can extend f_0 and g_0 to give mutually inverse homeomorphisms $\mathbb{R}^n_\infty \to S^n \to \mathbb{R}^n_\infty$ by putting $f(\infty) = e_0$ and $g(e_0) = \infty$. □

The map f used in the above proof is called *stereographic projection*.

Now recall (from Definition 4.1) that a map $f \colon X \to Y$ is said to be *proper* if it is continuous, and the preimage under f of any compact subset $L \subseteq Y$ is compact in X.

Proposition 18.21. [prop-proper-cpfn]
Let $f \colon X \to Y$ be a function between locally compact Hausdorff spaces. Define $f_\infty \colon X_\infty \to Y_\infty$ by $f_\infty(x) = f(x)$ for all $x \in X$, and $f_\infty(\infty) = \infty$. Then f_∞ is continuous if and only if f is proper.

Proof. First suppose that f is proper. Consider an open subset $V \subseteq Y_\infty$. If V is open in Y then we have $(f_\infty)^{-1}(V) = f^{-1}(V)$, which is open in X because f is continuous. Otherwise, we must have $V = Y_\infty \setminus L$ for some compact set $L \subseteq Y$. As f is proper, we see that the preimage $K = f^{-1}(L) \subseteq X$ is also compact. We find that $(f_\infty)^{-1}(V) = X_\infty \setminus K$, which is open in X_∞. It follows that f_∞ is continuous as claimed. We leave it to the reader to check that the argument is reversible. □

Remark 18.22. [rem-proper-cpfn]
We can give a categorical formulation of the above proposition as follows. Let **LCHP** be the category whose objects are locally compact Hausdorff spaces, and whose morphisms are proper maps. Let $\mathbf{CH}_*$ denote the category whose objects are compact Hausdorff spaces equipped with a specified basepoint, and whose morphisms are basepoint-preserving continuous maps. If $X \in \mathbf{LCHP}$ then we can take ∞ as the basepoint in X_∞ and thus regard X_∞ as an object of $\mathbf{CH}_*$. The proposition tells us that the construction $X \mapsto X_\infty$ gives a functor $\mathbf{LCHP} \to \mathbf{CH}_*$. More precisely, we can let $\mathbf{CH}'_*(Y,Z)$ be the set of morphisms for which the preimage of the basepoint consists only of the basepoint. This defines a wide subcategory $\mathbf{CH}'_* \subseteq \mathbf{CH}_*$, and our functor gives an equivalence $\mathbf{LCHP} \to \mathbf{CH}'_*$, with inverse $Y \mapsto Y \setminus \{\text{basepoint}\}$.

Proposition 18.23. [prop-collapse-space]
Let X be a compact Hausdorff space, and let Y be a closed subspace of X. Let X/Y be the quotient space where Y is collapsed to a point (as in Example 5.55), with the quotient topology. Then there is a canonical homeomorphism between X/Y and $(X \setminus Y)_\infty$.

Proof. Let $q \colon X \to X/Y$ be the quotient map. Define $f \colon X \to (X \setminus Y)_\infty$ as follows: if $x \in Y$ then $f(x) = \infty$, otherwise $x \in X \setminus Y \subset (X \setminus Y)_\infty$ and we put $f(x) = x$. We claim that f is continuous. To see this, consider an open set $V \subseteq (X \setminus Y)_\infty$. If V is actually an open set in $X \setminus Y$, then $f^{-1}(V) = V$, which is open. Otherwise we have $V = (X \setminus Y)_\infty \setminus L$ for some compact subset $L \subseteq X \setminus Y$. It then follows that $f^{-1}(V) = X \setminus L$, which is again open in X. This shows that f is continuous as claimed. Moreover, if x and x' lie in the same equivalence class then either $x = x'$ or $\{x, x'\} \subseteq Y$ and in either case we have $f(x) = f(x')$. It follows using Proposition 5.61 that the induced map $\overline{f} \colon X/Y \to (X \setminus Y)_\infty$ is continuous. Here X/Y is compact (by Corollary 10.21) and $(X \setminus Y)_\infty$ is Hausdorff (by Proposition 18.17) and $\overline{f}$ is a bijection (by inspection) so $\overline{f}$ is a homeomorphism (by Proposition 10.22). $\square$

Remark 18.24. [`rem-N-infty`]
Note that the discrete space $\mathbb{N}$ is locally compact and Hausdorff, so $\mathbb{N}_\infty$ is compact and Hausdorff. From the definitions we see that $U \subseteq \mathbb{N}_\infty$ is open iff it is either contained in $\mathbb{N}$ or has finite complement.

For a more concrete model, consider the space $T = \{2^{-n} : n \in \mathbb{N}\} \cup \{0\} \subset [0,1]$. We can define a map $f\colon \mathbb{N}_\infty \to T$ by $f(n) = 2^{-n}$ and $f(\infty) = 0$. This is a continuous bijection from a compact space to a Hausdorff space, so it is a homeomorphism.

There is an interesting relationship between $\mathbb{N}_\infty$ and the theory of convergent sequences, as follows.

Proposition 18.25. [`prop-N-infty`]
Let X be a topological space, and let $f\colon \mathbb{N}_\infty \to X$ be an arbitrary function. Then f is continuous if and only if the sequence $(f(n))_{n\in\mathbb{N}}$ converges to $f(\infty)$.

Proof. Consider an open set $U \subseteq X$. If $f(\infty) \notin U$ then $f^{-1}(U)$ is a subset of $\mathbb{N}$, and every subset of $\mathbb{N}$ is open in $\mathbb{N}_\infty$. On the other hand, if $f(\infty) \in U$ then $f^{-1}(U)$ is not contained in $\mathbb{N}$, so it is open in $\mathbb{N}_\infty$ iff it has finite complement, iff there exists N such that $f(n) \in U$ for all $n \geq N$. The claim is clear from this. $\square$

Remark 18.26. [`rem-lch-component-relation`]
We can use an example based on the above space T to show that Proposition 14.12 and Corollary 14.13 do not extend to locally compact Hausdorff spaces. Put $X = (T \times [0,1]) \setminus \{(0,1/2)\}$, and note that this is locally compact and Hausdorff. The connected components are the sets $A_n = \{2^{-n}\} \times [0,1]$ together with the sets $B = \{0\} \times [0,1/2)$ and $C = \{0\} \times (1/2,1]$. Any clopen subset $F \subseteq X$ must be the disjoint union of some of these components. If F meets B, then (because it is open), it must meet A_n for all large n. It must therefore contain A_n for all large n, and because it is closed, it must contain C. Thus, the intersection of the clopen sets containing $(0,0)$ is $B \cup C$, which is strictly larger than the component B, so Proposition 14.12 does not extend to this case. Similarly, we can define an equivalence relation

$$E = (B \times B) \cup (C \times C) \cup \bigcup_n (A_n \times A_n)$$

as in Corollary 14.13. We find that $(2^{-n},0)$ and $(2^{-n},1)$ are related for all n, but that $(0,0)$ and $(0,1)$ are not, so E is not closed in $X \times X$.

Now consider the space $\mathbb{C}_\infty$. This is known as the *Riemann sphere*; it is homeomorphic to S^2 by Proposition 18.20.

Lemma 18.27. [`lem-riemann-inverse`]
There is a homeomorphism $\chi\colon \mathbb{C}_\infty \to \mathbb{C}_\infty$ given by $\chi(0) = \infty$ and $\chi(\infty) = 0$ and $\chi(z) = 1/z$ for $z \in \mathbb{C}^\times$.

Proof. First, it is clear that $\chi^2 = 1$, so χ is a bijection and is its own inverse. Next, we claim that the restricted map $\chi\colon \mathbb{C}^\times \to \mathbb{C}^\times$ is continuous. This can be proved by the same method as used for $\mathbb{R}$ in Proposition 3.10, or deduced from Corollary 3.11 using the formula

$$\chi(x+iy) = \frac{x-iy}{x^2+y^2}.$$

Now consider an open set $U \subseteq \mathbb{C}_\infty$. As $\chi\colon \mathbb{C}^\times \to \mathbb{C}^\times$ is continuous, we see that the set $\chi^{-1}(U) \setminus \{0,\infty\}$ is open. Suppose that $0 \in \chi^{-1}(U)$, or equivalently $\infty \in U$. By the definition of the topology on $\mathbb{C}_\infty$, we see that the set $K = \mathbb{C}_\infty \setminus U$ is compact and therefore bounded in $\mathbb{C}$. We can thus find $R > 0$ such that $K \subseteq B_R(0)$, so U contains all $z \in \mathbb{C}$ with $|z| > R$, so $OB_{1/R}(0) \subseteq \chi^{-1}(U)$, so 0 is in the interior of $\chi^{-1}(U)$. Suppose instead that $\infty \in \chi^{-1}(U)$, or equivalently that $0 \in U$. As U is open, there exists $\epsilon > 0$ such that $OB_\epsilon(0) \subseteq U$. It follows that $\chi^{-1}(U)$ contains the set $\mathbb{C}_\infty \setminus B_{1/\epsilon}(0)$, so ∞ is in the interior. It follows that $\chi^{-1}(U)$ is open in all cases. As U was arbitrary we see that χ is continuous. As χ is self-inverse, it is therefore a homeomorphism. $\square$

Proposition 18.28. [prop-hopf-quotient]
Consider the map $\eta\colon \mathbb{C}^2 \setminus \{0\} \to \mathbb{C}_\infty$ *given by*

$$\eta(x,y) = \begin{cases} x/y & \text{if } y \neq 0 \\ \infty & \text{if } y = 0. \end{cases}$$

Then η *is surjective, continuous and open (and thus is a quotient map). Moreover, we have* $\eta(x,y) = \eta(u,v)$ *iff there exists* $w \in \mathbb{C}^\times$ *such that* $w.(x,y) = (u,v)$*. (This* w *is unique if it exists.)*

Proof. For every $x \in \mathbb{C}$ we have $\eta^{-1}(\{x\}) = \{(xw,w) \;:\; w \in \mathbb{C}^\times\}$, and we also have $\eta^{-1}(\{\infty\}) = \{(w,0) \;:\; w \in \mathbb{C}^\times\}$. It follows that η is surjective, and that $\eta(x,y) = \eta(u,v)$ iff there exists $w \in \mathbb{C}^\times$ such that $w.(x,y) = (u,v)$.

Now put $U_0 = \mathbb{C} \times \mathbb{C}^\times$ and $U_1 = \mathbb{C}^\times \times \mathbb{C}$; these are open sets whose union is $\mathbb{C}^2 \setminus \{0\}$. Let η_i be the restriction of η to U_i. The map η_0 is just $(x,y) \mapsto x/y$, and this is continuous by the evident complex analog of Corollary 3.11. We also have $\eta_1(x,y) = \chi\eta_0(y,x)$ and χ is a homeomorphism, so η_1 is continuous. As the sets U_i are open and cover $\mathbb{C}^2 \setminus \{0\}$ we can conclude (by Proposition 5.9) that η is continuous.

Now define $\zeta_0\colon U_0 \to U_0$ by $\zeta_0(x,y) = (x/y,y)$. This is a homeomorphism (with inverse $\zeta_0^{-1}(u,y) = (uy,y)$), and we have $\eta_0 = \pi_0 \circ \zeta_0\colon \mathbb{C} \times \mathbb{C}^\times \to \mathbb{C}$. We also know from Corollary 5.30 that π_0 is an open map, and it follows that η_0 is also an open map. Using the relation $\eta_1(x,y) = \chi\eta_0(y,x)$ again we see that η_1 is also an open map. For any open set $V \subseteq \mathbb{C}^2 \setminus \{0\}$ we have $\eta(V) = \eta_0(V \cap U_0) \cup \eta_1(V \cap U_1)$, and using this we see that η itself is open. It follows by Proposition 4.8 that η is a quotient map. □

Remark 18.29. [rem-hopf-fibration]
We can restrict η to the subspace $\{(x,y) \in \mathbb{C}^2 \;:\; |x|^2 + |y|^2 = 1\}$, which can be identified with S^3; this gives a map $\eta_1\colon S^3 \to S^2$, called the *Hopf fibration*. For $z \in \mathbb{C}$ we find that

$$\eta_1^{-1}(\{z\}) = \{u.(z,1)/\sqrt{1+|z|^2} \;:\; u \in S^1 \subset \mathbb{C}\},$$

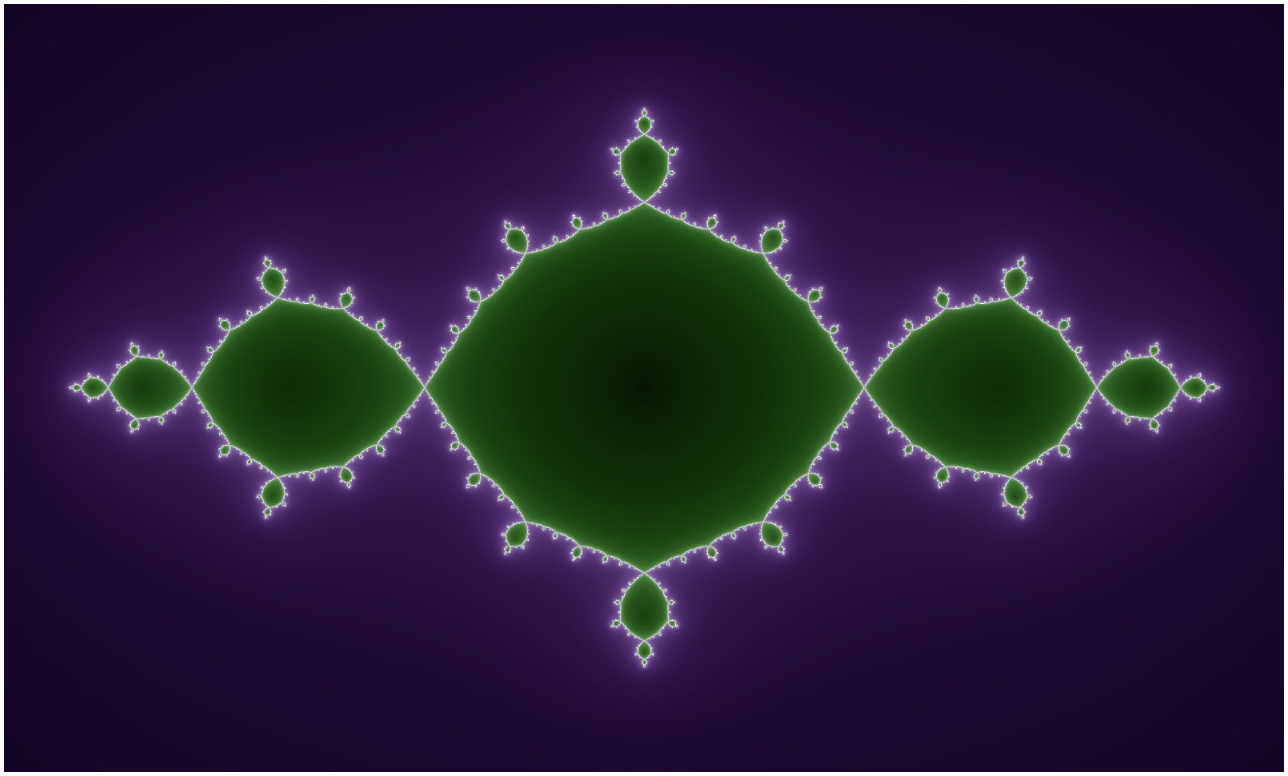
whereas $\eta_1^{-1}(\{\infty\}) = \{(u,0) \;:\; u \in S^1\}$. This implies that η_1 is a continuous surjection between compact Hausdorff spaces, so it is automatically a quotient map.

Proposition 18.30. [prop-rational-riemann]
Let $u(z)$ *and* $v(z)$ *be polynomials with complex coefficients, not both identically zero. Then there is a unique continuous map* $f\colon \mathbb{C}_\infty \to \mathbb{C}_\infty$ *satisfying* $f(z) = u(z)/v(z)$ *for all* $z \in \mathbb{C}$ *such that* $v(z) \neq 0$*, and also* $f(z) = \infty$ *whenever* $v(z) = 0 \neq u(z)$*.*

Proof. If $u = 0$ we must take f to be the constant function with value zero, and if $v = 0$ we must take f to be the constant function with value ∞. For the rest of the proof we assume that neither u nor v is identically zero. It is then standard that we can write $u(z) = \overline{u}(z)w(z)$ and $v(z) = \overline{v}(z)w(z)$ for some polynomials $\overline{u}(z)$, $\overline{v}(z)$ and $w(z)$ such that $\overline{u}(z)$ and $\overline{v}(z)$ are coprime. This implies that $\overline{u}$ and $\overline{v}$ have no common roots, so we can define $\tilde{f}_0\colon \mathbb{C} \to \mathbb{C}^2 \setminus \{0\}$ by $\tilde{f}_0(z) = (\overline{u}(z), \overline{v}(z))$. We then put $f_0 = q \circ \tilde{f}_0\colon \mathbb{C} \to \mathbb{C}_\infty$; this is clearly continuous. Now let d be the maximum of the degrees of $\overline{u}$ and $\overline{v}$. Put $u^*(z) = z^d u(1/z)$ and $v^*(z) = z^d v(1/z)$. We find that these are again polynomials with no common roots, so we can define $\tilde{f}_1\colon \mathbb{C} \to \mathbb{C}^2 \setminus \{0\}$ by $\tilde{f}_1(z) = (u^*(z), v^*(z))$. Using this, we define a continuous map $f_1\colon \mathbb{C}_\infty \setminus \{0\} \to \mathbb{C}_\infty$ by $f_1 = q \circ \tilde{f}_1 \circ \chi$. It is straightforward to check that f_0 and f_1 agree on the open set $\mathbb{C}^\times = \mathbb{C} \cap (\mathbb{C}_\infty \setminus \{0\})$ where both are defined. They can thus be patched together to give a map $f\colon \mathbb{C}_\infty \to \mathbb{C}_\infty$, which is continuous by Proposition 5.9. Note that if $z \in \mathbb{C}$ with $v(z) \neq 0$ then we must have $\overline{v}(z), w(z) \neq 0$ and $f(z) = f_0(z) = \overline{u}(z)/\overline{v}(z) = u(z)/v(z) \in \mathbb{C}$. Similarly, if $v(z) = 0 \neq u(z)$ we must have $\overline{v}(z) = 0$ and $\overline{u}(z), w(z) \neq 0$, so $f(z) = q(\overline{u}(z), 0) = \infty$. Thus, our map f has the stated properties. Let $g\colon \mathbb{C}_\infty \to \mathbb{C}_\infty$ be another continuous map with the stated properties. As v is not identically zero we see that the set $Z = \{z \in \mathbb{C} \;:\; v(z) = 0\}$ is finite, and $f = g$ on the set $\mathbb{C} \setminus Z$ which is dense in $\mathbb{C}_\infty$, so we have $f = g$. Thus f is unique as claimed. □

Remark 18.31. [defn-julia]
Let $f\colon \mathbb{C}_\infty \to \mathbb{C}_\infty$ be as in the above Proposition, and put $A = \{f^n \;:\; n \in \mathbb{N}\} \subseteq C(\mathbb{C}_\infty, \mathbb{C}_\infty)$. Here f^n denotes the n'th iterate, so $f^3 = f \circ f \circ f$, for example. By Proposition 17.31, there is a largest open set $F(f) \subseteq \mathbb{C}_\infty$ such that $A|_{F(f)}$ is equicontinuous. This is called the *Fatou set* of f, and the complement

$J(f) = \mathbb{C}_\infty \setminus F(f)$ is called the *Julia set* of f. These sets are typically highly complex and fractal. The white part of the picture below is the Julia set for the polynomial $f(x) = x^2 - 1$:



More generally, we can consider the polynomials $f_c(z) = z^2 + c$. Recall that the Mandelbrot set M is the set of those c for which the sequence $(f_c^n(0))_{n\in\mathbb{N}}$ is bounded. It can be shown that $c \in M$ if and only if the Julia set $J(f_c)$ is connected. If $c \notin M$ it can be shown that $J(f_c)$ is totally disconnected and equal to its own boundary.

Remark 18.32. [`rem-real-ends`]
If we work over the reals, it is sometimes more natural to use the set $\mathbb{R}_{\pm\infty} = \mathbb{R} \cup \{-\infty, \infty\}$ rather than $\mathbb{R}_\infty = \mathbb{R} \cup \{\infty\}$. We can topologise this by declaring that the family

$$\sigma = \{[-\infty, b) \ : \ b \in \mathbb{R}\} \cup \{(a, \infty] \ : \ a \in \mathbb{R}\}$$

is a subbasis of open sets. This makes $\mathbb{R}_{\pm\infty}$ a compact Hausdorff space, homeomorphic to the interval $[0, 1]$. One can check that every polynomial $f(x) \in \mathbb{R}[x]$ extends to give a continuous map $\overline{f} \colon \mathbb{R}_{\pm\infty} \to \mathbb{R}_{\pm\infty}$. Using this and the compactness of $\mathbb{R}_{\pm\infty}$, we can deduce that the map $f \colon \mathbb{R} \to \mathbb{R}$ is closed.

However, rational functions do not work as well in this context. For example, the rational function $f(x) = 1/x$ has a continuous extension $\mathbb{R}_\infty \to \mathbb{R}_\infty$ (with $0 \to \infty$) but not a continuous extension $\mathbb{R}_{\pm\infty} \to \mathbb{R}_{\pm\infty}$ (because the options $0 \mapsto +\infty$ and $0 \mapsto -\infty$ both lead to problems).

Another important application of the one-point compactification is the Pontrjagin-Thom construction, which we now discuss.

Definition 18.33. [`defn-pontrjagin-thom`]
Let $f \colon X \to Y$ be an open embedding of locally compact Hausdorff spaces. We define $f^! \colon Y_\infty \to X_\infty$ by

$$f^!(y) = \begin{cases} x & \text{if } y \in Y \text{ and } y = f(x) \text{ for some } x \in X \\ \infty & \text{if } y = \infty \text{ or } y \in Y \setminus f(X). \end{cases}$$

(The first clause is well-defined because f is assumed to be injective.)

Proposition 18.34. [prop-pt-cts]
The map $f^!\colon Y_\infty \to X_\infty$ is continuous.

Proof. Let U be an open set in X_∞. First consider the case where $U \subseteq X$. We then see from the definitions that $(f^!)^{-1}(U) = f(U) \subseteq Y$. This is open in Y because $f\colon X \to Y$ is assumed to be an open embedding. It is therefore also open in Y_∞, as required.

Now suppose instead that $\infty \in U$. This means that $U = X_\infty \setminus K$ for some compact set $K \subseteq X$, and thus that $(f^!)^{-1}(U) = Y_\infty \setminus (f^!)^{-1}(K)$. We again see from the definitions that $(f^!)^{-1}(K) = f(K)$, which is a compact subset of Y. It follows once more that $(f^!)^{-1}(U)$ is open in Y_∞, as required. □

We will see a number of examples of this when we discuss manifolds in Section 20. There will also be applications to homotopy theory in Section 27.

19. Examples from linear algebra

We will now take a more systematic look at various spaces related to linear algebra. As far as possible we will work with abstract vector spaces rather than $\mathbb{R}^n$ or $\mathbb{C}^n$; this turns out to have many conceptual and technical advantages.

19.1. Spaces of linear maps. Recall (from Definition 12.21) that a *Banach space* is a normed vector space that is complete with respect to the corresponding metric.

19.2. Inner products. Many constructions will involve inner products or hermitian products so we start by reviewing some facts about these.

Definition 19.1. [defn-inner-product]
Let V be a vector space over $\mathbb{R}$. An *inner product* on V is a rule giving a number $\langle x, y\rangle \in \mathbb{R}$ for each $x, y \in V$ such that

IP0: $\langle sx + ty, z\rangle = s\langle x, z\rangle + t\langle y, z\rangle$ for all $s, t \in \mathbb{R}$ and $x, y, z \in V$.
IP1: $\langle x, y\rangle = \langle y, x\rangle$ for all $x, y \in V$.
IP2: $\langle x, x\rangle \geq 0$ for all x, with equality if and only if $x = 0$.

In this context, we write $\|x\| = \sqrt{\langle x, x\rangle}$.

The Cauchy-Schwartz inequality and the triangle inequality continue to hold in this context:

Lemma 19.2. [lem-triangle-general]
Let V be a vector space over $\mathbb{R}$ equipped with an inner product. Then for $u, v \in V$ we have $|\langle u, v\rangle| \leq \|u\|\|v\|$ and $\|u + v\| \leq \|u\| + \|v\|$. (Thus, the rule $x \mapsto \|x\|$ gives a norm in the sense of Definition 3.31.)

Proof. Both claims are clear if $v = 0$, so we may assume that $v \neq 0$ and therefore $\langle v, v\rangle > 0$. Put

$$f(t) = \langle u - tv, u - tv\rangle = \langle u, u\rangle - 2t\langle u, v\rangle + t^2\langle v, v\rangle.$$

By axiom IP2 we have $f(t) \geq 0$ for all t. Now put $t_0 = \langle u, v\rangle/\langle v, v\rangle$. We find that

$$f(t_0) = \langle u, u\rangle - 2\frac{\langle u, v\rangle^2}{\langle v, v\rangle} + \frac{\langle u, v\rangle^2\langle v, v\rangle}{\langle v, v\rangle^2} = \langle u, u\rangle - \frac{\langle u, v\rangle^2}{\langle v, v\rangle}.$$

As this is nonnegative we have $\langle u, v\rangle^2 \leq \langle u, u\rangle\langle v, v\rangle$, or equivalently $|\langle u, v\rangle| \leq \|u\|\|v\|$ as claimed.

This in turn gives

$$\begin{aligned}\|u + v\|^2 = \langle u + v, u + v\rangle = \langle u, u\rangle + \langle v, v\rangle + 2\langle u, v\rangle &= \|u\|^2 + \|v\|^2 + 2\langle u, v\rangle \\ \leq \|u\|^2 + \|v\|^2 + 2\|u\|\|v\| &= (\|u\| + \|v\|)^2,\end{aligned}$$

so $\|u + v\| \leq \|u\| + \|v\|$. □

Definition 19.3. [defn-hermitian-product]
Let V be a vector space over $\mathbb{C}$. A *hermitian product* on V is a rule giving a number $\langle x, y\rangle \in \mathbb{C}$ for each $x, y \in V$ such that

IP0: $\langle sx + ty, z\rangle = s\langle x, z\rangle + t\langle y, z\rangle$ for all $s, t \in \mathbb{C}$ and $x, y, z \in V$.
IP1: $\langle x, y\rangle = \overline{\langle y, x\rangle}$ for all $x, y \in V$.

IP2: $\langle x, x\rangle$ is real and nonnegative for all x, and is zero iff $x = 0$.

In this context, we again write $\|x\| = \sqrt{\langle x, x\rangle}$.

Remark 19.4. [`rem-conjugate-linear`]

Axiom IP1 already implies that $\langle x, x\rangle$ is real, so it is not strictly necessary to include that in IP2. Using IP0 and IP1 we also see that

$$\langle x, sy + tz\rangle = \overline{s}\langle x, y\rangle + \overline{t}\langle x, z\rangle$$

for all $s, t \in \mathbb{C}$ and $x, y, z \in V$.

Remark 19.5. [`rem-hermitian-inner`]

Let V be a vector space over $\mathbb{C}$. If $\langle\cdot,\cdot\rangle$ is a hermitian product on V, then we can define an inner product on V (regarded as a real vector space) by $\langle x, y\rangle' = \mathrm{Re}(\langle x, y\rangle)$. Conversely, if $\langle\cdot,\cdot\rangle'$ is an inner product, one can check that the rule

$$\langle x, y\rangle = \langle x, y\rangle' + i\langle x, iy\rangle'$$

gives a hermitian product on V. Moreover, these constructions are inverse to each other, so they give a bijection between inner products and hermitian products.

Definition 19.6. [`defn-hilbert-space`]

A *Hilbert space* is a vector space V over $\mathbb{R}$ or $\mathbb{C}$ equipped with an inner product or hermitian product such that V is complete with respect to the metric $d(x, y) = \|x - y\| = \sqrt{\langle x - y, x - y\rangle}$.

Proposition 19.7. [`prop-finite-hilbert`]

Let V be a finite-dimensional vector space over $\mathbb{R}$ or $\mathbb{C}$ equipped with an inner product or hermitian product. Then V is a Hilbert space.

Proof. The claim does not involve the complex structure (if there is one) so we need only discuss the real case. We note that V is isomorphic to $\mathbb{R}^n$, so we may assume that $V = \mathbb{R}^n$ (with a possibly nonstandard inner product). Corollary 12.15 tells us that $\mathbb{R}^n$ is complete with respect to the metric

$$d_\infty(x, y) = \max(|x_1 - y_1|, \ldots, |x_n - y_n|).$$

Part (b) of Proposition 10.34 implies that the metric coming from the inner product is strongly equivalent to d_∞, and the claim follows. □

Proposition 19.8. [`prop-l-two-I-hilbert`]

Let I be an arbitrary set, and consider the vector space $l^2(I)$ defined in Remark 12.20. Then $l^2(I)$ is a Hilbert space in a natural way.

Proof. Let x be a map from I to $\mathbb{R}$. Recall that $\|x\|$ is defined as the supremum of the numbers $\|x\|_J = \sqrt{\sum_{j\in J} x_j^2}$, where J runs over the finite subsets of I, and $l^2(I)$ is defined to be the set of those x for which $\|x\| < \infty$. Remark 12.20 shows that $l^2(I)$ is a complete normed vector space. The only issue is to prove that the norm comes from an inner product. To do this, we define

$$\langle x, y\rangle = (\|x + y\|^2 - \|x - y\|^2)/4.$$

This formulation makes it clear that $\langle x, y\rangle$ is well-defined and finite. Now put

$$\langle x, y\rangle_J = (\|x + y\|_J^2 - \|x - y\|_J^2)/4.$$

A straightforward calculation shows that this is the same as $\sum_{j\in J} x_j y_j$, so the function $(x, y) \mapsto \langle x, y\rangle_J$ satisfies axioms IP0 and IP1. We also have $\langle x, x\rangle_J \geq 0$, with equality iff $x|_J = 0$.

Now suppose we have numbers u_J defined for all finite subsets $J \subseteq I$, and a number $a \in \mathbb{R}$. We write $a = \lim_J u_J$ if for every $\epsilon > 0$ there exists J such that $|a - u_K| < \epsilon$ whenever $K \supseteq J$. It is not hard to see that $\|x\| = \lim_J \|x\|_J$ and thus that $\langle x, y\rangle = \lim_J \langle x, y\rangle_J$. From this it follows that the function $(x, y) \mapsto \langle x, y\rangle$ also satisfies IP0 and IP1. Moreover, we see from the definitions that $\langle x, x\rangle = \|x\|^2$ and we already know that $\|x\|$ is a norm so IP2 also holds. □

Definition 19.9. [defn-orthogonal-complement]

Let V be a vector space over $\mathbb{R}$ equipped with an inner product, or a vector space over $\mathbb{C}$ equipped with a hermitian product. Let W be a vector subspace of V. We then put

$$W^\perp = \{v \in V \;:\; \langle v, w\rangle = 0 \text{ for all } w \in W\},$$

and call this the *orthogonal complement* of W.

Remark 19.10. [rem-orthogonal-complement]

In the complex case, there are apparently two possible definitions of $W^\perp$: we could define it using the given hermitian product, or using the inner product $\langle x, y\rangle' = \mathrm{Re}(\langle x, y\rangle)$. Provided that W is a complex subspace, these are actually the same. Indeed, if $\langle v, w\rangle' = 0$ for all $w \in W$ then we also have $\langle v, iw\rangle' = 0$ (because $iw \in W$) and so $\langle v, w\rangle = \langle v, w\rangle' + i\langle v, iw\rangle' = 0$.

Remark 19.11. [rem-complement-closed]

We claim that $W^\perp$ is always closed, for any subspace $W \leq V$. Indeed, if $v \not\in W^\perp$ then we can find $w \in W$ and $\epsilon > 0$ with $|\langle v, w\rangle| > \epsilon$. This in particular means that $\|w\| > 0$ so it is legitimate to define $\delta = \epsilon/\|w\| > 0$. Using the Cauchy-Schwartz inequality we find that $\|u\| < \delta$ we have $|\langle u, w\rangle| < \epsilon$ so $\langle v + u, w\rangle \neq 0$. This means that v is not a closure point of $W^\perp$, as required.

Proposition 19.12. [prop-hilbert-closest]

Let V be a Hilbert space, let $C \subseteq V$ be a closed convex subset, and let x be a point of V. Write $d(x, C) = \inf\{d(x, y) \;:\; y \in C\}$. Then there is a unique point $y \in C$ with $d(x, y) = d(x, C)$.

Proof. Put $r = d(x, C)$. For each $\epsilon \geq 0$ put $C_\epsilon = \{y \in C \;:\; d(x, y)^2 \leq r^2 + \epsilon\}$. By the definition of r we have $C_\epsilon \neq \emptyset$ for $\epsilon > 0$, and the claim is that C_0 consists of a single point.

We first claim that for $y, z \in C_\epsilon$ we have $d(y, z)^2 \leq 4\epsilon$. Indeed, as C is convex we see that the point $u = (y + z)/2$ lies in C, so $d(x, u)^2 \geq r^2$. One can also check by expanding everything out that

$$\langle y - z, y - z\rangle = 2\langle y - x, y - x\rangle + 2\langle z - x, z - x\rangle - 4\langle u - x, u - x\rangle,$$

or equivalently

$$d(y, z)^2 = 2d(y, x)^2 + 2d(z, x)^2 - 4d(u, x)^2 \leq 2(r^2 + \epsilon) + 2(r^2 + \epsilon) - 4r^2 = 4\epsilon$$

as claimed. This implies that C_0 has at most one point. We have also remarked that $C_\epsilon \neq \emptyset$ for $\epsilon > 0$, so we can choose $y_n \in C_{2^{-n}}$ for all $n \geq 0$. The above argument shows that for $i, j \geq n$ we have $d(y_i, y_j) \leq \sqrt{4.2^{-n}} = 2^{1-n/2}$, so the sequence $(y_i)_{i \geq 0}$ is Cauchy. As V is complete it follows that the sequence converges to some point $y \in V$. As C is closed we have $y \in C$, and it is clear that $d(x, y) = \lim_{n\to\infty} d(x, y_n) = r$ so $y \in C_0$ as required. □

Corollary 19.13. [prop-orthogonal-splitting]

Let V be a Hilbert space, and let W be a closed vector subspace. Then $V = W \oplus W^\perp$ and $W^{\perp\perp} = W$.

Proof. First, if $x \in W \cap W^\perp$ then we must have $\langle x, x\rangle = 0$ so $x = 0$. Thus, it will be enough to prove that $V = W + W^\perp$. Let x be an arbitrary element of V. As W is a vector subspace it is certainly convex, so there is a unique element $y \in W$ with $d(x, y) = d(x, W)$. We claim that the element $z = x - y$ lies in $W^\perp$. To see this, consider an element $w \in W$ and a number $t \in \mathbb{R}$. We have $y - tw \in W$ so $d(x, y - tw)^2 \geq d(x, W)^2 = d(x, y)^2$, or equivalently $\langle z + tw, z + tw\rangle \geq \langle z, z\rangle$ or $\|z\|^2 t^2 + \langle z, w\rangle t \geq 0$. As this holds for all t, it is easy to see that we must have $\langle z, w\rangle = 0$. (Consider the case $t = -\langle z, w\rangle/(2\|z\|^2)$.) We now have $x = y + z \in W + W^\perp$ as required.

Next, it is tautological that $W \leq W^{\perp\perp}$. In the opposite direction, suppose we have $x \in W^{\perp\perp} \leq V$. By the above we can write $x = y + z$ for some $y \in W$ and $z \in W^\perp$. By assumption we have $\langle x, W^\perp\rangle = 0$, so in particular we have $\langle x, z\rangle = 0$. On the other hand, we have $y \in W$ and $z \in W^\perp$, so when we expand out $\langle x, z\rangle = \langle y + z, z\rangle$ we just get $\|z\|^2$. As $\langle x, z\rangle = 0$ we conclude that $z = 0$, so $x = y \in W$. This proves that $W^{\perp\perp} \leq W$ as required. □

Definition 19.14. [defn-hilbert-projector]

Let V be a Hilbert space, and let W be a closed vector subspace. We write π_W for the linear map $V \to V$ given by $\pi_V(y + z) = y$ for all $y \in W$ and $z \in W^\perp$.

Corollary 19.15 (The Riesz Representation Theorem). [prop-riesz]

Let V be a Hilbert space over $\mathbb{R}$. Then there is an isometric isomorphism $\lambda\colon V \to V^*$ given by $\lambda(a)(v) = \langle v, a\rangle$. More explicitly, for any continuous linear map $\phi\colon V \to \mathbb{R}$ there is a unique element $a \in V$ such that $\phi(v) = \langle v, a\rangle$ for all $v \in V$; moreover, we have $\|\phi\|_{op} = \|a\|$.

Proof. First, the Cauchy-Schwartz inequality tells us that $\|\lambda(a)(v)\| \leq \|a\|\|v\|$ for all v, and that this bound is attained when $v = a$. It follows that $\lambda(a) \in V^*$ with $\|\lambda(a)\|_{\mathrm{op}} = \|a\|$. It is also clear that the map $\lambda\colon V \to V^*$ is linear, and as $\|\lambda(a)\|_{\mathrm{op}} = \|a\|$ we see that it is also injective. The main point is to prove that it is surjective. It is clear that zero is in the image, so consider instead a nonzero element $\phi \in V^*$. Put $W = \ker(\phi) = \phi^{-1}\{0\}$. As ϕ is linear and continuous, we see that W is a closed vector subspace, so $V = W \oplus W^\perp$. As $\phi \neq 0$ we have $V \neq W$ and so $W^\perp \neq 0$. We can thus choose $b \in W^\perp$ with $\|b\| = 1$. Put $t = \phi(b) \in \mathbb{R} \setminus \{0\}$. Note that for any $z \in W^\perp$ we have $\phi(z - t^{-1}\phi(z)b) = 0$ so $z - t^{-1}\phi(z)b \in W \cap W^\perp = 0$ so $z = t^{-1}\phi(z)b$. This means that $W^\perp = \mathbb{R}b$. Next, put $\psi = \phi - \lambda(tb)$; it will suffice to show that $\psi = 0$. Now $V = W \oplus W^\perp = W \oplus \mathbb{R}b$ so it will be enough to prove that $\psi(W) = 0$ and $\psi(b) = 0$. For the first, we note that $\phi(W) = 0$ and that $tb \in W^\perp$ so $\langle W, tb\rangle = 0$ so $\psi(W) = 0$. For the second, we have $\phi(b) = t$ and $\langle b, tb\rangle = t\langle b, b\rangle = t$ so $\psi(b) = 0$. □

Remark 19.16. [rem-riesz-complex]

The Riesz Representation Theorem is also valid over $\mathbb{C}$, but with one small wrinkle. We can still define $\lambda\colon V \to V^*$ by $\lambda(a)(v) = \langle v, a\rangle$, and this is $\mathbb{R}$-linear and an isometric isomorphism, but it is not $\mathbb{C}$-linear. Instead, it is conjugate-linear, in the sense that $\lambda(ta) = \overline{t}\lambda(a)$ for all $t \in \mathbb{C}$ and $a \in V$.

Definition 19.17. [defn-linear-adjoint]

Let V and W be vector spaces over $\mathbb{R}$ equipped with inner products, and let $f\colon V \to W$ and $g\colon W \to V$ be linear maps. We say that g is *adjoint* to f if we have $\langle f(v), w\rangle = \langle v, f(w)\rangle$ for all $v \in V$ and $w \in W$. Similarly, if V and W are complex vector spaces equipped with hermitian products, and $f\colon V \to W$ and $g\colon W \to V$ are $\mathbb{C}$-linear maps, we say that g is *adjoint* to f if $\langle f(v), w\rangle = \langle v, g(w)\rangle$ for all $v \in V$ and $w \in W$. By the proposition below, there is at most one such g, so we can legitimately denote it by f^*.

Proposition 19.18. [prop-adjoint-unique]

Given a linear map $f\colon V \to W$, there is at most one adjoint $g\colon W \to V$. If f is continuous and V is a Hilbert space, then there is exactly one adjoint, and it is also continuous, with $\|g\|_{op} = \|f\|_{op}$.

Proof. First, suppose that g_0 and g_1 are both adjoint to f. Put $h(w) = g_0(w) - g_1(w)$, so for all $v \in V$ and $w \in W$ we have

$$\langle v, h(w)\rangle = \langle v, g_0(w)\rangle - \langle v, g_1(w)\rangle = \langle f(v), w\rangle - \langle f(v), w\rangle = 0.$$

In particular, this holds for $v = h(w)$, so we have $\|h(w)\|^2 = \langle h(w), h(w)\rangle = 0$, so $h(w) = 0$, so $g_0 = g_1$.

Now suppose that f is continuous and V is a Hilbert space. Let $\mathbb{K}$ denote $\mathbb{R}$ or $\mathbb{C}$ as appropriate. For any $w \in W$ we have a linear map $\phi_w\colon V \to \mathbb{K}$ given by $\phi_w(v) = \langle f(v), w\rangle$. This is continuous with $\|\phi_w\|_{\mathrm{op}} \leq \|f\|_{\mathrm{op}}\|w\|$, because $\|\phi_w(v)\| \leq \|f(v)\|\|w\| \leq \|f\|_{\mathrm{op}}\|v\|\|w\|$. By the Riesz representation theorem, there is a unique element $g(w) \in V$ such that $\phi_w(v) = \langle v, g(w)\rangle$ for all $v \in V$. If we have elements $w_0, w_1 \in W$ and scalars $t_0, t_1 \in \mathbb{K}$ we note that

$$\langle v, t_0 g(w_0) + t_1 g(w_1)\rangle = \overline{t_0}\langle v, g(w_0)\rangle + \overline{t_1}\langle v, g(w_1)\rangle = \overline{t_0}\langle f(v), w_0\rangle + \overline{t_1}\langle f(v), w_1\rangle = \langle f(v), t_0 w_0 + t_1 w_1\rangle,$$

so we must have $t_0 g(w_0) + t_1 g(w_1) = g(t_0 w_0 + t_1 w_1)$. Thus, the map $g\colon W \to V$ is linear. Moreover, as the Riesz map λ is an isometry we have $\|g(w)\| = \|\phi_w\|_{\mathrm{op}} \leq \|f\|_{\mathrm{op}}\|w\|$. It follows that g is continuous with $\|g\|_{\mathrm{op}} \leq \|f\|_{\mathrm{op}}$. On the other hand, the relationship between f and g is symmetrical, so we also have $\|f\|_{\mathrm{op}} \leq \|g\|_{\mathrm{op}}$. □

Example 19.19. [eg-matrix-adjoint]

Let A be an $m \times n$ matrix, and define $m_A\colon \mathbb{R}^n \to \mathbb{R}^m$ by $m_A(v) = Av$, so $m_A(v)_j = \sum_i A_{ij} v_j$. If we interpret $\mathbb{R}^n$ as the space of column vectors of length n, then the standard inner product is just $\langle u, v\rangle = u^T v$. From this perspective it is clear that $\langle Au, v\rangle = \langle u, A^T v\rangle$, which shows that $m_A^* = m_{A^T}$. In the analogous complex case we have $\langle u, v\rangle = u^\dagger v$ and $m_A^* = m_{A^\dagger}$.

Definition 19.20. [defn-orthonormal]

Let V be a vector space over $\mathbb{R}$ equipped with an inner product, or a vector space over $\mathbb{C}$ equipped with a hermitian product. An *orthonormal sequence* in V means a finite or infinite sequence $(v_0, v_1, \dots)$ such that $\langle v_i, v_j \rangle = 0$ for all $i \neq j$, and $\|v_i\| = 1$ for all i. An *orthonormal basis* is an orthonormal sequence that is also a basis.

Remark 19.21. [rem-orthonormal-complex]

Suppose that V is a complex vector space, and that the sequence $(v_0, v_1, \dots)$ is orthonormal with respect to a hermitian product $\langle \cdot, \cdot \rangle$. One can then check that the sequence $(v_0, iv_0, v_1, iv_1, \dots)$ is orthonormal with respect to the inner product $\langle x, y \rangle = \mathrm{Re}(\langle x, y \rangle)$.

Proposition 19.22. [prop-gram-schmidt]

Let V be a vector space over $\mathbb{R}$ equipped with an inner product, or a vector space over $\mathbb{C}$ equipped with a hermitian product, and suppose that V has finite dimension. Then V admits an orthonormal basis.

Proof. This is is just the well-known Gram-Schmidt procedure. As V has finite dimension, we can choose a basis, say $u_0, \dots, u_{n-1}$. Let U_i be the span of $\{u_0, \dots, u_{i-1}\}$. We define $v_0, \dots, v_{n-1}$ recursively by

$$v_i = u_i - \sum_{j<i} \frac{\langle u_i, v_j \rangle}{\langle v_j, v_j \rangle} v_j.$$

(In particular, we start with $v_0 = u_0$.) One can check by induction that $(v_0, \dots, v_{i-1})$ is a basis for U_i (so in particular the v_j are nonzero) and $\langle v_i, v_j \rangle = 0$ for $j < i$. We then put $w_i = v_i / \|v_i\|$ and we find that $(w_0, \dots, w_{n-1})$ is the required orthonormal basis for V. □

Proposition 19.23. [prop-projector-formula]

Let $(v_0, \dots, v_{n-1})$ be a finite orthonormal sequence in V, and let W be the subspace spanned by $v_0, \dots, v_n$. Then W is closed in V, and the projector $\pi_W \colon V = W \oplus W^\perp \to W$ (as in Definition 19.14) is given by

$$\pi_W(x) = \sum_{i=0}^{n-1} \langle x, v_i \rangle v_i.$$

We thus have

$$\|x\|^2 = \|\pi_{W^\perp}(x)\|^2 + \sum_{i=0}^{n-1} |\langle x, v_i \rangle|^2 \geq \sum_{i=0}^{n-1} |\langle x, v_i \rangle|^2.$$

Proof. First note that W can be considered as an inner product space in its own right, and it is complete by Proposition 19.7, so it is closed in V.

Now consider an element $x \in V$, and put $t_i = \langle x, v_i \rangle$ and $y = \sum_{i=0}^{n-1} t_i v_i \in W$ and $z = x - y$. As the v_k are orthonormal, the only nonzero term in $\langle y, v_j \rangle$ is $\langle t_j v_j, v_j \rangle = t_j$. This implies that $\langle z, v_j \rangle = \langle x, v_j \rangle - \langle y, v_j \rangle = t_j - t_j = 0$. As the elements v_j span W we deduce that $z \in W^\perp$. As $x = y + z$ with $y \in W$ and $z \in W^\perp$ we see that $\pi_W(x) = y$ and $\pi_{W^\perp}(x) = z$. Next, as the elements v_k are orthogonal to each other and to z, we get no cross-terms when we expand out the inner product $\langle x, x \rangle = \langle z + \sum_i t_i v_i, z + \sum_i t_i v_i \rangle$. Moreover, most of the remaining terms have the form $\langle t_i v_i, t_i v_i \rangle = t_i \overline{t_i} \langle v_i, v_i \rangle = |t_i|^2$. The conclusion is that $\|x\|^2 = \|z\|^2 + \sum_i |t_i|^2$, as claimed. □

Corollary 19.24. [cor-parseval]

Let $(v_0, v_1, v_2, \dots)$ be an infinite orthonormal sequence in V. Let $W[n]$ be the span of $v_0, \dots, v_{n-1}$, and let W be the closure of $\bigcup_n W[n]$. Then W is a vector subspace of V. Moreover, for any $x \in V$ the sequence $(\pi_{W[n]}(x))_{n \in \mathbb{N}}$ converges to $\pi_W(x)$, and we have

$$\|x\|^2 = \|\pi_{W^\perp}(x)\|^2 + \sum_{i=0}^{\infty} |\langle x, v_i \rangle|^2 \geq \sum_{i=0}^{\infty} |\langle x, v_i \rangle|^2.$$

(This is known as Parseval's inequality.*)*

Proof. First, put $W' = \bigcup_n W[n] = \text{span}\{v_i \;:\; i \in \mathbb{N}\}$, so W is the closure of W'. Let $\mathbb{K}$ be $\mathbb{R}$ or $\mathbb{C}$ as appropriate, and suppose we have $s, t \in \mathbb{K}$. We can define $f\colon V \times V \to V$ by $f(a,b) = sa + tb$. This is easily seen to be continuous, so $f^{-1}(W)$ is closed in $V \times V$. It is also clear that $W' \times W' \subseteq f^{-1}(W') \subseteq f^{-1}(W)$, so $f^{-1}(W)$ contains the set $\overline{W' \times W'} = W \times W$. Using this we see that W is a vector subspace of V.

Now consider an element $x \in V$. Put $t_i = \langle x, v_i\rangle$ and $y_n = \sum_{i<n} t_i v_i = \pi_{W[n]}(x)$ and $z_n = x - y_n = \pi_{W[n]^\perp}(x)$. From the proposition we know that $\|y_n\|^2 = \sum_{i<n} |t_i|^2 \leq \|y\|^2$. It follows that the supremum $r = \sup\{\|y_n\| \;:\; n \in \mathbb{N}\}$ is at most $\|y\|$. Suppose we are given $\epsilon > 0$. There must then exist n such that $\|y_n\|^2 > r^2 - \epsilon$. Now for $m \geq n$ we have

$$\|y_m - y_n\|^2 = \|\sum_{n \leq i < m} t_i v_i\|^2 = \sum_{n \leq i < m} |t_i|^2 = \|y_m\|^2 - \|y_n\|^2 \leq r^2 - (r^2 - \epsilon) = \epsilon.$$

Using this, we see that the sequence $(y_n)_{n\in\mathbb{N}}$ is Cauchy. It therefore converges to some point $y \in V$. In fact, as each y_n lies in $W[n] \leq W$ and W is closed we see that $y \in W$. We also see that the element $z = x - y$ is the limit of the elements z_n. Moreover, for $n \geq m$ we see that $z_n \in W[n]^perp \leq W[m]^\perp$. As $W[m]^\perp$ is closed and $z_n \to z$ we see that $z \in W[m]^\perp$. Equivalently, we have $W[m] \leq (\mathbb{K}z)^\perp$. As this holds for all m and $(\mathbb{K}z)^\perp$ is closed we deduce that $W' \leq (\mathbb{K}z)^\perp$ and then that $W \leq (\mathbb{K}z)^\perp$, or equivalently $z \in W^\perp$. As $x = y + z$ with $y \in W$ and $z \in W^\perp$ we have $y = \pi_W(x)$ and $z \in \pi_{W^\perp}(x)$. As $y_n \to y$ with $\|y_n\|^2 = \sum_{i<n} |t_i|^2$ we also have $\|y\|^2 = \sum_{i=0}^\infty |t_i|^2$. We can also expand out $\langle y+z, y+z\rangle$ to see that

$$\|x\|^2 = \|y\|^2 + \|z\|^2 \geq \|y\|^2 = \sum_i |t_i|^2,$$

which is Parseval's inequality. □

19.3. Finite-dimensional spectral theory. We now consider some theory of eigenvalues and eigenvectors for endomorphisms of finite-dimensional complex vector spaces. There is also a rich theory for the infinite-dimensional case, but that would take us further into the realms of functional analysis than we wish to go. In the finite-dimensional case we avoid many analytic technicalities but there are a number of interesting algebraic and topological points that not as well-known as they might be.

Definition 19.25. [`defn-euclidean-space`]
A *euclidean space* is a finite-dimensional vector space over $\mathbb{R}$ equipped with an inner product (or equivalently, a finite-dimensional real Hilbert space). A *hermitian space* is a finite-dimensional vector space over $\mathbb{C}$ equipped with a hermitian product (or equivalently, a finite-dimensional complex Hilbert space).

Definition 19.26. Let V and W be Hermitian spaces. We define a Hermitian product on $\text{Hom}(V, W)$ by $\langle\alpha, \beta\rangle = \text{trace}(\beta^*\alpha)$. (If we choose orthonormal bases for V and W and let A and B be the matrices corresponding to α and β, we find that $\langle\alpha, \beta\rangle = \sum_{i,j} A_{ij}\overline{B_{ij}}$, which makes it clear that this is indeed a Hermitian product.) We write $\|\alpha\|_2 = \sqrt{\text{trace}(\alpha^*\alpha)}$ for the corresponding norm.

Recall that we also have another norm

$$\|\alpha\|_{\text{op}} = \max\{\|\alpha(v)\| \;:\; \|v\| \leq 1\}.$$

(We previously wrote this as a supremum rather than a maximum, but we are now working in a finite-dimensional context where the ball $B(V)$ is compact, so the supremum is attained.)

Example 19.27. Let $\alpha \in \text{End}(\mathbb{C}^2)$ be given by the matrix $\begin{bmatrix} 1 & 2a \\ 0 & 1 \end{bmatrix}$ with $a \geq 0$. Then $\|\alpha\|_2 = 2 + 4a^2$, and a nice exercise in calculus shows that

$$\|\alpha\|_\infty = \max\{(\cos(t) + 2a\sin(t))^2 + \sin(t)^2 \;:\; t \in \mathbb{R}\} = \frac{\sqrt{a^2+1} + a}{\sqrt{a^2+1} - a}.$$

In fact, we have

$$(\cos(t) + 2a\sin(t))^2 + \sin(t)^2 = \tfrac{1}{2}(r + r^{-1}) - \tfrac{1}{2}(r - r^{-1})\cos(2t + \arctan(1/a)),$$

where $r = (\sqrt{a^2+1} + a)/(\sqrt{a^2+1} - a)$.

We know from Proposition 10.34 that there exist constants k, K (depending only on the dimensions of V and W) such that $k\|\alpha\|_{\mathrm{op}} \leq \|\alpha\|_2 \leq K\|\alpha\|_{\mathrm{op}}$ for all α. We can give specific constants as follows:

Proposition 19.28. *For any $\alpha\colon V \to W$ we have $\|\alpha\|_\infty \leq \|\alpha\|_2 \leq \sqrt{\dim(V)}\|\alpha\|_\infty$.*

Proof. Choose a unit vector $v_1 \in V$ with $\|\alpha(v_1)\| = \|\alpha\|_\infty$, and extend it to give an orthonormal basis $v_1, \dots, v_n$ for V. Choose an orthonormal basis $w_1, \dots, w_m$ for W, and put $A_{ij} = \langle \alpha(v_i), w_j \rangle$, so $\alpha(v_i) = \sum_j A_{ij} w_j$. It follows that $\alpha^*(w_j) = \sum_i \overline{A_{ij}} v_i$ and thus that $\|\alpha\|_2^2 = \sum_{i,j} |A_{ij}|^2$. However, we also find that $\|\alpha(v_i)\|^2 = \sum_j |A_{ij}|^2$ so $\|\alpha\|_2^2 = \sum_{i=1}^n \|\alpha(v_i)\|^2$. Here $\|\alpha(v_i)\|^2 \leq \|\alpha\|_{\mathrm{op}}^2$ for all i, with equality when $i = 1$. It follows that $\|\alpha\|_{\mathrm{op}}^2 \leq \|\alpha\|_2^2 \leq n\|\alpha\|_{\mathrm{op}}^2$ as claimed. $\square$

Proposition 19.29. [`prop-linear-isometry`]
Let $\mathbb{K}$ be $\mathbb{R}$ or $\mathbb{C}$, and let $\alpha\colon V \to W$ be a $\mathbb{K}$-linear map between finite-dimensional Hilbert spaces over $\mathbb{K}$. Then α is an isometry if and only if $\alpha^*\alpha = 1_V$. If so, then α is an isometric isomorphism iff $\alpha\alpha^* = 1_W$ iff $\dim(V) = \dim(W)$.

Proof. First suppose that $\alpha^*\alpha = 1$. For $v \in V$ we then have $\|\alpha(v)\|^2 = \langle \alpha(v), \alpha(v) \rangle = \langle v, \alpha^*\alpha(v) \rangle = \langle v, v \rangle = \|v\|^2$, so $\|\alpha(v)\| = \|v\|$. This gives $d(\alpha(v), \alpha(v')) = \|\alpha(v - v')\| = \|v - v'\| = d(v, v')$, so α is an isometry.

Conversely, suppose that α is an isometry, so $\|\alpha(v)\|^2 = \|v\|^2$ for all $v \in V$. Observe that

$$\|v + v'\|^2 - \|v - v'\|^2 = \langle v, v' \rangle + \langle v', v \rangle = 2\mathrm{Re}(\langle v, v' \rangle).$$

Similarly, we have

$$\|\alpha(v) + \alpha(v')\|^2 - \|\alpha(v) - \alpha(v')\|^2 = 2\mathrm{Re}(\langle \alpha(v), \alpha(v') \rangle) = 2\mathrm{Re}(\langle v, \alpha^*\alpha(v') \rangle).$$

By comparing these, we see that $\mathrm{Re}(\langle v, v' \rangle) = \mathrm{Re}(\langle v, \alpha^*\alpha(v') \rangle)$ for all v and v', or equivalently $\mathrm{Re}(\langle v, v' - \alpha^*\alpha(v) \rangle) = 0$. If we take $v = v' - \alpha^*\alpha(v')$ then there is no imaginary part and we see that $\|v' - \alpha^*\alpha(v')\|^2 = 0$, so $v' = \alpha^*\alpha(v')$. As v' was arbitrary we have $\alpha^*\alpha = 1_V$ as claimed.

For the rest of the proof we assume that $\alpha^*\alpha = 1$. It is then standard linear algebra that α is an isomorphism iff $\dim(V) = \dim(W)$, and if so, then α^* must be a right inverse as well as a left inverse. $\square$

Definition 19.30. [`defn-isometry-spaces`]
If V and W are euclidean spaces, we write $\mathcal{L}(V, W)$ for the space of $\mathbb{R}$-linear isometries $\alpha\colon V \to W$ (considered as a topological subspace of the finite-dimensional vector space $\mathrm{Hom}(V, W)$). Similarly, if V and W are hermitian spaces, we write $\mathcal{H}(V, W)$ for the space of $\mathbb{C}$-linear isometries $\alpha\colon V \to W$.

Proposition 19.31. *The spaces $\mathcal{L}(V, W)$ and $\mathcal{H}(V, W)$ are compact Hausdorff.*

Proof. The description $\mathcal{H}(V, W) = \{\alpha \in \mathrm{Hom}_{\mathbb{C}}(V, W) \ : \ \alpha^*\alpha = 1\}$ makes it clear that $\mathcal{H}(V, W)$ is a closed subspace of $\mathrm{Hom}(V, W)$, so it is Hausdorff. For $\alpha \in \mathcal{H}(V, W)$ we also have $\|\alpha\|_2 = \sqrt{\mathrm{trace}(\alpha^*\alpha)} = \sqrt{\dim(V)}$, so $\mathcal{H}(V, W)$ is bounded and thus compact. The argument for $\mathcal{L}(V, W)$ is essentially the same. $\square$

Definition 19.32. Let V be a Hermitian space, and let α be an endomorphism of V. We put

$$\begin{aligned} \mathrm{spec}(\alpha) &= \{ \text{ eigenvalues of } \alpha \} = \{\lambda \in \mathbb{C} \ : \ \alpha - \lambda \text{ is not invertible } \} \\ \rho(\alpha) &= \text{ the spectral radius of } \alpha = \max\{|\lambda| \ : \ \lambda \in \mathrm{spec}(\alpha)\}. \end{aligned}$$

Lemma 19.33. *For any α we have $\rho(\alpha) \leq \|\alpha\|_\infty$.*

Proof. Just choose an eigenvalue λ with $|\lambda| = \rho(\alpha)$, and then a unit vector v with $\alpha(v) = \lambda v$; then $\rho(\alpha) = |\lambda| = \|\alpha(v)\| \leq \|\alpha\|_\infty$. $\square$

Remark 19.34. Let α be a nonzero map represented by an upper triangular matrix with zeros on the diagonal. Then $\mathrm{spec}(\alpha) = \{0\}$ so $\rho(\alpha) = 0$ but $\|\alpha\|_2,\ \|\alpha\|_\infty > 0$. Thus, neither norm can be bounded above by a constant multiple of the spectral radius.

Definition 19.35. We say that an endomorphism $\alpha\colon V \to V$ is *normal* if $\alpha^*\alpha = \alpha\alpha^*$.

Remark 19.36. Recall that

- α is unitary if it is invertible, with $\alpha^* = \alpha^{-1}$
- α is hermitian (or self-adjoint) if $\alpha^* = \alpha$
- α is antihermitian (or anti self-adjoint) if $\alpha^* = -\alpha$.

It is clear that α is normal in all these cases.

Proposition 19.37. *α is normal if and only if there exists an orthonormal basis for V consisting of eigenvectors for α.*

Proof. Suppose that $v_1, \dots, v_n$ is an orthonormal basis for V with $\alpha(v_i) = \lambda_i v_i$. We then define $\beta\colon V \to V$ by $\beta(v_i) = \overline{\lambda_i} v_i$, and check that $\langle \alpha(v_i), v_j \rangle = \delta_{ij}\lambda_i = \langle v_i, \beta(v_j)\rangle$, so $\beta = \alpha^*$. We also see that $\alpha\beta = \beta\alpha$, so α is normal.

Conversely, suppose that α is normal. Let λ be an eigenvalue, and put $W = \ker(\alpha - \lambda) \neq 0$. For $w \in W$ we have $\alpha\alpha^*(w) = \alpha^*\alpha(w) = \lambda\alpha^*(w)$, so $\alpha^*(w) \in W$. It follows that for $x \in W^\perp$ we have $\langle \alpha(x), W\rangle = \langle x, \alpha^*(W)\rangle \leq \langle x, W\rangle = 0$, so $\alpha^*(x) \in W^\perp$. This means that α preserves the splitting $V = W \oplus W^\perp$, and so restricts to give a normal operator on $W^\perp$. By induction on dimension we can choose an orthonormal basis of $W^\perp$ consisting of eigenvectors for α, and we can combine this with an arbitrary orthonormal basis of W to get the required basis of V. □

Corollary 19.38. *Let α be normal, with eigenvalues $\lambda_1, \dots, \lambda_n$ (repeated according to multiplicity). Then*

$$\|\alpha\|_2 = \left(\sum_i |\lambda_i|^2\right)^{1/2}$$
$$\|\alpha\|_\infty = \rho(\alpha) = \max\{|\lambda_1|, \dots, |\lambda_n|\}.$$

Proof. We can choose an orthonormal basis $v_1, \dots, v_n$ with $\alpha(v_i) = \lambda_i v_i$. The formula for $\|\alpha\|_2$ follows immediately. Now consider a vector $v = \sum_i x_i v_i \in V$. We have $\|v\|^2 = \sum_i |x_i|^2$ and

$$\|\alpha(v)\|^2 = \sum_i |\lambda_i|^2 |x_i|^2 \leq \sum_i \rho(\alpha)^2 |x_i|^2 = \rho(\alpha)^2 \|v\|^2,$$

so $\|\alpha\|_\infty^2 \leq \rho(\alpha)^2$. The reverse inequality is given by Lemma 19.33. □

Proposition 19.39. *Let $\alpha\colon V \to V$ be self-adjoint, and consider the set*

$$R(\alpha) = \{\langle v, \alpha(v)\rangle \ : \ v \in S(V)\}$$

(known as the numerical range *of α). Then all eigenvalues of α are real, and $R(\alpha) = [\rho_-(\alpha), \rho_+(\alpha)]$, where $\rho_-(\alpha)$ and $\rho_+(\alpha)$ are respectively the smallest and largest eigenvalues.*

Proof. Choose an orthonormal basis $v_1, \dots, v_n$ with $\alpha(v_i) = \lambda_i v_i$ say. We then have

$$\overline{\lambda_i} = \langle v_i, \lambda_i v_i\rangle = \langle v_i, \alpha(v_i)\rangle = \langle \alpha^*(v_i), v_i\rangle = \langle \alpha(v_i), v_i\rangle = \lambda_i,$$

so λ_i is real as claimed. We can order the basis so that $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$, so that $\rho_-(\alpha) = \lambda_1$ and $\rho_+(\alpha) = \lambda_n$. Now consider an element $v = \sum_i x_i v_i \in V$. We have

$$\langle v, \alpha(v)\rangle = \langle \sum_i x_i v_i, \sum_j \lambda_j x_j v_j\rangle = \sum_i \lambda_i |x_i|^2 \leq \sum_i \lambda_n |x_i|^2 = \lambda_n \|v\|^2.$$

It follows that $R(\alpha) \subseteq (-\infty, \lambda_n]$. By a similar argument we have $R(\alpha) \subseteq [\lambda_1, \infty)$, so $R(\alpha) \subseteq [\lambda_1, \lambda_n]$. By taking $v = \cos(\theta)v_1 + \sin(\theta)v_n$ for $0 \leq \theta \leq \pi/2$ we see that $[\lambda_1, \lambda_n] \subseteq R(\alpha)$, which completes the proof. □

Definition 19.40. We write $w(V)$ for the space of self-adjoint endomorphisms of V. We say that $\alpha \in w(V)$ is *positive* if $\langle v, \alpha(v)\rangle > 0$ for all $v \neq 0$, or equivalently $\rho_-(\alpha) > 0$, or equivalently $\operatorname{spec}(\alpha) \subset (0, \infty)$. We write $w^+(V)$ for the space of positive operators. We also define *nonnegative* operators in the analogous way.

Remark 19.41. For a self-adjoint operator α we have $\|\alpha\|_\infty = \rho(\alpha) = \max(|\rho_-(\alpha)|, |\rho_+(\alpha)|)$. For a positive operator this reduces to $\|\alpha\|_\infty = \rho(\alpha) = \rho_+(\alpha)$.

Corollary 19.42. *For any linear map $\alpha\colon V \to W$, we have $\|\alpha\|_\infty = \|\alpha^*\alpha\|_\infty^{1/2} = \sqrt{\rho(\alpha^*\alpha)}$.*

PROOF. Apply Proposition 19.39 to the nonnegative operator $\alpha^*\alpha$, noting that $\langle v, \alpha^*\alpha(v)\rangle = \langle \alpha(v), \alpha(v)\rangle = \|\alpha(v)\|^2$. □

19.4. Compact convex sets. Let V be a finite-dimensional normed vector space over $\mathbb{R}$. Recall that a set $C \subseteq V$ is said to be *convex* if $(1-t)x + ty \in C$ whenever $x, y \in C$ and $t \in [0,1]$. It is straightforward to check that the ball $B(V) = \{x \in V : \|x\| \leq 1\}$ is compact and convex, and using Proposition 10.34 we see that it is homeomorphic to the standard ball $B^{\dim(V)}$. We next explain a generalisation of this.

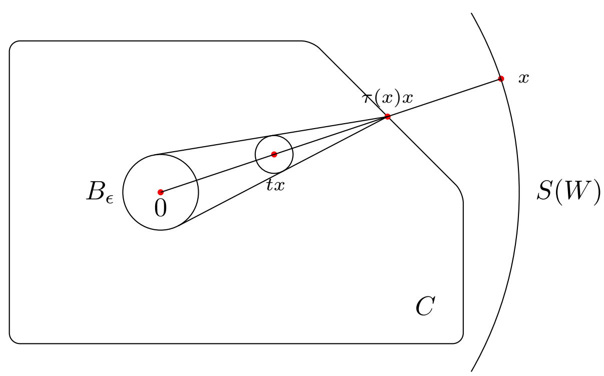
PROPOSITION 19.43. [prop-compact-convex-i]
Let C be a compact convex subset of V such that $0 \in \mathrm{int}(C)$. Then C is homeomorphic to $B(V)$ and thus to the standard ball $B^{\dim(V)}$. More precisely, there is a unique homeomorphism $f\colon C \to B(V)$ satisfying $f(tx) = t\,f(x)$ for all $x \in C$ and $t \in [0,1]$.

PROOF. First, as $0 \in \mathrm{int}(C)$ we can choose some $\epsilon > 0$ such that $OB_\epsilon(0) \subseteq C$. Moreover, as C is compact the function $x \mapsto \|x\|$ must be bounded on C, so we can choose $R > 0$ with $C \subseteq B_R(0)$.

Now put $\partial(C) = C \setminus \mathrm{int}(C)$ as usual. This is closed in the compact set C, so it is again compact. As $0 \in \mathrm{int}(C)$ we see that $\partial(C) \subseteq V \setminus \{0\}$, so we can define a continuous map $f_1\colon \partial(C) \to S(V)$ by $f_1(x) = x/\|x\|$.

We next claim that f_1 is surjective. To see this, consider a point $x \in S(V)$, define a continuous map $m_x\colon [0,\infty) \to V$ by $m_x(t) = tx$, and put $T_x = m_x^{-1}(C)$. As $\|m_x(t)\| = t$ and $OB_\epsilon(0) \subseteq C \subseteq B_R(0)$ we have $[0,\epsilon) \subseteq T_x \subseteq [0,R]$. As C is convex, the same is true of T_x. It now follows easily that $T_x = [0,\tau(x)]$ for some number $\tau(x) > 0$. Put $g_1(x) = \tau(x)x \in C$. By the definition of $\tau(x)$ we have $g_1(x) + \epsilon x \notin C$ for all $\epsilon > 0$, and this means that $g_1(x) \notin \mathrm{int}(C)$. We have thus defined a map $g_1\colon S(V) \to \partial(C)$ (not obviously continuous) and it is clear that $f_1 g_1 = 1$ so f_1 is surjective.

Next, for $0 \leq t < 1$ we note that the set $tg_1(x) + (1-t)OB_\epsilon(0)$ is an open neighbourhood of $tg_1(x)$ which is contained in C by convexity, so $tg_1(x) \in \mathrm{int}(C)$. This means that $g_1(x)$ is the unique positive multiple of x that lies in $\partial(C)$. It follows easily that the map $f_1\colon \partial(C) \to S(V)$ is injective as well as surjective. Moreover, $\partial(C)$ is compact and $S(V)$ is Hausdorff, so f_1 is actually a homeomorphism. It follows that g_1 is inverse to f_1 and is continuous.



Now define $g\colon B(V) \to C$ by

$$g(x) = \begin{cases} \|x\|\, g_1(x/\|x\|) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0. \end{cases}$$

To check that this is continuous, consider the map $m\colon [0,1] \times S(V) \to B(V)$ given by $m(t,x) = tx$. This is a continuous surjection of compact Hausdorff spaces, and thus a quotient map. We have $g(m(t,x)) = t\,g_1(x)$ for all t and x, so $g \circ m$ is continuous, so g is continuous. Using the fact that $T_x = [0,\tau(x)]$ we see that g is a bijection. As the source and target are compact Hausdorff, it is therefore a homeomorphism. We write f for the inverse. By construction we have $g(tx) = t\,g(x)$ when $0 \leq t \leq 1$, and it follows that f has the same property. We leave it to the reader to check that it is uniquely determined by this. □

We can generalise slightly further as follows.

PROPOSITION 19.44. [prop-compact-convex]
Let V be a finite-dimensional vector space, and let C be a nonempty compact convex subset of V. Then C is homeomorphic to B^n for some $n \leq \dim(V)$.

PROOF. Choose a point $c_0 \in C$. Let W be the linear span of all the vectors $x - c_0$ with $x \in C$. Put $X = c_0 + W$, and note that this is the same as $x + W$ for any $x \in C$. Choose $c_1, \dots, c_n \in C$ such that the vectors $c_i - c_0$ give a basis for W. Put $b = (\sum_{i=0}^n c_i)/(n+1)$, and note that this lies in C by convexity, so $X = b + W$. We can now define a homeomorphism $f \colon \mathbb{R}^n \to X$ by

$$f(t) = b + \sum_{i=1}^n t_i(c_i - c_0) = \left(\frac{1}{n+1} - \sum_{i=1}^n t_i\right) c_0 + \sum_{i=1}^n \left(\frac{1}{n+1} + t_i\right) c_i.$$

If the numbers t_i are sufficiently small, then all the coefficients $1/(n+1) - \sum_i t_i$ and $1/(n+1) + t_i$ will be positive, and their sum is one, so we have $f(t) \in C$ by convexity. It now follows that $f^{-1}(C)$ is a compact convex subset of $\mathbb{R}^n$ with 0 in the interior, so it is homeomorphic to B^n. We also have a homeomorphism $f \colon f^{-1}(C) \to C$, so $C \simeq B^n$. □

19.5. More balls and spheres. In this section we let V denote a finite-dimensional real Hilbert space. To avoid notational clutter we will generally write v^2 for $\|v\|^2$.

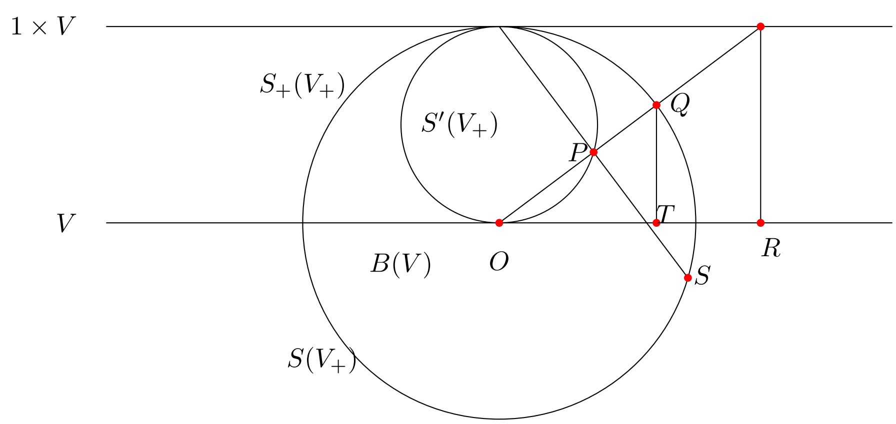
Associated to V we have a number of different spaces that are homeomorphic to S^n or B^m. It will be helpful to have a systematic catalogue of these. We put

$$\begin{aligned}
V_+ &= \mathbb{R} \oplus V \\
S^V &= V \cup \{\infty\} = \text{ the one-point compactification of } V \\
S(V_+) &= \{(t, v) \in V_+ \;:\; t^2 + v^2 = 1\} \\
S_+(V_+) &= \{(t, v) \in S(V_+) \;:\; t \geq 0\} \\
S'(V_+) &= \{(t, v) \in V_+ \;:\; (t - 1/2)^2 + v^2 = 1/4\} \\
&= \{(t, v) \in V_+ \;:\; t(1 - t) = v^2\}.
\end{aligned}$$

As in Example 5.55, we use the notation X/Y for the quotient space of X where Y is collapsed to a single point. In particular, this construction gives us spaces $S_+(V_+)/S(V)$ and $B(V)/S(V)$. We can define homeomorphisms

$$S'(V_+) \simeq S_+(V_+)/S(V) \simeq S^V \simeq S(V_+) \simeq B(V)/S(V)$$

by letting the points P, Q, R, S and T in the following diagram correspond to each other.



Formulae can be read off from the following table:

P	Q	R	S	T
$S'(V_+)$	$S_+(V_+)/S(V)$	S^V	$S(V_+)$	$B(V)/S(V)$
(p,u)	$(p,u)/\sqrt{p}$	u/p	$(2p-1,2u)$	$u/\sqrt{p}$
$q(q,v)$	(q,v)	v/q	$(2q^2-1,2qv)$	v
$(1,w)/(1+w^2)$	$(1,w)/\sqrt{1+w^2}$	w	$(w^2-1,2w)/(1+w^2)$	$w/\sqrt{1+w^2}$
$(1+s,x)/2$	$(1+s,x)/\sqrt{2(1+s)}$	$x/(1+s)$	(s,x)	$x/\sqrt{2(1+s)}$
$(1-y^2,y\sqrt{1-y^2})$	$(\sqrt{1-y^2},y)$	$y/\sqrt{1-y^2}$	$(1-2y^2,2y\sqrt{1-y^2})$	y

For example, the last line means that if we have a point $T \in B(V)/S(V)$ corresponding to an element $y \in B(V)$, then the corresponding point $P \in S'(V_+)$ is given by $P = (1-y^2, y\sqrt{1-y^2})$. We will leave all verifications to the reader. Note that the homeomorphism $S^V \simeq S(V_+)$ is stereographic projection, which we already met in Proposition 18.20.

We now give yet another model of the same homeomorphism type.

Proposition 19.45. [prop-sphere-double]
Let E denote the equivalence relation on $B(V_+)$ given by

$$E = \{(a,a) \ : \ a \in B(V_+)\} \cup \{((x,v),(-x,v)) \ : \ (x,v) \in S(V_+)\}.$$

Then there is a homeomorphism $f\colon B(V_+)/E \to S(\mathbb{R}^2 \oplus V)$ given by

$$f(x,v) = \left(\frac{2x^2+v^2-1}{\sqrt{1-v^2}}, 2x\frac{\sqrt{1-v^2-x^2}}{\sqrt{1-v^2}}, v \right)$$

(or $f(x,v) = (0,0,v)$ if $v^2 = 1$).

Remark 19.46. [rem-sphere-double]
In the case $V = \mathbb{R}$, one can check that f can be described geometrically as follows. We first use the map $(y,v) \mapsto (\sqrt{1-y^2-v^2}, y, v)$ from B^2 to the right-hand hemisphere in S^2; then we apply the map given in spherical polar coordinates by $(\theta,\phi) \mapsto (2\theta,\phi)$. The general case is obtained by reinterpreting the formulae from this case in a straightforward way.

Proof. It is straightforward algebra to check that when $v^2 \neq 1$ we have $\|f(x,v)\|^2 = 1$. Note that the last entry in $f(x,v)$ is just v so if v^2 is close to 1 then the other two entries must be small; this proves that the stated rule for $v^2 = 1$ gives a continuous extension of the formula. If $(x,v) \in S(V_+)$ then the second entry in $f(x,v)$ is zero and the first entry is clearly an even function of x; thus f respects our equivalence relation. We next show that f is surjective. Put

$$\begin{aligned} U &= \{(s,t,u) \in S(\mathbb{R}^2 \oplus V) \ : \ s < \sqrt{s^2+t^2}\} \\ &= \{(s,t,u) \in S(\mathbb{R}^2 \oplus V) \ : \ s < 0 \text{ or } t \neq 0\} \end{aligned}$$

and define $g\colon U \to \mathbb{R}^2 \oplus V$ by

$$g(s,t,v) = \left(\frac{t}{\sqrt{2}} \left(1 - \frac{s}{\sqrt{s^2+t^2}} \right)^{-1/2}, v \right).$$

If we let x be the first entry in $g(s,t,v)$ and put $r = \sqrt{s^2+t^2} = \sqrt{1-v^2}$, we find after some manipulation that $1 - x^2/r^2 = (1-s/r)/2$. From this one can derive the formulae $f(g(s,t,v)) = (s,t,v)$ and $g(f(x,v)) = (x,v)$, and after considering carefully the range of validity of these calculations we see that f and g give a homeomorphism from the interior of $B(V_+)$ to U. When $x^2 + v^2 = 1$ the formula for f reduces to $f(x,v) = (|x|, 0, v)$. It follows easily from this that the induced map $B(V_+)/E \to S(\mathbb{R}^2 \oplus V)$ is a continuous bijection from a compact space to a Hausdorff space, so it is a homeomorphism. $\square$

Remark 19.47. **Tidy this up**
I think we can also define a homeomorphism $B(V_+)/S_+(V_+) \to B(V_+)$ by

$$f(t,u) = \left(1+t-m, \sqrt{\frac{m-t}{2+m-t}}u\right) \qquad\qquad m = \sqrt{1-u^2}.$$

Note here that $t \in [-m,m]$ and $(t,u) \in S_+(V_+)$ iff $t = m$. The inverse is

$$f^{-1}(r,v) = (r-1+\sqrt{1-u^2}, u) \qquad\qquad u = \sqrt{\frac{3-r}{1-r}}v.$$

This restricts to give a homeomorphism $S(V_+)/S_+(V_+) \simeq S(V_+)$.

As well as the spheres $S(V)$ themselves, products such as $X = S(V_1) \times \cdots \times S(V_n)$ are also important examples for a variety of purposes. One interesting fact is that X admits an embedding in $\mathbb{R}^{\dim(X)+1}$, as we now explain.

Definition 19.48. [`defn-sphere-embedding`]
Let V be a euclidean space. We define

$$j(V)\colon (-1,1) \times S(V_+) \to V \times (-1,1)$$

by

$$j(V)(s,t,v) = \tfrac{1}{2}(1+s)(v,t).$$

More generally, suppose we have inner product spaces $V_1, \dots, V_n$. We define (recursively) a map

$$j(\underline{V})\colon (-1,1) \times \prod_{k\le n} S(V_{k+}) \to \left(\prod_{k\le n} V_k\right) \times (-1,1)$$

as the composite

$$(-1,1) \times \prod_{k\le n} S(V_{k+}) \xrightarrow{j(V_0,\dots,V_{n-1})\times 1} \prod_{k<n} V_k \times (-1,1) \times S(V_{n+}) \xrightarrow{1\times j(V_n)} \prod_{k\le n} V_k \times (-1,1).$$

Proposition 19.49. [`prop-sphere-embedding`]
The map $j(\underline{V})$ is an open embedding. It therefore restricts to give an embedding of $\prod_k S(V_{k_+}) \simeq \prod_k S^{V_k}$ in $(\prod_k V_k) \times \mathbb{R}$.

Proof. It will suffice to show that $j(V)$ is an open embedding. In fact, it gives a homeomorphism from $(-1,1) \times S(V_+)$ to $OB(V_+) \setminus \{0\}$, with inverse given by

$$(v,p) \mapsto (2\sqrt{v^2+p^2}-1, p/\sqrt{v^2+p^2}, v/\sqrt{v^2+p^2}).$$

□

Now consider a complex vector space V. If we have a hermitian product on V we can define $S(V)$ as before and all the above discussion still applies. It is also interesting to consider the situation where we have a perfect symmetric bilinear form $(\cdot,\cdot)\colon V \otimes V \to \mathbb{C}$ and the space $\widetilde{S}(V) = \{v \in V \;:\; (v,v) = 1\}$. In particular, if W is a real inner product space then we can consider the symmetric bilinear form on $\mathbb{C} \otimes W$ given by

$$(x+iy, u+iv) = (\langle x,u\rangle - \langle y,v\rangle) + (\langle x,v\rangle + \langle y,u\rangle)i,$$

and the space $\widetilde{S}(\mathbb{C} \otimes W)$. We also put

$$TS(W) = \{(u,v) \in S(W) \times W \;:\; \langle u,v\rangle = 0\}.$$

Proposition 19.50. *There is a homeomorphism $f\colon TS(W) \to \widetilde{S}(\mathbb{C} \otimes W)$ given by*

$$f(u,v) = \sqrt{1+v^2}\,u + iv$$
$$f^{-1}(x+iy) = (x/\|x\|, y).$$

When we have introduced the relevant concepts in Section 27 we will deduce that $\widetilde{S}(\mathbb{C}\otimes W)$ is homotopy equivalent to the sphere $S(W)$.

Proof. We have $(x+iy, x+iy) = x^2 - y^2 + 2i\langle x, y\rangle$, so $x + iy \in \widetilde{S}(\mathbb{C}\otimes W)$ if and only if $\langle x, y\rangle = 0$ and $x^2 = 1 + y^2 \geq 1$. Given this, one can check directly that the given formulae have the required effect. $\square$

20. Manifolds

Definition 20.1. [defn-manifold]
Let M be a topological space. We say that M is *locally euclidean of dimension n* if every point is contained in an open set that is homeomorphic to $\mathbb{R}^n$. Any such open set U is called a *chart domain*, and a choice of homeomorphism $f\colon U \to \mathbb{R}^n$ is called a *chart*. A *(topological) manifold of dimension n* (or *n-manifold*) is a Hausdorff, second countable space M that is locally euclidean of dimension n.

Remark 20.2. [rem-centred-chart]
Consider a chart $f\colon U \to \mathbb{R}^n$ and a point $a \in U$. We say that the chart is *centred at a* if $f(a) = 0$. For any point $a \in U$ we can define a new chart $g\colon U \to \mathbb{R}^n$ by $g(x) = f(x) - f(a)$, and this is centred at a.

Example 20.3. [eg-vector-manifold]
The most basic example is that $\mathbb{R}^n$ itself is an n-manifold. Similarly, if V is any n-dimensional real vector space (with the linear topology) then every linear isomorphism $f\colon V \to \mathbb{R}^n$ is a homeomorphism, and it follows that V is again an n-manifold. More generally, if M is an arbitrary n-manifold then it is often more natural and convenient to exhibit homeomorphisms from open subsets of M to various vector spaces that need not be explicitly identified with $\mathbb{R}^n$. We will still refer to such homeomorphisms as charts.

Proposition 20.4. *Let M be an n-dimensional manifold, and let N be an open subset of M. Then N is also an n-dimensional manifold.*

Proof. First, N is Hausdorff and second countable by Propositions 6.4 and 2.64. Now consider a point $a \in N$. As M is a manifold, there exists a chart $f\colon U \xrightarrow{\simeq} \mathbb{R}^n$ with $a \in U$. After replacing f by $f - f(a)$ if necessary, we may assume that $f(a) = 0$. Now $f(N \cap U)$ is an open subset of $\mathbb{R}^n$ containing 0, so it contains $OB_\epsilon(0)$ for some $\epsilon > 0$. Put $V = f^{-1}(OB_\epsilon(0))$, which is an open neighbourhood of a contained in $N \cap U$. Define $p\colon OB_\epsilon(0) \to \mathbb{R}^n$ by $p(x) = x/\sqrt{1 - \|x\|^2/\epsilon^2}$, and note that this is a homeomorphism with $p^{-1}(y) = y/\sqrt{1 + \|y\|^2/\epsilon^2}$. It follows that the map $g = p \circ f\colon V \to \mathbb{R}^n$ is a homeomorphism. This means that N is locally euclidean, as required. $\square$

Example 20.5. [eg-open-manifolds]
The space $M_n(\mathbb{R})$ is homeomorphic to $\mathbb{R}^{n^2}$, so it is a manifold of dimension n^2. The subset $GL_n(\mathbb{R})$ is open in $M_n(\mathbb{R})$, so it is also a manifold. The space $GL_n(\mathbb{C})$ is a manifold of dimension $2n^2$, for similar reasons. The set

$$F_n(\mathbb{C}) = \{z \in \mathbb{C}^n \ : \ z_i \neq z_j \text{ for all } i \neq j\}$$

(previously considered in Example 8.12) is open in $\mathbb{C}^n = \mathbb{R}^{2n}$, so it is a $2n$-manifold.

Example 20.6. [eg-sphere-manifold]
The sphere S^n is an n-manifold. Indeed, it is a subspace of the second countable Hausdorff space $\mathbb{R}^{n+1}$, so it is Hausdorff and second countable. Next, we recall from Proposition 18.20 that stereographic projection gives a homeomorphism $g\colon S^n \setminus \{e_n\} \to \mathbb{R}^n$. If we split $\mathbb{R}^{n+1}$ as $\mathbb{R}^n \oplus \mathbb{R}$, the formula is just $g(u,t) = u/(1-t)$. We have a similar homeomorphism $h\colon S^n \setminus \{-e_n\} \to \mathbb{R}^n$ given by $h(u,t) = g(u,-t) = u/(1+t)$. As the open sets $S^n \setminus \{e_n\}$ and $S^n \setminus \{-e_n\}$ cover S^n, this proves that S^n is locally euclidean.

More generally, for any unit vector $a \in S^n$ we have an n-dimensional vector space $T_a = \{v \in \mathbb{R}^{n+1} : \langle v, a\rangle = 0\}$, and there is a homeomorphism $h_a\colon S^n \setminus \{a\} \to T_a$ given by $h_a(x) = (x - \langle x, a\rangle a)/(1 + \langle x, a\rangle)$. This is a chart with $h_a(a) = 0$.

Example 20.7. [eg-On-manifold]
Consider the orthogonal matrix group $O_n = \{A \in M_n(\mathbb{R}) \ : \ A^T A = 1\}$; we will show that this is a manifold. It is a subspace of the second countable Hausdorff space $M_n(\mathbb{R}) \simeq \mathbb{R}^{n^2}$, so it is Hausdorff and second countable. To construct charts, we consider the space o_n of antisymmetric $n \times n$ matrices, so

$o_n = \{B \in M_n(\mathbb{R}) \;:\; B^T + B = 0\}$. This is a vector space of dimension $(n^2 - n)/2$ over $\mathbb{R}$. Proposition 20.8 below will give a homeomorphism $f\colon o_n \to U$, where U is a certain open neighbourhood of I in O_n. For an arbitrary point $A \in O_n$ we can define a homeomorphism from o_n to a neighbourhood of A by $B \mapsto f(B)A$. It follows that O_n is locally euclidean, as required.

Proposition 20.8. [prop-cayley]
Put

$$U = \{A \in O_n \;:\; A + I \textit{ is invertible }\} = \{A \in O_n \;:\; \det(A + I) \neq 0\}.$$

Then there is a homeomorphism $f\colon o_n \to U$ *given by* $f(B) = (I + B)(I - B)^{-1}$*, with inverse* $f^{-1}(A) = (A - I)(A + I)^{-1}$.

Proof. First, we claim that if $B \in o_n$ then $I + tB$ is invertible for all $t \in \mathbb{R}$. Indeed, if $(I + tB)v = 0$ then $v = -tBv$ so

$$\|v\|^2 = \langle v, v\rangle = \langle -tBv, v\rangle = \langle v, -tB^T v\rangle = \langle v, tBv\rangle = \langle v, -v\rangle = -\|v\|^2,$$

so $v = 0$ as required. We can therefore define $f\colon o_n \to GL_n(\mathbb{R})$ by $f(B) = (I+B)(I-B)^{-1}$. Note also that $(I + B)^T = I - B$, so $((I + B)^{-1})^T = ((I + B)^T)^{-1} = (I - B)^{-1}$ and similarly $((I - B)^{-1})^T = (I + B)^{-1}$. This gives

$$f(B)^T f(B) = (I + B)^{-1}(I - B)(I + B)(I - B)^{-1} = (I + B)^{-1}(I + B)(I - B)(I - B)^{-1} = I,$$

which proves that $f(B) \in O_n$. One can also check that $(I - B)/2$ is an inverse for $f(B) + I$, so $f\colon o_n \to U$. Next, we can certainly define $g\colon U \to M_n(\mathbb{R})$ by $g(A) = (A-I)(A+I)^{-1}$. We then have $(A+I)g(A) = A-I$, and we can take transposes to get $g(A)^T(A^T + I) = A^T - I$. We now multiply on the right by A, recalling that $AA^T = A^T A = I$ to get $g(A)^T(A + I) = -(A - I)$, so $g(A)^T = -(A - I)(A + I)^{-1} = -g(A)$. We can thus regard g as a map $U \to o_n$. It is now a matter of straightforward algebra to check that g is inverse to f. □

Proposition 20.9. [prop-manifold-local]
Any manifold is locally compact Hausdorff and locally path-connected.

Proof. Let M be a manifold. For any point $a \in M$, we can choose a chart $f\colon U \to \mathbb{R}^n$ centred at a. We then put $U_r = f^{-1}(OB_r(0))$, and note that f gives a homeomorphism $U_r \to OB_r(0)$, so U_r is open in M and path-connected. It is easy to see that every neighbourhood of a contains U_r for some $r > 0$. We also see that the set $f^{-1}(B_r(0))$ is compact and contains U_r, so U_r is precompact, so M is locally compact. It is also Hausdorff by the definition of a manifold. □

Example 20.10. [eg-double-line]
Put $X = \mathbb{R} \times \{-1, 1\}$ and define an equivalence relation $E \subseteq X^2$ by

$$E = \{((x, a), (y, b)) \;:\; x = y \text{ and } (a = b \text{ or } x \neq 0).\}$$

Put $Y = X/E$, and let $q\colon X \to Y$ be the quotient map. Note that there is a map $p\colon Y \to \mathbb{R}$ given by $p([x, a]) = x$. If $x \neq 0$ then we have $(x, 1)E(x, -1)$ and so $p^{-1}\{x\}$ contains a single point. However, $(0, 1)$ and $(0, -1)$ are not E-equivalent, so $p^{-1}\{0\}$ consists of two points. Now define $f_+, f_-\colon \mathbb{R} \to Y$ by $f_\pm(x) = q(x, \pm 1)$, and then put $U_\pm = f_\pm(\mathbb{R})$. We find that $U_\pm$ is open and $f_\pm\colon \mathbb{R} \to U_\pm$ is a homeomorphism and $Y = U_+ \cup U_-$. It follows that Y is locally euclidean and second countable. However, there is no Hausdorff pair separating $q(0, 1)$ from $q(0, -1)$, so Y is not a Hausdorff space, and thus is not a manifold.

21. Ultrafilters

We now introduce the theory of ultrafilters. This can be seen as a cure for the fact that sequentially closed sets need not be closed, so information about sequences is insufficient to determine a topology. We will define a notion of convergence for ultrafilters, and relate it to the notion of convergence of sequences. It will turn out that information about ultrafilters and their limits *does* determine a topology, and using this we will be able to strengthen various results that are typically proved using sequences.

Definition 21.1. [`defn-filter-subbase`]
Recall from Definition 10.11 that a set $\mathcal{F}$ of subsets of X is said to have the *finite intersection property* if for each finite list $S_1, \dots, S_n$ with each $S_k \in \mathcal{F}$ we have $S_1 \cap \dots \cap S_n \neq \emptyset$. In the case $n = 0$ we interpret $\bigcap_i S_i$ as X, so for $\mathcal{F}$ to have the finite intersection property we require in particular that $X \neq \emptyset$. Such a set $\mathcal{F}$ will be called a *family with the finite intersection property* or an *FFIP* or a *filter subbase*.

Example 21.2. [`eg-ffip`]

(a) The set $\{(a, \infty) : a \in \mathbb{R}\}$ is an FFIP on $\mathbb{R}$.
(b) The set $\{S \subseteq \mathbb{N} : \mathbb{N} \setminus S \text{ is finite}\}$ is an FFIP on $\mathbb{N}$.
(c) The set $\{S \subseteq \mathbb{C} : \mathbb{C} \setminus S \text{ is compact}\}$ is an FFIP on $\mathbb{C}$.
(d) For any topological space X and $x \in X$, the set
$$\mathcal{N}_x = \{ \text{ neighbourhoods of } x \text{ in } X\}$$
has FIP.
(e) If $(x_n)_{n \in \mathbb{N}}$ is a sequence in X then the set
$$\mathcal{F} = \{S \subseteq X : \exists N \quad \{x_N, x_{N+1}, \dots\} \subseteq S\}$$
has FIP.

Definition 21.3. [`defn-ultrafilter`]
An *ultrafilter* (or UF) on X is a set $\mathcal{W}$ of subsets of X which is a maximal FFIP. In other words:

U0: $\mathcal{W}$ has the FIP.
U1: If $\mathcal{W} \subseteq \mathcal{W}'$ and $\mathcal{W}'$ has FIP then $\mathcal{W}' = \mathcal{W}$.

We write βX for the set of all ultrafilters on X.

Example 21.4. [`eg-fixed-ultrafilter`]
For any element $x \in X$, we put
$$\mathcal{W}_x = \{S \subseteq X : x \in S\}.$$
If $S_1, \dots, S_n \in \mathcal{W}_x$ then $S_1 \cap \dots \cap S_n$ contains x, so it is nonempty; so $\mathcal{W}_x$ has FIP. Now let $\mathcal{W}'$ be an FFIP with $\mathcal{W}_x \subseteq \mathcal{W}'$. Note that $\{x\} \in \mathcal{W}_x \subseteq \mathcal{W}'$, so $\{x\} \in \mathcal{W}'$. Thus, for any $S \in \mathcal{W}'$ we must have $S \cap \{x\} \neq \emptyset$, so $x \in S$, so $S \in \mathcal{W}_x$. This proves that $\mathcal{W}_x$ is an ultrafilter on X. Ultrafilters of this type are called *fixed ultrafilters*; a *free ultrafilter* is an ultrafilter that is not fixed. We can now define $\eta \colon X \to \beta X$ by $\eta(x) = \mathcal{W}_x$.

We will show (using Zorn's Lemma) that every filter is contained in an ultrafilter, which means that free ultrafilters are plentiful. However, one cannot specify a particular free ultrafilter (even on the set $\mathbb{N}$) without making an infinite number of arbitrary choices, which is impossible in practice. For a formal statement of this impossibility, see [**2**].

Proposition 21.5. [`prop-ultrafilter-omni`]
Let $\mathcal{W}$ be an ultrafilter on a set X.

(a) *If $S \in \mathcal{W}$ and $T \supseteq S$ then $T \in \mathcal{W}$.*
(b) *If $S_k \in \mathcal{W}$ for each k then $S_1 \cap \dots \cap S_n \in \mathcal{W}$.*
(c) *If $S \subseteq X$ then either $S \in \mathcal{W}$ or $S^c \in \mathcal{W}$ (but not both).*
(d) *If $T \subseteq X$ and $T \cap S \neq \emptyset$ for every $S \in \mathcal{W}$ then $T \in \mathcal{W}$.*
(e) *If $S_1 \cup \dots \cup S_n \in \mathcal{W}$ then $S_k \in \mathcal{W}$ for some k.*
(f) *$X \in \mathcal{W}$*

Proof. (a) Write
$$\mathcal{W}' = \{T \subseteq X : \exists S \in \mathcal{W},\ S \subseteq T\}.$$
Clearly $\mathcal{W} \subseteq \mathcal{W}'$. Moreover, $\mathcal{W}'$ has FIP. Indeed, if $T_1, \dots T_n \in \mathcal{W}'$ then there are sets $S_1, \dots S_n \in \mathcal{W}$ with $T_k \supseteq S_k$ and so
$$T_1 \cap \dots T_n \supseteq S_1 \cap \dots S_n \neq \emptyset$$
It follows by maximality of $\mathcal{W}$ that $\mathcal{W}' = \mathcal{W}$, hence the claim.

(b) Similarly, write

$$\mathcal{W}' = \{S_1 \cap \dots S_n \ : \ n \in \mathbb{N}, S_k \in \mathcal{W}\}$$

This contains $\mathcal{W}$ and has FIP so equals $\mathcal{W}$ as required.

(c) Suppose $S \notin \mathcal{W}$, so $\mathcal{W}' = \mathcal{W} \cup \{S\} \neq \mathcal{W}$. Thus, as $\mathcal{W}$ is maximal among families with FIP, we see that $\mathcal{W}'$ cannot have FIP. This means that there are sets $T_1, \dots T_n$ in $\mathcal{W}$ such that

$$T_1 \cap \dots T_n \cap S = \emptyset$$

It follows from (b) that the set $T = T_1 \cap \dots T_n$ is an element of $\mathcal{W}$. As $T \cap S = \emptyset$, we have $T \subseteq S^c$. Using (a), we find that $S^c \in \mathcal{W}$, as required. We cannot have both $S \in \mathcal{W}$ and $S^c \in \mathcal{W}$ as $S \cap S^c = \emptyset$ and $\mathcal{W}$ is supposed to have FIP.

(d) Suppose that $T \subseteq X$ and $T \cap S \neq \emptyset$ for every $S \in \mathcal{W}$. Using (b) we see that $\mathcal{W} \cup \{T\}$ has FIP and so equals $\mathcal{W}$ by maximality; thus $T \in \mathcal{W}$.

(e): Suppose that $S = \bigcup_{k=1}^n S_k \in \mathcal{W}$, so $S^c = \bigcap_k S_k^c \notin \mathcal{W}$. By claim (b), we must have $S_k^c \notin \mathcal{W}$ for some k. Using (c) we deduce that $S_k \in \mathcal{W}$ as required.

(f): This is the case $n = 0$ of (b).

□

Corollary 21.6. [`cor-free-infinite`]
If $\mathcal{U}$ is a free ultrafilter, then every set in $\mathcal{U}$ is infinite.

Proof. Suppose that $\mathcal{U}$ contains a finite set, say $S = \{x_1, \dots, x_n\}$. We can write this as $S = \{x_1\} \cup \dots \cup \{x_n\}$ and use part (e) of the proposition to see that $\{x_i\} \in \mathcal{U}$ for some i. Now if $T \in \mathcal{W}_{x_i}$ then $\{x_i\} \subseteq T$ so $T \in \mathcal{U}$ by part (a). This proves that $\mathcal{W}_{x_i} \subseteq \mathcal{U}$, but $\mathcal{W}_{x_i}$ is maximal so $\mathcal{U} = \mathcal{W}_{x_i}$, so $\mathcal{U}$ is not free. □

Proposition 21.7. [`prop-ultrafilter-test`]
Suppose that $\mathcal{W}$ has FIP. Then $\mathcal{W}$ is an ultrafilter iff for each $S \subseteq X$ we have $S \in \mathcal{W}$ or $S^c \in \mathcal{W}$.

Proof. One half of this is Proposition 21.5(c). Conversely, suppose $\mathcal{W}$ has FIP and contains S or S^c for each $S \subseteq X$. Consider a family $\mathcal{W}' \supseteq \mathcal{W}$ that still has has FIP. Consider a set $S \in \mathcal{W}'$. By assumption we have either $S \in \mathcal{W}$ or $S^c \in \mathcal{W}$. In the latter case we would have $S, S^c \in \mathcal{W}'$, which would contradict the FIP for $\mathcal{W}'$. We must therefore have $S \in \mathcal{W}$. This holds for all $S \in \mathcal{W}'$, so $\mathcal{W}' = \mathcal{W}$. Thus $\mathcal{W}$ is an ultrafilter. □

We can now reformulate the notion of an ultrafilter in a way which is often useful.

Definition 21.8. [`defn-boolean-character`]
Let X be a set. We write PX for the set of all subsets of X. A *character* of PX is a map $\xi\colon PX \to \{0,1\}$ such that

(p) $\xi(X) = 1$
(q) $\xi(\emptyset) = 0$
(r) $\xi(S \cap T) = \min(\xi(S), \xi(T))$ for all S, T.
(s) $\xi(S^c) = 1 - \xi(S)$ for all S.

We write $\beta' X$ for the set of all characters of PX.

Proposition 21.9. [`prop-boolean-characters`]
There is a canonical bijection $F\colon \beta' X \to \beta X$ given by $F(\xi) = \xi^{-1}\{1\}$.

Proof. Consider an arbitrary map $\xi\colon PX \to \{0,1\}$. Put $\mathcal{W} = \xi^{-1}\{1\}$, so we have $\xi(S) = 1$ for $S \in \mathcal{W}$, and $\xi(S) = 0$ for $S \notin \mathcal{W}$. The conditions in Definition 21.8 translate as follows:

(p) $X \in \mathcal{W}$
(q) $\emptyset \notin \mathcal{W}$
(r) We have $S \cap T \in \mathcal{W}$ iff $S \in \mathcal{W}$ and $T \in \mathcal{W}$
(s) For all $S \subseteq X$ we have $S \in \mathcal{W}$ or $S^c \in \mathcal{W}$ (but not both).

If $\mathcal{W}$ is an ultrafilter then these properties follow easily from Proposition 21.5. Conversely, suppose that (p) to (s) are satisfied. We claim that for $S_1, \dots, S_n \in \mathcal{W}$ we have $\bigcap_i S_i \neq \emptyset$. In view of (q), it will suffice to show that $\bigcap_i S_i \in \mathcal{W}$. The case $n = 0$ is (p), the case $n = 1$ is clear, and the case $n = 2$ is (r). The case $n > 2$ now follows by induction. This means that $\mathcal{W}$ has FIP, and using (s) together with Proposition 21.7 we see that $\mathcal{W}$ is an ultrafilter. □

Recall that βX denotes the set of all ultrafilters on X.

PROPOSITION 21.10. [`prop-beta-functor`]
Let X and Y be arbitrary sets, and $f\colon X \to Y$ be any function. Then for any ultrafilter $\mathcal{W}$ on X, the family

$$f_*(\mathcal{W}) = \{T \subseteq Y \;:\; f^{-1}(T) \in \mathcal{W}\}$$

*is an ultrafilter on Y. Thus, this construction gives a map $f_*\colon \beta X \to \beta Y$. Moreover, we have $(1_X)_*(\mathcal{W}) = \mathcal{W}$, and for $g\colon Y \to Z$ we have $(gf)_*(\mathcal{W}) = g_*(f_*(\mathcal{W}))$. Thus, we have a functor $\beta\colon$ **Sets** $\to$ **Sets**.*

PROOF. Suppose we have sets $T_1, \dots, T_n \in f_*(\mathcal{W})$. This means that the sets $f^{-1}(T_i)$ are all in $\mathcal{W}$, and $\mathcal{W}$ has FIP, so $\bigcap_i f^{-1}(T_i) \neq \emptyset$. If we choose $x \in \bigcap_i f^{-1}(T_i)$ we find that $f(x) \in \bigcap_i T_i$, so $\bigcap_i T_i \neq \emptyset$. This means that $f_*(\mathcal{W})$ has FIP. Now consider a set $T \subseteq Y$. As $\mathcal{W}$ is an ultrafilter, Proposition 21.7 tells us that either $f^{-1}(T) \in \mathcal{W}$ or $f^{-1}(T)^c \in \mathcal{W}$. After noting that $f^{-1}(T)^c = f^{-1}(T^c)$, we deduce that either T or T^c lies in $f_*(\mathcal{W})$. This shows that $f_*(\mathcal{W})$ is an ultrafilter on Y. We have thus defined a map $f_*\colon \beta X \to \beta Y$. It is trivial that $(1_X)_*$ is the identity. Now suppose we have another function $g\colon Y \to Z$. We have

$$\begin{aligned} U \in (gf)_*(\mathcal{W}) &\Leftrightarrow (gf)^{-1}(U) \in \mathcal{W} \\ &\Leftrightarrow f^{-1}(g^{-1}(U)) \in \mathcal{W} \\ &\Leftrightarrow g^{-1}(U) \in f_*(\mathcal{W}) \\ &\Leftrightarrow U \in g_*(f_*(\mathcal{W})), \end{aligned}$$

so this construction is functorial. □

We next turn to the problem of proving that ultrafilters exist.

DEFINITION 21.11. [`defn-ffip-chain`]
A *chain* of FFIPs is a set $\mathcal{L}$ of FFIP's on X such that whenever $\mathcal{F}, \mathcal{G} \in \mathcal{L}$ we have either $\mathcal{F} \subseteq \mathcal{G}$ or $\mathcal{G} \subseteq \mathcal{F}$. In other words, $\mathcal{L}$ is linearly ordered by inclusion.

PROPOSITION 21.12. [`prop-ffip-chain`]
If $\mathcal{L}$ is a chain of FFIPs on X then the set

$$\mathcal{F} = \bigcup_{\mathcal{G} \in \mathcal{L}} \mathcal{G} = \{S \subseteq X \;:\; \exists \mathcal{G} \in \mathcal{L} \quad S \in \mathcal{G}\}$$

has FIP.

PROOF. Suppose $S_1, \dots, S_n \in \mathcal{F}$. Then there are sets $\mathcal{G}_1, \dots, \mathcal{G}_n$ in $\mathcal{L}$ with $S_k \in \mathcal{G}_k$ for each k. As $\mathcal{L}$ is a chain, for each k and l we have $\mathcal{G}_k \subseteq \mathcal{G}_l$ or $\mathcal{G}_l \subseteq \mathcal{G}_k$. By changing the indexing if neccessary, we may assume that

$$\mathcal{G}_1 \subseteq \mathcal{G}_2 \subseteq \cdots \subseteq \mathcal{G}_n.$$

Thus $S_k \in \mathcal{G}_n$ for each k. Moreover, $\mathcal{L}$ is assumed to be a set of FFIPs and $\mathcal{G}_n \in \mathcal{L}$ so $\mathcal{G}_n$ has FIP. Thus $S_1 \cap \cdots \cap S_n \neq \emptyset$, as required. □

THEOREM 21.13. [`thm-ultrafilters-exist`]
If $\mathcal{F}$ is an FFIP on X then there exists an ultrafilter $\mathcal{W}$ on X with $\mathcal{F} \subseteq \mathcal{W}$.

PROOF. Apply Zorn's Lemma (Theorem 35.10) to the poset of FFIPs that contain $\mathcal{F}$. □

PROPOSITION 21.14. [`prop-beta-image`]
Let $f\colon X \to Y$ be a function. Then the image of the map $f_*\colon \beta X \to \beta Y$ is given by

$$f_*(\beta X) = \{\mathcal{V} \in \beta Y \;:\; T \cap f(X) \neq \emptyset \text{ for all } T \in \mathcal{V}\}.$$

In particular, if f is surjective then so is f_*.

Proof. First suppose that $\mathcal{V} = f_*(\mathcal{U})$ for some ultrafilter $\mathcal{U}$, and that $T \in \mathcal{V}$. This means that $f^{-1}(T) \in \mathcal{U}$, so in particular $f^{-1}(T) \neq \emptyset$, so $f(f^{-1}(T)) \neq \emptyset$. However, it is elementary that $f(f^{-1}(T)) = T \cap f(X)$, so $T \cap f(X) \neq \emptyset$ as required.

Conversely, suppose that $\mathcal{V} \in \beta Y$ and that $T \cap f(X) \neq \emptyset$ for all $T \in \mathcal{V}$. Put $\mathcal{U}_0 = \{f^{-1}(T) \; : \; T \in \mathcal{V}\}$. We find that the sets in $\mathcal{U}_0$ are all nonempty, and using the fact that $\bigcap_i f^{-1}(T_i) = f^{-1}(\bigcap_i T_i)$ we deduce that the family $\mathcal{U}_0$ has FIP. It follows that there is an ultrafilter $\mathcal{U} \in \beta X$ with $\mathcal{U}_0 \subseteq \mathcal{U}$. Consider a set $T \subseteq Y$. If $T \in \mathcal{V}$ then $f^{-1}(T) \in \mathcal{U}_0 \subseteq \mathcal{U}$ so $T \in f_*(\mathcal{U})$. This shows that $\mathcal{V} \subseteq f_*(\mathcal{U})$ but $\mathcal{V}$ is an ultrafilter so it is maximal among families with FIP, so we must have $\mathcal{V} = f_*(\mathcal{U})$.

If f is surjective then $f(X) = Y$ and we deduce that $f_*(\beta X) = \beta Y$, so f_* is also surjective. Alternatively, for each $y \in Y$ we can choose $g(y) \in X$ such that $f(g(y)) = y$. This gives a map $g \colon Y \to X$ with $fg = 1_Y$ so $f_* g_* = 1_{\beta Y}$, and it follows again that f_* is surjective. $\square$

Proposition 21.15. [`prop-beta-induced-epi`]
If $f \colon X \to Y$ is surjective and $\mathcal{U} \in \beta X$, then $f_*(\mathcal{U}) = \{f(S) \; : \; S \in \mathcal{U}\}$.

Proof. Put $\mathcal{V} = \{f(S) \; : \; S \in \mathcal{U}\}$. If $T \in f_*(\mathcal{U})$ then the set $S = f^{-1}(T)$ is an element of $\mathcal{U}$, and as f is surjective we have $T = f(f^{-1}(T)) = f(S)$, so $T \in \mathcal{V}$. Conversely, suppose that $T \in \mathcal{V}$, so $T = f(S)$ for some $S \in \mathcal{U}$. We then have $S \subseteq f^{-1}(f(S)) = f^{-1}(T)$, so $f^{-1}(T) \in \mathcal{U}$ by Proposition 21.5(a), so $T \in f_*(\mathcal{U})$. $\square$

We now consider convergence of ultrafilters.

Definition 21.16. [`defn-neighbourhood-filter`]
Let X be a topological space. For $x \in X$ we put

$$\mathcal{N}_x = \{ \text{ open neighbourhoods of } x\} = \{ \text{ open sets } U \; : \; x \in U\}.$$

If we are given a subbasis σ for the topology on X, we also put

$$\mathcal{N}'_x = \{ \text{ subbasic open neighbourhoods of } x\} = \{U \in \sigma \; : \; x \in U\}.$$

Definition 21.17. [`defn-ultrafilter-conv`]
Let X be a topological space, and let σ be a subbasis for the topology. An ultrafilter $\mathcal{W}$ *converges* to $x \in X$ if the following equivalent conditions hold:

(p): $\mathcal{N}_x \subseteq \mathcal{W}$
(q): $\mathcal{N}'_x \subseteq \mathcal{W}$
(r): $x \in \bigcap_{S \in \mathcal{W}} \overline{S}$
(s): For all $S \in \mathcal{W}$ and $U \in \mathcal{N}'_x$ we have $S \cap U \neq \emptyset$.

If so, we write $\mathcal{W} \to x$ and say that x is a *limit* of $\mathcal{W}$.

Proof of equivalence. Clearly (p) implies (q). By the definition of a subbasis, every neighbourhood of x contains a finite intersection of subbasic neighbourhoods of x. Using parts (a) and (b) of Proposition 21.5, we deduce that (q) implies (p), so (p) and (q) are equivalent.

Now suppose that (p) holds. Consider a set $S \in \mathcal{W}$. If $x \not\in \overline{S}$ then the set $U = X \setminus \overline{S}$ is an open neighbourhood of x, so it must lie in $\mathcal{W}$, but $U \cap S = \emptyset$, which contradicts the FIP for $\mathcal{W}$. We must therefore have $x \in \overline{S}$. As S was an arbitrary element of $\mathcal{W}$, we have $x \in \bigcap_{S \in \mathcal{W}} \overline{S}$ as required, so (p) implies (r).

Now suppose that (r) holds. This means that every open neighbourhood of x meets every set in S. In particular, this holds for subbasic open neighbourhoods, which means that (s) holds.

Finally, suppose that (s) holds. Consider a set $U \in \mathcal{N}'_x$, so $U \cap S \neq \emptyset$ for all $S \in \mathcal{W}$. Using part (d) of Proposition 21.5 we deduce that $U \in \mathcal{W}$. This means that $\mathcal{N}'_x \subseteq \mathcal{W}$, so (q) holds. $\square$

Example 21.18. [`eg-ultrafilter-conv`]

(a) The fixed ultrafilter $\mathcal{W}_x$ converges to x.
(b) If X is discrete and $\mathcal{W} \to x$ then $\mathcal{W} = \mathcal{W}_x$. Indeed, $\{x\}$ is a neighbourhood of x so $\{x\} \in \mathcal{W}$. If $S \in \mathcal{W}_x$ then $\{x\} \subseteq S$ so $S \in \mathcal{W}$ by Proposition 21.5(a). This means that $\mathcal{W}_x \subseteq \mathcal{W}$ but $\mathcal{W}_x$ is maximal so $\mathcal{W} = \mathcal{W}_x$.

(c) Put
$$\mathcal{F} = \subseteq \mathbb{C} \ : \ \mathbb{C} \setminus S \text{ is compact}\}.$$
Theorem 21.13 tells us that there exist ultrafilters $\mathcal{W}$ with $\mathcal{W} \supseteq \mathcal{F}$. We claim that no such ultrafilter can converge. Indeed, for any $x \in \mathbb{C}$ we can consider the set $S = \mathbb{C} \setminus B_1(x)$; this lies in $\mathcal{F}$ and therefore in $\mathcal{W}$, but $x \notin \overline{S}$ so criterion (r) in Definition 21.17 tells us that $\mathcal{W} \not\to x$.

Proposition 21.19. [`prop-sequence-ultrafilter`]
Consider a sequence $\underline{x} = (x_n)_{n \in \mathbb{N}}$ in X, and put
$$x_{\geq N} = \{x_k \ : \ k \geq N\}$$
$$\mathcal{F} = \{S \subseteq X \ : \ x_{\geq N} \subseteq S \text{ for some } N \in \mathbb{N}\}.$$
Then $\mathcal{F}$ has FIP. Moreover, the sequence $\underline{x}$ converges to a iff every ultrafilter containing $\mathcal{F}$ converges to a.

Proof. First suppose we have sets $S_1, \dots, S_n \in \mathcal{F}$. This means that there are natural numbers $N_1, \dots, N_n$ such that $x_k \in S_i$ whenever $k \geq N_i$. It follows that if we put $N = \max(x_1, \dots, x_n)$ then $x_N \in \bigcap_i S_i$, so $\bigcap_i S_i \neq \emptyset$. This means that $\mathcal{F}$ has FIP.

Now suppose that $\underline{x}$ converges to a. This means precisely that every open neighbourhood U of a contains $x_{\geq N}$ for some N, and so has $U \in \mathcal{F}$. This means that $\mathcal{N}_x \subseteq \mathcal{F}$, so for every ultrafilter $\mathcal{W} \supseteq \mathcal{F}$ we have $\mathcal{N}_x \subseteq \mathcal{W}$ or equivalently $\mathcal{W} \to x$.

Suppose instead that $\underline{x}$ does not converge to a. This means that there is an open neighbourhood U of a that does not contain $x_{\geq N}$ for any N. Now for $S \in \mathcal{F}$ we have $S \supseteq x_{\geq N}$ for some N, and $x_{\geq N} \not\subseteq U$, so $S \cap U^c \neq \emptyset$. Using this we see that the family $\mathcal{F} \cup \{U^c\}$ has FIP, so there exists an ultrafilter $\mathcal{W}$ with $\mathcal{F} \cup \{U^c\} \subseteq \mathcal{W}$. Clearly $U \notin \mathcal{W}$, so $\mathcal{W} \not\to a$. □

Proposition 21.20. [`prop-ultrafilter-closure`]
Let X be a topological space, let Y be a subset, and let $i\colon Y \to X$ be the inclusion map. Then for $x \in X$ we have $x \in \overline{Y}$ if and only if there is an ultrafilter $\mathcal{W}$ on Y with $i_*(\mathcal{W}) \to x$.

Proof. Suppose that $x \in \overline{Y}$. It follows that the family $\mathcal{F} = \{U \cap Y \ : \ U \in \mathcal{N}_x\}$ consists of nonempty sets, and it is also closed under taking intersections, so it has FIP. We can thus choose an ultrafilter $\mathcal{W}$ on Y with $\mathcal{W} \supseteq \mathcal{F}$. After noting that $U \cap Y = i^{-1}(U)$ we see that $\mathcal{N}_x \subseteq i_*(\mathcal{W})$, so $i_*(\mathcal{W}) \to x$ as required.

Conversely, suppose that $x \notin \overline{Y}$. It follows that the set $U = X \setminus \overline{Y}$ is an open neighbourhood of x. Moreover, we have $i^{-1}(U) = \emptyset$, so U cannot be in $i_*(\mathcal{W})$ for any $\mathcal{W}$, so we cannot have $i_*(\mathcal{W}) \to x$. □

Proposition 21.21. [`prop-ultrafilter-continuity`]
Let $f\colon X \to Y$ be a function between topological spaces. Then f is continuous iff it has the following property: for every ultrafilter $\mathcal{W} \in \beta X$ converging to $a \in X$, the ultrafilter $f_*(\mathcal{W})$ converges to $f(a)$.

Proof. We will temporarily say that f is *ultrafilter-continuous* if it has the stated property.

Suppose that f is continuous. Consider an ultrafilter $\mathcal{W} \in \beta X$ that converges to $a \in X$. Let V be an open neighbourhood of $f(a)$ in Y. By continuity, the set $f^{-1}(V)$ is an open neighbourhood of a in X, and $\mathcal{W} \to x$, so we must have $f^{-1}(V) \in \mathcal{W}$. This means that $V \in f_*(\mathcal{W})$. As V was an arbitrary open neighbourhood of $f(a)$, this means that $f_*(\mathcal{W}) \to f(a)$ as required. Thus, f is ultrafilter-continuous.

Conversely, suppose that f is not continuous. This means that we can find a point $a \in X$ and an open neighbourhood V of $f(a)$ such that the set $f^{-1}(V)$ is not a neighbourhood of a. Put $S = f^{-1}(V)^c = f^{-1}(V^c)$. Consider an open neighbourhood U of a. As $f^{-1}(V)$ is not a neighbourhood of a, we cannot have $U \subseteq f^{-1}(V)$, and it follows that $U \cap S \neq \emptyset$. This means that the family $\mathcal{N}_a \cup \{S\}$ has FIP, so we can choose an ultrafilter $\mathcal{W}$ with $\mathcal{W} \supseteq \mathcal{N}_a \cup \{S\}$. By construction we have $\mathcal{W} \to a$, but $V^c \in f_*(\mathcal{W})$ so $V \notin f_*(\mathcal{W})$ so $f_*(\mathcal{W}) \not\to f(a)$. This shows that f is not ultrafilter-continuous. □

Proposition 21.22. [`prop-ultrafilter-product`]
Consider a family of spaces $(X_i)_{i \in I}$ with product $X = \prod_{i \in I} X_i$ and projections $\pi_i\colon X \to X_i$. Consider an ultrafilter $\mathcal{W} \in \beta X$ and a point $x \in X$. Then $\mathcal{W} \to x$ iff $(\pi_i)_*(\mathcal{W}) \to x_i$ for all i.

Proof. We use the standard subbasis for the product topology, and recall from Definition 21.17 that $\mathcal{W} \to x$ iff $\mathcal{W}$ contains every subbasic open neighbourhood of x. These subbasic neighbourhoods are precisely the sets $\pi_i^{-1}(U_i)$, where U_i runs over the open neighbourhoods of x_i in X_i. We have $\pi_i^{-1}(U_i) \in \mathcal{W}$ iff $U_i \in (\pi_i)_*(\mathcal{W})$, and this holds for all U_i iff $(\pi_i)_*(\mathcal{W}) \to x_i$. $\square$

PROPOSITION 21.23. [prop-ultrafilter-hausdorff]
The space X is Hausdorff if and only if every ultrafilter converges to at most one point.

PROOF. Suppose that X is Hausdorff, that $\mathcal{W} \to x$, and that $y \neq x$. Then there are disjoint open sets U, V with $x \in U$ and $y \in V$. As $\mathcal{W} \to x$, we must have $U \in \mathcal{W}$. As $U \cap V = \emptyset$ and $\mathcal{W}$ has FIP, this means that $V \notin \mathcal{W}$. This means that $\mathcal{W} \not\to y$, as required.

Conversely, suppose that ultrafilter limits are unique. Suppose that x and y do not have disjoint neighbourhoods. Then the family $\mathcal{N}_x \cup \mathcal{N}_y$ has FIP, so there is an ultrafilter $\mathcal{W} \supseteq \mathcal{N}_x \cup \mathcal{N}_y$. This means that $\mathcal{W} \to x$ and $\mathcal{W} \to y$, so by hypothesis $x = y$. This shows that X is Hausdorff. $\square$

THEOREM 21.24. [thm-ultrafilter-compact]
Let X be a topological space, and let σ be a subbasis for the topology. The following are equivalent:

(a) *X is compact.*
(b) *Every covering of X by subbasic open sets has a finite subcover.*
(c) *Every ultrafilter on X has a limit.*

The equivalence of (a) and (b) above is called *Alexander's subbasis theorem.*

PROOF. It is immediate that (a) implies (b). Suppose that (b) holds, and let $\mathcal{W}$ be an ultrafilter on X. Suppose that $\mathcal{W}$ has no limit. By criterion (q) in Definition 21.17, each point $x \in X$ has a subbasic open neighbourhood $U_x \in \sigma$ with $U_x \notin \mathcal{W}$. The sets U_x give a covering of X of X by subbasic open sets, so by assumption there is a finite list $x_1, \dots, x_n$ with $X = U_{x_1} \cup \dots \cup U_{x_n}$. Now part (f) of Proposition 21.5 tells us that $X \in \mathcal{W}$, so part (e) tells us that $U_{x_i} \in \mathcal{W}$ for some i, but this is false by our choice of U_x. Thus, $\mathcal{W}$ must have a limit after all. This proves that (b) implies (c).

Finally, suppose (c) holds. We need to prove that X is compact. Proposition 10.12 gave the following criterion: if $\mathcal{F}$ is any family of closed sets with FIP, we must show that $\bigcap_{F \in \mathcal{F}} F \neq \emptyset$. By Theorem 21.13, we can choose an ultrafilter $\mathcal{W}$ with $\mathcal{F} \subseteq \mathcal{W}$. By assumption, there is a point $x \in X$ such that $\mathcal{W} \to x$. By criterion (r) in Definition 21.17, we have $x \in \overline{S}$ for all $S \in \mathcal{W}$. In particular, we have $x \in F$ for all $F \in \mathcal{F}$, so $x \in \bigcap_{F \in \mathcal{F}} F$ as required. $\square$

THEOREM 21.25 (Tychonov). [thm-tychonov]
Let $(X_i)_{i \in I}$ be a family of compact spaces. Then the product $X = \prod_I X_i$ is compact.

PROOF. Theorem 21.24 tells us that every ultrafilter on X_i has a limit; we must show that the same holds for X. If $\mathcal{W}$ is an ultrafilter on X, then for each i we can choose a limit x_i for $(\pi_i)_*(\mathcal{W})$, and then put these points together to get a point $x = (x_i)_{i \in I} \in X$. Proposition 21.22 then tells us that $\mathcal{W} \to x$, as required. $\square$

REMARK 21.26. [rem-FIX-product]
Suppose we have a set I and a compact space X, and we let $F(I, X)$ denote the set of all functions $u \colon I \to X$. This can be identified with $\prod_{i \in I} X$, so we have a product topology on $F(I, X)$, and Tychonov's theorem tells us that this is compact. In this notation, the projection maps $\pi_J \colon \prod_{i \in I} X \to \prod_{j \in J} X$ become the restriction maps $u \mapsto u|_J$. Thus, for each finite subset $J \subseteq I$ and each open set $U \subseteq F(J, X)$ we have a basic open set $V = \{u \in F(I, X) \;:\; u|_J \in U\}$ in $F(I, X)$.

One important consequence of Tychonov's theorem is that certain spaces of linear functionals on infinite-dimensional vector spaces are compact. One example of this is the space of probability measures on a compact Hausdorff space, which we now describe.

DEFINITION 21.27. [defn-probability-measure]
Let X be a compact Hausdorff space. A *probability measure* on X is a map $\mu \colon C(X) \to \mathbb{R}$ with the following properties:

PM0: μ is linear: for all $a, b \in \mathbb{R}$ and $u, v \in C(X)$ we have $\mu(au + bv) = a\mu(u) + b\mu(v)$.
PM1: μ is positive: if $u(x) \geq 0$ for all $x \in X$ then $\mu(u) \geq 0$.

PM2: μ is normalised: for the constant function with value 1 we have $\mu(1) = 1$.

We will write $PM(X)$ for the set of all probability measures on X.

Example 21.28. [`eg-probability-measure`]
We can define probability measures $\delta, \lambda, \theta, \rho_n$ on $[0, 1]$ by

$$\begin{aligned} \delta(u) &= u(0) \\ \lambda(u) &= \int_0^1 u(x)\,dx \\ \theta(u) &= \int_0^1 2x\,u(x)\,dx \\ \rho_n(u) &= \sum_{k=0}^{n-1} u(k/n)/n. \end{aligned}$$

We now want to introduce a topology on $PM(X)$, by regarding it as a subspace of a certain infinite product. For each element $f \in C(X)$, we put

$$\begin{aligned} b(f) &= \min\{f(x) \;:\; x \in X\} \in \mathbb{R} \\ t(f) &= \max\{f(x) \;:\; x \in X\} \in \mathbb{R}. \end{aligned}$$

It is clear that $b(f) \leq t(f)$, so we have a compact interval $[b(f), t(f)] \subset \mathbb{R}$. This construction gives a family of spaces indexed by the elements of the set $C(X)$, so we can form the product

$$Z = \prod_{f \in C(X)} [b(f), t(f)] = \{\alpha\colon C(X) \to \mathbb{R} \;:\; b(f) \leq \alpha(f) \leq t(f) \text{ for all } f \in C(X)\}.$$

Tychonov's theorem tells us that Z is compact, and it is also easily seen to be Hausdorff.

Proposition 21.29. [`prop-PM-CH`]
$PM(X)$ is a closed subspace of Z, and thus is compact Hausdorff.

Proof. Let μ be a probability measure. Note that if c is constant we have $\mu(c) = c.\mu(1) = c$. For any function f, we can take $c = b(f)$ to see that $\mu(f - b(f)) = \mu(f) - b(f)$. On the other hand, the function $f - b(f)$ is everywhere nonnegative, so by axiom PM1 we have $\mu(f - b(f)) \geq 0$. Putting these together, we see that $\mu(f) \geq b(f)$, and essentially the same argument shows that $\mu(f) \leq t(f)$, so $\mu(f) \in [b(f), t(f)]$. It follows that $PM(X)$ is a subset of Z.

Now let μ be an arbitrary element of Z. If $f \in C(X)$ has $f(x) \geq 0$ for all x then $b(f) \geq 0$, and $\mu(f) \in [b(f), t(f)]$ so $\mu(f) \geq 0$. Thus, axiom PM1 is satisfied. Similarly, we have $\mu(1) \in [b(1), t(1)] = \{1\}$, so PM2 is satisfied. However, PM0 is not satisfied automatically. Consider a pair of functions $f, g \in C(X)$ and a pair of constants $a, b \in \mathbb{R}$. Define $\Delta(a, b, f, g)\colon Z \to \mathbb{R}$ by

$$\Delta(a, b, f, g)(\mu) = \mu(af + bg) - a\mu(f) - b\mu(g).$$

Then define

$$L(a, b, f, g) = \Delta(a, b, f, g)^{-1}\{0\} = \{\mu \in Z \;:\; \mu(af + bg) = a\mu(f) + b\mu(g)\}.$$

If $\mu \in Z$ we note that μ is a probability measure iff it is linear iff $\mu(af + bg) = a\mu(f) + b\mu(g)$ for all a, b, f, g. This means that

$$PM(X) = \bigcap_{(a,b,f,g) \in \mathbb{R}^2 \times C(X)^2} L(a, b, f, g).$$

We also note that $\Delta(a, b, f, g)$ can also be described as

$$\Delta(a, b, f, g) = \pi_{af+bg} - a\,\pi_f - b\,\pi_g\colon \prod_{h \in C(X)} [b(h), t(h)] \to \mathbb{R},$$

and this description makes it clear that $\Delta(a, b, f, g)$ is continuous with respect to the product topology. It follows that $L(a, b, f, g)$ is closed in Z, and $PM(X)$ is the intersection of all these closed sets, so it is again closed. $\square$

Remark 21.30. [lem-measure-convergence]

After adjusting the notation in Proposition 5.35, we obtain the following statement: a sequence $(\mu_n)_{n\in\mathbb{N}}$ in $PM(X)$ converges to ν iff $\mu_n(f) \to \nu(f)$ for each $f \in C(X)$. For example, consider the measures ρ_n and λ on $[0,1]$ described in Example 21.28, so $\rho_n(f)$ is the approximation to $\lambda(f) = \int_0^1 f(x)\,dx$ given by the rectangle rule. Fix $f \in C([0,1])$, and suppose we are given $\epsilon > 0$. Proposition 12.49 tells us that f is uniformly continuous, so we can choose n such that $|f(x) - f(y)| < \epsilon$ whenever $|x - y| \leq 1/n$. It follows easily from this that when $m \geq n$ we have

$$|\lambda(f) - \rho_m(f)| = \left|\sum_{k=0}^{m-1} \int_{k/m}^{(k+1)/m} f(x) - f(k/m)\,dx\right| \leq \sum_{k=0}^{m-1} \int_{k/m}^{(k+1)/m} |f(x) - f(k/m)|\,dx < \epsilon.$$

As ϵ was arbitrary we deduce that $\rho_n(f) \to \lambda(f)$ as $n \to \infty$. It follows that $\rho_n \to \lambda$ in $PM([0,1])$.

21.1. Ultrapowers. We now digress slightly to explain another application of ultrafilters. This may help to develop the reader's intuition, even though we do not know of any direct connection with topology.

Definition 21.31. [defn-ultrapower]

Consider a free ultrafilter $\mathcal{U}$ on $\mathbb{N}$. For any two sequences $x, y \in \prod_{n\in\mathbb{N}} X$ we put

$$T(x,y) = \{n \in \mathbb{N} \;:\; x_n = y_n\}.$$

We then define a relation E on $\prod_{n\in\mathbb{N}} X$ by

$$E = \{(x,y) \;:\; T(x,y) \in \mathcal{U}\}.$$

It is clear that $T(x,x) = \mathbb{N}$ and $T(x,y) = T(y,x)$ and that $T(x,y) \cap T(y,z) \subseteq T(x,z)$, and it follows from this that E is an equivalence relation. We put $X^* = (\prod_n X)/E$, and call this the *ultrapower* of X. For $x \in X$ we let $i(x)$ denote the equivalence class of the constant sequence $(x, x, \dots)$, so $i \colon X \to X^*$.

Remark 21.32. [rem-fixed-ultrapower]

Recall that for each $n \in \mathbb{N}$ we have a fixed ultrafilter $\mathcal{W}_n = \{S \subseteq \mathbb{N} \;:\; n \in S\}$; above we assumed that $\mathcal{U}$ is free, so it does not have the form $\mathcal{W}_n$ for any n. If we did allow $\mathcal{U} = \mathcal{W}_n$, then the projection $\pi_n \colon \prod_n X \to X$ would induce a bijection $X^* \to X$, which would be uninteresting.

In model theory it is common to consider ultrapowers of many different sets X with different types of structure. Here we will restrict attention to the ultrapower $\mathbb{R}^*$, whose elements are called *hyperreals.* Note that this involves a choice of ultrafilter, but it turns out that there are many interesting statements that do not depend on this choice.

Proposition 21.33. [prop-hyperreal-field]

The set $I = \{x \in \prod_{\mathbb{N}} \mathbb{R} \;:\; xE0\}$ is a maximal ideal in $\prod_{\mathbb{N}} \mathbb{R}$, and $\mathbb{R}^*$ can be identified with $(\prod_{\mathbb{N}} \mathbb{R})/I$, so it is a field in a natural way. Moreover, the map $i \colon \mathbb{R} \to \mathbb{R}^*$ is a homomorphism of fields.

Proof. It is clear that $T(x+y,0)$ and $T(x-y,0)$ contain $T(x,0)\cap T(y,0)$, so if $x,y \in I$ then $x \pm y \in I$. It is also clear that $0 \in I$, so I is an additive subgroup of $\prod_{\mathbb{N}} \mathbb{R}$. Similarly, if $x \in I$ and $y \in \prod_{\mathbb{N}} \mathbb{R}$ then $T(xy,0) \supseteq T(x) \in \mathcal{U}$ so $T(xy,0) \in \mathcal{U}$ so $xy \in I$. This means that I is an ideal, so the quotient group $\mathbb{R}' = (\prod_{\mathbb{N}} \mathbb{R})/I$ has a natural ring structure. Moreover, we have $T(x,y) = T(x-y,0)$ so xEy iff $x - y \in I$, so $\mathbb{R}' = \mathbb{R}^*$.

Now suppose that $x \notin I$, so $T(x,0) \notin \mathcal{U}$, so the complement $S = \mathbb{N} \setminus T(x,0) = \{n \;:\; x_n \neq 0\}$ does lie in $\mathcal{U}$. Define $y \in \prod_{\mathbb{N}} \mathbb{R}$ by

$$y_n = \begin{cases} 1/x_n & \text{if } n \in S \\ 0 & \text{if } n \notin S. \end{cases}$$

Now $T(xy,1) = S \in \mathcal{U}$, so $[y]$ is an inverse for $[x]$. This proves that $\mathbb{R}^*$ is a field, so I is a maximal ideal. The rest is clear. □

Proposition 21.34. [prop-hyperreal-order]

For $x, y \in \prod_{\mathbb{N}} \mathbb{R}$ put $U(x,y) = \{n \;:\; x_n \leq y_n\}$. Then there is a total order on $\mathbb{R}^*$ given by $[x] \leq [y]$ iff $U(x,y) \in \mathcal{U}$. Moreover, this makes $\mathbb{R}^*$ into a nonarchimedean ordered field (as in Definition 34.4), and the homomorphism $i \colon \mathbb{R} \to \mathbb{R}^*$ preserves the order.

Proof. First, suppose we have four sequences $x, y, x', y' \in \prod_{\mathbb{N}} \mathbb{R}$. It is then easy to see that

$$U(x,y) \supseteq T(x,x') \cap T(y,y') \cap U(x',y')$$
$$U(x',y') \supseteq T(x,x') \cap T(y,y') \cap U(x,y).$$

It follows that if xEx' and yEy' (so $T(x,x'), T(y,y') \in \mathcal{U}$) then ($U(x,y) \in \mathcal{U}$ iff $U(x',y') \in \mathcal{U}$). We therefore have a well-defined relation on $\mathbb{R}^*$ given by ($[x] \leq [y]$ iff $U(x,y) \in \mathcal{U}$). As $U(x,x) = \mathbb{N}$ we see that $[x] \leq [x]$. If $[x] \leq [y]$ and also $[y] \leq [x]$ we see that $U(x,y), U(y,x) \in \mathcal{U}$, and $T(x,y) = U(x,y) \cap U(y,x)$ so $T(x,y) \in \mathcal{U}$, so $[x] = [y]$. Moreover, if $[x] \leq [y]$ and $[y] \leq [z]$ then we have $U(x,y) \cap U(y,z) \subseteq U(x,z)$ so $U(x,z) \in \mathcal{U}$ so $[x] \leq [z]$. This proves that we have a partial order on $\mathbb{R}^*$.

To see that this is total, note that we always have either $U(x,y) \in \mathcal{U}$ (in which case $[x] \leq [y]$) or $U(x,y)^c \in \mathcal{U}$. In the latter case we note that $U(x,y)^c \subseteq U(y,x)$, so $U(y,x) \in \mathcal{U}$ and $[y] \leq [x]$.

Now suppose we have $[a] \leq [b]$ and $[c] \leq [d]$, so $U(a,b), U(c,d) \in \mathcal{U}$. It is clear that $U(a,b) \cap U(c,d) \subseteq U(a+c, b+d)$ and thus that $[a+c] \leq [b+d]$. Similarly, we have $U(a,b) \cap U(0,c) \subseteq U(ac, bc)$, so if $[a] \leq [b]$ and $[0] \leq [c]$ then $[ac] \leq [bc]$. This means that $\mathbb{R}^*$ is an ordered field.

Now suppose we have numbers $x, y \in \mathbb{R}$, and we let $c(x)$ and $c(y)$ be the corresponding constant sequences, so $i(x) = [c(x)]$. If $x \leq y$ then $U(c(x), c(y)) = \mathbb{N} \in \mathcal{U}$, so $i(x) \leq i(y)$; this shows that i is order-preserving.

Now let ω denote the sequence $(0, 1, 2, 3, \dots)$. Recall from Corollary 21.6 that $\mathcal{U}$ does not contain any finite sets, so in particular it does not contain the set $U(\omega, i(n)) = \{0, 1, \dots, n\}$, so $[\omega] \not\leq i(n)$. This shows that $\mathbb{R}^*$ is not archimedean. $\square$

Definition 21.35. [`defn-finite-hyperreals`]
We say that a hyperreal $a \in \mathbb{R}^*$ is *finite* if $-i(n) \leq a \leq i(n)$ for some $n \in \mathbb{N}$. We say that a is *infinitesimal* if $-i(1/n) \leq a \leq i(1/n)$ for all $n > 0$. We write F for the set of finite hyperreals, and I for the subset of infinitesimals.

Proposition 21.36. [`prop-standard-part`]
The set F is a subring of $\mathbb{R}^*$, and I is an ideal in F. Moreover, there is a canonical isomorphism $F/I \simeq \mathbb{R}$.

Proof. It is clear that $i(x)$ is finite for all $x \in \mathbb{R}$; in particular, $i(0)$ and $i(1)$ are finite. Now suppose that a and b are finite, say $-i(n) \leq a \leq i(n)$ and $-i(m) \leq b \leq i(m)$. It follows that $-i(n+m) \leq a \pm b \leq i(n+m)$ and $-i(nm) \leq ab \leq i(nm)$, so $a \pm b$ and ab are also finite. This means that F is an $\mathbb{R}$-algebra.

Now suppose that c and d are infinitesimal. For all $k > 0$ we then have $-i(1/2k) \leq c, d \leq i(1/2k)$, so $-i(1/k) \leq c + d \leq i(1/k)$; this means that $c + d$ is also infinitesimal. Now suppose we have $a \in F$, so $-i(n) \leq a \leq i(n)$ say. As c is infinitesimal we have $-i(1/kn) \leq c \leq i(1/kn)$, and it follows that $-i(1/k) \leq ac \leq i(1/k)$. This means that ac is again infinitesimal. Using this we see that I is an ideal in $\mathbb{R}$.

Next, for $a \in \mathbb{R}^*$ we put $L(a) = \{x \in \mathbb{R} : i(x) \leq a\}$. If $-i(n) \leq a \leq i(n)$ then $L(a)$ contains $-n$ and is bounded above by n, so we can define $\lambda(a) = \sup(L(a)) \in \mathbb{R}$. This gives a map $\lambda\colon F \to \mathbb{R}$.

Suppose that $\lambda(a) = 0$. This means that for all $n > 0$ the number $-1/n$ is not an upper bound for $L(a)$, so there exists $x \in L(a)$ with $-1/n < x$, and it follows that $-i(1/n) \leq a$. On the other hand, as 0 is an upper bound for $L(a)$ we have $1/n \notin L(a)$ so $a < i(1/n)$. Putting these facts together, we see that $a \in I$. This argument can easily be reversed to prove the converse, so we have $\lambda(a) = 0$ iff $a \in I$. Next, it is clear that $\lambda(a - i(t)) = \lambda(a) - t$ for all $t \in \mathbb{R}$. By taking $t = \lambda(a)$, we see that $\lambda(a - i(\lambda(t))) = 0$, so $a - i(\lambda(t)) \in I$. From this it is easy to check that the composite $\mathbb{R} \xrightarrow{i} F \to F/I$ is an isomorphism. $\square$

Remark 21.37. [`rem-nonstandard-analysis`]
The hyperreal framework can be used to make respectable various heuristic manipulations with infinite and infinitesimal quantities. For example, given a smooth function $f\colon \mathbb{R} \to \mathbb{R}$ we can define an extension $f^*\colon \mathbb{R}^* \to \mathbb{R}^*$ by

$$f^*([x]) = [(f(x_0), f(x_1), f(x_2), \dots)].$$

We then find that when a is finite and ϵ is a nonzero infinitesimal, we have

$$(f^*(a+\epsilon) - f^*(a))/\epsilon = (f')^*(a) + \text{ an infinitesimal }.$$

This approach is called *nonstandard analysis*. There is an extensive literature, to which [**3**] will serve as an introduction.

21.2. The Stone-Cech compactification. In Proposition 21.10, we explained how to regard β as a functor $\mathbf{Sets} \to \mathbf{Sets}$. We will now introduce a topology on βX, making it a compact Hausdorff space, called the *Stone-Cech compactification* of X. We will show that the set of fixed ultrafilters is an open, dense, discrete subset of βX, which can be identified with the original set X. We will also show that for any function $f\colon X \to Y$, the induced map $f_*\colon \beta X \to \beta Y$ is continuous, so β gives a functor from sets to compact Hausdorff spaces. In fact, this functor is left adjoint to a suitable forgetful functor.

Recall that Proposition 21.9 gives a canonical bijection $F\colon \beta' X \to \beta' X$, where $\beta' X$ is the set of characters of PX. In particular, $\beta' X$ is a subset of the set $F(PX,\{0,1\}) \simeq \prod_{S\in PX}\{0,1\}$, which is a compact Hausdorff space by the Tychonov theorem.

PROPOSITION 21.38. [`prop-chars-compact`]
The set $\beta' X$ is closed in $F(PX,\{0,1\})$, and so is compact and Hausdorff. Moreover, the sets

$$D'_S = \{\xi \in \beta' X \;:\; \xi(S) = 1\}$$

are both open and closed in $\beta' X$, and they form a basis for the topology.

PROOF. Put

$$\begin{aligned}
Z_0 &= \{\xi \in F(PX,\{0,1\}) \;:\; \xi(X) = 1\} \\
Z_1 &= \{\xi \in F(PX,\{0,1\}) \;:\; \xi(\emptyset) = 0\} \\
Z_2(S,T) &= \{\xi \in F(PX,\{0,1\}) \;:\; \xi(S\cap T) = \min(\xi(S),\xi(T))\} \\
Z_2(S) &= \{\xi \in F(PX,\{0,1\} \;:\; \xi(S) + \xi(S^c) = 1\},
\end{aligned}$$

so

$$\beta' X = Z_0 \cap Z_1 \cap \bigcap_{S,T\in PX} Z_2(S,T) \cap \bigcap_{S\in PX} Z_2(S).$$

It will suffice to show that all of the terms in this intersection are closed. For each $S \in PX$ we have a projection map

$$\pi_S\colon F(PX,\{0,1\}) \to \{0,1\} \subset \mathbb{Z},$$

which is continuous by the definition of the product topology. It follows that we can define another continuous map $f_{S,T}\colon F(PX,\{0,1\}) \to \mathbb{Z}$ by

$$f_{S,T}(\xi) = \xi(S\cap T) - \min(\xi(S),\xi(T)).$$

We then have $Z_2(S,T) = f_{S,T}^{-1}\{0\}$, which proves that $Z_2(S,T)$ is closed. A similar argument shows that the sets $Z_3(S)$ are closed, as are the sets $Z_0 = \pi_X^{-1}\{1\}$ and $Z_1 = \pi_\emptyset^{-1}\{0\}$. This means that $\beta' X$ is closed, and thus is compact and Hausdorff as claimed.

Next, let p_S be the restriction of π_S to $\beta' X$, and put

$$\begin{aligned}
\sigma_0 &= \{\pi_S^{-1}(U) \;:\; S \in PX,\ U \subseteq \{0,1\}\} \\
\sigma_1 &= \{A \cap \beta' X \;:\; A \in \sigma_0\} = \{p_S^{-1}(U) \;:\; S \in PX,\ U \subseteq \{0,1\}\} \\
\sigma_2 &= \{D_S \;:\; S \in PX\} = \{p_S^{-1}\{1\} \;:\; S \in PX\}.
\end{aligned}$$

It is standard that σ_0 is a subbasis for the product topology on $F(PX,\{0,1\})$, and it follows that σ_1 is a subbasis for the topology on $\beta' X$. Now $p_S^{-1}(\emptyset) = \emptyset$ and $p_S^{-1}(\{0,1\}) = \beta' X$, and using the condition $\xi(S^c) = 1 - \xi(S)$ we see that $p_S^{-1}\{0\} = p_{S^c}^{-1}\{1\} = D_{S^c}$. Using this, it follows that σ_2 is still a subbasis for the topology. It follows that the family of all finite intersections $\bigcap_{i=1}^n D_{S_i}$ is a basis for the topology. However, using the condition $\xi(S\cap T) = \min(\xi(S),\xi(T))$ we see that $\xi(\bigcap_i S_i) = \min(\xi(S_1),\dots,\xi(S_n))$ and so $\bigcap_i D_{S_i} = D_{\bigcap_i S_i}$. As this is already in σ_2, we see that σ_2 is actually a basis, as claimed. ☐

COROLLARY 21.39. [`cor-beta-topology`]
The sets $D_S = \{\mathcal{W} \in \beta X \;:\; S \in \mathcal{W}\}$ form a basis for a topology on βX, with respect to which it is compact and Hausdorff.

Proof. We have a bijection $F\colon \beta' X \to \beta X$ as in Proposition 21.9, and we declare that $U \subseteq \beta X$ is open iff $F^{-1}(U)$ is open in $\beta' X$. This gives a topology with respect to which F is a homeomorphism, so βX is compact Hausdorff. From the definitions we see that $F^{-1}(D_S) = D'_S$, so the sets D_S form a basis for the topology. $\square$

Now recall that we defined a map $\eta\colon X \to \beta X$ by

$$\eta(x) = \mathcal{W}_x = \{S \subseteq X \ : \ x \in S\}.$$

Proposition 21.40. [prop-beta-density]
The map η is injective, and $\eta(X)$ is discrete, open and dense in βX.

Proof. First, if $x \neq y$ then $\{x\} \in \eta(x)$ but $\{x\} \notin \eta(y)$, so $\eta(x) \neq \eta(y)$. This proves that η is injective.

Next, it is straightforward to check that $D_{\{x\}} = \{\eta(x)\}$, so the singleton $\{\eta(x)\}$ is open. It follows that for any subset $Y \subseteq X$, the set $\eta(Y) = \bigcup_{y\in Y}\{\eta(y)\} = \bigcup_{y \in Y} D_{\{y\}}$ is open in βX. This means that $\eta(X)$ is open and discrete.

Next, consider a nonempty open set $U \subseteq \beta X$. As this is nonempty, we can choose $\mathcal{W} \in U$. As U is open, there must be a basic open set D_S such that $\mathcal{W} \in D_S \subseteq U$. As $D_\emptyset = \emptyset$, we must have $S \neq \emptyset$. We can thus choose $x \in S$, and we find that $S \in \eta(x)$, so $\eta(x) \in D_S \subseteq U$, so U meets $\eta(X)$. As U was an arbitrary nonempty open set, this means that $\eta(X)$ is dense. $\square$

Proposition 21.41. [prop-beta-topological-functor]
*For any function $f\colon X \to Y$, the induced map $f_*\colon \beta X \to \beta Y$ is continuous. Thus, we can regard β as a functor from **Sets** to the category **CompHaus** of compact Hausdorff spaces and continuous maps.*

Proof. It will suffice to check that for each subset $T \subseteq Y$, the preimage $f_*^{-1}(D_T)$ is open in βX. Consider an ultrafilter $\mathcal{U}$ on X. We have $\mathcal{U} \in f_*^{-1}(D_T)$ iff $f_*(\mathcal{U}) \in D_T$ iff $T \in f_*(\mathcal{U})$ iff $f^{-1}(T) \in \mathcal{U}$ iff $\mathcal{U} \in D_{f^{-1}(T)}$. This means that $f_*^{-1}(D_T) = D_{f^{-1}(T)}$, which is open in βX as required. $\square$

Now consider a compact Hausdorff space A. Proposition 21.23 and Theorem 21.24 tell us that every ultrafilter $\mathcal{W}$ on A has a unique limit in A, which we denote by $\epsilon(\mathcal{W})$.

Proposition 21.42. [prop-limit-continuous]
The map $\epsilon\colon \beta A \to A$ is continuous. More precisely, for every open set $V \subseteq A$, the preimage $\epsilon^{-1}(V)$ is the union of all the basic open sets D_S for which $\overline{S} \subseteq V$.

Proof. Let U be the union of all D_S with $\overline{S} \subseteq V$. Suppose that $\mathcal{W} \in \epsilon^{-1}(V)$. This means that $\mathcal{W}$ converges to some point $a \in V$, or in other words that $\mathcal{N}_a \subseteq \mathcal{W}$. As compact Hausdorff spaces are regular, we can choose an open set S with $a \in S \subseteq \overline{S} \subseteq V$. Now $S \in \mathcal{N}_a$ so $S \in \mathcal{W}$ so $\mathcal{W} \in D_S$ and $\overline{S} \subseteq V$, so $\mathcal{W} \in U$.

Conversely, suppose that $\mathcal{W} \notin \epsilon^{-1}(V)$, so $\mathcal{W}$ converges to some point $a \notin V$. Now if $S \subseteq A$ with $\overline{S} \subseteq V$ we see that $\overline{S}^c$ is an open neighbourhood of a, so $\overline{S}^c \in \mathcal{W}$. As $S \cap \overline{S}^c = \emptyset$, we must have $S \notin \mathcal{W}$, so $\mathcal{W} \notin D_S$. As S was arbitrary this means that $\mathcal{W} \notin U$. We deduce that $\epsilon^{-1}(V) = U$ as claimed. This is clearly open in βA, so the map $\epsilon\colon \beta A \to A$ is continuous. $\square$

Proposition 21.43. [prop-beta-adjoint]
*The functor $\beta\colon$ **Sets** $\to$ **CompHaus** is left adjoint to the forgetful functor $U\colon$ **CompHaus** $\to$ **Sets**.*

Proof. We have already defined a function $\eta\colon X \to U\beta X$ by $\eta(x) = \mathcal{W}_x$. We claim that this is natural, or equivalently that for every function $f\colon X \to Y$ we have $f_*(\mathcal{W}_x) = \mathcal{W}_{f(x)}$. Indeed, we have $T \in f_*(\mathcal{W}_x)$ iff $f^{-1}(T) \in \mathcal{W}_x$ iff $x \in f^{-1}(T)$ iff $f(x) \in T$ iff $T \in \mathcal{W}_{f(x)}$ as required. This can be displayed diagramatically as follows:

$$\begin{tikzcd} X \arrow[r, "f"] \arrow[d, "\eta_X"'] & Y \arrow[d, "\eta_Y"] \\ U\beta X \arrow[r, "Uf_*"'] & U\beta Y. \end{tikzcd}$$

Next, we have already defined continuous maps $\epsilon\colon \beta UA \to A$ for all compact Hausdorff spaces A. (Previously we wrote the domain without a U, but it was implicit.) We claim that this is natural, or in other words that for every continuous map $g\colon A \to B$ and every ultrafilter $\mathcal{W}$ on A, we have $\epsilon(f_*(\mathcal{W})) = f(\epsilon(\mathcal{W}))$. This is just a restatement of Proposition 21.21.

We now need to check the triangular identities. The first of these says that the composite

$$UA \xrightarrow{\eta_{UA}} U\beta UA \xrightarrow{U\epsilon_A} UA$$

is the identity, or more explicitly, that the fixed ultrafilter $\mathcal{W}_a$ converges to a. This is immediate from the definitions. The second says that the composite

$$\zeta = (\beta X \xrightarrow{(\eta_X)_*} \beta U\beta X \xrightarrow{\epsilon_{\beta X}} \beta X)$$

is the identity. Here ζ is continuous and βX is Hausdorff so the set $Z = \{\mathcal{W} \in \beta X \ :\ \zeta(\mathcal{W}) = \mathcal{W}\}$ is closed in βX. We need to show that Z is all of βX. As $\eta(X)$ is dense in βX, it will suffice to show that $\eta(X) \subseteq Z$, or equivalently that $(U\zeta)\eta = \eta\colon X \to U\beta X$. For this, we consider the diagram

$$\begin{tikzcd}
X \arrow[r, "\eta_X"] \arrow[d, "\eta_X"'] & U\beta X \arrow[d, "\eta_{U\beta X}"'] \arrow[dr, "1"] & \\
U\beta X \arrow[r, "U(\eta_X)_*"'] & U\beta U\beta X \arrow[r, "U\epsilon_{\beta X}"'] & U\beta X.
\end{tikzcd}$$

The square is the naturality diagram for η in the case where $Y = U\beta X$ and $f = \eta_X$; it is therefore commutative. The triangle is the first triangular identity in the case $A = \beta X$, so it is also commutative. The composite along the bottom is $U\zeta$. It is now clear that $(U\zeta)\eta = \eta$ as required. □

Remark 21.44. [rem-beta-adjunction]
The adjunction gives a natural bijection $\mathbf{Sets}(X, UA) \simeq \mathbf{CompHaus}(\beta X, A)$ for all sets X and compact Hausdorff spaces A as in Proposition 36.60. By working through the definitions we see that a function $f\colon X \to UA$ corresponds to the continuous map $f^\#\colon \beta X \to A$ given by

$$f^\#(\mathcal{W}) = \epsilon(f_*(\mathcal{W})) = \text{the unique limit point of the ultrafilter } f_*(\mathcal{W}).$$

21.3. The Vietoris space. We now have the tools necessary to introduce a topology on the set of closed subsets of X, at least when X is compact Hausdorff.

Definition 21.45. [defn-vietoris]
Let X be a compact Hausdorff space, with topology τ say. Let $V(X)$ denote the set of all closed subsets of X. For any open set $U \in \tau$, we put

$$\begin{aligned}
s(U) &= \{K \in V(X) \ :\ K \subseteq U\} \\
m(U) &= \{K \in V(X) \ :\ K \cap U \neq 0\} \\
\sigma' &= \{s(U) \ :\ U \in \tau\} \cup \{m(U) \ :\ U \in \tau\}.
\end{aligned}$$

(Mnemonic: $s(U)$ stands for "subset of U", and $m(U)$ for "meets U"). The family σ' is a subbasis for a topology τ' on $V(X)$, called the *Vietoris topology*. We call $V(X)$ (equipped with this topology) the *Vietoris space* for X.

Lemma 21.46. [lem-vietoris-relations]
The sets defined above have the following properties:
(a) $s(U) \cap s(V) = s(U \cap V)$
(b) $s(U) \cap m(V) = s(U) \cap m(U \cap V)$
(c) *If* $U \subseteq V$ *then* $m(U) \cap m(V) = m(U)$
(d) $s(U \cup V) = s(U) \cup (s(U \cup V) \cap m(V))$.

Proof. Let K be a closed subset of X.
(a) It is clear that ($K \subseteq U$ and $K \subseteq V$) iff $K \subseteq U \cap V$, or in other words $K \in s(U) \cap s(V)$ iff $K \in s(U \cap V)$.

(b) Suppose that $K \in s(U) \cap m(V)$, so $K \subseteq U$ and there exists a point $x \in K \cap V$. As $K \subseteq U$ we have $K \cap V = K \cap U \cap V$, so $x \in K \cap (U \cap V)$, so $K \in s(U) \cap m(U \cap V)$. This shows that $s(U) \cap m(V) \subseteq s(U) \cap m(U \cap V)$, and the opposite inclusion is clear.

(c) Now suppose that $U \subseteq V$. It is clear that $K \cap U \subseteq K \cap V$, so if K meets U then it also meets V. This means that $m(U) \subseteq m(V)$, so $m(U) \cap m(V) = m(U)$.

(d) Suppose that $K \in s(U \cap V)$ but that $K \notin s(U)$. This means that $K \subseteq U \cap V$ but $K \not\subseteq U$, so K must meet V, so $K \in s(U \cup V)$. Using this, we see that $s(U \cup V) \subseteq s(U) \cup (s(U \cup V) \cap m(V))$, and the opposite inclusion is clear.

□

LEMMA 21.47. [lem-vietoris-basis]
Let β' be the collection of all sets of the form

$$s(U) \cap m(V_1) \cap \cdots \cap m(V_r)$$

where

- *The sets U and V_i are open in X*
- *We have $V_i \subseteq U$ for all i*
- *For $i \neq j$ we have $V_i \not\subseteq V_j$.*

Then β' is closed under finite intersections, and is a basis for the topology on $V(X)$.

PROOF. Let β'' be the collection of all sets of the form

$$s(U_1) \cap \cdots \cap s(U_q) \cap m(V_1) \cap \cdots \cap m(V_r),$$

where the sets U_i and V_j are open in X. From very general facts about bases and subbases we see that β'' is a basis for the topology, and it is clearly closed under finite intersections. It will therefore suffice to show that $\beta' = \beta''$. The inclusion $\beta' \subseteq \beta''$ is clear. In the opposite direction, suppose we have a set

$$A = s(U_1) \cap \cdots \cap s(U_q) \cap m(V_1) \cap \cdots \cap m(V_r) \in \beta''.$$

Put $U = \bigcap_i U_i$ and $V_i' = U \cap V_i$. Using parts (a) and (b) of Lemma 21.46 we see that

$$A = s(U) \cap m(V_1') \cap \cdots \cap m(V_r').$$

Next, if $V_i' \subseteq V_j'$ for some $i \neq j$, part (c) of Lemma 21.46 tells us that we can omit the term $m(V_j')$ without changing the intersection. After a finite number of steps of this type, we obtain an expression for A showing that it is a member of β'. □

THEOREM 21.48. [thm-vietoris-compact]
The space $V(X)$ is compact Hausdorff.

PROOF. We will use Alexander's subbasis theorem. Consider a covering of $V(X)$ by subbasic open sets:

$$V(X) = \bigcup_{i \in I} s(U_i) \cup \bigcup_{j \in J} m(V_j).$$

Write

$$K = X \setminus \bigcup_{j \in J} V_j.$$

Note that $K \in V(X)$, but $K \notin m(V_j)$ for any $j \in J$ so we must instead have $K \in s(U_i)$ for some $i \in I$. Thus $K \subseteq U_i$. Next, consider $K' = X \setminus U_i \in V(X)$. Note that $K' \subseteq X \setminus K = \bigcup_J V_j$ and K' is compact so $K' \subseteq \bigcup_{J'} V_j$ for some finite set $J' \subseteq J$.

Now consider an arbitary element $L \in V(X)$. Either $L \subseteq U_i$ (so $L \in s(U_i)$) or $L \cap K' \neq \emptyset$. In the latter case we have $L \cap \bigcup_{J'} V_j \neq \emptyset$ so $L \cap V_j \neq \emptyset$ for some $j \in J'$, so $L \in m(V_j)$. Either way, we have

$$L \in s(U) \cup \bigcup_{J'} m(V_j).$$

As L was an arbitrary element of $V(X)$ we deduce that

$$V(X) = s(U) \cup \bigcup_{J'} m(V_j).$$

This is the required finite subcover, proving that $V(X)$ is compact.

Finally, we prove that $V(X)$ is Hausdorff. Suppose $K, L \in Z$ with $K \neq L$. Without loss of generality, there is an element $x \in K \setminus L$. As compact Hausdorff spaces are regular, we can choose disjoint open sets U and V with $x \in U$ and $L \subseteq V$. Then $K \in m(U)$ and $L \in s(V)$ and $m(U) \cap s(V) = \emptyset$ as required. $\square$

Proposition 21.49. [`prop-vietoris-metric`]
Let X be a compact metric space. Then the Vietoris topology on $V(X)$ is the same as the metric topology coming from the Hausdorff metric $\overline{d}$ discussed in Section 12.3.

Proof. Let τ' be the Vietoris topology and let τ'' be the metric topology. Consider a subset $A \subseteq V(X)$ and an element $K \in A$. It will suffice to prove that A is a τ'-neighbourhood of K iff it is a τ''-neighbourhood of K.

Suppose that A is a τ'-neighbourhood of K, so there exists a set $B = s(U) \cap m(V_1) \cap \cdots \cap m(V_r)$ as in Lemma 21.47 such that $K \in B \subseteq A$. As $K \in B$ we have $K \subseteq U$ and we can choose $x_i \in K \cap V_i$ for each i. Now let $\epsilon > 0$ be small enough that $OB_\epsilon(x_i) \subseteq V_i$ for all i, and also $\epsilon < \overline{d}(K, U^c)$. Now suppose we have $L \in V(X)$ with $\overline{d}(K, L) < \epsilon$. On the one hand this means that for $y \in L$ we have $d(y, K) < \epsilon$ but for $z \in U^c$ we have $d(z, K) \geq d(U^c, K) > \epsilon$; it follows that $L \subseteq U$. On the other hand, for each $x \in K$ we have $d(x, L) < \epsilon$. In particular, we have $d(x_i, L) < \epsilon$, so we can choose $y_i \in L$ with $d(x_i, y_i) < \epsilon$. As $OB_\epsilon(x_i) \subseteq V_i$ this gives $y_i \in L \cap V_i$, so $L \cap V_i \neq \emptyset$, so $L \in m(V_i)$. This means that $L \in B \subseteq A$ whenever $d(K, L) < \epsilon$, so A is a τ''-neighbourhood of K.

Conversely, suppose we start with the assumption that A is a τ''-neighbourhood of K. This means that there exists $\epsilon > 0$ such that $L \in A$ whenever $d(K, L) < \epsilon$. Put $U = \{x \,:\, d(x, K) < \epsilon\}$, which is an open set containing K. Choose an $\epsilon/2$-net $\{x_1, \ldots, x_r\}$ for K, and put $V_i = OB_{\epsilon/2}(x_i)$ and $B = s(U) \cap \bigcap_i m(V_i)$. It is clear that B is a τ'-neighbourhood of K; we claim that it is contained in A. Indeed, suppose that $L \in B$. This firstly means that $L \subseteq U$, so $d(y, K) < \epsilon$ for all $y \in L$. Next, suppose we have $x \in K$. By construction we have $d(x, x_i) < \epsilon/2$ for some i. As $L \in m(V_i) = m(OB_{\epsilon/2}(x_i))$ we can choose $y \in L$ with $d(x_i, y) < \epsilon/2$. We now have $d(x, y) < \epsilon$, so $d(x, L) < \epsilon$. It now follows that

$$d(K, L) = \max(\max\{d(x, L) \,:\, x \in K\}, \max\{d(y, K) \,:\, y \in L\}) < \epsilon,$$

so $L \in A$ as required. This proves that A is a τ'-neighbourhood of K. $\square$

22. Paracompactness and Partitions of Unity

In this section we will discuss the notion of paracompactness. This is a rather weak condition, satisfied by the great majority of spaces that arise in practice, although it is quite strenuous to prove this. However, it is very useful. There are many topological problems that can always be solved on sufficiently small open sets; one then wants to patch together these local solutions (which may not be unique) to get a global solution. It will turn out that this is much easier when the total space is paracompact.

Definition 22.1. [`defn-paracompact`]
Let X be a topological space.

(a) A *cover* of X is a family $\mathcal{U} = (U_i)_{i \in I}$ of subsets of X whose union is the whole of X. An *open cover* is a cover, each of whose sets is open.
(b) We say that a cover $\mathcal{U}$ as above is *locally finite* if for each point $x \in X$ there is an open neighbourhood N such that $\{i \,:\, U_i \cap N \neq \emptyset\}$ is finite.
(c) For any function $\phi\colon X \to \mathbb{R}$, the *support* of ϕ is the closure of the set $\{x \,:\, \phi(x) \neq 0\}$.
(d) A *partition of unity subordinate to* $\mathcal{U}$ is a family of continuous functions $\phi_i\colon X \to [0,1]$ such that $\mathrm{supp}(\phi_i) \subseteq U_i$, and $(\mathrm{supp}(\phi_i))_{i \in I}$ is locally finite, and $\sum_i \phi_i = 1$.
(e) We say that $\mathcal{U}$ is *numerable* if there exists a partition of unity subordinate to $\mathcal{U}$.
(f) Now let $\mathcal{V} = (V_j)_{j \in J}$ be another cover. We say that $\mathcal{V}$ is a *refinement* of $\mathcal{U}$ if for each $j \in J$ there exists $i \in I$ with $V_j \subseteq U_i$.
(g) We say that X is *paracompact* if every open cover has a locally finite open refinement.

Partitions of unity are often a useful tool for patching together local solutions to obtain a global solution. We will prove as Theorem 22.22 that every open cover of a paracompact Hausdorff space is numerable.

We start with the following simple result, which illustrates the meaning and importance of local finiteness.

LEMMA 22.2. [lem-sum-support]
Let $(\phi_i)_{i\in I}$ be a family of continuous functions $X\to[0,\infty)$, and suppose that the family $(\mathrm{supp}(\phi_i))_{i\in I}$ is locally finite. Then the sum $\phi=\sum_i\phi_i$ is finite and continuous, and $\mathrm{supp}(\phi)=\bigcup_i\mathrm{supp}(\phi_i)$.

PROOF. For any subset $T\subseteq X$ we put

$$I_T=\{i\in I\ :\ \mathrm{supp}(\phi_i)\cap T\neq\emptyset\}.$$

Consider a point $x\in X$. By assumption, there is an open neighbourhood N of x such that I_N is finite. Now $\phi|_N$ can be written as the finite sum $\sum_{i\in I_N}\phi_i|_N$, so $\phi|_N$ is finite and continuous. It follows (by Proposition 5.9(a)) that ϕ is finite and continuous.

Next, we have $\phi_i\leq\phi$ for all i, so $\bigcup_i\mathrm{supp}(\phi_i)\subseteq\mathrm{supp}(\phi)$. Suppose instead that $x\not\in\bigcup_i\mathrm{supp}(\phi_i)$. Choose N as before. As I_N is finite, the set $M=N\setminus\bigcup_{i\in I_N}\mathrm{supp}(\phi_i)$ is open, and it contains x, and $\phi|_M=0$; it follows that $x\not\in\mathrm{supp}(\phi)$. □

COROLLARY 22.3. [cor-numerable-refinement]
If an open cover has a numerable refinement, then it is itself numerable.

PROOF. Let $(U_i)_{i\in I}$ be an open cover, let $(V_j)_{j\in J}$ be a numerable refinement, and let $(\psi_j)_{j\in J}$ be a partition of unity subordinate to $(V_j)_{j\in J}$. The refinement condition means that we can choose a function $u\colon J\to I$ such that $V_j\subseteq U_{u(j)}$ for all j. Put $J[i]=\{j\ :\ u(j)=i\}$, so J is the disjoint union of the sets $J[i]$. As the family $(\mathrm{supp}(\psi_j))_{j\in J}$ is locally finite, the same is true of the subfamily $(\mathrm{supp}(\psi_j))_{j\in J[i]}$. It follows from the lemma that the function $\phi_i=\sum_{j\in J[i]}\psi_j$ is continuous, with

$$\mathrm{supp}(\phi_i)=\bigcup_{j\in J[i]}\mathrm{supp}(\psi_j)\subseteq\bigcup_{j\in J[i]}V_j\subseteq U_i.$$

For any set $T\subseteq X$, we put

$$\begin{aligned}I_T&=\{i\in I\ :\ \mathrm{supp}(\phi_i)\cap T\neq\emptyset\}\\ J_T&=\{j\in J\ :\ \mathrm{supp}(\psi_j)\cap T\neq\emptyset\}.\end{aligned}$$

It is now clear that $I_T=u(J_T)$. For any point $x\in X$, we can choose an open neighbourhood N such J_N is finite, and it follows that I_N is also finite. This means that the family $(\mathrm{supp}(\phi_i))_{i\in I}$ is locally finite. It is also clear that

$$\sum_{i\in I}\phi_i=\sum_{i\in I}\sum_{j\in J[i]}\psi_j=\sum_{j\in J}\psi_j=1,$$

so we have a partition of unity subordinate to the cover $(U_i)_{i\in I}$ as required. □

It turns out that most naturally occurring spaces are paracompact, but substantial work is required to prove this. The simplest case is as follows.

PROPOSITION 22.4. [prop-compact-paracompact]
Any compact Hausdorff space is paracompact.

PROOF. Any open cover has a finite subcover, which is automatically a locally finite open refinement. □

PROPOSITION 22.5. [prop-exhaustion-paracompact]
Suppose that X is Hausdorff and can be written as $X=\bigcup_n X_n$ where X_n is open, $\overline{X_n}$ is compact and $\overline{X_n}\subseteq X_{n+1}$. Then X is paracompact.

PROOF. Consider an open cover $\mathcal{U}=(U_i)_{i\in I}$. As $\overline{X_n}$ is compact, we can choose a finite set $I_n\subseteq I$ such that $X_n\subseteq\bigcup_{i\in I_n}U_i$. Now put $V_n=X_n\setminus\overline{X_{n-2}}$ (or $V_n=X_n$ for $n\in\{0,1\}$). These sets are clearly open. As $\overline{X_{n-2}}\subseteq X_{n-1}$ we see that $X_n\setminus X_{n-1}\subseteq V_n$, so $X_n\subseteq\bigcup_{m\leq n}V_m$, so the sets V_m cover X. Moreover, for $m\geq n+2$ we have $V_n\subseteq\overline{X_{m-2}}$ and $V_m\subseteq X_n\subseteq X_{m-2}\subseteq\overline{X_{m-2}}^c$ so $V_n\cap V_m=\emptyset$.

Now consider the family $\mathcal{W}=(V_n\cap U_i)_{n\in\mathbb{N},\,i\in I_n}$. It is clear that these sets are open, and that each of them is contained in some set U_i. Consider a point $x\in X$. We have seen that $X=\bigcup_n V_n$, so $x\in V_n$ for some n. This means that $x\in\overline{X_n}$, which is covered by $(U_i)_{i\in I_n}$, so there is a set $V_n\cap U_i$ in $\mathcal{W}$ that contains x. Moreover V_n is an open neighbourhood of x, and using the fact that $V_n\cap V_m=\emptyset$ for $m\geq n+2$ we see that V_n meets only finitely many sets in $\mathcal{W}$. Thus, $\mathcal{W}$ is a locally finite open cover that refines $\mathcal{U}$. □

Corollary 22.6. [`cor-lch-paracompact`]
Suppose that X is locally compact Hausdorff and second countable. Then X is paracompact.

Proof. It will suffice to construct a chain of subsets X_n as in Proposition 22.5. As X is second countable, there is a countable basis β for the topology. Put

$$\gamma = \{U \in \beta \ : \ \overline{U} \text{ is compact } \}.$$

This is a countable collection of precompact open sets; we claim it covers X. Indeed, suppose $x \in X$. Then as X is locally compact, there is a neighbourhood W of x such that $\overline{W}$ is compact. As β is a basis, there is a set $U \in \beta$ with $x \in U \subseteq W$. As $\overline{U}$ is closed in $\overline{W}$, it is compact, so $U \in \gamma$. Thus, for any x there is a set $U \in \gamma$ with $x \in U$ as claimed.

Now enumerate γ as $\gamma = \{V_n \ : \ n > 0\}$. We shall define recursively precompact open sets X_n such that

$$V_n \subseteq X_n \subseteq \overline{X_n} \subseteq X_{n+1}$$

Indeed, we can take $X_0 = \emptyset$. Suppose we have defined sets $X_0, \dots X_n$ satisfying the requirements. Then $\overline{X_n}$ is compact and covered by γ (as the whole space is) so

$$\overline{X_n} \subseteq V_{k_1} \cup \dots V_{k_m}$$

say. We take

$$X_{n+1} = V_{n+1} \cup V_{k_1} \cup \dots V_{k_m}$$

and observe that this is precompact because each V_k is.

This procedure gives us X_n for all n. As $V_n \subseteq X_n$ and the V_n cover X, we see that $\bigcup_n X_n = X$ as required. □

Corollary 22.7. [`cor-manifold-paracompact`]
Every topological manifold is paracompact Hausdorff.

Proof. Manifolds are locally compact Hausdorff by Proposition 20.9, and second countable by definition. □

It is convenient to digress slightly at this point to prove another result with the same hypotheses as Proposition 22.5.

Proposition 22.8. [`prop-exists-proper`]
Suppose that X is Hausdorff and can be written as $X = \bigcup_n X_n$ where X_n is open, $\overline{X_n}$ is compact and $\overline{X_n} \subseteq X_{n+1}$. Then there exists a proper map $f \colon X \to \mathbb{R}$.

Proof. Consider the sets

$$A_n = \overline{X_n} \setminus X_{n-1}$$
$$C_n = X_{n+1} \setminus \overline{X_{n-2}}.$$

Note that A_n and $\overline{C_n}$ are closed and compact, that C_n is open in X, and that $A_n \subseteq C_n \subseteq \overline{C_n}$. As $\overline{C_n}$ is compact Hausdorff, it is normal, so we can choose a set B_n that is open in $\overline{C_n}$ with

$$A_n \subseteq B_n \subseteq \overline{B_n} \subseteq C_n \subseteq \overline{C_n}.$$

As B_n is open in $\overline{C_n}$ and $B_n \subseteq C_n$ and C_n is open in X we see that B_n is open in X. Similarly, the symbol $\overline{B_n}$ above officially refers to the closure in $\overline{C_n}$ but that is the same as the closure in X because $\overline{C_n}$ is closed in X. Now, by Urysohn's Lemma we can find $\phi_n \colon \overline{C_n} \to [0,1]$ with $\phi_n(x) = 1$ for $x \in A_n$ and $\phi_n(x) = 0$ on $\overline{C_n} \setminus B_n$. We extend ϕ_n over all of x by putting $\phi_n(x) = 0$ for $x \notin \overline{C_n}$. Now ϕ_n is continuous on the closed set $\overline{C_n}$, and it is constant on the closed set $X \setminus B_n$, and X is the union of these two closed sets. It follows that $\phi_n \colon X \to [0,1]$ is continuous. Note also that $\operatorname{supp}(\phi_n) \subseteq C_n$, but by construction we have $C_n \cap C_m = \emptyset$ when $|n-m| \geq 3$, so the family of supports is locally finite. This means that the function $\phi = \sum_n n\phi_n$ is continuous. I claim that it is also proper. Indeed, suppose that $K \subseteq \mathbb{R}$ is compact, so $K \subseteq [-n,n]$ say. As $\phi \geq m\phi_m = m$ on A_m, we see that $\phi^{-1}(K) \subseteq A_1 \cup \dots A_n$ which is compact. Moreover, $\phi^{-1}(K)$ is closed by continuity. As a closed subset of a compact set, it is itself compact. □

Theorem 22.9. [thm-metric-paracompact]
Let X be a metric space; then X is paracompact.

The proof, which relies on an elaborate system of preparatory results, will be given after Lemma 22.17.

Definition 22.10. If $V \subseteq X^2$, we write $V^T = \{(y,x) \ : \ (x,y) \in V\}$. We say that V is *symmetric* if $V = V^T$. We also say that V is *reflexive* if it contains the diagonal $\Delta = \{(x,x) \ : \ x \in X\}$.

Proposition 22.11. [prop-closed-refinement-metric]
Let $(U_i)_{i\in I}$ be an open cover of X. Then there is a family of sets $(D_j)_{j\in J}$ such that

(a) *Each set D_j is closed in X.*
(b) *Each set D_j is contained in U_i for some i.*
(c) *X is the union of the sets D_j.*
(d) *For each point $x \in X$ there is an open neighbourhood N such that N meets only finitely many of the sets D_j.*

Moreover, there is a set $V \subseteq X^2$ such that

(e) *V is open.*
(f) *V is symmetric and reflexive.*
(g) *For each $x \in X$, the set $V[x] = \{y \ : \ (x,y) \in V\}$ is contained in U_i for some i.*

Framework of proof. By Theorem 35.27, we may assume that the index set I is well-ordered. We then make the following definitions.

$$\begin{aligned}
A_{ni} &= \{a \in X \ : \ d(a,x) < 2^{-n} \implies x \in U_i\} \\
B_{ni} &= A_{ni} \setminus \bigcup_{j<i} A_{n+1,j} \\
C_{ni} &= \{c \in X \ : \ d(c,b) < 2^{n-3} \text{ for some } b \in B_{ni}\} \\
D_{ni} &= \overline{C_{ni}} \setminus \left(\bigcup_{m<n} \bigcup_{j} C_{mj} \right) \\
V &= \bigcap_{n,i} ((U_i \times U_i) \cup (D_{ni}^c \times D_{ni}^c)) \subseteq X^2.
\end{aligned}$$

The properties of these sets will be established in Lemmas 22.12 to 22.16. In particular, we will show in Lemma 22.15 that the family $(D_{ni})_{n\in\mathbb{N},i\in I}$ has the announced properties (a) to (d), and in Lemma 22.16 that V has properties (e) to (g). $\square$

Lemma 22.12. [lem-pc-A]

(a) *$A_{ni} \subseteq U_i$.*
(b) *We have $U_i = \bigcup_n A_{ni}$, and thus $X = \bigcup_{n,i} A_{ni}$.*
(c) *A_{ni} is closed in X.*
(d) *For all $a \in A_{ni}$ and $a' \notin A_{n+1,i}$ we have $d(a,a') \geq 2^{-n-1}$.*

Proof.

(a) If $a \in A_{ni}$ we can take $x = a$ in the definition to see that $a \in U_i$.
(b) Consider a point $a \in U_i$. As U_i is open, it contains the open ball around a of radius 2^{-n} for some n; we then have $a \in A_{ni}$. This proves that U_i is the union of the sets A_{ni}, and $X = \bigcup_i U_i$ by assumption, so $X = \bigcup_{n,i} A_{ni}$.
(c) Suppose that a lies in the closure of A_{ni}. Consider a point $x \in X$ with $d(a,x) < 2^{-n}$. As $a \in \overline{A_{ni}}$ we can find $a' \in A_{ni}$ with $d(a,a') < 2^{-n} - d(a,x)$. We then find that $d(a',x) < 2^{-n}$, so $x \in U_i$ by the definition of A_{ni}. This shows that a lies in A_{ni}, so A_{ni} is closed.

(d) Suppose that $a \in A_{ni}$ and $a' \notin A_{n+1,i}$. As $a' \notin A_{n+1,i}$ there exists x with $d(a,x) < 2^{-n-1}$ but $x \notin U_n$. As $a \in A_{ni}$ we must have $d(a,x) \geq 2^{-n}$. It follows that

$$d(a,a') \geq d(a,x) - d(a',x) \geq 2^{-n} - 2^{-n-1} = 2^{-n-1}$$

as claimed.

□

Lemma 22.13. [lem-pc-B]

(a) *$B_{ni} \subseteq \overline{B_{ni}} \subseteq A_{ni} \subseteq U_i$.*
(b) *$X = \bigcup_{n,i} B_{ni}$.*
(c) *If $b \in \overline{B_{ni}}$ and $b' \in \overline{B_{nj}}$ with $i \neq j$ then $d(b,b') \geq 2^{-n-1}$.*

Proof.

(a) By definition we have $B_{ni} \subseteq A_{ni}$ and A_{ni} is closed so $\overline{B_{ni}} \subseteq A_{ni}$. We saw in Lemma 22.12(a) that $A_{ni} \subseteq U_i$.
(b) Consider a point $b \in X$. As the sets U_i cover X and I is well-ordered, there is a smallest index i such that $b \in U_i$. As $U_i = \bigcup_n A_{ni}$ we have $b \in A_{ni}$ for some n. If $j < i$ we have $b \notin U_j$, so certainly $b \notin A_{n+1,j}$. We thus have $b \in B_{n,i}$ as required.
(c) First, as the space $\{(x,x') : d(x,x') \geq 2^{-n-1}\}$ is closed in X^2, it will suffice to prove the claim for $b \in B_{ni}$ and $b' \in B_{nj}$. We may also assume by symmetry that $j < i$. This means that $b \notin A_{n+1,j}$ and $b' \in A_{nj}$, so the claim follows from Lemma 22.12(d).

□

Lemma 22.14. [lem-pc-C]

(a) *C_{ni} is open in X.*
(b) *$B_{ni} \subseteq C_{ni} \subseteq \overline{C_{ni}} \subseteq A_{n+1,i} \subseteq U_i$.*
(c) *$X = \bigcup_{n,i} C_{ni}$.*
(d) *If $c \in \overline{C_{ni}}$ and $c' \in \overline{C_{nj}}$ with $i \neq j$ then $d(c,c') \geq 2^{-n-2}$.*

Proof.

(a) This is clear, as C_{ni} is the union of all open balls of radius 2^{-n-3} with centre in B_{ni}.
(b) It is clear from the definition that $B_{ni} \subseteq C_{ni}$. Next, if $c \in C_{ni}$ we can find $b \in B_{ni} \subseteq A_{ni}$ with $d(c,b) < 2^{-n-3}$, but for $a' \notin A_{n+1,i}$ we have $d(a',b) \geq 2^{-n-1}$ by Lemma 22.12(d); so we must have $c \in A_{n+1,i}$. This proves that $C_{ni} \subseteq A_{n+1,i}$, and $A_{n+1,i}$ is closed so $\overline{C_{ni}} \subseteq A_{n+1,i}$. We also have $A_{n+1,i} \subseteq U_i$ by Lemma 22.12(a).
(c) We have just shown that $C_{ni} \supseteq B_{ni}$, and $X = \bigcup_{n,i} B_{ni}$ by Lemma 22.13(b), so $X = \bigcup_{n,i} C_{ni}$.
(d) First, as the space $\{(x,x') : d(x,x') \geq 2^{-n-2}\}$ is closed in X^2, it will suffice to prove the claim for $c \in C_{ni}$ and $c' \in C_{nj}$. We can then find $b \in B_{ni}$ and $b' \in B_{nj}$ with $d(c,b) < 2^{-n-3}$ and $d(c',b') < 2^{-n-3}$. Lemma 22.13(c) tells us that $d(b,b') \geq 2^{-n-1}$. It follows that

$$d(c,c') \geq d(b,b') - d(b,c) - d(b',c') > 2^{-n-1} - 2^{-n-3} - 2^{-n-3} = 2^{-n-2}.$$

as required.

□

Lemma 22.15. [lem-pc-D]

(a) *D_{ni} is closed in X.*
(b) *$D_{ni} \subseteq \overline{C_{ni}} \subseteq A_{n+1,i} \subseteq U_i$.*
(c) *$X = \bigcup_{n,i} D_{ni}$.*
(d) *For each $x \in X$ there is an open neighbourhood N that meets only finitely many of the sets D_{ni}.*

Proof.

(a) As the sets C_{mj} are open, this is clear from the definition.
(b) By definition we have $D_{ni} \subseteq \overline{C_{ni}}$, and the rest is part (b) of the previous lemma.
(c) Consider a point $x \in X$. By part (c) of the previous lemma, there exist pairs (m, j) with $d \in C_{mj}$. If we choose such a pair with m as small as possible, we find that $x \in C_{mj} \setminus \bigcup_{p<m} \bigcup_k C_{pk} \subseteq D_{mj}$.
(d) Choose m and j as in (c), and put

$$N = \{y \in C_{mj} \ : \ d(x, y) < 2^{-m-3}\},$$

which is an open neighbourhood of x. If $n > m$ then $D_{ni} \subseteq C_{mj}^c$ by the definition of D_{ni}, and so N does not meet D_{ni}. Suppose instead that $n \leq m$. For any $y, z \in N$ we have $d(y, z) \leq d(y, x) + d(x, z) < 2^{-m-2}$, but if $y \in \overline{C_{nk}}$ and $z \in \overline{C_{nl}}$ with $k \neq l$ then $d(y, z) \geq 2^{-n-2} \geq 2^{-m-2}$ by Lemma 22.14(d). It follows that N meets at most one of the sets $\overline{C_{ni}}$, and $D_{ni} \subseteq \overline{C_{ni}}$, so N meets at most one of the sets D_{ni}. This applies for $n = 0, 1, \ldots, m$, so N meets at most $m + 1$ of the sets D_{ni} altogether.

□

Lemma 22.16. [`lem-pc-V`]

(e) *V is open.*
(f) *The diagonal $\Delta = \{(x, x) \ : \ x \in X\}$ is a subset of V.*
(g) *For each $x \in X$, the set $V[x] = \{y \ : \ (x, y) \in V\}$ is contained in U_i for some i.*

Proof. (e) Consider a point $(x, y) \in V$. Choose open neighbourhoods N and M of x and y as in Lemma 22.15(d), so there is a finite subset $T \subseteq \mathbb{N} \times I$ such that for $(n, i) \notin T$ we have $D_{ni} \cap N = D_{ni} \cap M = \emptyset$, or equivalently $N \times M \subseteq D_{ni}^c \times D_{ni}^c$. It follows that

$$V \cap (N \times M) = \bigcap_{(n,i) \in T} ((U_i^2 \cup (D_{ni}^c)^2) \cap (N \times M)),$$

which is a *finite* intersection of open sets and so is open. It follows that V is a neighbourhood of (x, y). As (x, y) was arbitrary, this means that V is open.
(f) Now consider a point $(x, x) \in \Delta$. For each (n, i) we have $D_{ni} \subseteq U_i$, so either $x \in U_i$ or $x \in D_{ni}^c$, so $(x, x) \in U_i^2 \cup (D_{ni}^c)^2$. It follows that $(x, x) \in V$, so V is reflexive. It is also clearly symmetric.
(g) Consider an arbitrary point $x \in X$. By Lemma 22.15(c) we can find (n, i) such that $x \in D_{ni}$. Now if $y \in V[x]$ then $(x, y) \in V \subseteq (U_i^2 \cup (D_{ni}^c)^2)$, so either ($x \in U_i$ and $y \in U_i$) or ($x \in D_{ni}^c$ and $y \in D_{ni}^c$). The second possibility is excluded because $x \in D_{ni}$, so $y \in U_i$. Thus $V[x] \subseteq U_i$ as required.

□

Lemma 22.17. [`lem-root-neighbourhood`]
Let V be an open, reflexive and symmetric subset of X^2. Then there is an open, reflexive and symmetric subset $W \subseteq V$ such that whenever $(u, v) \in W$ and $(v, w) \in W$ we have $(u, w) \in V$.

The basic example here is that if $V = \{(x, y) \ : \ d(x, y) < \epsilon\}$, we can take $W = \{(x, y) \ : \ d(x, y) < \epsilon/2\}$.

Proof. We may assume that $d(x, x') < \infty$ for all x and x'. We will use the metric on X^2 given by

$$d((x, y), (x', y')) = \max(d(x, x'), d(y, y')).$$

If $V = X^2$ we can just take $W = V$. We will therefore assume that $V \neq X^2$, so it is meaningful to define $f \colon X \to \mathbb{R}$ by

$$f(x) = d((x, x), V^c) = \inf\{d((x, x), (y, z)) \ : \ (y, z) \notin V\}.$$

As V^c is closed and does not meet Δ, Lemmas 12.53 and 12.54 tell us that f is continuous and strictly positive. We now put

$$W = \{(u, v) \in X^2 \ : \ d((u, v), (x, x)) < f(x)/3 \text{ for some } x \in X\}.$$

This can also be described as the union of the open balls of radius $f(x)/3$ around the points $(x,x) \in \Delta$, so it is clearly open, reflexive and symmetric. Suppose that $(u,v) \in W$ and $(v,w) \in W$, so there exist $x,y \in X$ such that $d((u,v),(x,x)) < f(x)/3$ and $d((v,w),(y,y)) < f(y)/3$, or equivalently

$$d(u,x) < f(x)/3 \qquad d(v,x) < f(x)/3 \qquad d(v,y) < f(y)/3 \qquad d(w,y) < f(y)/3.$$

By symmetry we may assume that $f(x) \leq f(y)$. It then follows from the above that $d(u,y) < 2f(x)/3 + f(y)/3 \leq f(y)$, and thus that $d((u,w),(y,y)) < f(y)$, so $(u,w) \in V$. □

Proof of Theorem 22.9. Consider an open cover $(U_i)_{i\in I}$ of X. We can apply Proposition 22.11 to this cover, giving a family of closed sets $(D_j)_{j\in J}$ with the properties (a) to (d) described there. Property (d) means that we can cover X by a family $(N_k)_{k\in K}$ of open sets such that each N_k meets only finitely many of the sets D_j. We can then apply Proposition 22.11 again to this new family, giving an open, reflexive and symmetric set $V \subseteq X^2$ such that each open set $V[x]$ is contained in some N_k. Choose W as in Lemma 22.17. For each j choose $\phi(j) \in I$ such that $D_j \subseteq U_{\phi(j)}$, and put

$$E_j = \{x \in U_{\phi(j)} \;:\; (x,y) \in W \text{ for some } y \in D_j\} = U_{\phi(j)} \cap \bigcup_{y \in D_j} W[y].$$

We claim that these sets form a locally finite open cover of X refining the original cover $(U_i)_{i\in I}$. Indeed, it is clear that E_j is open and contained in $U_{\phi(j)}$. Note that if $x \in D_j$ then $x \in U_{\phi(j)}$ and also $(x,x) \in \Delta \subseteq W$ so $x \in E_j$. This means that $D_j \subseteq E_j$ and the sets D_j cover X, so the sets E_j also cover X. This just leaves local finiteness. Consider a point $z \in X$, so the set $W[z]$ is an open neighbourhood of z. Suppose that $W[z]$ meets E_j, say $x \in E_j \cap W[z]$. As $x \in E_j$ there exists $y \in D_j$ such that $(x,y) \in W$. As $x \in W[z]$ we also have $(z,x) \in W$. By the defining property of W we therefore have $(z,y) \in V$, so $y \in V[z] \cap D_j$, so $V[z] \cap D_j \neq \emptyset$. However, the defining property of V means that $V[z]$ meets only finitely many of the sets D_j. It follows that $W[z]$ meets only finitely many of the sets E_j, as required. □

Proposition 22.18. [`prop-paracompact-normal`]
If X is paracompact and Hausdorff then it is normal.

We will deduce this from the following weaker statement:

Lemma 22.19. [`lem-paracompact-regular`]
If X is paracompact and Hausdorff then X is regular.

Proof. Let Y be a closed set in X, and let z be a point in $X \setminus Y$. For each $y \in Y$ we can choose disjoint open sets U_y and V_y with $y \in U_y$ and $z \in V_y$ (so $z \notin \overline{U_y}$). The sets U_y together with $X \setminus Y$ give an open cover of X, which must have a locally finite open refinement. This will consists of a locally finite family $(A_i)_{i\in I}$ of open subsets of $X \setminus Y$, together with another locally finite family of open sets $(B_j)_{j\in J}$ such that each $B_j \subseteq U_{y(j)}$ for some $y(j) \in Y$. Choose an open neighbourhood N of z such that the set $J_0 = \{j \;:\; N \cap B_j \neq \emptyset\}$ is finite. Put $M = N \cap \bigcap_{j\in J_0} \overline{U_{y(j)}}^c$, and observe that this is an open neighbourhood of z. Put $L = \bigcup_{j\in J} B_j$, and observe that this is open. Moreover, as the sets A_i and B_j cover all of X, and none of the sets A_i meet Y, we see that $Y \subseteq L$. Also, the intersection $L \cap M$ is the union of the sets $B_j \cap M \subseteq B_j \cap N$. If $j \notin J_0$ then this is empty. If $j \in J_0$ then $B_j \subseteq U_{y(j)}$ and $M \subseteq \overline{U_{y(j)}}^c$ so again $B_j \cap M = \emptyset$. We conclude that $L \cap M = \emptyset$, so we have disjoint open sets containing Y and x respectively as required. □

Proof of Proposition 22.18. Let Y and Z be disjoint closed sets in X. As X is regular, for each $z \in Z$ we can choose disjoint open sets U_z and V_z with $Y \subseteq U_z$ and $z \in V_z$. The sets V_z together with Z^c therefore give an open cover of X, which must have a locally finite refinement. This will consists of a locally finite family $(A_i)_{i\in I}$ of open subsets of Z^c, together with another locally finite family of open sets $(B_j)_{j\in J}$ such that each B_j is contained in $V_{z(j)}$ for some $z(j) \in Z$. Put $M = \bigcup_{j\in J} B_j$, and observe that this is open and contains Z. Put $L = \bigcap_{j\in J} \overline{B_j}^c$, and observe that this is disjoint from M. If $y \in Y$ then for each $j \in J$ we have $B_j \subseteq V_{z(j)} \subseteq U_{z(j)}^c$ and $U_{z(j)}^c$ is closed so $\overline{B_j} \subseteq U_{z(j)}^c$, so $Y \subseteq U_{z(j)} \subseteq \overline{B_j}^c$. It follows that $Y \subseteq L$. Finally, suppose that $x \in L$. We can choose an open neighbourhood N of x such that the set $J_0 = \{j \in J \;:\; N \cap B_j \neq \emptyset\}$ is finite. Put $L_0 = \bigcap_{j\in J_0} \overline{B_j}^c$ and $N_0 = N \cap L_0$. As J_0 is finite we see that

L_0 and N_0 are open. It is clear that $L \subseteq L_0$ and so $N \cap L \subseteq N_0$. We claim that the opposite inclusion also holds. Indeed, if $j \in J_0$ then $N_0 \subseteq \overline{B_j}^c$ by the definition of L_0. If $j \in J \setminus J_0$ then $N \cap B_j = \emptyset$ by the definition of J_0, so B_j is contained in the closed set N^c, so $\overline{B_j} \subseteq N^c$, so $N_0 \subseteq N \subseteq \overline{B_j}^c$ again. Thus $N_0 \subseteq \overline{B_j}^c$ for all $j \in J$, which means that $N_0 \subseteq L$ as required. As $N_0 = N \cap L$ we see that $x \in N_0 \subseteq L$, so x is in the interior of L. As x was an arbitrary point of L, this proves that L is open. Now L and M are disjoint open sets containing Y and Z respectively, as required for normality. □

Proposition 22.20. [`prop-closed-refinement`]
Let X be a normal space, and let $(U_i)_{i \in I}$ be a locally finite open cover. Then there exist open subsets V_i such that $V_i \subseteq \overline{V_i} \subseteq U_i$ and $\bigcup_i V_i = X$.

Proof. The case where $I = \{1, \ldots, n\}$ was covered by Lemma 14.7. More generally, Theorem 35.27 tells us that we can choose a well-ordering of I, and in view of Proposition 35.28 we may then assume that $I = S(\lambda) = \{\alpha \ : \ \alpha < \lambda\}$ for some ordinal λ. We can then follow the method of Lemma 14.7 but using transfinite recursion. More specifically, put $U_\alpha^+ = \bigcup_{\beta > \alpha} U_\beta$, so $U_\alpha^+ \supseteq U_\beta^+$ whenever $\alpha \leq \beta$. We will define sets V_β (for all $\beta < \lambda$) such that the following statements hold for all β:

A_β: $\overline{V_\beta} \subseteq U_\beta$
B_β: $\bigcup_{\alpha \leq \beta} V_\alpha \cup U_\beta^+ = X$.

Suppose that V_α has been defined for all $\alpha < \beta$, and that A_α and B_α hold for all $\alpha < \beta$. Put $U'_\beta = U_\beta^+ \cup \bigcup_{\alpha < \beta} V_\alpha$. We claim that $U'_\beta \cup U_\beta = X$. To see this, consider a point $x \in X$ and put $J_x = \{\mu < \lambda \ : \ x \in U_\mu\}$, which is finite by our local finiteness assumption. It therefore has a largest element, say μ_x.

- If $\mu_x < \beta$ then B_{μ_x} holds, which means that either $x \in V_\alpha$ for some $\alpha \leq \mu_x < \beta$, or $x \in U_\gamma$ for some $\gamma > \mu_x$. The latter is impossible by the definition of μ_x, so we must have $x \in \bigcup_{\alpha < \beta} V_\alpha \subseteq U'_\beta$.
- If $\mu_x = \beta$ then $x \in U_\beta$.
- If $\mu_x > \beta$ then $x \in U_{\mu_x} \subseteq U_\beta^+ \subseteq U'_\beta$.

In all cases we have $x \in U'_\beta \cup U_\beta$, as claimed. This means that the set $F'_\beta = X \setminus U'_\beta$ is closed and contained in the open set U_β. By Proposition 14.6(b), we can choose an open set V_β with $F'_\beta \subseteq V_\beta \subseteq \overline{V_\beta} \subseteq U_\beta$. As $\overline{V_\beta} \subseteq U_\beta$, we see that statement A_β holds. As $F'_\beta \subseteq V_\beta$, we see that $V_\beta \cup U'_\beta = X$, or equivalently that statement B_β also holds. This completes the recursion setup, so we have sets V_β as indicated for all β.

Now note that for all $x \in X$ the statement B_{μ_x} implies that $x \in V_\alpha$ for some $\alpha \leq \mu_x$, so the sets V_α cover X. Given this, we see that the sets $\overline{V_k}$ give a closed cover of X refining the original open cover $(U_i)_{i \in I}$. As the original cover was locally finite and the refining cover has the same indexing, we see that $(\overline{V_i})_{i \in I}$ is also locally finite. □

Corollary 22.21. [`cor-normal-numerable`]
Let X be a normal space. Then every locally finite open cover is numerable.

Proof. Let $(U_i)_{i \in I}$ be a locally finite open cover. Choose open sets V_i as in the proposition. By Proposition 14.6(b) we can then choose open sets W_i with $\overline{V_i} \subseteq W_i \subseteq \overline{W_i} \subseteq U_i$. Using Urysohn's Lemma we can then choose continuous functions $\psi_i \colon X \to [0, 1]$ with $\psi_i = 1$ on $\overline{V_i}$ and $\psi_i = 0$ on $X \setminus W_i$. This means that $\mathrm{supp}(\psi_i) \subseteq \overline{W_i} \subseteq U_i$, so in particular the family of supports is locally finite. Put $\psi(x) = \sum_i \psi_i(x)$. Using Lemma 22.2 we see that this is finite and continuous. As the sets V_i cover X, we also have $\psi(x) \geq 1$ for all x. We can thus define $\phi_i = \psi_i / \psi$, and we find that these functions give the required partition of unity. □

Theorem 22.22. [`thm-paracompact-numerable`]
Let X be a paracompact Hausdorff space. Then every open cover of X is numerable.

Proof. We know from Proposition 22.18 that X is normal, so every locally finite open cover is numerable by Corollary 22.21. Every open cover has a locally finite refinement (by the definition of paracompactness) and so is numerable by Corollary 22.3. □

23. CGWH spaces

In this section, we introduce a certain full subcategory **CGWH** of the category of topological spaces. The great majority of spaces arising in nature will lie in **CGWH**, and we will also show that limits, colimits and mapping spaces have better behaviour in **CGWH** than in **Spaces**. Because of this, it is standard to use **CGWH** as a foundation for work in homotopy theory, as we will explain later on.

In order to set up the definitions, it will be convenient to think of a topology as being specified by its *closed* sets, as in Remark 2.14. We will typically write ζ for the collection of all closed sets.

Definition 23.1. [`defn-CG`]
Let X be a topological space. A *test map* for X is a continuous map $u\colon K \to X$, where K is a compact Hausdorff space. We say that a subset $Y \subseteq X$ is *k-closed* if $u^{-1}(Y)$ is closed in K for every such test map. We write $k(\zeta)$ for the collection of k-closed sets. It is easy to check that this is a topology on X, and that $\zeta \subseteq k(\zeta)$. We write kX for the set X equipped with the topology $k(\zeta)$. We say that X is *compactly generated* if $kX = X$. We write **CG** for the category of compactly generated spaces and continuous maps.

Definition 23.2. [`defn-WH`]
A topological space X is *weakly Hausdorff* if for every test map $u\colon K \to X$, the image $u(K)$ is closed in X. We write **WH** for the category of weakly Hausdorff spaces and continuous maps.

The connection with the ordinary Hausdorff condition will not become fully apparent until we have proved some preparatory results. For the best available statement, the reader should compare Proposition 23.27 with Proposition 6.6.

Definition 23.3. [`defn-CGWH`]
We write **CGWH** for the full subcategory of **Spaces** consisting of the spaces that are both compactly generated and weakly Hausdorff.

The next two propositions imply that the vast majority of spaces in common use are CGWH.

Proposition 23.4. [`prop-hausdorff`]
Any Hausdorff space is weakly Hausdorff.

Proof. If X is Hausdorff and $u\colon K \to X$ is a test map then $u(K)$ is a compact subset of a Hausdorff space and thus is closed by Proposition 10.16. $\square$

Lemma 23.5. [`lem-WH-omni`]
Suppose that X is weakly Hausdorff.

(a) *Every point is closed (so X is T_1).*
(b) *If $u\colon K \to X$ is a test map then $u(K)$ is compact Hausdorff with respect to the subspace topology.*
(c) *A subset $Y \subseteq X$ is k-closed if and only if $Y \cap K$ is closed in K for every subset $K \subseteq X$ that is compact Hausdorff with respect to the subspace topology.*

Proof. For part (a), take K to be a single point in Definition 23.2. Next, let $u\colon K \to X$ be a test map, and put $L = u(K)$, which is closed in X by hypothesis. If $F \subseteq K$ is closed then it is compact Hausdorff so $u(F)$ is also closed in X and thus in K, so $u\colon K \to L$ is a closed map. If $a, b \in K$ and $a \neq b$ then $u^{-1}\{a\}$ and $u^{-1}\{b\}$ are disjoint closed subspaces of the compact Hausdorff space K, so they have disjoint neighbourhoods, say U and V. Put

$$U' = \{x \in X \ :\ u^{-1}\{x\} \subseteq U\} = \{x \in X \ :\ u^{-1}\{x\} \cap (K \setminus U) = \emptyset\} = X \setminus u(K \setminus U)$$
$$V' = \{x \in X \ :\ u^{-1}\{x\} \subseteq V\} = \{x \in X \ :\ u^{-1}\{x\} \cap (K \setminus V) = \emptyset\} = X \setminus u(K \setminus V).$$

From the first description we see that $a \in U'$ and $b \in V'$ and $U' \cap V' \cap L = \emptyset$. We also note that $K \setminus U$ is compact so $u(K \setminus U)$ is closed in X so $U' = X \setminus u(K \setminus U)$ is open in X. Similarly, V' is open, so we have found disjoint open neighbourhoods of a and b in L. This shows that L is Hausdorff. This proves (b), and (c) follows easily. $\square$

Proposition 23.6. [`prop-metric-cg`]
Every first countable space is compactly generated. In particular, every metric space is CGWH.

Proof. Let X be a first countable space, and Y a k-closed subset. We must show that Y is closed. By Proposition 2.72, it will suffice to prove that Y is sequentially closed. Consider a sequence $(y_n)_{n\in\mathbb{N}}$ in Y converging to a point $x \in X$. We can define a map $u\colon \mathbb{N}_\infty \to X$ by $u(n) = y_n$ for $n \in \mathbb{N}$, and $u(\infty) = x$. This is continuous (for the one-point compactification topology on $\mathbb{N}_\infty$) by Proposition 18.25. As $\mathbb{N}_\infty$ is compact Hausdorff and u is continuous and Y is k-closed, we see that $u^{-1}(Y)$ is closed in $\mathbb{N}_\infty$. We have $\mathbb{N} \subseteq u^{-1}(Y)$ by assumption and $\mathbb{N}$ is dense in $\mathbb{N}_\infty$ so $u^{-1}(Y) = \mathbb{N}_\infty$. In particular we have $\infty \in u^{-1}(Y)$, so $x = u(\infty) \in Y$. This means that Y is sequentially closed as required.

If X is a metric space then it first countable and Hausdorff, so it is compactly generated by the argument above, and weakly Hausdorff by Proposition 23.4. $\square$

Proposition 23.7. [prop-lch-cg]
Every locally compact Hausdorff space is CGWH.

Proof. Let X be a locally compact Hausdorff space, and Y a k-closed subset. Suppose that $x \in \overline{Y}$; we need to show that $x \in Y$. As X is locally compact, x has a neighbourhood U such that the set $K = \overline{U}$ is compact. If V is a neighbourhood of x then so is $V \cap K$, and $x \in \overline{Y}$, so $V \cap K \cap Y \neq \emptyset$; this shows that $x \in \overline{K \cap Y}$. On the other hand, as Y is k-closed and the inclusion $j\colon K \rightarrowtail X$ is continuous we see that $K \cap Y = j^{-1}(Y)$ is closed in K. Thus $x \in Y$ as required.

This shows that X is compactly generated, and it is weakly Hausdorff by Proposition 23.4. $\square$

Lemma 23.8. [lem-k-idempotent]
Let K be a compact Hausdorff space, let X be an arbitrary space (with topology ζ) and let $u\colon K \to X$ be a function. Then u is continuous with respect to ζ iff it is continuous with respect to $k(\zeta)$.

Proof. First suppose that u is continuous with respect to $k(\zeta)$. This means that for every $F \in k(\zeta)$, the preimage $u^{-1}(F)$ is closed in K. As $\zeta \subseteq k(\zeta)$, we deduce that u is also continuous with respect to ζ. Conversely, suppose that u is continuous with respect to ζ. This means that u is one of the maps involved in the definition of $k(\zeta)$, so it is tautological that $u^{-1}(F)$ is closed for all $F \in k(\zeta)$. Thus, u is continuous with respect to $k(\zeta)$ as well. $\square$

Corollary 23.9. [cor-k-idempotent]
For any space X we have $k^2X = kX$ and thus kX is compactly generated.

Proof. Let ζ be the original topology on X. By definition we have $F \in k^2(\zeta)$ iff $u^{-1}(F)$ is closed for every compact Hausdorff K and every map $u\colon K \to X$ that is continuous with respect to $k(\zeta)$. We have seen that these are precisely the same as the maps that are continuous with respect to ζ, so $k^2(\zeta) = k(\zeta)$ as required. $\square$

Corollary 23.10. [cor-CGWH-adjoint]
Let X and Y be topological spaces, and consider an arbitrary function $f\colon X \to Y$.

(a) *If f is continuous as a map $X \to Y$, then it is also continuous as a map $kX \to kY$.*
(b) *In particular, if X is compactly generated and f is continuous as a map $X \to Y$, then it is continuous as a map $X \to kY$.*
(c) *Conversely, if X is compactly generated and f is continuous as a map $X \to kY$, then it is continuous as a map $X \to Y$.*

Proof. Let ζ be the original topology on X, and let ξ be the original topology on Y.

(a) Suppose that f is continuous for ζ and ξ. Consider a set $G \in k(\xi)$ and the preimage $f^{-1}(G) \subseteq X$. For any test map $u\colon K \to X$ we note that $fu\colon K \to Y$ is a test map for Y, so $(fu)^{-1}(G)$ is closed in K. This set is the same as $u^{-1}(f^{-1}(G))$, so we see that $f^{-1}(G) \in k(\zeta)$. This means that f is continuous for $k(\zeta)$ and $k(\xi)$.
(b) If X is compactly generated then $kX = X$ so (b) is a special case of (a).
(c) If f is continuous as a map $X \to kY$, then we see that $f^{-1}(G) \in \zeta$ for all $G \in k(\xi)$. As $\xi \subseteq k(\xi)$ this means that $f^{-1}(G) \in \zeta$ for all $G \in \xi$, so f is continuous as a map $X \to Y$.

$\square$

Corollary 23.11. *Let $j\colon$ **CG** $\to$ **Spaces** be the inclusion functor; then k gives a functor **Spaces** $\to$ **CG** that is right adjoint to j.*

Proof. Parts (b) and (c) of the previous corollary give $\mathbf{CG}(X, kY) = \mathbf{Spaces}(jX, Y)$. □

Proposition 23.12. [prop-test]
Let X be a compactly generated space and Y an arbitrary space. Then a function $f\colon X \to Y$ is continuous if and only if $fu\colon K \to Y$ is continuous for all test maps $u\colon K \to X$.

Proof. Suppose that fu is continuous for all K and u. Let $F \subseteq Y$ be closed. For any K and u, the continuity of uf means that $f^{-1}(u^{-1}F) = (uf)^{-1}F$ is closed. This shows that $f^{-1}F$ is k-closed and thus closed, which means that f is continuous. The opposite implication is trivial. □

Proposition 23.13. [prop-topologies]
If ζ and ξ are topologies on X with $\zeta \subseteq \xi$ then $k(\zeta) \subseteq k(\xi)$.

Proof. Consider a set $F \in k(\zeta)$; we must show that F is k-closed with respect to ξ. Consider a compact Hausdorff space K and a ξ-continuous map $u\colon K \to X$. As $\zeta \subseteq \xi$ we see that u is also ζ-continuous. As F is k-closed with respect to ζ we see that $u^{-1}(F)$ is closed in K. Thus $F \in k(\xi)$. □

23.1. Products, coproducts and quotients.

Proposition 23.14. [prop-quot-cg]
If X is a compactly generated space and E is an equivalence relation on X then the quotient $Y = X/E$ is compactly generated.

Proof. Write $q\colon X \to Y$ be the quotient map. Let Z be a k-closed subset of Y. By Corollary 23.10, the function q is continuous when thought of as a map $X \to kY$, so $q^{-1}(Z)$ is closed in X. By the definition of the quotient topology, we conclude that Z is closed in Y. □

Proposition 23.15. [prop-coprod-CGWH]
If $(X_i)_{i\in I}$ is a family of compactly generated spaces then their disjoint union $X = \coprod_i X_i$ is compactly generated. Moreover, if each X_i is also weakly Hausdorff, then the same is true of X.

Proof. Let $Z \subseteq X$ be k-closed. Then Z has the form $\coprod_i Z_i$, where $Z_i = Z \cap X_i$, and it is sufficient to check that Z_i is closed in X_i. As X_i is compactly generated, it is enough to check that Z_i is k-closed in X_i. Consider a test map $u\colon K \to X_i$. Then the composite $v = (K \xrightarrow{u} X_i \rightarrowtail X)$ is continuous and $u^{-1}(Z_i) = v^{-1}(Z)$, which is closed because Z is k-closed in X.

Now suppose that each X_i is also weakly Hausdorff. Consider a test map $u\colon K \to X$. As X_i is both open and closed in X, we see that the preimage $K_i = u^{-1}(X_i)$ is both open and closed in K. The sets K_i form a disjoint open cover of the compact space K, so all but finitely many of them must be empty. Moreover, as K_i is also closed it is compact, and X_i is weakly Hausdorff so $u(K_i) \subseteq X_i$ is closed in X_i. The image $u(K)$ is the disjoint union of the sets $u(K_i)$, so it is also closed (by the definition of the coproduct topology). □

Corollary 23.16. [cor-coprod-CGWH]
*The category **CG** has coproducts, given by disjoint unions with the usual coproduct topology. Moreover, the same construction also gives coproducts in **CGWH**.* □

Definition 23.17. [defn-prod-CG]
Given two spaces X and Y, we shall write $X \times_0 Y$ for the product space equipped with the usual product topology. This need not be compactly generated even if X and Y are. We thus define $X \times Y = k(X \times_0 Y)$. Similarly, given an indexed family of (possibly infinitely many) spaces X_i we write $\prod_{0,i} X_i$ for their product under the usual topology and $\prod_i X_i = k \prod_{0,i} X_i$.

Proposition 23.18. [prop-prod-CG]
*Let $(X_i)_{i\in I}$ be a family of compactly generated spaces. Then the projection maps $\pi_i\colon \prod_i X_i \to X_i$ are continuous. Moreover, for any compactly generated space W, a map $f\colon W \to \prod_i X_i$ is continuous if and only if each component $f_i = \pi_i \circ f$ is continuous. (This means that $\prod_i X_i$ is the product of the objects X_i in the category **CG**.)*

PROOF. First, the projections π_i are continuous as maps $\prod_{0,i} X_i \to X_i$, so they are continuous as maps $\prod_i X_i \to X_i$ by Corollary 23.10(a). Thus, if $f\colon W \to \prod_i X_i$ is continuous, the same is true of the composites $f_i = \pi_i \circ f$. Conversely, suppose we start with the assumption that all the maps f_i are continuous and that $kW = W$. It follows from Proposition 5.16 that f is continuous as a map $W \to \prod_{0,i} X_i$, so it is also continuous as a map $W \to \prod_i X_i$ by Corollary 23.10(b). □

It is also true that CG-products of CGWH spaces are again CGWH, so the category **CGWH** has products. However, the most efficient proof is somewhat indirect; it will be given as Corollary 23.29.

PROPOSITION 23.19. [prop-prod-lch]
If X is a locally compact Hausdorff space and Y is compactly generated then $X \times_0 Y$ is compactly generated and thus $X \times Y = X \times_0 Y$.

PROOF. In this proof we will say that subsets of $X \times_0 Y$ are open or closed if they have that property with respect to the ordinary product topology. Suppose that $Z \subseteq X \times_0 Y$ is k-closed; we need to check that it is closed in the ordinary product topology. Suppose that $(x, y) \notin Z$. The map $i_y\colon x' \mapsto (x', y)$ is continuous with respect to the product topology, and hence also with respect to the compactly generated topology, by Corollary 23.10. It follows that the set $i_y^{-1}(Z) = \{x' \in X \;:\; (x', y) \in Z\}$ is closed in X. Thus, by Proposition 18.4, we can choose a precompact open neighbourhood U of x in X such that $\overline{U} \cap i_y^{-1} Z = \emptyset$, or equivalently $\overline{U} \times \{y\} \subseteq Z^c$. Now put

$$V = \{y' \in Y \;:\; \overline{U} \times \{y'\} \subseteq Z^c,$$

so $y \in V$. We claim that V is open in Y. As Y is compactly generated, it will suffice to show that $u^{-1}(V)$ is open in K for every test map $u\colon K \to Y$. Note that the map $1 \times u\colon \overline{U} \times K \to X \times_0 Y$ is a test map, and Z is assumed to be k-closed, so the preimage $Z' = (1 \times u)^{-1}(Z)$ is closed in $\overline{U} \times K$ and thus is compact. Let $\pi\colon \overline{U} \times K \to K$ be the projection, so $\pi(Z')$ is compact and therefore closed, so $K \setminus \pi(Z')$ is open. Moreover, we have $b \in K \setminus \pi(Z')$ iff $((a, b) \notin Z'$ for all $a \in \overline{U})$ iff $((a, u(b)) \notin Z$ for all $a \in \overline{U})$ iff $\overline{U} \times \{u(b)\} \subseteq Z^c$ iff $b \in u^{-1}(V)$, so $u^{-1}(V)$ is open as required. This completes the proof that V is open in Y, so $U \times V$ is an open neighbourhood of (x, y) in $X \times_0 Y$ which does not meet Z. It follows that Z is closed in $X \times_0 Y$, as required. □

PROPOSITION 23.20. [prop-countable-prod]
Suppose that X and Y are both first countable (and thus compactly generated by Proposition 23.6). Then $X \times_0 Y$ is also first countable and therefore compactly generated, so $X \times Y = X \times_0 Y$

PROOF. Given a point $(x, y) \in X \times Y$ we choose a countable basis of neighbourhoods U_i for x in X, and a countable basis of neighbourhoods V_j for $y \in Y$. Then the sets $U_i \times V_j$ give a countable basis of neighbourhoods for $(x, y) \in X \times Y$. □

23.2. Cartesian closure. In this section, all spaces are assumed to be compactly generated unless otherwise specified. We next introduce a topology on the set of all continuous maps from X to Y.

DEFINITION 23.21. [defn-compact-open]
Let X and Y be compactly generated spaces. For any test map $u\colon K \to X$ and any open set $U \subseteq Y$, we write

$$W(u, K, U) = \{ \text{ continuous maps } f\colon X \to Y \;:\; fu(K) \subseteq U\}.$$

If K is a compact Hausdorff subspace of X and $u\colon K \to X$ is the inclusion then we write $W(K, U)$ for $W(u, K, U)$. We write $C_0(X, Y)$ for the set of maps $f\colon X \to Y$, equipped with the smallest topology for which the sets $W(u, K, U)$ are open (this is called the *compact-open topology*). We also write $C(X, Y) = kC_0(X, Y)$.

REMARK 23.22. [rem-CXZ-closed]
If Z is a closed subspace of Y then $C(X, Z)$ is closed in $C(X, Y)$, because $C(X, Z) = \bigcap_x W(\{x\}, Z^c)^c$.

LEMMA 23.23. [lem-star-cont]
If $g\colon Y \to Z$ is continuous, then so is the map $g_*\colon C(X, Y) \to C(X, Z)$ defined by $g_*(t) = g \circ t$. If $f\colon W \to X$ is continuous then so is the map $f^*\colon C(X, Y) \to C(W, Y)$ defined by $f^*(t) = t \circ f$.

Proof. Consider a test map $u: K \to X$ and an open subset $U \subseteq Z$. We then have

$$(g_*)^{-1}W(u,K,U) = \{t: X \to Y \ : \ gtu(K) \subseteq U\} = \{t: X \to Y \ : \ tu(K) \subseteq g^{-1}(U)\} = W(u,K,g^{-1}U).$$

It follows easily that the preimage of any open subset of $C_0(X,Z)$ is open in $C_0(X,Y)$ and thus in $C(X,Y)$; by Corollary 23.10 we conclude that g_* is continuous as a map $C(X,Y) \to C(X,Z)$.

Now consider a test map $v: L \to W$, and an open subset $V \subseteq Y$. Then

$$(f^*)^{-1}W(v,L,V) = \{t: X \to Y \ : \ tfv(L) \subseteq V\} = W(fv,L,V).$$

By the same logic, we conclude that $f^*: C(X,Y) \to C(W,Y)$ is continuous. □

Proposition 23.24. [prop-ev-inj]
Define functions $\mathrm{ev}_{X,Y}: X \times C(X,Y) \to Y$ *and* $\mathrm{inj}_{X,Y}: Y \to C(X, X \times Y)$ *by*

$$\mathrm{ev}(x,f) = f(x)$$
$$\mathrm{inj}(y)(x) = (x,y).$$

Then ev *and* inj *are continuous.*

Proof. We first consider inj. By Corollary 23.10, it is enough to show that inj is continuous as a map $Y \to C_0(X, X \times Y)$, or equivalently that $\mathrm{inj}^{-1} W(u,K,U)$ is open in Y for every $u: K \to X$ and every open set $U \subseteq X \times Y$. As Y is compactly generated, it is equivalent to check that $v^{-1}\,\mathrm{inj}^{-1} W(u,K,U)$ is open in L for every test map $v: L \to Y$. Note that $u \times v: K \times L \to X \times Y$ is a test map, so $(u \times v)^{-1}(U)$ is open in $K \times L$, so $\{b \in L \ : \ K \times \{b\} \subseteq (u \times v)^{-1}U\}$ is open in L by the Tube Lemma. It is easy to check that this set is the same as $v^{-1}\,\mathrm{inj}^{-1} W(u,K,U)$, which completes the proof.

We next consider the evaluation map $\mathrm{ev}: X \times C(X,Y) \to Y$. Consider an open set $U \subseteq Y$, and a test map $u: K \to X \times C(X,Y)$. It will be enough to show that $V = u^{-1}\,\mathrm{ev}^{-1} U$ is open in K. Let $v: K \to X$ and $w: K \to C(X,Y)$ be the two components of u, so $V = \{a \in K \ : \ w(a)(v(a)) \in U\}$. Suppose that $a \in V$. As $w(a) \circ v: K \to Y$ is continuous, we can choose a compact neighbourhood L of a in K such that $w(a)(v(L)) \subseteq U$. This means that $w(a) \in W(v,L,U) \subseteq C(X,Y)$. As $w: K \to C(X,Y)$ is continous, the set $N = w^{-1}(W(v,L,U))$ is a neighbourhood of a in K. If $b \in N \cap L$ then $w(b)(v(b)) \in w(b)(v(L)) \subseteq U$, so $b \in V$. Thus the neighbourhood $N \cap L$ of a is contained in V. This shows that V is open, as required. □

Proposition 23.25. [prop-adj-homeo]
There is a natural homeomorphism $\mathrm{adj}_{X,Y,Z}: C(X, C(Y,Z)) \to C(X \times Y, Z)$ *given by* $\mathrm{adj}(f)(x,y) = f(x)(y)$. *Thus, the category* ***CG*** *is cartesian closed.*

Proof. Write $D(X,Y)$ for the set of all (possibly discontinuous) functions $X \to Y$. It is clear that there is a bijection between functions $f: X \to D(Y,Z)$ and functions $g: X \times Y \to Z$ defined by $g(x,y) = f(x)(y)$. We first claim that g is continuous if and only if

(1) $f(x): Y \to Z$ is continuous for each $x \in X$, so that f can be considered as a function $X \to C(X,Y)$.
(2) f is continuous when considered as a function $X \to C(X,Y)$.

Indeed, if f satisfies these conditions then g is the composite

$$X \times Y \xrightarrow{f \times 1} C(Y,Z) \times Y \xrightarrow{\mathrm{ev}} Y$$

which is continuous by Proposition 23.24. Conversely, suppose that g is continuous. If $x \in X$ then we have a continuous map $i_x: Y \to X \times Y$ defined by $i_x(y) = (x,y)$ and $f(x) = g \circ i_x$ so $f(x)$ is continuous. Moreover, f is the composite

$$X \xrightarrow{\mathrm{inj}} C(Y, X \times Y) \xrightarrow{g_*} C(Y,Z)$$

which is continuous by Lemma 23.23 and Proposition 23.24.

It follows from the above that we have a bijection $\mathrm{adj}: C(X, C(Y,Z)) \to C(X \times Y, Z)$, which already means that **CG** is cartesian closed.

However we still need to show that adj is a homeomorphism. This is true by a purely formal argument, which we now explain. We know that the evaluation map $\mathrm{ev}_{X,C(Y,Z)}: X \times C(X, C(Y,Z)) \to C(Y,Z)$ is continuous, as is $\mathrm{ev}_{Y,Z}: Y \times C(Y,Z) \to Z$. It follows that the composite

$$\mathrm{ev}_{Y,Z} \circ (1_Y \times \mathrm{ev}_{X,C(Y,Z)}): Y \times X \times C(X, C(Y,Z)) \to Z$$

is continuous. It follows that the adjoint map

$$C(X, C(Y, Z)) \to C(X \times Y, Z)$$

is also continuous. However, this last map is just adj itself. Thus, adj is continuous.

Similarly, we know that the evaluation map $X \times Y \times C(X \times Y, Z) \to Z$ is continuous. It follows that the adjoint map $X \times C(X \times Y, Z) \to C(Y, Z)$ is continuous. Applying the same argument again, we see that the adjoint map $C(X \times Y, Z) \to C(X, C(Y, Z))$ is continuous. However, this last map is just the inverse of adj. This proves that adj is a homeomorphism.

We give a second proof which may be found more conceptual. For any space W, we have natural bijections

$$\begin{aligned} C(W, C(X, C(Y, Z))) &\xrightarrow{\mathrm{adj}_{W,X,C(Y,Z)}} C(W \times X, C(Y, Z)) \\ &\xrightarrow{\mathrm{adj}_{W\times X,Y,Z}} C(W \times X \times Y, Z) \\ &\xleftarrow{\mathrm{adj}_{W,X\times Y,Z}} C(W, C(X \times Y, Z)). \end{aligned}$$

This means that $C(X, C(Y, Z))$ and $C(X \times Y, Z)$ represent the same contravariant functor from spaces to sets, and it follows by Yoneda's Lemma that there is a natural homeomorphism

$$C(X, C(Y, Z)) = C(X \times Y, Z).$$

Applying Yoneda's Lemma just comes down to considering the cases $W = C(X, C(Y, Z))$ and $W = C(X \times Y, Z)$, as we did previously. $\square$

Proposition 23.26. [prop-map-metric]
If X is compact Hausdorff and Y is a metric space then $C(X, Y)$ is a metric space, with metric $d(f, g) = \max_{x\in X} d(f(x), g(x))$.

Proof. Let ζ be the compact-open topology on $C(X, Y)$, so the official topology on $C(X, Y)$ is $k(\zeta)$. Next, observe that the definition of $d(f, g)$ makes sense: the map $x \mapsto d(f(x), g(x))$ is a continuous real-valued function on a compact Hausdorff space, so it has a maximum. It is easy to check that $d(f, g)$ defines a metric; we write ξ for the resulting topology, so $\xi = k(\xi)$ by Proposition 23.6.

Suppose that we have a subbasic open set $W(u, K, U)$ for ζ, and a point $f \in W(u, K, U)$, so that $f: X \to Y$ and $fu(K) \subseteq U$. Then $a \mapsto d(fu(a), U^c)$ is a continuous, strictly positive, real-valued function on the compact space K, so it has a lower bound $\epsilon > 0$. It is easy to see that the open ball $OB_\epsilon(f) = \{g : d(g, f) < \epsilon\}$ is contained in $W(u, K, U)$. It follows that $W(u, K, U)$ is open with respect to ξ, so $\zeta \subseteq \xi$. It follows using Proposition 23.13 that $k(\zeta) \subseteq k(\xi) = \xi$.

Conversely, consider a point $f \in C(X, Y)$ and an open ball $OB_\epsilon(f)$ around f. The sets $f^{-1}(OB_{\epsilon/3}(y))$ (as y runs over Y) form an open cover of X. We may therefore choose finitely many points $y_1, \dots, y_n \in Y$ such that the sets $f^{-1}(OB_{\epsilon/3}(y_i))$ cover X. We write $K_i = f^{-1}(B_{\epsilon/3}(y_i))$ and $U_i = OB_{\epsilon/2}(y_i)$, so that the K_i are compact and cover X, and $f(K_i) \subseteq U_i$. It is clear that the set $N = \bigcap_i W(K_i, U_i)$ is a neighbourhood of f in the compact-open topology. We claim that $N \subseteq OB_\epsilon(f)$. Indeed, suppose that $g \in N$ and $x \in X$; we need to show that $d(f(x), g(x)) < \epsilon$. We know that $x \in K_i$ for some i, so both $f(x)$ and $g(x)$ lie in $OB_{\epsilon/3}(y_i)$, and the required inequality follows immediately. This shows that $OB_\epsilon(f)$ is a neighbourhood of f in ζ, so $\xi \subseteq \zeta \subseteq k(\zeta) \subseteq k(\xi) = \xi$. required. $\square$

Proposition 23.27. [prop-diag]
A compactly generated space X is weakly Hausdorff if and only if the diagonal subspace $\Delta_X = \{(x, x) : x \in X\}$ is closed in $X \times X$.

Proof. Suppose that X is weakly Hausdorff. First, observe that every one-point set $\{x\} \subset X$ is certainly a continuous image of a compact Hausdorff space and thus is closed in X, so X is T_1. Next, consider a test map $u = (v, w): K \to X \times X$. It will be enough to show that the set $u^{-1}(\Delta_X) = \{a \in K : v(a) = w(a)\}$ is closed in K. Suppose that $a \notin u^{-1}(\Delta_X)$, so $v(a) \neq w(a)$. Then the set $U = \{b : v(b) \neq w(a)\}$ is an open neighbourhood of a (because $\{w(a)\}$ is closed in X). Now K is compact Hausdorff and therefore regular, so there is an open neighbourhood V of a in K such that $\overline{V} \subseteq U$, or equivalently $w(a) \notin v(\overline{V})$. This means that a lies in the set $W = w^{-1}(v(\overline{V})^c)$. The weak Hausdorff condition implies that $v(\overline{V})$ is closed in X and thus W is open in K. We claim that $(V \cap W) \cap u^{-1}\Delta_X = \emptyset$. Indeed, if $b \in V \cap W$ then $v(b) \in v(\overline{V})$ but $w(b) \in v(\overline{V})^c$ by the definition of W, so $v(b) \neq w(b)$, so $u(b) = (v(b), w(b)) \notin \Delta_X$. This shows that $u^{-1}(\Delta_X)$ is closed in K, as required.

Conversely, suppose that Δ_X is closed in $X \times X$. Let $u \colon K \to X$ be a test map. Given any other test map $v \colon L \to X$, we define $M = \{(a, b) \in K \times L \;:\; u(a) = v(b)\} \subseteq K \times L$. This can also be described as $(u \times v)^{-1}\Delta_X$, so it is closed in $K \times L$ and thus compact. It follows that the projection $\pi_L(M)$ is compact and thus closed in L. However, it is easy to see that $\pi_L(M) = v^{-1}(u(K))$. Thus shows that $u(K)$ is k-closed in X, and hence closed. This means that X is weakly Hausdorff. □

Corollary 23.28. [cor-equaliser-closed]
If X and Y are CGWH and $f, g \colon X \to Y$ are continuous then the set

$$\operatorname{eq}(f, g) = \{x \;:\; f(x) = g(x)\} = (f, g)^{-1}(\Delta_Y)$$

is closed in X. □

Corollary 23.29. [cor-prod-wh]
*The product of an arbitrary family of CGWH spaces (with the CG topology) is CGWH, and thus gives a product in the category **CGWH**.*

Proof. Consider a product $X = \prod_i X_i$ of CGWH spaces with projection maps $\pi_i \colon X \to X_i$. We know that Δ_{X_i} is closed in $X_i \times X_i$ and $\pi_i \times \pi_i \colon X \times X \to X_i \times X_i$ is continuous so the set $D_i = (\pi_i \times \pi_i)^{-1}\Delta_{X_i}$ is closed in $X \times X$. It is clear that $\Delta_X = \bigcap_i D_i$, so Δ_X is closed in $X \times X$ and X is weakly Hausdorff. □

Proposition 23.30. [prop-distrib-CGWH]
Suppose we have two families of CGWH spaces, say $(X_i)_{i \in I}$ and $(Y_j)_{j \in J}$. Then the evident bijection

$$\coprod_{i,j}(X_i \times Y_j) \to \left(\coprod_i X_i\right) \times \left(\coprod_j Y_j\right)$$

is a homeomorphism.

Proof. Put $X = \coprod_i X_i$ and $Y = \coprod_j Y_j$ and $Z_{ij} = X_i \times Y_j$, so the claim is that $\coprod_{i,j} Z_{ij} = X \times Y$. In one direction, we can take the product of the inclusion $X_i \to X$ with the inclusion $Y_j \to Y$ to get a continuous map $Z_{ij} \to X \times Y$ for all i and j, and by the universal property of the coproduct these fit together to give a continuous map $f \colon \coprod_{i,j} Z_{ij} \to X \times Y$. By inspection of definitions we see that this is a bijection.

Next, for an arbitrary space T, we note that there are natural bijections

$$\begin{aligned}
\mathbf{CGWH}(X \times Y, T) = \mathbf{CGWH}(X, C(Y, T)) &= \mathbf{CGWH}(\coprod_i X_i, C(Y, T)) \\
&= \prod_i \mathbf{CGWH}(X_i, C(Y, T)) = \prod_i \mathbf{CGWH}(X_i \times Y, T) \\
&= \prod_i \mathbf{CGWH}(Y, C(X_i, T)) = \prod_i \mathbf{CGWH}(\coprod_j Y_j, C(X_i, T)) \\
&= \prod_i \prod_j \mathbf{CGWH}(Y_j, C(X_i, T)) = \prod_{i,j} \mathbf{CGWH}(X_i \times Y_j, T) \\
&= \mathbf{CGWH}(\coprod_{i,j} Z_{ij}, T).
\end{aligned}$$

In other words, we have a natural isomorphism between the functors represented by $X \times Y$ and $\coprod_{i,j} Z_{ij}$, and by the Yoneda Lemma this must come from a homeomorphism between these spaces. By taking $T = X \times Y$ and chasing the identity map $1_{X \times Y}$ through the above chain of equalities, we see that the resulting homeomorphism is just f. □

The following proposition is a good example of something which is usually true for the ordinary product of general spaces, but requires messy hypotheses. In our compactly generated context, it is true without restriction.

Proposition 23.31. [prop-quot]
Let X and Y be compactly generated spaces and E an equivalence relation on X. Let E' be the equivalence relation on $X \times Y$ defined by $(x_0, y_0)E'(x_1, y_1)$ if and only if $x_0 E x_1$ and $y_0 = y_1$. Then the natural bijection $(X \times Y)/E' \to (X/E) \times Y$ is a homeomorphism.

Proof. Let $q\colon X \to X/E$ and $q'\colon X \times Y \to (X \times Y)/E'$ be the quotient maps. We have a continuous map $q \times 1\colon X \times Y \to (X/E) \times Y$, which evidently respects the equivalence relation E', so we have an induced continuous map $f\colon (X \times Y)/E' \to (X/E) \times Y$. On the other hand, the adjoint of q' is a continuous map $X \to C(Y, (X \times Y)/E')$ respecting E, so we get an induced map $g^{\#}\colon X/E \to C(Y, (X \times Y)/E')$. The adjoint of this is a continuous map $g\colon (X/E) \times Y \to (X \times Y)/E'$. It is easy to check that f and g are just the evident bijections and thus that $fg = 1$ and $gf = 1$. □

Proposition 23.32. [prop-quot-prod]
If $f\colon W \to X$ and $g\colon Y \to Z$ are quotient maps of compactly generated spaces, then so is $f \times g\colon W \times Y \to X \times Z$.

Proof. It is immediate from Proposition 23.31 that $f \times 1_Y\colon W \times Y \to X \times Y$ and $1_X \times g\colon X \times Y \to X \times Z$ are quotient maps, and $f \times g = (1_X \times g) \circ (f \times 1_Y)$. □

We can now prove an analogue of Proposition 14.11 for CGWH spaces.

Corollary 23.33. [cor-quot-wh]
Let E be an equivalence relation on a compactly generated space X. Then X/E is weakly Hausdorff if and only if E is closed in $X \times X$. (Here we are identifying a relation R on X with the set $\{(x, y) \;:\; xRy\} \subseteq X \times X$.)

Proof. We know from Proposition 23.27 that X/E is weakly Hausdorff if and only if $\Delta_{X/E}$ is closed in $X/E \times X/E$. Let $q\colon X \to X/E$ be the quotient map, so Proposition 23.32 tells us that $q \times q\colon X \times X \to X/E \times X/E$ is a quotient map, so $\Delta_{X/E}$ is closed if and only if $(q \times q)^{-1}(\Delta_{X/E})$ is closed in $X \times X$. It is easy to see that $(q \times q)^{-1}(\Delta_{X/E}) = E$, so we conclude that X/E is weakly Hausdorff if and only if E is closed in $X \times X$. □

Proposition 23.34. [prop-wh-quot]
*Let X be a compactly generated space. Then there is a smallest closed equivalence relation E on X. If we write $hX = X/E$ then h defines a functor $\mathcal{G} \to$ **CGWH**, which is left adjoint to the inclusion **CGWH** $\to$ **CG**. In other words, any map from X to a CGWH space factors uniquely through hX.*

Proof. Let $\mathcal{R}$ be the set of all equivalence relations R on X such that R is closed as a subset of $X \times X$. (There is at least one such relation, namely $R = X \times X$.) One can then check that the set $E = \bigcap_{R \in \mathcal{R}} R$ is an equivalence relation and is closed in $X \times X$; clearly it is the smallest such. Corollary 23.33 tells us that $hX = X/E$ is CGWH. If Y is a CGWH space and $f\colon X \to Y$ is continuous then it is not hard to see that the set
$$R = \{(x, x') \;:\; f(x) = f(x')\} = (f \times f)^{-1}(\Delta_Y)$$
is a closed equivalence relation on X. It follows that $E \subseteq R$, and thus that f factors through a unique continuous map $hX \to Y$. This implies that h is a functor and is left adjoint to the inclusion of CGWH spaces in compactly generated spaces. □

Corollary 23.35. [prop-colimits]
The category of CGWH spaces has colimits for all diagrams, obtained by applying the functor h to the colimit as calculated in the category of all spaces. □***Refer to the appendix***

Proposition 23.36. [prop-cartesian-closed]
If X is compactly generated and Y is CGWH then $C(X, Y)$ is CGWH. Thus, the category ***CGWH*** *is cartesian closed.*

Proof. By definition, $C(X,Y)$ is compactly generated; we need only check that it is weakly Hausdorff, or equivalently that $\Delta_{C(X,Y)}$ is closed in $C(X,Y)\times C(X,Y)$. Define $\mathrm{ev}_x\colon C(X,Y)\to Y$ by $\mathrm{ev}_x(f)=f(x)$. For any open set $U\subseteq Y$ we have $\mathrm{ev}_x^{-1}(U)=W(\{x\},U)$, which is open in $C(X,Y)$, so ev_x is continuous. Moreover, we have $\Delta_{C(X,Y)}=\bigcap_x(\mathrm{ev}_x\times\mathrm{ev}_x)^{-1}\Delta_Y$, which is closed because Δ_Y is. □

23.3. Subspaces.

Definition 23.37. [defn-subspace]
Let X be a CGWH space (with topology ζ), and let Y be a subset of X. Let ζ_Y^0 denote the ordinary subspace topology on Y (so $\zeta_Y^0=\{F\cap Y\;:\;F\in\zeta\}$) and put $\zeta_Y=k(\zeta_Y^0)$; we call this the CGWH subspace topology. Let $i_Y\colon Y\to X$ be the inclusion map. As $\zeta_Y^0\subseteq\zeta_Y$ we see that i_Y is continuous with respect to ζ_Y and ζ.

Lemma 23.38. [lem-open-ordinary]
If Y is open or closed then $\zeta_Y=\zeta_Y^0$.

Proof. The case where Y is closed is easy, so we will assume instead that Y is open.

Suppose that $F\in\zeta_Y$, and put $V=Y\setminus F$. We must prove that F is closed in the ordinary subspace topology on the open set Y, or equivalently that V is open in X. Consider a test map $u\colon K\to X$; as X is compactly generated, it will be enough to show that $u^{-1}(V)$ is open in K. Suppose that $a\in u^{-1}(V)$. As Y is open in X, we know that $u^{-1}(Y)$ is an open neighbourhood of a in K. As K is compact Hausdorff and thus regular, we can choose an open neighbourhood N of a in K such that $\overline{N}\subseteq u^{-1}(Y)$. Now $\overline{N}$ is a compact Hausdorff space, and u restricts to give a continuous map $v\colon\overline{N}\to X$ with image contained in Y. Thus, as $F\in\zeta|_Y$, we see that $v^{-1}(F)$ is closed in $\overline{N}$, so the set $u^{-1}(V)\cap\overline{N}=v^{-1}(V)$ is open in $\overline{N}$. This means that $u^{-1}(V)\cap N$ is open in N and thus in K, so $u^{-1}(V)$ is a neighbourhood of a in K. This proves that $u^{-1}(V)$ is open in K, as required. □

Definition 23.39. [defn-inclusion]
A continuous map $i\colon Y\to X$ of CGWH spaces is an *inclusion* if it is injective, and the resulting map $Y\to i(Y)$ is a homeomorphism if $i(Y)$ is given the CGWH subspace topology inherited from X. Using Lemma 23.38 we see that an inclusion sends closed sets in Y to closed sets in X iff the image $i(Y)$ is closed in X. If so, we say that i is a *closed inclusion*. If Y is just a subset of X and $i(y)=y$ for all Y, we say that i is the *identity inclusion*.

Lemma 23.40. [lem-inc-detect]
Let $i\colon Y\to X$ be a continuous injective map of CGWH spaces. Then i is an inclusion if and only if it has the following property: if T is CGWH and $f\colon T\to Y$ is such that $if\colon T\to X$ is continuous, then f is continuous.

Proof. Let $P(i)$ denote the statement that i has the property mentioned above.

Let Y be a subset of X (with the CGWH subspace topology) and let $i_Y\colon Y\to X$ be the identity inclusion. Suppose that $f\colon T\to Y$ is such that i_Yf is continuous. This immediately implies that f is continuous with respect to ζ_Y^0, but T is CGWH so this is equivalent to continuity with respect to ζ_Y (by Corollary 23.10). Thus $P(i_Y)$ holds, and it follows that $P(i)$ holds for any inclusion.

Conversely, suppose we have an injective map $i\colon Y\to X$ such that $P(i)$ holds. Put $Y'=i(Y)$ and give this the CGWH subspace topology. Let $f\colon Y\to Y'$ be the bijection induced by i, so $i_{Y'}f=i$ and $if^{-1}=i_{Y'}$. Note that $P(i_{Y'})$ holds by the previous paragraph, and it follows that f is continuous. We are given that $P(i)$ holds, and it follows that f^{-1} is continuous. Thus f is a homeomorphism, so i is an inclusion. □

Corollary 23.41. [cor-retract]
Let $Y\xrightarrow{i}X\xrightarrow{r}Y$ be continuous maps of CGWH spaces such that $ri=1_Y$. Then i is a closed inclusion and r is a quotient map.

Proof. Let $f\colon T\to Y$ be such that if is continuous; then $f=rif$ is continuous as well. The lemma therefore tells us that i is an inclusion. One checks that

$$i(Y)=\{x\in X\;:\;ir(x)=x\}=(ir,1_X)^{-1}(\Delta_X),$$

so $i(Y)$ is closed in X, so i is a closed inclusion. Clearly r is surjective. If $F \subseteq Y$ is such that $r^{-1}F$ is closed, then the set $F = (ri)^{-1}(F) = i^{-1}(r^{-1}(F))$ is itself closed. This shows that r is a quotient map. $\square$

PROPOSITION 23.42. [`prop-limits`]
The category of CGWH spaces has limits for all small diagrams, and they are preserved by the forgetful functor to sets.

PROOF. Suppose we have a diagram $\{X_i\}$ of CGWH spaces. The limit calculated in the category of sets is a certain subset $X \subseteq \prod_i X_i$. We give $\prod_i X_i$ the CGWH product topology, and observe (using Corollary 23.28) that X is then a closed subspace. If we give X the subspace topology, it is easy to check that this makes it the limit in the CGWH category. $\square$

PROPOSITION 23.43. [`prop-inc-detect`]
Let $X \xrightarrow{i} Y \xrightarrow{j} Z$ be continuous maps of CGWH spaces.

(a) *If i and j are inclusions then so is ji.*
(b) *If i and j are closed inclusions then so is ji.*
(c) *If ji is an inclusion then so is i.*
(d) *If ji is a closed inclusion then so is i.*

PROOF. Parts (a) and (b) are easy, and part (c) follows from Lemma 23.40. Thus, in (d) we know that i is an inclusion and we just have to check that it is closed. By assumption, $ji(X)$ is closed in Z, so the set $Y' = j^{-1}(ji(X))$ is closed in Y, and it clearly contains $i(X)$. Let $k\colon Y' \to Y$ be the inclusion, and let i' be i regarded as a map to Y', so $ki' = i$. Next, if $y \in Y'$ then $j(y) = ji(x)$ for some x, and this x is unique because ji is injective, so we can denote it by $r(y)$. This gives a map $r\colon Y' \to X$ with $jir = jk$. In particular, jir is continuous and ji is an inclusion so r is continuous. We also have $jiri' = jki' = ji$, and ji is injective, so $ri' = 1_X$. It follows from Corollary 23.41 that i' is a closed inclusion, and k is also a closed inclusion, so $i = ki'$ is a closed inclusion as claimed. The maps considered are indicated in the following diagram:

$$\begin{tikzcd}
& Y' \arrow[r, tail, "k"] \arrow[dl, "r"'] & Y \arrow[d, "j"] \\
X \arrow[r, equal] & X \arrow[u, tail, "i'"] \arrow[ur, tail, "i"] \arrow[r, tail, "ji"'] & Z
\end{tikzcd}$$

$\square$

PROPOSITION 23.44. [`prop-inc-prod`]
Let X, Y and Z be CGWH spaces, and let $X \xrightarrow{i} Y$ be an inclusion. Then the map $i \times 1\colon X \times Z \to Y \times Z$ is again an inclusion. If i is closed then so is $i \times 1$.

PROOF. Suppose that $(u, v)\colon W \to X \times Z$ and that the map $(i \times 1) \circ (u, v) = (iu, v)\colon W \to Y \times Z$ is continuous. This means that iu and v are continuous, and i is an inclusion, so u is also continuous, so (u, v) is continuous. This proves that $i \times 1$ is an inclusion. If i is closed then $i(X)$ is closed in Y so $(i \times 1)(X \times Z) = i(X) \times Z$ is closed in $Y \times Z$, so $i \times 1$ is closed (by the remarks in Definition 23.39). $\square$

PROPOSITION 23.45. [`prop-inc-pullback`]
Suppose we have a pullback square as shown, in which i is an inclusion. Then i' is also an inclusion. Moreover, if i is closed then so is i'.

$$\begin{tikzcd}
X' \arrow[r, "i'"] \arrow[d, "f'"'] & Y' \arrow[d, "f"] \\
X \arrow[r, tail, "i"'] & Y.
\end{tikzcd}$$

PROOF. Suppose we have a map $u\colon W \to X'$ such that $i'u$ is continuous. We then see that the map $if'u = fi'u$ is continuous, and i is an inclusion, so $f'u$ is continuous. As $i'u$ and $f'u$ are continuous, the pullback property tells us that u is continuous. This proves that i' is an inclusion. Now suppose that i is closed. Then one checks that $i'(X') = f^{-1}i(X)$, which is closed in Y' because f is continuous and i is closed. It follows that i' is a closed inclusion. $\square$

Lemma 23.46. [lem-pullback]
If the diagram

$$\begin{tikzcd} W \arrow[r, tail, "f"] \arrow[d, "g"'] & X \arrow[d, "h"] \\ Y \arrow[r, "k"'] & Z \end{tikzcd}$$

is a pullback of sets and f is a closed inclusion, then it is a pullback of spaces.

Proof. We need to show that W is topologised as a subspace of $X \times Y$, or equivalently that the map $W \xrightarrow{(f,g)} X \times Y$ is a closed inclusion. Let $p\colon X \times Y \to X$ be the projection. Then $p \circ (f, g) = f$ is a closed inclusion, so (f, g) is a closed inclusion by Proposition 23.43(d). $\square$

Proposition 23.47. [prop-push-pull]
Suppose we have a pushout square as shown, in which i is a closed inclusion. Then j is also a closed inclusion, and the square is a pullback. Moreover, the pushout is created in the category of sets.

$$\begin{tikzcd} W \arrow[r, tail, "i"] \arrow[d, "f"'] & X \arrow[d, "g"] \\ Y \arrow[r, "j"'] & Z \end{tikzcd}$$

Proof. We may assume that W is a closed subspace of X and that i is the identity inclusion. We first analyse the pushout of i and f in the category of sets. Put $Z' = (X \setminus W) \amalg Y$ (just considered as a set, for the moment). Let $p\colon X \amalg Y \to Z'$ be given by the identity on the subset $Z' \subseteq X \amalg Y$, and by f on the subset $W \subseteq X \subseteq X \amalg Y$. Let $g'\colon X \to Z'$ and $j'\colon Y \to Z'$ be the restrictions of p. One can check directly that the square

$$\begin{tikzcd} W \arrow[r, tail, "i"] \arrow[d, "f"'] & X \arrow[d, "g'"] \\ Y \arrow[r, "j'"'] & Z' \end{tikzcd}$$

is a pushout in the category of sets. We give Z' the unique topology for which p is a quotient map. We will need to show that this topology is CGWH, or equivalently that the equivalence relation

$$E = \mathrm{eq}(p) = \{(a, b) \in (X \amalg Y)^2 \ : \ p(a) = p(b)\}$$

is closed in $X \amalg Y$. For this we put $G = \{(w, y) \in W \times Y \ : \ f(w) = y\}$, which is closed in $W \times Y$, and thus also in $X \times Y$. Put $G' = \{(y, x) \ : \ (x, y) \in G\} \subseteq_C Y \times X$ and

$$\begin{aligned} E_0 &= \mathrm{eq}(f) = \{(w, w') =\in W^2 \ : \ f(w) = f(w')\} \subseteq_C W^2 \subseteq_C X^2 \\ E_1 &= (\Delta_X \cup E_0) \amalg G \amalg G' \amalg \Delta_Y \subseteq_C X^2 \amalg (X \times Y) \amalg (Y \times X) \amalg Y^2 = (X \amalg Y)^2. \end{aligned}$$

Then E_1 is visibly closed, and one checks directly that $E = E_1$. Thus Z' is CGWH, so it is also the pushout in the category $\mathcal{U}$, so we can identify Z' with Z. It follows that j is injective and that for any closed set $F \subseteq Y$ we have

$$p^{-1}j(F) = f^{-1}(F) \amalg F \subseteq X \amalg Y,$$

which is closed. Thus j is a closed inclusion. Moreover, we now see that the square is a pullback of sets, so it is a pullback of spaces by Lemma 23.46. $\square$

Proposition 23.48. [prop-quot-pullback]
Consider a pullback square in $\mathcal{U}$ as shown, in which q is a quotient map. Then p is also a quotient map.

$$\begin{tikzcd} W \arrow[r, "f"] \arrow[d, "p"'] & X \arrow[d, two heads, "q"] \\ Y \arrow[r, "g"'] & Z \end{tikzcd}$$

Proof. First consider the special case where g is a closed inclusion. Proposition 23.45 tells us that f is also a closed inclusion. As the pullback is created in the category of sets, we can check by a small diagram chase that p is surjective and that $q^{-1}g(F) = fp^{-1}(F)$ for all subsets $F \subseteq Y$. Now suppose that $p^{-1}(F)$ is closed. As f is a closed inclusion we deduce that $q^{-1}g(F) = fp^{-1}(F)$ is closed, and as q is a quotient map this means that $g(F)$ is closed. As g is injective we have $F = g^{-1}g(F)$, so F is closed. This proves that p is a quotient map, as claimed.

For the general case, it is formal that the square below is also a pullback:

$$\begin{tikzcd} W \arrow[r, "{(f,p)}"] \arrow[d, "p"'] & X \times Y \arrow[d, two heads, "q\times 1"] \\ Y \arrow[r, "{(g,1)}"'] & Z \times Y \end{tikzcd}$$

Here $q \times 1$ is a quotient map by Proposition 23.32, and $(g, 1)$ is a closed inclusion by Corollary 23.41, so p is a quotient map by the special case already considered. □

Proposition 23.49. [prop-inc-map]

Let X, Y and Z be CGWH spaces, and let $X \xrightarrow{i} Y$ be an inclusion. Then the map $i_*: C(Z, X) \to C(Z, Y)$ is again an inclusion. Moreover, if i is closed then so is i_*.

Proof. Consider a map $w: W \to C(Z, X)$ such that $i_* \circ w: W \to C(Z, Y)$ is continuous. This means that the adjoint map $\mathrm{adj}^{-1}(i_* \circ w): W \times Z \to Y$ is continuous, but this is the same as the composite

$$W \times Z \xrightarrow{\mathrm{adj}^{-1}(w)} X \xrightarrow{i} Y,$$

and i is an inclusion, so $\mathrm{adj}^{-1}(w)$ is continuous, so w is continuous. This proves that i_* is an inclusion. If i is closed then the set $i_*(C(Z, X)) = C(Z, i(X))$ is closed in $C(Z, Y)$ by Remark 23.22, so i_* is a closed inclusion. □

Definition 23.50. [defn-collapse]

Let X be a CGWH space, and let Y be a closed subspace. Define X/Y to make the square on the left below a pushout:

$$\begin{tikzcd} Y \amalg 0 \arrow[r, tail, "i"] \arrow[d, "c"'] & X \amalg 0 \arrow[d, two heads, "q"] \\ 1 \arrow[r, tail, "z"'] & X/Y \end{tikzcd} \qquad \begin{tikzcd} Y \arrow[r, tail, "i"] \arrow[d] & X \arrow[d, two heads] \\ 1 \arrow[r, tail, "z"'] & X/Y \end{tikzcd}$$

By Proposition 23.47, the pushout is created in the category of sets, and the square is also a pullback. If $Y = \emptyset$ then $X/Y = X \amalg \{0\}$; otherwise, one checks that the right hand square is also a pushout and a pullback. We also let $p: X\amalg\{0\} \to X/Y$ be the constant map with value z, and note that $\{x \in X \ : \ p(x) = q(x)\} = Y\amalg\{0\}$.

Remark 23.51. [rem-collapse-closed]

If $F \subseteq X$ is a closed set then $q^{-1}(q(F))$ is either F (if $F \cap Y = \emptyset$) or $F \cup Y$. Either way $q^{-1}(q(F))$ is closed, and q is a quotient map, so $q(F)$ is closed in X/Y. This shows that q is a closed map.

Proposition 23.52. [prop-collapse-prod]

Suppose we have a CGWH space X, a closed subspace Y, and another CGWH space Z. Then $(X \times Z)/(Y \times Z) = (X/Y) \times Z$.

PROOF. This is essentially Proposition 23.32 applied to the quotient map $X \to X/Y$. $\square$

24. Limits and regularity

From now on we will take a somewhat more categorical viewpoint, and investigate the relationship between various kinds of limits and colimits in the category $\mathcal{U}$ of CGWH spaces. We will also change our terminology slightly: the word "space" will refer to a CGWH space unless we explicitly say otherwise.

24.1. Regularity. Let $\mathcal{C}$ be a category with finite limits and colimits. Then any parallel pair of maps $f, g\colon A \to B$ has an equaliser $\mathrm{eq}(f,g) \to A$ and a coequaliser $B \to \mathrm{coeq}(f,g)$. We also write $\mathrm{eq}(f)$ for the equaliser of $f\pi_1, f\pi_2\colon A^2 \to B$ and $\mathrm{coeq}(f)$ for the dual thing. Equivalently, we have pullback and pushout squares as shown.

$$\begin{tikzcd} \mathrm{eq}(f) \arrow[r] \arrow[d, tail] & B \arrow[d, tail, "\Delta"] \\ A \times A \arrow[r, "f\times f"'] & B \times B \end{tikzcd} \qquad \begin{tikzcd} A \amalg A \arrow[r, "f \amalg f"] \arrow[d, two heads, "\nabla"'] & B \amalg B \arrow[d, two heads] \\ A \arrow[r] & \mathrm{coeq}(f) \end{tikzcd}$$

or

$$\begin{tikzcd} \mathrm{eq}(f) \arrow[r] \arrow[d] & A \arrow[d, "f"] \\ A \arrow[r, "f"'] & A \end{tikzcd} \qquad \begin{tikzcd} A \arrow[r, "f"] \arrow[d, "f"'] & B \arrow[d] \\ B \arrow[r] & \mathrm{coeq}(f) \end{tikzcd}$$

In other words, $\mathrm{eq}(f) = A \times_B A$ and $\mathrm{coeq}(f) = B \amalg_A B$.

Recall that a map $f\colon B \to C$ is said to be a *regular epimorphism* if it is the coequaliser of some pair of maps $A \rightrightarrows B$, or equivalently if it is the coequaliser of the obvious pair of maps $\mathrm{eq}(f) \rightrightarrows B$. Dually, f is a *regular monomorphism* if it is the equaliser of some pair of maps $C \rightrightarrows D$.

We say that a category is *regular* if every map can be factored as a regular epimorphism followed by a monomorphism, and pullbacks of regular epimorphisms are regular epimorphisms. *Coregularity* is defined dually, and *biregular* means regular and coregular.

THEOREM 24.1. [`thm-regular-spaces`]

(a) *A map in $\mathcal{U}$ is a monomorphism if and only if it is injective, and an epimorphism if and only if it has dense image.*
(b) *A map in $\mathcal{U}$ is a regular monomorphism if and only if it is a closed inclusion.*
(c) *A map in $\mathcal{U}$ is a regular epimorphism if and only if it is a quotient map.*
(d) *A product, coproduct or composite of (regular) monomorphisms is a (regular) monomorphism.*
(e) *A coproduct, finite product or composite of (regular) epimorphisms is a (regular) epimorphism.*
(f) *$\mathcal{U}$ is biregular*

PROOF OF THEOREM 24.1. (a): By definition, $i\colon A \to B$ is a monomorphism if and only if the induced map of sets $i_*\colon \mathcal{U}(X,A) \to \mathcal{U}(X,B)$ is injective for all X. This clearly holds if f is injective; for the converse, take X to be a single point, so that $\mathcal{U}(X,A)$ is just the underlying set of A.

Now suppose that $r\colon A \to B$ has dense image; we claim that r is an epimorphism. Indeed, suppose we have two maps $g, h\colon B \to X$ with $gr = hr$. Then the set

$$\mathrm{eq}(g,h) = \{b \in B \;:\; g(b) = h(b)\} = (g \times h)^{-1}\Delta_X$$

is a closed subspace of B containing the image of r. As this image is dense, we have $C = B$ and $g = h$. Thus r is an epimorphism.

Conversely, suppose that $r\colon A \to B$ is an epimorphism. Let A' be the closure of the image of r, and apply Definition 23.50 to A'. This gives a pair of maps $p, q\colon B \to B/A'$ and a point $z \in B/A'$ with $p^{-1}\{z\} = B$ and $q^{-1}\{z\} = A'$. It follows that $pr = qr$, but r is epi, so $p = q$. This means that $A' = q^{-1}\{z\} = p^{-1}\{z\} = B$, so $r(A)$ is dense as claimed.

(b): Any regular monomorphism $i\colon A \to B$ is the equaliser of some parallel pair of arrows $g, h\colon B \to X$. The equaliser was originally constructed by taking the closed set $\{b \in B \;:\; g(b) = h(b)\} = (g \times h)^{-1}\Delta$ and giving it the subspace topology; so i is a closed inclusion.

Conversely, let $i\colon A \to B$ be the inclusion of a closed subspace. Then i is the equaliser of the maps $p, q\colon B \to B/A$ in Definition 23.50, so it is a regular monomorphism.

(c): Let $r\colon A \to B$ be a regular epimorphism, so r is the coequaliser of some pair of arrows $g, h\colon X \to A$. There exist closed equivalence relations $R \subseteq A^2$ such that the set $S = \{(g(x), h(x)) \;:\; x \in X\}$ is contained in R (for example, $R = A^2$). Let R be the intersection of all such relations. Clearly R itself is a closed equivalence relation, so we have a quotient map $q\colon A \to B' = A/R$, and $B' \in \mathcal{U}$ by Corollary 23.33. This is easily seen to be a coequaliser of g and h, so it can be identified with r, so r is a quotient map.

Conversely, let $r\colon A \to B$ be a quotient map, so $B = A/R$ for some equivalence relation R, which must be a closed subspace of A^2 because B is weakly Hausdorff. It is then clear that r is a coequaliser for the two projections $\pi_0, \pi_1\colon R \to A$, so it is a regular epimorphism.

(d): In any category, it is trivial that a product of (regular) monomorphisms is a (regular) monomorphism. The corresponding facts for coproducts and composites follow from parts (a) and (b) and the explicit construction of coproducts.

(e): The statements about coproducts are again formal, as is the fact that a composite of epimorphisms is epi. It follows easily from (c) that composites of regular epimorphisms are regular epimorphisms. Proposition 23.32 tells us that finite products of quotient maps are quotient maps, and we now know that quotient maps are the same as regular epimorphisms.

(f): Consider a map $f\colon A \to B$. Write $R = \mathrm{eq}(f)$, which is a closed equivalence relation on A. Let C be the closure of the image of f, topologised as a subspace of B. We then have a factorisation of f as a composite

$$A \to A/R \to C \to B.$$

The factorisation $A \to A/R \to B$ displays f as a regular epimorphism followed by a monomorphism. The factorisation $A \to C \to B$ displays f as an epimorphism followed by a regular monomorphism.

Proposition 23.48 now tells us that pullbacks of regular epis are regular epi, so $\mathcal{U}$ is regular. Proposition 23.47 tells us that pushouts of regular monos are regular mono, so $\mathcal{U}$ is coregular. □

24.2. Filtered colimits. The general theory of filtered diagrams is explained in Section 36.8. We now discuss how this works out in the category of CGWH spaces.

Lemma 24.2. [`lem-filtered`]
Let $\{X_i\}_{i \in I}$ be a filtered diagram of closed inclusions of spaces, with colimit X. Then the underlying set of X is the colimit of the underlying sets of the X_i, and the maps $X_i \to X$ are closed inclusions.

Proof. We first claim that if $u, v\colon i \to j$ in I then $u_* = v_*\colon X_i \to X_j$. Indeed, if w is as in axion (c) above then certainly $w_* u_* = w_* v_*$, but w_* is assumed to be a closed inclusion, so $u_* = v_*$.

Next suppose we have $i, j, m \in I$ and there exist maps $i \xrightarrow{u} m \xleftarrow{v} j$. We put

$$R_{ij}^m = \{(x, y) \in X_i \times X_j \;:\; u_*(x) = v_*(y)\},$$

which is closed in $X_i \times X_j$. By the previous paragraph, this is independent of the choice of u and v. In the case $m = i = j$ we can take $u = v = 1_i$ to see that $R_{ii}^i = \Delta_{X_i}$.

If there exists a morphism $f\colon m \to m'$ then, using the fact that $f_*\colon X_m \to X_{m'}$ is injective we see that $R_{ij}^m = R_{ij}^{m'}$. Even if there is no map $m \to m'$ we can certainly choose m'' with maps $m \to m'' \leftarrow m'$ so we still have $R_{ij}^m = R_{ij}^{m'}$. We write R_{ij} for this set. If $(x, y) \in R_{ij}$ and $(y, z) \in R_{jk}$ then, by choosing on object m that admits maps from all of i, j and k, we see that $(x, z) \in R_{ik}$.

Now put $T = \coprod_i X_i$, so $T^2 = \coprod_{i,j} X_i \times X_j$. As R_{ij} is closed in $X_i \times X_j$ we see that the set $R = \coprod_{i,j} R_{ij}$ is closed in T^2. It is also an equivalence relation, so we have a CGWH space $X = T/R$. Let $q\colon T \to X$ be the quotient map, and let $f_i\colon X_i \to X$ be the obvious map. It is now straightforward to check that these form a universal cone, so $\varinjlim_i X_i = X$, and this colimit is created in the category of sets. As $R_{ii} = \Delta_{X_i}$ we see that f_i is injective. Suppose we have a closed set $F \subseteq X_i$; we claim that $f_i(F)$ is closed in X. It will suffice to show that $q^{-1}(f_i(F))$ is closed in T, or that $X_j \cap q^{-1}(f_i(F))$ is closed in X_j for all j. To see this choose an object m and maps $i \xrightarrow{u} m \xleftarrow{v} j$, giving closed inclusions $X_i \xrightarrow{u_*} X_j \xleftarrow{v_*} X_k$. One checks that $X_j \cap q^{-1}(f_i(F)) = v_*^{-1}(u_*(F))$, which is closed in X_j as required. This proves that f_i is a closed inclusion. $\square$

Definition 24.3. [defn-strongly-filtered]
We say that a filtered diagram $\{A_i\}$ of closed inclusions is *strongly filtered* if every compact subset of $A = \varinjlim_i A_i$ lies in the image of some A_i.

Remark 24.4. [rem-not-strongly-filtered]
A convergent sequence together with its limit is compact. Using this, one sees that any compact metric space A is the filtered colimit of its countable compact subspaces. This diagram is not strongly filtered unless A is countable.

Lemma 24.5. [lem-pointy]
A sequence of closed inclusions is strongly filtered. The diagram of finite subcomplexes of a CW complex is strongly filtered. More generally, let $\{A_i\}$ be a directed system of subsets of $A = \varinjlim_i A_i$, and suppose that there are disjoint sets B_j such that each A_i is the union of a finite set of B's. Then the family $\{A_i\}$ is strongly filtered.

Proof. This argument is well-known, but it seems a little less well-known exactly what one needs to make it work.

The first and second claims follow from the third, by taking $B_i = A_i \setminus A_{i-1}$ in the first case, or taking the B's to be the open cells in the second. So suppose that the A's and B's are as in the third case. Suppose that $C \subseteq A$ is compact. For each j such that $B_j \cap C \neq \emptyset$, choose $b_j \in B_j \cap C$. Let D be the set of these b_j's. For any $E \subseteq D$ and any i, we see that $E \cap A_i$ is finite and thus closed in A_i. It follows that E is closed in A. As this holds for all $E \subseteq D$, we conclude that D is a discrete closed subset of the compact set C, hence D is finite. Thus C is contained in some finite union of B's. As the diagram of A's is directed, we conclude that $C \subseteq A_i$ for some i. $\square$

Lemma 24.6. [lem-sub-pointy]
Let $\{A_i\}$ be a directed family of closed subsets of B, and write $A = \bigcup_i A_i \subseteq B$. Consider the conditions

(a) *A is closed in B, and is homeomorphic to $\varinjlim_i A_i$.*
(b) *For any compact set $C \subseteq B$, we have $C \cap A = C \cap A_i$ for some i.*

Then (b) implies (a), and the converse holds if $\{A_i\}$ is strongly filtered.

Proof. (b)⇒(a): Let $C \subseteq B$ be compact. Then $C \cap A = C \cap A_i$ for some i, and this is closed in C because A_i is closed in B. Thus A is compactly closed and thus closed in B. Similarly, suppose that $D \subseteq A$ is such that $D \cap A_i$ is closed for all i. Then for any compact C we can choose i such that $C \cap A = C \cap A_i$, so $D \cap C = D \cap C \cap A_i$, and this is again closed in C; thus D is closed in B. Thus, the subspace topology on A coincides with the colimit topology.

(a)⇒(b): Suppose that $\{A_i\}$ is strongly filtered and (a) holds. Suppose that $C \subseteq B$ is compact. As A is closed, we see that $C \cap A$ is compact. As $A = \varinjlim_i A_i$ is a strongly filtered colimit, we see that $C \cap A \subseteq A_i$ for some i, so $C \cap A = C \cap A_i$. $\square$

Lemma 24.7. [lem-filtered-mappings]
If X is compact, then the functor $C(X, -)$ preserves strongly filtered colimits of closed inclusions.

Proof. Let $\{A_i\}$ be a strongly filtered diagram with colimit A. As $C(X, -)$ has a left adjoint, it preserves regular monos, so $\{C(X, A_i)\}$ is a diagram of closed inclusions. Any map $X \to A$ has compact image; by the definition of a strongly filtered colimit, it therefore factors through some A_i. This means that the natural map $b\colon \varinjlim_i C(X, A_i) \to C(X, A)$ is a continuous bijection. Now let W be another compact space. Proposition 23.25 gives us a diagram

$$\begin{tikzcd}
\varinjlim C(W, C(X, A_i)) \arrow[r] \arrow[d, "a"', "\simeq"] & C(W, \varinjlim C(X, A_i)) \arrow[r, "b_*"] & C(W, C(X, A)) \\
\varinjlim C(W \times X, A_i) \arrow[rr, "c"'] & & C(W \times X, A) \arrow[u, "\simeq", "d"']
\end{tikzcd}$$

in which a and d are homeomorphisms. The maps a and d are bijective for formal reasons. The map c is bijective by the previous paragraph, as $W \times X$ is compact. As $b\colon \varinjlim C(X, A_i) \to C(X, A)$ is bijective, the map b_* is mono. Commutativity of the diagram now shows that b is also epi, so it is a bijection. As any space is a colimit of compact spaces, we conclude that $C(W, \varinjlim C(X, A_i)) = C(W, C(X, A))$ for noncompact W's as well. Thus $\varinjlim C(X, A_i)$ is homeomorphic to $C(X, A)$ by Yoneda's lemma.

Now suppose that $K \subseteq C(X, A)$ is compact. Then the image of the compact space $K \times X$ under the evaluation map is a compact subspace of A, hence contained in some A_k; it follows that $K \subseteq C(X, A_k)$. Thus, the diagram $\{C(X, A_i)\}$ is strongly filtered. □

Lemma 24.8. [lem-sequence-pullback]
Consider a diagram of the form

$$\begin{tikzcd}
A_0 \arrow[r, tail, "i_0"] \arrow[d, tail, "f_0"'] & A_1 \arrow[r, tail, "i_1"] \arrow[d, tail, "f_1"'] & A_2 \arrow[r, tail, "i_2"] \arrow[d, tail, "f_2"'] & \dots \\
B_0 \arrow[r, tail, "j_0"'] & B_1 \arrow[r, tail, "j_1"'] & B_2 \arrow[r, tail, "j_2"'] & \dots
\end{tikzcd}$$

in which all maps are closed inclusions and all squares are pullbacks of sets (and therefore of spaces, by lemma 23.46). Write A_∞ and B_∞ for the evident colimits. Then the diagram

$$\begin{tikzcd}
A_i \arrow[r, tail] \arrow[d, tail, "f_k"'] & A_\infty \arrow[d, tail, "f_\infty"] \\
B_i \arrow[r, tail] & B_\infty
\end{tikzcd}$$

is a pullback square of closed inclusions.

Proof. Using Lemma 24.2, we see that the square is a pullback of sets, that the horizontal maps are closed inclusions, and that the vertical maps are injective. It follows from Lemma 23.46 that it is a pullback of spaces. The left hand vertical is a closed inclusion by assumption. If $C \subseteq A_\infty$ is closed then $f_\infty(C) \cap B_k = f_k(C \cap A_k)$ by the pullback property, and this is clearly closed in B_k. As $B_\infty = \varinjlim_k B_k$, we conclude that $f_\infty(C)$ is closed in B_∞, so f_∞ is a closed inclusion. □

Corollary 24.9. [cor-bifiltered]
Let $\{A_{k,l}\}$ be a diagram of closed inclusions indexed by $\mathbb{N}^2$, such that each square

$$\begin{tikzcd}
A_{k,l} \arrow[r] \arrow[d] & A_{k+1,l} \arrow[d] \\
A_{k,l+1} \arrow[r] & A_{k+1,l+1}
\end{tikzcd}$$

is a pullback. Define $A_{\infty,l}$, $A_{k,\infty}$ and $A_{\infty,\infty}$ to be the obvious colimits. Then the resulting diagram indexed by $(\mathbb{N} \cup \{\infty\})^2$ again consists of pullback squares of closed inclusions.

24.3. β-epimorphisms. For any set S we let βS be the space of ultrafilters on S, or equivalently the maximal ideal space of the Banach algebra $C^*(S)$ of bounded functions $S \to \mathbb{R}$. This is called the Stone-Cech compactification of the discrete space S. The basic point is that βS is a compact Hausdorff space containing S as a discrete open subspace, and that any function $u\colon S \to X$ (where X is compact Hausdorff) extends uniquely to a continuous map $\beta S \to X$. In other words, β is left adjoint to the forgetful functor from compact Hausdorff spaces to sets. (In fact, a theorem of Manes tells us that the category of such spaces is equivalent to the category of algebras for β, considered as a monad in the category of sets.) All this is covered in [**1**].

Definition 24.10. [`defn-beta-epi`]
We say that a map $f\colon X \to Y$ of CGWH spaces is a β-epimorphism if it satisfies the following equivalent conditions.

(a) For every set S and map $u\colon \beta S \to Y$ there is a map $v\colon \beta S \to X$ with $fv = u$.
(b) If $L \subseteq Y$ is compact then there is a compact set $K \subseteq Y$ such that $f(K) = L$.
(c) If $L \subseteq Y$ is compact then there is a compact set $K \subseteq Y$ such that $f(K) \supseteq L$.

Proof of equivalence. Suppose that (a) holds and that we are given a compact set $L \subseteq Y$. We have a counit map $\beta L \twoheadrightarrow L$ which we compose with the inclusion $L \rightarrowtail Y$ to get $u\colon \beta L \to Y$, which we can lift to a map $v\colon \beta L \to X$. We then put $K = v(\beta L)$ and observe that this is compact and has $f(K) = L$.

It is trivial that (b) implies (c). On the other hand, given K satisfying (c) we can take $K' = K \cap f^{-1}(L)$ to see that (b) holds as well.

Now suppose that (b) holds and we are given $u\colon \beta S \to Y$. Put $L = u(\beta S)$; this is a compact Hausdorff subset of Y, so we can choose a compact Hausdorff set $K \subseteq X$ with $f(K) = L$. Now choose a function $v\colon S \to K$ lifting the composite $S \rightarrowtail \beta S \xrightarrow{u} K$, and extend over βS using the universal property. This gives the required lifting of u to X. $\square$

Proposition 24.11. [`prop-beta-regular`]
If f is a β-epimorphism, then it is a regular epimorphism.

Proof. We can take L to be a point in (b) to see that f is surjective. Suppose that $Z \subseteq Y$ is such that $f^{-1}Z$ is closed in X; it will be enough to check that Z is closed in Y, for then f is a quotient map and thus a regular epi. As Y is CGWH it is enough to check that $Z \cap L$ is closed when L is compact. We can choose a compact set $K \subseteq X$ with $f(K) = L$ and then $Z \cap L = f(f^{-1}(Z) \cap K)$ but $f^{-1}(Z)$ is closed so $f^{-1}(Z) \cap K$ is compact Hausdorff so $f(f^{-1}(Z) \cap K)$ is closed as required. $\square$

Lemma 24.12. [`lem-beta-epi`]
The class of β-epimorphisms is closed under products, pullbacks and composition. $\square$

Lemma 24.13. [`lem-metric-beta-epi`]
If X is a metric space and $\emptyset \neq W \subseteq X$ is closed and $f\colon X \to Y := X/W$ is the projection then f is β-epi.

Proof. Suppose that $L \subseteq Y$ is compact. Put $L' = L \setminus \{0\}$ and $K' = f^{-1}L'$; one checks that $f\colon K' \to L'$ is a homeomorphism. Let K be the closure of K' in X and choose $w \in W$, so $f(K \cup \{w\}) = L \cup \{0\}$. It is thus enough to show that K is compact, or equivalently that K' is totally bounded.

Fix $\epsilon > 0$. Let S be a subset of K' with the property that $d(a,b) > \epsilon/2$ when $a, b \in S$ and $a \neq b$. Any Cauchy sequence in S is clearly eventually constant, and it follows that S is a discrete closed subspace of X. Also $S \cap W = \emptyset$ so $f^{-1}f(S) = S$ and f is a quotient map so $f(S)$ is closed so $U := f(S)^c$ is open in Y. For each $s \in S$, choose a small open ball B_s of radius at most $\epsilon/4$ around s such that $B_s \cap W = \emptyset$ and put $V_s = f(B_s)$, which is an open neighbourhood of $f(s)$ in Y. The sets U and V_s cover Y and so some finite subcollection covers L. However $f(s) \in L$ and V_s is the only set in the cover that contains $f(s)$ so each V_s must be in the finite subcover so S must be finite.

It follows from this that we can choose a maximal set S with the stated properties, and it is easy to see that such an S is an ϵ-net for K'. Thus K' is totally bounded, as required. $\square$

Lemma 24.14. [`lem-CW-beta-epi`]

If X is a CW complex and $\emptyset \neq W \subseteq X$ is a subcomplex and $f\colon X \to Y := X/W$ is the projection then f is β-epi.

Proof. Suppose that $L \subseteq X/W$ is compact. Then L is contained in a finite subcomplex, so we may assume that it is itself a finite subcomplex. Each open cell lifts to an open cell in X whose closure is a finite complex and thus compact. It follows easily that we can choose a compact subcomplex $K \subseteq X$ with $f(K) = L$, as required. □

Definition 24.15. [`defn-proper`]
We say that a surjective map $f\colon X \to Y$ is *proper* if it satisfies the following equivalent conditions.

(a) f is closed and has compact fibres.
(b) The preimage of any compact set is compact.

(It is immediate from (b) that a proper map is β-epi and thus regular epi.)

Proof of equivalence. Suppose that (a) holds and that $L \subseteq Y$ is compact, and put $K = f^{-1}(L)$; we need to show that K is compact. Let $\{F_i\}$ be a collection of nonempty closed subsets of K that is closed under finite intersections. It suffices to show that $\bigcap_i F_i \neq \emptyset$. The sets $f(F_i)$ are closed in the compact space L and any finite intersection of them is nonempty, so $\bigcap_i f(F_i)$ contains some point y say. The sets $F_i \cap f^{-1}\{y\}$ are thus nonempty closed subsets of the compact space $f^{-1}\{y\}$ and are closed under finite intersections, so $\bigcap_i F_i \cap f^{-1}\{y\} \neq \emptyset$ so $\bigcap_i F_i \neq \emptyset$ as required.

Conversely, suppose that (b) holds. It is immediate that f has compact fibres. Suppose that $F \subseteq X$ is closed; we need to show that $f(F)$ is closed. As Y is CGWH, it suffices to show that $f(F) \cap L$ is closed when $L \subseteq Y$ is compact, but then $K = f^{-1}(L)$ is compact in X so $F \cap K$ is compact so $f(F) \cap L = f(F \cap K)$ is closed as required. □

Proposition 24.16. [`prop-proper-bij`]
A proper continuous bijection of CGWH spaces is a homeomorphism.

Proof. Let $f\colon Y \to Z$ be a proper continuous bijection. As remarked above, f is regular epi and clearly also mono and thus an isomorphism. For a more direct argument, we need only show that f^{-1} is continuous or equivalently that f is closed. Let $X \subseteq Y$ be closed, and let $L \subseteq Z$ be compact. It will suffice to show that $f(X) \cap L$ is closed. As f is proper we see that $K := f^{-1}(L)$ is compact so $X \cap K$ is compact, and images of compact sets are always compact so the set $f(X) \cap L = f(X \cap K)$ is compact, and thus closed as required. **check that this does not need strong Hausdorff property**. □

Remark 24.17. [`rem-not-proper`]
The obvious map $[0,1) \amalg [0,1] \to [0,1]$ is a split epimorphism (and thus a β-epi and a regular epi) with compact fibres, but is not proper.

25. Based spaces

Definition 25.1. [`defn-pointed-category`]
Let $\mathbf{CGWH}_*$ denote the category of CGWH spaces equipped with a specified basepoint $0_X \in X$. The morphisms from X to Y are the continuous maps $f\colon X \to Y$ for which $f(0_X) = 0_Y$. We will use the symbol 0 to denote the constant map $X \to Y$ taking everything to 0_Y, and also for the based space whose only point is the basepoint.

Definition 25.2. [`defn-base-adjunction`]
If X is an unbased space we write X_+ for $X \amalg \{0\}$, with the new point 0 taken as the basepoint. If Y is a based space we write Y_- for Y regarded as an unbased space. We then have an evident adjunction $\mathbf{CGWH}_*(X_+, Y) = \mathbf{CGWH}(X, Y_-)$.

Remark 25.3. [`rem-collapse-pointed`]
Given a space $X \in \mathbf{CGWH}$ and a closed subspace Y, we can form $X/Y = (X \amalg \{0\})/E$ as in Definition 23.50, and take the image of 0 as the basepoint. Most often we will do this when X is a based space and Y contains the basepoint. In that case we can regard X/Y as a quotient of X, and then $0_{X/Y}$ is the image of 0_X.

Remark 25.4. [rem-limits-pointed]
If $(X_i)_{i\in I}$ is any diagram in $\mathbf{CGWH}_*$ then we can form the inverse limit $X = \varprojlim_i X_i$ in $\mathbf{CGWH}$. The points $(0_{X_i})_i$ define a point 0_X of this inverse limit, and we take this as the basepoint in X. This makes X into the inverse limit of the diagram in $\mathbf{CGWH}_*$. More formally, we can say that $\mathbf{CGWH}_*$ has all small limits, and they are created by the forgetful functor $\mathbf{CGWH}_* \to \mathbf{CGWH}$. In particular, the categorical product of based spaces X and Y is just the space $X \times Y$ with the basepoint $0_{X\times Y} = (0_X, 0_Y)$.

Before discussing colimits, we need a small preliminary.

Definition 25.5. [defn-connected-category]
Given a category I, we let $\sim$ be the smallest equivalence relation on $\mathrm{obj}(I)$ such that $i \sim j$ whenever there is a morphism from i to j. We write $\pi_0(I)$ for the set of equivalence classes, and we say that I is *connected* if $\pi_0(X)$ is a singleton. For example, the usual indexing categories for coequalisers, pushouts and quotients by group actions are connected, but those for products are not.

Remark 25.6. [rem-connected-limit]
Let $\mathcal{C}$ be a category with finite products and coproducts, let X be an object of $\mathcal{C}$, and regard it as a constant functor $I \to \mathcal{C}$. Then one checks that $\varinjlim_I X = \coprod_{i\in\pi_0(I)} X$ and $\varprojlim_I X = \prod_{i\in\pi_0(I)} X$. In particular, if I is connected then $\varinjlim_I X = X = \varprojlim_I X$.

Remark 25.7. [rem-colimits-pointed]
Let $X : I \to \mathbf{CGWH}_*$ be a diagram of pointed spaces, and let X' be the colimit in $\mathbf{CGWH}$. Let 0 denote the constant I-diagram whose value is the one-point space, so $\varinjlim_I 0 = \pi_0(I)$ (with the discrete topology). There are evident maps $0 \to X \to 0$ of I-diagrams, giving maps $\pi_0(I) \to X' \to \pi_0(I)$ in $\mathbf{CGWH}$. The composite is the identity on $\pi_0(I)$, showing that $\pi_0(I)$ is embedded as a closed subspace in X'. One checks that $X'/\pi_0(I)$ is the colimit of X in $\mathbf{CGWH}_*$, showing that $\mathbf{CGWH}_*$ has all small colimits. Moreover, if I is connected we see that colimits for I-diagrams are created in $\mathbf{CGWH}$. In particular, pushouts and coequalisers are the same whether calculated in $\mathbf{CGWH}_*$ or in $\mathbf{CGWH}$. However, the categorical coproduct in $\mathbf{CGWH}_*$ is the wedge product, defined by

$$X \vee Y = (X \amalg Y)/(\{0_X\} \amalg \{0_Y\})$$

Note that there is a canonical map $k\colon X \vee Y \to X \times Y$ given by $k(x) = (x, 0_Y)$ for $x \in X$ and $k(y) = (0_X, y)$ for $y \in Y$.

Theorem 25.8. [thm-regular-pointed]
Let $f\colon A \to B$ be a morphism in $\mathbf{CGWH}_*$.

(a) *f is a monomorphism if and only if it is injective, and an epimorphism if and only if it has dense image.*
(b) *f is a regular monomorphism if and only if it is a homeomorphism of A with a closed subset of B (with the usual subspace topology), or in other words a closed inclusion.*
(c) *f is a regular epimorphism if and only if it is surjective and B has the quotient topology.*
(d) *A coproduct, product or composite of (regular) monomorphisms is a (regular) monomorphism.*
(e) *A coproduct, finite product or composite of (regular) epimorphisms is a (regular) epimorphism.*
(f) *$\mathbf{CGWH}_*$ is biregular.*

Proof. In part (e), it is categorical nonsense that coproducts preserve coequalisers and so preserve regular epimorphisms. In part (d), we leave aside for the moment the claim that coproducts preserve (regular) monomorphisms. The remaining claims involve only products, equalisers and coequalisers, all of which are created in $\mathbf{CGWH}$. Moreover, the proofs given for Theorem 24.1 also use only these constructs. Thus, everything goes through as before.

We now return to (d). Consider injective maps $i\colon A \to B$ and $j\colon C \to D$. By inspection of the construction we see that $i \vee j\colon A \vee C \to B \vee D$ is injective. Part (a) tells us that monomorphisms are precisely the injective maps, so coproducts preserve monomorphisms. Now suppose that i and j are regular monomorphisms, or in other words closed inclusions. We have a commutative diagram as follows, in $i \amalg j$ is a closed inclusion, and q is also a closed map by Remark 23.51.

$$\begin{tikzcd} A \amalg C \arrow[r, "i \amalg j"] \arrow[d, two heads, "p"'] & B \amalg D \arrow[d, two heads, "q"] \\ A \vee C \arrow[r, "i \vee j"'] & B \vee D \end{tikzcd}$$

Given a closed subset $F \subseteq A \vee C$, one checks directly (separating the cases $0 \in F$ and $0 \notin F$) that $(i \vee j)(F) = q((i \amalg j)(p^{-1}(F)))$, which is again closed. If follows that $i \vee j$ is a closed inclusion, as claimed. $\square$

Proposition 25.9. [`prop-wedge-closed`]
The maps $X \xrightarrow{i} X \vee Y \xleftarrow{j} Y$ and $X \vee Y \xrightarrow{k} X \times Y$ are closed inclusions.

Proof. Let $X \xleftarrow{p} X \times Y \xrightarrow{q} Y$ be the projections. Note that $pki = 1_X$ and $qkj = 1_Y$. This means that i, j, ki and kj are all split monomorphisms, hence regular monomorphisms, hence closed inclusions. If $A \subseteq X \vee Y$ is closed then $B = i^{-1}(A)$ is closed in X and $C = j^{-1}(A)$ is closed in Y, so $k(A) = ki(B) \cup kj(C)$ is closed in $X \times Y$. Thus k is closed as claimed. $\square$

25.1. The smash product.

Definition 25.10. [`defn-smash`]
We define $X \wedge Y = (X \times Y)/(X \vee Y)$ and $S^0 = \{0, 1\}$ (with 0 as the basepoint). The underlying sets satisfy

$$(X \wedge Y) \setminus \{0\} = (X \setminus \{0\}) \times (Y \setminus \{0\}).$$

Remark 25.11. [`rem-one-point-smash`]
Let X and Y be locally compact Hausdorff, so we have compactifications X_∞ and Y_∞ as in Definition 18.15. Using Lemma 18.19 one checks that $X_\infty \wedge Y_\infty$ is homeomorphic to $(X \times Y)_\infty$.

For example, we know that $S^n \simeq \mathbb{R}^n_\infty$ by stereographic projection (Proposition 18.20). It follows that

$$S^n \wedge S^m \simeq \mathbb{R}^n_\infty \wedge \mathbb{R}^m_\infty \simeq (\mathbb{R}^n \times \mathbb{R}^m)_\infty = \mathbb{R}^{n+m}_\infty \simeq S^{n+m}.$$

For a different kind of example, suppose that X and Y are already compact. Then X_∞ and Y_∞ can be identified with X_+ and Y_+, and the conclusion is that $X_+ \wedge Y_+ = (X \times Y)_+$.

Remark 25.12. [`rem-collapse-smash`]
Suppose we have CGWH spaces X and Y, with closed subspaces A and B. By a comparison of universal properties, one can check that there is a homeomorphism

$$\frac{X}{A} \wedge \frac{Y}{B} \simeq \frac{X \times Y}{(X \times B) \cup (A \times Y)}.$$

Proposition 25.13. [`prop-smash-symmon`]
There are natural homeomorphisms $S^0 \wedge X = X = X \wedge S^0$ and $X \wedge Y = Y \wedge X$ and $(X \wedge Y) \wedge Z = X \wedge (Y \wedge Z)$.

Proof. The only point requiring a little attention is the associativity isomorphism. Let $q_{X,Y} \colon X \times Y \to X \wedge Y$ be the quotient map. Using Proposition 23.32 we see that the composite

$$X \times Y \times Z \xrightarrow{q_{X,Y} \times 1} (X \wedge Y) \times Z \xrightarrow{q_{X \wedge Y, Z}} (X \wedge Y) \wedge Z$$

is a quotient map. This identifies $(X \wedge Y) \wedge Z$ with $(X \times Y \times Z)/A$, where

$$A = \{(x, y, z) \in X \times Y \times Z \ :\ x = 0_X \text{ or } y = 0_Y \text{ or } z = 0_Z\}.$$

We identify $X \wedge (Y \wedge Z)$ with the same space, by a symmetrical argument. $\square$

Proposition 25.14. [`prop-wedge-smash`]

The square

$$\begin{tikzcd} X \vee Y \arrow[r, tail, "k"] \arrow[d] & X \times Y \arrow[d, two heads, "q"] \\ 0 \arrow[r] & X \wedge Y \end{tikzcd}$$

is both a pushout and a pullback. Moreover k is a closed inclusion, and q is a closed quotient map.

Proof. The map k is a closed inclusion by Proposition 25.9. The map q is a quotient map by definition, and is closed by Remark 23.51. The square is a pushout by definition, and a pullback by inspection or by Proposition 23.47. □

Proposition 25.15. [`prop-smash-connected-limit`]
Let X be a based space. Then the functor $Y \mapsto X \wedge Y$ preserves equalisers (but not pullbacks or products).

Proof. Consider a pair of arrows $f, g\colon V \to W$ with equaliser $j\colon U \to V$. Consider a point $b \in X \wedge V$ with $(1 \wedge f)(b) = (1 \wedge g)(b)$. If $b = 0$ then $b = (1 \wedge j)(0)$. Otherwise $b = x \wedge v$ for a unique pair $(x, v) \in (X \setminus \{0\}) \times (V \setminus \{0\})$, in which case the equation $(1 \wedge f)(b) = (1 \wedge g)(b)$ gives $x \wedge f(v) = x \wedge g(v)$. As $x \neq 0$ this gives $f(v) = g(v)$ (even if $f(v) = 0$ or $g(v) = 0$). This means that $v = j(u)$ for a unique $u \in U$, so $b = (1 \wedge j)(x \wedge u)$. This shows that $1 \wedge j$ is the equaliser of $1 \wedge f$ and $1 \wedge g$ in the category of sets. Now consider the diagram

$$\begin{tikzcd} X \times U \arrow[r, tail, "1\times j"] \arrow[d, two heads, "p"'] & X \times V \arrow[d, two heads, "q"] \\ X \wedge U \arrow[r, tail, "1\wedge j"'] & X \wedge V \end{tikzcd}$$

It is formal that products preserve regular monomorphisms, so $1 \times j$ is a closed inclusion. The map q is closed by Remark 23.51. Given a closed subset $F \subseteq X \wedge U$ we have $F = pp^{-1}(F)$ (because p is surjective) so $(1 \wedge j)(F) = (1 \wedge j)pp^{-1}(F) = q((1 \times j)(p^{-1}(F)))$, which is closed. Thus $1 \times j$ is a closed inclusion, so it is the equaliser in $\mathbf{CGWH}_*$.

On the other hand, if we let n denote a set with n points then smashing with 2_+ does not preserve the pullback of the maps $1_+ \to 0 \leftarrow 1_+$. □

Corollary 25.16. [`cor-smash-inc`]
The functor $Y \mapsto X \wedge Y$ preserves closed inclusions.

Proof. Closed inclusions are the same as regular monomorphisms, or in other words maps that can be written as an equaliser of some fork. □

Definition 25.17. We put $F(X, Y) = \{f \in C(X, Y) \ : \ f(0_X) = 0_Y\}$, topologised as a closed subspace of $C(X, Y)$. We take the constant map with value 0_Y as a basepoint in $F(X, Y)$.

Remark 25.18. To see that $F(X, Y)$ is closed one can go back to Definition 23.21 and note that $F(X, Y) = W(\{0_X\}, \{0_Y\}^c)^c$. More abstract arguments are also possible.

Proposition 25.19. [`prop-F-adjoint`]
There are natural homeomorphisms $F(X, F(Y, Z)) = F(X \wedge Y, Z)$.

Proof. As $F(Y, Z)$ is a subspace of $C(Y, Z)$, we see that $F(X, F(Y, Z))$ is a subset of $F(X, C(Y, Z))$, which is a subset of $C(X, C(Y, Z))$, which bijects naturally with $C(X \times Y, Z)$. A function $f\colon X \times Y \to Z$ corresponds to an element of $F(X, F(Y, Z))$ iff (a) each of the functions $f(x, -)\colon Y \to Z$ preserves basepoints, and (b) the map $f(0_X, -)\colon Y \to Z$ is the zero map. These mean that $f(x, y) = 0_Z$ if $x = 0_X$ or $y = 0_Y$, or in other words that $f(X \vee Y) = 0_Z$, so f factors through $(X \times Y)/(X \vee Y) = X \wedge Y$. We thus arrive at a bijection $F(X, F(Y, Z)) = F(X \wedge Y, Z)$. One can show using the Yoneda Lemma that this is in fact a homeomorphism, as in the proof of Proposition 23.25. □

Corollary 25.20. [`cor-smash-colimits`]
The functor $(-) \wedge Y$ is left adjoint to $F(Y, -)$. Thus

(a) $(-)\wedge Y$ *preserves all colimits in* $\boldsymbol{CGWH}_*$*, and thus preserves regular epimorphisms.*
(b) $F(Y,-)$ *preserves all limits in* $\boldsymbol{CGWH}_*$*, and thus preserves regular monomorphisms.* $\square$

Two important special cases are as follows:

Definition 25.21. [`defn-sigma-omega`]
Let X be a based space. We put $I=[0,1]$ (with 0 as the basepoint), and we take $I/\partial(I)=[0,1]/\{0,1\}$ as our model of S^1. We then put

$$CX = I\wedge X \qquad\qquad \Sigma X = S^1\wedge X$$
$$PX = F(I,X) \qquad\qquad \Omega X = F(S^1,X).$$

These are called the *cone*, *suspension*, *path space* and *loop space* of X, respectively.

Remark 25.22. [`rem-sigma-omega`]
As a special case of Proposition 25.19 we have natural homeomorphisms $F(CX,Y)=F(X,PY)$ and $F(\Sigma X,Y)=F(X,\Omega Y)$.

25.2. Mapping spaces and filtered colimits.

Lemma 25.23. [`lem-filtered-based-mappings`]
If X *is compact, then the functor* $F(X,-)$ *preserves strongly filtered colimits of closed inclusions.*

Proof. This can be deduced from Lemma 24.7, or proved by the same line of argument. $\square$

Lemma 25.24. [`lem-A-FBBA`]
The unit map $\eta\colon A\to F(B,B\wedge A)$ *is a closed embedding unless* $B=0$.

Proof. Suppose that $B\neq 0$, and choose $b\in B$ with $b\neq 0$. Define $i\colon S^0\to B$ by $i(0)=0$ and $i(1)=b$; this is a closed inclusion. One checks that the following diagram commutes:

$$\begin{tikzcd} A \arrow[r, "\eta"] \arrow[dr, tail, "i\wedge 1"'] & F(B,B\wedge A) \arrow[d, "i^*"] \\ & B\wedge A. \end{tikzcd}$$

The map $i\wedge 1$ is a closed inclusion by Corollary 25.16, so η is a closed inclusion by Proposition 23.43(d). $\square$

Corollary 25.25. [`cor-inc-adj`]
If $f\colon A\wedge B\to C$ *is a closed inclusion and* $B\neq 0$ *then the adjoint map* $f^\#\colon A\to F(B,C)$ *is a closed inclusion.*

Proof. The map $f^\#$ is the composite

$$A\xrightarrow{\eta} F(B,A\wedge B)\xrightarrow{f_*} F(B,C).$$

We have just shown that η is a closed inclusion and $F(B,-)$ preserves closed inclusions so f_* is a closed inclusion; the claim follows. $\square$

Corollary 25.26. [`cor-OnSn`]
For any A*, the maps* $\Omega^n\Sigma^n A\to\Omega^{n+1}\Sigma^{n+1}A\to QA$ *are closed inclusions.*

Proof. After replacing A by $\Sigma^n A$ and taking $B=S^1$, the lemma tells us that $\Sigma^n A\to\Omega\Sigma^{n+1}A$ is a closed inclusion. The functor Ω^n preserves regular monos, so $\Omega^n\Sigma^n A\to\Omega^{n+1}\Sigma^{n+1}A$ is a closed inclusion. As $QA=\varinjlim_n \Omega^n\Sigma^n A$, the rest follows from Lemma 24.2. $\square$

Corollary 25.27. [`cor-OnSnmap`]
If $A\to B$ *is a closed embedding, then the diagram*

$$\begin{tikzcd} \Omega^n\Sigma^n A \arrow[r, tail] \arrow[d, tail] & \Omega^{n+1}\Sigma^{n+1}A \arrow[r, tail] \arrow[d, tail] & QA \arrow[d, tail] \\ \Omega^n\Sigma^n B \arrow[r, tail] & \Omega^{n+1}\Sigma^{n+1}B \arrow[r, tail] & QB \end{tikzcd}$$

consists of pullback squares of closed inclusions.

Proof. We know by Corollary 25.26 that the horizontal maps are closed inclusions. Corollary 25.16 tells us that $\Sigma^n A \to \Sigma^n B$ is a closed inclusion, and it follows that $\Omega^n\Sigma^n A \to \Omega^n\Sigma^n B$ is a closed inclusion. The left hand square is a pullback of sets, by inspection. The claim now follows from Lemma 24.8. $\square$

26. Examples

26.1. $\mathbb{R}^\infty$. We regard $\mathbb{R}^n$ as a subspace of $\mathbb{R}^{n+1}$ in the obvious way, and put $\mathbb{R}^\infty = \varinjlim_n \mathbb{R}^n$. Each $\mathbb{R}^n$ is a metric space and thus CGWH, so $\mathbb{R}^\infty$ is CGWH. If we let K_n be the closed ball of radius n in $\mathbb{R}^n$ then K_n is compact and $\mathbb{R}^\infty = \bigcup_n K_n$. Thus $\mathbb{R}^\infty$ is σ-compact (ie a countable union of compact sets). It follows that every open cover has a countable subcover. Let U be an open set with compact closure. Then $U \subseteq \mathbb{R}^n$ for some n and $U = U \cap \mathbb{R}^{n+1}$ is open in $\mathbb{R}^{n+1}$; it follows that $U = \emptyset$. This shows that $\mathbb{R}^\infty$ is not locally compact.

26.2. $[0,\omega_1)$. Let ω_1 be the first uncountable ordinal, and give the set $[0,\omega_1)$ the order topology. This is an open subspace of the compact Hausdorff space $[0,\omega_1]$, so it is locally compact Hausdorff and thus CGWH. It is easy to see that a subset is countable iff bounded iff precompact. It follows easily that $[0,\omega_1)$ is neither separable nor σ-compact. The open sets $[0,\alpha)$ (with α countable) form an open cover with no countable subcover, so $[0,\omega_1)$ is not Lindelöf. I think that every countable open cover has a finite subcover, ie $[0,\omega_1)$ is countably compact.

26.3. The long line. Let α be any ordinal, and put $L^+(\alpha) = \alpha \times [0,1)$, ordered lexicographically, so $(\beta,s) < (\gamma,t)$ iff ($\beta<\gamma$ or ($\beta=\gamma$ and $s<t$). We give this the order topology, with a subbase consisting of sets $D(x) = \{y \;:\; y<x\}$ and $U(x) = \{y \;:\; y>x\}$. This is easily seen to be Hausdorff. We also define $L(\alpha)$ to be the quotient of $\{1,-1\} \times L^+(\alpha)$ in which $(-1,0,0)$ is identified with $(1,0,0)$.

The spaces $L^+(\omega_1)$ and $L(\omega_1)$ are called the *long ray* and the *long line*.

If $\alpha < \omega_1$ then α is countable so we can choose a bijection $d\colon \alpha \to \mathbb{N}$, and define $f\colon L(\alpha) \to [0,1)$ by $f(\beta,s) = s2^{-d(\beta)} + \sum_{\gamma<\beta} 2^{-d(\gamma)}$. This is an order-isomorphism and thus a homeomorphism. It follows that $L^+(\omega_1)$ is locally homeomorphic to $[0,1)$. However, it is easy to see that every countable subset of $L^+(\omega_1)$ is contained in $L^+(\alpha)$ for some countable ordinal α, so $(\alpha^+,1/2)$ is not in the closure. This means that $L^+(\omega_1)$ is not separable, and so is not homeomorphic to $[0,1)$.

Check that it is not paracompact or second countable.

Note that the long line is Hausdorff and locally homeomorphic to $\mathbb{R}$, so it would count as a topological manifold, if we had not included second countability as part of the definition.

26.4. $C(I,I)$. The space $C(I,I)$ is separable, because the set of piecewise linear functions with rational breakpoints and slopes is a countable dense subset. Put

$$F_n = \{f\colon I \to I \;:\; f(0)=0 \text{ and } f(t)=1 \text{ for } t \geq 1/n\}.$$

This is a decreasing sequence of closed sets with empty intersection. It follows that the sets $U_n = F_n^c$ form a countable open cover with no finite subcover, showing that $C(I,I)$ is not countably compact and thus not σ-compact.

26.5. $(\beta\mathbb{N}) \setminus \{\omega\}$. Let $\beta\mathbb{N}$ be the Stone-Cech compactification of $\mathbb{N}$, and choose a point $\omega \in (\beta\mathbb{N}) \setminus \mathbb{N}$ (so ω is a free ultrafilter on $\mathbb{N}$). Put $X = (\beta\mathbb{N}) \setminus \{\omega\}$; this is clearly a locally compact Hausdorff space. For any $T \subseteq \mathbb{N}$ we put $V(T) = \{\alpha \in \beta\mathbb{N} \;:\; T \in \alpha\}$. These sets satisfy $V(S \cup T) = V(S) \cup V(T)$ and $V(S \cap T) = V(S) \cap V(T)$ and $V(T) \cap \mathbb{N} = T$. They form a basis for the topology on $\beta\mathbb{N}$, consisting of compact clopen sets. It follows that $\{V(T) \;:\; T \notin \omega\}$ is a basis of open sets in X, which is closed under finite unions and intersections. It follows in turn that every compact subspace of X is contained in some $V(T)$ with $T \notin \omega$.

Lemma 26.1. [`lem-cech-chain`]

Consider a descending chain $T_0 \supseteq T_1 \supseteq \dots$ of subsets of $\mathbb{N}$. Then either $T_n = \emptyset$ for some n, or $X \cap \bigcap_n V(T_n) \neq \emptyset$.

Proof. First suppose that $\bigcap_n T_n$ is nonempty, containing a number a say. The image of a under the embedding $\mathbb{N} \to \beta\mathbb{N}$ is the ultrafilter $\alpha = \{T \subseteq \mathbb{N} \,:\, a \in T\}$, which certainly lies in $\bigcap_n V(T_n)$. Moreover, as $\omega \in (\beta\mathbb{N}) \setminus \mathbb{N}$ we have $\alpha \neq \omega$ and so $\alpha \in X \cap \bigcap_n V(T_n)$, as required.

Now suppose that $\bigcap_n T_n = \emptyset$. If T_n is finite for some n, it is easy to see that $T_n = \emptyset$ for some larger n. Thus, we may assume that T_n is always infinite. It follows that we can choose numbers $a_1, a_2, \dots, b_1, b_2, \dots$, all of them distinct, such that $a_n, b_n \in T_n$. Put $A = \{a_n \,:\, n > 0\}$ and $B = \{b_n \,:\, n > 0\}$ so the sets $A \cap T_n$ and $B \cap T_n$ are nonempty for all n. As $A \cap B = \emptyset$ and ω is an ultrafilter we must have $A \notin \omega$ or $B \notin \omega$; we assume wlog that $A \notin \omega$. Now put

$$\phi = \{S \subseteq \mathbb{N} \,:\, S \supseteq A \cap T_n \text{ for some } n\}.$$

This is a proper filter, so we can choose an ultrafilter α with $\phi \subseteq \alpha$. Clearly $A \in \phi$ and $A \notin \omega$ so $\alpha \neq \omega$ so $\alpha \in X$. For each n we have $T_n \in \phi$ so $T_n \in \alpha$ so $\alpha \in V(T_n)$. This shows that $\alpha \in X \cap \bigcap_n V(T_n)$ as claimed. □

Lemma 26.2. *X is not σ-compact.*

Proof. Any compact subset of X is contained in $V(T)$ for some T with $T \notin \omega$. It will thus suffice to show that for any chain $T_1 \subseteq T_2 \subseteq \dots$ with $T_n \notin \omega$ for all n, we have $\bigcup_n V(T_n) \neq X$. As $V(T)^c = V(T^c)$, it is equivalent to show that $X \cap \bigcap_n V(T_n^c) \neq \emptyset$. This will follow from Lemma 26.1, if we can show that $T_n^c \neq \emptyset$, or equivalently $T_n \neq \mathbb{N}$. This holds because $T_n \notin \omega$, whereas $\mathbb{N}$ lies in every ultrafilter. □

Corollary 26.3. *X is not Lindelöf.*

Proof. The sets $V(T)$ (with $T \notin \omega$) form an open cover of X. If X were Lindelöf, then there would be a countable subcover, and as the sets $V(T)$ are compact, this would mean that X was σ-compact, giving a contradiction. □

Lemma 26.4. [`lem-cone-nbhd`]
Let U be an open subset of $I \times X$ containing $0 \times X$. Then there exists $m \in \mathbb{N}$ such that $[0, 2^{-m}) \times \mathbb{N} \subseteq U$.

Proof. Put

$$T_m := \{n \in \mathbb{N} \,:\, [0, 2^{-m}) \times \{n\} \subseteq U\}.$$

We clearly have $T_m \subseteq T_{m+1}$. For any $n \in \mathbb{N}$ we know that U is open and contains $(0, n)$, so we have $[0, 2^{-m}) \times \{n\} \subseteq U$ for $m \gg 0$; this shows that $\bigcup_m T_m = \mathbb{N}$, so $\bigcap_m T_m^c = \emptyset$. More generally, consider $\alpha \in X$. We again know that U is a neighbourhood of $(0, \alpha)$, so there exists $S \in \alpha$ and $m \in \mathbb{N}$ with $[0, 2^{-m}) \times V(S) \subseteq U$. As $V(S) \cap \mathbb{N} = S$ we deduce that $[0, 2^{-m}) \times S \subseteq U$ and so $S \subseteq T_m$, so $V(S) \subseteq V(T_m)$, so $\alpha \in V(T_m)$. This shows that $X \subseteq \bigcup_m V(T_m)$, so $X \cap \bigcap_m V(T_m^c) = \emptyset$. Lemma 26.1 thus tells us that for $m \gg 0$ we have $T_m^c = \emptyset$ and so $T_m = \mathbb{N}$. This means that $[0, 2^{-m}) \times \mathbb{N} \subseteq U$, as required. □

Corollary 26.5. [`cor-not-locomp`]
The space $Y := (I \times X)/(0 \times X) = I \wedge X_+$ is not locally compact.

Proof. Let N be a neighbourhood of the cone point in Y, and let $\widetilde{N}$ be its preimage in $I \times X$. Then claim 26.4 tells us that $[0, 2^{-m}) \times X \subseteq \widetilde{N}$ for some m. It follows that the set $L = \{2^{-m-1} \wedge x \,:\, x \in X\}$ is contained in N. Moreover, L is closed in Y and is not compact, so N cannot be compact. □

Corollary 26.6. *Put*

$$\widetilde{K} = \{(t, n) \in I \times \mathbb{N} \,:\, t \leq 2^{-n}\} \subset I \times X,$$

and let K be the image of $\widetilde{K}$ in Y. Then K is compact, but there is no compact set $L \subseteq X$ with $K \subseteq I \wedge L_+$.

Proof. Let $\{U_i\}_{i \in I}$ be a family of open sets in Y that cover K. Then some set U_{i_0} contains the basepoint, and thus (by Claim 26.4) contains $[0, 2^{-m}) \wedge X_+$ for some m. It follows easily that $K \setminus U_{i_0}$ is compact, and thus is covered by some finite list $U_{i_1}, \dots, U_{i_r}$, so K is covered by $U_{i_0}, \dots, U_{i_r}$. Thus, K is compact as claimed.

If $K \subseteq I \wedge L_+$ with L closed we see that $\mathbb{N} \subseteq L$, but $\mathbb{N}$ is dense so we must have $L = X$, so in particular L is not compact. □

26.6. A non-regular space. Let ζ be the lattice of closed sets for the usual topology on $\mathbb{R}$. For $k \geq 0$, write $S_k = \{1/n \ : \ n > k\}$, and write $S = S_0$. Define

$$\begin{aligned} (1) \qquad \zeta' &= \{F \cup T \ : \ F \in \zeta \text{ and } T \subseteq S\} \\ (2) \qquad &= \{F \cup T \ : \ F \in \zeta \text{ and } T \subseteq S_k\} \qquad \text{(for any } k \geq 0) \\ (3) \qquad &= \{G \subseteq \mathbb{R} \ : \ G \cap S^c \in \zeta|_{S^c}\} \end{aligned}$$

(where $\zeta|_{S^c}$ means the ordinary subspace topology on S^c).

Proposition 26.7. *ζ' is a compactly generated weak Hausdorff topology on $\mathbb{R}$. Moreover,*

$$c(\zeta') = \{K \in c(\zeta) \ : \ K \cap S \text{ is finite } \} = \{K \in c(\zeta) \ : \ K \cap S_k = \emptyset \text{ for } k \gg 0\},$$

and for any $K \in c(\zeta')$ we have $\zeta'|_K = \zeta|_K$. However, ζ' is not regular (so there is a closed set and a point which cannot be separated by disjoint neighbourhoods).

Proof. The last description of ζ' makes it clear that it is a topology. As $\zeta \subseteq \zeta'$, we see that ζ' is Hausdorff, and that $c(\zeta') \subseteq c(\zeta)$. Suppose that $K \in c(\zeta')$. For any $T \subseteq S$ we have $T \in \zeta'$, so $T \cap K$ is closed in K. Thus $S \cap K$ is compact and discrete, hence finite as claimed; thus $S_k \cap K = \emptyset$ for $k \gg 0$. Using this and the second description of ζ', we see that $\zeta'|_K = \zeta|_K$.

Conversely, suppose that $K \in c(\zeta)$ and that $K \cap S_k = \emptyset$ for some k. Then the second description of ζ shows that $\zeta|_K = \zeta'|_K$, so that K is also compact under ζ'. This verifies all the claims about $c(\zeta')$.

We next prove that ζ' is compactly generated. Let F be compactly closed for ζ'; we must show that $F \in \zeta'$. Let G be the closure of F in the ordinary topology ζ (so $F \subseteq G$). If $G = F$, then $F \in \zeta \subset \zeta'$, so we are done; so suppose that $G \neq F$. We claim that $G = F \cup \{0\}$ and $0 \notin F$. Indeed, suppose that $0 \neq x \in G$. If x has the form $1/n$ then set $k = n$, otherwise set $k = 0$. We can then choose a sequence x_i in $F \setminus S_k$ converging to x in the usual metric. Moreover, $K = \{x_i \ : \ i \geq 0\} \cup \{x\}$ is compact under ζ', so by assumption $F \cap K$ is closed in $\zeta'|_K = \zeta|_K$. It follows that $x \in F$. This (with $G \neq F \subset G$) implies immediately that $G = F \cup \{0\}$ and $0 \notin F$. If $(-1/n, 1/n) \cap F$ is not contained in S for any n, then we may choose $x_n \in (-1/n, 1/n) \cap F$ (so that $x_n \to 0$) and proceed as as above to deduce that $0 \in F$, contrary to assumption; thus $T = (-1/n, 1/n) \cap F \subseteq S$ for some n. Write $F' = F \setminus (-1/n, 1/n) = G \setminus (-1/n, 1/n)$, so that $F' \in \zeta$. Thus $F = F' \cup T \in \zeta'$ as required.

Finally, it is easy to see that the closed sets $\{0\}$ and S cannot be separated by disjoint open neighbourhoods in ζ'. Thus, ζ' is not regular. $\square$

26.7. Bad sequential colimits.

Proposition 26.8. *There is a diagram*

$$\begin{tikzcd} W_0 \arrow[r, two heads] \arrow[d, tail] & W_1 \arrow[r, two heads] \arrow[d, tail] & W_2 \arrow[r, two heads] \arrow[d, tail] & \cdots \\ X_0 \arrow[r, two heads] & X_1 \arrow[r, two heads] & X_2 \arrow[r, two heads] & \cdots \end{tikzcd}$$

in which the vertical maps are closed inclusions, but the induced map $\varinjlim_n W_n \to \varinjlim_n X_n$ is not injective. (It follows that finite limits do not commute with sequential colimits in general.)

Proof. Put $X' = [0,1] \times \{0,1\}$ and $W' = \{0\} \times \{0,1\}$. Define an equivalence relation R_n on X' by $(s,a)R_n(t,b)$ if $s = t$ and ($a = b$ or $s \geq 2^{-n}$). Put $X_n = X'/R_n$ and $W_n = W'$. One finds that the induced map $W_n \to X_n$ is a closed inclusion, but $\varinjlim_n W_n = W'$ and $\varinjlim_n X_n = I$ but the induced map $W' \to I$ sends both points of W' to 0. $\square$

26.8. Irregular evaluation.

Proposition 26.9. *There is a path connected based space X for which the evaluation map $\epsilon \colon PX \to X$ is not a quotient map.*

Proof. Put

$$\begin{aligned} a &= (0,0,0) \\ b_n &= (0,n,1) \text{ for } n < \infty \\ b_\infty &= (0,-1,1) \\ c_n &= (2^{-n},0,2) \text{ for } n < \infty \\ c_\infty &= (0,0,2). \end{aligned}$$

Note that

(a) $d(b_i,b_j) > 1$ whenever $i \neq j$
(b) $c_n \to c_\infty$ as $n \to \infty$
(c) The line segments $(a,b_i]$ and $(b_i,c_i]$ are all disjoint.

Let X be the union of all the line segments $[a,b_i]$ and $[b_i,c_i]$, and take a as the basepoint. Put $U = (b_\infty, c_\infty]$, which is not open in X because $c_n \to c_\infty$. Put $V = \epsilon^{-1}U \subseteq PX$. I'm fairly sure that this is open in PX, proving the claim. □

26.9. Discontinuity of the Pontrjagin-Thom construction. Given a locally compact Hausdorff space U, we write U_∞ for the one-point compactification. Given an open embedding $i\colon U \to V$ of such spaces, we define $i^\bullet\colon V_\infty \to U_\infty$ by $i^\bullet(i(u)) = u$ and $i^\bullet(v) = \infty$ for v not in the image of i. We write $\mathrm{Emb}(U,V)$ for the space of open embeddings from U to V, with the Kellification of the subspace topology inherited from $C(U,V)$. We then define a function $\phi\colon \mathrm{Emb}(U,V) \to F(V_\infty, U_\infty)$ by $\phi(i) = i^\bullet$. I think that this is often continuous, for example when U and V are manifolds. Here, however, we will give an example where ϕ is *not* continuous.

First, we take $U = V = \prod_{k=0}^\infty \{0,1\}$. We define $f_k\colon U \to U$ by

$$f_k(x)_i = \begin{cases} x_i & \text{if } i < k \\ 0 & \text{if } i = k \\ x_{i-1} & \text{if } i > k. \end{cases}$$

This is isomorphic to the product of 1_U with the inclusion $\{0\} \to \{0,1\}$, so it is an open embedding. I claim that $f_k \to 1$ in $C(U,U)$. To see this, note that $C(U,U) \simeq \prod_i C(U,\{0,1\})$, and convergence in product spaces is detected termwise, so it suffices to show that $\pi_i \circ f_k \to \pi_i$ as $k \to \infty$, for all $i \geq 0$. This is clear because $\pi_i \circ f_k = \pi_i$ when $k > i$.

Now put $e_k = 1$ for all k, giving an element $e \in U$. For any $x \in U$ we have $f_k(x)_k = 0 \neq e_k$, so e is not in the image of f_k, so $f_k^\bullet(e) = \infty$. Thus $f_k^\bullet(e) \not\to e = 1^\bullet(e)$, so $f_k^\bullet \not\to 1^\bullet$.

26.10. Bad pullbacks. We remarked earlier that although **CGWH** is regular (so regular epis are preserved by pullback), coequaliser diagrams need not be preserved by pullback. We now give an example of this behaviour. Let X and Y be CGWH spaces, and let U be a dense open subspace of Y with $Z = Y \setminus U$. Consider the following diagram, in which all the maps are the evident inclusions and projections:

$$\begin{tikzcd} \emptyset \arrow[r, shift left] \arrow[r, shift right] \arrow[d] & X \times Z \arrow[r] \arrow[d] & Z \arrow[d] \\ X^2 \times U \arrow[r, shift left] \arrow[r, shift right] & X \times Y \arrow[r] & Y. \end{tikzcd}$$

We claim that the bottom line is a coequaliser. To see this, let R be the smallest equivalence relation on $X \times Y$ such that $(x,u)R(x',u)$ whenever $x,x' \in X$ and $u \in U$. As subsets of $X^2 \times Y^2 = (X \times Y)^2$, one can check that

$$R = (\Delta_X \times \Delta_Y) \cup (X^2 \times \Delta_U).$$

The set $\overline{R} = X^2 \times \Delta_Y$ is a closed equivalence relation and R is dense in $\overline{R}$ so $\overline{R}$ is the smallest closed equivalence relation containing R. Clearly also $(X \times Y)/\overline{R} = Y$, and it follows that the bottom line is a coequaliser as claimed. The top row is obtained by pulling back the bottom row along the map $Z \to Y$.

Although the right hand map on the top row is a regular epimorphism (as required for regularity) the row itself is not a coequaliser diagram.

27. Basics of homotopy theory

DEFINITION 27.1. [`defn-homotopy`]
Let f_0 and f_1 be continuous maps from a topological space X to another space Y. A *homotopy* from f_0 to f_1 is a path from f_0 to f_1 in the space $C(X,Y)$. If there exists such a homotopy, we say that f_0 and f_1 are *homotopic* and write $f_0 \simeq f_1$.

REMARK 27.2. [`rem-continuous-family`]
A homotopy can be thought of as a family of maps $(f_t)_{t\in[0,1]}$ such that the map $t \mapsto f_t$ (from $[0,1]$ to $C(X,Y)$) is continuous. Given such a family, we can define a single map $F\colon [0,1]\times X \to Y$ by $F(t,x) = f_t(x)$, and Proposition 23.25 tells us that continuity of F is equivalent to continuity of $t \mapsto f_t$.

REMARK 27.3. [`rem-homotopy-classes`]
As in Definition 8.1 we see that the relation of being homotopic is an equivalence relation. The equivalence classes are called *homotopy classes* of maps. The set $\pi_0 C(X,Y)$ of homotopy classes is also denoted by $[X,Y]$. If X and Y are based spaces we also write $[X,Y]_*$ for $\pi_0 F(X,Y)$.

EXAMPLE 27.4. [`eg-linear-homotopy`]
Suppose we have maps $f_0, f_1\colon X \to Y$, where Y is a subset of a real vector space V (with the subspace topology). For $(t,x) \in [0,1]\times X$ we can then define $F(t,x) = (1-t)f_0(x) + tf_1(x) \in V$. If it happens that $F(t,x) \in Y$ for all t and x, then we have a homotopy between f_0 and f_1, which we call the *linear homotopy*. It is distressingly easy to fall into the trap of writing such formulae without verifying that $F(t,x) \in Y$.

In the case where $Y = V$, of course, there is nothing to check: any two maps from X to V are homotopic by a linear homotopy.

EXAMPLE 27.5. [`eg-sphere-homotopies`]
Define maps $f_i\colon S^2 \to S^2$ by

$$\begin{aligned} f_0(x,y,z) &= (x,y,z) \\ f_1(x,y,z) &= (-x,y,z) \\ f_2(x,y,z) &= (-x,-y,z). \end{aligned}$$

There is a homotopy between f_0 and f_2 given by

$$F(t,(x,y,z)) = (\cos(\pi t)x - \sin(\pi t)y, \sin(\pi t)x + \cos(\pi t)y, z).$$

(In other words, at time t we rotate about the z-axis by an angle πt.) However, it turns out that f_0 is not homotopic to f_1. It is possible but difficult to prove this directly. One approach is as follows. Consider a map $F\colon [0,1]\times S^2 \to S^2$, and put $C = F^{-1}\{(0,0,1)\} \subseteq [0,1]\times S^2$. If we merely asume that F is continuous, then C could be very complicated; it could even be fractal, for example, or a knot with infinitely many loops. However, after adjusting F by an arbitrarily small amount, we can arrange that F is continuously differentiable, and that C is a smooth curve, and that a certain kind of coincidental vanishing of derivatives does not happen anywhere on C. Then, by considering how the derivatives vary along C, one can check that it is impossible for F to be a homotopy from f_0 to f_1. A much simpler and more general approach is to use invariants from algebraic topology, such as homotopy or homology groups. **Will we give further details anywhere?**

EXAMPLE 27.6. [`eg-maps-to-sphere`]
Consider two maps $f_0, f_1\colon X \to S^n$ such that $f_0(x) + f_1(x) \neq 0$ for all x. Define $F\colon [0,1]\times X \to \mathbb{R}^{n+1}$ by

$$F(t,x) = (1-t)f_0(x) + tf_1(x) \in \mathbb{R}^{n+1}.$$

Except in the trivial case where $f_0 = f_1$, this will not give a homotopy from f_0 to f_1, simply because the values $F(t,x)$ will not lie in S^n. However, we see as in Example 8.8 that $F(t,x)$ is never zero, so we can regard F as a continuous map from X to $\mathbb{R}^{n+1}\setminus\{0\}$. There is a continuous map $r\colon \mathbb{R}^{n+1}\setminus\{0\} \to S^n$ given by $r(v) = v/\|v\|$, so we can define $G = r \circ F\colon [0,1] \times X \to S^n$. We then have $G(0,x) = r(f_0(x)) = f_0(x)/\|f_0(x)\| = f_0(x)$ and similarly $G(1,x) = f_1(x)$, so this gives a homotopy from f_0 to f_1.

The following principle will often be useful for specifying homotopy classes of maps.

Proposition 27.7. [`prop-connected-parameters`]
Let P be a path connected space. Suppose we have a continuous map $F\colon P \times X \to Y$, and we define $f_p\colon X \to Y$ by $f_p(x) = F(p,x)$. Then the homotopy class of f_p is independent of $p \in P$.

Proof. The rule $p \mapsto f_p$ gives a map $F^{\#}\colon P \to C(X,Y)$, which is continuous by Proposition 23.25. For any two points $p, q \in P$ we can choose a path $u\colon [0,1] \to P$ with $u(0) = p$ and $u(1) = q$, and then the composite $F^{\#} \circ u$ (given by $t \mapsto f_{u(t)}$) is a homotopy from f_p to f_q. $\square$

Lemma 27.8. [`lem-composite-homotopy`]
Suppose we have maps

$$\begin{tikzcd} X \arrow[r, shift left, "f_0"] \arrow[r, shift right, "f_1"'] & Y \arrow[r, shift left, "g_0"] \arrow[r, shift right, "g_1"'] & Z \end{tikzcd}$$

where $f_0 \simeq f_1$ and $g_0 \simeq g_1$. Then $g_0 f_0 \simeq g_1 f_1$.

Proof. Choose a homotopy F from f_0 to f_1, and a homotopy G from g_0 to g_1, and then put $H(t,x) = G(t, F(t,x))$; this gives a homotopy from $g_0 f_0$ to $g_1 f_1$. $\square$

Corollary 27.9. [`cor-homotopy-category`]
*There is a well-defined composition operation $[Y,Z] \times [X,Y] \to [X,Z]$, given by $([g],[f]) \mapsto [gf]$. We can thus define a category **hCGWH**, whose objects are spaces and whose morphisms are homotopy classes of map. There is a functor $U\colon$ **CGWH** $\to$ **hCGWH** given by $U(X) = X$ on objects, and $U(f) = [f]$ on morphisms. Similarly, there is a category $\boldsymbol{hCGWH}_*$ whose objects are based CGWH spaces and whose morphism sets are the sets $[X,Y]_*$.* $\square$

Definition 27.10. [`defn-homotopy-equivalence`]
A map $f\colon X \to Y$ is a *homotopy equivalence* if there exists a *homotopy inverse* $g\colon Y \to X$ with $gf \simeq 1_X$ and $fg \simeq 1_Y$. We say that X and Y are *homotopy equivalent* if there exists a homotopy equivalence between them. We say that X is *contractible* if it is homotopy equivalent to a single point.

Before giving some examples, it will be convenient to reformulate the notion of contractibility a little.

Definition 27.11. A *contraction* of a space X is a map $h\colon [0,1] \times X \to X$ such that $h(1,x) = x$ for all x, and $h(0,x)$ is independent of x.

Lemma 27.12. [`lem-contractible`]
A space X is contractible if and only if it has a contraction.

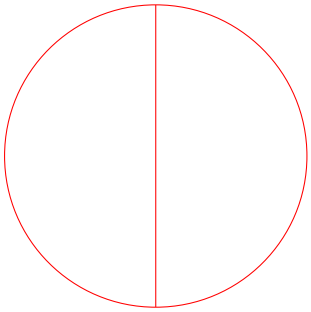
Proof. Let $h\colon [0,1] \times X \to X$ be a contraction. This means that there is a point $x_0 \in X$ such that $h(0,x) = x_0$ for all $x \in X$. Let Y denote the singleton space $\{0\}$. Let f be the unique map from X to Y, given by $f(x) = 0$ for all x. Let g be the map from Y to X given by $g(0) = x_0$. Then fg is equal (and therefore homotopic) to 1_Y, and h gives a homotopy from gf to 1_X. This proves that f is a homotopy equivalence, as required. The converse is essentially the same. $\square$

Example 27.13. [`eg-linear-contraction`]
Let V be any vector space; then the map $h(t,x) = tx$ defines a contraction of V. More generally, let X be any subset of V, and let x_0 be a point of X. We say that X is *star-shaped around* x_0 if for all $x \in X$, the line segment from x to x_0 is contained wholly in X. If so, then the formula $h(t,x) = tx + (1-t)x_0$ gives a contraction of X.

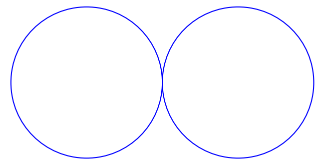
Example 27.14. [`eg-punctured-vector-space`]
Let V be a vector space with inner product, and put $V^\times = V \setminus \{0\}$. Let $i\colon S(V) \to V^\times$ be the inclusion, and define $r\colon V^\times \to S(V)$ by $r(v) = v/\|v\|$. Then $ri = 1_{S^n}$, and we can define a homotopy from $1_{V^\times}$ to ir by $h(t,v) = v/\|v\|^t$. This shows that i and r are mutually inverse homotopy equivalences.

Example 27.15. We next claim that $\mathbb{R} \setminus \mathbb{Z}$ is homotopy equivalent to $\mathbb{Z}$. Indeed, we can define $f\colon \mathbb{Z} \to \mathbb{R} \setminus \mathbb{Z}$ by $f(n) = n + 1/2$, and we can define $g\colon \mathbb{R} \setminus \mathbb{Z} \to \mathbb{Z}$ by $g(x) = n$ whenever $n < x < n+1$. Then gf is equal to the identity, and fg is linearly homotopic to the identity.

Example 27.16. For an example where neither fg nor gf is equal to the identity, consider the spaces X and Y pictured below:

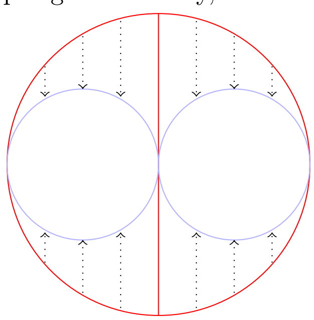


X

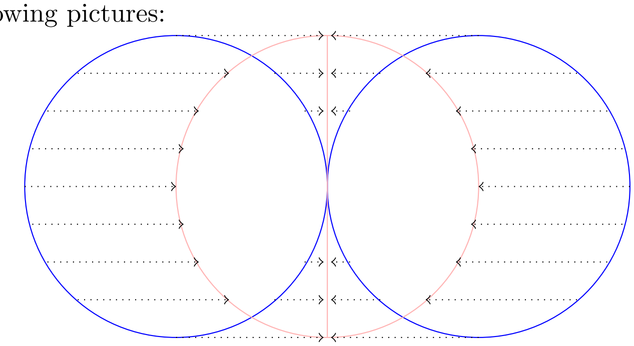


Y

We can define $f\colon X \to Y$ by collapsing vertically, and we can define $g\colon Y \to X$ by doubling and then collapsing horizontally, as illustrated by the following pictures:



f



g

The composites fg and gf are then homotopic to the respective identity maps. The reader should be able to visualise the required homotopies, but we will not attempt to give formulae.

One problem in understanding the homotopy sets $[X, Y]$ is that there may exist maps $f\colon X \to Y$ that are very wild, as with the space-filling curves discussed in Section 11. However, in many cases this does not cause trouble, because one can find a well-behaved map g that is very close to f and is homotopic to it. There are a number of different results of this type **(refer to simplicial approximation theorem)**; here we will prove one that is quite simple and flexible.

Definition 27.17. [`defn-polynomial`]
Let X be a subspace of $\mathbb{R}^n$, and let Y be a subspace of $\mathbb{R}^m$. We say that a map $f\colon X \to Y$ is *polynomial* if there exist polynomials $p_j(x_0, \dots, x_{n-1})$ (for $0 \leq j < m$) such that

$$f(x) = (p_0(x), \dots, p_{m-1}(x))$$

for all $x \in X$. We also say that a subset $W \subseteq X$ is *Zariski closed* in X if there are polynomials $w_0(x), \dots, w_{r-1}(x)$ such that

$$W = \{x \in X \;:\; w_k(x) = 0 \text{ for all } k\}.$$

Lemma 27.18. [`lem-zariski-closed`]

(a) *If W is Zariski closed in X, then there is a single polynomial $w(x)$ such that $w(x) \geq 0$ for all $x \in X$, and $W = X \cap w^{-1}\{0\}$.*
(b) *If W_0 and W_1 are Zariski closed in X, then so are $W_0 \cap W_1$ and $W_0 \cup W_1$.*

Proof.

(a) If W is Zariski closed, then there exist polynomials $w_0(x), \dots, w_{r-1}(x)$ such that

$$W = \{x \in X \;:\; w_k(x) = 0 \text{ for all } k\}.$$

The polynomials $w(x) = \sum_k w_k(x)^2$ then has the stated properties.

(b) By part (a) there are nonnegative polynomials w_0 and w_1 with $X \cap w_i^{-1}\{0\} = W_i$ for $i = 0, 1$. If we put $w_2(x) = w_0(x) + w_1(x)$ and $w_3(x) = w_0(x)w_1(x)$ we find that $X \cap w_2^{-1}\{0\} = W_0 \cap W_1$ and $X \cap w_3^{-1}\{0\} = W_0 \cup W_1$.

□

PROPOSITION 27.19. [prop-polynomial-approx]
Let X be a compact subspace of $\mathbb{R}^n$, and let Y be an open subspace of $\mathbb{R}^m$. Use the norm $\|y\| = \max(|y_0|, \dots, |y_{m-1}|)$ and the corresponding metric on $\mathbb{R}^m$.

(a) *For any continuous map $f\colon X \to Y$ and any $\epsilon > 0$ there exists a polynomial $g\colon X \to Y$ with $d(f, g) < \epsilon$.*
(b) *Let $W \subseteq X$ be Zariski closed, and suppose that $f|_W$ is already polynomial. Then in (a) we can choose g such that $g|_W = f|_W$.*
(c) *Any continuous map $f\colon X \to Y$ is homotopic to a polynomial map.*
(d) *Suppose that $f, g\colon X \to Y$ are polynomial maps that are homotopic. Then there exists a homotopy $h\colon [0,1] \times X \to Y$ between them that is itself a polynomial map.*

PROOF. (a) First, as $f(X)$ is compact and Y is open, the number

$$\eta = d(f(X), Y^c) = \inf\{d(f(x), z) \;:\; x \in X,\ z \in \mathbb{R}^m \setminus Y\}$$

is strictly positive. Next, the Stone-Weierstrass Theorem (Theorem 17.10) easily implies that the set of polynomial functions $X \to \mathbb{R}$ is dense in $C(X, \mathbb{R})$. We can thus find polynomials g_k for $0 \le k < m$ with $d(f_k, g_k) < \min(\eta, \epsilon)$ for all k. We can combine these to give a map $g\colon X \to \mathbb{R}^m$ with $d(f, g) < \min(\eta, \epsilon)$. As $d(f, g) < \eta$ we see that $g(X) \subseteq Y$, as required.

(b) Now let W be a Zariski closed subset of X, so we can choose a nonnegative polynomial function $w\colon X \to \mathbb{R}_+$ with $X \cap w^{-1}\{0\} = W$. Note that as X is compact, the set $w(X)$ must be bounded. After dividing by a positive constant if necessary, we may assume that $w(X) \subseteq [0, 1]$.

Now suppose we have a continuous function $f\colon X \to Y$ and that there exists a polynomial $p\colon \mathbb{R}^n \to \mathbb{R}^m$ such that $f|_W = p|_W$. Suppose we are also given $\epsilon > 0$, and we define η as before and put $\epsilon' = \min(\eta, \epsilon)/2$. By part (a), we can find a polynomial function $q\colon X \to Y$ with $\|f(x) - q(x)\| < \epsilon'$ for all $x \in X$. In particular, we see that the polynomial $r(x) = p(x) - q(x)$ has $\|r(w)\| < \epsilon'$ for all $w \in W$. Put $Z = \{z \in X \;:\; \|r(z)\| \ge \epsilon'\}$ and $R = \sup\{\|r(x)\| \;:\; x \in X\}$. As Z is compact and disjoint from W and $W = X \cap w^{-1}\{0\}$ we see that the number $\zeta = \min\{w(z) \;:\; z \in Z\}$ lies in $(0, 1]$. If N is large enough we will have $(1 - \zeta)^N < \epsilon'/R$ and also $0 \le (1 - \zeta)^N < 1$. It follows that the polynomial $s(x) = (1 - w(x))^N r(x)$ has $\|s(x)\| < \epsilon'$ for all $x \in X$, and $s(w) = r(w) = p(w) - q(w)$ for all $w \in W$. This means that the function $g = (s + q)|_X$ is polynomial with $g|_W = f|_W$ and $d(f, g) < 2\epsilon' = \min(\epsilon, \eta)$, so $g(X) \subseteq Y$ as required.

(c) If we choose g as in the proof of (a), then g is a polynomial that is linearly homotopic to f.

(d) Suppose that $f, g\colon X \to Y$ are polynomial maps, and that $k\colon [0,1] \times X \to Y$ is a homotopy between them. The set $\{0, 1\} \times X = \{(t, x) \in \mathbb{R} \times X \;:\; t(t - 1) = 0\}$ is Zariski closed in $[0, 1] \times X$, so by (b) we can find a polynomial map $h\colon [0, 1] \times X \to Y$ that agrees with k on $\{0, 1\} \times X$. This is the required polynomial homotopy from f to g.

□

28. Coverings and the fundamental groupoid

DEFINITION 28.1. [defn-path-homotopy]
Let X be a CGWH space. We write $\mathrm{Path}(X)$ for the space $C([0,1], X)$ of paths in X. We define continuous maps $\sigma, \tau\colon \mathrm{Path}(X) \to X$ (called *source* and *target*) by $\sigma(u) = u(0)$ and $\tau(u) = u(1)$.

Given points $x, y \in X$, we also put

$$\begin{aligned}\mathrm{Path}(X)(x,y) &= \{\text{ paths in } X \text{ from } x \text{ to } y\ \} \\ &= \{u \in \mathrm{Path}(X) \ :\ \sigma(u) = x \text{ and } \tau(u) = y\}.\end{aligned}$$

Suppose we have two paths $u_0, u_1 \in \mathrm{Path}(X)(x, y)$. A *path-homotopy* from u_0 to u_1 is a continuous map $U\colon [0,1]^2 \to X$ with

(a) $U(s,0) = x$ and $U(s,1) = y$ for all s
(b) $U(0,t) = u_0(t)$ and $U(1,t) = u_1(t)$ for all t.

Remark 28.2. [rem-path-homotopy]
Just as in Remark 27.2, a path-homotopy U gives a family of paths $u_s \in \mathrm{Path}(X)(x,y)$ by $u_s(t) = U(s,t)$, and these give a path from u_0 to u_1 in the space $\mathrm{Path}(X)(x,y)$. Using this we see that the relation of path homotopy is an equivalence relation on the set $\mathrm{Path}(X)(x,y)$.

Definition 28.3. [defn-Pi-one]
We write $\Pi_1(X)(x,y)$ for the set of path-homotopy classes of paths from x to y in X. Equivalently, we have

$$\Pi_1(X)(x,y) = \pi_0(\mathrm{Path}(X)(x,y)).$$

We also write $\pi_1(X,x)$ for $\Pi_1(X)(x,x)$. If X has a specified basepoint then we just write $\pi_1(X)$ for $\pi_1(X, 0_X)$.

Remark 28.4. [rem-pi-one-omega]
We can identify S^1 with $[0,1]/\{0,1\}$ in an obvious way, and take the collapsed point as the basepoint. For any based space X we then have

$$\Omega X = F(S^1, X) = \{u \in \mathrm{Path}(X) \ :\ u(0) = u(1) = 0_X\} = \mathrm{Path}(X)(0_X, 0_X)$$

and therefore $\pi_1(X) = [S^1, X]_* = \pi_0(\Omega X)$.

Proposition 28.5. [prop-path-homotopy]

(a) *Suppose we have paths $u_0, u_1 \in \mathrm{Path}(X)(a,b)$, and paths $v_0, v_1 \in \mathrm{Path}(X)(b,c)$. If u_0 is path-homotopic to u_1 and v_0 is path-homotopic to v_1 then the join $v_0 * u_0$ is path-homotopic to $v_1 * u_1$.*
(b) *Suppose we have paths $u \in \mathrm{Path}(X)(a,b)$ and $v \in \mathrm{Path}(X)(b,c)$ and $w \in \mathrm{Path}(X)(c,d)$. Then $w * (v * u)$ is path-homotopic to $(u * v) * w$.*
(c) *If c_a is the constant path with value x, then any path $u \in \mathrm{Path}(X)(a,b)$ is path-homotopic to $u * c_a$ and to $c_b * u$.*
(d) *If $\overline{u}$ denotes the reverse path $\overline{u}(t) = u(1-t)$, then $\overline{u} * u$ is path-homotopic to c_a, and $u * \overline{u}$ is path-homotopic to c_b.*

Proof.

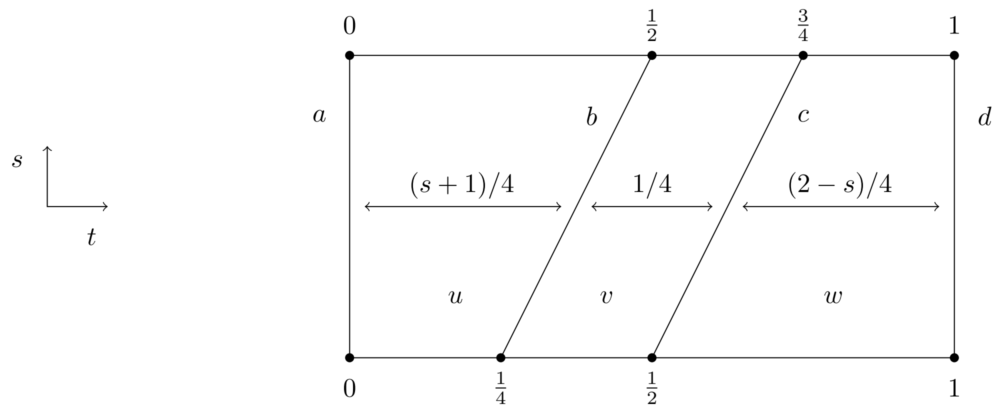
(a) Let U be a path homotopy from u_0 to u_1, and let V be a path homotopy from v_0 to v_1. Put

$$W(s,t) = \begin{cases} U(s,2t) & \text{if } 0 \le t \le 1/2 \\ V(s,2t-1) & \text{if } 1/2 \le t \le 1 \end{cases}$$

(so the corresponding paths w_s are just $w_s = v_s * u_s$). It is clear by patching (Proposition 5.9(b)) that W is continuous, and it gives the required path homotopy.

(b) Define $F\colon [0,1]^2 \to X$ by

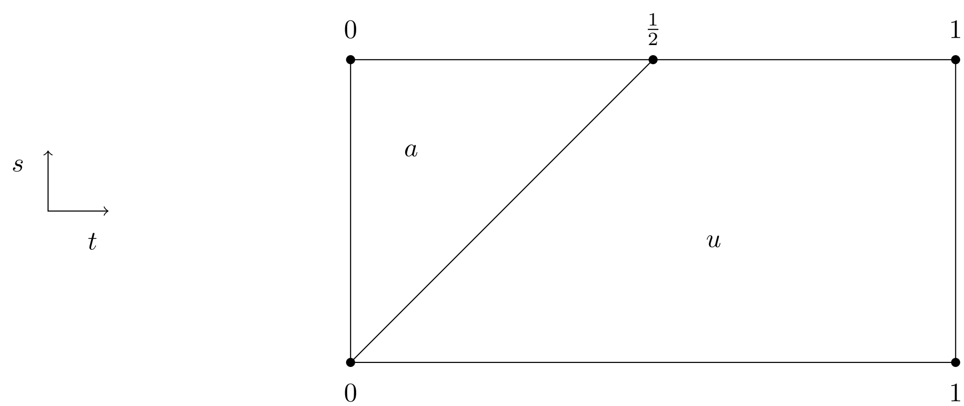
$$F(s,t) = \begin{cases} u(4t/(s+1)) & \text{if } 0 \le 4t \le s+1 \\ v(4t-s-1) & \text{if } s+1 \le 4t \le s+2 \\ w((4t-s-2)/(2-s)) & \text{if } s+2 \le 4t \le 4. \end{cases}$$



This gives the required path-homotopy. It corresponds to a family of paths $(f_s)_{s\in[0,1]}$, where f_s follows u at speed $4/(s+1)$, then follows v at speed 4, then follows w at speed $4/(2-s)$.

(c) To see that $u * c_a \simeq u$, use the path-homotopy

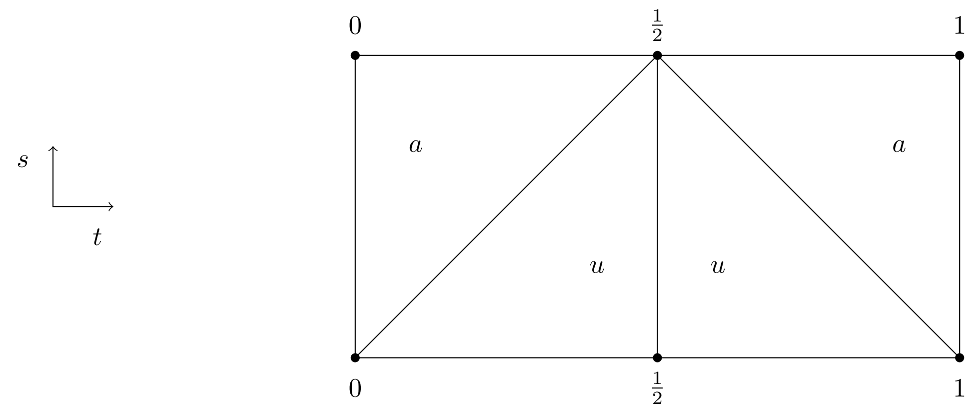
$$F(s,t) = \begin{cases} a & \text{if } 0 \le 2t \le s \\ u((2t-s)/(2-s)) & \text{if } s \le 2t \le 2. \end{cases}$$



This corresponds to a family of paths $(f_s)_{s\in[0,1]}$, where f_s waits at a for time $s/2$, and the follows u at speed $1/(1-s/2)$ for the remaining time. The proof that $c_b * u \simeq u$ is similar.

(d) To see that $\overline{u} * u \simeq c_a$, use the path-homotopy

$$F(s,t) = \begin{cases} a & \text{if } 0 \le 2t \le s \\ u(2t-s) & \text{if } s \le 2t \le 1 \\ u(2-s-2t) & \text{if } 1 \le 2t \le 2-s \\ a & \text{if } 2-s \le t \le 2. \end{cases}$$



This corresponds to a family of paths $(f_s)_{s\in[0,1]}$, where f_s waits at a for time $s/2$, and then starts to follow u at speed 2 for time $(1-s)/2$, then reverses direction and runs back to a and waits there again for a further period of length $s/2$. The proof that $u * \overline{u} \simeq c_b$ is similar.

□

Corollary 28.6. [`cor-Pi-one`]
There is a category $\Pi_1(X)$, whose objects are the points of X, and whose morphisms from x to y are the path-homotopy classes of paths from x to y. The reverse of a path gives an inverse for the morphism corresponding to that path, so every morphism is an isomorphism, or in other words the category $\Pi_1(X)$ is a groupoid. It is called the fundamental groupoid *of X.*

Remark 28.7. It is cumbersome to distinguish rigorously between paths and path-homotopy classes of paths, so we will sometimes abuse notation by blurring this distinction.

Example 28.8. [`eg-convex-groupoid`]
Let X be a convex subset of a vector space V. For any two points $x, y \in X$ we have a path u from x to y given by $u(t) = (1-t)x + ty$. If v is any other path from x to y, we have a path homotopy between u and v given by $H(s,t) = (1-s)u(t) + sv(t)$. Thus, there is a unique morphism in $\Pi_1(X)$ from x to y, so $\Pi_1(X)$ is an indiscrete category (as in Example 36.9).

Example 28.9. [`eg-circle-groupoid`]
Define a groupoid G as follows. The objects are the nonzero complex numbers, and the morphisms from z to w are the complex numbers a with $e^a = w/z$, with composition given by addition. In particular, for any z we have $G(z,z) = 2\pi i\mathbb{Z}$.

We can define a functor $I\colon \Pi_1(\mathbb{C}^\times) \to G$ as follows: on objects we have $I(z) = z$, and on morphisms we have

$$I([u]) = \int_{t=0}^{1} \frac{u'(t)}{u(t)} dt.$$

We are glossing over some technicalities here. If the path u is differentiable then the above integral expression is meaningful, and if v is another differentiable path and there is a differentiable path-homotopy between them, then $\int_0^1 u'/u = \int_0^1 v'/v$ by a standard version of Cauchy's integral theorem. There are various different ways to remove the differentiability condition, but we will not discuss them here. Once we have developed the theory of covering spaces, it will be easy to give an alternative definition of I and to prove that it is an isomorphism of categories.

Recall that $\pi_1(X,a)$ is defined to be $\Pi_1(X)(a,a)$. As $\Pi_1(X)$ is a groupoid, we see that this is a group under composition. We therefore have a one-object category $b\pi_1(X,a)$ as in Example 36.7.

Proposition 28.10. [`prop-Pi-one-connected`]
Suppose that X is path connected, and that we have a chosen basepoint $a \in X$. Then $\Pi_1(X)$ is equivalent to $b\pi_1(X,a)$.

Proof. We can identify $b\pi_1(X, a)$ with the full subcategory of $\Pi_1(X)$ whose only object is the point a. We thus have an inclusion functor $J: b\pi_1(X, a) \to \Pi_1(X)$, which is visibly full and faithful. As X is assumed to be path connected, we can choose, for each $x \in X$, a path m_x from a to x. In the case $x = a$, we will choose the constant path. This means that for each object x of $\Pi_1(X)$ we have an isomorphism from $J(a)$ to x, which means that J is also essentially surjective, and thus is an equivalence of categories. We can construct an inverse functor $R: \Pi_1(X) \to b\pi_1(X, a)$ as follows. On objects, we have no choice but to define $R(x) = a$ for all x. For any morphism $u \in \Pi_1(X)(x, y)$, we put $R(u) = m_y u m_x^{-1}$. If v is another morphism from y to z we have

$$R(v)R(u) = (m_z v m_y^{-1})(m_y u m_x^{-1}) = m_z(vu)m_x^{-1} = R(vu),$$

and similarly R preserves identity morphisms, so it is a functor. We have $RJ = 1$, and the morphisms $m_x: a = JR(x) \to x$ give a natural isomorphism $JR \to 1$. □

Definition 28.11. [defn-simply-connected]
We say that X is *simply connected* if it is nonempty and has the following equivalent properties:
(a) For all $x, y \in X$ we have $|\Pi_1(X)(x, y)| = 1$.
(b) $\Pi_1(X)$ is an indiscrete category.
(c) X is path connected, and $\pi_1(X, x) = \{1\}$ for all $x \in X$.
(d) X is path connected, and $\pi_1(X, x) = \{1\}$ for some $x \in X$.

Proof of equivalence. Conditions (b) is included just as a reminder of terminology; it means the same as (a). It is also clear by definition that X is path connected iff $|\Pi_1(X)(x, y)| > 0$ for all x and y. Suppose this holds, and that we have points $x, y, x', y' \in X$. We can then choose paths $p \in \Pi_1(X)(x, x')$ and $q \in \Pi_1(X)(y, y')$, and define maps

$$\alpha: \Pi_1(X)(x, y) \to \Pi_1(X)(x', y') \qquad \alpha(u) = qup^{-1}$$
$$\beta: \Pi_1(X)(x', y') \to \Pi_1(X)(x, y) \qquad \beta(v) = q^{-1}up.$$

It is clear that these are inverse to each other, so $|\Pi_1(X)(x, y)| = |\Pi_1(X)(x', y')|$. Given this, it is clear that (a), (c) and (d) are equivalent. □

Proposition 28.12. [prop-Pi-one-functor]
Any continuous map $f: X \to Y$ gives rise to a functor

$$f_* = \Pi_1(f): \Pi_1(X) \to \Pi_1(Y),$$

given by $f_*(x) = f(x)$ on objects and $f_*([u]) = [f \circ u]$ on morphisms. Moreover:
(a) *Any homotopy from f_0 to f_1 gives rise to a natural isomorphism $(f_0)_* \simeq (f_1)_*$.*
(b) *For any maps $X \xrightarrow{f} Y \xrightarrow{g} Z$ we have $(g_* f_*) = g_* f_*$, and moreover $(1_X)_*$ is the identity functor on $\Pi_1(X)$.*

Proof. The first thing to check is that f_* is well-defined on morphisms. This works because if U is a path-homotopy from u_0 to u_1, then $f \circ U$ is a path-homotopy from $f \circ u_0$ to $f \circ u_1$. Next, whenever we have a path u from a to b, and a path v from b to c, we note that

$$(f_*(v * u))(t) = ((f_* v) * (f_* u))(t) = \begin{cases} f(u(2t)) & \text{if } 0 \leq t \leq 1/2 \\ f(v(2t-1)) & \text{if } 1/2 \leq t \leq 1. \end{cases}$$

This shows that we have a functor. Now suppose we have a homotopy F from f_0 to f_1. For each point $a \in X$, we have a path p_a from $f_0(a)$ to $f_1(a)$, given by $p_a(t) = F(t, a)$. This gives an element $\phi_a = [p_a] \in \Pi_1(Y)(f_0(a), f_1(a))$. We claim that these form a natural transformation from $(f_0)_*$ to $(f_1)_*$. Equivalently, for any path u from a to b in X, we claim that the diagram

$$\begin{tikzcd}
f_0(a) \arrow[r, "\phi_a"] \arrow[d, "(f_0)_*[u]"'] & f_1(a) \arrow[d, "(f_1)_*[u]"] \\
f_0(b) \arrow[r, "\phi_b"'] & f_1(b)
\end{tikzcd}$$

commutes in the category $\Pi_1(Y)$. Equivalently, we have paths $p_b * (f_0 \circ u)$ and $(f_1 \circ u) * p_a$ from $f_0(a)$ to $f_1(b)$, and we claim that they are path-homotopic. A suitable path-homotopy is given by

$$V(s,t) = \begin{cases} F(0, u(2t)) = f_0(u(2t)) & \text{if } 0 \leq 2t \leq s \\ F(2t - s, u(s)) & \text{if } s \leq 2t \leq s+1 \\ F(1, u(2t-1)) = f_1(u(2t-1)) & \text{if } s+1 \leq 2t \leq 2. \end{cases}$$

This proves claim (a), and (b) is clear. □

Corollary 28.13. *If X is homotopy equivalent to Y, then $\Pi_1(X)$ and $\Pi_1(Y)$ are equivalent categories. In particular, if X is contractible then $\Pi_1(X)$ is an indiscrete category, so X is simply connected.* □

Definition 28.14. [`defn-covering`]
Let $p\colon X \to Y$ be a continuous map of spaces. We say that p is a *covering map* (or that X is a *covering space* of Y) if for each point $y \in Y$ there is an open neighbourhood V of y, a discrete space F and a map $f\colon p^{-1}(V) \to F$ such that the combined map $(p, f)\colon p^{-1}(V) \to V \times F$ is a homeomorphism. We will say that such a set V is *trivially covered* by q, and that the map f is a *trivialisation*.

Remark 28.15. [`rem-coverings-open`]
It is easy to see that covering maps are open maps, and thus that surjective covering maps are quotient maps.

Example 28.16. [`eg-product-covering`]
If F is discrete and Y is arbitrary, then the projection $p\colon Y \times F \to Y$ is clearly a covering. (For any y we can take $V = Y$ and $f(y,t) = t$.) More generally, if we have a family of spaces $(Y_i)_{i \in I}$ and a family of discrete spaces $(F_i)_{i \in I}$ then the evident map

$$p\colon \coprod_i (Y_i \times F_i) \to \coprod_i Y_i$$

is a covering.

Proposition 28.17. [`prop-exp-covering`]
The exponential map $\exp\colon \mathbb{C} \to \mathbb{C}^\times$ *is a covering.*

Proof. Put

$$\begin{aligned} T_1 &= \{x + iy \in \mathbb{C} \;:\; -\pi < y < \pi\} \\ U_1 &= \mathbb{C} \setminus \{x + (2n+1)\pi i \;:\; x \in \mathbb{R},\; n \in \mathbb{Z}\} = T_1 + 2\pi i \mathbb{Z} \\ V_1 &= \mathbb{C} \setminus (-\infty, 0]. \end{aligned}$$

For each $x \in U_1$ there is a unique integer $n = f_1(x)$ such that $(2n-1)\pi < \mathrm{Im}(x) < (2n+1)\pi$, and this function $f_1\colon U_1 \to \mathbb{Z}$ is continuous. We will take for granted the following facts from complex analysis:

(a) $\exp(x + y) = \exp(x)\exp(y)$.
(b) The map $\exp\colon \mathbb{C} \to \mathbb{C}^\times$ is surjective. (If u is any differentiable path from 1 to y avoiding 0 and $x = \int_0^1 u'(t)/u(t)dt$ then $\exp(x) = y$.)
(c) exp gives a homeomorphism $\mathbb{R} \to (0, \infty)$. (We have $\exp(x) > 1$ for $x > 0$ by inspection, so exp is strictly increasing on $\mathbb{R}$ by (a), and $\exp(\mathbb{R}) = (0, \infty)$ by the argument in (b), and this is almost enough.)
(d) We have $\exp(x) = 1$ iff x is an integer multiple of $2\pi i$. (Point (a) shows that $\exp^{-1}\{1\}$ is an additive subgroup of $\mathbb{C}$, and using complex conjugation together with (c) we see that it is contained in $i\mathbb{R}$. After studying the behaviour of $\exp(x)$ near $x = 0$ we conclude that $\exp^{-1}\{1\}$ must be discrete, so it is the set of integer multiples of $2\pi i$ for some number π, which we find by numerical calculation to be approximately 3.14. It is best to take this as the primary definition of π, and to deduce facts about the area of circles and so on as consequences.)
(e) $\exp(i\pi) = -1$. (It follows from (a) and (d) that $\exp(i\pi)^2 = 1$, and from (d) that $\exp(i\pi) \neq 1$.)

From these facts we see that

$$\exp^{-1}(-\infty,0] = \{(x+(2n+1)\pi i \ : \ x\in\mathbb{R},\ n\in\mathbb{Z}\}$$

and thus that $\exp^{-1}(V_1)=U_1$. We also see that $\exp\colon T_1\to V_1$ is a homeomorphism, and thus that the map $(\exp,f_1)\colon U_1\to V_1\times\mathbb{Z}$ is a homeomorphism. This provides the required data for the case $y=1$. For a general element $y\in\mathbb{C}^\times$, we can choose x_0 with $\exp(x_0)=y$ and then put $U_y=y.U_1$ and $V_y=\exp^{-1}(U_y)=x_0+V_1$ and $f_y(x)=f_1(x-x_0)$. This gives a homeomorphism $(\exp,f_y)\colon U_y\to V_y\times\mathbb{Z}$, as required. □

Proposition 28.18. [`prop-group-covering`]
Let X be a space with an action of a group G. Suppose that for all $x\in X$ there is an open neighbourhood U_1 of x such that $gU_1\cap U_1=\emptyset$ for all $g\in G\setminus\{1\}$. Then the quotient map $q\colon X\to X/G$ is a covering.

Proof. Let y be a point in X/G. We can then find $x\in X$ with $q(x)=y$, and U_1 as in the statement of the proposition. Put $V=q(U_1)\subseteq X/G$. As in Lemma 5.72 we see that gU_1 is open for all G and that $q^{-1}(V)=\bigcup_g gU_1$ and that V is open. Moreover, when $g\neq h$ we also have $g^{-1}h\neq 1$ so $U_1\cap g^{-1}hU_1=\emptyset$ so $gU_1\cap hU_1=g(U_1\cap g^{-1}hU_1)=\emptyset$. This means that $q^{-1}(V)$ is the disjoint union of the open sets gU_1, so we can define a continuous map $f\colon q^{-1}(V)\to G$ by $f(x)=g$ for all $x\in gU_1$. It is then clear that the map $(q,f)\colon q^{-1}(V)\to V\times G$ is a continuous bijection and an open map, so it is a homeomorphism. □

Corollary 28.19. [`cor-group-covering`]
Let X be a regular space with a free action of a finite group G; then the projection $q\colon X\to X/G$ is a covering map. (In particular, this works for metric spaces by Proposition 14.8, and for locally compact Hausdorff spaces by Proposition 18.4.)

Proof. Consider a point $x\in X$. Put $W=X\setminus\{gx \ : \ g\neq 1\}$. This is open, and as the action is free we see that $x\in W$. As X is locally compact Hausdorff, we can find a neighbourhood V of X such that $\overline{V}\subseteq W$. Note that if $g\neq 1$ then $g^{-1}\neq 1$ so $g^{-1}x\notin\overline{V}$ so $x\notin g\overline{V}$. It follows that the set $U=V\setminus\bigcup_{g\neq 1}\overline{V}$ is an open neighbourhood of U, and by construction we have $U\cap gU=\emptyset$ for $g\neq 1$. The claim therefore follows by Proposition 28.18. □

Example 28.20. [`eg-torus-covering`]
We can let $\mathbb{Z}^n$ act on $\mathbb{R}^n$ by translation. For any $x\in\mathbb{R}^n$ we can let U_1 be the open ball of radius $1/2$ centred at x, and we find that he condition in Proposition 28.18 is satisfied. It follows that the quotient map $\mathbb{R}^n\to\mathbb{R}^n/\mathbb{Z}^n$ is a covering map. Here $\mathbb{R}^n/\mathbb{Z}^n$ is homeomorphic to $(\mathbb{R}/\mathbb{Z})^n$ and thus to $(S^1)^n$, otherwise known as an n-dimensional torus.

Example 28.21. [`eg-lens-covering`]
In Example 5.70 we defined

$$L(p_1,\dots,p_n)=S^{2n-1}/C_d,$$

where the group $C_d<\mathbb{C}^\times$ acts on $S^{2n-1}\subset\mathbb{C}^n$ by the rule

$$z.(x_1,\dots,x_n)=(z^{p_1}x_1,\dots,z^{p_n}x_n).$$

If the integers p_i are all coprime to d then the action is free and so the quotient map

$$q\colon S^{2n-1}\to L(p_1,\dots,p_n)$$

is a covering map.

Example 28.22. [`eg-elliptic-modular`]
Let $U=\{z\in\mathbb{C} \ : \ \mathrm{Im}(z)>0\}$ be the upper half-plane in $\mathbb{C}$. The group $SL_2(\mathbb{R})$ acts on U by the rule

$$\left[\begin{smallmatrix} a & b\\ c & d\end{smallmatrix}\right].z=\frac{az+b}{cz+d}.$$

Note that the subgroup $Z=\{I,-I\}$ acts as the identity, so there is an induced action of the quotient group $PSL_2(\mathbb{R})=SL_2(\mathbb{R})/Z$. We will be interested in the subgroup

$$G=\left\{\left[\begin{smallmatrix} a & b\\ c & d\end{smallmatrix}\right]\in SL_2(\mathbb{Z}) \ : \ \left[\begin{smallmatrix} a & b\\ c & d\end{smallmatrix}\right]=\left[\begin{smallmatrix} 1 & 0\\ 0 & 1\end{smallmatrix}\right] \pmod 2\right\}<SL_2(\mathbb{R})$$

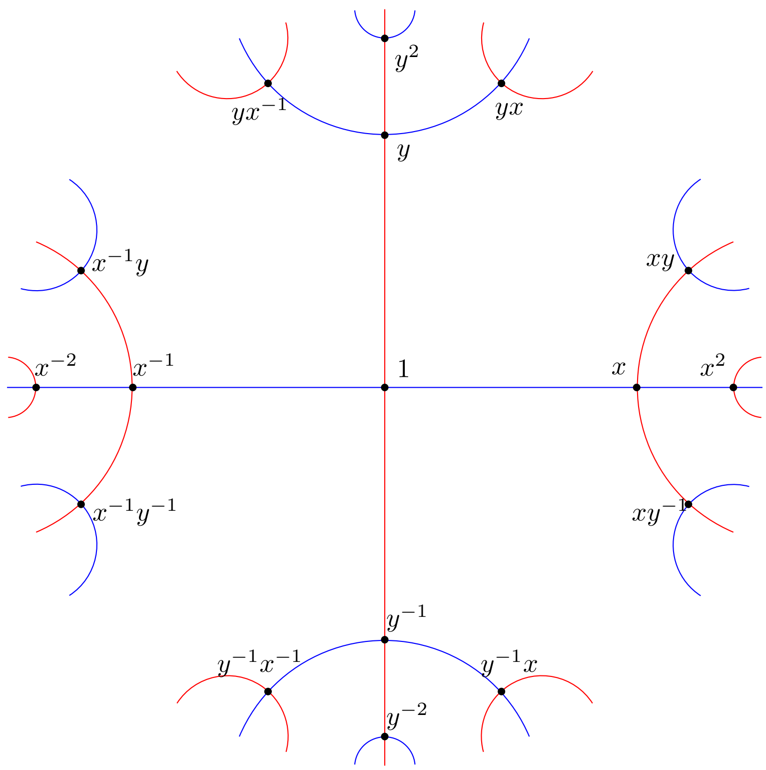
and $\overline{G} = G/Z < PSL_2(\mathbb{R})$. There is then a holomorphic covering map $\lambda\colon U \to \mathbb{C} \setminus \{0,1\}$ called the *elliptic modular function*. This satisfies $\lambda(g.z) = \lambda(z)$ for all $g \in G$, so there is an induced map $\overline{\lambda}\colon U/G \to \mathbb{C} \setminus \{0,1\}$, which is actually a homeomorphism. However, it would take us too far afield to prove these facts. **Find a good reference.**

Example 28.23. [`eg-tree`]
Let G be the free group generated by elements x and y. This means that each element $g \in G$ is a list $(a_1, \dots, a_r)$ where each a_i is one of the symbols x, y, x^{-1} and y^{-1}, and x never occurs next to x^{-1}, and y never occurs next to y^{-1}. The empty list is allowed, and gives the identity element of G. The product of $(a_1, \dots, a_r)$ and $(b_1 \dots, b_s)$ is defined by taking the concatenated list $(a_1, \dots, a_r, b_1, \dots, b_s)$ and repeatedly removing adjacent pairs of the form (x, x^{-1}), (x^{-1}, x), (y, y^{-1}) or (y^{-1}, y) until none are left.

Now consider the space

$$X_0 = G \amalg (G \times \{x, y\} \times [0,1])$$

and form the quotient space X where $(g, u, 0) \in (G \times \{x, y\} \times [0, 1])$ is identified with $g \in G$, and $(g, u, 1)$ is identified with $gu \in G$. This can be pictured as a graph, with one vertex for each element of G, and edges linking g to gx and gy. Part of X can be displayed as follows:

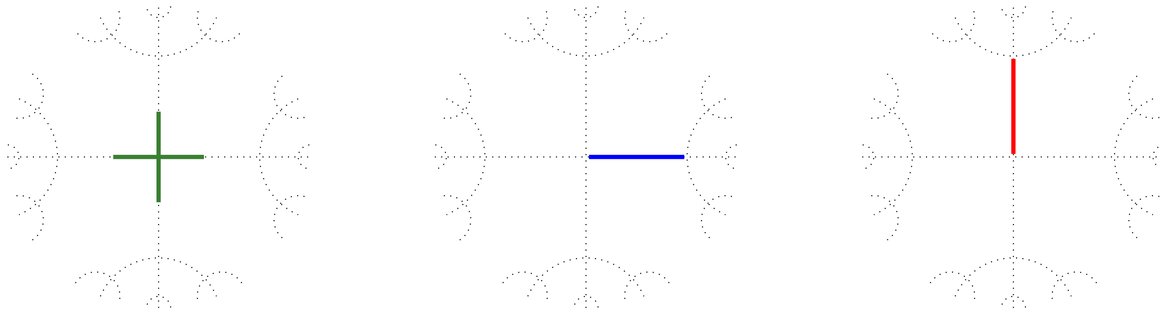


The group G acts on X by $g.[h, u, t] = [gh, u, t]$. The orbit space X/G is the quotient of $\{x, y\} \times [0, 1]$ where $(x, 0)$, $(x, 1)$, $(y, 0)$ and $(y, 1)$ are all identified together. This is homeomorphic to the figure-eight space

$$Y = \{z \in \mathbb{C} \;:\; |z - 1| = 1 \text{ or } |z + 1| = 1\}$$

by the map sending $[x, t]$ to $1 - e^{2\pi i t}$ and $[y, t]$ to $e^{2\pi i t} - 1$. Now put

$$\begin{aligned}
U_0 =& \{[1, x, t] \;:\; 0 \le t < 1/2\} \cup \{[1, y, t] \;:\; 0 \le t < 1/2\} \cup \\
& \{[x^{-1}, x, t] \;:\; 1/2 < t \le 1\} \cup \{[y^{-1}, y, t] \;:\; 1/2 < t \le 1\} \\
U_1 =& \{[1, x, t] \;:\; 0 < t < 1\} \\
U_2 =& \{[1, y, t] \;:\; 0 < t < 1\}.
\end{aligned}$$



U_0 $\qquad$ U_1 $\qquad$ U_2

One can see that $gU_i \cap hU_i = \emptyset$ whenever $g \neq h$, and that every point in X lies in some set gU_i. It follows that the criterion in Proposition 28.18 is satisfied, and so the quotient map $q\colon X \to X/G$ is a covering map. We could do essentially the same thing with a free group on r generators for any r; then the space X would be a graph in which every vertex lies on $2r$ different edges, and X/G would be a union of r circles, all meeting in a single point.

Proposition 28.24. [`prop-lifting`]
Let $q\colon X \to Y$ be a covering map. Let $v\colon [0,1] \to Y$ be a path, and let $a \in X$ be a point such that $q(a) = v(0)$. Then there is a unique path $u\colon [0,1] \to X$ with $u(0) = a$ and $q \circ u = v$.

The path u is called the *lifting* of v with initial point a.

Proof. As q is a covering map, Y is a union of open sets that are trivially covered by q, and so $[0,1]$ is the union of the preimages under v of such sets. This gives an open covering of the compact metric space $[0,1]$, and any such covering has a Lebesgue number $\epsilon > 0$ by Theorem 12.28. Choose n such that $1/n < \epsilon$. Then each the sets $v([i/n,(i+1)/n])$ must lie in some open set $V_i \subseteq Y$ that is trivially covered by q, so we can choose a discrete set F_i and a continuous map $f_i\colon q^{-1}(V_i) \to F_i$ such that the map $(q, f_i)\colon q^{-1}(V_i) \to V_i \times F_i$ is a homeomorphism. Define points $a_i \in q^{-1}(V_i)$ and $b_i \in F_i$ and maps $u_i\colon [i/n,(i+1)/n] \to X$ recursively by

$$\begin{aligned} a_0 &= a \\ b_i &= f_i(a_i) \\ u_i(t) &= (q, f_i)^{-1}(v(t), b_i) \\ a_{i+1} &= u_i((i+1)/n). \end{aligned}$$

Note that $q(a_{i+1}) = v((i+1)/n) \in V_i \cap V_{i+1}$ as required. By construction we have $u_i((i+1)/n) = a_{i+1} = u_{i+1}((i+1)/n)$, so there is a unique map $u\colon [0,1] \to X$ that agrees with u_i on $[i/n,(i+1)/n]$ for all i, and this is continuous with $q \circ u = v$.

Now suppose we have another path $u'\colon [0,1] \to X$ with $u'(0) = a$ and $q \circ u' = v$. Put $a'_i = u'(i/n) \in q^{-1}(V_i)$ and $b'_i = f_i(a'_i) \in F_i$. Note that $f_i \circ u'$ gives a continuous map from the interval $[i/n,(i+1)/n]$ to the discrete space F_i, so it is necessarily constant, with value b'_i say. It follows that $u'(t) = (q, f_i)^{-1}(v(t), b'_i)$ for $t \in [i/n,(i+1)/n]$. By assumption we have $u'(0) = a_0 = u(0)$ and so $b'_0 = f_0(a_0) = b_0$, so $u' = u$ on $[0,1/n]$, so $a'_1 = a_1$, so $b'_1 = b_1$. Extending this by induction in the obvious way, we find that $u' = u$. $\square$

Proposition 28.25. [`prop-covering-homotopy`]
Let $q\colon X \to Y$ be a covering map. Let $K\colon [0,1]^2 \to Y$ be a path-homotopy from v_0 to v_1. Let $a \in X$ be a point with $q(a) = v_0(0) = v_1(0)$, and let u_0 and u_1 be the lifts of v_0 and v_1 with initial point a. Then $u_0(1) = u_1(1)$ and there is a unique path-homotopy $H\colon [0,1]^2 \to X$ from u_0 to u_1 with $q \circ H = K$.

Proof. Put $\overline{a} = v_0(0) = v_1(0) = q(a) \in Y$. As K is a path-homotopy, we actually have $v_s(0) = \overline{a}$ for all s. There is thus a unique path $u_s\colon [0,1] \to X$ with $u_s(0) = a$ and $q \circ u_s = v_s$. We put $H(s,t) = u_s(t)$, so $H(s,0) = a$ and $q \circ H = K$. From this definition it is clear that H is continuous on each vertical line $\{s\} \times [0,1]$, but not that it is continuous as a function on all of $[0,1]^2$. To prove this, we need to subdivide $[0,1]^2$.

Choose an integer $n > 0$ and put

$$Q_{ij} = [i/n, (i+1)/n] \times [j/n, (j+1)/n] \subseteq [0,1]^2$$

for $i, j, \in \{0, 1, \dots, n-1\}$. As in the previous proof, if we take n large enough then for each i and j we can choose an open set $V_{ij} \subseteq Y$ containing $K(Q_{ij})$ such that V_{ij} is trivially covered by q, and then we can choose a trivialisation $f_{ij} \colon q^{-1}(V_{ij}) \to F_{ij}$. By the standard patching argument, it will suffice to prove that $H|_{Q_{ij}}$ is continuous for all i and j, which we will do by induction on j. As (q, f_{ij}) is a homeomorphism and the map $q \circ H = K$ is continuous by assumption, it will suffice to prove that the map

$$g_{ij} = f_{ij} \circ H|_{Q_{ij}} \colon Q_{ij} \to F_{ij}$$

is constant on Q_{ij}. As H is continuous on vertical lines, we see that g_{ij} is constant on vertical lines in Q_{ij}. It will thus be enough to show that it is constant on the horizontal line $[i/n, (i+1)/n] \times \{j/n\}$. If $j = 0$ this is clear because $H(s, 0) = a$, and if $j > 0$ then it follows from the inductive assumption that $H_{Q_{i,j-1}}$ is continuous. Thus, the map $H \colon [0,1]^2 \to X$ is continuous as required.

Now let $\overline{b}$ be the common endpoint of all the paths v_s, and put $b = H(0,1) \in X$, so $q(b) = \overline{b}$. We can now define a path $w \colon [0,1] \to X$ by $w(s) = H(s,1)$, and we have $q(w(s)) = \overline{b}$ for all s. Thus w and the constant path c_b are two lifts of the constant path $c_{\overline{b}}$ with the same initial point, so they must be the same. This means that $H(s,1) = b$ for all s, so H is a path homotopy as claimed.

If H' is any other path homotopy with $H'(s,0) = a$ and $q \circ H' = K$ then the path $u'_s(t) = H'(s,t)$ is a lift of v_s starting with a, so it must be the same as u_s; so $H' = H$. □

Corollary 28.26. [`cor-covering-functor`]

*Let $q \colon X \to Y$ be a covering map, and put $F_q(y) = q^{-1}\{y\}$ for all $y \in Y$. Then for each path v from y_0 to y_1 in Y there is a function $v_* \colon F_q(y_0) \to F_q(y_1)$ defined as follows: for $x_0 \in F_q(y_0)$ we let u be the unique lift of v such that $u(0) = x_0$, then we put $v_*(x_0) = u(1)$. Moreover, this construction gives a functor $F_q \colon \Pi_1(Y) \to$ **Sets**.*

Proof. It is clear from Proposition 28.24 that we have a well-defined function $F_q(y_0) \to F_q(y_1)$ for any path v, and Proposition 28.25 tells us that this depends only on the path-homotopy class of v. If v is constant then the lift u is also constant, and it follows that $(c_y)_* = 1_{F_q(y)}$. Now suppose we have a second path v' from y_1 to y_2. Write $x_1 = v_*(x_0)$ and $x_2 = v'_*(x_1)$. By definition, this means that there is a path u from x_0 to x_1 with $q \circ u = v$, and there is a path u' from x_1 to x_2 with $q \circ u' = v'$. Thus, $u' * u$ is a path from x_0 to x_2 with $q \circ (u' * u) = v' * v$, so $(v' * v)_*(x_0) = x_2$. This proves that $(v' * v)_* = v'_* \circ v_*$, we have a well-defined functor as claimed. □

Remark 28.27. [`rem-Pi-one-connected`]

Suppose that Y is path connected, with a chosen basepoint $a \in Y$. For any functor $F \colon \Pi_1(Y) \to \mathbf{Sets}$, we have a set $F(a)$ with an action of the group $\pi_1(Y, a) = \Pi_1(Y)(a, a)$. Moreover, using Proposition 28.10 we see that every set with an action of $\pi_1(Y, a)$ comes from an essentially unique functor.

Remark 28.28. [`rem-trivialised-path`]

Suppose we have an open set $V \subseteq X$ and a trivialisation $(q, g) \colon q^{-1}(V) \to V \times G$, so g gives a bijection $F_q(y) \to G$ for each $y \in V$. Now suppose we have a path $v \colon [0,1] \to V$ with $v(0) = y_0$ and $v(1) = y_1$. We claim that the diagram

$$\begin{tikzcd} F_q(y_0) \arrow[r, "v_*"] \arrow[d, "g"', "\simeq"] & F_q(y_1) \arrow[d, "\simeq"', "g"] \\ G \arrow[r, equal] & G \end{tikzcd}$$

commutes. Indeed, given any point $x_0 \in F_q(y_0)$ we can put $a = g(x_0) \in G$ and $u(t) = (q, g)^{-1}(v(t), a)$ and $x_1 = u(1)$. This gives a path $u \colon [0,1] \to X$ with $q \circ u = v$ and $u(0) = x_0$ and $u(1) = x_1$, so $v_*(x_0) = x_1$. It is also built in to the definition of u that $g(u(t)) = a$ for all t, so in particular $g(x_0) = g(x_1)$ as required.

Definition 28.29. [`defn-locally-simply-connected`]

We say that a space Y is *locally simply connected* if the simply connected open subsets form a basis for the topology.

Example 28.30. [eg-locally-simply-connected]

As open balls in $\mathbb{R}^n$ are simply connected, we see that all open sets in $\mathbb{R}^n$ are locally simply connected. More generally, an *n-dimensional topological manifold* is a Hausdorff space X where every point has a neighbourhood homeomorphic to $\mathbb{R}^n$. (One further condition is often imposed but we can ignore it here.) Any topological manifold is locally simply connected, and this covers a large class of interesting spaces. Next, say that $X \subseteq \mathbb{R}^n$ is *locally star-shaped* if for all $x \in X$ there exists $\epsilon_x > 0$ such that $OB_{\epsilon_x}(x) \cap X$ is star-shaped around x (as in Example 27.13). If so, the sets $OB_\delta(x) \cap X$ (for $\delta < \epsilon_x$) form a basis for the topology consisting of open sets that are contractible and therefore simply connected. This covers another large class of naturally occurring spaces.

Proposition 28.31. [prop-classify-coverings]

Let $q\colon X \to Y$ and $q'\colon X' \to Y$ be covering maps, and suppose that Y is locally simply connected. Then the natural transformations $F_q \to F_{q'}$ are the same as the continuous maps $f\colon X \to X'$ with $q'f = q$.

Proof. Let $f\colon X \to X'$ be continuous with $q'f = q$. For $x \in F_q(y)$ we then have $q'(f(x)) = q(x) = y$, so $f(x) \in F_{q'}(y)$. Thus, f gives a system of maps $f_y\colon F_q(y) \to F_{q'}(y)$, and we claim that these give a natural transformation. In other words, we claim that for every path v from y_0 to y_1 in Y, the following diagram commutes:

$$\begin{tikzcd}
F_q(y_0) \arrow[r, "v_*"] \arrow[d, "f_{y_0}"'] & F_q(y_1) \arrow[d, "f_{y_1}"] \\
F_{q'}(y_0) \arrow[r, "v_*"'] & F_{q'}(y_1).
\end{tikzcd}$$

To see this, consider a point $x_0 \in F_q(y_0)$, so $x_0 \in X$ and $q(x_0) = y_0$. There is then a unique path $u\colon [0,1] \to X$ with $u(0) = x_0$ and $q \circ u = v$. Put $x_1 = u(1)$, so by definition we have $v_*(x_0) = x_1$. Now $f_{y_0}(x_0)$ is just $f(x_0)$, and $v_*(f_{y_0}(x_0))$ is the endpoint of the unique path $u'\colon [0,1] \to X'$ with $u'(0) = f(x_0)$ and $q' \circ u' = v$. One can check that $f \circ u$ has the defining properties of u', so

$$v_*(f_{y_0}(x_0)) = u'(1) = f(u(1)) = f(x_1) = f_{y_1}(v_*(x_0))$$

as required.

Conversely, suppose we have a natural map $F_q \to F_{q'}$, given by a family of maps $f_y\colon F_q(y) \to F_{q'}(y)$ for all $y \in Y$. As X is the disjoint union of the sets $q^{-1}\{y\} = F_q(y)$ (and similarly for X') we see that the maps f_y fit together to give a unique map $f\colon X \to X'$ with $q' \circ f = q$. We claim that this map f is continuous. To see this, consider a point $b \in Y$. For a sufficiently small open neighbourhood V of b, we can choose trivialisations $(q,g)\colon q^{-1}(V) \xrightarrow{\simeq} V \times G$ and $(q',g')\colon (q')^{-1}(V) \xrightarrow{\simeq} V \times G'$. After replacing V by a smaller open set if necessary, we may assume that V is simply connected. Let $\phi\colon G \to G'$ denote the composite

$$G \xrightarrow{g^{-1}} F_q(b) \xrightarrow{f_b} F_{q'}(b) \xrightarrow{g'} G'.$$

For any $y \in V$, let v_y denote the unique morphism $b \to y$ in $\Pi_1(V)$. Consider the following diagram:

$$\begin{tikzcd}
G \arrow[r, "g^{-1}", "\simeq"'] \arrow[d, equal] & F_q(b) \arrow[r, "f_b"] \arrow[d, "(v_y)_*"'] & F_{q'}(b) \arrow[r, "g'", "\simeq"'] \arrow[d, "(v_y)_*"] & G' \arrow[d, equal] \\
G \arrow[r, "\simeq", "g^{-1}"'] & F_q(y) \arrow[r, "f_y"'] & F_{q'}(y) \arrow[r, "\simeq", "g'"'] & G'.
\end{tikzcd}$$

The outer squares commute by Remark 28.28, and the middle one commutes because the maps f_y are assumed to be natural. The composite along the top is by definition ϕ, so we see that $g' \circ f_y = \phi \circ g$. This means that the following square commutes:

$$\begin{tikzcd}
q^{-1}(V) \arrow[r, "f"] \arrow[d, "(q,g)"'] & (q')^{-1}(V) \arrow[d, "(q',g')"] \\
V \times G \arrow[r, "1\times\phi"'] & V \times G'.
\end{tikzcd}$$

The vertical maps are homeomorphisms, and the bottom map is visibly continuous, so $f\colon q^{-1}(V) \to (q')^{-1}(V)$ is continuous. As X can be covered by open sets of the form $q^{-1}(V)$ arising here, we see that $f\colon X \to X'$ is continuous as claimed. $\square$

Proposition 28.32. [`prop-construct-coverings`]

*Let Y be locally simply connected, and let $T\colon \Pi_1(Y) \to$ **Sets** be a functor. Then there is a covering space $q\colon X \to Y$ such that F_q is naturally isomorphic to T.*

Proof. Put

$$X = \{(y,a) \;:\; y \in Y \text{ and } a \in T(y)\},$$

and define $q\colon X \to Y$ by $q(y,a) = y$. The main problem is to introduce a suitable topology on X. For any open set $V \subseteq Y$, we let $j\colon V \to Y$ denote the inclusion, and we let $S(V)$ denote the inverse limit of the composite functor

$$\Pi_1(V) \xrightarrow{\Pi_1(j)} \Pi_1(X) \xrightarrow{T} \mathbf{Sets}\,.$$

More explicitly, an element $s \in S(V)$ is a family of elements $s_y \in T(y)$ for all $y \in V$, such that $v_*(s_{y_0}) = s_{y_1}$ for every path v from y_0 to y_1 in V. For any such s, we put

$$U(V,s) = \{(y, s_y) \;:\; y \in V\} \subseteq X.$$

We will show that these sets form a basis for a topology on X, with respect to which q is a covering map.

Consider a point $y \in Y$, and a simply connected neighbourhood V of y. For any two points $z_0, z_1 \in V$ there is a unique morphism $v_{z_0,z_1}\colon z_0 \to z_1$ in $\Pi_1(V)$. Define

$$f\colon q^{-1}(V) = \{(z,a) \;:\; z \in V,\ a \in T(z)\} \to T(y)$$

by $f(z,a) = (v_{zy})_*(a) = (v_{yz})_*^{-1}(a)$. It is then clear that $(q,f)\colon q^{-1}(V) \to V \times T(y)$ is a bijection, with inverse $(z,b) \mapsto (z, (v_{yz})_*(b))$. For any $a \in T(y)$ we can thus define an element $s^a \in S(V)$ by $s^a_z = (v_{yz})_*(a)$. In particular, we have $s^a_y = a$, so $(y,a) \in U(V,s^a)$. More generally, we see that $q^{-1}(V) = \coprod_{a \in T(y)} U(V,s^a)$. This shows that the sets $U(V,s)$ cover X.

Now suppose we have sets $U(V,s)$ and $U(V',s')$; we need to understand $U(V,s) \cap U(V',s')$. Put

$$W = \{y \in V \cap V' \;:\; s(y) = s'(y) \in T(y)\}.$$

We claim that this is open. Indeed, if $y \in W$ then we can choose a simply connected open set W' with $y \in W' \subseteq W$. Then for any $z \in W'$ there is a path v from y to z in W', so by the assumed property of s and s' we have $s_z = v_*(s_y)$ and $s'_z = v_*(s'_y)$. As $y \in W$ we have $s_y = s'_y$, so $s_z = s'_z$. This shows that $W' \subseteq W$ as required. Now put $t = s|_W = s'|_W$; it is clear that $U(V,s) \cap U(V',s') = U(W,t)$. We thus have a topological basis, and thus a topology on X. When V is simply connected the previous paragraph gives $q^{-1}(V) = \bigcup_a U(V,s^a)$, so $q^{-1}(V)$ is open. As the simply connected open sets form a basis for the topology on Y, we deduce that $q\colon X \to Y$ is continuous. We also have $f^{-1}\{a\} = U(V,s^a)$, so the map $f\colon q^{-1}(V) \to T(y)$ is continuous, as is the combined map $(q,f)\colon q^{-1}(V) \to V \times T(y)$. We claim that (q,f) is also an open map. To see this, consider an open set $A \subseteq q^{-1}(V)$ and a point $(z,b) \in A$. Put $a = f(z,b) = (v_{zy})_*(b) \in T(y)$. As A is open, there exists an open subset $V' \subseteq V$ and an element $s' \in S(V')$ such that $(z,b) \in U(V',s') \subseteq A$. After shrinking V' if necessary, we may assume that it is simply connected. Note that s' agrees with s^a at z, and every point in V' can be connected to z by a path in V', so $s' = s^a|_{V'}$. It follows that

$$(q,f)(A) \supseteq (q,f)(U(V',s')) = V' \times \{a\},$$

which is a neighbourhood of $(q,f)(z,b)$. This proves that (q,f) is open as well as continuous and bijective, so it is a homeomorphism. This means that q is a covering as claimed.

Now consider an arbitrary path $v\colon [0,1] \to Y$. As T is a functor we have an induced map $v_*\colon T(v(0)) \to T(v(1))$. As $q\colon X \to Y$ is a covering with $q^{-1}\{y\} = T(y)$, we get another map $T(v(0)) \to T(v(1))$ defined by path lifting; we will temporarily use the notation $v_\bullet$ for this. It is clear by construction that $v_* = v_\bullet$ when $v([0,1])$ is contained in a simply connected open set. In general, we note that the sets $v^{-1}(V)$ (for $V \subseteq Y$ open and simply connected) form an open cover of $[0,1]$. Choose N large enough that $1/(N-1)$ is a Lebesgue number for this cover. Define $v_i(t) = v((i+t)/N)$ for $i \in \{0,1,\dots,N-1\}$ and $t \in [0,1]$; we then see that $v_i([0,1])$ is contained in a simply connected open set, so $(v_i)_* = (v_i)_\bullet$ for all i. Moreover, v is path-homotopic to the join $v_{N-1} * \dots * v_0$, so we also have $v_* = v_\bullet$. This means that F_q and T are the same functor. More precisely, we have $F_q(y) = \{(y,a) \,:\, a \in T(y)\}$ and the maps $\pi_y(y,a) = a$ give a natural isomorphism $\pi\colon F_q \to T$. $\square$

Remark 28.33. For the last few results we have assumed that Y is locally simply connected. It is sometimes useful to note that we can make do with a slightly weaker condition. For $V \subseteq Y$ and $y, z \in V$ put

$$\mathcal{A}(V)(y,z) = \mathrm{img}(\Pi_1(V)(y,z) \to \Pi_1(X)(y,z)).$$

Thus, every element of $\mathcal{A}(V)(y,z)$ is represented by a path from y to z in V, and two paths represent the same element iff they are path-homotopic in X. This construction gives a subcategory $\mathcal{A}(V) \subseteq \Pi_1(X)$, which is neither wide nor full in general. Thus, any functor $F\colon \Pi_1(X) \to \mathbf{Sets}$ can be restricted to give a functor $\mathcal{A}(V) \to \mathbf{Sets}$. Let us say that V is *relatively simply connected* if $\mathcal{A}(V)$ is indiscrete (so all morphism sets $\mathcal{A}(V)(x,y)$ have size one). It is easy to see that (V is simply connected) $\implies$ (V is relatively simply connected) $\implies$ (V is path connected). Moreover, if $W \subseteq V$ and V is relatively simply connected and W is path connected, then W is relatively simply connected.

We say that X is *semilocally simply connected* if the relatively simply connected open sets form a basis for the topology. One can check that Propositions 28.31 and 28.32 remain valid if we assume only that Y is semilocally simply connected. Indeed, we can just replace $\Pi_1(V)$ by $\mathcal{A}(V)$ whenever it appears (implicitly or explicitly) in the proofs.

29. Simplicial complexes

We now digress slightly to discuss a particular class of spaces called simplicial complexes, which are important in many applications.

Definition 29.1. [`defn-ASC`]

An *(abstract) simplicial complex* K consists of a set $\mathrm{vert}(K)$ (whose elements are called *vertices*) together with a set $\mathrm{simp}(K)$ of subsets of $\mathrm{vert}(K)$ (called *simplices*) such that:

- ASC0: Every simplex is a finite, nonempty set.
- ASC1: If σ is a simplex and τ is a nonempty subset of σ then τ is also a simplex.
- ASC2: If v is a vertex then $\{v\}$ is a simplex.

A *k-simplex* is a simplex σ with $|\sigma| = k+1$ (so the 0-simplices biject with the vertices). We say that K is *finite* if there are only finitely many vertices, or equivalently, only finitely many simplices. If K has simplices of arbitrarily large size, we say that it is *infinite-dimensional*; otherwise, the *dimension* of K is the largest n such that K has an n-simplex.

If K and L are simplicial complexes, a *simplicial map* from K to L is a function $f\colon \mathrm{vert}(K) \to \mathrm{vert}(L)$ such that for every simplex $\sigma \in \mathrm{simp}(K)$, the image $f(\sigma) = \{f(v) \,:\, v \in \sigma\}$ is a simplex of L. We write **ASC** for the category of abstract simplicial complexes and simplicial maps.

Definition 29.2. [`defn-restricted-complex`]

Let K be an abstract simplicial complex, and let W be a subset of $\mathrm{vert}(K)$. We write $K|_W$ for the abstract simplicial complex with $\mathrm{vert}(K|_W) = W$ and

$$\mathrm{simp}(K|_W) = \{\sigma \in \mathrm{simp}(K) \,:\, \sigma \subseteq W\}.$$

Subcomplexes of this form are called *full subcomplexes* of K.

Definition 29.3. [`defn-realisation`]

Let K be an abstract simplicial complex. The *geometric realisation* of K (written $G(K)$ or $|K|$) is the set

$$G(K) = \{x\colon \mathrm{vert}(K) \to [0,1] \,:\, \mathrm{supp}(x) \in \mathrm{simp}(K),\ \sum_{v \in \mathrm{vert}(K)} x(v) = 1\},$$

where $\mathrm{supp}(x) = \{v \,:\, x(v) > 0\}$. Note here that the condition $\mathrm{supp}(x) \in K$ implies that $\mathrm{supp}(x)$ is finite, so the expression $\sum_{v \in \mathrm{vert}(K)} x(v)$ really only involves a finite sum.

For $v \in \mathrm{vert}(K)$, we have a map $e_v\colon \mathrm{vert}(K) \to [0,1]$ given by

$$e_v(w) = \begin{cases} 1 & \text{if } w = v \\ 0 & \text{otherwise} \end{cases}$$

This is clearly a point of $G(K)$. We will often identify e_v with v.

If $|\operatorname{vert}(K)| = n < \infty$, we give $G(K)$ the obvious topology as a subspace of $\operatorname{Map}(\operatorname{vert}(K), \mathbb{R}) \simeq \mathbb{R}^n$. If K is infinite, we declare that $U \subseteq G(K)$ is open iff $U \cap G(K|_W)$ is open for all finite sets $W \subseteq \operatorname{vert}(K)$; this defines a CGWH topology on $G(K)$.

Remark 29.4. In some places we will need a metric on $G(K)$, and by default we will always use this one:

$$d(x, y) = \sum_{v \in V} |x(v) - y(v)|.$$

If K is finite then the metric topology is the obvious one. If K is infinite then the standard topology on $G(K)$ is obtained from the metric topology by applying the functor $k()$ from Definition 23.1. Note also that for all x and y we have

$$d(x, y) = \sum_{v \in V} |x(v) - y(v)| \leq \sum_{v \in V} x(v) + y(v) = 2.$$

Example 29.5. [`eg-simplex-realisation`]

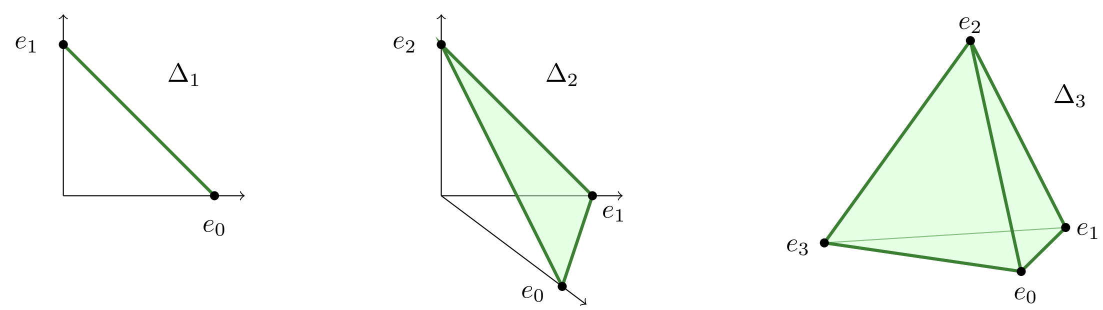
For any finite set I we can define an abstract simplicial complex Δ_I^a with vertex set I by declaring that *every* nonempty subset of V is a simplex. We write $\Delta_I = G(\Delta_I^a)$, so

$$\Delta_I = \{x \colon I \to [0, 1] \ : \ \sum_i x(i) = 1\}.$$

We will often consider the case $I = [n] = \{0, \ldots, n\}$, in which case we write Δ_n^a and Δ_n rather than $\Delta_{[n]}^a$ and $\Delta_{[n]}$. We also use the streamlined notation

$$\Delta_n = \{(x_0, \ldots, x_n) \in [0, 1]^{n+1} \ : \ \sum_i x_i = 1\}.$$

We call Δ_n^a the *abstract n-simplex*, and Δ_n the *geometric n-simplex*. Note that Δ_0 is a single point, and we can draw pictures of Δ_1, Δ_2 and Δ_3 as follows.



If K is an arbitrary simplicial complex and $\sigma = \{v_0, \ldots, v_n\}$ is an n-simplex of K then we will write

$$|\sigma| = G(K|_\sigma) = \{x \in G(K) \ : \ \operatorname{supp}(x) \subseteq \sigma\}.$$

This is clearly homeomorphic to Δ_n. Moreover, $G(K)$ is the union of the sets $|\sigma|$, and we have $|\sigma| \cap |\tau| = |\sigma \cap \tau|$ (provided we interpret $|\emptyset|$ as $\emptyset$).

Example 29.6. [`eg-simplex-boundary`]

For $n > 0$ we also define a complex $\partial \Delta_n^a$ with vertex set $[n] = \{0, \ldots, n\}$ again, but with

$$\operatorname{simp}(\partial \Delta_n^a) = \{\sigma \subseteq [n] \ : \ \sigma \neq \emptyset \text{ and } \sigma \neq [n]\}.$$

We find that

$$G(\partial \Delta_n^a) = \{x \in \Delta_n \ : \ x_i = 0 \text{ for some } i\},$$

which is the boundary of Δ_n in an evident sense; we denote it by $\partial \Delta_n$. There is a homeomorphism $f \colon \partial \Delta_n \to S^{n-1}$ given by

$$f(x_0, \ldots, x_n) = \frac{(x_1 - x_0, x_2 - x_1, \ldots, x_n - x_{n-1})}{\sqrt{\sum_i (x_{i+1} - x_i)^2}}.$$

Example 29.7. [`eg-discrete-simplicial`]
We say that K is *discrete* if the only simplices are the singletons, so $\mathrm{simp}(K) = \{\{v\} \ : \ v \in \mathrm{vert}(K)\}$. If so, then $G(K)$ is just $\mathrm{vert}(K)$ considered as a space with the discrete topology.

Definition 29.8. [`defn-affine-map`]
Let K be a simplicial complex, and let f be a map from $\mathrm{vert}(K)$ to a finite-dimensional real vector space P. We define $f_*\colon G(K) \to P$ by

$$f_*(x) = \sum_{v \in \mathrm{vert}(K)} x(v) f(v).$$

We call this the *affine extension* of f, and we say that f_* is an *affine map* from $G(K)$ to P. If f_* gives a homeomorphism from $G(K)$ to some subspace $X \subseteq P$, we say that f_* is an *affine triangulation* of X. We will write f rather than f_* in cases where this is unlikely to cause confusion.

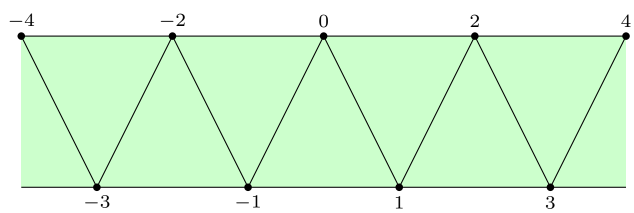
Example 29.9. [`eg-infinite-band`]
For each $r \geq 0$ we can define a simplicial complex $K(r)$ with $\mathrm{vert}(K(r)) = \mathbb{Z}$ and

$$\mathrm{simp}(K(r)) = \{\sigma \subset \mathbb{Z} \ : \ \max(\sigma) - \min(\sigma) \leq r\}.$$

When $r = 0$, the only simplices are the singletons $\{n\}$, so $G(K(0)) = \{e_n \ : \ n \in \mathbb{Z}\}$ and this is homeomorphic to the discrete space $\mathbb{Z}$.

When $r = 1$, the only simplices are the singletons and the sets $\epsilon_n = \{n, n+1\}$ for $n \in \mathbb{Z}$. If we let $f\colon \mathbb{Z} \to \mathbb{R}$ be the inclusion, we find that the affine extension $f_*\colon G(K(1)) \to \mathbb{R}$ is a homeomorphism, sending $|\epsilon_n|$ to the interval $[n, n+1]$.

Now take $r = 2$, so the 0-simplices have the form $\{n\}$, the 1-simplices have the form $\{n, n{+}1\}$ or $\{n, n{+}2\}$, and the 2-simplices have the form $\{n, n+1, n+2\}$. We can define $g\colon \mathbb{Z} \to \mathbb{R}^2$ by $g(n) = (n, (-1)^n)$ and we find that the affine extension gives a homeomorphism $G(K(2)) \to \mathbb{R} \times [-1, 1]$.



Example 29.10. [`eg-octahedron`]
Take $\mathrm{vert}(K) = \{1, 2, 3, -1, -2, -3\}$ and

$$\mathrm{simp}(K) = \{\sigma \subseteq \mathrm{vert}(K) \ : \ \sigma \neq \emptyset, \ \sigma \cap (-\sigma) = \emptyset\}.$$

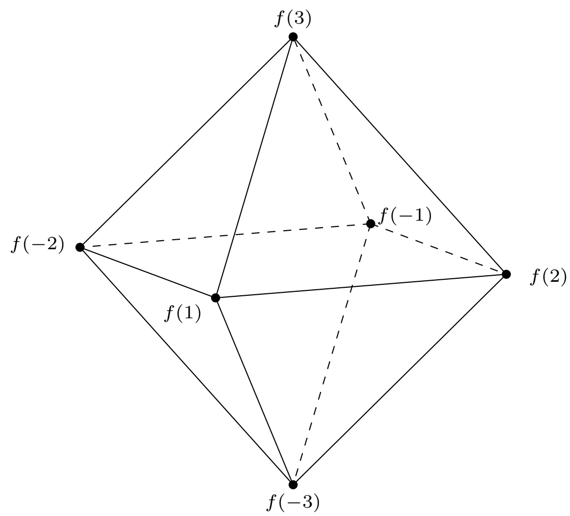
There are then eight 2-simplices, as follows:

$$\begin{array}{llll} \sigma_0 = \{1,2,3\} & \sigma_1 = \{-1,2,3\} & \sigma_2 = \{1,-2,3\} & \sigma_3 = \{-1,-2,3\} \\ \sigma_4 = \{1,2,-3\} & \sigma_5 = \{-1,2,-3\} & \sigma_6 = \{1,-2,-3\} & \sigma_7 = \{-1,-2,-3\}. \end{array}$$

We can define $f\colon \mathrm{vert}(K) \to \mathbb{R}^3$ by

$$\begin{array}{ccc} f(1) = (1,0,0) & f(2) = (0,1,0) & f(3) = (0,0,1) \\ f(-1) = (-1,0,0) & f(-2) = (0,-1,0) & f(-3) = (0,0,-1) \end{array}$$

and we find that the affine extension gives a homeomorphism from $G(K)$ to the octahedron:

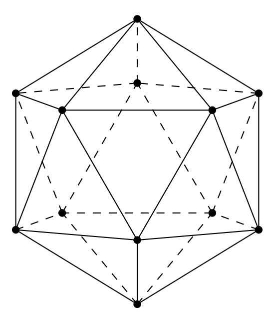


EXAMPLE 29.11. [eg-icosahedron]

Put $\tau = (1+\sqrt{5})/2$. We have four points in $\mathbb{R}^3$ of the form $(0, \pm 1, \pm\tau)$, and by adding in all cyclic permutations of these we obtain a set V containing twelve points in total. One can check that for any $x, y \in V$ we have

$$\|x - y\| \in \{0, 2, \sqrt{5}+1, \sqrt{10+2\sqrt{5}}\}.$$

We can now define a simplicial complex K with $\mathrm{vert}(K) = V$ and $\mathrm{simp}(K) = \{\sigma \subseteq V \;:\; \mathrm{diam}(\sigma) \leq 2\}$. We have a map $f\colon G(K) \to \mathbb{R}^3$ given by $f(x) = \sum_v x(v).v$, and one can check that this is an embedding, and that the image is an icosahedron.



We can give a more algebraic definition of this simplicial complex, as follows. Let r denote the five-cycle $(1\;2\;3\;4\;5)$, and let C denote the conjugacy class of r in A_5, which has $|C| = 12$. (Note here that r and r^2 are conjugate in Σ_5 but not in A_5; there are 24 five-cycles in all, of which 12 are conjugate to r, and the other 12 to r^2.) We now have a simplicial complex L with vertex set C, where a subset $\sigma \subseteq C$ is a simplex iff for all distinct $g, h \in \sigma$, the permutation $g^{-1}h$ is a three-cycle. The group A_5 acts on C by conjugation, and this action preserves simplices, so we get an induced action on L and $G(L)$. It turns out that L is isomorphic to K. We will not give a full proof, but we will record some relevant formulae. There is an embedding

$\rho\colon A_5 \to SO(3)$ with

$$\begin{aligned}
\rho((1\ 2)(3\ 4)) &= \left[\begin{smallmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{smallmatrix}\right] \\
\rho((1\ 3)(2\ 4)) &= \left[\begin{smallmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{smallmatrix}\right] \\
\rho((2\ 3)(4\ 5)) &= \frac{1}{2}\left[\begin{smallmatrix} 1/\tau & -\tau & 1 \\ -\tau & -1 & 1/\tau \\ -1 & 1/\tau & -\tau \end{smallmatrix}\right] \\
\rho((1\ 2\ 3)) &= \left[\begin{smallmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{smallmatrix}\right] \\
\rho(1\ 2\ 3\ 4\ 5) &= \frac{1}{2}\left[\begin{smallmatrix} \tau & -1 & -1/\tau \\ 1 & 1/\tau & \tau \\ -1/\tau & -\tau & 1 \end{smallmatrix}\right].
\end{aligned}$$

One can check that this action preserves the set V. For each five-cycle $c \in C$ there is a unique vector $v(c)$ such that $\|v(c)\| = \sqrt{1+\tau^2}$ and $\rho(c)$ is a clockwise rotation through $2\pi/5$ around $v(c)$. One can check that this vector lies in V, and that the map $v\colon C \to V$ is an A_5-equivariant bijection. Moreover, one can show that $v(\sigma)$ is a simplex if and only if σ is a simplex, so v is an isomorphism of simplicial complexes.

Remark 29.12. [`rem-realisation-functor`]
Let $f\colon K \to L$ be a morphism of simplicial complexes. We can then define $G(f)\colon G(K) \to G(L)$ by

$$G(f)(x)(w) = \sum_{f(v)=w} x(v).$$

Equivalently, we have a map $\mathrm{vert}(K) \to G(L) \subset \mathrm{Map}(\mathrm{vert}(L), \mathbb{R})$ given by $v \mapsto e_{f(v)}$, and $G(f)$ is the affine extension of this. One can check that this construction makes G into a functor $\mathbf{ASC} \to \mathbf{Spaces}$.

Remark 29.13. [`rem-asc-coprod`]
If K and L are abstract simplicial complexes, we can define a new complex $K \amalg L$ as follows: the vertex set is just $\mathrm{vert}(K) \amalg \mathrm{vert}(L)$, and a subset $\sigma \subseteq \mathrm{vert}(K) \amalg \mathrm{vert}(L)$ is a simplex if either

(a) $\sigma \subseteq \mathrm{vert}(K)$ and σ is a simplex of K; or
(b) $\sigma \subseteq \mathrm{vert}(L)$ and σ is a simplex of L.

We then have a bijection $\mathrm{simp}(K \amalg L) \simeq \mathrm{simp}(K) \amalg \mathrm{simp}(L)$. If we have maps $K \xrightarrow{f} M \xleftarrow{g} L$ of simplicial complexes then we can combine them in an obvious way to get a map $K \amalg L \to M$, and using this we see that $K \amalg L$ is a categorical coproduct for K and L. Given a (possibly infinite) family of complexes K_i, we can define a coproduct $\coprod_i K_i$ in the same way. It is clear that $G(\coprod_i K_i) = \coprod_i G(K_i)$.

Remark 29.14. [`rem-asc-prod`]
The category $\mathbf{ASC}$ also has products, constructed as follows. Suppose we have abstract simplicial complexes K and L, and we let α and β denote the projections

$$\mathrm{vert}(K) \xleftarrow{\alpha} \mathrm{vert}(K) \times \mathrm{vert}(L) \xrightarrow{\beta} \mathrm{vert}(L).$$

We then have an abstract simplicial complex $K \boxtimes L$ with $\mathrm{vert}(K \boxtimes L) = \mathrm{vert}(K) \times \mathrm{vert}(L)$ and

$$\begin{aligned}
\mathrm{simp}(K \boxtimes L) &= \{\rho \subseteq \mathrm{vert}(K) \times \mathrm{vert}(L) \;:\; \alpha(\rho) \in \mathrm{vert}(K),\ \beta(\rho) \in \mathrm{vert}(L)\} \\
&= \{\rho \;:\; \rho \subseteq \sigma \times \tau \text{ for some } \sigma \in \mathrm{simp}(K),\ \tau \in \mathrm{simp}(L)\}.
\end{aligned}$$

One can check that this is a categorical product for K and L. However, we will not denote it by $K \times L$ because we want to avoid confusion with the categorical product for *ordered* simplicial complexes, which will be introduced below. For that product we will have a homeomorphism $G(K \times L) = G(K) \times G(L)$, but for $K \boxtimes L$ we have only a weaker statement which we now discuss.

Proposition 29.15. [`prop-asc-product`]
There are natural maps

$$G(K) \times G(L) \xrightarrow{\mu} G(K \boxtimes L) \xrightarrow{\nu} G(K) \times G(L)$$

with $\nu\mu = 1$. There is also a natural homotopy from $\mu\nu$ to the identity, so μ and ν are homotopy equivalences.

Proof. The definitions are $\mu(x,y)(u,v) = x(u)y(v)$, and $\nu(z) = (G(\alpha)(z), G(\beta)(z))$, where α and β are the projections from $K \boxtimes L$ to K and L. More explicitly, we have $\nu(z) = (m,n)$, where $m(u) = \sum_v z(u,v)$ and $n(v) = \sum_u z(u,v)$. As $\sum_u x(u) = \sum_v y(v) = 1$ we see that $\nu\mu = 1$. Now consider a point $z \in G(K \boxtimes L)$, so $\operatorname{supp}(z) \subseteq \sigma \times \tau$ for some $\sigma \in \operatorname{simp}(K)$ and $\tau \in \operatorname{simp}(L)$. If $\nu(z) = (m,n)$ we find that $\operatorname{supp}(m) \subseteq \sigma$ and $\operatorname{supp}(n) \subseteq \tau$ so $\operatorname{supp}(\mu(m,n))$ is also contained in $\sigma \times \tau$. Thus, the linear path in $\operatorname{Map}(\operatorname{vert}(K) \times \operatorname{vert}(L), \mathbb{R})$ from $\mu\nu(z)$ to z lies wholly in the set $|\sigma \times \tau| \subseteq G(K \boxtimes L)$, so we have a linear homotopy $\mu\nu \simeq 1$ as required. □

There is one further way of combining two simplicial complexes that is useful for various applications.

Definition 29.16. [`defn-simplicial-join`]
Let K and L be abstract simplicial complexes. We define a new complex $K * L$ (called the *join* of K and L) as follows. The vertex set is $\operatorname{vert}(K) \amalg \operatorname{vert}(L)$. A subset ρ is a simplex if either

(a) $\rho \subseteq \operatorname{vert}(K)$ and ρ is a simplex of K; or
(b) $\rho \subseteq \operatorname{vert}(L)$ and ρ is a simplex of L; or
(c) $\rho = \sigma \amalg \tau$ for some $\sigma \in \operatorname{simp}(K)$ and $\tau \in \operatorname{simp}(L)$.

Example 29.17. [`eg-simplex-join`]
It is clear that $\Delta^a_I * \Delta^a_J = \Delta^a_{I \amalg J}$. Given $n, m \geq 0$ we have a bijection $f\colon [n] \amalg [m] \to [n+m+1]$ given by $f(i) = i$ on $[n]$, and $f(j) = n+1+j$ on $[m]$. Using this we obtain an isomorphism $\Delta_n * \Delta_m \simeq \Delta_{n+m+1}$.

Example 29.18. [`eg-octahedron-multiple`]
Consider the discrete complex K with vertex set $\{1,-1\}$, and the n-fold iterated join $K^{*n} = K * \cdots * K$. There is an evident way to identify the set

$$\operatorname{vert}(K^{*n}) = \{1,-1\} \amalg \cdots \amalg \{1,-1\}$$

with $\{1,\dots,n,-1,\dots,-n\}$, and by induction we see that $\operatorname{simp}(K^{*n}) = \{\sigma \;:\; \sigma \cap (-\sigma) = \emptyset\}$. As in Example 29.10 we can define $f\colon \operatorname{vert}(K^{*n}) \to \mathbb{R}^n$ by $f(\pm k) = \pm e_k$ and this gives an affine embedding $f\colon G(K^{*n}) \to \mathbb{R}^n$. We find that 0 is not in the image, so we can define $g\colon G(K^{*n}) \to S^{n-1}$ by $g(x) = f(x)/\|x\|$. This is a homeomorphism, with

$$g^{-1}(y)(\pm k) = \begin{cases} |y_k|/(\sum_i |y_i|) & \text{if } \pm y_k \geq 0 \\ 0 & \text{if } \pm y_k \leq 0. \end{cases}$$

We can also define a join operation for topological spaces, and prove that it is compatible with our join operation for simplicial complexes.

Definition 29.19. [`defn-join`]
Let X and Y be topological spaces. Define $f\colon X \times \{0,1\} \times Y \to X \amalg Y$ by $f(x,0,y) = x$ and $f(x,1,y) = y$. Define $X * Y$ to be the pushout in the following square:

$$\begin{tikzcd} X \times \{0,1\} \times Y \arrow[r, tail, "\text{inc}"] \arrow[d, "f"'] & X \times [0,1] \times Y \arrow[d] \\ X \amalg Y \arrow[r] & X * Y. \end{tikzcd}$$

More explicitly, $X * Y$ is the quotient space of $X \amalg Y \amalg (X \times [0,1] \times Y)$ in which all points of the form $(x,0,y)$ are identified with x, and all points of the form $(x,1,y)$ are identified with Y.

Remark 29.20. [`rem-empty-join`]
Provided that X and Y are nonempty, we can regard $X * Y$ as a quotient of $X \times [0,1] \times Y$. However, we have $\emptyset * Y = Y$ and $X * \emptyset = X$ whereas $X \times [0,1] \times Y$ is empty if X or Y is empty.

Example 29.21. [`eg-sphere-join`]
Let U and V be finite dimensional real vector spaces with inner products. We then have an evident inclusion $i\colon S(U) \amalg S(V) \to S(U \oplus V)$ and a surjective map $m\colon S(U) \times [0,1] \times S(V) \to S(U \oplus V)$ given by $m(u,t,v) = (\sqrt{t}u, \sqrt{1-t}v)$. One can check that these induce a homeomorphism $S(U) * S(V) \to S(U \oplus V)$.

Remark 29.22. **This probably needs to be moved.**
Suppose that X and Y are based, and put

$$\begin{aligned} TX &= \{[x,t,0_Y] \ : \ x \in X, \ t \in [0,1]\} \\ TY &= \{[0_X,t,y] \ : \ y \in Y, \ t \in [0,1]\} \\ T &= TX \cup TY. \end{aligned}$$

The map $[x,t,y] \mapsto t \wedge x \wedge y$ then gives a homeomorphism $(X * Y)/T \to \Sigma(X \wedge Y)$. Moreover, the spaces TX and TY are contractible, as is the space $TX \cap TY = \{[0_X,t,0_Y] \ : \ t \in [0,1]\}$. Under mild cofibrancy assumptions we can deduce that T is contractible and $X * Y \to \Sigma(X \wedge Y)$ is a homotopy equivalence.

Proposition 29.23. [`prop-join-realisation`]
For finite simplicial complexes K and L there is a natural homeomorphism $G(K) * G(L) = G(K * L)$.

Remark 29.24. A suitable version of this is true for infinite complexes, but we defer the details.

Proof. Any map $z\colon \operatorname{vert}(K * L) = \operatorname{vert}(K) \amalg \operatorname{vert}(L) \to [0,1]$ can be regarded as a pair (x,y), where $x\colon \operatorname{vert}(K) \to [0,1]$ and $y\colon \operatorname{vert}(L) \to [0,1]$. From the definitions we see that $G(K * L)$ is the set of pairs (x,y) such that

(a) Either $x = 0$ or $\operatorname{supp}(x) \in \operatorname{simp}(K)$; and
(b) Either $y = 0$ or $\operatorname{supp}(y) \in \operatorname{simp}(L)$; and
(c) $\sum_{u \in \operatorname{vert}(K)} x(u) + \sum_{v \in \operatorname{vert}(L)} y(v) = 1$.

The subspace where $y = 0$ is a copy of $G(K)$, and the subspace where $x = 0$ is a copy of $G(L)$. We can also define a map

$$m\colon G(K) \times [0,1] \times G(L) \to G(K * L)$$

by $m(x,t,y) = ((1-t)x, ty)$. By combining these we obtain a map

$$G(K) \amalg G(L) \amalg (G(K) \times [0,1] \times G(L)) \to G(K * L)$$

which is compatible with the relevant equivalence relation and so induces a map $\overline{m}\colon G(K) * G(L) \to G(K * L)$. If $(x,y) \in G(K * L)$ with $x \neq 0$ and $y \neq 0$, we can define $t = \sum_v y(v)$ and $x' = x/(1-t)$ and $y' = y/t$; we find that (x',t,y') is the unique point in $G(K) \times [0,1] \times G(L)$ that is sent by m to (x,y). Using this we see that $\overline{m}$ is a continuous bijection from a compact space to a Hausdorff space, so it is a homeomorphism. □

Definition 29.25. [`defn-contiguous`]
Let $f, g\colon K \to L$ be morphisms of ordered simplicial complexes. We say that f and g are *directly contiguous* if for all simplices σ in K, the set $f(\sigma) \cup g(\sigma)$ is a simplex in L. We say that f and g are *contiguous* if there is a list $f = p_0, p_1, \dots, p_r = g$ such that p_i and p_{i+1} are directly contiguous for all i. This defines an equivalence relation on $\mathbf{ASC}(K,L)$. We write $[f]$ for the contiguity class of f.

Lemma 29.26. [`lem-composite-contiguous`]
Suppose we have maps

$$K \underset{f_1}{\overset{f_0}{\rightrightarrows}} L \underset{g_1}{\overset{g_0}{\rightrightarrows}} M$$

where f_0 and f_1 are contiguous, and g_0 and g_1 are contiguous. Then $g_0 f_0$ and $g_1 f_1$ are also contiguous.

Proof. We can easily reduce to the case where f_0 and f_1 are directly contiguous, and g_0 and g_1 are also directly contiguous. Consider a simplex σ of K. Then the set $\tau = f_0(\sigma) \cup f_1(\sigma)$ is a simplex of L, so the set $\rho = g_0(\tau) \cup g_1(\tau)$ is a simplex of M. It is clear that $g_0 f_0(\sigma) \cup g_1 f_1(\sigma)$ is a nonempty subset of ρ, so it is again a simplex. Thus $g_0 f_0$ and $g_1 f_1$ are directly contiguous. □

Corollary 29.27. [`cor-contiguity-category`]
If we put $\boldsymbol{cASC}(K,L) = \boldsymbol{ASC}(K,L)/$contiguity, then there is a well-defined composition operation

$$\boldsymbol{cASC}(L,M) \times \boldsymbol{cASC}(K,L) \to \boldsymbol{cASC}(K,M),$$

given by $([g],[f]) \mapsto [gf]$. We thus have a category $\boldsymbol{cASC}$, whose objects are abstract simplicial complexes and whose morphisms are contiguity classes of maps. There is a functor $U\colon \boldsymbol{ASC} \to \boldsymbol{cASC}$ given by $U(K) = K$ on objects, and $U(f) = [f]$ on morphisms. □

Proposition 29.28. [prop-contiguous]
If f and g are contiguous, then $G(f)$ and $G(g)$ are homotopic.

Proof. As homotopy is an equivalence relation, it will suffice to treat the case where f and g are directly contiguous. In that case we can define

$$h\colon [0,1]\times G(K)\to \mathrm{Map}(\mathrm{vert}(L),[0,1])$$

by $h(t,x)=(1-t)G(f)(x)+t\,G(g)(x)$. Note that the support of x is a simplex, say σ. It follows that $\mathrm{supp}(h(t,x))\subseteq f(\sigma)\cup g(\sigma)$, which is a simplex in L by assumption, so $h(t,x)\in G(L)$. We thus have a map $h\colon [0,1]\times G(K)\to G(L)$, which is the required homotopy. ☐

Corollary 29.29. *There is a functor $\overline{G}\colon$ **cASC** $\to$ **hSpaces** given by $\overline{G}(K)=G(K)$ on objects, and by $\overline{G}([f])=[G(f)]$ on morphisms. The following diagram commutes:*

$$\begin{array}{ccc} \boldsymbol{ASC} & \xrightarrow{G} & \boldsymbol{Spaces} \\ {\scriptstyle U}\downarrow & & \downarrow{\scriptstyle U} \\ \boldsymbol{cASC} & \xrightarrow[\overline{G}]{} & \boldsymbol{hSpaces}. \end{array}$$

Proof. This is now clear. ☐

Remark 29.30. [rem-need-subdivision]
As simplices are rather inflexible, the map $\overline{G}\colon \mathbf{cASC}(K,L)\to[G(K),G(L)]$ is rarely a bijection. However, we can make it closer to a bijection by subdividing the simplices of K. This process will be discussed in more detail below.

We next give a construction giving a different criterion for maps of simplicial complexes to be homotopic.

Construction 29.31. [cons-simplicial-ELCX]
Let K be an abstract simplicial complex. For $x,y\in G(K)$ we can regard x and y as maps $\mathrm{vert}(K)\to[0,1]$ and take the pointwise minimum to get a map $\min(x,y)\colon \mathrm{vert}(K)\to[0,1]$. We then put $m(x,y)=\sum_v \min(x,y)(v)$, and note that this gives a continuous map $m\colon G(K)^2\to[0,1]$ with $m(x,x)=1$ for all x. It follows that the set

$$U=\{(x,y)\in G(K)^2\ :\ m(x,y)>0\}=\{(x,y)\in G(K)^2\ :\ \mathrm{supp}(x)\cap\mathrm{supp}(y)\neq\emptyset\}$$

is open. We define $p\colon U\to G(K)$ by $p(x,y)=\min(x,y)/m(x,y)$, and note that $\mathrm{supp}(p(x,y))=\mathrm{supp}(x)\cap\mathrm{supp}(y)$. We then define $h\colon[0,1]\times U\to G(K)$ by

$$h(t,x,y)=\begin{cases}(1-2t)x+2tp(x,y) & \text{for } 0\leq t\leq 1/2\\ (2-2t)p(x,y)+(2t-1)y & \text{for } 1/2\leq t\leq 1.\end{cases}$$

Note that for $t\leq 1/2$ we have $\mathrm{supp}(h(t,x,y))\subseteq\mathrm{supp}(x)$, and for $t\geq 1/2$ we have $\mathrm{supp}(h(t,x,y))\subseteq\mathrm{supp}(y)$, so this does indeed land in $G(K)$ as advertised. Note also that $h(0,x,y)=x$ and $h(1,x,y)=y$, so h gives a homotopy between the two projections $\pi_0,\pi_1\colon U\to G(K)$.

Lemma 29.32. [lem-diagonal-nbhd]
The open set U can also be described as $\{(x,y)\ :\ d(x,y)<2\}$.

Proof. As in Remark 29.4 we always have $d(x,y)\leq 2$, and this is only an equality if all the inequalities $|x(v)-y(v)|\leq x(v)+y(v)$ are equalities. That can only happen if the supports of x and y are disjoint, and the claim follows. ☐

Proposition 29.33. [lem-simplicial-ELCX]
Let K and L be simplicial complexes, and let f and g be maps from $G(K)$ to $G(L)$ such that for all $x\in G(K)$ we have $\mathrm{supp}(f(x))\cap\mathrm{supp}(g(x))\neq\emptyset$, or equivalently $d(f(x),g(x))<2$. Then f and g are homotopic.

Proof. Just use the homotopy $(t,x)\mapsto h(t,f(x),g(x))$, where h is as in Construction 29.31. ☐

Remark 29.34. [`rem-simplicial-ELCX`]

One could think of applying this proposition in the case where both f and g are simplicial maps, say $f = G(p)$ and $g = G(q)$ for maps $p, q \colon K \to L$ of abstract simplicial complexes. However, for distinct vertices v and w we have $d(v, w) = 2$, so the condition $d(G(p), G(q)) < 2$ can only be satisfied if $p = q$, which is not interesting. However, the proposition can still be useful if f is simplicial and g is not.

There is a close relationship between simplicial complexes and partially ordered sets. We recall the main definitions:

Definition 29.35. [`defn-poset`]

A *partial order* on a set P is an relation on P (written $p \leq q$) such that

PO0: For all $p \in P$ we have $p \leq p$
PO1: For all $p, q, r \in P$, if $p \leq q$ and $q \leq r$ then $p \leq r$.
PO2: For all $p, q \in P$, if $p \leq q$ and $q \leq p$ then $p = q$.

A *partially ordered set* or *poset* is a set with a specified partial order. A map $f \colon P \to Q$ of posets is *monotone* if whenever $p \leq p'$ in P, we also have $f(p) \leq f(p')$ in Q. We write **POSets** for the category of partially ordered sets and monotone maps.

Definition 29.36. [`defn-chain`]

Let P be a partially ordered set, and let C be a subset of P. We say that C is a *chain* if for all $p, q \in C$ we have either $p \leq q$ or $q \leq p$. We say that the partial order is *total* if the whole set P is a chain.

Definition 29.37. [`defn-OSC`]

An *ordered simplicial complex* is a simplicial complex K with a specified partial ordering on $\mathrm{vert}(K)$ such that every simplex is a chain. If K and L are ordered simplicial complexes, a *morphism* from K to L will mean a morphism $f \colon K \to L$ of simplicial complexes such that for each $\sigma \in \mathrm{simp}(K)$, the restriction $f \colon \sigma \to f(\sigma)$ is monotone. We write **OSC** for the category of ordered simplicial complexes.

Remark 29.38. [`rem-OSC-ASC`]

For any simplicial complex K we can choose a total order on $\mathrm{vert}(K)$, and this makes K an ordered simplicial complex. Thus, the forgetful functor **OSC** $\to$ **ASC** is surjective on objects.

Remark 29.39. [`rem-trimming`]

Because morphisms in **OSC** are not required to be monotone everywhere, it is possible for two ordered simplicial complexes to be isomorphic even if the posets of vertices are not isomorphic. We can eliminate this issue as follows. Let K be an ordered simplicial complex, and let x and y be vertices of K. We write $x \lll y$ if there is a sequence $x = w_0 \leq w_1 \leq \cdots \leq w_r = y$ with $\{w_i, w_{i+1}\} \in \mathrm{vert}(K)$ for $i = 0, \dots, r-1$. (We allow the degenerate case where $r = 0$ and $x = y$.) One can check that this gives a new partial order on $\mathrm{vert}(K)$, and that every simplex is still a chain. We thus have a new ordered simplicial complex K' with the same vertices and simplices, but with this modified order. The identity function on vertices can be regarded as a morphism $K \to K'$ or as a morphism $K' \to K$, so K and K' are isomorphic. Morphisms $K' \to L$ in **ASC** are automatically monotone everywhere. We refer to the construction $K \to K'$ as *trimming*, and say that K is *trimmed* if $K' = K$.

Definition 29.40. [`defn-poset-OSC`]

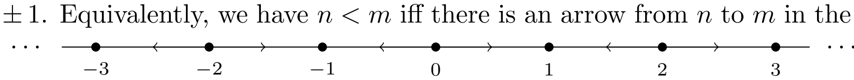
Let P be a partially ordered set. We can then regard P as an ordered simplicial complex by taking $\mathrm{vert}(P) = P$ and

$$\mathrm{simp}(P) = \{ \text{ finite, nonempty chains in } P\}.$$

If $f \colon P \to Q$ is a map of posets that is monotone on every chain in P, then it is clearly monotone. Using this, we see that $\mathbf{POSets}(P, Q) = \mathbf{OSC}(P, Q)$, so we can regard **POSets** as a full subcategory of **OSC**.

Example 29.41. [`eg-poset-Z`]

We can introduce a nonstandard order on $\mathbb{Z}$ as follows: we have $n \leq m$ if and only if $n = m$, or n is even and $m = n \pm 1$. Equivalently, we have $n < m$ iff there is an arrow from n to m in the following picture:



Thus, the chains are precisely the sets $\{n\}$ and $\{n, n+1\}$, and we deduce that $G(\mathbb{Z}) = \mathbb{R}$.

Example 29.42. [eg-poset-polygon]

Consider the following picture:

$$\begin{tikzcd}
a & b \arrow[l] \arrow[d] \\
c \arrow[u] \arrow[r] & d
\end{tikzcd}$$

We can partially order the set $P = \{a, b, c, d\}$ by declaring that $x < y$ iff there is an arrow from x to y in the diagram. The diagram itself then displays a homeomorphism from $G(P)$ to the boundary of a square.

More generally, we can partially order the set $Q_n = \mathbb{Z}/n \times \{0, 1\}$ by declaring that $(a, i) < (b, j)$ iff $i = 0$ and $j = 1$ and $b \in \{a, a+1\}$. We find that $G(Q_n)$ is homeomorphic to the boundary of a $2n$-gon (and thus to S^1). The above picture corresponds to the case $n = 2$.

Example 29.43. [eg-poset-cube]

Let I be a finite set, and let P be the set of subsets of I. We partially order this set by the rule $J \leq K$ iff $J \subseteq K$. This gives us a simplicial complex, and thus a space $G(P)$. We can define $f\colon P \to \mathrm{Map}(I, \mathbb{R})$ by

$$f(J) = \chi_J = \text{ characteristic function of } J.$$

We claim that the affine extension gives a homeomorphism $f_*\colon G(P) \to \mathrm{Map}(I, [0,1])$. Indeed, for each $J \subseteq I$ we can define $g_J\colon \mathrm{Map}(I, [0,1]) \to [0,1]$ by

$$g_J(u) = \max(0, \min(u(j) \ :\ j \in J) - \max(u(k) \ :\ k \not\in k)).$$

We interpret $\min(\emptyset)$ as 1 and $\max(\emptyset)$ as 0, so $g_\emptyset(u) = 1 - \max(u)$ and $g_I(u) = \min(u)$. These formulae make it clear that g_J is well-defined and continuous. To understand them more clearly, note that we have a finite subset $u(I) \cup \{0,1\} \subseteq [0,1]$, so we can list the elements in reverse order as $u(I) \cup \{0,1\} = \{s_0, \ldots, s_{r+1}\}$ with $1 = s_0 > \cdots > s_r > s_{r+1} = 0$. Put $K_i = u^{-1}\{s_i\}$; these sets are disjoint, and nonempty for $1 \leq i \leq r$. For $0 \leq i \leq r$ we put $J_i = K_0 \cup \cdots \cup K_i$ and $t_i = s_i - s_{i+1}$. We find that $J_0 \subset \cdots \subset J_r$ and $0 < t_i \leq 1$ and $\sum_i t_i = 1$. We also find that $g_{J_i}(u) = t_i$, and that $g_J(u) = 0$ if J is not one of the sets J_i. It follows that there is a map $g\colon \mathrm{Map}(I, [0,1]) \to G(P)$ given by $g(u)(J) = g_J(u)$, and that this is inverse to f_*.

Example 29.44. [eg-subgroup-posets]

Fix a prime p. For any finite group Γ, we let $\mathcal{P}(\Gamma)$ denote the poset of nontrivial p-subgroups of Γ, and we let $\mathcal{E}(\Gamma)$ denote the subposet of nontrivial elementary abelian p-subgroups. There are a number of interesting results relating the group theory of Γ to the homotopy theory of the spaces $G(\mathcal{P}(\Gamma))$ and $G(\mathcal{E}(\Gamma))$, some of which we will examine later. In particular, it can be shown that the inclusion $G(\mathcal{E}(\Gamma)) \to G(\mathcal{P}(\Gamma))$ is a homotopy equivalence.

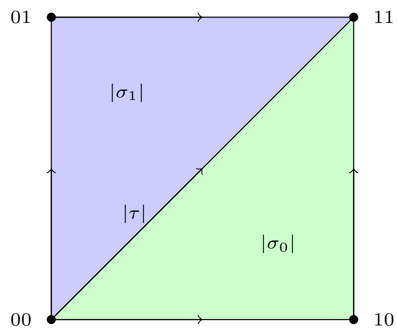
Remark 29.45. [rem-ASC-limits]

One advantage of working with ordered complexes is that in that context we can define products. Specifically, if P and Q are partially ordered sets, we can define a partial order on $P \times Q$ by $(p, q) \leq (p', q')$ iff ($p \leq p'$ and $q \leq q'$). If T is another poset, we find that a map $(f, g)\colon T \to P \times Q$ is monotone iff both $f\colon T \to P$ and $g\colon T \to Q$ are monotone, so $P \times Q$ is a product of P and Q in the sense of category theory. More generally, if we have a finite diagram consisting of partially ordered sets P_i and maps $u\colon P_i \to P_j$, we can form the limit set $Q \subseteq \prod_i P_i$ and order it by declaring that $q \leq q'$ iff $q_i \leq q'_i$ in P_i for all i. This makes Q a limit for the diagram in **POSets**.

Now suppose instead that we have a diagram of ordered simplicial complexes K_i. It will be harmless to assume that they are all trimmed, so all the maps of vertex sets are monotone. Let V be the limit of the vertex sets, with the partial order just discussed. We say that $\sigma \subseteq V$ is a simplex iff the projection $\pi_i(\sigma) \subseteq \mathrm{vert}(K_i)$ is a simplex for all i, and also σ is a chain. This gives us an ordered simplicial complex L with $\mathrm{vert}(L) = V$, and we see that this is a limit of the diagram.

Remark 29.46. [`rem-product-triangulation`]

We will prove later that $G\colon \mathbf{OSC} \to \mathbf{Spaces}$ preserves finite limits, so in particular $G(K \times L) = G(K) \times G(L)$. For the moment we just illustrate two cases of this. First, take $K = L = \Delta_1^a$. We then have $\text{vert}(K \times L) = \{00, 01, 10, 11\}$, so the maximal simplices are $\sigma_0 = \{00, 10, 11\}$ and $\sigma_1 = \{00, 01, 11\}$, with intersection $\tau = \sigma_0 \cap \sigma_1 = \{00, 11\}$. Thus $G(K \times L)$ is the union of the triangles $|\sigma_0|$ and $|\sigma_1|$, which share the common edge $|\tau|$. It can be displayed as follows:



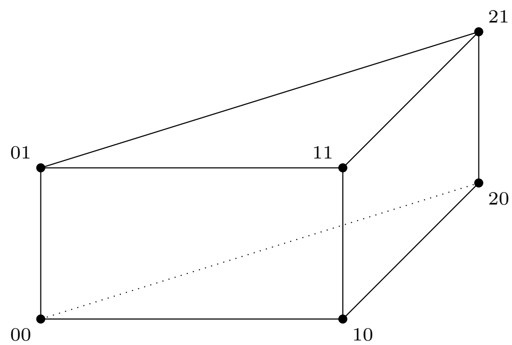
Now instead take $K = \Delta_2^a$ and $L = \Delta_1^a$ so

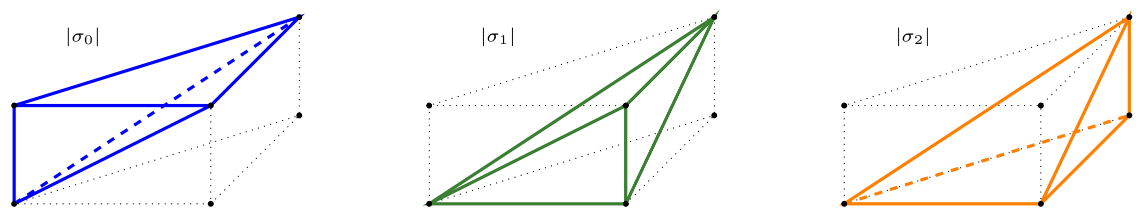
$$\text{vert}(K \times L) = \{00, 01, 10, 11, 20, 21\}$$

and there are three maximal simplices:

$$\sigma_0 = \{00, 01, 11, 21\} \qquad \sigma_1 = \{00, 10, 11, 21\} \qquad \sigma_2 = \{00, 10, 20, 21\}.$$

The product $|K| \times |L|$ decomposes as the union of three tetrahedra $|\sigma_i|$ as shown below.





Definition 29.47. For any simplicial complex K, we write sK for the set of simplices of K, regarded as a partially ordered set by inclusion. We call this the *(abstract) barycentric subdivision* of K. For each $\sigma \in sK$ we have a point $b(\sigma) \in G(K) \subseteq \text{Map}(\text{vert}(K), \mathbb{R})$ given by

$$b(\sigma)(v) = \begin{cases} 1/|\sigma| & \text{if } v \in \sigma \\ 0 & \text{otherwise.} \end{cases}$$

This is called the *barycentre* of σ. By affine extension we obtain a map $b_*\colon G(sK) \to \text{Map}(\text{vert}(K), \mathbb{R})$.

Proposition 29.48. [`prop-barycentric`]

The map b_* gives a homeomorphism $G(sK) \to G(K)$.

Proof. The general case reduces easily to the case where K is finite, so we restrict attention to that case from now on. Put $I = \mathrm{vert}(K)$, so $G(K) \subseteq \mathrm{Map}(I, [0,1])$. It is clear that each barycentre $b(\sigma)$ lies in $\mathrm{Map}(I, [0,1])$, and this set is convex, so it contains the whole image of b_*. Recall from Example 29.43 that we have maps $g_J \colon \mathrm{Map}(I, [0,1]) \to [0,1]$ such that $u = \sum_J g_J(u)\chi_J$ for all $u \in \mathrm{Map}(I, [0,1])$. Now suppose that $u \in G(K)$, so $\mathrm{supp}(u) \in \mathrm{simp}(K)$. From the definitions it is clear that $g_J(u) = 0$ unless $J \subseteq \mathrm{supp}(u)$, in which case J is either empty or a simplex. As $\chi_\emptyset = 0$, the corresponding term can be omitted. For any simplex σ we note that $b(\sigma) = \chi_\sigma/|\sigma|$, and we put $h_\sigma(u) = |\sigma|\, g_\sigma(u)$. We now find that for $u \in G(K)$ we have $u = \sum_{\sigma \in sK} h_\sigma(u)\, b(\sigma)$. From this it follows that

$$\sum_\sigma h_\sigma(u) = \sum_\sigma h_\sigma(u) \sum_{x\in\sigma} \frac{1}{|\sigma|} = \sum_x \sum_\sigma h_\sigma(u)\, b_\sigma(x) = \sum_x u(x) = 1$$

(and thus that $h_\sigma(u) \leq 1$ for all σ). We also saw in Example 29.43 that the set $\{\sigma \;:\; h_\sigma(u) > 0\}$ is a chain in sK. We can therefore define $h\colon G(K) \to G(sK)$ by $h(u)(\sigma) = h_\sigma(u)$, and we find that $b_* \circ h$ is the identity. A similar argument shows that $h \circ b_*$ is also the identity, as required. □

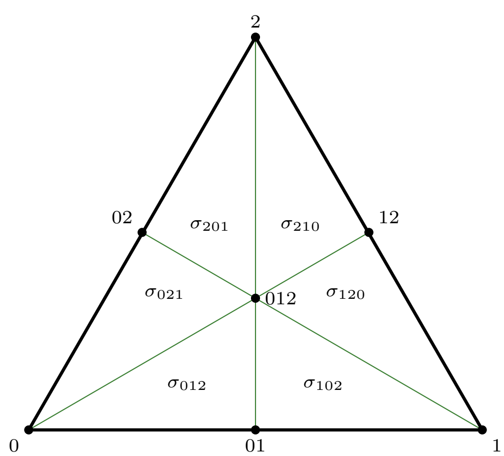
Example 29.49. Consider the case $K = \Delta_2^a$. We write 02 for the simplex $\{0, 2\}$ regarded as a vertex of $s\Delta_2^a$, and similarly for the other vertices, so

$$\mathrm{vert}(s\Delta_a^2) = \{0, 1, 2, 01, 02, 12, 012\}.$$

There are six maximal simplices (one for each permutation of $\{0, 1, 2\}$) as follows:

$$\begin{aligned}
\sigma_{012} &= \{0, 01, 012\} & |\sigma_0| &= \{x \in \Delta_2 \;:\; x_0 \geq x_1 \geq x_2\} \\
\sigma_{102} &= \{1, 01, 012\} & |\sigma_1| &= \{x \in \Delta_2 \;:\; x_1 \geq x_0 \geq x_2\} \\
\sigma_{120} &= \{1, 12, 012\} & |\sigma_2| &= \{x \in \Delta_2 \;:\; x_1 \geq x_2 \geq x_0\} \\
\sigma_{210} &= \{2, 12, 012\} & |\sigma_3| &= \{x \in \Delta_2 \;:\; x_2 \geq x_1 \geq x_0\} \\
\sigma_{201} &= \{2, 02, 012\} & |\sigma_4| &= \{x \in \Delta_2 \;:\; x_2 \geq x_0 \geq x_1\} \\
\sigma_{021} &= \{0, 02, 012\} & |\sigma_5| &= \{x \in \Delta_2 \;:\; x_0 \geq x_2 \geq x_1\}
\end{aligned}$$

The homeomorphism b_* can be displayed as follows:



Proposition 29.50. [prop-subdivision-mu]
For ordered simplicial complexes K, there is a natural map $\mu\colon sK \to K$ given by $\mu(\sigma) = \max(\sigma)$, and $G(\mu)\colon G(sK) \to G(K)$ is homotopic to the homeomorphism b_*, so it is a homotopy equivalence.

Proof. First, as each simplex $\sigma \in \mathrm{vert}(sK)$ is a nonempty chain, it certainly has a unique largest element, so we have a well-defined map $\mu\colon \mathrm{vert}(sK) \to \mathrm{vert}(K)$ as indicated. Now suppose we have a simplex ω in sK, or equivalently, a chain $\omega = (\sigma_0 \subset \cdots \subset \sigma_r)$ of simplices in K. Put $v_i = \mu(\sigma_i) \in \sigma_i \subseteq \sigma_r$ and $\tau = \mu(\sigma) = \{v_0, \ldots, v_r\}$. As $\tau \subseteq \sigma_r$ and σ_r is a simplex, we see that τ is also a simplex. Thus, μ is a morphism of ordered simplicial complexes. It therefore gives a map $G(\mu)\colon G(sK) \to G(K)$. We can thus define a map

$$h\colon [0,1] \times G(sK) \to \mathrm{Map}(|(K), \mathbb{R})$$

by $h(t,x) = (1-t)b_*(x) + t\,G(\mu)(x)$. If we can prove that $h(t,x) \in G(K)$ for all t, we will have the required homotopy. We may assume that $x \in |\omega|$ for some ω as above, with $x(\sigma_i) = t_i$ say. We then have $b_*(x) = \sum_i t_i b(\sigma_i)$ whereas $G(\mu)(x) = \sum_i t_i v_i$. The points $b(\sigma_i)$ and v_i all lie in the convex set $|\sigma_r| \subseteq G(K)$, and it follows that $h(t,x) \in |\sigma_r| \subseteq G(K)$ for all $t \in [0,1]$ as required. □

Remark 29.51. [`rem-barycentric-natural`]

Let $f\colon K \to L$ be a map of simplicial complexes. For any simplex σ of K we have a simplex $f(\sigma)$ of L, and if $\sigma \subseteq \tau$ then $f(\sigma) \subseteq f(\tau)$. Thus, the construction $K \mapsto sK$ gives a functor $\mathbf{ASC} \to \mathbf{OSC}$. Next, if f is injective we find that $G(f)\colon G(K) \to G(L)$ sends $b(\sigma)$ to $b(f(\sigma))$, and thus that the naturality diagram

$$\begin{tikzcd} G(sK) \arrow[r, "b_*"] \arrow[d, "G(sf)"'] & G(K) \arrow[d, "G(f)"] \\ G(sL) \arrow[r, "b_*"'] & G(L) \end{tikzcd}$$

commutes. However, this does not work when f is not injective. For example, for $0 < k < n$ we have a map $f\colon \Delta_n^a \to \Delta_2^a$ given by $f(0) = \cdots = f(k) = 0$ and $f(k+1) = \cdots = f(n) = 1$, and the induced map $\Delta_n \to \Delta_1$ sends the barycentre to the point $((k+1)/(n+1), (n-k)/(n+1))$, which is generally not the same as the barycentre at $(1/2, 1/2)$.

It is clear that the simplices for $G(sK)$ are in some sense smaller than those for K. It will be useful to quantify this. Consider a maximal simplex $\omega \in \mathrm{simp}(s\Delta_n^a)$, so

$$\omega = \{\{i_0\}, \{i_0, i_1\}, \{i_0, i_1, i_2\}, \dots \{i_0, \dots, i_n\}\}$$

for some enumeration of the set $\{0, \dots, n\}$ as $\{i_0, \dots, i_n\}$. The map $k \mapsto i_k$ gives a homeomorphism $\Delta_n \to |\omega| \subseteq G(s\Delta_n^a)$, which we compose with b_* to get a map $\theta_\omega\colon \Delta_n \to \Delta_n$.

Proposition 29.52. [`prop-tht-contraction`]

If we use the metric on Δ_n given by

$$d(x,y) = \sum_{i=0}^{n} |x_i - y_i|$$

then $d(\theta_\omega(x), \theta_\omega(y)) \leq \frac{n}{n+1} d(x,y)$.

Proof. Nothing interesting depends on ω, so we may assume that

$$\omega = \{\{0\}, \{0,1\}, \{0,1,2\}, \dots, \{0, \cdots, n\}\}.$$

In this context we write θ for θ_ω. If e_p denotes the p'th standard basis vector for $\mathbb{R}^{n+1}$ (for $0 \leq p \leq n$) then we have $\theta(e_p) = (\sum_{q \leq p} e_q)/(p+1)$, or $\theta(x)_i = \sum_{j=i}^{n} x_j/(j+1)$. Using this formula, we extend θ to give a linear automorphism of $\mathbb{R}^{n+1}$.

Put $w_i = x_i - y_i$ (so $\sum_i w_i = 0$) and $\|w\| = \sum_i |w_i| = d(x,y)$. The claim is then that $\|\theta(w)\| \leq \frac{n}{n+1}\|w\|$. In the special case where $w = e_q - e_p$ with $p \neq q$ we have $\|w\| = 2$, and we claim that $\|\theta(w)\| \leq 2n/(n+1)$.

It will be harmless to assume that $p \leq q$, and we then have

$$\begin{aligned}
\theta(w) &= \sum_{i=0}^{q} \frac{e_i}{q+1} - \sum_{i=0}^{p} \frac{e_i}{p+1} \\
&= \sum_{i=p+1}^{q} \frac{1}{q+1} e_i - \sum_{i=0}^{p} \frac{q-p}{(p+1)(q+1)} e_i \\
\|\theta(w)\| &= \sum_{i=p+1}^{q} \frac{1}{q+1} + \sum_{i=0}^{p} \frac{q-p}{(p+1)(q+1)} = 2\frac{q-p}{q+1} \\
&= 2\left(1 - \frac{p+1}{q+1}\right) \leq 2\left(1 - \frac{1}{n+1}\right) = 2\frac{n}{n+1}
\end{aligned}$$

as required. Now consider the general case again. Put $J = \{i \;:\; w_i > 0\}$ and $K = \{i \;:\; w_i < 0\}$ and $a = \sum_{j \in J} w_j$. To avoid trivialities, we may assume that $J, K \neq \emptyset$ and so $a > 0$. As $\sum_i w_i = 0$ we also have $\sum_{k \in K} w_k = -a$ and so $d(x,y) = \sum_i |w_i| = 2a$. Now put $b_{jk} = -w_j w_k$ (for $(j,k) \in J \times K$) so $\sum_j b_{jk} = -a w_k$ and $\sum_k b_{jk} = a w_j$ and $\sum_{j,k} b_{jk} = a^2$. It is now straightforward to check that $aw = \sum_{j,k} b_{jk}(e_j - e_k)$, so

$$a\|\theta(w)\| \leq \sum_{j,k} b_{jk} \|\theta(e_j - e_k)\| \leq \sum_{j,k} b_{jk} \frac{2n}{n+1} = \frac{2na^2}{n+1}.$$

After dividing by a and recalling that $\|w\| = 2a$ we find that $\|\theta(w)\| \leq \frac{n}{n+1}\|w\|$ as claimed. □

Definition 29.53. [`defn-diameter`]
Let X be a metric space, and let Y be a subset of X. Then the *diameter* of Y is

$$\operatorname{diam}(Y) = \sup\{d(a,b) \;:\; a, b \in Y\}.$$

Corollary 29.54. [`cor-subdivision-diameter`]
Let K be a simplicial complex of dimension n, and use the metric $d(x,y) = \sum_v |x(v) - y(v)|$ on $G(K)$. Then for any simplex σ of $s^r K$, the image $b_*^r|\sigma| \subseteq G(K)$ has diameter at most $2(n/(n+1))^r$.

Proof. In the case $r = 0$ the claim is just that every simplex of K has diameter at most 2, which is clear because

$$d(x,y) = \sum_v |x(v) - y(v)| \leq \sum_v (x(v) + y(v)) = \sum_v x(v) + \sum_v y(v) = 2.$$

The general case follows by induction using the proposition. □

For ordered simplicial complexes we can give a different description of the geometric realisation, which is convenient for various purposes.

Definition 29.55. [`defn-monotone-regular`]
Let P be a partially ordered set. We say that a monotone map $u \colon [0,1) \to P$ is *regular* if

(a) $\operatorname{img}(u)$ is finite; and
(b) for all $s \in [0,1)$ there exists $t > s$ such that u is constant on $[s,t)$.

Remark 29.56. [`rem-adjust-regular`]
Let $u \colon [0,1) \to P$ be a monotone map, and suppose that $x \in \operatorname{img}(u)$. Then $u^{-1}\{x\}$ is a nonempty convex subset of $[0,1)$, so it must have the form $[a,b]$ or $[a,b)$ or $(a,b]$ or (a,b) for some a and b. Condition (b) just requires that $u^{-1}\{x\}$ must have the form $[a,b)$ for all $x \in \operatorname{img}(u)$. If u satisfies (a) but not (b), we note that $u(t+\epsilon)$ is independent of ϵ for small $\epsilon > 0$. We call this value $u'(t)$, and we find that u' is a regular map $[0,1) \to P$ with $u'(t) = u(t)$ for all but finitely many values of t. We also use the shorthand $u(t+0^+)$ for $u'(t)$.

Definition 29.57. [`defn-G-prime`]
Let K be an ordered simplicial complex. We put

$$G'(K) = \{u \colon [0,1) \to \operatorname{vert}(K) \;:\; \operatorname{img}(u) \text{ is a simplex, and } u \text{ is monotone and regular }\}.$$

For $u,v \in G'(K)$ we put $D(u,v) = \{t \;:\; u(t) \neq v(t)\}$. This is a finite union of intervals, and we write $d(u,v)$ for their total length. This gives a metric on $G'(K)$. If K is finite we give $G'(K)$ the metric topology. If K is infinite, we declare that a set $U \subseteq G'(K)$ is open iff $U \cap G'(L)$ is open in $G'(L)$ for all finite subcomplexes $L \subseteq K$.

If $f\colon K \to L$ is a morphism of ordered simplicial complexes, we define $G'(f)\colon G'(K) \to G'(L)$ by $G'(f)(u) = f \circ u$. This makes G' into a functor **OSC** $\to$ **Spaces**.

Proposition 29.58. [`prop-G-G-prime`]
There is a natural homeomorphism $\lambda\colon G'(K) \to G(K)$ given by $\lambda(u)(a) = \mathrm{len}(u^{-1}\{a\})$.

Proof. It is easy to reduce the general case to the case where K is finite, so we restrict attention to that case from now on.

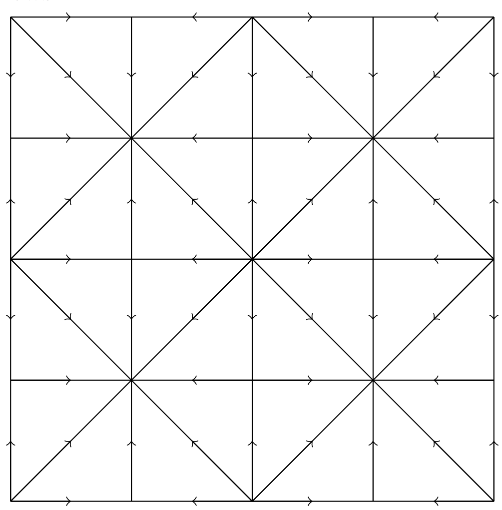
Consider an arbitrary point $u \in G'(K)$. The image of u is then a simplex, and thus a nonempty chain, say $\mathrm{img}(u) = \{a_0,\dots,a_r\}$ with $a_0 < \dots < a_r$. As u is monotone and regular, we see that there are numbers $0 = s_0 < \dots < s_{r+1} = 1$ with $u^{-1}\{a_i\} = [s_i, s_{i+1})$, so $\lambda(u)(a_i) = s_{i+1} - s_i$ and $\lambda(u)(b) = 0$ for $b \notin \{a_0,\dots,a_r\}$. This shows that $\lambda(u) \in G(K)$, as required.

In the opposite direction, suppose we have a simplex $\sigma = \{a_0,\dots,a_r\}$ with $a_0 < \dots < a_r$. Consider a point $x \in |\sigma|$, with $x(a_i) = t_i$ say. We put $s_i = \sum_{j<i} t_j$ and define $\mu_\sigma(x) \in G'(K)$ by $\mu_\sigma(x)(t) = a_i$ for $s_i \leq t < s_{i+1}$. If x and x' are nearby points in $|\sigma|$ then the corresponding numbers s_i and s'_i will be close, so the set $D(\mu_\sigma(x), \mu_\sigma(x'))$ will be a union of at most $2r+2$ very short intervals, so $d(\mu_\sigma(x), \mu_\sigma(x'))$ will be small. It follows that μ defines a continuous map $|\sigma| \to G'(K)$. The maps for different simplices σ are compatible, so they patch together to give a continuous map $\mu\colon G(K) \to G'(K)$, which is clearly inverse to λ. Now μ is a continuous bijection from a compact space to a Hausdorff space, so it is a homeomorphism by Proposition 10.22. Thus, λ is also a homeomorphism, as claimed. $\square$

Corollary 29.59. [`cor-limit-realisation`]
*The functors $G, G'\colon$ **OSC** $\to$ **Spaces** preserve finite limits.*

Proof. This is clear from the definitions for G', and the statement for G follows using the proposition. $\square$

Example 29.60. [`eg-poset-Zr`]
Consider $\mathbb{Z}$ with the nonstandard partial order described in Example 29.41, so $G'(\mathbb{Z}) \simeq G(\mathbb{Z}) \simeq \mathbb{R}$. This gives a partial order on $\mathbb{Z}^d$, for which $G'(\mathbb{Z}^d) \simeq G(\mathbb{Z}^d) \simeq \mathbb{R}^d$ by the corollary. In the case $d = 2$, we can draw a finite part of the picture as follows:



Corollary 29.61. [`cor-order-homotopy`]
Let $f, g\colon P \to Q$ be monotone maps of posets, and suppose that $f(p) \leq g(p)$ for all $p \in P$. Then $G(f)$ and $G(g)$ are homotopic.

Proof. Regard $\{0,1\}$ as a poset in the obvious way, and define $h\colon \{0,1\}\times P \to Q$ by $h(0,p) = f(p)$ and $h(1,p) = g(p)$. Our assumptions imply that h is a monotone map, so it induces a map

$$G(h)\colon G(\{0,1\}) \times G(P) = [0,1]\times G(P) \to G(Q)$$

of spaces; this is the required homotopy. □

Example 29.62. [eg-largest-element]
Suppose that P has a largest element, say $a \in P$ with $a \geq p$ for all $p \in P$. We can then let $f\colon P \to P$ be the identity, and let $g\colon P \to P$ be the constant map with value a; we find from this that $G(P)$ is contractible. Similarly, if P has a smallest element then $G(P)$ is again contractible.

Now let Γ b a finite group, and let p be a prime. As in Example 29.44, we let $\mathcal{P} = \mathcal{P}(\Gamma)$ be the poset of nontrivial p-subgroups of Γ, and let $\mathcal{E} = \mathcal{E}(\Gamma)$ be the subposet of nontrivial elementary abelian subgroups. We will illustrate the theory that we have just developed by proving some results (due to Quillen) about the topology of $G(\mathcal{P})$ and $G(\mathcal{E})$.

Lemma 29.63. [lem-normal-p-subgroup]
If Γ has a nontrivial normal p-subgroup, then $G(\mathcal{P})$ is contractible. In particular, this holds if Γ is itself a nontrivial p-group.

Proof. Let N be a nontrivial normal p-subgroup. If P is any other nontrivial p-subgroup, it is standard that the set $NP = \{xy \;:\; x \in N, y \in P\}$ is a subgroup of Γ of order $|N||P|/|N \cap P|$, so in particular it is a nontrivial p-subgroup. It is clear that when $P \leq Q$ we have $P \leq NP \leq NQ$ and also $N \leq NP$. We thus have poset maps $f, g, h\colon \mathcal{P} \to \mathcal{P}$ given by $f(P) = P$, and $g(P) = NP$, and $h(P) = N$. These satisfy $f \leq g \geq h$, so $G(f)$, $G(g)$ and $G(h)$ are all homotopic by Corollary 29.61. As f is the identity and h is constant, this gives a contraction of $G(\mathcal{P})$. □

Proposition 29.64. [prop-elem-ab]
The inclusion $G(\mathcal{E}) \to G(\mathcal{P})$ is a homotopy equivalence.

Proof. We have an inclusion map $i\colon \mathcal{E} \to \mathcal{P}$, and also maps μ as in Proposition 29.50. As μ is natural, we see that square below commutes:

$$\begin{tikzcd} s\mathcal{E} \arrow[r, "si"] \arrow[d, "\mu"'] & s\mathcal{P} \arrow[d, "\mu"] \\ \mathcal{E} \arrow[r, "i"'] & \mathcal{P}. \end{tikzcd}$$

Now consider a point $\sigma = \{P_0, \dots, P_d\} \in s\mathcal{P}$, so the P_k are p-subgroups of Γ with $1 < P_0 < \cdots < P_r$. It is a standard result about p-groups that the centre ZP_k is nontrivial, and if z is a nontrivial element then some power z^{p^k} will have order p. It follows that the group

$$E_k = \{z \in ZP_k \;:\; z^p = 1\}$$

is nontrivial, and elementary abelian. If $k \leq j$, we do not have $E_k \leq E_j$ in general. However, we do have $E_k \leq G_j$ and E_j is central in G_j so E_k commutes with E_j. As this holds for all j and k, we find that the group

$$\alpha(\sigma) = \alpha(\{P_0, \dots, P_d\}) = E_0 E_1 \cdots E_d$$

is nontrivial and elementary abelian. If $\sigma, \tau \in s\mathcal{P}$ with $\sigma \subseteq \tau$, it is clear that $\alpha(\sigma) \leq \alpha(\tau)$. We can thus regard α as a map $s\mathcal{P} \to \mathcal{E}$ of posets. Note that

$$i\alpha(\sigma) = \prod_{k=1}^{d} E_k \leq P_d = \mu(\sigma),$$

so Corollary 29.61 tells us that $G(i)G(\alpha)$ is homotopic to $G(\mu)$. On the other hand, if the groups P_k are all elementary abelian then $E_k = P_k$ and so $\prod_k E_k = P_d$; this shows that $\alpha \circ (si) = \mu\colon s\mathcal{E} \to \mathcal{E}$. We now see that the diagram

$$\begin{tikzcd}
G(s\mathcal{E}) \arrow[r, "G(si)"] \arrow[d, "G(\mu)"', "\simeq"] & G(s\mathcal{P}) \arrow[d, "\simeq"', "G(\mu)"] \arrow[dl, "G(\alpha)"'] \\
G(\mathcal{E}) \arrow[r, "G(i)"'] & G(\mathcal{P}).
\end{tikzcd}$$

commutes up to homotopy, and the vertical maps are homotopy equivalences. It follows formally that all maps in the diagram, including $G(\alpha)$, are homotopy equivalences. In detail, let $g\colon G(\mathcal{E}) \to G(s\mathcal{E})$ and $h\colon G(\mathcal{P}) \to G(s\mathcal{P})$ be homotopy inverses for the corresponding maps μ. By composing the relation $G(\alpha)G(si) \simeq G(\mu)$ with g, we obtain $G(\alpha)G(si)g \simeq 1_{G(\mathcal{E})}$. By composing the relation $G(i)G(\alpha) \simeq G(\mu)$ with h, we obtain $hG(i)G(\alpha) \simeq 1_{G(s\mathcal{P})}$ By combining these, we find that

$$hG(i) = hG(i) \circ 1_{G(\mathcal{E})} \simeq hG(i)G(\alpha)G(si)g \simeq 1_{G(s\mathcal{P})}G(si)g = G(si)g.$$

If we call this map k, we deduce that $kG(\alpha) \simeq 1_{G(s\mathcal{P})}$ and $G(\alpha)k \simeq 1_{G(\mathcal{E})}$, so k is the required homotopy inverse for $G(\alpha)$. ☐

We next consider a different kind of subdivision, which is often more convenient that the barycentric version.

Definition 29.65. [`defn-cubic-subdivision`]
For any poset P, we put $cP = \{(p,q) \in P^2 \ : \ p \leq q\}$, with the following ordering:

$$(p,q) \leq (p',q') \text{ iff } p' \leq p \leq q \leq q'.$$

Equivalently, we can define an "interval" $[p,q] = \{x \ : \ p \leq x \leq q\}$, and then we have $(p,q) \leq (p',q')$ iff $[p,q] \subseteq [p',q']$. We call cP the *cubic subdivision* of P.

We also define a poset map $\pi\colon cP \to P$ by $\pi(p,q) = q$. There is an obvious way to make c a functor from posets to posets, and then π is a natural map.

Proposition 29.66. *There is a natural homeomorphism $\phi\colon G'(cP) \to G'(P)$ that is naturally homotopic to $G'(\pi)$. In particular, $G'(\pi)$ is a homotopy equivalence.*

Proof. Any poset map $v\colon [0,1) \to cP$ can be written in the form $v(x) = (v_0(x), v_1(x))$ where $v_i\colon [0,1) \to P$. Here v_0 reverses the order and v_1 preserves it, and $v_0(x) \leq v_1(x)$ for all x. Using the notation of Remark 29.56, we define $\phi(v)\colon [0,1) \to P$ by

$$\phi(v)(x) = \begin{cases} v_0(1-2x-0^+) & \text{if } 0 \leq x < 1/2 \\ v_1(2x-1) & \text{if } 1/2 \leq x < 1. \end{cases}$$

One can check that this preserves order and is regular. Next, given a poset map $u\colon [0,1) \to P$ we define $\psi(u)\colon [0,1) \to cP$ by

$$\psi(u)(x) = (u((1-x-0^+)/2), u((1+x)/2)),$$

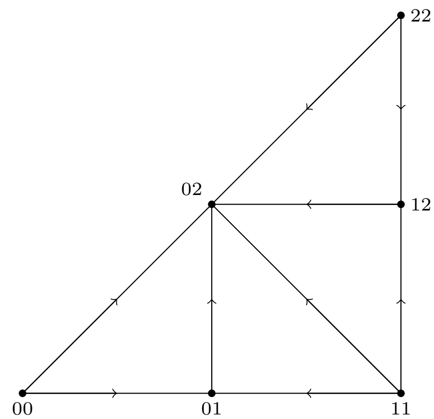
which is again a regular order-preserving map map. We next claim that $\phi(\psi(u))(x) = u(x)$, and $\psi(\phi(v))(x) = v(x)$. Indeed, it is easy to see from the formulae that this is true for all but finitely many values of x, and all maps considered are regular, so this is enough.

Finally, we need to show that ϕ is homotopic to $G'(\pi)$, or equivalently that $\lambda\phi \simeq \lambda G'(\pi)\colon G'(cP) \to G(P)$, where λ is as in Proposition 29.58. Put $\widehat{v}_0(x) = v_0(1-x-0^+)$, so $\widehat{v}_0 \in G'(P)$. One checks that $\lambda\phi(v) = (\lambda(\widehat{v}_0) + \lambda(v_1))/2$, whereas $\lambda G'(\pi)(v) = \lambda(v_1)$. We define $h\colon [0,1] \times G'(cP) \to \mathrm{Map}(P,\mathbb{R})$ by

$$h(t,v) = \frac{1-t}{2}\lambda(\widehat{v}_0) + \frac{1+t}{2}\lambda(v_1).$$

It follows that $\mathrm{supp}(h(t,v)) \subseteq v_0([0,1)) \cup v_1([0,1))$, and this is a chain in P, so h is a map from $[0,1] \times G'(cP) \to G(P)$, which gives the required homotopy. ☐

Example 29.67. [`eg-cubic-triangle`]
The cubic subdivision of Δ_2 can be displayed as follows:



(We have just written 02 for the vertex $(0,2) \in c\Delta_2^a$ and so on.)

Remark 29.68. [`rem-cubic-natural`]
Unlike the situation in Remark 29.51, the map ϕ is natural with respect to all maps of posets, not just injective ones.

Remark 29.69. [`rem-cubic-limits`]
From the definitions it is clear that $c(P \times Q) = c(P) \times c(Q)$. More generally, the functor $c\colon \mathbf{POSets} \to \mathbf{POSets}$ preserves all finite limits.

29.1. Simplicial approximation.

Definition 29.70. [`defn-ostar`]
Let K be a simplicial complex, and let v be a vertex of K. The *open star* of v in K is the set

$$\mathrm{ostar}_K(v) = \{x \in G(K) \;:\; x(v) > 0\}.$$

Remark 29.71. Let us identify v with the corresponding point $e_v \in G(K)$, defined by $e_v(v) = 1$ and $e_v(w) = 0$ for $w \neq v$.

Lemma 29.72. *Let $p\colon K \to L$ be a morphism of abstract simplicial complexes, and let $f\colon |K| \to |L|$ be a continuous map such that $f(\mathrm{ostar}_K(v)) \subseteq \mathrm{ostar}_L(p(v))$ for all $v \in \mathrm{vert}(K)$.*

Recall that Proposition 29.48 gives a homeomorphism $b_*\colon G(sK) \to G(K)$ for any abstract simplicial complex K. By iterating this, we obtain a homeomorphism $b_*^r\colon G(s^rK) \to G(K)$.

Proposition 29.73. [`prop-SAT`]
Let K and L be simplicial complexes, with K finite, and let $f\colon G(K) \to G(L)$ be a continuous map. Then for sufficiently large r there exists a map $p\colon s^rK \to L$ of abstract simplicial complexes such that f is homotopic to $G(p) \circ (b_*^r)^{-1}$.

Proof. We will use the metric on $G(K)$ given by $d(x,y) = \sum_v |x(v) - y(v)|$. The metric topology is the appropriate one, because K is finite.

For each vertex $w \in L$, put $V_w = \{y \in G(L) \;:\; y(v) > 0\} \subseteq G(L)$ and $U_w = f^{-1}(V_w) \subseteq G(K)$. The sets U_w then form an open cover of the compact metric space $G(K)$, so there is a Lebesgue number $\epsilon > 0$ by Theorem 12.28. Let n be the dimension of K, and choose r large enough that $2(n/(n+1))^r < \epsilon$. Put $g = f \circ b_*^r$ ∎

30. CW complexes

We now discuss CW complexes, which are spaces that are built from balls and spheres in a well-controlled way.

30.1. Basic definitions and examples.

Definition 30.1. Let X be a CGWH space. A *CW structure* on X consists of a set A, a map $d\colon A \to \mathbb{N}$, and a quotient map

$$\phi\colon \coprod_{a\in A} [0,1]^{d(a)} \to X$$

satisfying the conditions listed below. These are formulated in terms of the sets

$$\begin{aligned} E_a &= \phi(\{a\} \times [0,1]^{d(a)}) \\ DE_a &= \phi(\{a\} \times \partial([0,1]^{d(a)})) \\ OE_a &= \phi(\{a\} \times (0,1)^{d(a)}). \end{aligned}$$

When $d(a) = 0$ these are interpreted as $DE_a = \emptyset$ and $OE_a = E_a$. The conditions are:

CW0: ϕ restricts to give a bijection $\coprod_a (0,1)^{d(a)} \to X$ (so the sets OE_a are disjoint and cover X).

CW1: For each $a \in A$ the set

$$A_a = \{b \;:\; OE_b \cap DE_a \neq \emptyset\}$$

is finite, and for all $b \in A_a$ we have $d(b) < d(a)$.

A *finite CW structure* is a CW structure for which the set A is finite. The sets E_a are called the *cells* of the CW structure. A *(finite) CW complex* is a space equipped with a specified (finite) CW structure. The *dimension* is the maximum value of d (or ∞, if d is unbounded).

Remark 30.2. [`rem-carrier`]

As a special case of axiom CW1 we have $a \notin A_a$, so $OE_a \cap DE_a = \emptyset$. It also follows directly from the axioms that

$$DE_a \subseteq \bigcup_{b\in A_a} OE_b \subseteq \bigcup_{b\in A_a} E_b.$$

If we put $A_a^+ = A_a \cup \{a\}$ we therefore have

$$E_a \subseteq \bigcup_{b\in A_a^+} OE_b \subseteq \bigcup_{b\in A_a^+} E_b.$$

Remark 30.3. [`rem-ball-models`]

We have given the definition in terms of the standard cubes $[0,1]^d$, but it is often more convenient to work with different spaces. Suppose we have a system of spaces $(B_a)_{a\in A}$ with B_a homeomorphic to $[0,1]^{d(a)}$, and a quotient map $\phi\colon \coprod_{a\in A} B_a \to X$. We can choose homeomorphisms $f_a\colon [0,1]^{d(a)} \to B_a$ and put $\phi' = \phi\circ\coprod_a f_a$. We will allow ourselves to say that ϕ is a CW structure if ϕ' is a CW structure. To be rigorous, we should check that this does not depend on the choice of the maps f_a. The only issue is to show that the image $f_a(\partial([0,1]^d)) \subseteq B_a$ does not depend on f, or equivalently, that every homeomorphism $g\colon [0,1]^d \to [0,1]^d$ preserves $\partial([0,1]^d)$. This is true but surprisingly hard to prove. **Reference?** Alternatively, we can avoid this issue by imposing restrictions on B_a and f_a. In most cases B_a will be a compact convex subset of some finite-dimensional vector space and no trouble will arise if we use the homeomorphism arising from Proposition 19.44.

Remark 30.4. [`rem-finite-CW`]

Suppose we have a Hausdorff space X and a continuous surjection $\phi\colon \coprod_{a\in A} B^{d(a)} \to X$ with A finite. It is then automatic (from Proposition 10.22(c)) that ϕ is a quotient map, and also that the sets A_a in CW1 are finite. Thus, we will have a finite CW structure iff CW0 and the second half of CW1 are satisfied.

For expository purposes, it is convenient to introduce a weaker notion as follows.

Definition 30.5. A *pre-CW structure* on X is a decomposition of X as a a disjoint union of subsets X_a such that each X_a is homeomorphic to OB^d (or equivalently, to $\mathbb{R}^d$) for some d.

A CW structure on X decomposes X as the disjoint union of the sets OE_a, and gives continuous bijections $\phi_a\colon OB^{d(a)} \to OE_a$ for all a. We will check later that ϕ_a is actually a homeomorphism, so any CW structure gives a pre-CW structure. Pre-CW structures are not very useful in themselves, but they can often be constructed with relatively little work, and then one can try to improve them to get genuine CW structures.

Example 30.6. [eg-CW-R]
The sets $\{n\}$ and $(n, n+1)$ (for $n \in \mathbb{Z}$) give a pre-CW structure on $\mathbb{R}$. To improve this to a CW structure, put

$$\begin{aligned} A &= \mathbb{Z} \times \{0,1\} \\ d(n,k) &= k \\ \phi(n,0,0) &= n \\ \phi(n,1,t) &= n+t. \end{aligned}$$

We next consider several different CW structures on the space S^n.

Example 30.7. [eg-CW-sphere-min]
Put $e_0 = (0,\dots,0,1) \in S^n$ and $X_0 = \{e_n\}$ and $X_n = S^n \setminus X_0$. Recall from Proposition 18.20 that X_n is homeomorphic to $\mathbb{R}^n$, so this gives a pre-CW structure with one 0-cell and one n-cell. To improve this to a CW structure, we regard $\mathbb{R}^{n+1}$ as $\mathbb{R}^n \times \mathbb{R}$ and define $\phi\colon B_2^0 \amalg B_2^n \to S^n$ by $\phi(0) = e_n$ on B_2^0 and

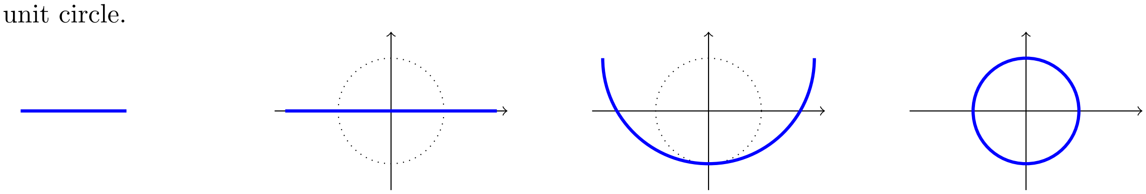
$$\phi(x) = (2x\sqrt{1 - \|x\|^2}, 2\|x\|^2 - 1)$$

on B_2^n. It is straightforward to check that

$$\|\phi(x)\|^2 = 4\|x\|^2(1 - \|x\|^2) + (2\|x\|^2 - 1)^2 = 1,$$

so this lands in S^n as claimed. Moreover, it restricts to give a homeomorphism $OB_2^n \to X_n$ with inverse $(s,y) \mapsto y/\sqrt{2(1-s)}$.

When $n = 1$ the geometric picture is as follows. The ball B_2^1 is just the interval $[-1,1]$, which we stretch to put the endpoints at $(\pm 2, 0)$, then bend to form a semicircle of radius two, then collapse horizontally onto the unit circle.



Using Remark 30.4 we see that ϕ gives the required CW structure. Note that the cell E_n is all of S^n, so this example illustrates the fact that the maps $\phi_a\colon B^{d(a)} \to E_a$ need not be homeomorphisms.

Example 30.8. [eg-CW-sphere-a]
Now instead put

$$\begin{aligned} A &= \{0,\dots,n\} \times \{1,-1\} \\ d(i,\epsilon) &= i \\ X_{i,\epsilon} &= \{x \in S^n \ : \ \epsilon x_i > 0,\ x_j = 0 \text{ for } j > i\}. \end{aligned}$$

Note that every point in S^n must have at least one nonzero coordinate, and by considering the position and sign of the last nonzero coordinate we see that S^n is the disjoint union of the sets $X_{i,\epsilon}$. Next, we can define

$$\phi_{i,\epsilon}\colon B_2^i \to S^n$$

by

$$\phi_{i,\epsilon}(x_0,\dots,x_{i-1}) = (x_0,\dots,x_{i-1},\epsilon\sqrt{1 - \textstyle\sum_{j<i} x_j^2}, 0, \dots, 0).$$

It is easy to see that this gives a homeomorphism $OB_2^i \to X_{i,\epsilon}$ and also that

$$\begin{aligned} \phi_{i,\epsilon}(B_2^i) &= \{x \in S^n \ : \ \epsilon x_i \geq 0,\ x_j = 0 \text{ for } j > i\} \\ &= \overline{X_{i,\epsilon}} = X_{i,\epsilon} \cup \bigcup_{j<i, \delta \in \{1,-1\}} X_{j,\delta}. \end{aligned}$$

We can combine these maps $\phi_{i,\epsilon}$ to give a map $\phi\colon \coprod_{(i,\epsilon)\in A} B_2^i \to S^n$, and using Remark 30.4 we see that this gives a CW structure.

Example 30.9. [eg-CW-RP]
Consider the space $\mathbb{R}P^n$. As in Examples 5.24 and 5.69, we have a quotient map $q\colon S^n \to \mathbb{R}P^n$ with $q(x) = q(y)$ iff $x = \pm y$. Put $A = \{0, \dots, n\}$ and $d(i) = i$. Define $\phi_i\colon B_2^i \to \mathbb{R}P^n$ by

$$\phi_i(x_0, \dots, x_{i-1}) = q(x_0, \dots, x_{i-1}, \sqrt{1 - \sum_{j<i} x_j^2}, 0, \dots, 0).$$

Equivalently, this can be written in the notation of Example 30.8 as $\phi_i = q \circ \phi_{i,+1}$. We can combine these maps to give a map $\phi\colon \coprod_{i \leq n} B_2^i \to \mathbb{R}P^n$. Note that if $x \in \mathbb{R}P^n$ then there exists $y \in S^n$ with $q(y) = x$, and if we insist that the last nonzero coordinate in y is positive, then y is unique. Using this, we see that ϕ gives a bijection $\coprod_i OB_2^i \to \mathbb{R}P^n$. We also see that the cell $E_i = \phi(\{i\} \times B_2^i)$ is just the evident copy of $\mathbb{R}P^i$ in $\mathbb{R}P^n$, which is the same as $\phi(\{i+1\} \times \partial B_2^{i+1})$. We therefore have a CW structure on $\mathbb{R}P^n$.

We next want to construct a CW structure on the orthogonal group $O(n)$ and various related spaces. This is somewhat more elaborate, so we will do it in stages.

Proposition 30.10. *Let V be a finite-dimensional real vector space equipped with an inner product, and put*

$$\mathbb{R}P(V) = \{A \in \mathrm{End}(V) \ :\ A^2 = A^T = A,\ \mathrm{trace}(A) = 1\}$$
$$O(V) = \{B \in \mathrm{End}(V) \ :\ B^T B = I\}.$$

Then there is a natural embedding $\rho\colon \mathbb{R}P(V) \to O(V)$ given by $\rho(A) = I - 2A$.

Proof. **Write this** □

Fix up and merge the next two examples

Example 30.11. [eg-CW-On]
Consider the space $O(n+1)$ of $(n+1)\times(n+1)$ real matrices Q with $Q^TQ = I$. We will construct a CW structure on $O(n+1)$. First, suppose we have $A \in \mathbb{R}P^n \subseteq M_{n+1}(\mathbb{R})$ and we put $\rho(A) = I - 2A$. As $A^T = A^2 = A$ we have

$$\rho(A)^T\rho(A) = (I - 2A)^2 = I - 4A + 4A^2 = I,$$

so $\rho\colon \mathbb{R}P^n \to O(n+1)$. We can interpret this more geometrically in terms of the map $q\colon S^n \to \mathbb{R}P^n$: if we put $Q = \rho(Q(u))$ then $Qu = -u$ but $Qv = v$ for all $v \in u^\perp$, so Q is just the reflection across the hyperplane $v^\perp$.

Next, let $\mu_r\colon O(n+1)^r \to O(n+1)$ be the multiplication map

$$\mu_r(Q_1, \dots, Q_r) = Q_1 Q_2 \cdots Q_r.$$

Put $N = \{0, 1, \dots, n\}$ and $A = \{$ subsets of $N\}$, and define $d\colon A \to \mathbb{N}$ by $d(a) = \sum_{i\in a} i$. For any $a \in A$ we can write $a = \{i_1, \dots, i_r\}$ for some $r \geq 0$ and some sequence $i_1 > \cdots > i_r$. We put $B_a = \prod_t B_2^{i_t}$, and note that this is homeomorphic in an obvious way to $B^{d(a)}$. Let $\phi_i\colon B_2^i \to \mathbb{R}P^i \subseteq \mathbb{R}P^n$ be as in Example 30.9, and define $\psi_a\colon B_a \to O(n+1)$ to be the composite

$$\prod_t B_2^{i_t} \xrightarrow{\prod_t \phi_{i_t}} (\mathbb{R}P^n)^r \xrightarrow{\rho^r} O(n+1)^r \xrightarrow{\mu_r} O(n+1).$$

We claim that the resulting map $\psi\colon \coprod_a B_a \to O(n+1)$ is a CW structure.

Example 30.12. Consider the space $O(n+1)$ of $(n+1)\times(n+1)$ real matrices Q with $Q^TQ = I$. We will construct a CW structure on $O(n+1)$. First, for $m \leq n$ and $Q' \in O(m)$ we have a matrix

$$Q = \left[\begin{array}{c|c} Q' & 0 \\ \hline 0 & I \end{array}\right] \in O(n+1),$$

and this construction identifies $O(m)$ with the subgroup

$$\{Q \in O(n+1) \ :\ Qe_i = e_i \text{ for all } i > m\} \leq O(n+1).$$

(Here we are numbering the standard basis of $\mathbb{R}^{n+1}$ as $e_0, \dots, e_n$.) For any $x \in S^n \setminus \{e_n\}$ let H_x be the hyperplane orthogonal to $x - e_n$, and let $r(x) \in O(n+1)$ be the reflection across H_x. We can thus define

$$s\colon (S^n \setminus \{e_n\}) \times O(n) \to O(n+1)$$

by $s(x,Q) = r(x)Q$. Here $Qe_n = e_n$ so $s(x,Q)e_n = r(x)e_n = x$, and as $x \neq e_n$ we have $s(x,Q) \notin O(n)$. For an arbitrary $R \in O(n+1) \setminus O(n)$ we can put $x = Re_n \in S^n \setminus \{e_n\}$ and $Q = r(x)^{-1}R$ and we find that $Q \in O(n+1)$ and $Qe_n = e_n$, so $Q \in O(n)$, and $s(x,Q) = R$. Using this we see that s gives a bijection $(S^n \setminus \{e_n\}) \times O(n) \to O(n+1) \setminus O(n)$. Note also that $S^n \setminus \{e_n\}$ is homeomorphic to $\mathbb{R}^n$ by stereographic projection, so we have decomposed $O(n+1)$ as the union of $O(n)$ with a space homeomorphic to $\mathbb{R}^n \times O(n)$. We may assume by induction that we have a pre-CW structure decomposing $O(n)$ as a union of subspaces homeomorphic to $\mathbb{R}^d$ for various d, and it follows that we have a decomposition of $O(n+1)$ as a union of spaces homeomorphic to $\mathbb{R}^d$ or $\mathbb{R}^{n+d}$. More specifically, we claim that for each subset $J \subseteq \{0, \dots, n\}$ the decomposition of $O(n+1)$ has a cell of dimension $d(J) = \sum_{j \in J} j$; this can be checked by induction on n.

We next explain how to improve this to a genuine CW structure. First, for $x \in B^k$ we define $q(x) \in M_k(\mathbb{R})$ by $q(x)_{ij} = x_i x_j$. Equivalently, if we view x as a column vector then $q(x) = xx^T$. We find that $q(x)^T = q(x)$ and $q(x)^2 = \|x\|^2 q(x)$ and $q(x)y = \langle x, y \rangle x$. Next, for $x \in B^k_2$ we can define a matrix $p_k(x) \in M_{k+1}(\mathbb{R})$ by

$$p_k(x) = \left[\begin{array}{c|c} I - 2q(x) & 2\sqrt{1 - \|x\|^2}\; x \\ \hline 2\sqrt{1 - \|x\|^2}\; x^T & 2\|x\|^2 - 1 \end{array} \right]$$

It is clear that $p_k(x)^T = p_k(x)$. We claim that $p_k(x)^2 = 1$. Indeed, the top left entry in $p_k(x)^2$ is

$$(I - 2q(x))^2 + 2\sqrt{1 - \|x\|^2}x.2\sqrt{1 - \|x\|^2}x^T.$$

Using $q(x)^2 = \|x\|^2 q(x)$ and $xx^T = q(x)$ this simplifies down to I. The top right entry is

$$2\sqrt{1 - \|x\|^2}x^T.2\sqrt{1 - \|x\|^2}x + (2\|x\|^2 - 1)^2.$$

Here $x^T x = \|x\|^2$ and everything cancels down to zero. The bottom left entry is zero by symmetry, and the bottom right entry simplifies to 1 in a similar way. We therefore see that $p_k(x) \in O(k+1)$, which we regard in the usual way as a subgroup of $O(n+1)$. Next, any subset $J \subseteq \{1, \dots, n\}$ can be written as $J = \{j_1, \dots, j_r\}$ with $j_1 < \dots < j_r$. We put $B[J] = \prod_{j \in J} B^j_2 \simeq B^{d(J)}$ and we define

$$m_J\colon \prod_{j \in J} B^j \to O(n+1)$$

by

$$m_J(x_{j_1}, \dots, x_{j_r}) = p_{j_r}(x_{j_r}) \cdots p_{j_1}(x_{j_1}).$$

These maps fit together to give a map $m\colon \coprod_J B[J] \to O(n+1)$. This can be combined with the standard homeomorphisms $B^j \to B^j_2$ to give a map $\phi\colon \coprod_J B^{d(J)} \to O(n+1)$. We claim that this is a CW structure. To prove this, define $f_k\colon B^k_2 \to S^k$ by $f_k(x) = p_k(x)e_k$. By inspecting the formulae, we see that this is essentially the same map as in Example 30.7, so it is surjective with $f_k^{-1}\{e_k\} = S^{k-1}$, and it induces a homeomorphism $\mathring{B}^k_2 \to S^k \setminus \{e_k\}$. It follows that the map $(x,Q) \mapsto p_k(x)Q$ gives a homeomorphism $\mathring{B}^k_2 \times O(k) \to O(k+1) \setminus O(k)$, and an inductive argument based on this shows that ϕ gives a bijection $\coprod_J \mathring{B}^{d(J)} \to O(n+1)$. In particular, this means that $\phi\colon \coprod_J B^{d(J)} \to O(n+1)$ is surjective. It follows by Remark 30.4 that ϕ is a quotient map, so the only thing left to check is that the cell boundaries work correctly as in axiom CW2. **It's not clear that this is actually correct.**

Example 30.13. **Explain Schubert cells**

Example 30.14. **Mention Morse theory**

Example 30.15. **Simplicial complexes are CW complexes. Also simplicial sets, but that should probably come later.**

Example 30.16. **Algebraic varieties, o-minimal stuff?**

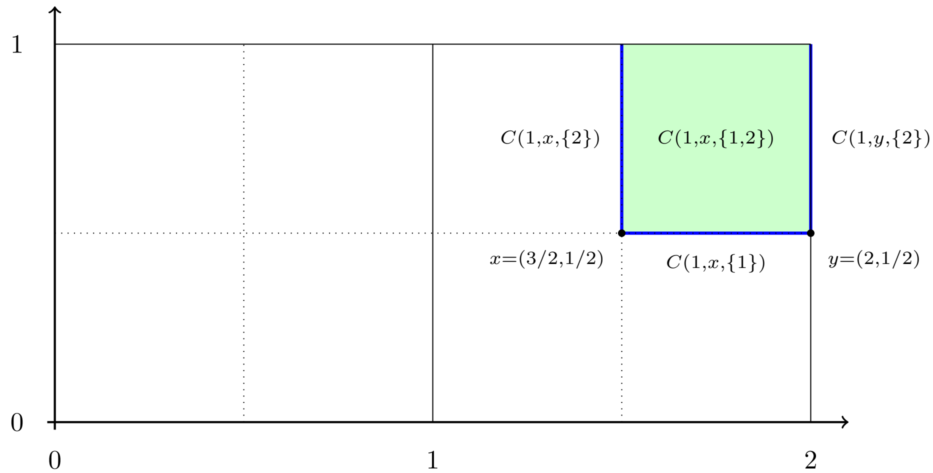
30.2. Open subsets of $\mathbb{R}^n$. We now construct CW structures on all open subsets of $\mathbb{R}^n$.

Definition 30.17. [`defn-enr-cubes`]
Suppose we have an integer $k \in \mathbb{N}$, a point $a \in 2^{-k}\mathbb{Z}^n$ and a subset $J \subseteq \{1, \dots, n\}$. We put

$$C(k,a,J) = \{x \in \mathbb{R}^n \ : \ a_i < x_i < a_i + 2^{-k} \text{ for } i \in J \text{ and } x_i = a_i \text{ for } i \notin J\}$$
$$\mathcal{C}(k) = \{C(k,a,J) \ : \ a \in 2^{-k}\mathbb{Z}^n,\ J \subseteq \{1, \dots, n\}\}.$$

In the case where $J = \{1, \dots, n\}$ we will omit it and write

$$C(k,a) = C(k,a,\{1,\dots,n\}) = \prod_{i=1}^{n} (a_i, a_i + 2^{-k}).$$



Sets of the form $C(k,a,J)$ are called *dyadic cubes*; those of the form $C(k,a)$ are *full dyadic cubes*. We say that the *base* is a, and *set of directions* is J.

Remark 30.18. [`rem-cube-diameter`]
Everywhere in this section we will use the metric

$$d(x,y) = d_\infty(x,y) = \max(|x_1 - y_1|, \dots, |x_n - y_n|)$$

on $\mathbb{R}^n$. With respect to this metric the diameter of $C(k,a,J)$ is 2^{-k} except in the case $J = \emptyset$, when we have $C(k,a,\emptyset) = \{a\}$ and the diameter is zero.

Lemma 30.19. [`lem-cubes-cover`]
For fixed k, the set $\mathbb{R}^n$ is the disjoint union of the cubes in $\mathcal{C}(k)$.

Proof. Consider a point $x \in \mathbb{R}^n$. It is clear that for each i there is a unique element $a_i \in 2^{-k}\mathbb{Z}$ such that $a_i \leq x_i < a_i + 2^{-k}$. If we put $J = \{i \ : \ x_i = a_i\}$, we find that $x \in C(k,a,J)$, and that this is the unique cube in $\mathcal{C}(k)$ containing x. □

Lemma 30.20. [`lem-nesting`]
Consider a pair of dyadic cubes, say $A = C(j,a,J)$ and $B = C(k,b,K)$. Then precisely one of the following holds:

(a) $A \cap B = \emptyset$
(b) $A = B$
(c) $A \subset B$ *and* $j > k$ *and* $\operatorname{diam}(A) < \operatorname{diam}(B)$ *and* $J \subseteq K$
(d) $A \supset B$ *and* $j < k$ *and* $\operatorname{diam}(A) > \operatorname{diam}(B)$ *and* $J \supseteq K$.

Proof. We may assume without loss that $A \cap B \neq \emptyset$ and $j \geq k$; we must then prove that case (b) or (c) holds. Choose $x \in A \cap B$. For $i \notin K$ we have $x_i = b_i \in 2^{-k}\mathbb{Z} \subseteq 2^{-j}\mathbb{Z}$, but for $i \in J$ we then have $x_i \in (a_i, a_i + 2^{-j})$ so $x_i \notin 2^{-j}\mathbb{Z}$. It follows that $J \cap K^c = \emptyset$, or in other words $J \subseteq K$. Thus:

(1) If $i \in J$ we have $x_i \in (a_i, a_i + 2^{-j})$ and also $x_i \in (b_i, b_i + 2^{-k})$. As these are dyadic intervals and $j \geq k$ it follows easily that $(a_i, a_i + 2^{-j}) \subseteq (b_i, b_i + 2^{-k})$, and that $a_i = b_i$ if $j = k$.
(2) If $i \in K \setminus J$ we have $x_i = b_i \in 2^{-k}\mathbb{Z}$ and also $a_i < x_i < a_i + 2^{-j}$. This can only happen if $j > k$.
(3) If $i \notin K$ then also $i \notin J$, so $a_i = x_i = b_i$.

It is clear from this that $A \subseteq B$. For the inclusion to be strict we must have a strict inclusion in (1) or case (2) must arise, and in either case $j > k$. Moreover, as the inclusion is strict B cannot just be a single point, so $\operatorname{diam}(B) = 2^{-k}$ whereas $\operatorname{diam}(A) \in \{0, 2^{-j}\}$, so $\operatorname{diam}(A) < \operatorname{diam}(B)$, so (c) holds. If the inclusion is not strict then (b) holds, as required. □

Corollary 30.21. [`cor-nesting`]
Let $\mathcal{A}$ be any family of dyadic cubes. Let X be the union of all the cubes in $\mathcal{A}$, and let $\mathcal{M}$ be the set of those cubes that are maximal in $\mathcal{A}$ with respect to inclusion. Then X is the disjoint union of the cubes in $\mathcal{M}$.

Proof. First, it is clear from Lemma 30.20 that if two different dyadic cubes overlap, then one is contained in the other, so at most one of them can be maximal in $\mathcal{A}$. Thus, the cubes in $\mathcal{M}$ are disjoint. Now consider a point $x \in X$, so the set $\mathcal{A}_x = \{A \in \mathcal{A} \,:\, x \in A\}$ is nonempty. The set of diameters of dyadic cubes is $\{0\} \cup \{2^{-k} \,:\, k \geq 0\}$, and every subset of this set has a largest element. Thus, we can choose a cube $A \in \mathcal{A}_x$ of maximal diameter. Suppose we have $B \in \mathcal{A}$ with $A \subseteq B$; then clearly $x \in B$, so $B \in \mathcal{A}_x$, so $\operatorname{diam}(B) = \operatorname{diam}(A)$. By inspecting the four possibilities in Lemma 30.20, we see that $B = A$. This means that $A \in \mathcal{M}$. It follows that X is the union of the cubes in $\mathcal{M}$, as claimed. □

Definition 30.22. Given a dyadic cube $A = C(k, a, J)$, we put

$$\mathcal{N}(A) = \{C(k, b, K) \,:\, K \supseteq J,\ b_i = a_i \text{ for } i \in J \text{ or } i \notin K,\ b_i \in \{a_i - 2^{-k}, a_i\} \text{ for } i \in K \setminus J\}$$
$$\nu(A) = \{x \in \mathbb{R}^n \,:\, a_i < x_i < a_i + 2^{-k} \text{ for } i \in J,\ a_i - 2^{-k} < x_i < a_i + 2^{-k} \text{ for } i \notin J\}.$$

Lemma 30.23. [`lem-nu-cubes`]
The set $\nu(A)$ is open in $\mathbb{R}^n$ and has diameter at most 2^{1-k}. It is the disjoint union of the cubes in $\mathcal{N}(A)$. Moreover, if $B \in \mathcal{N}(A)$ then $\mathcal{N}(B) \subseteq \mathcal{N}(A)$ and $\nu(B) \subseteq \nu(A)$.

Proof. It is clear that $\nu(A)$ is open, with diameter 2^{-k} if $J = \{1, \dots, n\}$, or 2^{1-k} otherwise. Consider a point $x \in \nu(A)$. Put $K = \{i \,:\, x_i \neq a_i\}$, so $J \subseteq K$. If $a_i - 2^{-k} < x_i < a_i$ then necessarily $i \in K \setminus J$ and we put $b_i = a_i - 2^{-k}$; otherwise we put $b_i = a_i$. We then find that $x \in C(k, b, K) \in \mathcal{N}(A)$. Thus, $\nu(A)$ is contained in the union of the cubes in $\mathcal{N}(A)$. The reverse inclusion is clear from the definitions.

Now suppose that $B = C(k, b, K) \in \mathcal{N}(A)$ and $C = C(k, c, L) \in \mathcal{N}(B)$. We then have $L \supseteq K \supseteq J$.

- If $i \in J$ then i is also in K so $a_i = b_i = c_i$.
- If $i \in K \setminus J$ then $b_i \in \{a_i, a_i - 2^{-k}\}$ and $b_i = c_i$ so so $c_i \in \{a_i, a_i - 2^{-k}\}$.
- If $i \in L \setminus K$ then $a_i = b_i$ and $c_i \in \{b_i, b_i - 2^{-k}\}$ so $c_i \in \{a_i, a_i - 2^{-k}\}$.
- If $i \notin L$ then also $i \notin K$ so $a_i = b_i = c_i$.

It follows that $C \in \mathcal{N}(A)$. This means that $\mathcal{N}(B) \subseteq \mathcal{N}(A)$ and so $\nu(B) \subseteq \nu(A)$. □

Definition 30.24. For any open set $U \subseteq \mathbb{R}^n$ we put

$$\mathcal{A}(U) = \{ \text{ dyadic cubes } A \,:\, \overline{\nu(A)} \subseteq U\}$$
$$\mathcal{M}(U) = \{ \text{ maximal elements in } \mathcal{A}(U)\}.$$

Proposition 30.25. [`prop-open-CW`]
Let U be an open subset of $\mathbb{R}^n$. Then U is the disjoint union of the cubes in $\mathcal{M}(U)$. Moreover, for any point $x \in U$ there is an open neighbourhood N such that the set $\{A \in \mathcal{M}(U) \,:\, \overline{A} \cap N \neq \emptyset\}$ is finite.

Proof. Consider a point $x \in U$. As U is open, there exists $\epsilon > 0$ such that $OB_\epsilon(x) \subseteq U$. Choose k such that $2^{1-k} < \epsilon$. By Lemma 30.19 we see that there is a unique cube $A \in \mathcal{C}(k)$ such that $x \in A$. Put $N = \nu(A)$, which is an open neighbourhood of x of diameter at most $2^{1-k} < \epsilon$. It follows that $\overline{\nu(A)} \subseteq U$, so $A \in \mathcal{A}(U)$. More generally, for $B \in \mathcal{N}(A)$ we have $\nu(B) \subseteq \nu(A)$ so $\overline{\nu(B)} \subseteq U$ so $B \in \mathcal{A}(U)$; in other words, we have $\mathcal{N}(A) \subseteq \mathcal{A}(U)$. Next, for each $B \in \mathcal{N}(A)$ we can choose a cube $\phi(B) \in \mathcal{A}(U)$ of maximal diameter such that $B \subseteq \phi(B)$. Using Lemma 30.20 again we see that $\phi(B) \in \mathcal{M}(U)$. In particular, we have $x \in A \subseteq \phi(A) \in \mathcal{M}(U)$, but x was an arbitrary point of U, so the sets in $\mathcal{M}(U)$ cover U. They are also disjoint by Corollary 30.21.

Now consider a cube $C \in \mathcal{M}(U)$ such that $\overline{C}$ meets N. As N is open this means that C itself must meet N, but N is the disjoint union of the sets in $\mathcal{N}(A)$, so there is some cube $B \in \mathcal{N}(A)$ such that C meets B. This means that C also meets the larger set $\phi(B)$. Now C and $\phi(B)$ are both in $\mathcal{M}(U)$ and the sets in $\mathcal{M}(U)$ are disjoint, so we must have $C = \phi(B)$. Thus, there are only finitely many possibilities for C. □

30.3. Topological properties. Consider a space X with CW structure $\phi\colon \coprod_{a\in A} B^{d(a)} \to X$.

Lemma 30.26. [lem-cell-closure]
For all $a \in A$, the cell E_a is the closure of OE_a.

Proof. As $\phi_a\colon B^{d(a)} \to X$ is continuous and $B^{d(a)}$ is compact, the weak Hausdorff condition means that $\phi_a(B^{d(a)}) = E_a$ is closed. It follows that $\overline{OE_a} \subseteq E_a$. On the other hand, the set $\phi_a^{-1}(\overline{OE_a})$ is closed in $B^{d(a)}$ by continuity, and it contains the dense set $\mathring{B}^{d(a)}$, so it must be all of $B^{d(a)}$. This means that $E_a = \phi(B^{d(a)}) \subseteq \overline{OE_a}$. □

Lemma 30.27. [lem-cell-intersection]
If $a, b \in A$ with $d(b) < d(a)$ then $OE_a \cap E_b = \emptyset$.

Proof. By remark 30.2, we have $E_b \subseteq \bigcup_{c\in A_b^+} OE_c$, so it will suffice to prove that $OE_c \cap OE_a = \emptyset$ for $c \in A_b^+$. By axiom CW1 we have $d(c) \le d(b) < d(a)$, so $c \ne a$, so $OE_c \cap OE_a = \emptyset$ by axiom CW0. □

Lemma 30.28. [lem-CW-closure-test]
Suppose that $Y \subseteq X$, and that $\phi_a^{-1}(Y)$ is closed in $B^{d(a)}$ for all a such that $Y \cap OE_a \ne \emptyset$. Then Y is closed in X.

Proof. Put $T_a = \phi_a^{-1}(Y) \subseteq B^{d(a)}$. As ϕ is a quotient map, it will suffice to show that T_a is closed for all a, which we will do by induction on $d(a)$. If $d(a) = 0$ then $B^{d(a)}$ is just a point so all subsets are closed. Suppose instead that $d(a) > 0$. If T_a meets $\mathring{B}^{d(a)}$ then Y meets OE_a, so T_a is closed by hypothesis. Suppose instead that $T_a \cap \mathring{B}^{d(a)} = \emptyset$, so

$$Y \cap E_a \subseteq DE_a \subseteq \bigcup_{b\in A_a} E_b.$$

Put $Z = \bigcup_{b\in A_a} \phi_b(T_b)$. By induction we may assume that each T_b is closed in $B^{d(b)}$ and so is compact, so Z is closed in X. Using $Y \cap E_a \subseteq \bigcup_b E_b$ we also see that $Y \cap E_a = Z \cap E_a$, so $Y \cap E_a$ is closed in X, so the set $T_a = \phi_a^{-1}(Y \cap E_a)$ is closed in $B^{d(a)}$ as required. □

Corollary 30.29. [cor-CW-closure-test]
Suppose that $Y \subseteq X$, and that $Y \cap E_a$ is closed in X for all a such that $Y \cap OE_a \ne \emptyset$. Then Y is closed in X.

Proof. As $\phi_a^{-1}(Y) = \phi_a^{-}(Y \cap E_a)$, this follows from the lemma. □

Lemma 30.30. [lem-open-cell]
Let X be a CW complex, with CW structure $\phi\colon \coprod_{a\in A} B^{d(a)} \to X$. Consider the subset $OE_a = \phi_a(\mathring{B}^{d(a)}) \subseteq X$ with the subspace topology. Then the map $\phi_a\colon \mathring{B}^{d(a)} \to OE_a$ is a homeomorphism.

Proof. It is clear from the definitions that $\phi_a\colon \mathring{B}^{d(a)} \to OE_a$ is a continuous bijection, so it will suffice to check that this map is closed. Consider a subset $F \subseteq \mathring{B}^{d(a)}$ that is closed in $\mathring{B}^{d(a)}$. Put $F_1 = F \cup \partial(B^{d(a)})$, so F_1 is closed in $B^{d(a)}$ and thus is compact. It follows that $\phi_a(F_1)$ is closed in X. Now $\phi_a(F_1) = \phi_a(F) \cup DE_a$, and $DE_a \cap OE_a = \emptyset$ (by Remark 30.2) so $\phi_a(F) = \phi_a(F_1) \cap OE_a$, and this is closed in OE_a as required. □

Definition 30.31. [defn-CW-subcomplex]
A *subcomplex* of a CW complex (X, A, d, ϕ) is a system (X', A', d', ϕ') such that

SCW0: X' is a subspace of X, and A' is a subset of A.
SCW1: $X' = \bigcup_{a \in A'} E_a$, and $A' = \{a \in A \;:\; OE_a \cap X' \neq \emptyset\}$.
SCW2: d' is the restriction of d to A', and ϕ' is the restriction of ϕ to $\coprod_{a \in A'} B^{d(a)}$.

A *finite subcomplex* means a subcomplex (X', A', d', ϕ') where the set A' is finite.

REMARK 30.32. [`rem-gives-a-subcomplex`]
If we are given a subspace $X' \subseteq X$, it is clear that there is at most one way to define A', d' and ϕ' so that (X', A', d', ϕ') is a subcomplex. Similarly, if we are given A' then there is at most one way to define X', d' and ϕ'.

PROPOSITION 30.33. [`prop-CW-subcomplex`]
Let (X', A', d', ϕ') be a subcomplex of (X, A, d, ϕ). Then

(a) *As a set, X' is the disjoint union of the sets OE_a for $a \in A'$.*
(b) *If $a \in A'$ then $A_a \subseteq A'$.*
(c) *X' is closed in X.*
(d) *The map $\phi' \colon \coprod_{a \in A'} B^{d(a)} \to X'$ gives a CW structure on X'.*

PROOF.

(a) As X is the disjoint union of the sets OE_a, we see that X' is the disjoint union of the sets $X' \cap OE_a$. If $X' \cap OE_a$ is empty then we can ignore it. It is is nonempty then (by SCW1) we have $a \in A'$ and $OE_a \subseteq E_a \subseteq X'$, so $X' \cap OE_a = OE_a$. The claim follows.
(b) Suppose that $a \in A'$ and $b \in A_a$, so OE_b meets E_a. By SCW1 we have $E_a \subseteq X'$, so OE_b meets X', so $b \in A'$ by SCW1 again.
(c) By Corollary 30.29, it is enough to show that $X' \cap E_a$ is closed whenever $X' \cap OE_a \neq \emptyset$, or equivalently $a \in A'$. In that case we have $E_a \subseteq X'$ so $X' \cap E_a = E_a$, which is closed by Lemma 30.26.
(d) We first claim that ϕ' is a quotient map. Equivalently, we claim that if $Y \subseteq X'$ and $\phi_a^{-1}(Y)$ is closed in $B^{d(a)}$ for all $a \in A'$, then Y is closed in X (and therefore also in X'). This follows easily from Lemma 30.28, because if $OE_a \cap Y \neq \emptyset$ then necessarily $a \in A'$. Axiom CW0 is now clear from (a), and CW1 for X' follows from CW1 for X.

□

PROPOSITION 30.34. [`prop-gives-a-subcomplex`]

(a) *A subset $A' \subseteq A$ gives a subcomplex iff we have $A_a \subseteq A'$ for all $a \in A'$.*
(b) *A subspace $X' \subseteq X$ gives a subcomplex iff we have $E_a \subseteq X'$ for all a such that $OE_a \cap X' \neq \emptyset$.*

PROOF.

(a) One direction is given by part (b) of Proposition 30.33. For the converse, suppose that $A' \subseteq A$ and that $A_a \subseteq A'$ for all $a \in A'$. Put $X' = \bigcup_{a \in A'} E_a$. It is clear that if $a \in A'$ then OE_a meets X'. Conversely, suppose that $a \in A$ and that OE_a meets X'. This means that OE_a must meet E_b for some $b \in A'$, but Remark 30.2 tells us that $E_b \subseteq \bigcup_{c \in A_b^+} OE_c$, so OE_a must meet OE_c for some $c \in A_b^+$. By axiom CW0 this gives $a = c$ so $a \in A_b^+$ so either $a = b \in A'$ or $a \in A_b \subseteq A'$; either way, we have $a \in A'$. We now define d' and ϕ' in accordance with SCW2 and we find that all axioms are satisfied.
(b) First suppose that X' gives a subcomplex, so there exist A', d' and ϕ' such that SCW0, SCW1 and SCW2 are satisfied. If $OE_a \cap X' \neq \emptyset$ then SCW1 gives $a \in A'$ and then $E_a \subseteq X'$, as required.

    Conversely, suppose we have a subspace $X' \subseteq X$ with the property that $E_a \subseteq X'$ whenever $OE_a \cap X' \neq \emptyset$. If we put $A' = \{a \;:\; OE_a \cap X' \neq \emptyset\}$ and $X'' = \bigcup_{a \in A} E_a$. Our condition on X' ensures that $X'' \subseteq X'$. On the other hand, if $x \in X'$ then certainly $x \in OE_a$ for some $a \in A$, and $x \in OE_a \cap X'$ so $OE_a \cap X' \neq \emptyset$ so $a \in A'$, so $x \in E_a \subseteq X''$, so $x \in X''$. This proves that $X' \subseteq X''$, so $X'' = X'$. This means that SCW1 is satisfied. We can now define d' and ϕ' in accordance with SCW2 and we see that we have a subcomplex, as required.

□

Corollary 30.35. [`cor-subcomplex-lattice`]
The intersection of any family of subcomplexes is a subcomplex. Similarly, the union of any family of subcomplexes is also a subcomplex.

Proof. This follows easily from criterion (b) in Proposition 30.33. □

Now suppose we have a subspace $Y \subseteq X$. We can let $\mathcal{F}$ be the family of all subcomplexes that contain Y, and then let X' be the intersection of that family. Then X' is again a subcomplex, and it clearly contains Y, so X' is the smallest element in $\mathcal{F}$. This validates the following definition.

Definition 30.36. [`defn-carrier`]
For any subset $Y \subseteq X$, the *carrier* of Y is the smallest subcomplex of X that contains Y.

Lemma 30.37. [`lem-finite-carrier`]
For each $a \in A$, the carrier of E_a is a finite subcomplex of X.

Proof. We argue by induction on $d(a)$. If $d(a) = 0$ then $E_a = OE_a$ and this is a single point. It follows directly from the definitions that E_a is by itself a subcomplex, so it is its own carrier and the claim holds. If $d(a) > 0$ we put

$$X' = E_a \cup \bigcup_{b \in A_a} \operatorname{carrier}(E_b).$$

Using Proposition 30.34 we see that this is a subcomplex of X, and it clearly contains E_a, so it contains the carrier of E_a. For $b \in A_a$ we have $d(b) < d(a)$ so by induction the carrier of E_b is finite, and A_a is finite, so X' is finite, so the carrier of E_a is finite. (It is not hard to see that X' is actually equal to the carrier of E_a, but we do not need that.) □

Proposition 30.38. [`prop-finite-carrier`]
If $Y \subseteq X$ is compact, then the carrier of Y is finite.

Proof. Put $A' = \{a \in A \;:\; Y \cap OE_a \neq \emptyset\}$, and for each $a \in A'$ choose a point $y_a \in Y \cap OE_a$. Put $Z = \{y_a \;:\; a \in A'\} \subseteq Y$. As the sets OE_a (for $a \in A$) are all disjoint, we see that $|Z \cap OE_a| \leq 1$, and thus that the intersection of Z with any finite subcomplex is finite. Lemma 30.37 tells us that E_a is contained in a finite subcomplex, so $E_a \cap Z$ is finite, so $E_a \cap Z$ is closed in X. It follows by Corollary 30.29 that Z is closed in X, and therefore also in Y. As Y is compact, the same is true of Z. More generally, the same argument shows that every subset of Z is closed, so Z is discrete as well as compact, so it must be finite. On the other hand, Z bijects with A' by construction, so A' is finite. Now put $X' = \bigcup_{a \in A'} \operatorname{carrier}(E_a)$. This is a finite subcomplex by Lemma 30.37. For any $y \in Y$ we have $y \in OE_a$ for some $a \in A$, and then $a \in A'$ by the definition of A', so $E_a \subseteq X'$, so $y \in X'$. It follows that X' is a finite subcomplex containing Y, so the carrier of Y is finite. □

Corollary 30.39. [`cor-finite-compact`]
A subcomplex $X' \subseteq X$ is a finite subcomplex iff it is compact.

Proof. Let A' be the indexing set for X', so $X' = \bigcup_{a \in A'} E_a$. The sets E_a are compact, so if A' is finite then X' is compact. Conversely, if X' is compact then the carrier of X' is finite, but X' is a subcomplex so it is its own carrier. □

Definition 30.40. [`lem-skeleton`]
For any $k \geq 0$, the *k-skeleton* of X is the set

$$\operatorname{skel}_k(X) = \bigcup_{d(a) \leq k} E_a \subseteq X.$$

Using part (a) of Proposition 30.34, we see that this is a subcomplex of X.

Proposition 30.41. [`prop-skeleta`]

The space $\mathrm{skel}_0(X)$ *is discrete, and for any* $k > 0$ *we have a pushout square*

$$\begin{tikzcd} \coprod_{d(a)=k} \partial(B^k) \arrow[r, tail] \arrow[d, "\phi"'] & \coprod_{d(a)=k} B^k \arrow[d, "\phi"] \\ \mathrm{skel}_{k-1}(X) \arrow[r, tail] & \mathrm{skel}_k(X). \end{tikzcd}$$

Moreover, X *is the colimit of the sequence*

$$\mathrm{skel}_0(X) \rightarrowtail \mathrm{skel}_1(X) \rightarrowtail \mathrm{skel}_2(X) \rightarrowtail \mathrm{skel}_3(X) \rightarrowtail \mathrm{skel}_4(X) \rightarrowtail \cdots$$

Proof. First, we see using Corollary 30.29 that every subset of $\mathrm{skel}_0(X)$ is closed, so $\mathrm{skel}_0(X)$ is discrete as claimed. Now suppose that $k > 0$, and form a pushout square

$$\begin{tikzcd} \coprod_{d(a)=k} \partial(B^k) \arrow[r, tail, "i"] \arrow[d, "\phi"'] & \coprod_{d(a)=k} B^k \arrow[d, "\psi"] \\ \mathrm{skel}_{k-1}(X) \arrow[r, "j"'] & P \end{tikzcd}$$

in the category of CGWH spaces. It is clear that i is a closed inclusion, so Proposition 23.47 tells us that j is also a closed inclusion and that the square is also a pushout in the category of sets. In that category we have $B^k = \partial(B^k) \amalg \mathring{B}^k$, and it follows that

$$P = \mathrm{skel}_{k-1}(X) \amalg \coprod_{d(a)=k} \mathring{B}^k$$

as sets, and thus that the evident comparison map $\xi\colon P \to \mathrm{skel}_k(X)$ is a continuous bijection. Our claim is that ξ is a homeomorphism, and it will suffice to show that it is a quotient map. Consider a subset $Y \subseteq \mathrm{skel}_k(X)$ such that $\xi^{-1}(Y)$ is closed in P. It follows that the set $Y \cap \mathrm{skel}_{k-1}(X) = j^{-1}\xi^{-1}(Y)$ is closed in $\mathrm{skel}_{k-1}(X)$ and thus also in X. It also follows that $\psi^{-1}\xi^{-1}(Y)$ is closed, or equivalently that $\phi_a^{-1}(Y)$ is closed in B^k for all $a \in A$ with $d(a) = k$. Given this, we can use Lemma 30.28 to see that Y is closed in X, so ξ is a quotient map as claimed.

Next, it is clear from axiom CW0 that X is the union of the subspaces $\mathrm{skel}_k(X)$, and using Lemma 30.28 again we see that a subset $Y \subseteq X$ is closed if and only if $Y \cap \mathrm{skel}_k(X)$ is closed in $\mathrm{skel}_k(X)$ for all k. It follows that a map $f\colon X \to Z$ is continuous iff $f|_{\mathrm{skel}_k(X)}$ is continuous for all k, and using this we see that X has the defining property for the colimit of the spaces $\mathrm{skel}_k(X)$. □

Proposition 30.42. [`prop-coprod-CW`]
The coproduct of any family of CW complexes has a natural structure as a CW complex.

Proof. Suppose we have CW complexes (X_i, A_i, d_i, ϕ_i) for all $i \in I$. We put $X = \coprod_i X_i$ and $A = \coprod_i A_i$, and we regard X_i as a subspace of X and A_i as a subset of A in the obvious way. We let $d\colon A \to \mathbb{N}$ be the map given by d_i on A_i, and we let

$$\phi\colon \coprod_{a\in A} B^{d(a)} = \coprod_{i\in I} \coprod_{a\in A_i} B^{d_i(a)} \to \coprod_{i\in I} X_i = X$$

be the coproduct of the maps ϕ_i. We leave it to the reader to check that this gives a CW structure. □

Proposition 30.43. [`prop-prod-CW`]
The product in ***CGWH*** *of any two CW complexes has a natural structure as a CW complex.*

Proof. Let (X_0, A_0, d_0, ϕ_0) and (X_1, A_1, d_1, ϕ_1) be CW complexes. Put $X = X_0 \times X_1$ and $A = A_0 \times A_1$. Define $d\colon A \to \mathbb{N}$ by $d(a_0, a_1) = d_0(a_0) + d_1(a_1)$, so

$$B^{d(a_0,a_1)} = B^{d_0(a_0)} \times B^{d_1(a_1)}.$$

The maps $\phi_{i,a_i}\colon B^{d_i(a_i)} \to X_i$ (for $i = 0, 1$) give a map

$$\phi_{a_0,a_1} = \phi_{0,a_0} \times \phi_{1,a_1}\colon B^{d(a_0,a_1)} = B^{d_0(a_0)} \times B^{d_1(a_1)} \to X_0 \times X_1 = X.$$

These fit together to give a map $\phi\colon \coprod_{a\in A} B^{d(a)} \to X$. The claim is that (X, A, d, ϕ) is a CW complex. First note that Proposition 23.30 allows us to identify $\coprod_a B^{d(a)}$ with $\left(\prod_{a_0} B^{d_0(a_0)}\right) \times \left(\prod_{a_1} B^{d_1(a_1)}\right)$ and thus ϕ with $\phi_0 \times \phi_1$. The maps ϕ_0 and ϕ_1 are quotient maps by hypothesis, so ϕ is a quotient map by Proposition 23.32. Next, we can identify $\mathring{B}^{d(a)}$ with $\mathring{B}^{d_0(a_0)} \times \mathring{B}^{d_1(a_1)}$, and thus identify the restricted map $\phi'\colon \coprod_a \mathring{B}^{d(a)} \to X$ with the product of the corresponding maps $\phi'_0\colon \coprod_{a_0} \mathring{B}^{d_0(a_0)} \to X_0$ and $\phi'_1\colon \coprod_{a_1} \mathring{B}^{d_1(a_1)} \to X_1$. Here ϕ'_0 and ϕ'_1 are bijections by hypothesis, so ϕ' is a bijection, which is axiom CW0 for X. Next, for any pair $a = (a_0, a_1) \in A$ we see from the definitions that

$$\begin{aligned} E_a &= E_{a_0} \times E_{a_1} \\ OE_a &= OE_{a_0} \times OE_{a_1} \\ DE_a &= (DE_{a_0} \times E_{a_1}) \cup (E_{a_0} \times DE_{a_1}). \end{aligned}$$

This means that for any other pair $b = (b_0, b_1) \in A$, we have

$$OE_b \cap E_a = (OE_{b_0} \cap E_{a_0}) \times (OE_{b_1} \cap E_{a_1}).$$

By considering when this is nonempty, we deduce that

$$\begin{aligned} A_a^+ &= (A_0)_{b_0}^+ \times (A_1)_{b_1}^+ \\ A_a &= ((A_0)_{b_0}^+ \times (A_1)_{b_1}^+) \setminus \{a\} \end{aligned}$$

and axiom CW1 follows from this. $\square$

30.4. Paracompactness.

30.5. Homotopical properties.

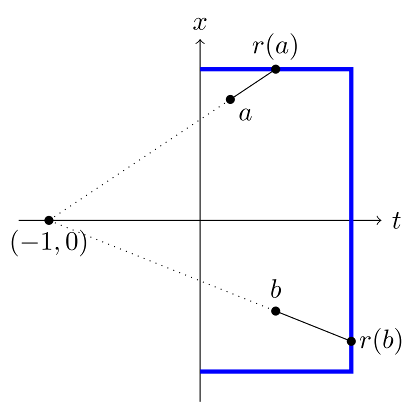
Lemma 30.44. [lem-ndr-cell]
There is a continuous retraction

$$r\colon [0,1] \times B^n \to ([0,1] \times \partial(B^n)) \cup (\{1\} \times B^n)$$

given by

$$r(t,x) = \begin{cases} \left(1, \frac{2x}{1+t}\right) & \text{if } 0 \le \|x\| \le \frac{1+t}{2} \\ \left(\frac{1+t-\|x\|}{\|x\|}, \frac{x}{\|x\|}\right) & \text{if } \frac{1+t}{2} \le \|x\| \le 1. \end{cases}$$

Proof. Geometrically, this is just radial projection from the point $(-1, 0)$:



For a formal proof, put

$$\begin{aligned} X &= \{(t,x) \ : \ 0 \le \|x\| \le (1+t)/2\} \\ Y &= \{(t,x) \ : \ (1+t)/2 \le \|x\| \le 1\} \\ C &= ([0,1] \times \partial(B^n)) \cup (\{1\} \times B^n) \end{aligned}$$

These sets are both closed. The first clause above defines a continuous map $r_X\colon X \to \{1\} \times B^n \subseteq C$. (There is no problem with the denominator, because $1 + t \geq 1$ on X.) Similarly, we have $\|x\| \geq 1/2$ on Y, so the second clause gives a continuous map $r_Y\colon Y \to [0,1] \times \partial(B^n) \subseteq C$. For $(t,x) \in X \cap Y$ we have $\|x\| = (1+t)/2$ so $r_X(t,x) = (1, x/\|x\|) = r_Y(t,x)$, so there is a well-defined map $r\colon [0,1] \times B^n \to C$ as described. It is clear that $r(1,x) = (1,x)$, and that $r(t,x) = (t,x)$ whenever $\|x\| = 1$, so $r|_C = 1_C$, so r is a retraction as claimed. $\square$

LEMMA 30.45. *Let X be a CW complex, and let X' be a subcomplex. Suppose that X is n-dimensional and that* $\mathrm{skel}_{n-1}(X) \subseteq X'$. *Then there is a natural retraction*

$$r\colon [0,1] \times X \to ([0,1] \times X') \cup (\{1\} \times X).$$

PROOF. More formally, we have a CW complex (X, A, d, ϕ) where $d(a) \leq n$ for all a, and a subcomplex (X', A', d', ϕ') such that $a \in A'$ whenever $d(a) < n$. Put $T = A \setminus A'$, so $d(t) = n$ for all $t \in T$. By a straightforward adaptation of Proposition 30.41, we have a pushout square

$$\begin{tikzcd} \coprod_t \partial(B^n) \arrow[r, rightarrowtail] \arrow[d, "\phi"'] & \coprod_t B^n \arrow[d, "\phi"] \\ X' \arrow[r, rightarrowtail] & X. \end{tikzcd}$$

Unfinished $\square$

30.6. Spaces homotopy equivalent to CW complexes.

31. Euclidean neighbourhood retracts

DEFINITION 31.1. [`defn-subeuclidean`]
Let X be a topological space. We say that X is *subeuclidean* if it is locally compact, and it admits an embedding $i\colon X \to \mathbb{R}^n$ for some n.

REMARK 31.2. [`rem-subeuclidean`]
Here we mean an embedding as defined in Definition 4.1, so i gives a homeomorphism from X to the set $i(X) \subseteq \mathbb{R}^n$ with its subspace topology. In particular, this means that $i(X)$ is locally compact and thus locally closed (by Proposition 18.5).

LEMMA 31.3. [`lem-subeuclidean-closed`]
If X is subeuclidean then there is a closed embedding $j\colon X \to \mathbb{R}^{n+1}$ for some n.

PROOF. By definition, there must exist a map $i\colon X \to \mathbb{R}^n$ that is an embedding but need not be closed. However, $i(X)$ is locally closed, so it has the form $U \cap F$ for some open set U and some closed set F. We can define a continuous map $f\colon \mathbb{R}^n \to \mathbb{R}$ by $f(z) = d(z, U^c) = \inf\{d(z,z') \;:\; z' \in U^c\}$ as in Definition 12.52. We then put $G = \{(z,t) \in \mathbb{R}^n \times \mathbb{R} \;:\; f(z)t = 1\}$, which is easily seen to be closed in $\mathbb{R}^{n+1}$. If $(z,t) \in G$ then $f(z) \neq 0$ and so $z \in U$. We therefore have a projection $p\colon G \to U$ given by $p(z,t) = z$. In the other direction, as f is strictly positive on U we can define a continuous map $g\colon U \to G$ by $g(z) = (z, 1/f(z))$. These maps are inverse to each other, so g gives a homeomorphism from the open set $U \subseteq \mathbb{R}^n$ to the closed set $G \subseteq \mathbb{R}^{n+1}$. Now put $j = gi$, considered as a map $X \to \mathbb{R}^{n+1}$. We have $pj = i$, which is an embedding, so j is an embedding by Proposition 4.10. Moreover, as $i(X)$ is closed in U, we see that $j(X) = g(i(X))$ is closed in G and therefore in $\mathbb{R}^{n+1}$. It follows that j is a closed embedding, as required. $\square$

LEMMA 31.4. [`lem-union-subeuclidean`]
Suppose that X is a locally compact Hausdorff space that can be covered by finitely many subeuclidean open subsets. Then X is itself subeuclidean.

PROOF. By hypothesis we have $X = \bigcup_{i=1}^k X_i$ say, where X_i is open and subeuclidean. The claim is trivial if $X_i = X$ for some i, so we assume that all the subsets X_i are proper, so $X_i^c \neq \emptyset$. Proposition 18.23 tells us that there is a quotient map $q_i\colon X \to X_i \cup \{\infty\}$ that acts as the identity on X_i and sends X_i^c to ∞. We claim that q_i is a closed map. To see this, let F be a closed subset of X and put $F_i = q_i(F)$. We need to show that F_i is closed, but q_i is a quotient map, so it will suffice to show that $q_i^{-1}(F_i)$ is closed in

X. By inspection we find that $q_i^{-1}(F_i)$ is equal to F (if $F \subseteq X_i$) or $F \cup X_i^c$ (if $F \not\subseteq X_i$) and in either case it is closed as required.

Now put $X' = \prod_i (X_i \cup \{\infty\})$ and define $q\colon X \to X'$ by $q(x) = (q_1(x), \dots, q_k(x))$. We claim that this is an embedding. We will test this using closed sets, as in Remark 4.2. First suppose that $q(x) = q(x')$, so $q_i(x) = q_i(x')$ for all i. As the sets X_i cover X, we have $x \in X_i$ for some i and then $q_i(x) = x \neq \infty$. By assumption we then have $q_i(x') = x \neq \infty$, and this means that $x' \in X_i$ and then that $x' = x$. This shows that q is injective. Now suppose that $F \subseteq X$ is closed. We put $F_i = q_i(F)$ as before, so F_i is closed in $X_i \cup \{\infty\}$. We then put $F' = \prod_i F_i$, which is closed in X' by Proposition 5.34. It is clear that $F \subseteq q^{-1}(F') = \bigcap_i q_i^{-1}(F_i)$, and we claim that this is actually an equality. Indeed, if $x \in q^{-1}(F') \subseteq X$ we can choose i such that $x \in X_i$, and then we find that $x \in q_i^{-1}(F_i) = q_i^{-1}(q_i(F))$ so $q_i(x) = q_i(x')$ for some $x' \in F$. On the other hand, as $x \in F_i$ we have $q_i(x) = x \neq \infty$, so we see that $x' = x$ and so $x \in F$ as required. This completes the proof that q is an embedding.

Next, by Lemma 31.3 we can choose closed embeddings $f_i\colon X_i \to \mathbb{R}^{n_i-1}$ say. Now, if $K \subseteq \mathbb{R}^{n_i-1}$ is compact then the set $K' = K \cap f_i(X_i)$ is closed in K and therefore compact, and the map $f_i\colon X_i \to f_i(X_i)$ is a homeomorphism so the set $f_i^{-1}(K) = f_i^{-1}(K')$ is again compact. This means that f_i is proper, so it extends to give a continuous injective map $\overline{f}_i\colon X_i \cup \{\infty\} \to \mathbb{R}^{n_i-1} \cup \{\infty\}$. As the source and target of $\overline{f}_i$ are compact Hausdorff, we see from Proposition 10.22 that $\overline{f}_i$ is an embedding.

Next, Proposition 18.20 identifies $\mathbb{R}^{n_i-1}$ with S^{n_i-1} and so gives an embedding $g_i\colon \mathbb{R}^{n_i-1} \cup \{\infty\} \to \mathbb{R}^{n_i}$. The composite $g_i\overline{f}_i$ is then also an embedding. We now put $n = \sum_i n_i$ and $h = \prod_i g_i\overline{f}_i\colon X' \to \mathbb{R}^n$, which is an embedding by Remark 5.33. Now $qh\colon X \to \mathbb{R}^n$ is the required embedding. □

Definition 31.5. [`defn-locally-contractible`]
We say that X is *locally contractible* if for all $x \in X$ and every neighbourhood U of X, there is a smaller neighbourhood V with $x \in V \subseteq U$, and a map $h\colon [0,1] \times V \to U$ with $h(0,v) = x$ for all $v \in V$, and $h(1,v) = v$. In this situation we say that V *contracts to* x *inside* U.

Example 31.6. [`eg-locally-contractible`]
Let X be open in $\mathbb{R}^n$. We then claim that X is locally contractible. Indeed, given any $x \in X$ and any neighbourhood U, we observe that U is open in $\mathbb{R}^n$ and so contains the set $V = OB_\epsilon(x)$ for some $\epsilon > 0$. We can then define $h\colon [0,1] \times V \to V \subseteq U$ by $h(t,y) = x + t(y - x)$, and this is easily seen to have the required properties.

Remark 31.7. With notation as in Definition 31.5, we see that for each $v \in V$ the function $t \mapsto h(t,v)$ gives a path γ_v from x to v, so every point in V lies in the same path component as x. It does not quite follow that V is path connected, because the path $\gamma_v\colon [0,1] \to U$ might stray outside V. There are a number of subtle distinctions that can be made here, but we will not discuss them.

Definition 31.8. [`defn-enr`]
Let X be a locally compact Hausdorff space. We say that X is a *euclidean neighbourhood retract* (or *ENR*) if it satisfies the following equivalent conditions.

(a) There is an integer $n \geq 0$, an open set $U \subseteq \mathbb{R}^n$ and continuous maps $X \xrightarrow{j} U \xrightarrow{p} X$ with $pj = 1_X$.
(b) There is an integer $n \geq 0$ and an embedding $X \xrightarrow{j} \mathbb{R}^n$ (so X is subeuclidean). Moreover, for any such j there is an open set $U \subseteq \mathbb{R}^n$ containing $j(X)$ and a continuous map $p\colon U \to X$ with $pj = 1_X$.
(c) X is subeuclidean and locally contractible.

The proof of equivalence will be split between several lemmas below.

Lemma 31.9. [`lem-enr-ab`]
In Definition 31.8, conditions (a) and (b) are equivalent.

Proof. It is clear that (b) implies (a). Conversely, suppose that (a) holds, and choose U, j and p as specified there. Suppose we have another embedding $k\colon X \to \mathbb{R}^m$. As this is an embedding, it can be regarded as a homeomorphism $X \to k(X)$. We therefore have a continuous map $r = (k(X) \xrightarrow{k^{-1}} X \xrightarrow{j} \mathbb{R}^n)$, with components $r_i\colon k(X) \to \mathbb{R}$ for $i = 1, \dots, n$.

Next, as X is locally compact and k is an embedding we see that $k(X)$ is locally compact and thus locally closed, so it has the form $F \cap W$ for some open subset $W \subseteq \mathbb{R}^m$ and some closed subset $F \subseteq \mathbb{R}^m$. (Here we have used Proposition 18.5.) Now W can be regarded as a metric space, so it is normal by Proposition 14.8, and $k(X)$ is closed in W. The Tietze Extension Theorem (Theorem 17.5) therefore applies, so the maps $r_i: k(X) \to \mathbb{R}$ can be extended over W. We will use the same notation for the extended maps $r_i: W \to \mathbb{R}$. These combine to give a continuous map $r: W \to \mathbb{R}^n$ with $rk = j: X \to \mathbb{R}^n$. Put $V = r^{-1}(U)$, which is open in W and contains $k(X)$. As $r(V) \subseteq U$ and $p: U \to X$ we have a map $q = pr|_V: V \to X$. For $x \in X$ we have $qk(x) = prk(x) = pj(x) = x$. This proves that (b) holds. $\square$

LEMMA 31.10. [`lem-enr-ac`]
In Definition 31.8, condition (a) implies condition (c).

PROOF. Suppose that (a) holds, and that we are given a point $x \in X$ and an open neighbourhood W of x in X. Choose U, j and p as in (a). After shifting everything by a translation if necessary, we may assume that $j(x) = 0$. As j gives a homeomorphism $X \to j(X)$, it will be harmless to replace X by $j(X)$ and thus assume that $0 \in X \subseteq U \subseteq \mathbb{R}^n$ and that $p: U \to X$ with $p|_X = 1_X$. As U is open in $\mathbb{R}^n$ and W is open in X we can choose $\epsilon > 0$ such that $OB_\epsilon(0) \subseteq U$ and $OB_\epsilon(0) \cap X \subseteq W$. As $p: U \to X$ is continuous, we can choose δ such that $0 < \delta < \epsilon$ and $OB_\delta(0) \subseteq p^{-1}(OB_\epsilon(0))$. Put $V = OB_\delta(0) \cap X \subseteq W$. If $t \in [0,1]$ and $y \in V$ then $ty \in OB_\delta(0) \subseteq U$ so $p(ty)$ is defined and lies in X. Moreover, as $OB_\delta(0) \subseteq p^{-1}(OB_\epsilon(0))$ we see that $p(ty) \in OB_\epsilon(0) \cap X \subseteq W$. We can thus define $h: [0,1] \times V \to W$ by $h(t,y) = p(ty)$. We then have $h(0,y) = p(0) = 0$ (because $0 \in X$ and $p|_X = 1_X$). We also have $h(1,y) = p(y) = y$, because $y \in V \subseteq X$. This gives the required local contraction. $\square$

All that is left is to prove that (c) implies (a). To do this we will need to subdivide $\mathbb{R}^n$ into small cubes. **Redo this using the CW structures on open subsets of $\mathbb{R}^n$ discussed previously.**

32. The category of arrows

We next define some auxiliary constructions that will be needed for our analysis of fibrations and cofibrations. Let $\mathbf{CGWH}^\downarrow$ be the category whose objects are diagrams in $\mathbf{CGWH}$ of shape $A = (A_0 \xrightarrow{u} A_1)$, and whose morphisms from $(A_0 \xrightarrow{u} A_1)$ to $(B_0 \xrightarrow{v} B_1)$ are the commutative squares

$$\begin{tikzcd} A_0 \arrow[r, "f_0"] \arrow[d, "u"'] & B_0 \arrow[d, "v"] \\ A_1 \arrow[r, "f_1"'] & B_1. \end{tikzcd}$$

Given objects A and B as above, we put

$$P = (A_0 \times B_1) \cup_{A_0 \times B_0} (A_1 \times B_0),$$

so we have a diagram

$$\begin{tikzcd} A_0 \times B_0 \arrow[rr] \arrow[dd] & & A_1 \times B_0 \arrow[dd] \\ & P \arrow[ur] \arrow[dl] \arrow[dr] & \\ A_0 \times B_1 \arrow[rr] & & A_1 \times B_1, \end{tikzcd}$$

in which the top left triangle is a distorted pushout square. We write

$$A \square B = (P \to A_1 \times B_1) \in \mathbf{CGWH}^\downarrow.$$

This can also be characterised by a universl property, as follows. If we have a third object $C \in \mathbf{CGWH}^{\downarrow}$ then we have two commutative squares as shown:

$$\begin{tikzcd} A_0 \times B_0 \arrow[r] \arrow[d] & A_1 \times B_0 \arrow[d] \\ A_0 \times B_1 \arrow[r] & A_1 \times B_1 \end{tikzcd} \qquad \begin{tikzcd} C_0 \arrow[r, "1"] \arrow[d, "1"'] & C_0 \arrow[d] \\ C_0 \arrow[r] & C_1. \end{tikzcd}$$

Maps $A \square B \to C$ in $\mathbf{CGWH}^{\downarrow}$ biject with maps between the above squares, or equivalently with compatible systems of maps $A_i \times B_j \to C_{ij}$ for $i, j \in \{0, 1\}$. Using this one can check that $\mathbf{CGWH}^{\downarrow}$ is symmetric monoidal under the above product. The unit is the object $(\emptyset \to 1) \in \mathbf{CGWH}^{\downarrow}$, and the triple product is characterised by the fact that maps $A \square B \square C \to D$ biject with compatible systems of maps $A_i \times B_j \times C_j \to D_{ijk}$ in $\mathbf{CGWH}$.

Next, observe that any map $g \colon B_1 \to C_0$ gives a map $(gv, wg) \colon B \to C$ in $\mathbf{CGWH}^{\downarrow}$, as indicated in the following diagram:

$$\begin{tikzcd} B_0 \arrow[r, "gv"] \arrow[d, "v"'] & C_0 \arrow[d, "w"] \\ B_1 \arrow[r, "wg"'] \arrow[ur, "g"] & C_1. \end{tikzcd}$$

By considering this construction for all possible g, we obtain a diagram as follows:

$$\begin{tikzcd} C(B_1, C_0) \arrow[rr, "v^*"] \arrow[dd, "w_*"'] \arrow[dr] & & C(B_0, C_0) \arrow[dd, "w_*"] \\ & \mathbf{CGWH}^{\downarrow}(B, C) \arrow[ur] \arrow[dl] & \\ C(B_1, C_1) \arrow[rr, "v^*"'] & & C(B_0, C_1). \end{tikzcd}$$

The bottom right triangle is a distorted pullback square, by the definition of $\mathbf{CGWH}^{\downarrow}(B, C)$. We define

$$F(B, C) = (C(B_1, C_0) \to \mathbf{CGWH}_1(B, C)) \in \mathbf{CGWH}^{\downarrow}.$$

Proposition 32.1. [prop-box-f]
There is a natural bijection $\mathbf{CGWH}^{\downarrow}(A, F(B, C)) = \mathbf{CGWH}^{\downarrow}(A \square B, C)$, making $\mathbf{CGWH}^{\downarrow}$ a closed symmetric monoidal category.

Proof. A morphism from A to $F(B, C)$ consists of a map $f_{0,1} \colon A_0 \to C(B_1, C_0)$ together with a map $A_1 \to \mathbf{CGWH}^{\downarrow}(B, C)$, which in turn consists of maps $f_{1,0} \colon A_1 \to C(B_0, C_0)$ and $f_{1,1} \colon A_1 \to C(B_1, C_1)$. These three maps are subject to some compatibility conditions. Adjointly, we have maps $f^{\#}_{0,1} \colon A_0 \times B_1 \to C_0$ and $f^{\#}_{1,0} \colon A_1 \times B_0 \to C_0$ and $f^{\#}_{1,1} \colon A_1 \times B_1 \to C_1$. One of the compatibility conditions implies that the diagram

$$\begin{tikzcd} A_0 \times B_0 \arrow[r, "u \times 1"] \arrow[d, "1 \times v"'] & A_1 \times B_0 \arrow[d, "f^{\#}_{10}"'] \\ A_0 \times B_1 \arrow[r, "f^{\#}_{01}"'] & C_0 \end{tikzcd}$$

commutes; we write $f^{\#}_{0,0}$ for the resulting map $A_0 \times B_0 \to C_0$. We thus have compatible maps $f^{\#}_{i,j} \colon A_i \times B_j \to C_{ij}$ for all $i, j \in \{0, 1\}$, and thus a map $A \square B \to C$ as discussed previously. We leave it to the reader to check that this construction is bijective. $\square$

Definition 32.2. [defn-orthogonal]

Suppose we have two objects $A = (A_0 \xrightarrow{u} A_1)$ and $B = (B_0 \xrightarrow{v} B_1)$ in $\mathbf{CGWH}^{\downarrow}$. We say that A (or u) is *left orthogonal* to B (or v), or that B is *right orthogonal* to A, if for every commutative square as shown (without the h) there exists a map h making the whole diagram commute.

$$\begin{tikzcd} A_0 \arrow[r, "f_0"] \arrow[d, "u"'] & B_0 \arrow[d, "v"] \\ A_1 \arrow[r, "f_1"'] \arrow[ur, dotted, "h"] & B_1 \end{tikzcd}$$

(We say that h *fills in the square.*) It is equivalent to say that the map $F(A,B) \in \mathbf{CGWH}^{\downarrow}$ is surjective.

Given a class M of maps, we write ${}^{\perp}M$ for the class of maps that are left orthogonal to every map in M, and $M^{\perp}$ for the class of maps that are right orthogonal to every map in M.

Example 32.3. (a) If $M = \{\emptyset \to 1\}$ then $M^{\perp}$ is the class of surjective maps.

(b) If M is the class of all maps of the form $\emptyset \to B$, then $M^{\perp}$ is the class of all split surjections. Dually, if M is the class of all maps of the form $A \to 1$, then ${}^{\perp}M$ is the class of all split monomorphisms.

(c) If M is the class of all maps, then ${}^{\perp}M$ is the class of homeomorphisms. Indeed, if $(A_0 \xrightarrow{u} A_1) \in {}^{\perp}M$ then we can consider $B = (A_0 \to 1)$ to see that u is a split monomorphism, and then $1 \to A_1/A_0$ to see that u is also surjective. By a dual argument, $M^{\perp}$ is also the class of homeomorphisms.

33. Fibrations, cofibrations and lifting properties

33.1. Cofibrations.

Definition 33.1. [defn-cofibration]

Let $j\colon X \to Y$ be a map. Define $\mathrm{Cyl}(X) = I \times X$, and let $i_1\colon X \to \mathrm{Cyl}(X)$ be the map $i_1(x) = (0, x)$. Let $\mathrm{Cyl}(j) = (I \times X) \cup_X Y$ be the following pushout:

$$\begin{tikzcd} X \arrow[r, "i_1"] \arrow[d, "j"'] & \mathrm{Cyl}(X) \arrow[d] \\ Y \arrow[r] & \mathrm{Cyl}(j). \end{tikzcd}$$

This is called the *mapping cylinder* of i. There is an evident map $k\colon (I \times X) \cup_X Y \to I \times Y$, which maps $I \times X$ by $1 \times j$ and Y by i_1. We say that j is a *cofibration* if there is a map $r\colon I \times Y \to (I \times X) \cup_X Y$ such that $rk = 1$. This means that $I \times Y$ can be pushed down continuously onto the subspace $(I \times X) \cup_X Y$.

A cofibration which is a homotopy equivalence is called an *acyclic cofibration.* We write cof and acf for the classes of cofibrations and acyclic cofibrations.

Remark 33.2. The map $k\colon \mathrm{Cyl}(j) \to \mathrm{Cyl}(X)$ above can also be thought of as the smashout map $(1 \rightarrowtail I)\square j$. Thus j is a cofibration if and only if $(1 \rightarrowtail I)\square j$ is a split monomorphism.

The idea is that a cofibration is a homotopically well-behaved inclusion. The following theorem gives a convenient test.

Theorem 33.3. [thm-ndr]

A map $i\colon X \to Y$ is a cofibration if and only if it is a closed inclusion (so we can think of X as a closed subspace of Y) and there are maps $u\colon Y \to I$ and $h\colon I \times Y \to Y$ such that

(1) $u^{-1}\{0\} = X$
(2) $h_1 = 1_Y$
(3) $h_t|_X = 1_X$ *for all* t.
(4) $h_0(y) \in X$ *for all* y *such that* $u(y) < 1$.

Here $h_t\colon Y \to Y$ is defined by $h_t(y) = h(t, y)$.

Moreover, i is an acyclic cofibration if and only if we can choose u, h such that $u(y) \le 1/2$ for all y, and thus $h_1(Y) = X$.

Example 33.4. [eg-cofibrations]

(1) A smooth closed embedding of manifolds is a cofibration.
(2) The inclusion of a subcomplex in a simplicial complex is a cofibration.
(3) The inclusion of a closed subvariety in a real or complex projective variety is a cofibration.
(4) For an example in the other direction, consider the "Hawaiian earrings" space E, which is the union of the circles of radius $1/n$ centred at $(1/n, 0) \in \mathbb{R}^2$ as n runs from 1 to ∞. The inclusion of the one-point space $\{(0,0)\}$ in E is not a cofibration.

The following fact turns out to be crucial.

Proposition 33.5. [prop-smashout-cof]
The smashout of two cofibrations is a cofibration. If either one is a homotopy equivalence, then so is the smashout. The crossmap of a cofibration and a fibration is a fibration. If either one is a homotopy equivalence, then so is the crossmap.

The following *homotopy extension property* of cofibrations is very useful.

Proposition 33.6. *Let X be a closed subspace of a space Y, such that the inclusion map $i\colon X \to Y$ is a cofibration. Suppose we are given a map $f\colon Y \to Z$, and a homotopy $g_t\colon X \to Z$ ending with $g_1 = f|_X$. Then the homotopy can be extended to give a homotopy $h_t\colon Y \to Z$ with $h_t|_X = g_t$ and $h_1 = f$.*

Conversely, if X is any closed subspace of Y with this homotopy extension property, then the inclusion $X \to Y$ is a cofibration.

33.2. Fibrations. The dual concept is that of a fibration.

Definition 33.7. [defn-fibration]
Let $q\colon E \to B$ be a map. As before we define

$$\mathrm{Path}(B) = C(I, B) = \{ \text{ continuous paths } \omega\colon I \to B\}.$$

Let $p_1\colon \mathrm{Path}(B) \to B$ be the map $p_1(\omega) = \omega(1)$, and define the *mapping path space* $\mathrm{Path}(q) = C(I,B) \times_B E$ by the following pullback:

$$\begin{tikzcd} \mathrm{Path}(i) \arrow[r] \arrow[d] & \mathrm{Path}(B) \arrow[d, "p_1"] \\ E \arrow[r, "q"'] & B. \end{tikzcd}$$

In other words,

$$\mathrm{Path}(q) = \{(\omega, e) \in C(I,B) \times E \ : \ \omega(1) = q(e)\},$$

so a point of $\mathrm{Path}(q)$ is a path in B together with a lift of the final point to E. There is an evident map $r\colon \mathrm{Path}(E) \to \mathrm{Path}(q)$, given by $r(\alpha) = (q \circ \alpha, \alpha(1))$. We say that q is a *(Hurewicz) fibration* if there is a map $l\colon \mathrm{Path}(q) \to \mathrm{Path}(E)$ such that $rl = 1$. This means that $\alpha = l(\omega, e)$ is a path in E which is a lift of ω (in the sense that $q \circ \alpha = \omega$), ending at $\alpha(1) = e$, which is our given lift of $\omega(1)$. Such a map l is called a *lifting function* for q. A fibration which is also a homotopy equivalence is called an *acyclic fibration*. We write fib and afb for the classes of fibrations and acyclic fibrations.

Remark 33.8. The map $r\colon \mathrm{Path}(E) \to \mathrm{Path}(q)$ defined above can also be thought of as the crossmap $F(1 \rightarrowtail I, q)$. Thus, q is a fibration if and only if $F(1 \rightarrowtail I, q)$ is a split epimorphism.

The idea is that a fibration is a homotopically well-behaved projection map.

Example 33.9. [eg-bundles]
Under mild technical conditions (for example, if the base space is a metric space or a CW complex — see Section 33.7), every locally-trivial fibre bundle is a fibration. Here are some interesting examples of locally-trivial fibre bundles.

(1) Covering maps; in particular, quotients by discrete group actions; for example, the covering of a compact Riemann surface of genus greater than one by the unit disc.

(2) Maps of the configuration spaces $F_k\mathbb{C}$.
(3) The fibration $O(n-1) \to O(n) \xrightarrow{q} S^{n-1}$, where $O(n)$ is the space of $n \times n$ orthogonal matrices, and $q(A) = Ae_n$, where e_n is the last basis vector in $\mathbb{R}^n$. (When we say that $F \to E \xrightarrow{q} B$ is a fibration, we mean that q is a fibration and $F = \mathrm{fib}(q)$.)
(4) The fibration $\mathcal{P}^{n-1} \to H \xrightarrow{q} \mathcal{P}^m$, where $m \leq n$, H is Milnor's hypersurface $\{([z],[w]) \in \mathcal{P}^n \times \mathcal{P}^m : \sum_{i=0}^m z_i w_i = 0\}$, and $q([z],[w]) = [w]$.
(5) The fibration $S^1 \to S^{2n+1} \xrightarrow{q} \mathcal{P}^n$. The case $n = 1$ is especially interesting. There we can identify S^3 with $\{(z,w) \in \mathbb{C}^2 : |z|^2 + |w|^2 = 1\}$ and $\mathcal{P}^1$ with $\mathbb{C} \cup \{\infty\}$, and q with the map $(z,w) \mapsto z/w$. As $\mathcal{P}^1$ is also homeomorphic to S^2, we get a fibration $S^1 \to S^3 \to S^2$.
(6) It can be shown that any surjective submersion of manifolds is a locally trivial bundle. (A smooth map $f\colon M \to N$ is said to be a *submersion* if the induced map of tangent spaces $T_xM \to T_{f(x)}N$ is surjective for all $x \in M$.)

EXAMPLE 33.10. [`eg-fibrations`]
We now list some fibrations which are not locally trivial bundles.

(1) For any space X, there is a fibration $\Omega X \to PX \xrightarrow{p_1} X$.
(2) For any $n > 0$ there is a 2-local fibration $S^n \xrightarrow{\eta} \Omega S^{n+1} \xrightarrow{H} \Omega S^{2n+1}$, called the EHP fibration. This really means that there is a map $H\colon \Omega S^{n+1} \to \Omega S^{2n+1}$ and a map $f\colon S^n \to PH$ which induces an isomorphism $\pi_*(S^n)/2 \simeq \pi_*(PH)/2$, and that the composite $S^n \to PH \to \Omega S^{n+1}$ is η.
(3) It can be shown that the homotopy fibre of the inclusion $X \vee Y \to X \times Y$ is homotopy equivalent to $\Sigma((\Omega X) \wedge (\Omega Y))$.

EXAMPLE 33.11. [`eg-not-fibration`]
If $B = \mathbb{R}$ and

$$E = \{(x,y) \in \mathbb{R}^2 \ : \ (x \leq 0 \text{ and } |y| = |x|) \text{ or } (x \geq 0 \text{ and } y = 0)\}$$

then the vertical projection $E \to B$ is not a fibration.

THEOREM 33.12. [`thm-orthogonal`]

(1) *A map is a cofibration if and only if it is left orthogonal to all maps of the form $p_1\colon \mathrm{Path}(B) \twoheadrightarrow B$, if and only if it is left orthogonal to every acyclic fibration. In symbols:*

$$\mathrm{cof} = {}^{\perp}\{ \text{ maps of the form } p_1\colon \mathrm{Path}(B) \twoheadrightarrow B\} = {}^{\perp}\,\mathrm{afb}\,.$$

(2) *A map is a fibration if and only if it is right orthogonal to all maps of the form $i_1\colon X \rightarrowtail \mathrm{Cyl}(X)$, if and only if it is right orthogonal to every acyclic cofibration. In symbols:*

$$\mathrm{fib} = \{ \text{ maps of the form } i_1\colon X \rightarrowtail \mathrm{Cyl}(X)\}^{\perp} = \mathrm{acf}^{\perp}\,.$$

We also have $\mathrm{acf} = {}^{\perp}\,\mathrm{fib}$ *and* $\mathrm{afb} = \mathrm{cof}^{\perp}$.

PROOF. **Prove this** □

COROLLARY 33.13. [`cor-orthogonal`]
The classes cof *and* acf *are closed under composition, disjoint unions, pushouts, retractions, and sequential colimits. The classes* fib *and* afb *are closed under compositions, products, pullbacks, retractions, and sequential inverse limits. All four classes contain all homeomorphisms.*

PROOF. **Prove this** □

PROPOSITION 33.14. [`prop-factor`]

Any map $f\colon X \to Y$ can be fitted into a natural commutative diagram as shown, in which the arrow marked cof *is a cofibration and so on.*

$$\begin{tikzcd}
X \arrow[r, tail, "\mathrm{cof}"] \arrow[d, "\mathrm{heq}"'] & \mathrm{Cyl}(f) \arrow[d, "\mathrm{heq}"] \\
\mathrm{Path}(f) \arrow[r, two heads, "\mathrm{fib}"'] & Y.
\end{tikzcd}$$

Proof. **Prove this** □

The following propositions are also useful.

Proposition 33.15. [`prop-map-cof`]
If X is compact and $j\colon Y \to Z$ is a cofibration then the map $j_*\colon C(X,Y) \to C(X,Z)$ is a cofibration.

Proof. **Prove this** □

Proposition 33.16. [`prop-pullback-cof`]
Suppose that we have a pullback square as follows, in which j is a cofibration and q is a fibration. Then j' is a cofibration and q' is a fibration.

$$\begin{tikzcd}
E' \arrow[r, "j'"] \arrow[d, "q'"'] & E \arrow[d, "q"] \\
B' \arrow[r, "j"'] & B.
\end{tikzcd}$$

Proof. **Prove this** □

Proposition 33.17. *Let $j\colon W \to X$ be a map. The following are equivalent:*

(a) *j is a cofibration.*
(b) *j is left orthogonal to all maps of the form $p_1\colon \mathrm{Path}(E) \to E$.*
(c) *j has the homotopy extension property: for any map $f\colon X \to E$ and any homotopy $g_t\colon W \to E$ ending with $g_1 = fj$, there is a homotopy $h_t\colon X \to E$ extending g_t (in the sense that $h_t \circ j = g_t$) and ending with $h_1 = f$.*

Proof. (a)⇒(c): Let j be a cofibration, so there is a retraction $r\colon I \times X \to I \times W \cup_W X$. Given maps $f\colon X \to E$ and $g\colon I \times W \to E$ as in (c) we define a map $k\colon I \times W \cup_W X \to E$ by $k(t,w) = g(t,w)$ on $I \times W$ and $k(x) = f(x)$ on X; this is consistent with the equivalence relation $(1,w) = jw$ because $g(1,w) = fj(w)$. We then define $h = kr\colon I \times X \to X$, and check that this is as required in (c).

(c)⇒(a): Suppose that (c) holds. Take $E = I \times W \cup_W X$, and let $f\colon X \to E$ and $g\colon I \times W \to E$ be the obvious maps. We then get a map $h\colon I \times X \to E$ such that $h(1,x) = f(x)$ and $h(t,jw) = (t,w)$ when $w \in W$. It follows that h is a retraction onto $I \times W \cup_W X$, so j is a cofibration.

(b)⇔(c): A square of the form

$$\begin{tikzcd}
W \arrow[r, "g^{\#}"] \arrow[d, "j"'] & \mathrm{Path}(E) \arrow[d, "p_1"] \\
X \arrow[r, "f"'] & E
\end{tikzcd}$$

is the same thing as a pair of maps $f\colon X \to E$, $g\colon I \times W \to E$ such that $g(1,w) = fj(w)$, via the usual translation $g(t,w) = g^{\#}(w)(t)$. A fill in map $h^{\#}\colon X \to \mathrm{Path}(E)$ with $p_1 h^{\#} = f$ and $h^{\#} j = g^{\#}$ is the same as a map $h\colon I \to E$ such that $h(1,x) = f(x)$ and $h(t,jw) = g(t,w)$. Thus (b) is just a translation of (c). □

Proposition 33.18. *Let $q\colon E \to B$ be a map. The following are equivalent:*

(a) *q is a fibration.*
(b) *q is right orthogonal to all maps of the form $i_1\colon X \to I \times X$.*

(c) *q has the homotopy lifting property: given a homotopy $g_t\colon X \to B$ and a map $f\colon X \to E$ which lifts g_1 (in the sense that $qf = g_1$) there is a lifted homotopy $h_t\colon X \to E$ with $qh_t = g_t$ and $h_1 = f$.*

Proof. Exercise. $\square$

33.3. Neighbourhood deformation retracts.

Definition 33.19. A closed subspace $W \subseteq X$ is a *neighbourhood deformation retract (NDR)* if there exist maps $u\colon X \to I$ and $h\colon I \times X \to X$ such that

(a) $W = u^{-1}\{0\}$.
(b) $h_1 = 1_X$.
(c) $h_t|_W = 1_W$ for all $t \in I$.
(d) $h_0(x) \in W$ for all $x \in X$ such that $u(x) < 1$.

We say that (u, h) is a *representation* of W as an NDR. We say that W is a *deformation retract (DR)* if we can choose h such that $h_1(X) \subseteq W$. This holds automatically if $u(x) < 1$ for all x, and conversely if $h_1(X) \subseteq W$ then we can replace u by $u/2$ and assume that $u < 1$ everywhere. Note also that in this case h_1 is a retraction of X onto W.

Proposition 33.20. *If $W \subseteq X$ and $Y \subseteq Z$ are NDRs, then so is $(W \times Z) \cup (X \times Y) \subseteq X \times Z$. Moreover, this is a DR if either $W \subseteq X$ or $Y \subseteq Z$ is.*

Proof. Let (u, h) and (v, k) represent W and Y as NDR's. Write $T = (W \times Z) \cup (X \times Y)$. Define $w\colon X \times Z \to I$ by $w(x, z) = \min(u(x), v(z))$; it is clear that $w^{-1}\{0\} = T$. Put

$$Q = \{(t, x, z, a, b) \in I \times X \times Z \times I^2 \ : \ (1 - a)u(x) = (1 - t)w(x, z) = (1 - b)v(z)\}$$

and define maps

$$I \times X \times Z \xleftarrow{\pi} Q \xrightarrow{\tilde{q}} X \times Z$$

by $\pi(t, x, z, a, b) = (t, x, z)$ and

$$\tilde{q}(t, x, z, a, b) = (h(a, x),\ k(b, z)).$$

Note that Q is closed in $I \times X \times Z \times I^2$. Suppose that $(t, x, z) \in I \times X \times Z$, and we seek a preimage (t, x, z, a, b) in Q.

- If $u(x) > 0$ then we note that $w(x, z) \leq u(x)$ so we can and must take $a = 1 - (1 - t)w(x, z)/u(x)$.
- On the other hand, if $u(x) = 0$ then $w(x, z) = 0$ and we can take a to be any point in I. In these cases we always have $x \in W$ and so $h(a, x) = x$ (independent of a).
- Similarly, if $v(z) > 0$ then we can and must take $b = 1 - (1 - t)w(x, z)/v(z)$. However, if $v(z) = 0$ then we can take b to be any element of I, and we always have $k(b, z) = z$.

In particular, we see that π is surjective. Standard theory of CGWH spaces shows that π is also a closed map, so it is a quotient map. The above comments also show that $\tilde{q}$ is constant on the fibres of π, so there is a unique map $q\colon I \times X \times Z$ satisfying $q\pi = \tilde{q}$. Because π is a quotient map, we see that q is continuous. We claim that the pair (w, q) represents $(X \times Z, T)$ as an NDR pair.

(a) We have already remarked that $w^{-1}\{0\} = T$.
(b) If $(1, x, z, a, b) \in Q$ then we must have $(1 - a)u(x) = 0$, so $a = 1$ or $u(x) = 0$ (which means that $x \in W$). In either case we have $h(a, x) = x$. Similarly, we have $k(b, z) = z$. This shows that $q_1 = 1_{X \times Z}$.
(c) If $(x, z) \in T$ then $w(x, z) = 0$. Thus, if $(t, x, z, a, b) \in Q$ then we again have $(1 - a)u(x) = (1 - b)v(z) = 0$ and so $h(a, x) = x$ and $k(b, z) = z$. This shows that $q_t|_T = 1_T$ for all t.
(d) Now suppose that $w(x, z) < 1$. This means that either ($u(x) \leq v(z)$ and $u(x) < 1$), or ($v(z) \leq u(x)$ and $v(z) < 1$). Suppose that the first of these holds, so $w(x, z) = u(x)$. If $(0, x, z, a, b) \in Q$ we must then have $(1 - a)u(x) = u(x)$, so either $a = 0$ or $u(x) = 0$. In either case, the axioms for (u, h) give $h(a, x) \in W$. Similarly, if our second alternative holds, then $k(b, z) \in Y$. Either way, we have $q(0, x, z) \in T$, as required.

This completes the proof that $(X \times Z, T)$ is an NDR pair.

If W is a DR of X then we may assume that $u < 1$ everywhere. It follows immediately that $w < 1$ everywhere and thus that T is a DR of $X \times Z$. Clearly this also applies if Y is a DR of Z. $\square$

PROPOSITION 33.21. *A map $j\colon W \to X$ is a cofibration if and only if it is a closed inclusion and jW is an NDR of X.*

PROOF. First suppose that j is a closed inclusion (so we can harmlessly think of W as a subspace of X) and that W is an NDR of X. It is easy to see that $\{1\}$ is a DR of I (take $u(s) = 1 - s$ and $h(t,s) = 1-t+ts$). It follows from Proposition 33.20 that $(1\times X)\cup(I\times W)$ is a DR of $I\times X$, so there is a map $r = h_1\colon I\times X \to 1\times X\cup I\times W$ which is the identity on $(1\times X)\cup(I\times W)$. As j is a closed inclusion, one can check that $\mathrm{Cyl}(j)$ is just the space $(1\times X)\cup(I\times W)$, so this map r is precisely what we need to show that j is a cofibration.

Conversely, suppose that $j\colon W\to X$ is a cofibration. One can check from the definitions that the map $W \xrightarrow{i_1} I\times W \to \mathrm{Cyl}(j)$ is always a closed inclusion. As j is a cofibration, the evident map $\mathrm{Cyl}(j)\to I\times X$ has a left inverse. As everything is weakly Hausdorff, it follows that $\mathrm{Cyl}(j)\to I\times X$ is also a closed inclusion, and thus that the composite map $W\to I\times X$ (sending w to $(1,j(w))$) is also a closed inclusion. It is not hard to conclude that j is a closed inclusion. We may therefore harmlessly think of W as a subspace of X, and of $\mathrm{Cyl}(j)$ as $(1\times X)\cup(I\times W)$. The retraction $r\colon I\times X\to(1\times X)\cup(I\times W)\subseteq I\times X$ thus has the form $r(t,x) = (v(t,x),h(t,x))$, where $v\colon I\times X\to I$ and $h\colon I\times X\to X$. Now v is adjoint to a continuous map $v^{\#}\colon X\to C(I,I)$, and the topology on $C(I,I)$ comes from the metric $d(f,g) = \sup\{|f(t)-g(t)| \;:\; t\in I\}$, and the identity map gives a point $i\in C(I,I)$. We can thus define a continuous map $u\colon X\to I$ by $u(x) = d(i,v^{\#}(x))$. We claim that (u,h) represents W as an NDR in X.

(a) If $x\in W$ then for all t we have $(v(t,x),h(t,x)) = r(t,x) = (t,x)$, so $v(t,x) = t$; this gives $u(x) = 0$. Conversely, suppose that $x\notin W$. This means that the point $r(1,x) = (1,x)$ does not lie in the closed set $I\times W$, so by continuity there exists $t<1$ such that $r(t,x)\notin(I\times W)$. However, we have $r(t,x)\in(I\times W)\cup(1\times X)$, so we must have $r(t,x)\in(1\times X)$, so $v(t,x) = 1\neq t$, so $u(x)>0$. This shows that $u^{-1}\{0\} = W$.
(b) We know that r is a retraction onto a subspace containing $1\times X$, and it follows immediately that $h_1 = 1_X$.
(c) We know that r is a retraction onto a subspace containing $I\times W$, and it follows immediately that $h_t|_W = 1_W$ for all t.
(d) Note that
$$u(x) = \sup\{v(t,x)-t \;:\; t\in I\} \geq v(0,x),$$
so if $u(x)<1$ then $v(0,x)<1$ so $r(0,x)\notin 1\times X$ so we must have $r(0,x)\in I\times W$, so $h(0,x)\in W$.

□

COROLLARY 33.22. [cor-cof-smashout]
A smashout of cofibrations is a cofibration.

PROOF. Let $j\colon W\to X$ and $k\colon Y\to Z$ be cofibrations. Then we can think of j and k as inclusions of subspaces, and their smashout is just the inclusion $W\times Z\cup X\times Y\to X\times Z$, so the claim follows from Proposition 33.20. □

PROPOSITION 33.23. [prop-acf-dr]
A map $j\colon W\to X$ is an acyclic cofibration if and only if it is a closed inclusion and jW is a DR of X.

PROOF. By proposition 33.21, we may assume that j is the inclusion of a closed subspace and that W is an NDR of X, represented by (u,h) say. If W is a DR we may assume that $h_1(X) = W$, and it is easy to check that $h_1\colon X\to W$ is a homotopy inverse for j, so that j is an acyclic cofibration.

For the converse, suppose that j is an acyclic cofibration. We then have a homotopy inverse $f\colon X\to W$ with $fj\simeq 1_W$ and $jf\simeq 1_X$. After extending the homotopy $fj\simeq 1_W$ over X (using the homotopy extension property of cofibrations) we may assume that $fj = 1_W$. Let $g_t\colon X\to X$ be a homotopy with $g_0 = 1_X$ and $g_1 = jf$. Define $P = \{0,1\}\times X\cup I\times W$ and $Q = I\times X$. It is easy to see that $\{0,1\}\subset I$ is an NDR, so Proposition 33.20 tells us that P is an NDR of Q, and thus that $1\times Q\cup I\times P$ is a retract of $I\times Q = I^2\times X$.

We define a map $h\colon 1\times Q\cup I\times P\to X$ by

$$\begin{aligned}
h(s,0,x) &= g(s,jf(x)) \\
h(1,t,x) &= g(1-t,x) \\
h(s,1,x) &= x \\
h(s,t,w) &= g(s(1-t),j(w)) && \text{for } w\in W.
\end{aligned}$$

Note that the first and second clauses are consistent because $g_0=jf$ and $fj=1_W$ so $g_0jf=jfjf=jf$. All other consistency checks are left to the reader. Because $1\times Q\cup I\times P$ is a retract of $I^2\times X$, we can extend h over all of $I^2\times X$ (just compose with the retraction). Having done this, we define $k(t,x)=h(0,t,x)$, so that $k\colon I\times X\to X$. We find that $k(1,x)=x$ for all x, that $k(t,w)=w$ for all t and all $w\in W$, and that $k(0,x)=f(x)\in W$ for all x. It follows that (u,k) represents W as a DR of X. □

Corollary 33.24. [`cor-acf-smashout`]
The smashout of a cofibration and an acyclic cofibration is an acyclic cofibration.

Proof. This is immediate from Propositions 33.20 and 33.23. □

Proposition 33.25. [`prop-acf-retract`]
If $j\colon W\to X$ is an acyclic cofibration then there is a diagram

$$\begin{tikzcd}
W \arrow[r, tail, "j"] \arrow[d, tail, "j"'] & X \arrow[r, two heads, "r"] \arrow[d, tail, "i_1"] & W \arrow[d, tail, "j"] \\
X \arrow[r, tail, "k"'] & I\times X \arrow[r, two heads, "s"'] & X,
\end{tikzcd}$$

in which $rj=1_W$ and $sk=1_X$. In other words, the map j is a retract of the map i_1.

Proof. We may assume that W is a closed subspace of X. Choose (u,h) representing W as a DR of X. Define $g\colon I\times X\to X$ by $g(t,x)=h(\max(t/ux,1),x)$. This is clearly continuous on $I\times(X\setminus W)$. Suppose that $(t,w)\in I\times W$ (so that $g(t,w)=w$) and that U is an open neighbourhood of w in X. Write $U'=\{x\in X\ :\ h(I\times\{x\})\subseteq U\}$. As in the proof of Proposition 33.20, we see that this is an open neighbourhood of w. Clearly $g(I\times U')\subseteq U$, and thus g is continuous at (t,w). This shows that g is continuous everywhere. One can check that (u,g) represents W as a DR of X, and that $g(t,x)=x$ whenever $t\geq ux$.

Now define

$$\begin{aligned}
k(x) &= (1-u(x),x) \\
r(x) &= g(0,x) \\
s(t,x) &= g(1-t,x).
\end{aligned}$$

One can check that the diagram commutes. □

33.4. Further orthogonality properties.

Corollary 33.26. [`prop-acf-perp-fib`]
If $j\colon W\to X$ is an acyclic cofibration and $q\colon E\to B$ is a fibration then j is left orthogonal to q.

Proof. Suppose we are given a diagram of the following form:

$$\begin{tikzcd}
W \arrow[r, "f"] \arrow[d, "j"'] & E \arrow[d, "q"] \\
X \arrow[r, "g"'] & B.
\end{tikzcd}$$

Choose maps k, r, s as in Proposition 33.25. The homotopy lifting property tells us that there is a map $h\colon I \times X \to E$ making the following diagram commute:

$$\begin{tikzcd} X \arrow[r, "fr"] \arrow[d, "i_1"'] & E \arrow[d, "q"] \\ I \times X \arrow[ur, "h"] \arrow[r, "gs"'] & B. \end{tikzcd}$$

It follows that we can add the map $hk\colon X \to E$ to the original diagram and it will still commute. This means that j is orthogonal to q. □

Proposition 33.27. [prop-afb-over]
If $q\colon E \to B$ is an acyclic fibration then it is homotopy equivalent over B to B.

Proof. As q is a homotopy equivalence, there is a map $e\colon B \to E$ such that $qe \simeq 1_B$ and $eq \simeq 1_E$. After lifting the homotopy $qe \simeq 1_B$ (using the homotopy lifting property of fibrations) we may assume that $qe = 1_B$. Choose a homotopy $g_t\colon E \to E$ with $g_0 = 1_E$ and $g_1 = eq$. Write $J = \{1\} \times I \cup I \times \{0,1\} \subset I^2$, and define maps $n\colon J \times E \to E$ and $m\colon I^2 \times E \to B$ by

$$\begin{aligned} n(s,0,x) &= eqg(s,x) \\ n(1,t,x) &= g(1-t,x) \\ n(s,1,x) &= x \\ m(s,t,x) &= qg(s(1-t),x). \end{aligned}$$

One can check that the following diagram commutes:

$$\begin{tikzcd} J \times E \arrow[r, "n"] \arrow[d, tail] & E \arrow[d, "q"] \\ I^2 \times E \arrow[r, "m"'] & B. \end{tikzcd}$$

The inclusion $J \times E \rightarrowtail I^2 \times E$ is an acyclic cofibration, so it is orthogonal to q by Proposition 33.26. Thus, there is a map $l\colon I^2 \times E \to E$ filling in the square. Define $h\colon I \times E \to E$ by $h(t,x) = l(0,t,x)$. Then $h(0,x) = eq(x)$ and $h(1,x) = x$ and $qh(t,x) = q(x)$, so h is a homotopy over B between eq and 1_E, as required. □

Corollary 33.28. [cor-afb-fibres]
If $q\colon E \to B$ is an acyclic fibration, then for each $b \in B$ the fibre $q^{-1}\{b\}$ is contractible.

Proof. **Prove this** □

Proposition 33.29. [prop-cof-perp-afb]
If $j\colon W \to X$ is a cofibration and $q\colon E \to B$ is an acyclic fibration, then j is left orthogonal to q.

Proof. Let e and h be as in the proof of Proposition 33.27. Given a diagram of the form

$$\begin{tikzcd} W \arrow[r, "f"] \arrow[d, "j"'] & E \arrow[d, "q"] \\ X \arrow[r, "g"'] & B, \end{tikzcd}$$

we consider the diagram

$$\begin{tikzcd} I \times W \cup 0 \times X \arrow[r, "k"] \arrow[d, tail] & E \arrow[d, "q"] \\ I \times X \arrow[r, "g \circ \mathrm{proj}"'] & B. \end{tikzcd}$$

The left hand vertical map is an acyclic cofibration, which is orthogonal to q by Proposition 33.26. There is thus a map $l\colon I\times X\to E$ filling in the square. One can check that the map $x\mapsto l(1,x)$ fills in the original square. $\square$

Proposition 33.30. [prop-cof-perp]
We have $\mathrm{acf}^{\perp}=\mathrm{fib}$ *and* $\mathrm{cof}^{\perp}=\mathrm{afb}$.

Proof. We have seen in Propositions 33.26 and 33.29 that acf $\perp$ fib and cof $\perp$ afb, so that $\mathrm{fib}\subseteq\mathrm{acf}^{\perp}$ and $\mathrm{afb}\subseteq\mathrm{cof}^{\perp}$. Suppose that $q\in\mathrm{acf}^{\perp}$. Recall that $\mathrm{Path}(q)=\{(\omega,e)\in\mathrm{Path}(B)\times E\ :\ \omega(1)=q(e)\}$. We define maps $f\colon\ \mathrm{Path}(q)\to E$ and $g\colon I\times\mathrm{Path}(q)\to B$ by $f(\omega,e)=e$ and $g(t,\omega,e)=\omega(t)$. This gives a diagram as follows:

$$\begin{tikzcd}
\mathrm{Path}(q) \arrow[r, "f"] \arrow[d, tail, "i_1"'] & E \arrow[d, "q"] \\
I\times\mathrm{Path}(q) \arrow[r, "g"'] & B.
\end{tikzcd}$$

As i_1 is clearly an acyclic cofibration, there is a map $m\colon I\times\mathrm{Path}(q)\to E$ filling in the square. One can check that the adjoint map $l=m^{\#}\colon\ \mathrm{Path}(q)\to\mathrm{Path}(E)$ (defined by $l(\omega,e)(t)=m(t,\omega,e)$) is a path-lifting function for q, so q is a fibration.

Now suppose that $q\in\mathrm{cof}^{\perp}$. The above shows that q is a fibration; we need to show that it is also a homotopy equivalence. By filling in the square on the left below, we get a map $e\colon B\to E$ with $qe=1_B$. We then fill in the right hand square (in which $g(0,x)=eq(x)$ and $g(1,x)=x$) to get a homotopy $h\colon I\times E\to E$ over B between eq and 1_E, as required.

$$\begin{tikzcd}
\emptyset \arrow[r, tail] \arrow[d, tail] & E \arrow[d, "q"] \\
B \arrow[r, "1"'] & B
\end{tikzcd}
\qquad
\begin{tikzcd}
\{0,1\}\times E \arrow[r, "g"] \arrow[d, tail] & E \arrow[d, "q"] \\
I\times E \arrow[r, "q\circ\mathrm{proj}"'] & B.
\end{tikzcd}$$

$\square$

Proposition 33.31. [prop-xmap-fib]
If $j\colon W\to X$ is a cofibration and $q\colon E\to B$ is a fibration then the crossmap $F(j,q)\colon C(X,E)\to C(W,E)\times_{C(W,B)}C(X,B)$ is a fibration. If j or q is acyclic then so is $F(j,q)$.

Proof. Let i be an acyclic cofibration. Then $i\square j$ is an acyclic cofibration and thus orthogonal to q, so $F(i\square j,q)$ is surjective. However, $F(i,F(j,q))=F(i\square j,q)$ so $F(i,F(j,q))$ is surjective, so i is orthogonal to $F(j,q)$. This holds for all $i\in\mathrm{acf}$, so $F(j,q)\in\mathrm{acf}^{\perp}=\mathrm{fib}$ as claimed. A similar argument shows that if $j\in\mathrm{acf}$ or $q\in\mathrm{afb}$ then $F(j,q)\in\mathrm{cof}^{\perp}=\mathrm{fib}$. $\square$

Corollary 33.32. [cor-restrict-fib]
If $j\colon W\to X$ is a cofibration and E is any space, then the restriction map $j^*\colon C(X,E)\to C(W,E)$ is a fibration. If j is acyclic then so is j^*.

Proof. Apply Proposition 33.31 to the map $q\colon E\to 0$. $\square$

Proposition 33.33. [prop-perp-fib]
We have $\mathrm{acf}={}^{\perp}\mathrm{fib}$ *and* $\mathrm{cof}={}^{\perp}\mathrm{afb}$.

Proof. We already know that $\mathrm{acf}\subseteq{}^{\perp}\mathrm{fib}$ and $\mathrm{cof}\subseteq{}^{\perp}\mathrm{afb}$. Suppose that $j\in{}^{\perp}\mathrm{afb}$. Let $f\colon I\times W\to\mathrm{Cyl}(j)$ and $g\colon X\to\mathrm{Cyl}(j)$ be the evident maps, so that $f(1,w)=gj(w)$. As usual, we write $f^{\#}\colon W\to$

Path Cyl(j) for the adjoint map, defined by $f^{\#}(w)(t) = f(t, w)$. This gives a commutative square as follows:

$$\begin{tikzcd} W \arrow[r, "f^{\#}"] \arrow[d, tail, "j"'] & \mathrm{Path}\,\mathrm{Cyl}(j) \arrow[d, "p_1"] \\ X \arrow[r, "g"'] & \mathrm{Cyl}(j). \end{tikzcd}$$

We know from Corollary 33.32 that the map p_1 is an acyclic fibration and thus is orthogonal to j, so there is a map $r^{\#} : X \to \mathrm{Path}\,\mathrm{Cyl}(j)$ filling in the square. One can check that the corresponding map $r : I \times X \to \mathrm{Cyl}(j)$ (defined by $r(t, x) = r^{\#}(t)(x)$) is a retraction, so that j is a cofibration.

Now suppose that $j \in {}^{\perp}\mathrm{fib}$. From the above, we know that j is a cofibration, and we need to show that it is also a homotopy equivalence. We first fill in the left hand diagram below to get a map $f : X \to W$ with $fj = 1_W$. We then define $c : W \to \mathrm{Path}(X)$ by $c(w)(t) = j(w)$, and apply Corollary 33.32 to the inclusion $\{0, 1\} \rightarrowtail I$ to see that $(p_0, p_1) : \mathrm{Path}(E) \to E \times E$ is a fibration. This means that we can fill in the right hand square below to get a map $h^{\#} : X \to \mathrm{Path}(X)$ whose adjoint is a homotopy between 1_X and fj under W. This shows that j is a homotopy equivalence, as claimed. □

33.5. Measured paths. So far we have dealt with paths $u : [0, 1] \to X$, which we can think of as having length one. It is sometimes convenient to work instead with paths of variable length (which will allow us to avoid awkward reparametrisation when we join paths together). The relevant definitions are as follows.

Definition 33.34. [`defn-mpath`]
Let X be a space. We put

$$\mathrm{MPath}(X) = \{(a, u) \;:\; a \in \mathbb{R}_+,\ u \in C(\mathbb{R}_+, X),\ u(t) = u(a) \text{ for all } t \geq a\}.$$

There is a closed subspace of $\mathbb{R}_+ \times C(\mathbb{R}_+, X)$, and we give it the subspace topology. The elements of $\mathrm{MPath}(X)$ are called *measured paths*. We also define continuous maps $\sigma, \tau : \mathrm{MPath}(X) \to X$ and $\lambda : \mathrm{MPath}(X) \to \mathbb{R}_+$ by $\sigma(a, u) = u(0)$ and $\tau(a, u) = u(a)$ and $\lambda(a, u) = a$. We call these the *source*, *target* and *length* of (a, u), respectively. We put

$$\begin{aligned} \mathrm{MPath}(X)(x, y) &= \{ \text{ measured paths in } X \text{ from } x \text{ to } y \ \} \\ &= \{(a, u) \in \mathrm{MPath}(X) \;:\; \sigma(a, u) = x,\ \tau(a, u) = y\}. \end{aligned}$$

Remark 33.35. [`rem-mpath`]
For a slightly different picture, we can put

$$\mathrm{MPath}'(X) = \{(r, \omega) \;:\; r \in \mathbb{R}_+,\ \omega \in C([0, r], X)\}.$$

It is clear that there is a bijection $\mathrm{MPath}'(X) \to \mathrm{MPath}(X)$ given by $(a, u) \mapsto (a, u \circ \min(-, a))$. Conceptually it is more natural to work with $\mathrm{MPath}'(X)$ but to define the topology and verify that various constructions are continuous it is easier to use the original definition.

Remark 33.36. [`rem-mpath-join`]
Suppose we have a measured path (a, u) from x to y, and a measured path (b, v) from y to z. We define $(b, v) * (a, u) = (a + b, w)$, where

$$w(t) = \begin{cases} u(t) & \text{if } 0 \leq t \leq a \\ v(t - a) & \text{if } a \leq t. \end{cases}$$

This gives a measured path from x to z, called the *join* of (a, u) and (b, v). It is not hard to see that this gives a continuous operation

$$* : \mathrm{MPath}(X) \times_X \mathrm{MPath}(X) \to \mathrm{MPath}(X)$$

that is strictly associative (not just up to homotopy). Moreover, if we put $1_x = (0, c_x)$ (where $c_x : \mathbb{R}_+ \to X$ is the constant map with value x) then we have $u = u * 1_x = 1_y * u$ for all $u \in \mathrm{MPath}(X)(x, y)$. We therefore have a category whose objects are the points of X and whose morphisms are the measured paths.

We can compare measured and unmeasured paths as follows:

Proposition 33.37. [prop-mpath-path]
There are continuous maps

$$\begin{aligned} &i\colon \mathrm{Path}(X) \to \mathrm{MPath}(X) && p\colon \mathrm{MPath}(X) \to \mathrm{Path}(X) \\ &h\colon [0,1] \times \mathrm{Path}(X) && k\colon [0,1] \times \mathrm{MPath}(X) \to \mathrm{MPath}(X) \end{aligned}$$

such that

$$\begin{aligned} h(0,u) &= pi(u) \\ h(1,u) &= u \\ k(0,(a,v)) &= ip(a,v) \\ k(1,(a,v)) &= (a,v) \\ \sigma(h(t,u)) &= \sigma(i(u)) = \sigma(u) \\ \sigma(k(t,(a,v))) &= \sigma(p(a,v)) = \sigma(a,v) \\ \tau(h(t,u)) &= \tau(i(u)) = \tau(u) \\ \tau(k(t,(a,v))) &= \tau(p(a,v)) = \tau(a,v). \end{aligned}$$

Specifically, we can take

$$\begin{aligned} i(u) &= (1, i_1(u)) & i_1(u)(s) &= u(\min(s,1)) \\ p(a,v)(s) &= v(s+as) \\ h(t,u)(s) &= u(\min(2s-ts,1)) \\ k(t,(a,v)) &= (1-t+at, k_1(t,(a,v))) & k_1(t,(a,v))(s) &= v(s+as-ast). \end{aligned}$$

Proof. Almost all of the claimed identities follow directly from the formulae. We will just explain two points that are less obvious. Firstly, to check that $k(0,(a,v)) = ip(a,v)$, we need to know that $v(s+as) = v(\min(s,1) + a\min(s,1))$ for all $(a,v) \in \mathrm{MPath}(X)$ and $s \geq 0$. This is clear when $s \leq 1$. If $s \geq 1$ then the claim is that $v(s+as) = v(1+a)$, and this again holds because $s+as, 1+a \geq a$ and v is constant on $[a,\infty)$ by assumption. Secondly, we need to show that $\tau(k(t,(a,v))) = \tau(a,v)$, or equivalently that $v((1-t+at)(1+a-at)) = v(a)$. As v is constant on $[a,\infty)$, it will suffice to check that the number $b = (1-t+at)(1+a-at) - a$ is nonnegative (for all $a \in \mathbb{R}_+$ and $t \in [0,1]$). One can check directly that

$$b = (1-t)(1+a^2t-at) = (1-t)(1+((a-\tfrac{1}{2})^2 - \tfrac{1}{4})t) \geq (1-t)(1-t/4) \geq 0$$

as required.

The only other thing to check is that the maps i, p, h and k are continuous. We will prove this for k; the other cases can be handled in a similar way and are left to the reader. As the first component of k is clearly continuous, we need only discuss k_1. We can define a map

$$k_1\colon [0,1] \times (\mathbb{R}_+ \times C(\mathbb{R}_+, X)) \to C(\mathbb{R}_+, X)$$

by the same formula as above, and it will be enough to check that this extended map is continuous. This is adjoint to the map

$$K\colon \mathbb{R}_+ \times [0,1] \times \mathbb{R}_+ \times C(\mathbb{R}_+, X) \to X$$

given by $K(s,t,a,v) = v(s+as-ast)$. We can define a continuous map $L\colon \mathbb{R}_+ \times [0,1] \times \mathbb{R}_+ \to \mathbb{R}_+$ by $L(s,t,a) = s+as-ast$ and then K is the composite

$$\mathbb{R}_+ \times [0,1] \times \mathbb{R}_+ \times C(\mathbb{R}_+, X) \xrightarrow{L\times 1} \mathbb{R}_+ \times C(\mathbb{R}_+, X) \xrightarrow{\mathrm{ev}} X.$$

This shows that K is continuous, so k_1 is also continuous, as required. □

Corollary 33.38. [cor-mpath-path]
For all x and y in X, the space $\mathrm{MPath}(X)(x,y)$ is homotopy equivalent to $\mathrm{Path}(X)(x,y)$. Thus, we have $\Pi_1(X)(x,y) = \pi_0\,\mathrm{MPath}(X)(x,y)$.

Proof. Because the maps i, p, h and k commute with σ and τ, they restrict to give the required maps and homotopies. □

Definition 33.39. [defn-hurewicz-fibration]
Consider a continuous map $q\colon E \to B$.

(a) Recall that
$$\mathrm{Path}(q) = \mathrm{Path}(B) \times_B E = \{(u,e) \in \mathrm{Path}(B) \times E \ :\ u(0) = q(e)\}.$$

(b) Analogously, we write
$$\mathrm{MPath}(q) = \mathrm{MPath}(B) \times_B E = \{(a,u,e) \in \mathrm{MPath}(B) \times E \ :\ u(0) = q(e)\}.$$

(c) Recall that a *lifting function* for q is a continuous map $l\colon\ \mathrm{Path}(q) \to \mathrm{Path}(E)$ such that $q(l(u,e)(s)) = u(s)$ for all $s \in [0,1]$ and $l(u,e)(0) = e$. Equivalently, we must have
$$q_* \circ l = \pi_0 \colon\ \mathrm{Path}(B) \times_B E \to \mathrm{Path}(B)$$
$$\sigma \circ l = \pi_1 \colon\ \mathrm{Path}(B) \times_B E \to E.$$

(d) Analogously, *measured lifting function* for q is a continuous map $m\colon\ \mathrm{MPath}(q) \to \mathrm{MPath}(E)$ such that
$$q_* \circ l = \pi_0 \colon\ \mathrm{MPath}(B) \times_B E \to \mathrm{MPath}(B)$$
$$\sigma \circ l = \pi_1 \colon\ \mathrm{MPath}(B) \times_B E \to E.$$
Equivalently, m must have the form $m(a,v,e) = (a, m_1(a,v,e))$ for some $m_1\colon\ \mathrm{MPath}(q) \to C(\mathbb{R}_+, E)$ with $q(m_1(a,v,e)(s)) = v(s)$ for all $s \in \mathbb{R}_+$ and $m_1(a,v,e)(0) = e$.

Proposition 33.40. [prop-measured-lifting]
The map q is a Hurewicz fibration iff there is a lifting function for q iff there is a measured lifting function for q.

Proof. The first iff is just the definition of a Hurewicz fibration, recorded as a reminder. The real point is the second iff.

If $m(a,u,e) = (a, m_1(a,u,e))$ is a measured lifting function then we can define a lifting function by $l(u,e) = m_1(1,u,e)$. Conversely, suppose we start with a lifting function l. We then define
$$m_1(a,v,e)(s) = l(p(a,v),e)(\min(s,a)/(1+a))$$
$$m(a,v,e) = (a, m_1(a,v,e)).$$
This has $m_1(a,v,e)(0) = l(p(a,v),e)(0) = e$ and
$$q(m_1(a,v,e)(s)) = q(l(p(a,v),e)(\min(s,a)/(1+a))) = p(a,v)(\min(s,a)/(1+a))) = v(\min(s,a)) = v(s).$$
It follows that m is a measured lifting function. $\square$

33.6. Schedules. From now on, we suppose that we have a set I and an open cover $(B_i)_{i\in I}$ for B. We also put $E_i = q^{-1}(B_i)$, and let $q_i\colon E_i \to B_i$ be the restriction of q. Suppose that we are given measured lifting functions $m_i\colon\ \mathrm{MPath}(B_i) \times_{B_i} E_i \to \mathrm{MPath}(E_i)$ for all i. We would like to combine these to construct a measured lifting function $m\colon\ \mathrm{Path}(B) \times_B E \to \mathrm{Path}(E)$. The basic idea is clear: given a measured path $(a,v) \in \mathrm{MPath}(B)$ and a lift $e \in E$ of $v(0)$, we can break v into shorter paths each of which is contained in some set B_i, then use the functions m_i to lift these shorter paths in such a way that they can be rejoined to give a lift of v. The problem is that we need to choose the breakpoints and the indices i in such a way that the lifted path depends continuously on the original data. To make this work, we need two extra ingredients.

(a) We will assume that the index set I is well-ordered. If it does not have an obvious well-ordering, we can use Theorem 35.27 to choose one.

(b) We need a partition of unity $(f_i)_{i\in I}$ subordinate to the cover $(B_i)_{i\in I}$. This means that the sets $C_i = \{b \in B \ :\ f_i(b) > 0\}$ are open with $\overline{C_i} \subseteq B_i$ and the family $(\overline{C_i})_{i\in I}$ is locally finite. It will be convenient to combine the maps f_i into a single map $f\colon I \times B \to [0,1]$ given by $f(i,b) = f_i(b)$.

Given these data, we can construct the combined lifting function m without any further arbitrary choices. However, we will need some preparatory theory, which we now start to develop.

Definition 33.41. [defn-schedule]
A *schedule* (with values in I) is a triple $s = (a, P, c)$, where:

(a) a is a nonnegative real number.
(b) P is an equivalence relation on the interval $(0,a]$ such that there are finitely many equivalence classes, and each class has the form $(x,y]$ for some x,y with $0 \leq x < y \leq a$.
(c) c is a function from $(0,a]$ to I that is constant on each equivalence class of P. (This means that there is an induced map $\overline{c}\colon (0,a]/P \to I$. We do not assume that this is injective.)

We write $\lambda(s) = a$ and call this the *length* of s. We also write $\mathrm{Sched}(I)$ for the set of all schedules. Note that there is a unique schedule of length zero.

We introduce a topology on $\mathrm{Sched}(I)$ as follows. First, we put

$$\mathrm{Sched}'(I) = \coprod_{n\in\mathbb{N}} (\mathbb{R}_+^n \times I^n) = \coprod_{n\in\mathbb{N}} \coprod_{\underline{i}\in I^n} \mathbb{R}_+^n,$$

and give this the obvious coproduct topology (which is locally compact Hausdorff and thus CGWH). Given $n \in \mathbb{N}$ and $\underline{a} = (a_1,\dots,a_n) \in \mathbb{R}_+^n$ and $\underline{i} = (i_1,\dots,i_n) \in I^n$ we have a schedule $s = \phi(n,\underline{a},\underline{i}) = (a,P,c)$ given by:

- $a = \sum_k a_k$
- The equivalence classes for P are the intervals $(b_k, b_{k+1}]$ (for $0 \leq k < n$), where $b_k = \sum_{j<k} a_j$.
- The map $c\colon (0,a] \to I$ takes the value i_k on the interval $(b_k, b_{k+1}]$.

This construction gives a surjective map

$$\phi\colon\ \mathrm{Sched}'(I) \to \mathrm{Sched}(I).$$

We declare that a set $F \subseteq \mathrm{Sched}(I)$ is closed iff $\phi^{-1}(F)$ is closed in $\coprod_n(\mathbb{R}_+^n \times I^n)$. Thus, ϕ is a quotient map with respect to this topology.

Proposition 33.42. *The space* $\mathrm{Sched}(I)$ *is CGWH.*

Proof. **Write this** □

Definition 33.43. [dfn-follows]
Suppose we have a measured path (a,v) and a schedule $s = (a,P,c)$ (with the same value of a). We say that v (or (a,v)) *follows* s if $v(t) \in \overline{C_{c(t)}}$ for all $t \in (0,a]$. We write $\mathrm{SPath}(B)$ for the set of quadruples (a,P,c,v) with this property, topologised as a subspace of $\mathrm{Sched}(I) \times \mathrm{MPath}(B)$. Elements of $\mathrm{SPath}(B)$ will be called *scheduled paths*. For $z = (a,P,c,v) \in \mathrm{SPath}(B)$ we put $\sigma(z) = \sigma(a,v) = v(0) \in B$ and $\tau(z) = \tau(a,v) = v(a) \in B$ and $\lambda(z) = a \in \mathbb{R}_+$ and $\pi(z) = (a,v) \in \mathrm{MPath}(B)$.

Remark 33.44. [rem-follows]
Suppose we have a point $z = (n,\underline{a},\underline{i}) \in \mathrm{Sched}'(I)$, and a measured path $(a',v) \in \mathrm{MPath}(B)$. Put $b_k = \sum_{j<k} a_j$ as usual. We see that (a',v) follows $\phi(z)$ iff $a' = \sum_k a_k$, and for all $k \in \{1,\dots,n\}$ we have ($a_k = 0$ or $v([b_k,b_{k+1}]) \subseteq \overline{C_k}$).

Proposition 33.45. [prop-spath-topology]
Put

$$\mathrm{SPath}'(B) = \{(z,a,v)\ :\ z \in \mathrm{Sched}'(I),\ (a,v) \in \mathrm{MPath}(B),\ \ \textit{and } (a,v) \textit{ follows } \phi(z)\}.$$

Then there is a pullback square as follows, in which the horizontal maps are quotient maps and the vertical maps are closed inclusions.

$$\begin{tikzcd}
\mathrm{SPath}'(B) \arrow[r, two heads, "(\phi\times 1)|_{\mathrm{SPath}'(B)}"] \arrow[d, tail] & \mathrm{SPath}(B) \arrow[d, tail] \\
\mathrm{Sched}'(I) \times \mathrm{MPath}(B) \arrow[r, two heads, "\phi\times 1"'] & \mathrm{Sched}(I) \times \mathrm{MPath}(B).
\end{tikzcd}$$

Proof. First, it is clear from the definitions that the vertical maps are injective, the horizontal maps are surjective, and the square is a pullback in the category of sets.

The map ϕ is a quotient map by the definition of the topology on $\mathrm{Sched}(I)$, so the bottom map $\phi \times 1$ is a quotient map by Proposition 23.32. Next, we claim that $\mathrm{SPath}'(B)$ is closed in $\mathrm{Sched}'(I) \times \mathrm{MPath}(B)$.

As $\mathrm{Sched}'(I)$ is a coproduct of copies of $\mathbb{R}_+^n$, it will suffice to check that the intersection of $\mathrm{SPath}'(B)$ with each copy of $\mathbb{R}_+^n \times \mathrm{MPath}(B)$ is closed. More explicitly, fix an integer $n \geq 0$ and a sequence $\underline{i} = (i_1, \ldots, i_n)$. We need to show that the set

$$Q_0 = \{(\underline{a}, a', v) \in \mathbb{R}_+^n \times \mathrm{MPath}(B) \;:\; (a', v) \text{ follows } \phi(n, \underline{a}, \underline{i})\}$$

is closed in $\mathbb{R}_+^n \times \mathrm{MPath}(B)$. It is clear that the larger set

$$Q_1 = \{(\underline{a}, a', v) \in \mathbb{R}_+^n \times \mathrm{MPath}(B) \;:\; a' = \sum_k a_k\}$$

is closed. Define $\alpha_k \colon Q_1 \to C(\mathbb{R}_+, X)$ by

$$\alpha_k(\underline{a}, a', v)(t) = v(\min(t, a_k) + \sum_{j<k} a_j).$$

By considering the adjoint map $Q_1 \times \mathbb{R}_+ \to X$, we see that this is continuous. We also know from Remark 23.22 that $C(\mathbb{R}_+, \overline{C_{i_k}})$ is closed in $C(\mathbb{R}_+, B)$. It follows that the set

$$Q_{2,k} = \alpha_k^{-1}(C(\mathbb{R}_+, \overline{C_{i_k}})) \cup \{(\underline{a}, a', v) \in Q_1 \;:\; a_k = 0\}$$

is closed in Q_1. We also see from Remark 33.44 that $Q_0 = \bigcap_{k=1}^n Q_{2,k}$, so Q_0 is closed in $\mathbb{R}_+^n \times \mathrm{MPath}(B)$, so $\mathrm{SPath}'(B)$ is closed in $\mathrm{Sched}'(I) \times \mathrm{MPath}(B)$ as claimed. Now $\mathrm{SPath}'(B) = (\phi \times 1)^{-1}(\mathrm{SPath}(B))$ and $\phi \times 1$ is a quotient map. We can thus conclude that $\mathrm{SPath}(B)$ is closed in $\mathrm{Sched}(I) \times \mathrm{MPath}(B)$. As we are topologising $\mathrm{SPath}(B)$ and $\mathrm{SPath}'(B)$ as subspaces of $\mathrm{Sched}(I) \times \mathrm{MPath}(B)$ and $\mathrm{Sched}'(I) \times \mathrm{MPath}(B)$, we see that the vertical maps are closed inclusions. It follows that the square is a pullback of spaces, and thus (by Proposition 23.48) that the top horizontal map is also a quotient map. ☐

Proposition 33.46. [`prop-scheduled-lifting`]
There is a canonical map $m\colon \mathrm{SPath}(B) \times_B E \to \mathrm{MPath}(E)$ with $\sigma(m(z, e)) = e$ and $q \circ m(z, e) = \pi(z)$. Thus, any map $\theta\colon \mathrm{MPath}(B) \to \mathrm{SPath}(B)$ with $\pi\theta = 1$ gives a measured lifting function $m \circ \theta$ for q.

Of course it is not obvious that any such map θ exists, but we will construct one below.

Proof. First, we will modify the lifting functions m_k slightly. Let Z be the space of measured paths of length zero in B. This is closed in $\mathrm{MPath}(B)$ and is homeomorphic to B. The space $\mathrm{MPath}(\overline{C_k})$ is also closed in $\mathrm{MPath}(B)$, and we put $\mathrm{MPath}_+(\overline{C_k}) = Z \cup \mathrm{MPath}(\overline{C_k})$. As $\overline{C_k} \subseteq B_k$ and lifting paths of length zero is trivial we see that there is an obvious way to define a lifting map $m_k\colon \mathrm{MPath}_+(\overline{C_k}) \to \mathrm{MPath}(E)$. This is continuous on Z and on $\mathrm{MPath}(\overline{C_k})$, and these sets are both closed, so m_k is continuous on $\mathrm{MPath}_+(\overline{C_k})$.

Now consider a point $(n, \underline{a}, \underline{i}, a', v) \in \mathrm{SPath}'(B)$ and an element $e_0 \in E$ with $q(e_0) = v(0)$. Put $(a, P, c) = \phi(n, \underline{a}, \underline{i}) \in \mathrm{Sched}(I)$. Write $b_k = \sum_{j<k} a_j$ so the equivalence classes for P are the intervals $(b_k, b_{k+1}]$ (except that some of these intervals may be empty, in which case they do not count as equivalence classes). We then define $v_k \colon \mathbb{R}_+ \to B$ by $v_k(t) = v(\min(b_k + t, b_{k+1}))$, so $(a_k, v_k) \in \mathrm{MPath}(B)$. Moreover, as v follows (a, P, c) we must have $v_k((0, a_k]) \subseteq \overline{C_k}$. If $a_k > 0$ then it follows easily that $(a_k, v_k) \in \mathrm{MPath}(\overline{C_k})$, and if $a_k = 0$ then $(a_k, v_k) \in Z$, so in all cases we have $(a_k, v_k) \in \mathrm{MPath}_+(\overline{C_k})$. We now define points $e_k \in E$ and $(a_k, u_k) \in \mathrm{MPath}(E)$ recursively by $(a_k, u_k) = m_k(a_k, v_k, e_{k-1})$ and $e_k = u_k(a_k)$. We then define $m'(n, \underline{a}, \underline{i}, a', v, e_0)$ to be the join of all the measured paths (a_k, u_k), which has the form (a, u) for some $u\colon \mathbb{R}_+ \to X$. This gives a function $m'\colon \mathrm{SPath}'(B) \times_B E \to \mathrm{MPath}(E)$. It is clear that any terms with $a_k = 0$ can be omitted without changing u, so this induces a well-defined function $m\colon \mathrm{SPath}(B) \times_B E \to \mathrm{MPath}(E)$ with $m(a, P, c, v, e_0) = (a, u)$. It is also clear that we have $q \circ u = v$ and $u(0) = e_0$. We thus have a lifting function as claimed, provided that m is continuous.

Recall that the map $\mathrm{SPath}'(B) \to \mathrm{SPath}(B)$ is a quotient map. It follows using Proposition 23.48 that the induced map $\mathrm{SPath}'(B) \times_B E \to \mathrm{SPath}(B) \times_B E$ is also a quotient map. Because of this, it will suffice to check that m' is continuous. Now fix $n \in \mathbb{N}$ and $\underline{i} \in I^n$, and put

$$Q = \{(t, \underline{a}, a', v, e_0) \in \mathbb{R}_+ \times \mathbb{R}_+^n \times \mathrm{MPath}(B) \times_B E \;:\; (a', v) \text{ follows } \phi(n, \underline{a}, \underline{i})\}.$$

Define $q\colon Q \to X$ by $q(t, \underline{a}, a', v) = u(t)$, where u is as constructed above. Thus, q is a restriction of the adjoint of the second component of m. By a straightforward argument, it will suffice to show that q is continuous. We now put

$$Q_k = \{(t, \underline{a}, a', vv, e_0) \in Q \;:\; b_k \leq t \leq b_{k+1}\}$$

for $1 \leq k \leq n$, and

$$Q_{n+1} = \{(t, \underline{a}, a', vv, e_0) \in Q \;:\; a \leq t\}.$$

This gives a finite list of closed subspaces whose union is Q, so it will suffice to prove that $q|_{Q_k}$ is continuous for all k. This restriction is built in a straightforward way from the maps m_k, so continuity is clear. □

33.7. Numerably local fibrations are fibrations. This section should be removed in favour of the previous one.

In this section we use the concept of a numerable open covering, as introduced in Section 22. We recall from that section that most commonly considered spaces are paracompact, and in that context, every open covering is numerable.

Definition 33.47. We say that a map $q\colon E \to B$ is a *numerably local fibration* if there is a numerable covering $B = \bigcup_{i\in I} B_i$ such that the maps $q\colon q^{-1}B_i \to B_i$ are all fibrations.

Definition 33.48. We say that a map $q\colon E \to B$ is a *numerable fibre bundle* if there is a numerable covering $B = \bigcup_{i\in I} B_i$ and homeomorphisms $f_i\colon B_i \times F_i \to q^{-1}B_i$ (for some collection of spaces F_i) such that $qf_i(b, x) = b$. It is easy to see that a numerable fibre bundle is a numerable fibration.

We now prove the following result:

Theorem 33.49. [`thm-loc-fibn`]
A numerable local fibration is a fibration. In particular, a numerable fibre bundle is a fibration. □

For the rest of these notes, we assume that $q\colon E \to B$ is a numerable local fibration, with a numerable cover $\{B_i\}$ as above and given functions $f_i\colon B \to I$ with $B_i = f_i^{-1}(0, 1]$. We will prove a number of lemmas and then prove the theorem at the end of the notes.

Definition 33.50. Given a subset $W \subseteq \mathrm{Path}(B)$, we define

$$F(W) = \{(\omega, e) \in U \times E \;:\; \omega(1) = q(e)\} \subseteq \mathrm{Path}(q)$$
$$G(W) = \{(\omega, t, e) \in U \times I \times E \;:\; \omega(t) = q(e)\}.$$

A *lifting function* over W is a map $l\colon F(W) \to \mathrm{Path}(E)$ such that $l(\omega, e)(1) = e$ and $q \circ l(\omega, e) = \omega$. An *extended lifting function* for W is a map $m\colon G(W) \to \mathrm{Path}(E)$ such that $m(\omega, t, e)(t) = e$ and $q\circ m(\omega, t, e) = \omega$.

Thus, $q\colon E \to B$ is a fibration if and only if there is a lifting function over the whole of $\mathrm{Path}(B)$.

Lemma 33.51. *Let B' be a subspace of B and $E' = q^{-1}B'$. Then the following are equivalent:*

(a) *$q\colon E' \to B'$ is a fibration.*
(b) *There is a lifting function over* $\mathrm{Path}(B')$.
(c) *There is an extended lifting function over* $\mathrm{Path}(B')$.

Proof. It is immediate from the definitions that (a)⇔(b). If m is an extended lifting function over $\mathrm{Path}(B')$ then $l(\omega, e) = m(\omega, 1, e)$ is a lifting function, so (c)⇒(b). Conversely, suppose that l is a lifting function. Suppose that $(\omega, t, e) \in G(\mathrm{Path}(B'))$, so $\omega\colon I \to B'$ and $q(e) = \omega(t)$. We then define

$$\omega_+(s) = \omega(\max(s - 1 + t, 0))$$
$$\omega_-(s) = \omega(\min(1 + t - s, 1)).$$

Thus ω_+ sits at $\omega(0)$ for a while then runs forwards to reach $\omega(t)$ when $s = 1$, and ω_- sits at $\omega(1)$ for a while and then runs backwards to reach $\omega(t)$ when $s = 1$. We next define

$$m(\omega, t, e)(s) = \begin{cases} 0 \leq s \leq t & l(\omega_+, e)(s + 1 - t) \\ t \leq s \leq 1 & l(\omega_-, e)(1 + t - s). \end{cases}$$

One can check that this gives an extended lifting function. □

It follows that we can choose extended lifting functions m_i over $\mathrm{Path}(B_i)$.

Definition 33.52. Let $\underline{i} = (i_1, \ldots, i_r)$ be a sequence of indices. We write $\text{len}(\underline{i}) = r$ and

$$W_{\underline{i}} = \{\omega \colon I \to B \;:\; \omega([(j-1)/r, j/r]) \subseteq B_{i_r} \text{ for } 1 \leq j \leq r\}.$$

We also define $f_{\underline{i}} \colon \text{Path}(B) \to I$ by

$$f_{\underline{i}}(\omega) = \min_{1 \leq j \leq r} \min_{(j-1)/r \leq t \leq j/r} f_{i_j}(\omega(t)).$$

One can check that this is a continuous map $\text{Path}(B) \to I$ and that $f_{i_j}(\omega) > 0$ if and only if $\omega \in W_{\underline{i}}$.

Lemma 33.53. *The sets $W_{\underline{i}}$ form an open cover of* $\text{Path}(B)$.

Proof. It is easy to see that $W_{\underline{i}}$ is open. Suppose that $\omega \in \text{Path}(B)$. For each $t \in I$ there is an index i_t and a number $\epsilon_t > 0$ such that $(t - \epsilon_t, t + \epsilon_t) \cap [0,1] \subseteq \omega^{-1} B_{i_t}$. The open sets $U_t = (t - \epsilon_t/2, t + \epsilon_t/2)$ cover $[0,1]$, so $[0,1] = U_{t_1} \cup \ldots \cup U_{t_n}$ for some finite sequence $t_1, \ldots, t_n$. Write $\epsilon = \min(\epsilon_{t_1}, \ldots, \epsilon_{t_n})$. It is not hard to check that for any open interval (a,b) of length at most ϵ, we have $\omega((a,b)) \subseteq B_i$ for some i. Thus if we choose r such that $1/r < \epsilon$ then there is a sequence $\underline{i} = (i_1, \ldots, i_r)$ such that $\omega([(j-1)/r, j/r]) \subseteq B_{i_j}$ for all j and thus $\omega \in W_{\underline{i}}$. □

Lemma 33.54. *There is an extended lifting function $m_{\underline{i}}$ over $W_{\underline{i}}$.*

Proof. Consider a point $(\omega, t, e) \in G(W_{\underline{i}})$, so that $\omega \in W_{\underline{i}}$ and $q(e) = \omega(t)$. Write $r = \text{len}(\underline{i})$ and let n be an integer such that $(n-1)/r \leq t \leq n/r$. There is only one such n unless t has the form m/r in which case n could be m or $m+1$; one can check that the constructions below do not depend on which we take in that case. For $1 \leq j \leq r$ we define a path $\omega_j \colon I \to B_{i_j}$ by

$$\omega_j(s) = \omega(\max(\min(s, (j-1)/r), j/r),$$

so that ω_j is constant at $\omega((j-1)/r)$ until $s = (j-1)/r$, then it moves to $\omega(j/r)$ when $s = j/r$, then it sits there until $s = 1$. We define points $e_j \in E$ for $j = 0, \ldots, r$ inductively by

$$\begin{aligned}
e_n &= m_{i_n}(\omega_n, t, e)(n/r) \\
e_{n-1} &= m_{i_n}(\omega_n, t, e)((n-1)/r) \\
e_{j+1} &= m_{i_{j+1}}(\omega_{j+1}, j/r, e_j)((j+1)/r) \qquad \text{for } n \leq j < r \\
e_{j-1} &= m_{i_{j-1}}(\omega_{j-1}, j/r, e_j)((j-1)/r) \qquad \text{for } 0 < j < n.
\end{aligned}$$

We then define $\alpha \colon I \to E$ by

$$\alpha(s) = \begin{cases} m_{i_j}(\omega_j, j/r, e_j)(s) & \text{if } s \in [(j-1)/r, r] \text{ and } j < n \\ m_{i_n}(\omega_n, t, e)(s) & \text{if } s \in [(n-1)/r, r] \\ m_{i_j}(\omega_j, (j-1)/r, e_{j-1})(s) & \text{if } s \in [(j-1)/r, r] \text{ and } n < j. \end{cases}$$

One can check that this is well-defined and continuous, and that the assignment $m_{\underline{i}}(\omega, t, e) = \alpha$ gives an extended lifting function for $W_{\underline{i}}$. □

We have now covered $\text{Path}(B)$ by open sets $W_{\underline{i}}$ such that there is an extended lifting function defined over each $\underline{i}$. We next need a way to glue together (extended) lifting functions over open sets V and W to get one over $V \cup W$. We first make a preliminary definition which does something a bit weaker.

Definition 33.55. Let W be a subset of $\text{Path}(B)$, let l be a lifting function over W and let m be an extended lifting function over W. For each $s \in [0,1]$ we define a "merged" lifting function by

$$M_s(l, m)(\omega, e)(t) = \begin{cases} l(\omega, e)(t) & \text{if } s \leq t \leq 1 \\ m(\omega, s, l(\omega, e)(s))(t) & \text{if } 0 \leq t \leq s. \end{cases}$$

In other words, we lift the section of ω from $t = s$ up to $t = 1$ using l. This gives us a lift $l(\omega, e)(s)$ of $\omega(s)$ and we feed this into m to lift the section of ω from $t = 0$ up to $t = s$. It is easy to check that $M_s(l, m)$ is indeed a lifting function. Moreover, $M_0(l, m) = l$ and $M_1(l, m) = m(-, 1, -)$.

Lemma 33.56. [lem-combine]
Let V, W be open subsets of $\mathrm{Path}(B)$*. Suppose that we have functions $g, f\colon \mathrm{Path}(B) \to I$ such that $g^{-1}(0,1] = V$ and $f^{-1}(0,1] = W$. Suppose that we also have a lifting function l over V and an extended lifting function m over W. Then there is a lifting function over $V \cup W$ which agrees with l over $V \setminus W$.*

Proof. Define $h\colon V \cup W \to I$ by $h = f/(f+g)$, and note that $\omega \in V$ if and only if $h(\omega) < 1$ and $\omega \in W$ if and only if $h(\omega) > 0$. Thus if $h(\omega) \leq 1/3$ we have $\omega \in V$, if $1/3 \leq h(\omega) \leq 2/3$ we have $h(\omega) \in U \cap V$, and if $h(\omega) \geq 2/3$ we have $\omega \in W$. We can thus define a lifting function over $V \cup W$ by

$$n(\omega, e) = \begin{cases} l(\omega, e) & \text{if } 0 \leq h(\omega) \leq 1/3 \\ M_{3h(\omega)-1}(l, m)(\omega, e) & \text{if } 1/3 \leq h(\omega) \leq 2/3 \\ m(\omega, 1, e) & \text{if } 2/3 \leq h(\omega) \leq 1. \end{cases}$$

If $\omega \in V \setminus W$ then $h(\omega) = 0$ and so $n(\omega, e) = l(\omega, e)$ as claimed. $\square$

This gives us a lifting function defined over any finite union of sets of the form $W_{\underline{i}}$. Unfortunately, we need to work a bit harder to get a lifting function defined over an infinite union of such sets.

Lemma 33.57. *For any r, the collection of sets $\{W_{\underline{i}} \;:\; \mathrm{len}(\underline{i}) = r\}$ is locally finite.*

Proof. Consider a path $\omega \in \mathrm{Path}(B)$. We need to produce a neighbourhood V of ω in $\mathrm{Path}(B)$ such that the set $S = \{(i_1, \dots, i_r) \;:\; V \cap W_{\underline{i}} \neq \emptyset\}$ is finite. We know that $\{B_i\}$ is locally finite, so for each $j = 1, \dots, r$ there is a neighbourhood V_j of $\omega(j/r)$ such that the set $S_j = \{i \;:\; V_j \cap B_i \neq \emptyset\}$ is finite. Write $V = \{\alpha \in \mathrm{Path}(B) \;:\; \alpha(j/r) \in V_j \text{ for } j = 1, \dots, r\}$. This is a neighbourhood of ω, and it is easy to see that for this V we have $S \subseteq S_1 \times \ldots \times S_r$, so S is finite as required. $\square$

Corollary 33.58. *The function $g_r(\omega) = r \sum_{\mathrm{len}(\underline{i}) < r} f_{\underline{i}}(\omega)$ is finite and continuous.*

Definition 33.59. We now define

$$f'_{\underline{i}} = \min(\max(0, f_{\underline{i}} - g_{\mathrm{len}(\underline{i})}), 1)\colon \mathrm{Path}(B) \to I$$

and

$$W'_{\underline{i}} = \{\omega \;:\; f'_{\underline{i}}(\omega) > 0\}.$$

It is easy to check that $W'_{\underline{i}} \subseteq W_{\underline{i}}$, so there is an extended lifting function over $W'_{\underline{i}}$.

Lemma 33.60. *The collection of sets $\{W'_{\underline{i}}\}$ is a numerable covering of* $\mathrm{Path}(B)$*.*

Proof. Consider a path $\omega \in \mathrm{Path}(B)$. Choose a set $W_{\underline{i}}$ such that $\omega \in W_{\underline{i}}$ with $r = \mathrm{len}(\underline{i})$ as small as possible. We then have $g_r(\omega) = 0$ and thus $f'_{\underline{i}}(\omega) = f_{\underline{i}}(\omega) > 0$, so $\omega \in W'_{\underline{i}}$. This means that the sets $W'_{\underline{i}}$ form an open cover of $\mathrm{Path}(B)$. We next need to show that our collection is locally finite. To do this, choose $N > r$ such that $1/N < f_{\underline{i}}(\omega)$, and note that $g_N(\omega) > 1$. Set $V = \{\alpha \in \mathrm{Path}(B) \;:\; g_N(\omega) > 1\}$, which is an open neighbourhood of ω. If $\alpha \in V$ then for $m \geq N$ we have $g_m(\alpha) > 1$ and thus $f'_{\underline{j}}(\alpha) = 0$ whenever $\mathrm{len}(\underline{j}) = m$. Thus, if V meets $W'_{\underline{j}}$ then we must have $\mathrm{len}(\underline{j}) < N$. We know that the collection $\{W_{\underline{j}} \;:\; \mathrm{len}(\underline{j}) < N\}$ is locally finite, so there is a neighbourhood V' of ω which meets only finitely many of the $W_{\underline{j}}$'s with $\mathrm{len}(\underline{j}) < N$. It follows that $V \cap V'$ meets only finitely many of the $W'_{\underline{j}}$'s. Thus $\{W'_{\underline{i}}\}$ is locally finite. We also have a map $f'_{\underline{i}}\colon \mathrm{Path}(B) \to I$ with $(f'_{\underline{i}})^{-1}(0,1] = W'_{\underline{i}}$. This shows that $\{W'_{\underline{i}}\}$ is a numerable covering. $\square$

Proof of Theorem 33.49. We will write the proof in terms of transfinite recursion; it can be rewritten to use Zorn's lemma instead if you prefer. We may assume that the collection of tuples $\underline{i}$ is well-ordered. (If the collection of indices i is already well-ordered, we can order the tuples lexicographically; if not, we can appeal to the axiom of choice to get a random well-ordering of the tuples.) Thus, after a slight change of notation we have a locally finite covering $\{W'_\alpha\}$ indexed by the ordinals $\alpha < \kappa$ for some fixed ordinal κ. We also have functions $f'_\alpha\colon \mathrm{Path}(B) \to I$ with $W'_\alpha = (f'_\alpha)^{-1}(0,1]$ and extended lifting functions m_α over W'_α. Because the family is locally finite, the function $g_\alpha = \sum_{\beta < \alpha} f'_\alpha$ is continuous. We write $V_\alpha = \bigcup_{\beta < \alpha} W'_\beta = (g_\alpha)^{-1}(0,1]$. We next define lifting functions l_α over V_α by transfinite recursion, such that when $\alpha < \beta$ and $\omega \in V_\alpha$ we have $l_\alpha(\omega,e) = l_\beta(\omega,e)$ unless $\omega \in W'_\gamma$ for some $\gamma \in [\alpha,\beta)$. (Note that there are only finitely many ordinals γ such that $\omega \in W'_\gamma$, so the lift $l_\alpha(\omega,e)$ will only change a finite number of times as α varies.)

As $V_0 = \emptyset$, the recursion starts. Given a successor ordinal $\alpha+$, we feed $l_\alpha, m_\alpha, f_\alpha$ and g_α into Lemma 33.56 to define a lifting function $l_{\alpha+}$ on $V_\alpha \cup W'_\alpha = V_{\alpha+}$ which agrees with l_α on $V_\alpha \setminus W'_\alpha$. Now consider a limit ordinal λ, and a point $(\omega,e) \in F(V_\lambda)$. Because $\{W'_\alpha\}$ is locally finite, we can choose a neighbourhood U of ω such that $S = \{\alpha \;:\; U \cap W'_\alpha \neq \emptyset\}$ is finite. Choose any ordinal α with $\max(S) < \alpha < \lambda$ and define $l_\lambda(\omega,e) = l_\alpha(\omega,e)$. This is independent of the choice of α because $l_\alpha(\omega,e) = l_\beta(\omega,e)$ unless $\omega \in W'_\gamma$ for some $\gamma \in [\alpha,\beta)$. Note that l_λ actually agrees with l_α on the neighbourhood $V_\alpha \cap U$ of ω and l_α is continuous so l_λ is continuous at (ω,e). As (ω,e) was arbitrary, we see that l_λ is continuous.

At the end of the recursion we have a lifting function defined over $V_\kappa = \mathrm{Path}(B)$, as required. □

34. Real and complex numbers

[`apx-real`]

34.1. The construction of the reals. Many results and examples in topology are based on real numbers, so for a rigorous treatment we need some background about the structure of $\mathbb{R}$. Most obviously, we need a definition of $\mathbb{R}$. Perhaps the most obvious approach would be to define a nonnegative real number to be a doubly infinite string of decimal digits, with all digits sufficiently far to the left being zero. One would need to modify this slightly to take account of the fact that $0.\dot{9} = 1$. Next, one would have to define addition and multiplication, by a careful specification of the usual algorithms for manipulating decimals. The full rules for carrying and borrowing are rather intricate, and it is very awkward to complete this approach. It is better to define $\mathbb{R}$ in terms of $\mathbb{Q}$ and $\mathbb{N}$ in a more abstract way, and derive facts about decimal representation at the end if necessary. In this appendix we will sketch some details.

As well as a construction of $\mathbb{R}$, it is useful to have an axiomatic characterisation of what we have constructed.

Definition 34.1. [`defn-ring`]

A *commutative ring* is a set R equipped with elements $0, 1 \in R$ and binary operations $+, \cdot \colon R \times R \to R$ such that:

- R0: The operation $+$ is commutative and associative, with 0 as a neutral element. Moreover, every element $a \in R$ has an inverse with respect to $+$, written as $-a$, so $a + (-a) = 0$.
- R1: The operation $\cdot$ is commutative and associative, with 1 as a neutral element.
- R2: For all $a, b, c \in R$ we have $a(b + c) = ab + ac$.

A *field* is a commutative ring with the following additional properties:

- F0: For and $a \in R \setminus \{0\}$ there exists $b \in R$ with $ab = 1$.
- F1: $1 \neq 0$.

Definition 34.2. [`defn-order`]

A *total order* on a set X is a relation on X (written $x \leq y$) such that

- TO0: For all $x \in X$ we have $x \leq x$
- TO1: For all $x, y, z \in X$, if $x \leq y$ and $y \leq z$ then $x \leq z$.
- TO2: For all $x, y \in X$, either $x \leq y$ or $y \leq x$. Moreover, if both of these hold then $x = y$.

Definition 34.3. [`defn-bound`]

Let X be a set equipped with a total order. Consider a subset $A \subseteq X$.

(a) An *upper bound* for A is an element $u \in X$ such that $a \leq u$ for all $a \in A$.
(b) A *lower bound* for A is an element $v \in X$ such that $v \leq a$ for all $a \in A$.
(c) A *least upper bound* (or *supremum*) for A is an upper bound u with the property that $u \leq u'$ for any other upper bound u'. It is clear that A has at most one supremum, which we denote by $\sup(A)$ if it exists.
(d) A *greatest lower bound* (or *infimum*) for A is a lower bound u with the property that $u \geq u'$ for any other lower bound u'. It is clear that A has at most one infimum, which we denote by $\inf(A)$ if it exists.

(e) We say that X is *order-complete* if every nonempty subset that has an upper bound, also has a supremum.

DEFINITION 34.4. [defn-ordered-field]
An *ordered field* is a field K equipped with a total order with the following compatibility properties:

OF0: If $a \leq b$ and $c \leq d$ then $a + c \leq b + d$.
OF1: If $a \leq b$ and $0 \leq c$ then $ac \leq bc$.

For $n \in \mathbb{N}$, we let $n.1_K$ denote the sum in K of n copies of the identity element 1_K. We say that K is *Archimedean* if it has the following additional property:

OF2: If $a \in K$ then there exists $n \in \mathbb{N}$ with $a < n.1_K$.

A *complete ordered field* means an ordered field that is order-complete.

The main result is as follows:

THEOREM 34.5. [thm-ordered-field]
There exists an Archimedean complete ordered field. Moreover, if K and L are two Archimedean complete ordered fields, then there is a unique field isomorphism $\phi\colon K \to L$, and this also satisfies $\phi(a) \leq \phi(b)$ if and only if $a \leq b$, so it is an isomorphism of ordered fields.

The rest of this appendix will constitute the proof.
We will take all facts about $\mathbb{N}$ and $\mathbb{Z}$ as given. We write $\mathbb{N}^+ = \mathbb{N} \setminus \{0\} = \{1, 2, 3, \dots\}$. We then introduce an equivalence relation on $\mathbb{Z} \times \mathbb{N}^+$ by $(a, b) \sim (c, d)$ iff $ad = bc$. We write a/b for the equivalence class of (a, b), and $\mathbb{Q}$ for the set of all equivalence classes. This can be made into a field by the usual rules $a/b + c/d = (ad + bc)/(bd)$ and $(a/b).(c/d) = (ac)/(bd)$. There is also an ordering given by $a/b \leq c/d$ iff $ad \leq bc$, which makes $\mathbb{Q}$ into an ordered field. However, it is not order-complete. We put

$$\mathbb{Q}^+ = \{q \in \mathbb{Q} \ : \ q > 0\} = \{a/b \ : \ a, b \in \mathbb{N}^+\}.$$

DEFINITION 34.6. [defn-dedekind]
A *Dedekind set* is a subset $a \subseteq \mathbb{Q}^+$ such that:

DS0: $a \neq \emptyset$
DS1: a has an upper bound
DS2: If $0 < p < q \in a$ with $p, q \in \mathbb{Q}$ then $p \in a$.
DS3: If $q \in a$ then there exists r with $q < r \in a$.

We write $\mathbb{R}^+$ for the collection of all Dedekind sets. For any $q \in \mathbb{Q}^+$ we put $\phi(q) = \{t \in \mathbb{Q} \ : \ 0 < t < q\}$, which is easily seen to be a Dedekind set. We write $1_{\mathbb{R}}$ for $\phi(1)$. We also define $a + b, ab \subseteq \mathbb{Q}^+$ by

$$a + b = \{r \in \mathbb{Q}^+ \ : \ r = p + q \text{ for some } p \in a \text{ and } q \in b\}$$
$$ab = \{r \in \mathbb{Q}^+ \ : \ r = pq \text{ for some } p \in a \text{ and } q \in b\}.$$

PROPOSITION 34.7. [prop-dedekind-product]
$\mathbb{R}^+$ is a group under multiplication, with $1_{\mathbb{R}}$ as the identity element.

PROOF. We first show that if $a, b \in \mathbb{R}^+$ then $ab \in \mathbb{R}^+$. Indeed, as a and b are nonempty, it is clear that $ab \neq \emptyset$. If x is an upper bound for a and y is an upper bound for b then xy is an upper bound for ab. Now suppose that $0 < p < q \in ab$. We must then have $q = uv$ for some $u \in a$ and $v \in b$. As $p < q$ we have $p = uw$ for some w with $0 < w < v$. As $0 < w < v \in b$ we must have $w \in b$, so $p = uw \in ab$. Moreover, as $v \in b$ we can choose x with $v < x \in b$, and put $r = ux$; then $q < r \in ab$. This shows that ab is a Dedekind set, so $ab \in \mathbb{R}^+$. It is clear that $ab = ba$. If we have a third element $c \in \mathbb{R}^+$, it is also clear that

$$(ab)c = \{pqr \ : \ p \in a, \ q \in b, \ r \in c\} = a(bc),$$

so multiplication is associative as well as commutative. Next, observe that

$$a.1_{\mathbb{R}} = \{pq \ : \ p \in a, \ 0 < q < 1\}.$$

For p and q as above, we note that $0 < pq < p \in a$, so $pq \in a$. This means that $a.1_{\mathbb{R}} \subseteq a$. On the other hand, if $p \in a$ then we can choose r such that $p < r \in a$. We then put $q = p/r$, so $0 < q < 1$, so $q \in 1_{\mathbb{R}}$.

As $r \in a$ and $q \in 1_{\mathbb{R}}$ we see that $p = rq \in a.1_{\mathbb{R}}$. This shows that $a \subseteq a.1_{\mathbb{R}}$, and thus $a = a.1_{\mathbb{R}}$, so $1_{\mathbb{R}}$ is an identity element for multiplication.

Next, for any $a \in \mathbb{R}^+$ we define

$$a^* = \{p \in \mathbb{Q}^+ \ : \ \text{there exists } x < 1 \text{ with } pq < x \text{ for all } q \in a\}.$$

We claim that this is a Dedekind set. Indeed, if u is an upper bound for a, then $(2u)^{-1} \in a^*$, so $a^* \neq \emptyset$. As a is nonempty we can choose $q \in a$, and then q^{-1} is an upper bound for a^*. It is clear from the definition that if $0 < m < p \in a^*$ then $m \in a^*$. Now suppose that $p \in a^*$, and choose $x < 1$ as in the definition, so $pq < x$ for all $q \in a$. Choose y with $1 < y < x^{-1}$; then $xy < 1$ and $(py)q < xy$ for all $q \in a$, so $p < py \in a^*$. Thus $a^* \in \mathbb{R}^+$ as claimed. We will now show that $a^*a = 1_{\mathbb{R}}$. It is clear from the definitions that when $p \in a^*$ and $q \in a$ we have $pq \in 1_{\mathbb{R}}$, so $a^*a \subseteq 1_{\mathbb{R}}$. For the reverse inclusion, consider a positive integer n, large enough that $1/n \in a$. As a has an upper bound, there will be a smallest value of k such that $k/n^2 \notin a$. As $1/n \in a$ we must have $k > n$. Put $p = n^2/(k+1)$ and $x = k/(k+1) = 1 - 1/(k+1)$. For $q \in a$ we must have $q < k/n^2$, so $pq < k/(k+1) = x < 1$, which shows that $p \in a^*$. By the definition of k we also have $(k-1)/n^2 \in a$ and so $p(k-1)/n^2 \in aa^*$, or in other words $1 - \frac{2}{k+1} \in a^*a$. As remarked above, we have $k > n$, so $1 - \frac{2}{n} < 1 - \frac{2}{k+1}$, so $1 - \frac{2}{n} \in a^*a$, so every rational r with $0 < r < 1 - 2/n$ lies in a^*a. This holds for all sufficiently large n, and it follows that every r with $0 < r < 1$ lies in a^*a, so $a^*a = 1_{\mathbb{R}}$ as claimed. □

Proposition 34.8. [prop-dedekind-sum]
Addition defines a commutative and associative operation on $\mathbb{R}^+$. Moreover, if $a, b, c \in \mathbb{R}^+$ and $a + b = a + c$ then $b = c$.

Proof. Suppose that $a, b \in \mathbb{R}^+$. As a and b are nonempty, it is clear that $a + b \neq \emptyset$. If x is an upper bound for a and y is an upper bound for b then $x + y$ is an upper bound for $a + b$. Now suppose that $0 < p < q \in a + b$. We can then write $q = u + v$ for some $u \in a$ and $v \in b$. Put $t = p/q < 1$. Then $0 < tu < u \in a$, so $tu \in a$. Similarly $0 < tv < v \in b$, so $tv \in b$. It follows that $tu + tv \in a + b$, but $tu + tv = t(u + v) = (p/q)q = p$, so $p \in a + b$. Finally, as $v \in b$ we can choose w with $v < w \in b$ and then the number $r = u + w$ has $q < r \in a + b$. This proves that $a + b$ is a Dedekind set, so $a + b \in \mathbb{R}^+$. It is clear that $a + b = b + a$ and that

$$a + (b + c) = \{p + q + r \ : \ p \in a, \ q \in b, \ r \in c\} = (a + b) + c,$$

so we have a commutative and associative operation.

Now suppose that $a + b = a + c$. Given $v \in b$, we can choose x with $v < x \in b$. We can then choose $n \in \mathbb{N}$ large enough that $1/n < x - v$ and also $1/n \in a$. Let k be the smallest integer such that $k/n \notin a$. This means that $(k-1)/n \in a$, so $(k-1)/n + x \in a + b = a + c$, so $(k-1)/n + x = u + w$ for some $u \in a$ and $w \in c$. By the definition of k we must have $u \leq k/n$, so

$$w = x + \frac{k-1}{n} - u \geq x + \frac{k-1}{n} - \frac{k}{n} = x - \frac{1}{n} > v.$$

Thus $v < w \in c$, so $v \in c$. This proves that $b \subseteq c$, and by symmetry we also have $c \subseteq b$, so $b = c$ as claimed. □

Proposition 34.9. [prop-dedekind-distrib]
If $a, b, c \in \mathbb{R}^+$ then $a(b + c) = ab + ac$.

Proof. By definition we have

$$\begin{aligned} a(b + c) &= \{pq + pr \ : \ p \in a, \ q \in b, \ r \in c\} \\ ab + ac &= \{p_0 q + p_1 r \ : \ p_0, p_1 \in a, \ q \in b, \ r \in c\}, \end{aligned}$$

and we have seen already that both of these are Dedekind sets. By taking $p_0 = p_1 = p$ we see that $a(b + c) \subseteq ab + ac$. By taking $p = \max(p_0, p_1)$ we see that every element of $ab + ac$ is less than or equal to some element of $a(b + c)$, and so lies in $a(b + c)$ because $a(b + c)$ is a Dedekind set. The claim follows. □

We next consider the order properties of $\mathbb{R}^+$.

Definition 34.10. [defn-R-plus-order]
For $a, b \in \mathbb{R}^+$ we write $a \leq b$ if and only if $a \subseteq b$. We also write $a < b$ if and only if $a \subseteq b$ and $a \neq b$.

Proposition 34.11. [prop-R-plus-complete]
This defines a total order on $\mathbb{R}^+$, with respect to which it is order-complete.

Proof. Suppose that $a, b \in \mathbb{R}^+$; we claim that either $a \leq b$ or $b \leq a$. If not, then we can choose $p \in a \setminus b$ and $q \in b \setminus a$. Clearly we cannot have $p = q$, so either $p < q$ or $q < p$. In the first case we have $0 < p < q \in b$ and b is a Dedekind set so $p \in b$, contrary to our assumption about p. In the second case we have $0 < q < p \in a$ and a is a Dedekind set, so $q \in a$, contrary to our assumption about a. Either way we have a contradiction, so we must have $a \leq b$ or $b \leq a$ after all. All other axioms in Definition 34.2 are clear, so we have a total order on $\mathbb{R}^+$.

Suppose that A is a nonempty subset of $\mathbb{R}^+$ that has an upper bound, say b. Each element $x \in A$ is a subset of $b \subseteq \mathbb{Q}^+$, so we can take the union of all these sets to get a set $a \subseteq b$. We claim that this is a Dedekind set. Indeed, as A is nonempty and each element of A is nonempty we see that $a \neq \emptyset$. As b is a Dedekind set there exists u such that $p < u$ for all $p \in b$, so clearly $p < u$ for all $p \in a$. Now suppose we have $0 < p < q \in a$. As $q \in a$ we must have $q \in x$ for some $x \in A$. As x is a Dedekind set and $0 < p < q \in x$ we have $p \in x \subseteq a$ and so $p \in A$. Moreover, as x is a Dedekind set and $q \in x$ there must also exist $r > q$ with $r \in x$ and therefore also $r \in a$. This proves that $a \in \mathbb{R}^+$, and a is clearly a supremum for A. □

Proposition 34.12. *Suppose that $a, b, c, d \in \mathbb{R}^+$ with $a \leq b$ and $c \leq d$. Then $a + c \leq b + d$ and $ac \leq bd$ and $a^{-1} \geq b^{-1}$.*

Proof. As the order is just given by subset inclusion, the first two claims are immediate from the definitions. We can take $c = d = (ab)^{-1}$ in the second claim to see that $b^{-1} \leq a^{-1}$. □

Corollary 34.13. [cor-inf]
If A is a nonempty subset of $\mathbb{R}^+$ that has a lower bound, then it has an infimum.

Proof. Put $B = \{a^{-1} \;:\; a \in A\}$. This is nonempty and bounded above, so it has a supremum by Proposition 34.11. One can then check that $\sup(B)^{-1}$ is an infimum for A. □

Proposition 34.14. [prop-R-plus-subtract]
For $a, b \in \mathbb{R}^+$ we have $a < b$ if and only if there exists $c \in \mathbb{R}^+$ with $b = a + c$.

Proof. First suppose that $b = a + c$. Choose any $t \in c$, and note that for $p \in a$ we have $p + t \in a + c = b$ so $0 < p < p + t \in b$, so $p \in b$. This proves that $a \subseteq b$. Next, choose n large enough that $1/n < t$ and also $1/n \in a$. Let k be the smallest positive integer such that $k/n \notin a$. Then $(k-1)/n \in a$ and $1/n \in c$ so $k/n \in a + c = b$, but $k/n \notin a$, so $a \neq b$. Thus $a < b$ as required.

Conversely, suppose that $a < b$. Put

$$c = \{q \;:\; \text{ there exists } x \in b \text{ with } p + q < x \text{ for all } p \in a\}.$$

We claim that c is a Dedekind set. Indeed, as $a < b$ we can choose $w \in b \setminus a$. As b is a Dedekind set we can then choose x with $w < x \in b$. We then put $q = (x - w)/2$. For $p \in a$ we must have $p \leq w$ and so $p + q < x$; it follows that $q \in c$, so $c \neq \emptyset$. Next, it is clear that $c \subseteq b$, and b has an upper bound, so c has an upper bound. Now suppose that $q \in c$, and let x be as in the definition. If $0 < q' < q$ then for all $p \in a$ we have $p + q' < p + q < x$, so again $q' \in c$. Finally, as b is a Dedekind set we can choose $y \in b$ with $x < y$, and it follows that $q < (y - x) + q \in c$. This shows that c is a Dedekind set as claimed, and it is clear that $a + c \subseteq b$. For the reverse inclusion, choose $n \in \mathbb{N}$ large enough that $2/n \in a \cap c \subset b$. Let j be the smallest integer such that $j/n \notin a$, and let k be the smallest integer such that $k/n \notin b$. As $2/n \in c$ we find that $k - j - 1 > 0$. We claim that the number $q = (k - j - 1)/n$ lies in c. Indeed, we certainly have $(k - 1)/n \in b$, and for $p \in a$ we have $p < j/n$ and so $p + q < j/n + (k - j - 1)/n = (k - 1)/n \in b$, so $q \in c$. We also have $(j - 1)/n \in a$ and so $(k - 2)/n = (j - 1)/n + q \in a + c$. Now consider an arbitrary element $r \in b$. By taking n large enough we can arrange that $r < (k - 2)/n$ and so $r \in a + c$. It follows that $a + c = b$ as required. □

Proposition 34.15. [prop-Q-R]
The map $\phi\colon \mathbb{Q}^+ \to \mathbb{R}^+$ satisfies $\phi(a + b) = \phi(a) + \phi(b)$ and $\phi(ab) = \phi(a)\phi(b)$. Moreover, we have $\phi(a) \leq \phi(b)$ if and only if $a \leq b$.

Proof. Left to the reader. □

We next want to define $\mathbb{R}$ in terms of $\mathbb{R}^+$. The idea is that any element $x \in \mathbb{R}$ can be written in the form $a - b$ for some $a, b \in \mathbb{R}^+$, with $a - b = c - d$ if and only if $a + d = b + c$. Formally, we will define $\mathbb{R}$ as the quotient of $\mathbb{R}^+ \times \mathbb{R}^+$ by an equivalence relation, and the equivalence class of (a, b) will correspond to $a - b$.

Definition 34.16. [`defn-R`]

We now introduce a relation on $\mathbb{R}^+ \times \mathbb{R}^+$ by $(a, b) \sim (c, d)$ if and only if $a + d = b + c$. This is clearly reflexive and symmetric. If $(a, b) \sim (c, d)$ and $(c, d) \sim (e, f)$ then $a + d = b + c$ and $c + f = d + e$. We can add these to get $(a + f) + (c + d) = (b + e) + (c + d)$, and then appeal to the last part of Proposition 34.8 to deduce that $a + f = b + e$, which means that $(a, b) \sim (e, f)$. Our relation is thus transitive, so it is an equivalence relation. We let $\mathbb{R}$ denote the set of equivalence classes, and write $[a, b]$ for the equivalence class of (a, b). We define a map $\lambda \colon \mathbb{R}^+ \to \mathbb{R}$ by $\lambda(a) = [a + 1_{\mathbb{R}^+}, 1_{\mathbb{R}^+}]$.

Proposition 34.17. [`prop-R-field`]

The set $\mathbb{R}$ has a unique ring structure for which the map λ preserves addition and multiplication. With this ring structure, $\mathbb{R}$ is a field. Moreover, $\mathbb{R}$ is the disjoint union of $\lambda(\mathbb{R}^+)$, $-\lambda(\mathbb{R}^+)$ and $\{0\}$.

Proof. We first define operations on $\mathbb{R}^+ \times \mathbb{R}^+$ as follows:

$$\begin{aligned}(a, b) + (c, d) &= (a + c, b + d)\\ (a, b)(c, d) &= (ac + bd, ad + bc).\end{aligned}$$

It is clear that addition is commutative and associative, and that multiplication is commutative. One can also check that

$$((a, b)(c, d))(e, f) = (ace + adf + bcf + bde, acf + ade + bce + bdf) = (a, b)((c, d)(e, f))$$

so multiplication is also associative. Another straightforward check gives the distributivity rule

$$(a, b)((c, d) + (e, f)) = (a, b)(c, d) + (a, b)(e, f).$$

Next, if $(a, b) \sim (a', b')$ and $(c, d) \sim (c', d')$ then $a + b' = a' + b$ and $c + d' = c' + d$, which gives

$$(a + c) + (b' + d') = (a + b') + (c + d') = (a' + b) + (c' + d) = (a' + c') + (b + d),$$

so $(a + c, b + d) \sim (a' + c', b' + d')$. It follows that there is an induced operation on $\mathbb{R}$ given by $[a, b] + [c, d] = [a + c, b + d]$, and this is again commutative and associative. Next, note that $(a + 1, b + 1) \sim (a, b)$, so the element $0_{\mathbb{R}} = [1, 1]$ is a neutral element for addition. Moreover, we find that $(1, 1) \sim (c, c)$ for all c, so $[a, b] + [b, a] = [a + b, a + b] = [1, 1] = 0_{\mathbb{R}}$, so $[b, a]$ is an additive inverse for $[a, b]$.

We now consider multiplication. Suppose that $(c, d) \sim (c', d')$, so $c + d' = c' + d$. It follows that

$$\begin{aligned}(ac' + bd') + (ad + bc) &= a(c' + d) + b(c + d')\\ &= a(c + d') + b(c' + d) = (ac + bd) + (ad' + bc'),\end{aligned}$$

so

$$(a, b)(c', d') = (ac' + bd', ad' + bc') \sim (ac + bd, ad + bc) = (a, b)(c, d).$$

By symmetry, if $(a, b) \sim (a', b')$ we also have $(a', b')(c', d') = (a, b)(c', d')$. It follows that there is a well-defined product on $\mathbb{R}$ given by

$$[a, b][c, d] = [ac + bd, ad + bc].$$

It follows from the corresponding properties of $\mathbb{R}^+ \times \mathbb{R}^+$ that this is commutative and associative and distributes over addition. We also note that $(2, 1)(a, b) = (2a + b, 2b + a) \sim (a, b)$, so the element $1_{\mathbb{R}} = [2, 1] = \lambda(1_{\mathbb{R}^+})$ is a neutral element for multiplication. We have thus made $\mathbb{R}$ into a commutative ring.

Next, note that

$$\begin{aligned}\lambda(a) + \lambda(b) &= [a + 1, 1] + [b + 1, 1] = [a + b + 2, 2] = [a + b + 1, 1] = \lambda(a + b)\\ \lambda(a)\lambda(b) &= [a + 1, 1][b + 1, 1] = [ab + a + b + 2, a + b + 2] = [ab + 1, 1] = \lambda(ab).\end{aligned}$$

Consider an arbitrary element $x = [a, b] \in \mathbb{R}$. If $a = b$ then $x = [1, 1] = 0_{\mathbb{R}}$. If $a < b$ then Proposition 34.14 tells us that $b = a + c$ for some c, so $x = [a, b] = [a, a + c] = [1, 1 + c] = -\lambda(c)$. Similarly, if $a > b$ then $a = b + d$ for some d, and so $x = \lambda(d)$. It follows that $\mathbb{R} = \{0\} \amalg \lambda(\mathbb{R}^+) \amalg -\lambda(\mathbb{R}^+)$ as claimed. We also know that every element $a \in \mathbb{R}^+$ has a multiplicative inverse $a^* \in \mathbb{R}^+$, so $\lambda(a)$ has inverse $\lambda(a^*)$, and $-\lambda(a)$ has inverse $-\lambda(a^*)$. It follows that $\mathbb{R}$ is a field. $\square$

Definition 34.18. For $x, y \in \mathbb{R}$ we declare that $x \leq y$ if $x = y$ or $y - x \in \lambda(\mathbb{R}^+)$.

Proposition 34.19. *This makes $\mathbb{R}$ into an Archimedean complete ordered field.*

Proof. First, it is clear from the definition that $x \leq x$. If $x < y$ and $y < z$ then $y = x + \lambda(a)$ and $z = y + \lambda(b)$ say, so $z = x + \lambda(a + b)$, so $x < z$. It follows easily that if $x \leq y \leq z$ then $x \leq z$. For an arbitrary pair $x, y \in \mathbb{R}$ we have $y - x \in \mathbb{R} = \{0\} \amalg \lambda(\mathbb{R}^+) \amalg -\lambda(\mathbb{R}^+)$, which implies that $x = y$ or $x < y$ or $y < x$. It follows that we have defined a total order on $\mathbb{R}$.

If $w < x$ and $y < z$ then $x = w + \lambda(a)$ and $z = y + \lambda(b)$ say, so $x + z = w + y + \lambda(a + b)$, so $w + y < x + z$. It follows that the order is compatible with addition. If we also have $y > 0$ then we can write $y = \lambda(c)$ say, so $xy - wy = \lambda(ac)$, so $wy < xy$. This means that the order is compatible with multiplication, so we have an ordered field.

Now suppose we have a nonempty subset $B \subset \mathbb{R}$, and an upper bound u for B. Choose $x \in B$, and put $A = \{a \in \mathbb{R}^+ \ : \ x - 1 + \lambda(a) \in B\}$. We have $1 \in A$ so $A \neq \emptyset$. We also have $u > x - 1$ so $u = x - 1 + \phi(t)$ for some t, and we find that t is an upper bound for A. It follows that A has a supremum in $\mathbb{R}^+$, and we find that $x - 1 + \lambda(\sup(A))$ is a supremum for B in $\mathbb{R}$.

All that is left is to check the Archimedean property. If $x \in -\lambda(\mathbb{R}^+)$ or $x = 0$ then $x < 1_{\mathbb{R}}$. If $x = \lambda(a)$ for some $a \in \mathbb{R}^+$ then we can choose a rational number q that is an upper bound for a, and then an integer n such that $q < n$, and we find that $x < n.1_{\mathbb{R}}$. □

Remark 34.20. [`rem-archimedean`]

It is useful to record some slight variants of the Archimedean property. First, for any $\epsilon > 0$ we claim that there is an integer $n > 0$ such that $1/n < \epsilon$. This is just because the Archimedean property gives us an integer $n > 0$ with $n > 1/\epsilon$. Next, an easy induction gives $2^n \geq n + 1$ for all $n \in \mathbb{N}$, so for all $x \in \mathbb{R}$ there exists $n \in \mathbb{N}$ with $2^n \geq x$.

Lemma 34.21. [`lem-ordered-field`]
Let K be an ordered field.

(a) *For $a \in K$ we have $a \geq 0$ if and only if $-a \leq 0$.*
(b) *For any $a \in K$ we have $a^2 \geq 0$.*

Proof. (a) If $a \geq 0$ we can add $-a \leq -a$ to that inequality (using axiom OF0) to see that $0 \geq -a$.

(b) We must either have $a = 0$ or $a > 0$ or $a < 0$. If $a = 0$ then the claim is clear. If $a > 0$ then we can multiply that inequality by a (using axiom OF1) to see that $a^2 > 0$. If $a < 0$ then $-a > 0$ so $(-a)^2 > 0$ by the previous case, but $(-a)^2 = a^2$ so $a^2 > 0$.

□

Proposition 34.22. *Let K be a complete Archimedean ordered field. Then there is a unique field isomorphism $\phi\colon \mathbb{R} \to K$, and moreover ϕ also preserves the order.*

Proof. First, as $1_K = 1_K^2$ we see from Lemma 34.21(b) that $1_K > 0$. It follows by induction that $n.1_K > 0$ for all $n \in \mathbb{N}^+$. We can thus define $\phi_0\colon \mathbb{Q} \to K$ by $\phi_0(p/n) = (p.1_K)(n.1_K)^{-1}$ for all $p \in \mathbb{Z}$ and $n \in \mathbb{N}^+$; it is easy to check that this is well-defined and is a homomorphism of fields and that it preserves order. Now put $K^+ = \{x \in K \ : \ x > 0\}$. For $a \in \mathbb{R}^+$, consider the set

$$\phi_0(a) = \{\phi_0(p) \ : \ p \in a\} \subseteq K^+.$$

This is nonempty, because a is. If u is an upper bound for a, then $\phi_0(u)$ is an upper bound for $\phi_0(a)$. As K is complete, it follows that $\phi_0(a)$ has a supremum, and we define $\phi^+(a) = \sup(\phi_0(a))$. This gives a map $\phi^+\colon \mathbb{R}^+ \to K^+$, which again preserves the order.

In the opposite direction, suppose we have $x \in K^+$. We then define $\psi^+(x) = \{p \in \mathbb{Q}^+ \ : \ \phi_0(p) < x\}$. We claim that this is a Dedekind set. Indeed, as K is Archimedean we can find $n \in \mathbb{N}$ with $x^{-1} < n.1_K$. It then follows that $1/n \in \psi^+(x)$, so $\psi^+(x) \neq \emptyset$. Similarly, we can find $m \in \mathbb{N}$ such that $x < m.1_K$, so m is an upper bound for $\psi^+(x)$. If $0 < p < q \in \psi^+(x)$ it is clear that $p \in \psi^+(x)$. If $q \in \psi^+(x)$ then $(x - \phi_0(q))^{-1} \in K^+$ and we can use the Archimedean axiom again to find $k \in \mathbb{N}$ with $(x - \phi_0(q))^{-1} < k.1_K$. This rearranges to give $\phi_0(q + 1/k) < x$, so $q < q + 1/k \in \psi^+(x)$. This proves that $\psi^+(x)$ is a Dedekind set as claimed, so we have defined $\psi^+\colon K^+ \to \mathbb{R}^+$. As ϕ_0 preserves order we find that $\psi^+\phi^+ = 1\colon \mathbb{R}^+ \to \mathbb{R}^+$. In the other direction, we see from the definitions that x is an upper bound for $\phi_0(\psi^+(x))$. Suppose that $0 < y < x$; we claim that y is not an upper bound for $\phi_0(\psi^+(x))$. Indeed, we can use the Archimedean axiom to find n such that $\phi_0(2/n) < x - y$, and then once more to see that there is a smallest integer m with $\phi_0(m) > nx$. As $x > x - y > \phi_0(2/n)$ we see that $m > 2$. We find that $\phi_0(m-2) < nx$ and so $(m-2)/n \in \psi^+(x)$ and $\phi_0((m-2)/n) \in \phi_0(\psi^+(x))$. Now $\phi_0(m/n) > x$ and $\phi_0(2/n) < x - y$ so $\phi_0((m-2)/n) > y$. It follows that y is not an upper bound, as claimed. This means that x is the least upper bound of $\phi_0(\psi^+(x))$, so $\phi^+\psi^+ = 1\colon K^+ \to K^+$.

We still need to check that ϕ and ψ respect the algebraic structure. □

34.2. Complex numbers. Write this

34.3. The exponential map. For completeness, we will outline a rigorous treatment of the exponential map. To make everything as self-contained as possible, we will allow ourselves to use some theory of metric spaces, but no differentiation or integration.

Definition 34.23. [`defn-exp-trunc`]
For all $x \in \mathbb{C}$ and $n \in \mathbb{N}$ we put $e_n(x) = \sum_{k=0}^{n} x^k/k!$. For $0 \le t < n+2$ we also put

$$r_n(t) = \frac{t^{n+1}}{(n+1)!}\frac{1}{1 - t/(n+2)} > 0.$$

Remark 34.24. [`rem-rn-inc`]
For $0 \le t < n+2$ we note that $1 - t/(n+2)$ is a positive and decreasing function of t, so $1/(1-t/(n+2))$ is positive and increasing, so $r_n(t)$ is also positive and increasing.

Lemma 34.25. [`lem-exp-gap`]
Suppose that $|x| \le R$. Then the terms $a_k = x^k/k!$ satisfy

$$|a_{k+j}| \le (|x|/(k+1))^j |a_k| \le (R/(k+1))^j |a_k|$$

for all $j \ge 0$. Moreover, if $R < n+2 \le m+2$ then

$$|e_m(x) - e_n(x)| \le r_n(|x|) \le r_n(R).$$

Proof. First, we have

$$|a_{k+j+1}| = |a_{k+j}|.\frac{|x|}{k+j+1} \le |a_{k+j}|.\frac{|x|}{k+1} \le |a_{k+j}|.\frac{R}{k+1}$$

and it follows by induction that

$$|a_{k+j}| \le (|x|/(k+1))^j |a_k| \le (R/(k+1))^j |a_k|$$

as claimed. Now suppose that $|x| \le R < n+2 \le m+2$ and put $\alpha = |x|/(n+2)$, so $0 \le \alpha < 1$. We then have $|a_{n+1+j}| \le |a_{n+1}|\alpha^j$, so

$$\begin{aligned}
|e_m(x) - e_n(x)| &= \left|\sum_{k=n+1}^{m} a_k\right| \le \sum_{k=n+1}^{m} |a_k| \le \sum_{j=0}^{m-n-1} |a_{n+1}|\alpha^j \\
&\le \frac{|x|^{n+1}}{(n+1)!}\frac{1-\alpha^{m-n}}{1-\alpha} \le \frac{|x|^{n+1}}{(n+1)!}\frac{1}{1-\alpha} \\
&= r_n(|x|)
\end{aligned}$$

as claimed. We also have $r_n(|x|) \le r_n(R)$ by Remark 34.24. □

Lemma 34.26. [`lem-rn-bound`]
For all $R \ge 0$ we have $r_n(R) \to 0$ as $n \to \infty$.

Proof. Suppose we are given $R \ge 0$ and $\epsilon > 0$. Put $a_n = R^n/n!$. Choose $N_0 \in \mathbb{N}$ such that $N_0 \ge 2R$, so $R/(N_0+1) < 1/2$. Lemma 34.25 then tells us that $a_{N_0+j} \le a_{N_0}/2^j$. Next, by Remark 34.20 we can choose $N_1 \in \mathbb{N}$ with $2^{N_1} > 2a_{N_0}/\epsilon$, and put $N = N_0 + N_1$. We find that $a_N < \epsilon/2$. We also have $R/(N+2) < 1/2$, so $1 - R/(N+2) > 1/2$, so $1/(1 - R/(N+2)) < 2$. It follows that $r_N(R) = a_N/(1 - R/(N+2)) < \epsilon$, and similarly $r_n(R) < \epsilon$ for all $n \ge N$, as required. □

Corollary 34.27. [cor-exp-cauchy]
For all $\epsilon > 0$ and $R > 0$ there exists $N \in \mathbb{N}$ such that $|e_n(x) - e_m(x)| < \epsilon$ whenever $n, m \geq N$ and $|x| \leq R$.

Proof. By Lemma 34.26 we can find N with $r_n(R) < \epsilon$ for all $n \geq N$, and the claim then follows from Lemma 34.25. □

Corollary 34.28. [cor-exp-exists]
There is a unique function $\exp\colon \mathbb{C} \to \mathbb{C}$ with the following property: whenever $|x| < n + 2$ we have $|\exp(x) - e_n(x)| \leq r_n(|x|)$.

Proof. Fix $x \in \mathbb{C}$. Corollary 34.27 shows that the sequence $(e_n(x))_{n \in \mathbb{N}}$ is a Cauchy sequence in the complete metric space $\mathbb{C}$, so it has a unique limit, which we take to be $\exp(x)$. The inequality $|\exp(x) - e_n(x)| \leq r_n(|x|)$ then follows from Lemmas 34.25 and 12.5. As $r_n(|x|) \to 0$ as $n \to \infty$, this inequality forces $\exp(x)$ to be the limit of $\{e_n(x)\}$, which gives uniqueness. □

Remark 34.29. [rem-exp-conjugate]
It is clear from the definition that $e_n(\overline{x}) = \overline{e_n(x)}$, and it follows by passing to the limit that $\exp(\overline{x}) = \overline{\exp(x)}$. In particular, if x is real then $\exp(x)$ is real.

Definition 34.30. [defn-trig]
For $z \in \mathbb{C}$ we put $\cos(z) = (\exp(iz) + \exp(-iz))/2$ and $\sin(z) = (\exp(iz) - \exp(-iz))/2$, so

$$\exp(iz) = \cos(z) + \sin(z) i.$$

Remark 34.31. [rem-trig]
For real x we note that $\exp(-ix) = \overline{\exp(ix)}$, so $\cos(x) = \mathrm{Re}(\exp(ix))$ and $\sin(x) = \mathrm{Im}(\exp(ix))$. In particular, $\cos(x)$ and $\sin(x)$ are real.

Lemma 34.32. *For all $x, y \in \mathbb{C}$ with $|x|, |y| \leq R$ we have*

$$|e_n(x)e_n(y) - e_n(x+y)| \leq e_{2n}(2R) - e_n(2R).$$

Proof. From the definitions we have

$$\begin{aligned}
e_n(x)e_n(y) &= \sum_{i,j=0}^{n} \frac{x^i y^j}{i!j!} \\
e_n(x+y) &= \sum_{k=0}^{n} \frac{(x+y)^k}{k!} = \sum_{k=0}^{n} \sum_{i=0}^{k} \left(\begin{array}{c} k \\ i \end{array} \right) \frac{x^i y^{k-i}}{k!} \\
&= \sum_{i+j \leq n} \frac{x^i y^j}{i!j!} \\
e_n(x)e_n(y) - e_n(x+y) &= \sum_{i,j \leq n,\ i+j > n} \frac{x^i y^j}{i!j!} \\
|e_n(x)e_n(y) - e_n(x+y)| &\leq \sum_{i,j \leq n,\ i+j > n} \frac{R^{i+j}}{i!j!} \leq \sum_{n < i+j \leq 2n} \frac{R^{i+j}}{i!j!} \\
&= \sum_{n < k \leq 2k} \frac{(2R)^k}{k!} = e_{2n}(2R) - e_n(2R).
\end{aligned}$$

□

Corollary 34.33. [cor-exp-hom]
For all $x, y \in \mathbb{C}$ we have $\exp(x)\exp(y) = \exp(x+y)$.

Proof. We know that $e_n(z) \to \exp(z)$ for all z. Using this and the continuity of multiplication and subtraction, we see that $e_(x)e_n(y) - e_n(x+y) \to \exp(x)\exp(y) - \exp(x+y)$. On the other hand, we see that $e_{2n}(2R) - e_n(2R)$ converges to $\exp(2R) - \exp(2R) = 0$. It therefore follows from the lemma that $e_n(x)e_n(y) - e_n(x+y) \to 0$, but limits are unique so $\exp(x)\exp(y) - \exp(x+y) = 0$. $\square$

Corollary 34.34. [cor-exp-nonzero]
For all $x \in \mathbb{C}$ we have $\exp(x) \neq 0$, and $\exp(-x) = 1/\exp(x)$.

Proof. It is clear that $e_n(0) = 1$ for all n, so $\exp(0) = 1$. It follows that $\exp(x)\exp(-x) = \exp(x + (-x)) = \exp(0) = 1$, so $\exp(x)$ is nonzero with inverse $\exp(-x)$. $\square$

Corollary 34.35. [cor-exp-cts]
The map $\exp\colon \mathbb{C} \to \mathbb{C} \setminus \{0\}$ is continuous.

Proof. First, Corollary 34.28 gives

$$|\exp(t) - 1| = |\exp(t) - e_0(t)| \leq r_0(|t|) = \frac{|t|}{1 - |t|/2}.$$

Now, if $|t| < 1$ then $1 - |t| > 1/2$ so $|\exp(t) - 1| \leq 2|t|$. Now consider an arbitrary point $x \in \mathbb{C}$ and a number $\epsilon > 0$. Put $\delta = \min(1, \epsilon/(2|\exp(x)|))$. If $|y - x| < \delta$ then $y = x + t$ for some t with $|t| < 1$ and $|\exp(t) - 1| \leq 2|t| < \epsilon/|\exp(x)|$ so $|\exp(y) - \exp(x)| = |\exp(t) - 1||\exp(x)| < \epsilon$, as required. $\square$

Lemma 34.36. [lem-exp-homeo-R]
The map $\exp$ restricts to give a strictly increasing homeomorphism $\mathbb{R} \to (0, \infty)$. (We will write $\log$ for the inverse map $(0, \infty) \to \mathbb{R}$.)

Proof. First, for $x \geq 0$ it is clear that $e_n(x) \geq e_1(x) = 1 + x$ for all n, and thus that $\exp(x) \geq 1 + x \geq 1$. We also have $\exp(x)\exp(-x) = \exp(0) = 1$, so $\exp(-x) = 1/\exp(x)$, so $1/(1+x) \leq \exp(-x) \leq 1$. It follows that for all $x \in \mathbb{R}$ we have $\exp(x) > 0$. Now suppose we have $x, y \in \mathbb{R}$ with $x < y$. The number $t = y - x$ is then strictly positive, so $\exp(t) \geq 1 + t > 1$, so $\exp(y) = \exp(t)\exp(x) > \exp(x)$. This proves that $\exp\colon \mathbb{R} \to (0, \infty)$ is strictly increasing (and therefore injective). It follows by Proposition 8.5 that $\exp(\mathbb{R})$ is a convex open subset of $(0, \infty)$, and that $\exp\colon \mathbb{R} \to \exp(\mathbb{R})$ is a homeomorphism. If $x \geq 0$ then $\exp(\mathbb{R})$ is convex and contains 1 and $\exp(x) \geq 1 + x$, so it must contain $1 + x$. It follows that $[1, \infty) \subseteq \exp(\mathbb{R})$, and using $\exp(-x) = 1/\exp(x)$ we deduce that $(0, 1] \subseteq \exp(\mathbb{R})$. It follows that $\exp$ gives a homeomorphism $\mathbb{R} \to (0, \infty)$ as claimed. $\square$

Lemma 34.37. [lem-abs-exp]
For all $z \in \mathbb{C}$ we have $|\exp(z)| = \exp(Re(z))$. In particular, for $y \in \mathbb{R}$ we have $|\exp(iy)| = 1$.

Proof. We have $\exp(z)\exp(-z) = \exp(0) = 1$, so $\exp(z) \neq 0$. For the more precise statement, put $x = \mathrm{Re}(z) = (z + \overline{z})/2$. We have

$$|\exp(z)|^2 = \exp(z)\overline{\exp(z)} = \exp(z)\epsilon(\overline{z}) = \exp(z + \overline{z}) = \exp(2x) = \exp(x)^2.$$

As $|\exp(z)|$ and $\exp(x)$ are both positive, we can deduce that $|\exp(z)| = \exp(x)$. In particular, for $z = iy$ we have $x = 0$ and so $|\exp(z)| = 1$. $\square$

Lemma 34.38. [lem-ker-discrete]
If $0 < |z| < 6/5$ then $\exp(z) \neq 1$.

Proof. Corollary 34.28 gives

$$|\exp(z) - 1 - z| = |\exp(z) - e_1(z)| \leq r_1(|z|),$$

so

$$|\exp(z) - 1| \geq |z| - r_1(|z|).$$

From the definitions we have

$$t - r_1(t) = t - \frac{t^2}{2 - 2t/3} = \frac{5t}{2}\,\frac{6/5 - t}{3 - t},$$

and this is strictly positive for $0 < t < 6/5$. $\square$

Lemma 34.39. [lem-exp-two-i]
The number $\exp(2i)$ has negative real part and positive imaginary part.

Proof. Corollary 34.28 gives $|\exp(2i)-e_5(2i)| < r_5(2)$, and direct calculation gives $e_5(2i) = 14i/15-1/3$ and $r_5(2) = 28/225 < 1/3$. The claim follows easily from this. $\square$

Corollary 34.40. [cor-pi-exists]
There is a unique number $\pi > 0$ such that $\exp(it) = 1$ if and only if $t \in 2\pi\mathbb{Z}$. Moreover, we have $3/5 \leq \pi < 4$.

Proof. First, as $|\exp(z)| = \exp(\mathrm{Re}(z))$ and $\exp\colon \mathbb{R} \to (0,\infty)$ is strictly increasing, we see that $\exp(it)$ can only be one if t is real. We will assume this from now on.

Put $K = \{t \in \mathbb{R} \,:\, \exp(it) = 1\}$ and $K^+ = K \cap (0,\infty)$. Lemma 34.38 tells us that $K \cap (-6/5, 6/5) = \{0\}$. More generally, if $s,t \in K$ then $s-t \in K$ so either $s = t$ or $|s-t| \geq 6/5$.

The function $\cos(t) = \mathrm{Re}(\exp(it))$ has $\cos(0) = 1$ and $\cos(2) < 0$, so there exists $\alpha \in (0,2)$ with $\cos(\alpha) = 0$. This means that $\exp(i\alpha) = iy$ for some y, but we also know that $|\exp(i\alpha)| = 1$, so $\exp(i\alpha) = \pm i$. It follows that $\exp(4\alpha i) = \exp(i\alpha)^4 = 1$, so $4\alpha \in K^+$. This means that K^+ is nonempty, and it is visibly bounded below by 0, so we can define $\pi = \inf(K^+)/2$. As $K \cap (-6/5, 6/5) = \{0\}$ we have $\pi \geq 3/5$. As $4\alpha \in K^+$ and $\alpha < 2$ we have $\pi < 4$. As distinct points in K are separated by at least $6/5$, we see that the interval $(2\pi - 3/5, 2\pi + 3/5)$ can contain at most one point in K. The only way this can be consistent with $2\pi = \inf(K^+)$ is if $K \cap (2\pi - 3/5, 2\pi + 3/5) = \{2\pi\}$. In particular, we must have $2\pi \in K$, so $\exp(2\pi i) = 1$. Now consider a general element $t \in K$. Let n be the largest integer with $n \leq t/2\pi$, so $2\pi n \leq t < 2\pi(n+1)$. Put $u = t - 2\pi n$. As t and 2π lie in K, we see that $u \in K$. As $0 \leq u < 2\pi$ and $2\pi = \inf(K^+)$ we must have $u = 0$, so $t = 2n\pi$, as claimed. $\square$

Lemma 34.41. [lem-tan]
Put $S^1 = \{z \in \mathbb{C} \;:\; |z| = 1\}$. There is a homeomorphism $f\colon S^1 \setminus \{-1\} \to \mathbb{R}$ given by $f(z) = (z-1)/(i(z+1))$, with inverse $f^{-1}(t) = (1+it)/(1-it)$. If $z = x+iy$ (with $x^2+y^2 = 1$) then $f(z) = y/(1+x)$, so $f(z)$ has the same sign as $\mathrm{Im}(z)$. Moreover, we have $f(\overline{z}) = f(1/z) = -f(z)$.

Proof. We can define $f\colon \mathbb{C} \setminus \{-1\} \to \mathbb{C}$ by the given formula, and then define $g\colon \mathbb{C} \setminus \{i\} \to \mathbb{C}$ by $g(t) = (1+it)/(1-it)$. It is straightforward algebra to check that $f(g(t)) = t$ and $(f(z)) = z$ in cases where everything is defined. If $z \in S^1 \setminus \{-1\}$ we have $\overline{z} = z^{-1}$ and using this we obtain $\overline{f(z)} = f(z)$, so $f(z)$ is real. If t is real we have $|g(t)| = |1+it|/|1-it| = 1$, so $g(t) \in S^1$, and it is clear that $1+it \neq -(1-it)$ so $g(t) \neq -1$. It follows that f and g give homeomorphisms as described. The identity $f(1/z) = -f(z)$ is straightforward algebra, as is the identity

$$f(x+iy) - \frac{y}{1+x} = \frac{i(1-x^2-y^2)}{(1+x)(1+x+iy)},$$

which reduces to $f(x+iy) = y/(1+x)$ for $x+iy \in S^1 \setminus \{-1\}$. Note that $1+x > 0$ here, so $f(x+iy)$ has the same sign as the imaginary part y, as claimed. $\square$

Proposition 34.42. [prop-tan]
Let f be as in Lemma 34.41, and put $g(t) = f(\exp(it))$. Then g gives a strictly increasing homeomorphism $(-\pi,\pi) \to \mathbb{R}$.

Proof. We have seen that $|\exp(it)| = 1$. To show that g is well-defined, we must check that $\exp(it) \neq -1$ for $t \in (-\pi,\pi)$. As $\exp(0) = 1$ and $\exp(-it) = 1/\exp(it)$, it will suffice to treat the case $0 < t < \pi$. Here we have $\exp(it)^2 = \exp(2it) \neq 1$ (by the definition of π) and so $\exp(it) \neq -1$ as required.

We next claim that g is injective. Suppose that $g(s) = g(t)$. As f is a homeomorphism, this means that $\exp(is) = \exp(it)$, so $t - s \in dir\pi\mathbb{Z}$. However, we have $|s|,|t| < \pi$ so $|t-s| < 2\pi$ so $s = t$ as required. It follows by Proposition 8.5 that g is either strictly increasing or strictly decreasing, and that it gives a homeomorphism $(-\pi,\pi) \to g(-\pi,\pi)$. We have seen that $\pi > 3/5$, so $1/2$ is in the domain of g. We have

$$|\exp(i/2) - (1+i/2)| = |\exp(i/2) - e_1(i/2)| \leq r_1(1/2) = 3/20,$$

and from this it follows that $\mathrm{Im}(\exp(i/2)) > 0$ and so $g(1/2) > 0$. We also have $g(0) = 0$, so it follows that g is strictly increasing. As t approaches $\pm\pi$ we see that the number $z = \exp(it)$ approaches $\exp(\pm i\pi) = -1$, so the absolute value of $g(t) = f(z) = (z-1)/(i(z+1))$ must become unboundedly large. As g is strictly increasing, we deduce that $g(t)$ tends to $+\infty$ as t approaches π from below, and that $g(t)$ approaches $-\infty$

as t approaches $-\pi$ from above. This means that $g((-\pi,\pi))$ is a convex open set that is unbounded in both directions, so it is all of $\mathbb{R}$. $\square$

Corollary 34.43. [cor-exp-homeo-circle]
The map $e(t) = \exp(it)$ gives a homeomorphism $e\colon (-\pi,\pi) \to S^1 \setminus \{-1\}$.

Proof. Combine Lemma 34.41 and Proposition 34.42. $\square$

Remark 34.44. [rem-pi-value]
We see from Proposition 34.42 that $\exp(i\pi/2)$ has positive imaginary part, and it satisfies $\exp(i\pi/2)^2 = \exp(i\pi) = -1$, so we must have $\exp(i\pi/2) = i$. We therefore have $g(\pi/2) = f(i) = 1$. We can use the explicit bound in Corollary 34.28 to prove that $g(3/2) < 1 < g(11/7)$ and deduce that $3 < \pi < 22/7$, for example. If we prefer to work with small angles so that convergence is more rapid, we can do the following. Consider the number

$$z = (\sqrt{6} + \sqrt{2} + \sqrt{6}i - \sqrt{2}i)/4.$$

One can check that $z^{12} = 1$ and that $\mathrm{Im}(z^k) > 0$ for $0 < k < 12$. It follows that $z = \exp(\pi i/12)$. One can also check that the number $u = f(z) = (z-1)/(i(z+1))$ is

$$u = \sqrt{6} + \sqrt{2} - \sqrt{3} - 2 \approx 0.13.$$

We have $\pi = 12g^{-1}(u)$, and we can locate this by computing $g(t)$ for numbers t close to 0.26. Of course, if we really want an efficient computation of an accurate value for π, it is better to develop more theory first.

Corollary 34.45. [cor-exp-homeo]
Put $U = \{x + iy \in \mathbb{C} \;:\; -\pi < y < \pi\}$. Then the exponential map restricts to give a homeomorphism $\exp\colon U \to \mathbb{C} \setminus (-\infty, 0]$.

Proof. Define $e\colon (-\pi,\pi) \to S^1 \setminus \{-1\}$ by $e(t) = \exp(it)$. Recall that we previously defined $\log\colon (0,\infty) \to \mathbb{R}$ to be the inverse of the homeomorphism $\exp\colon \mathbb{R} \to (0,\infty)$. We now extend this to define a map $\log\colon \mathbb{C} \setminus (-\infty, 0] \to U$ by

$$\log(z) = \log(|z|) + i\, e^{-1}(z/|z|).$$

It is now straightforward to check that this is continuous and is inverse to the exponential map. $\square$

Corollary 34.46. [cor-exp-open]
The map $\exp\colon \mathbb{C} \to \mathbb{C} \setminus \{0\}$ is open and surjective, as is the map $e\colon \mathbb{R} \to S^1$ given by $e(t) = \exp(it)$.

Proof. Consider a point $z \in \mathbb{C} \setminus \{0\}$. If z does not lie on the negative real axis, Corollary 34.45 immediately gives $w \in \mathbb{C}$ with $\exp(w) = z$. If z does lie on the negative real axis then $-z$ does not, so Corollary 34.45 gives $w \in \mathbb{C}$ with $\exp(w) = -z$ and so $\exp(w + i\pi) = z$. It follows that $\exp\colon \mathbb{C} \to \mathbb{C} \setminus \{0\}$ is surjective.

Now suppose we have an open set $V \subseteq \mathbb{C}$. The claim is that $\exp(V)$ is open in $\mathbb{C} \setminus \{0\}$ (or equivalently, in $\mathbb{C}$). Put

$$U_n = \{z \in \mathbb{C} \;:\; (n-1)\pi < \mathrm{Im}(z) < (n+1)\pi\},$$

and note that these sets give an open cover of $\mathbb{C}$. Corollary 34.45 tells us that $\exp\colon U_0 \to \mathbb{C}$ is open. For $z \in U_n$ we have $z - n\pi i \in U_0$ and $\exp(z) = (-1)^n \exp(z - n\pi i)$; it follows that $\exp\colon U_n \to \mathbb{C}$ is open. We now see that $\exp(V) = \bigcup_n \exp(V \cap U_n)$, which is an open set as required.

This proves that $\exp\colon \mathbb{C} \to \mathbb{C} \setminus \{0\}$ is open and surjective, and the proof for $e\colon \mathbb{R} \to S^1$ is essentially the same. $\square$

35. Set theory

[apx-sets]

35.1. Countability.

Definition 35.1. [defn-countable]
A set X is *countable* if the following equivalent conditions are satisfied:

(a) Either $X = \emptyset$, or there is a surjective map $f\colon \mathbb{N} \to X$.
(b) There is an injective map $g\colon X \to \mathbb{N}$.
(c) Either X is finite, or there is a bijection $h\colon X \to \mathbb{N}$.

Proof of equivalence. First suppose that (a) holds. It is also tautological that there is a unique map $\emptyset \to \mathbb{N}$ and that this is injective. We can thus restrict attention to the case where there is a surjective map $f\colon \mathbb{N} \to X$. This means that for each $x \in X$ we have a nonempty set $f^{-1}\{x\}$ of natural numbers, so this set must have a smallest element. We can therefore define $g\colon X \to \mathbb{N}$ by

$$g(x) = \min\{n \;:\; f(n) = x\}.$$

As $f(g(x)) = x$ we see that g is injective, so (b) holds.

Suppose instead that we start with (b). If X is finite then (c) holds trivially, so we can restrict attention to the case where X is infinite. As $g\colon X \to \mathbb{N}$ is injective, it follows that the set $g(X) \subseteq \mathbb{N}$ is also infinite. We can therefore list the elements in order as $g(X) = \{m_0, m_1, m_2, \dots\}$ say, with $m_0 < m_1 < m_2 < \cdots$. As g is injective, there is a unique element $x_k \in X$ with $g(x_k) = m_k$. For each $x \in X$ we have $g(x) \in g(X)$, so $g(x)$ must appear as m_k for some k, so $x = x_k$ for some k. If $x_j = x_k$ then $g(x_j) = g(x_k)$ or in other words $m_j = m_k$, but the elements m_i form a strictly increasing sequence, so we must have $j = k$. It follows that we have a bijection $h\colon X \to \mathbb{N}$ given by $h(x_k) = k$ (or equivalently, by $h(x) = |\{y \;:\; g(y) < g(x)\}|$). This proves that (b) implies (c), and it is clear that (c) implies (a). □

Proposition 35.2. [prop-countable-examples]
The sets $\mathbb{N}$, $\mathbb{N}^2$, $\mathbb{Z}$ and $\mathbb{Q}$ are countable, as is the set

$$\mathcal{P}_f(\mathbb{N}) = \{A \subseteq \mathbb{N} \;:\; A \text{ is finite } \}.$$

However, the set $\mathcal{P}(\mathbb{N})$ of all *subsets of $\mathbb{N}$ is uncountable, as is the set $\mathbb{R}$.*

Proof. First, $\mathbb{N}$ is trivially countable. Next, we can define a bijection $f\colon \mathbb{N} \to \mathbb{Z}$ by $f(2n) = n$ and $f(2n+1) = -n-1$; so $\mathbb{Z}$ is countable. Now define $g\colon \mathbb{N}^2 \to \mathbb{N}$ by

$$g(n,m) = \tfrac{1}{2}(n+m+1)(n+m) + n.$$

Note here that the first term $(n+m+1)(n+m)/2$ is just the number of pairs $(j,k) \in \mathbb{N}^2$ with $j+k < n+m$. Using this, we see that the pattern of values is as follows:

$$\begin{array}{c|ccccc}
 & \xrightarrow{\;n\;} & & & & \\
m\downarrow & 0 & 2 & 5 & 9 & 14 \\
 & 1 & 4 & 8 & 13 & \\
 & 3 & 7 & 12 & & \\
 & 6 & 11 & & & \\
 & 10 & & & &
\end{array}$$

It follows that g is a bijection, so $\mathbb{N}^2$ is countable. Now define $h\colon \mathbb{Z} \times \mathbb{N} \to \mathbb{Q}$ by $h(n,m) = n/(m+1)$; this is clearly surjective. The following composite therefore gives a surjection $\mathbb{N} \to \mathbb{Q}$:

$$\mathbb{N} \xrightarrow[\simeq]{g} \mathbb{N} \times \mathbb{N} \xrightarrow[\simeq]{f \times 1} \mathbb{Z} \times \mathbb{N} \xrightarrow{h} \mathbb{Q}.$$

We deduce that $\mathbb{Q}$ is also countable.

Next, define $b\colon \mathcal{P}_f(\mathbb{N}) \to \mathbb{N}$ by $b(A) = \sum_{a\in A} 2^a$. We will take it as given that every natural number has a unique binary representation, which means precisely that b is a bijection. (The conscientious reader may wish to give a more detailed proof by induction.) Thus, we see that $\mathcal{P}_f(\mathbb{N})$ is countable.

Now consider instead the larger set $\mathcal{P}(\mathbb{N})$ of all subsets of $\mathbb{N}$. Given any map $f\colon \mathbb{N} \to \mathcal{P}(\mathbb{N})$, we can consider the set $A = \{n \in \mathbb{N} \;:\; n \notin f(n)\}$. Fix a natural number m; we claim that $A \neq f(m)$. Indeed, we must either have $m \in f(m)$ or $m \notin f(m)$. By the definition of A, in the first case we have $m \notin A$, and in the second we have $m \in A$. Either way, we see that $f(m)$ is different from A, as claimed. As m was arbitrary we deduce that A is not in the image of f. Thus, f cannot be surjective, which proves that $\mathcal{P}(\mathbb{N})$ is uncountable. (This proof is called *Cantor's diagonal argument.*)

Finally, we can define $k\colon \mathcal{P}(\mathbb{N}) \to \mathbb{R}$ by $k(A) = \sum_{a\in A} 3^{-a}$. We leave it to the reader to check that this sum is convergent, so k is well-defined, and that k is injective. If $\mathbb{R}$ were countable we could choose an injective map $j\colon \mathbb{R} \to \mathbb{N}$ and then we could use jk to prove that $\mathcal{P}(\mathbb{N})$ is countable, which is a contradiction. Thus, $\mathbb{R}$ must actually be uncountable. □

Proposition 35.3.

(a) *If $f\colon X \to Y$ is injective and Y is countable then so is X.*
(b) *If $g\colon X \to Y$ is surjective and X is countable then so is Y.*
(c) *If X and Y are countable, then so is $X \times Y$. More generally, if we have a family $(X_i)_{i\in I}$ where the index set I is finite and each set X_i is countable, then the product $\prod_{i\in I} X_i$ is also countable.*
(d) *Suppose we have a set X and a family of subsets $(X_i)_{i\in I}$ where the index set I is countable, and each set X_i is countable. Then the union $\bigcup_{i\in I} X_i$ is also countable.*

Proof.

(a) As Y is countable there exists an injective map $j\colon Y \to \mathbb{N}$, so $jf\colon X \to \mathbb{N}$ is injective, so X is countable.
(b) If X is empty and there is a surjection $X \to Y$ then Y is also empty and therefore countable. Otherwise, as X is countable we can choose a surjective map $q\colon \mathbb{N} \to X$, and then the surjection $gq\colon \mathbb{N} \to Y$ shows that Y is countable.
(c) Suppose that X and Y are countable. To avoid trivialities, we assume that X and Y are also nonempty, so we can choose surjections $p\colon \mathbb{N} \to X$ and $q\colon \mathbb{N} \to Y$. We then have a surjection $p \times q\colon \mathbb{N}^2 \to X \times Y$ but $\mathbb{N}^2$ is countable so it follows from (b) that $X \times Y$ is countable.
(d) If any of the sets X_i are empty we can harmlessly remove them from the family. We therefore assume that each X_i is nonempty, so we can choose a surjection $f_i\colon \mathbb{N} \to X_i$. We then define $g\colon \mathbb{N} \times \mathbb{N} \to \bigcup_i X_i$ by $g(n,m) = f_n(m)$. This is surjective, and $\mathbb{N} \times \mathbb{N}$ is countable, so $\bigcup_i X_i$ is countable by (a).

□

Example 35.4. [`eg-algebraic-numbers`]

Recall that a number $z \in \mathbb{C}$ is said to be *algebraic* if there is a nonzero polynomial $f(t) \in \mathbb{Q}[t]$ with $f(z) = 0$; otherwise, we say that z is *transcendental.* Let $\overline{\mathbb{Q}}$ denote the set of algebraic numbers; we will show that this is countable. First, the set $\mathbb{Q}^n$ is countable by part (c) of the Proposition. Next, if we let $\mathbb{Q}[t]_{<n}$ denote the set of polynomials of degree less than n, we have a bijection $\mathbb{Q}^n \to \mathbb{Q}[t]_{<n}$ given by $(a_0, \ldots, a_{n-1}) \mapsto \sum_i a_i t^i$. This shows that $\mathbb{Q}[t]_{<n}$ is countable, so the set $\mathbb{Q}[t] = \bigcup_n \mathbb{Q}[t]_{<n}$ is countable by part (d). Finally, for each $f \in \mathbb{Q}[t] \setminus \{0\}$ the set $R_f = \{z \in \mathbb{C} \;:\; f(z) = 0\}$ is finite, so the set $\overline{Q} = \bigcup_f R_f$ is countable by another application of part (d).

As $\overline{\mathbb{Q}}$ is countable but $\mathbb{R}$ and $\mathbb{C}$ are uncountable, it is clear that most real or complex numbers are transcendental. For example, it is known that π and e are transcendental, but the proofs are quite hard.

Example 35.5. [`eg-definable-numbers`]

Now let D be the set of all complex numbers that are definable by some finite formula. To make this completely rigorous would require a large detour into mathematical logic, but the idea should be clear enough. A "finite formula" should mean a finite list taken from some agreed alphabet of mathematical symbols (including the quantifiers $\forall$ and $\exists$), subject to certain syntactic rules (so that brackets match properly, for example). We would then need to spell out further rules about how such formulae can be interpreted and manipulated. For example, the number e can be defined by the formula $e = \sum_{n=0}^{\infty} 1/n!$. We probably do not want to take infinite sums as a primitive part of the language but we can instead say that for all m we have $\sum_{n=0}^{m} 1/n! < e$, and that for all $x < e$ there exists m such that $\sum_{n=0}^{m} 1/n! > x$. Again, we probably do not want to take $n!$ as a primitive symbol, so we should ensure that our formal language is sufficiently expressive to encode the recursive definition of $n!$. Note also that any algebraic number can be defined by some formula such as "x is the unique real number with $x^5 - 3x + 1 = 0$ and $-3/2 < x < -1$".

The main point for our present purposes is this: provided that we have only a finite (or even countable) number of symbols in our alphabet, the set of finite formulae will be countable, so the set D of defineable numbers will be countable.

35.2. Ordinals. Here we give a brief outline of the theory of ordinals. More details can be found in books on axiomatic set theory.

(a) There is a class of objects called ordinals, of which the first few are

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

(b) There are too many ordinals for the class of ordinals to be a set; it is "roughly the same size" as the class of *all* sets.

(c) Algebraic operations with ordinals (e.g. ω^2) must be treated with caution. For example, $1 + \omega = \omega \neq \omega + 1$. The reason is essentially as follows: the ordinal $1 + \omega$ corresponds to what you get by adding an extra point at the beginning of the ordered set $\mathbb{N}$, which gives a new ordered set that is isomorphic to $\mathbb{N}$. On the other hand, $\omega + 1$ corresponds to what you get by adding an extra point at the end of $\mathbb{N}$, which gives a new ordered set that is not isomorphic to $\mathbb{N}$.

(d) There is a linear order relation on ordinals. In other words, for any pair of ordinals α and β precisely one of the alternatives $\alpha < \beta$, $\alpha = \beta$ and $\alpha > \beta$ is true.

(e) The ordinals are well-ordered by this relation — any nonempty collection S of ordinals has a least element α, so $\alpha \in S$ and $\alpha \leq \beta$ for any $\beta \in S$.

(f) For any ordinal κ, the collection $S(\kappa)$ of ordinals $\alpha < \kappa$ is a set.

(g) For any set X there is an ordinal κ and a bijection $S(\kappa) \to X$, so $X = \{x_\alpha \ : \ \alpha < \kappa\}$ say.

(h) For any set X there is an ordinal λ so large that there is no injective map $S(\lambda) \to X$.

(i) An ordinal α is a *successor ordinal* if and only if $\alpha = \beta + 1$ for some β if and only if there is no ordinal γ with $\beta < \gamma < \alpha$. A *limit ordinal* is an ordinal (such as ω) which is not a successor.

(j) Transfinite induction over ordinals is valid. Suppose we have a statement $P(\alpha)$ about ordinals α, and we can show that $P(\alpha)$ is true whenever $P(\beta)$ is true for all $\beta < \alpha$. Then $P(\alpha)$ is true for all α. Indeed, consider the collection S of ordinals for which P is false. If S were nonempty, it would have a least element α. This would mean that $P(\beta)$ holds for all $\beta < \alpha$, leading swiftly to a contradiction.

(k) Transfinite recursion is valid. We can define a function f of ordinals by specifying $f(\alpha)$ in terms of the values $f(\beta)$ for $\beta < \alpha$. (In particular, we must specify $f(0)$ in terms of no data at all, so $f(0)$ must be defined in a non-recursive manner.) This is really a special case of (j): we prove by induction on α that there is a unique function f_α defined on $S(\alpha)$ with the required proprties, and by uniqueness we have $f_\alpha|_{S(\beta)} = f_\beta$ for all $\beta < \alpha$ so the functions f_α fit together to give a single function f as required.

We now discuss the formal definitions used in axiomatic set theory, which are very compact and elegant but take some digestion. An ordinal is a particular kind of set. In particular, the numbers $0, 1, 2$ and 3 are identified with certain sets, as follows:

$$\begin{aligned} 0 &= \emptyset \\ 1 &= \{\emptyset\} = \{0\} \\ 2 &= \{\emptyset, \{\emptyset\}\} = \{0, 1\} \\ 3 &= \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\} = \{0, 1, 2\}. \end{aligned}$$

Note that 1 is both an element and a subset of 3, and similarly 2 is both an element and a subset of 3.

For any set α, an $\in$-*minimal element* in α is an element $s \in \alpha$ such that for all $t \in \alpha$ we have $s = t$ or $s \in t$. An ordinal is by definition a set α such that every nonempty subset $\alpha' \subseteq \alpha$ has an $\in$-minimal element. In particular, we can consider $\alpha' = \{x, y\}$ for some pair of distinct elements $x, y \in \alpha$, and we see that either $x \in y$ or $y \in x$. In conjunction with the standard axioms of set theory, this implies that the set α can be linearly ordered by the rule $x < y$ if and only if $x \in y$. One checks that every element of an ordinal is again an ordinal, and we can order the class of all ordinals by the rule $\alpha < \beta$ if and only if $\alpha \in \beta$. The set $S(\alpha)$ is thus actually equal to α. The successor ordinal $\alpha + 1$ is $\alpha \cup \{\alpha\}$

Discuss the von Neumann heierarchy

35.3. Zorn's lemma.

DEFINITION 35.6. [`defn-poset-zorn`]

A *partial order* on a set P is an relation on P (written $p \leq q$) such that

PO0: For all $p \in P$ we have $p \leq p$

PO1: For all $p, q, r \in P$, if $p \leq q$ and $q \leq r$ then $p \leq r$.

PO2: For all $p, q \in P$, if $p \leq q$ and $q \leq p$ then $p = q$.

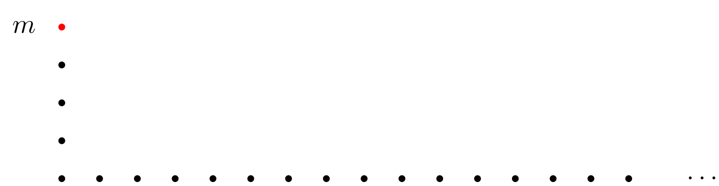
A *partially ordered set* or *poset* is a set with a specified partial order. We say that $m \in P$ is *largest* if for all $p \in P$ we have $p \leq m$. We say that $m \in P$ is *maximal* if whenever $p \in P$ and $m \leq p$ we have $p = m$.

REMARK 35.7. If m is largest in P then it is also the unique maximal element.

EXAMPLE 35.8. [`eg-maximal-elements`]

(a) Take $P = \mathbb{N}$ with its usual order. Then there is no maximal element and no largest element.
(b) Put $P = \{A \subseteq \mathbb{N} \,:\, |A| \leq 5\}$, ordered by inclusion (so we define $A \leq B$ to be true iff A is a subset of B). Then all sets of size five are maximal, but there is no largest element.
(c) Put $Q = \mathbb{N}^2$ with the order $(a, b) \leq (c, d)$ iff ($a \leq c$ and $b \leq d$). Then put

$$P = \{(0, b) \;:\; 0 \leq b \leq 4\} \cup \{(a, 0) \;:\; a \geq 0\}.$$



Then the point $m = (0, 4)$ is the unique maximal element, but it is not largest.
(d) Let X be a topological space and let Y be a subset of X. Let P be the set of all open subsets of Y, ordered by inclusion. Then P has a largest element, namely the interior of Y.

For many applications, it is useful to know that certain posets have maximal elements. We will prove a powerful result in this direction, known as Zorn's Lemma.

DEFINITION 35.9. [`defn-chain-bound`]

Let P be a partially ordered set. A subset $C \subseteq P$ is a *chain* if for all $p, q \in C$ we have either $p \leq q$ or $q \leq p$. An *upper bound* for C is an element $b \in P$ such that $p \leq b$ for all $p \in C$.

THEOREM 35.10 (Zorn's Lemma). [`thm-zorn`]

Let P be a partially ordered set in which every chain has an upper bound. Then P has a maximal element.

After some preliminaries, we will give two proofs of this. The first will be short and conceptual but it will rely on the theory of ordinals which we have not fully explained. The second will be self-contained but less easy to follow.

For historical reasons it is common to emphasise the fact that Zorn's Lemma depends on the Axiom of Choice, although in fact there are many other places in topology (and other fields of mathematics) where that axiom is used, often without explicit mention. We will at least give a formal statement here.

Definition 35.11. [defn-choice-function]
Let X be a set, and let $\mathcal{P}'(X)$ denote the set of nonempty subsets of X. A *choice function* for X is a function $c\colon \mathcal{P}'(X) \to X$ such that $c(A) \in A$ for all $A \in \mathcal{P}'(X)$. In other words, for every nonempty subset $A \subseteq X$, the function c gives a "chosen element" $c(A) \in A$. The *Axiom of Choice* says that every set has a choice function.

It is standard to take the Axiom of Choice as one of the basic assumptions of set theory. This is philosophically displeasing, because for many sets X we cannot specify a choice function explicitly, because there are too many arbitrary choices to make. However, in Remark 35.29 we will mention some ideas that may alleviate this displeasure.

Example 35.12. [eg-countable-choice]
We can define a choice function $c_{\mathbb{N}}\colon \mathcal{P}'(\mathbb{N}) \to \mathbb{N}$ by $c_{\mathbb{N}}(A) = \min(A)$. If we have a surjective map $f\colon \mathbb{N} \to X$ we can define a choice function $c_f\colon \mathcal{P}'(X) \to X$ by

$$c_f(A) = f(\min(f^{-1}(A))) = f(\min(\{n \;:\; f(n) \in A\})).$$

For most countable sets X that occur in practice, one can write down an explicit surjection $f\colon \mathbb{N} \to X$, so this procedure gives an explicit choice function.

Example 35.13. Suppose we have sets X and Y with choice functions c and d. Consider subset $A \subseteq X \times Y$. For $x \in X$ we put $A[x] = \{y \;:\; (x,y) \in A\}$, and then we put $\pi(A) = \{x \;:\; A[x] \neq \emptyset\}$. If $A \neq \emptyset$ we find that $\pi(A) \neq \emptyset$ so we can define

$$e(A) = (c(\pi(X)), d(A[c(\pi(X))])).$$

This gives a choice function for $X \times Y$. Less formally: to choose an element of A, we use c to choose an element x such that $A[x] \neq \emptyset$, then use d to choose $y \in A[x]$, then take (x,y) as our chosen element of A.

Example 35.14. [eg-choice-R]
Let $\mathcal{P}'_C(\mathbb{R})$ denote the set of nonempty closed subsets of $\mathbb{R}$. For any $A \in \mathcal{P}'_C(\mathbb{R})$ we can let n be the smallest integer such that $A \cap [-n,n] \neq \emptyset$, then let $c(A)$ be the largest element of $A \cap [-n,n]$ (which exists because $A \cap [-n,n]$ is compact). This defines a choice function for closed sets. Now consider instead the set $\mathcal{P}'_O(\mathbb{R})$ of nonempty open subsets of $\mathbb{R}$. For $B \in \mathcal{P}'_O(\mathbb{R})$ and $n \geq 1$ we put

$$F_n(B) = \{x \in \mathbb{R} \;:\; (x - 1/n, x + 1/n) \subseteq B\}.$$

One can check that this is closed and that B is the union of the sets $F_n(B)$, so in particular $F_n(B) \neq \emptyset$ for large n. Let m be the smallest integer such that $F_m(B) \neq \emptyset$ (so $F_m(B) \in \mathcal{P}'_C(\mathbb{R})$) and put $d(B) = c(F_m(B))$. This defines a choice function for open sets. By more elaborate constructions along similar lines, one can define choice functions for very large classes of subsets of $\mathbb{R}$, including all those that are likely to occur in applications. However, it does not seem to be possible to define a choice function for *all* nonempty subsets explicitly. **Extract reference from MathOverflow**

As an example of the use of the Axiom of Choice, we offer the following:

Proposition 35.15. [prop-epis-split]
Let $f\colon X \to Y$ be a surjective map. Then there is a map $g\colon Y \to X$ with $fg = 1_Y$.

Proof. Let c be a choice function for X. As f is surjective, we see that each set $f^{-1}\{y\} \subseteq X$ is nonempty, so we can define $g(y) = c(f^{-1}\{y\}) \in f^{-1}\{y\} \subseteq X$, so $g\colon Y \to X$ with $fg(y) = y$ as required. □

Normally we would just say "choose an element $g(y) \in f^{-1}\{y\}$ for each y" rather than referring explicitly to a choice function.

First proof of Zorn's Lemma. Let P be a poset in which every chain has an upper bound. In particular, the empty chain has an upper bound, so $P \neq \emptyset$. Let c be a choice function for P. We now put $f(0) = c(P) \in P$. If this is maximal then we are finished. Otherwise, we put $V(0) = \{p \;:\; p > f(0)\}$ and $f(1) = c(V(0))$ so $f(0) < f(1)$. If this is not maximal we put $V(2) = \{p \;:\; p > f(1)\}$ and $f(2) = c(V(2))$, and so on. It may be that after infinitely many steps we still do not have a maximal element. In that case we have a chain $C(\omega) = \{f(n) \;:\; n < \omega\}$, and we let $V(\omega)$ denote the set of upper bounds for this chain, and put $f(\omega) = c(V(\omega))$. If this is not maximal we put $V(\omega+1) = \{p \;:\; p > f(\omega)\}$ and $f(\omega+1) = c(V(\omega+1))$ and so on.

More formally, the construction is as follows. For each ordinal α we define $f(\alpha) \in P$ recursively as follows. Put $C(\alpha) = \{f(\beta) \;:\; \beta < \alpha\}$.

(a) If $C(\alpha)$ is not a chain, we put $f(\alpha) = c(P)$.
(b) Now suppose that $C(\alpha)$ is a chain, and let $U(\alpha)$ be the set of upper bounds for $C(\alpha)$, which is nonempty by assumption. If the set $V(\alpha) = U(\alpha) \setminus C(\alpha)$ is nonempty, we put $f(\alpha) = c(V(\alpha))$.
(c) This just leaves the case where $U(\alpha) \subseteq C(\alpha)$. If $u, v \in U(\alpha)$ then $u \in C(\alpha)$ and $v \in U(\alpha)$ so $u \leq v$, but also $v \leq u$ by symmetry, so $u = v$. Thus $U(\alpha)$ is a singleton, say $U(\alpha) = \{u\}$. In this case we put $f(\alpha) = u$.

We can now check by induction that $f(\alpha) \geq f(\beta)$ for all $\alpha \geq \beta$, and thus that clause (a) is never used. If clause (c) were never used we would have $f(\alpha) > f(\beta)$ for all $\alpha > \beta$, so f would be injective, contradicting fact (h) in Section 35.2. Thus for some α we must have $U(\alpha) \subseteq C(\alpha)$ and so $U(\alpha) = \{f(\alpha)\}$. If $f(\alpha) \leq p$ then it is clear that p is still an upper bound for $C(\alpha)$, so $p \in U(\alpha)$, so $p = f(\alpha)$. This proves that $f(\alpha)$ is maximal in P. □

We now start working towards an alternative proof that does not use ordinals.

Proposition 35.16. [`prop-zorn-special`]
Let X be a set, and let $\mathcal{Q}$ be a collection of subsets of X. We order $\mathcal{Q}$ by inclusion, making it a poset. Suppose that

(a) *Whenever $A \in \mathcal{Q}$ and $B \subseteq A$ we have $B \in \mathcal{Q}$.*
(b) *For every chain $\mathcal{C} \subseteq \mathcal{Q}$, the union*

$$\bigcup \mathcal{C} = \{x \in X \;:\; x \in A \text{ for some } A \in \mathcal{C}\}$$

is an element of $\mathcal{Q}$.

Then $\mathcal{Q}$ has a maximal element.

Proof. Let c be a choice function for X. For any set $A \in \mathcal{Q}$, we put

$$f(A) = \{x \in X \setminus A \;:\; A \cup \{x\} \in \mathcal{Q}\}.$$

If A is not maximal then we can choose $A' \in \mathcal{Q}$ with $A \subset A'$. Then, for any $x \in A' \setminus A$ we have $A \cup \{x\} \subseteq A' \in \mathcal{Q}$ so $A \cup \{x\} \in \mathcal{Q}$ by (a), so $x \in f(A)$, so $f(A) \neq \emptyset$. We can thus define $g\colon \mathcal{Q} \to \mathcal{Q}$ by

$$g(A) = \begin{cases} A & \text{if } A \text{ is maximal in } \mathcal{Q} \\ A \cup \{c(f(A))\} & \text{otherwise.} \end{cases}$$

Note that we can apply (b) to the empty chain to see that $\emptyset \in \mathcal{Q}$.

Next, we say that a subset $\mathcal{T} \subseteq \mathcal{Q}$ is a *tower* if

(p) Whenever $A \in \mathcal{T}$ we have $g(A) \in \mathcal{T}$.
(q) Whenever $\mathcal{C} \subseteq \mathcal{T}$ is a chain we have $\bigcup \mathcal{C} \in \mathcal{T}$.

For example, the set $\mathcal{Q}$ itself is a tower. Now put $A_0 = \emptyset$, $A_1 = g(\emptyset)$, $A_2 = g(A_1) = g^2(\emptyset)$ and so on. By the case $\mathcal{C} = \emptyset$ of (q) we have $A_0 \in \mathcal{T}$ for every tower $\mathcal{T}$, and thus $A_n \in \mathcal{T}$ for every tower $\mathcal{T}$ by induction using (p). Now put

$$\mathcal{T}_1 = \{A \in \mathcal{Q} \;:\; A \in \mathcal{T} \text{ for every tower } \mathcal{T}\}.$$

For example, we have $A_n \in \mathcal{T}_1$ for all n. From the definitions it is clear that $\mathcal{T}_1$ is itself a tower.

We would like to prove that $\mathcal{T}_1$ is a chain. We will use the following definitions:

(1) We say that a set $A \in \mathcal{T}_1$ is *comparable* if for all $B \in \mathcal{T}_1$ we have either $A \subseteq B$ or $B \subseteq A$.
(2) We let $\mathcal{T}_2 \subseteq \mathcal{T}_1$ be the set of comparable sets.
(3) For any comparable set A, we put

$$\mathcal{T}_3(A) = \{B \in \mathcal{T}_1 \;:\; B \subseteq A \text{ or } g(A) \subseteq B\}.$$

We will prove that $\mathcal{T}_3(A)$ is a tower, and thus that $\mathcal{T}_3(A) = \mathcal{T}_1$. Using this we will show that $\mathcal{T}_2$ is a tower, and thus that $\mathcal{T}_2 = \mathcal{T}_1$, or in other words that every set in $\mathcal{T}_1$ is comparable, or in other words that $\mathcal{T}_1$ is a chain.

First, let $\mathcal{C}$ be a chain contained in $\mathcal{T}_3(A)$, so for every set $B \in \mathcal{C}$ we have either $B \subseteq A$ or $g(A) \subseteq B$. If there is some set B with $g(A) \subseteq B$, then clearly $g(A) \subseteq \bigcup \mathcal{C}$. Otherwise, all the sets $B \in \mathcal{C}$ must have $B \subseteq A$, so $\bigcup \mathcal{C} \subseteq A$. This proves that $\mathcal{T}_3(A)$ has property (q). For property (p), consider a set $B \in \mathcal{T}_3(A)$. By the definition of $\mathcal{T}_3(A)$, one of the following three cases must occur:

(i) Suppose that B is a proper subset of A. Now $g(B) \in \mathcal{T}_1$ and A is assumed to be comparable, so either $g(B) \subseteq A$ or A is a proper subset of $g(B)$. In the latter case we would have $B \subset A \subset g(B)$ so $g(B) \setminus B$ would have size at least two, which is impossible by the definition of g. We must therefore have $g(B) \subseteq A$, so $g(B) \in \mathcal{T}_3(A)$.

(ii) Suppose that $B = A$; then $g(B) = g(A) \in \mathcal{T}_3(A)$.

(iii) Suppose that $g(A) \subseteq B$. As $B \subseteq g(B)$ we also have $g(A) \subseteq g(B)$, so $g(B) \in \mathcal{T}_3(A)$.

This shows that $\mathcal{T}_3(A)$ has property (p), so it is a tower, so it contains $\mathcal{T}_1$ by the definition of $\mathcal{T}_1$. On the other hand, $\mathcal{T}_3(A)$ is visibly a subset of $\mathcal{T}_1$, so $\mathcal{T}_3(A) = \mathcal{T}_1$. From this it is clear that $g(A)$ is comparable, so $\mathcal{T}_2$ has property (p).

Now consider a chain $\mathcal{C}$ contained in $\mathcal{T}_2$, and a set $B \in \mathcal{T}_1$. As every set $A \in \mathcal{C}$ is comparable, we have either $A \subseteq B$ or $B \subseteq A$. If $B \subseteq A$ for some $A \in \mathcal{C}$, it is clear that $B \subseteq \bigcup \mathcal{C}$. Otherwise we must have $A \subseteq B$ for all $A \in \mathcal{C}$, so $\bigcup \mathcal{C} \subseteq B$. Using this we see that $\bigcup \mathcal{C}$ is comparable. This means that $\mathcal{T}_2$ has property (q) as well as property (p), so it is a tower contained in $\mathcal{T}_1$, so it must be all of $\mathcal{T}_1$. This means that all elements of $\mathcal{T}_1$ are comparable, so $\mathcal{T}_1$ is a chain.

Now put $T = \bigcup \mathcal{T}_1$, so for all $A \in \mathcal{T}_1$ we have $A \subseteq T$. As $\mathcal{T}_1$ is both a tower and a chain, we see that $T \in \mathcal{T}_1$, and then that $g(T) \in \mathcal{T}_1$, so $g(T) \subseteq T$. By inspecting the definition of g, we deduce that T is maximal in $\mathcal{Q}$. □

REMARK 35.17. [`rem-ordinal-tower`]

If we allowed ourselves to use ordinals we could define $A_\alpha \in \mathcal{Q}$ recursively by $A_{\alpha+1} = g(A_\alpha)$, and $A_\lambda = \bigcup_{\alpha<\lambda} A_\alpha$ for limit ordinals α. We would then find that $\mathcal{T}_1$ is just the set of all A_α's.

SECOND PROOF OF ZORN'S LEMMA. Let P be a poset in which every chain has an upper bound. Let $\mathcal{P}$ denote the set of chains in P, ordered by inclusion. We claim that this satisfies the conditions of Proposition 35.16. Indeed, it is clear that every subset of a chain is a chain, so (a) holds. Now let $\mathcal{C}$ be a chain of chains, and suppose we have $p, q \in \bigcup \mathcal{C}$. As $p \in \bigcup \mathcal{C}$ there exists $C \in \mathcal{C}$ with $p \in C$. As $q \in \bigcup \mathcal{C}$ there exists $D \in \mathcal{C}$ with $q \in D$. As $\mathcal{C}$ is a chain we have either $C \subseteq D$ or $D \subseteq C$. Without loss of generality we may assume that $C \subseteq D$, so $p, q \in D$. As D is a chain we see that either $p \leq q$ or $q \leq p$. This proves that $\bigcup \mathcal{C}$ is a chain, so hypothesis (b) also holds. The proposition therefore tells us that there is a maximal chain $C \subseteq P$. By assumption, this chain has an upper bound, say m. This means that $C \cup \{m\}$ is a chain, but C is maximal, so we must have $m \in C$. Now suppose we have $m \leq p$. This means that p is another upper bound for C, so by the same logic $p \in C$, but m is an upper bound for C, so $p \leq m$, so $p = m$. Thus, m is maximal in P. □

We now explain a standard algebraic application of Zorn's Lemma, by way of example.

PROPOSITION 35.18. [`prop-maximal-ideal`]

Let R be a nonzero ring. Then there is an ideal $M < R$ that is maximal among all proper ideals.

PROOF. Let $\mathcal{P}$ denote the set of all proper ideals of R. Note that an ideal $I \leq R$ is proper iff $1 \notin I$, and that $0 \in \mathcal{P}$ because $R \neq 0$. Now let $\mathcal{C}$ be a chain in $\mathcal{P}$, and put $I = \{0\} \cup \bigcup \mathcal{C}$ (the zero only being necessary in the degenerate case $\mathcal{C} = \emptyset$). If $a, b \in \bigcup \mathcal{C}$ then we can find ideals $J, K \in \mathcal{C}$ with $a \in J$ and $b \in K$. As $\mathcal{C}$ is a chain we have either $J \leq K$ or $K \leq J$, and we can harmlessly assume the former. We then have $a, b \in K$ so $a \pm b \in K \subseteq I$ and similarly $ra, rb \in K \subseteq I$ for all $r \in R$. This proves that I is an ideal. For all $J \in \mathcal{C}$ we have $1 \notin J$, so it follows that $1 \notin I$, so $I \in \mathcal{P}$. It is clear that I is an upper bound for $\mathcal{C}$, so we have verified the conditions of Zorn's Lemma and the claim follows. □

DEFINITION 35.19. [`defn-hamel-basis`]

Let K be a field, and let V be a vector space (possibly not finite-dimensional) over K. Consider a set $X \subseteq V$. For any map $a\colon X \to K$, we put $\operatorname{supp}(a) = \{x \;:\; a(x) \neq 0\}$, and then we put

$$K\{X\} = \{a\colon X \to K \;:\; \operatorname{supp}(a) \text{ is finite }\}.$$

We define $\sigma_X\colon K\{X\} \to V$ by

$$\sigma_X(a) = \sum_{x\in X} a(x).x = \sum_{x\in\operatorname{supp}(a)} a(x).x$$

(The second expression makes it clear that this is a well-defined finite sum, but the first expression is conceptually more natural.) We say that X is a *basis* for V if σ_X is an isomorphism.

Proposition 35.20. [`prop-hamel-basis`]
Every vector space has a basis.

Proof. Let V be a vector space over K, and put

$$\mathcal{P} = \{X \subseteq V \;:\; \sigma_X \text{ is injective }\}.$$

We order this by inclusion and so regard it as a poset. Consider a chain $\mathcal{C} \subseteq \mathcal{P}$, and put $X = \bigcup\mathcal{C}$. Consider an element $a \in K\{X\}$ with $\sigma_X(a) = 0$. The support of a must be finite, say $\operatorname{supp}(a) = \{w_1, \ldots, w_n\}$. As $w_i \in X = \bigcup\mathcal{C}$, there exists $W_i \in \mathcal{C}$ with $w_i \in W_i$. As $\mathcal{C}$ is a chain, the sets $W_1, \ldots, W_n$ are all comparable, so after renumbering them if necessary we may assume that $W_1 \subseteq \cdots \subseteq W_n$. We then have $\operatorname{supp}(a) = \{w_1, \ldots, w_n\} \subseteq W_n$, which means that $\sigma_{W_n}(a|_{W_n}) = \sigma_X(a) = 0$. As $W_n \in \mathcal{C} \subseteq \mathcal{P}$ it follows that $a|_{W_n} = 0$, but $\operatorname{supp}(a) \subseteq W_n$, so $a = 0$. This proves that σ_X is injective, so $X \in \mathcal{P}$. It is clear that X is an upper bound for $\mathcal{C}$, so we have checked the hypothesis for Zorn's Lemma. It follows that there is a maximal element $X \in \mathcal{P}$. We claim that this is a basis. To see this, consider an element $v \in V$. If $v \in X$ we can define $a \in K\{X\}$ by $a(v) = 1$ and $a(x) = 0$ for $x \neq v$, and we have $\sigma_X(a) = v$. Suppose instead that $v \notin X$. As X is maximal in $\mathcal{P}$, we see that $\sigma_{X\cup\{v\}}$ cannot be injective, so there is a nonzero element $a \in K\{X \cup \{v\}\}$ with $\sigma_{X\cup\{v\}}(a) = 0$. If $a(v) = 0$ then we find that $a|_X$ is a nonzero element of $K\{X\}$ with $\sigma_X(a|_X) = 0$, contradicting the fact that $X \in \mathcal{P}$. We therefore have $a(v) \neq 0$, so we can put $b = -(a|_X)/a(v) \in K\{X\}$ and we find that $\sigma_X(b) = v$. Thus σ_X is surjective as required. $\square$

We conclude with another application that will be more directly useful for us.

Definition 35.21. [`defn-well-order`]
A *well-ordering* on a set I is a partial ordering with the following property: for every nonempty subset $J \subseteq I$, there is an element $j_0 \in J$ such that $j_0 \leq j$ for all $j \in J$. (In other words, every nonempty subset has a smallest element.) It is clear that such an element j_0 is unique if it exists.

Example 35.22. [`eg-well-order`]
The obvious ordering of $\mathbb{N}$ is a well-ordering, as is the obvious ordering on $\mathbb{N} \cup \{\infty\}$. The obvious ordering on $\mathbb{Z}$ is not a well-ordering, because the whole set does not have a smallest element. We can choose a bijection $f\colon \mathbb{N} \to \mathbb{Z}$ (for example, by setting $f(2n) = n$ and $f(2n + 1) = -n - 1$) and use this to transfer the standard ordering of $\mathbb{N}$ to a nonstandard ordering of $\mathbb{Z}$ that is a well-ordering. Alternatively, we can specify a well-ordering on $\mathbb{Z}$ by the rules

$$0 < 1 < 2 < 3 < 4 < \cdots < -1 < -2 < -3 < \cdots .$$

The obvious ordering on $[0, \infty)$ is not a well-ordering either, because the subset $(0, 1)$ does not have a smallest element.

Example 35.23. [`eg-ordinal-order`]
For any set α, we can try to define an ordering on α as follows: we have $x \leq y$ iff ($x = y$ or $x \in y$). As discussed in Section 35.2, an ordinal is precisely a set for which this rule defines a well-ordering.

Remark 35.24. [`rem-ordinal-ops`]
Let I and J be well-ordered sets. We can define an ordering on $I \times J$ as follows: we have $(i, j) < (i', j')$ iff $i < i'$, or ($i = i'$ and $j < j'$). This is called the *lexicographic ordering*. If A is a nonempty subset of $I \times J$ we can let i_0 be the smallest element of I such that A meets $\{i_0\} \times J$, then let j_0 be the smallest element of J such that $(i_0, j_0) \in \{i_0\} \times J$. We find that (i_0, j_0) is the smallest element in A, so $I \times J$ is again well-ordered.

Now consider instead the set $I \amalg J$. There is an obvious way to order this such that every element of I comes before every element of J. More explicitly, we declare that $k \leq k'$ if and only if

(a) $k, k' \in I$ and $k \leq k'$ with respect to the given order on I; or
(b) $k \in I$ and $k' \in J$; or
(c) $k, k' \in J$ and $k \leq k'$ with respect to the given order on J.

We write $I : J$ for the set $I \amalg J$ equipped with this relation, which is easily seen to be a well-ordering.

Addition and multiplication of ordinals are characterised by the following properties: for any ordinals α and β, the ordered set $\alpha \times \beta$ is isomorphic to the ordered set $\alpha\beta$, and the ordered set $\alpha : \beta$ is isomorphic to the ordered set $\alpha + \beta$.

Remark 35.25. [`rem-wo-restrict`]
If I is a well-ordered set and J is a subset of I, we can restrict the ordering on I to get an ordering on J, and it is clear that this is again a well-ordering.

Remark 35.26. [`rem-well-ordered-choice`]
If we have a well-ordering of I, we can use it to define a choice function $c\colon \mathcal{P}'(I) \to I$ just by taking $c(J)$ to be the smallest element in J.

Theorem 35.27. [`thm-well-order`]
Every set admits a well-ordering.

This is another highly non-constructive result. There does not seem to be any hope of exhibiting an explicit well-ordering of $\mathbb{R}$, for example. **Reference?**

Proof. In this proof we will use letters such as R (rather than the symbol $\leq$) for generic orderings. Fix a set I, and let $\mathcal{P}$ denote the set of pairs (J, R), where $J \subseteq I$ and R is a well-ordering of J. If $(J, R) \in \mathcal{P}$ and $K \subseteq J$ then we can restrict R to give a well-ordering of K, which we denote by $R|_K$. Consider two elements (K, S) and (J, R) in $\mathcal{P}$. We say that (K, S) is an *initial segment* in (J, R) if

(a) $K \subseteq J$ and $S = R|_K$.
(b) Whenever $j \in J$ and $k \in K$ and jRk we have $j \in K$.

We order $\mathcal{P}$ by declaring that $(K, S) \leq (J, R)$ iff (K, S) is an initial segment in (J, R).

Let $\mathcal{C}$ be a chain in $\mathcal{P}$. Put

$$J = \{k \in I \;:\; k \in K \text{ for some } (K, S) \in \mathcal{C}\}.$$

Consider a pair of points $j, k \in J$. We claim that there is an element $(K, S) \in \mathcal{C}$ such that j and k both lie in K. Indeed, from the definitions we see that there is an element $(K_0, S_0) \in \mathcal{C}$ with $j \in K_0$, and another element $(K_1, S_1) \in \mathcal{C}$ with $k \in K_1$. As $\mathcal{C}$ is a chain we have either $K_0 \subseteq K_1$ or $K_1 \subseteq K_0$. In the former case we take $(K, S) = (K_1, S_1)$, and in the latter case we take $(K, S) = (K_0, S_0)$.

Now suppose we have two different elements in $\mathcal{C}$, say (K, S) and (L, T), where $j, k \in K$ and also $j, k \in L$. As $\mathcal{C}$ is a chain we have either $(K, S) \leq (L, T)$ or $(L, T) \leq (K, S)$. In the former case we have $S = T|_K$ and in the latter we have $T = S|_L$. Either way, we see that jSk iff jTk. There is thus a well-defined relation R given by jRk iff (jSk for every $(K, S) \in \mathcal{C}$ with $j, k \in K$). It is easy to see that this is a total order, and that for all $(K, S) \in \mathcal{C}$ we have $S = R|_K$.

Consider a nonempty subset $A \subseteq J$. Choose any element $a \in A$, then choose an element $(K, S) = (K, R|_K) \in \mathcal{C}$ with $a \in K$, then let b be the smallest element in $A \cap K$ with respect to $R|_K$. We claim that b is actually smallest in A with respect to R. To see this, consider an arbitrary element $c \in A \subseteq J$. We can then find $(L, R|_L) \in \mathcal{C}$ such that $c \in L$. As $\mathcal{C}$ is a chain we either have $(L, R|_L) \leq (K, R|_K)$ or $(K, R|_K) \leq (L, R|_L)$. In the former case we have $c \in A \cap L \subseteq A \cap K$ so $b \leq c$ by the defining property of b. In the latter case we note that K is an initial segment of L, and it again follows that b is smallest in $A \cap L$ and so $b \leq c$. This shows that b is smallest in A as claimed, so R is a well-ordering of J. This means that (J, R) is an upper bound for $\mathcal{C}$, so we have checked the hypothesis of Zorn's Lemma. It follows that there is a maximal element $(J, R) \in \mathcal{P}$. We claim that J is actually equal to I. If not, we can choose any element $i \in I \setminus J$ and give the set $J' = J \cup \{i\}$ the obvious order R' for which i is largest in J'; this gives an element $(J', R') \in \mathcal{P}$ strictly larger than (J, R), which is a contradiction. Thus R is the required well-ordering of I itself. □

Proposition 35.28. [prop-woset-ordinal]
Let I be a well-ordered set. Then there is an ordinal λ and an order-preserving bijection $f\colon S(\lambda) \to I$.

Proof. Take a point $* \not\in I$ and define $f_0(\alpha) \in I \cup \{*\}$ for all ordinals α by the recursive rule

$$f_0(\alpha) = \begin{cases} \text{smallest element of } I \setminus f_0(S(\alpha)) & \text{if } I \setminus f_0(S(\alpha)) \neq \emptyset \\ * & \text{if } I \setminus f_0(S(\alpha)) = \emptyset. \end{cases}$$

It is clear by construction that if $\alpha < \beta$ and $f_0(\alpha), f_0(\beta) \neq *$ then $f_0(\alpha) \neq f_0(\beta)$. If we had $f_0(\alpha) \neq *$ for all α then f_0 would give injections $S(\lambda) \to I$ for all α, which would contradict fact (h) in Section 35.2. We can thus let λ be the least ordinal such that $f_0(\lambda) = *$, and we see that f_0 restricts to give an injective map $f\colon S(\lambda) \to I$. Moreover, as $f_0(\lambda) = *$ we have $I \setminus f(S(\lambda)) = \emptyset$ and so f is also surjective. Now suppose we have $\alpha < \beta < \lambda$. As f is injective, we see that $f(\alpha)$ and $f(\beta)$ are both elements of $I \setminus f(S(\alpha))$, but $f(\alpha)$ is by definition the smallest element of that set, so $f(\alpha) < f(\beta)$. Thus, the bijection $f\colon S(\lambda) \to I$ is order-preserving. $\square$

Remark 35.29. [rem-constructible]
Write this

36. Categories and functors

[apx-categories]

36.1. Basics.

Definition 36.1. [defn-category]
A *category* $\mathcal{C}$ consists of:

- A class $\mathrm{obj}(\mathcal{C})$ of *objects*.
- For each pair of objects X and Y, a set $\mathcal{C}(X,Y)$ of *morphisms* from X to Y. We write $f\colon X \to Y$ or $X \xrightarrow{f} Y$ to indicate that $f \in \mathcal{C}(X,Y)$.
- For each object X a morphism $1_X \in \mathcal{C}(X,X)$, called the *identity morphism* for X.
- For each triple of objects X, Y and Z and each pair of morphisms $f\colon X \to Y$ and $g\colon Y \to Z$, a morphism $g \circ f\colon X \to Z$, called the *composite* of g and f. Where convenient we will just write gf rather than $g \circ f$.

These must satisfy the following axioms:

C0: For all $X \xrightarrow{f} Y$ we have $f \circ 1_X = f = 1_Y \circ f$.
C1: For all $W \xrightarrow{e} X \xrightarrow{f} Y \xrightarrow{g} Z$ we have $g \circ (f \circ e) = (g \circ f) \circ e\colon W \to Z$.

Remark 36.2. [rem-in-obj]
By a slight abuse of notation, we will often write $X \in \mathcal{C}$ (rather than $X \in \mathrm{obj}(\mathcal{C})$) to indicate that X is an object of $\mathcal{C}$.

Example 36.3. [eg-cat-sets]
There is a category called **Sets**. The objects are sets, and the morphisms from X to Y are just the functions from X to Y. The identity morphism $1_X\colon X \to X$ is just the usual identity function, and the composition rule is just $(g \circ f)(x) = g(f(x))$.

Remark 36.4. [rem-proper-class]
Various paradoxes arise if one tries to talk about the set of all sets. Because of this, accounts of axiomatic set theory have been developed which specify carefully how sets can be described and constructed, and these rules mean that the collection of all sets cannot itself be regarded as a set; instead, it is a “proper class”. It is because of this that we say that a category has a class (rather than a set) of objects. We will not delve into these questions beyond a few passing remarks.

Example 36.5. [eg-cat-groups]
There is a category called **Groups**. The objects are groups, and the morphisms from G to H are the group homomorphisms. The identity morphisms and the composition rule are the same as for **Sets**. We can

define a category **Rings** (of rings and ring homomorphisms) and a category $\mathbf{Vect}_{\mathbb{R}}$ (of vector spaces over $\mathbb{R}$ and linear maps) in an analogous way.

Example 36.6. [`eg-cat-top`]

There is a category **Spaces** whose objects are topological spaces, and whose morphisms are continuous maps. Once again, the identity morphisms and the composition rule are the same as for **Sets**.

In some sense, examples like those above are the point of category theory. However, it turns out to be very useful to compare and relate such examples to much smaller categories of various types, some of which we introduce below.

Example 36.7. [`eg-bG`]

Fix a group G. We can define a category bG as follows. There is only one object, called 1. The set $bG(1,1)$ (of morphisms from 1 to 1) is just G itself. Thus, if $f, g \in bG(1,1) = G$ then we can use the group law of G to define $gf \in G$. The composition rule for bG is just $g \circ f = gf$. The identity morphism is the identity element of the group. Many authors refer to bG as "the group G, considered as a category" rather than using the notation bG.

Example 36.8. [`eg-sP`]

Let P be a partially ordered set, so we have a relation $\leq$ on P such that for all $x, y, z \in P$ we have

- $x \leq x$.
- If $x \leq y$ and $y \leq z$ then $x \leq z$.
- If $x \leq y$ and $y \leq x$ then $x = y$.

We can define a category sP as follows. The objects are just the elements of P. The morphism sets are

$$sP(x,y) = \begin{cases} \{(x,y)\} & \text{if } x \leq y \\ \emptyset & \text{if } x \not\leq y. \end{cases}$$

(The real point here is that there is a single morphism from x to y whenever $x \leq y$. It is not important what that morphism is, but we take it to be the pair (x,y).) The identity morphism 1_x is the pair (x,x), and the composition rule is $(y,z) \circ (x,y) = (x,z)$. Many authors refer to sP as "the poset P, considered as a category" rather than using the notation sP.

Example 36.9. [`eg-discrete-category`]

For any set X, we can define two different categories dX and eX. In both cases, the objects are the elements of X. In dX, the only morphisms are the identity morphisms 1_x for each $x \in X$. Categories of this type are called *discrete categories*. In eX, there is precisely one morphism u_{xy} from x to y for each pair of objects x and y. The composition rule is $u_{yz} \circ u_{xy} = u_{xz}$, and the identity morphism for x is u_{xx}. Categories of this type are called *indiscrete categories*.

Example 36.10. [`eg-cat-square`]

We can describe certain small categories by drawing pictures. For example, consider the pictures below:

$$\begin{tikzcd} 0 \arrow[r, "p"] \arrow[d, "q"'] & 1 \arrow[d, "r"] \\ 2 \arrow[r, "s"'] & 3 \end{tikzcd} \qquad\qquad \begin{tikzcd} \bullet \arrow[r] \arrow[d] & \bullet \arrow[d] \\ \bullet \arrow[r] & \bullet \end{tikzcd}$$

The picture on the left refers to a category with four objects (namely, the numbers $0, 1, 2, 3$) and nine morphisms. There are four morphisms $p\colon 0 \to 1$, $q\colon 0 \to 2$, $r\colon 1 \to 3$ and $s\colon 2 \to 3$. There is also a composite morphism $r \circ p = s \circ q\colon 0 \to 3$ and identity morphisms $1_0, 1_1, 1_2$ and 1_3, none of which are shown explicitly. Often when we use such categories it will not be necessary to name the objects or morphisms so we will just draw the picture on the right instead.

Example 36.11. [`eg-cat-mat`]

We can define a category $\mathbf{Mat}_R$ as follows. The objects are the natural numbers. The morphisms from n to m are the $m \times n$ matrices with entries in $\mathbb{R}$. The identity morphism 1_n is the $n \times n$ identity matrix, and the composition rule is $A \circ B = AB$.

Definition 36.12. Let $\mathcal{C}$ be a category. We can define a new category $\mathcal{C}^{\mathrm{op}}$ (called the *dual* of $\mathcal{C}$) as follows. The objects of $\mathcal{C}^{\mathrm{op}}$ are the same as the objects of $\mathcal{C}$. For each morphism $f \in \mathcal{C}(X,Y)$, there is a morphism $f^{\mathrm{op}} \in \mathcal{C}^{\mathrm{op}}(Y,X)$. The identities are just the morphisms 1_X^{op}, where 1_X is the identity morphism in $\mathcal{C}$. The composition rule is $f^{\mathrm{op}} \circ g^{\mathrm{op}} = (g \circ f)^{\mathrm{op}}$. In other words, $\mathcal{C}^{\mathrm{op}}$ is "the same as $\mathcal{C}$ but with the arrows turned around".

Remark 36.13. Whenever we have a definition or construction or theorem that works for all categories $\mathcal{C}$, we can generate a new definition or construction or theorem by applying the old one to $\mathcal{C}^{\mathrm{op}}$. This idea is called *duality*; when used judiciously it can save a lot of work.

36.2. Special classes of morphisms.

Definition 36.14. [defn-morphism-types]
Let $\mathcal{C}$ be a category, and let $f\colon X \to Y$ be a morphism in $\mathcal{C}$.

(a) Consider a morphism $g\colon Y \to X$. We say that g is a *left inverse* for f if $gf = 1_X$, and a *right inverse* for f if $fg = 1_Y$. We say that g is an *inverse* (or *two-sided inverse*, where emphasis is necessary) if it is both a left inverse and a right inverse.
(b) We say that f is a *monomorphism* if whenever we have maps $p, q\colon W \to X$ with $fp = fq\colon W \to Y$, we actually have $p = q$.
(c) We say that f is an *epimorphism* if whenever we have maps $r, s\colon Y \to Z$ with $rf = sf\colon X \to Z$, we actually have $r = s$.
(d) We say that f is a *bimorphism* if it is both a monomorphism and an epimorphism.
(e) We say that f is a *split monomorphism* if it has a left inverse, and a *split epimorphism* if it has a right inverse. If it has a two-sided inverse, we say that f is an *isomorphism*.
(f) We say that objects X and Y are *isomorphic* if there exists an isomorphism from X to Y. If so, we write $X \simeq Y$.

Remark 36.15. It is clear that f is a monomorphism (or epimorphism, or split monomorphism, or split epimorphism) in $\mathcal{C}$ if and only if f^{op} is an epimorphism (or monomorphism, or split epimorphism, or split monomorphism) in $\mathcal{C}^{\mathrm{op}}$.

We first check that the terminology in (e) above is sensible.

Lemma 36.16. *Any split monomorphism is a monomorphism, and any split epimorphism is an epimorphism.*

Proof. We will prove the first statement, and then the second one will follow by duality. If $f\colon X \to Y$ is a split monomorphism, then we can choose a left inverse $g\colon Y \to X$ with $gf = 1_X$. Now suppose we have $p, q\colon W \to X$ with $f \circ p = f \circ q$, as in the definition of monomorphisms. We then have $g \circ (f \circ p) = g \circ (f \circ q)$. Here $g \circ (f \circ p) = (g \circ f) \circ p = 1_X \circ p = p$, and similarly $g \circ (f \circ q) = q$. It follows that $p = q$ as required. □

Corollary 36.17. *Any isomorphism is a bimorphism.* □

We now discuss some examples.

Example 36.18. In the category bG every morphism is an isomorphism (and thus has all of the other properties mentioned above). In the category sP, every morphism is a bimorphism. The identity morphisms are isomorphisms, but no other morphism is a split monomorphism, a split epimorphism, or an isomorphism.

Proposition 36.19. [prop-sets-mono-epi]
Consider the category of sets.

(a) *The monomorphisms are the same as the injective maps.*
(b) *Every monomorphism $f\colon X \to Y$ is split, except when $X = \emptyset$ and $Y \neq \emptyset$.*
(c) *The epimorphisms are the surjective maps, and these are all split.*
(d) *The isomorphisms are the same as the bimorphisms which are the same as the bijective functions.*

Proof. (a) Consider an injective map $f\colon X \to Y$. Suppose we have $p, q\colon W \to X$ with $fp = fq$. This means that for all $w \in W$ we have $f(p(w)) = f(q(w))$, but f is injective, so we must have

$p(w) = q(w)$. This shows that $p = q$ as required, so f is a monomorphism. Conversely, suppose that f is a monomorphism. To show that it is injective, consider $x, x' \in X$ with $f(x) = f(x')$. Put $W = \{0\}$ and define $p, q\colon W \to X$ by $p(0) = x$ and $q(0) = x'$. Then $fp = fq$ but f is a monomorphism so $p = q$, so $x = x'$ as required.

(b) Now let $f\colon X \to Y$ be an injective map, or equivalently a monomorphism. If $X = Y = \emptyset$ then $1_\emptyset$ is an inverse for f, so f is split. Suppose instead that $X \neq \emptyset$. Choose a point $x_0 \in X$, and define $g\colon Y \to X$ by

$$g(y) = \begin{cases} x & \text{if } y = f(x) \text{ for some } x \in X \\ x_0 & \text{if } y \notin \operatorname{img}(f). \end{cases}$$

(Note that the first clause is well-defined because f is injective.) We then have $gf = 1_X$, so g is a left inverse for f.

(c) Consider a surjective map $f\colon X \to Y$. This means that for each $y \in Y$ there exists $x \in X$ with $f(x) = y$. We choose one such x and call it $g(y)$. (In general this may involve an infinite number of arbitrary choices, so to be set-theoretically rigorous we must appeal to the Axiom of Choice.) This gives a map $g\colon Y \to X$ with $fg = 1_Y$, so f is a split epimorphisms. Conversely, suppose that $f\colon X \to Y$ is an epimorphism. Put $Z = \{0, 1\}$ and define $r, s\colon Y \to Z$ by $r(y) = 1$ for all y, and

$$s(y) = \begin{cases} 1 & \text{if } y \in \operatorname{img}(f) \\ 0 & \text{if } y \notin \operatorname{img}(f). \end{cases}$$

We find that $rf = sf$, but f is assumed to be an epimorphism, so $r = s$. By inspecting the definition of r and s, this means that $\operatorname{img}(f)$ must be all of Y, so f is surjective.

(d) This is now clear from (a) to (c).

□

Proposition 36.20. [prop-rings-mono-epi]

Consider the category of commutative rings and ring homomorphisms. (Here all rings are implicitly assumed to have identity elements, and ring homomorphisms are required to preserve them.)

(a) *The monomorphisms are the same as the injective homomorphisms.*
(b) *Every split epimorphism is surjective.*
(c) *Every surjective homomorphism is an epimorphism.*
(d) *The quotient map $\mathbb{Z} \to \mathbb{Z}/2\mathbb{Z}$ is a surjective homomorphism that is not split.*
(e) *The inclusion $\mathbb{Z} \to \mathbb{Q}$ is an bimorphism that is not surjective (and so is not an isomorphism).*
(f) *The isomorphisms are precisely the bijective homomorphisms.*

Proof. (a) Every injective homomorphism is a monomorphism, by the same argument as for the category of sets. Conversely, let $f\colon R \to S$ be a monomorphism. Suppose we have $a, b \in R$ with $f(a) = f(b)$. Define $p, q\colon \mathbb{Z}[x] \to R$ by $p(\sum_i n_i x^i) = \sum_i n_i a^i$ and $q(\sum_i n_i x^i) = \sum_i n_i b^i$. These are homomorphisms with $p(x) = a$ and $q(x) = b$. We also have

$$fp(\sum_i n_i x^i) = \sum_i n_i f(a)^i = \sum_i n_i f(b)^i = fq(\sum_i n_i x^i),$$

so $fp = fq$. As f is a monomorphism, we deduce that $p = q$. In particular $p(x) = q(x)$, so $a = b$. This shows that f is injective.

(b) Suppose that $f\colon R \to S$ is a split epimorphism, so $fg = 1_S$ for some homomorphism $g\colon S \to R$. Then for all $b \in S$ we have $a = f(g(b))$, which shows that b is in the image of f. This means that f is surjective as claimed.

(c) Suppose that $f\colon R \to S$ is a surjective homomorphism. Consider a pair of homomorphisms $g, h\colon S \to T$ with $gf = hf$. For each $b \in S$ we can (by surjectivity) choose $a \in R$ with $f(a) = b$. It follows that $g(b) = gf(a) = hf(a) = h(b)$. As b was arbitrary we have $g = h$, as required.

(d) The quotient map $f\colon \mathbb{Z} \to \mathbb{Z}/2\mathbb{Z}$ is surjective, so it is an epimorphism by (c). Write $\overline{0}$ and $\overline{1}$ for the elements of $\mathbb{Z}/2\mathbb{Z}$. Any ring homomorphism $g\colon \mathbb{Z}/2\mathbb{Z} \to \mathbb{Z}$ would have to have $g(\overline{0}) = 0$ and $g(\overline{1}) = 1$ and thus $0 = g(\overline{0}) = g(\overline{1} + \overline{1}) = g(\overline{1}) + g(\overline{1}) = 1 + 1$, which is impossible. Thus, there are no morphisms $\mathbb{Z}/2\mathbb{Z} \to \mathbb{Z}$, so in particular f cannot have a right inverse.

(e) Let $f\colon \mathbb{Z} \to \mathbb{Q}$ be the inclusion, which is clearly not surjective. Suppose we have homomorphisms $g, h\colon \mathbb{Q} \to T$ for some ring T, with $gf = hf\colon \mathbb{Z} \to T$ say. This means that $g(a) = h(a)$ whenever $a \in \mathbb{Z} \subseteq \mathbb{Q}$. If $a \neq 0$ we note that $g(1/a)$ and $h(1/a)$ are both inverses for $g(a) = h(a)$, and inverses are unique, so $g(1/a) = h(1/a)$. Now consider a rational number $q \in \mathbb{Q}$, say $q = a/b$ for integers a, b with $b > 0$. We then have $g(a) = h(a)$ and $g(1/b) = h(1/b)$ by the above, so $g(q) = g(a)g(1/b) = h(a)h(1/b) = h(q)$. This means that $g = h$ as required, so f is an epimorphism and thus a bimorphism.

(f) Suppose that $f\colon R \to S$ is a bijective homomorphism, so there is a function $g\colon S \to R$ with $g(f(r)) = r$ for all $r \in R$ and $f(g(s)) = s$ for all $s \in S$. As f is a homomorphism we have $f(g(s)+g(s')) = f(g(s))+f(g(s')) = s+s'$. It follows that $g(s+s') = g(f(g(s)+g(s'))) = g(s)+g(s')$. Similar arguments show that $g(ss') = g(s)g(s')$ and $g(1) = 1$, so g is a ring homomorphism. Thus g is an inverse for f in **Rings**, so f is an isomorphism. The converse is trivial.

□

Proposition 36.21. [`prop-fields-mono-epi`]
*Let **Fields**$_p$ denote the category of fields of characteristic p (where p is zero or a prime number).*

(a) *All morphisms in **Fields**$_p$ are injective functions and thus are monomorphisms. The isomorphisms are precisely the bijective homomorphisms.*

(b) *If $f\colon K \to L$ is a morphism in **Fields**$_0$ then f is an epimorphism iff it is surjective iff it is an isomorphism.*

(c) *If $p > 0$ then we can define a homomorphism $\phi_K\colon K \to K$ by $\phi_K(a) = a^p$, and this is an epimorphism that need not be surjective.*

Proof.

(a) Let $f\colon K \to L$ be a field homomorphism. If $a \neq 0$ in K then we have $aa^{-1} = 1$ in K so $f(a)f(a^{-1}) = f(aa^{-1}) = f(1) = 1$ in L, so $f(a) \neq 0$. This shows that f is an injective function. Now suppose we have morphisms $g, h\colon M \to K$ with $fg = fh$. This means that for all $m \in M$ we have $f(g(m)) = f(h(m))$ but f is injective so $g(m) = h(m)$. As m was arbitrary we have $g = h$; so f is a monomorphism. We also see as in Proposition 36.20(f) that f is an isomorphism iff it is bijective.

(b) It would lead us too far astray to give a complete proof of this, but some indications are as follows. Let $f\colon K \to L$ be a homomorphism that is not surjective. It will be harmless to replace K by $f(K)$ and thus assume that f is just the inclusion of a subfield. Using Zorn's Lemma we can choose an intermediate field M such that $K \leq M \leq L$, and M is a purely transcendental extension of K and L is algebraic over M. Let $j\colon M \to \overline{M}$ be an algebraic closure for M. If L has finite degree $d > 1$ over M then Galois theory tells us that there are distinct maps $g_1, \dots, g_d\colon L \to \overline{M}$ extending j. As these agree on M we have $g_1 f = \cdots = g_d f$, so f is not an epimorphism. A similar approach works if L has infinite degree, but we need to use a version of Galois theory that covers infinite extensions, or an auxiliary argument with Zorn's lemma to pass from finite subextensions of L to L itself. This just leaves the case where $d = 1$, so $L = M$, so L is a purely transcendental extension of K. We thus have $L = K(x_i \,:\, i \in I)$ for some family of elements x_i, and we can define $g\colon L \to L$ by $g(x_i) = x_i + 1$ for all i. This is different from the identity but we have $gf = f$, so f is not an epimorphism.

(c) Now let K be a field of characteristic $p > 0$, and define $\phi_K\colon K \to K$ by $\phi_K(a) = a^p$. It is clear that $\phi_K(1) = 1$ and $\phi_K(ab) = \phi_K(a)\phi_K(b)$. One can also show (by considering prime factorisations of binomial coefficients) that $(x + y)^p = x^p + y^p \pmod{p}$, so $\phi_K(a + b) = \phi_K(a) + \phi_K(b)$ for all $a, b \in K$. This proves that ϕ_K is a homomorphism of fields.

Now suppose we have homomorphisms $g, h\colon K \to L$ with $g\phi_K = h\phi_K$. This means that for $a \in K$ we have $g(a^p) = h(a^p)$, or equivalently $g(a)^p = h(a)^p$, or equivalently $(g(a) - h(a))^p = 0$ in L. As L is a field we can only have $u^p = 0$ in L if $u = 0$, so $g(a) = h(a)$. Thus, ϕ_K is an epimorphism. If K is the field $(\mathbb{Z}/p)(t)$ of rational functions, then the image of ϕ_K is $(\mathbb{Z}/p)(t^p)$, so ϕ_K is not surjective.

□

We could ask for an analogue of Proposition 36.20 covering groups rather than rings. Most of this is easy and is left to the reader. We will just explain one point:

PROPOSITION 36.22. [`eg-group-epi`]
*Every epimorphism in the category **Groups** is surjective.*

PROOF. Let $p\colon G \to H$ be an epimorphism. Put $K = p(G)$, which is a subgroup of H. We can then form the set $H/K = \{hK \ :\ h \in K\}$ of cosets of K in H. For any $h \in H$ we have a bijection $m(h)\colon H/K \to H/K$ given by $m(h)(xK) = hxK$. Let a denote the coset K, and put $X = H/K \amalg \{b\}$ for some $b \notin H/K$. We extend $m(h)$ to give a map $X \to X$ by putting $m(h)(b) = b$. This construction gives a homomorphism $m\colon H \to \Sigma$, where Σ is the set of bijective maps $X \to X$, considered as a group under composition. Now define $s \in \Sigma$ by $s(a) = b$ and $s(b) = a$ and $s(x) = x$ for all $x \notin \{a, b\}$. Define $m'\colon H \to \Sigma$ by $m'(h) = s\,m(h)\,s$. As $s^2 = 1$ we have

$$m'(h_0)m'(h_1) = sm(h_0)s^2m(h_1)s = sm(h_0)m(h_1)s = sm(h_0h_1)s = m'(h_0h_1),$$

so m' is a homomorphism. It is easy to check that $m(h) = m'(h)$ iff $m(h)(a) = a$ iff $h \in K$. As $p(G) = K$, we deduce that $m \circ p = m' \circ p$. As p is assumed to be an epimorphism, it follows that $m(h) = m'(h)$ for all h, so $H = K$, so p is surjective. □

PROPOSITION 36.23. [`prop-inverses`]
Consider a morphism $f\colon X \to Y$ in a category $\mathcal{C}$.

(a) *If f is a monomorphism and a split epimorphism, then it is an isomorphism. More precisely, any right inverse for f is actually a two-sided inverse.*
(b) *If f is a split monomorphism and an epimorphism, then it is an isomorphism. More precisely, any left inverse for f is actually a two-sided inverse.*
(c) *If f is an isomorphism then it has a unique inverse, which is also the unique left inverse and the unique right inverse.*

PROOF. (a) Suppose that f is a monomorphism and a split epimorphism. The latter means that there is a right inverse morphism $g\colon Y \to X$ with $fg = 1_Y$. It follows that $fgf = 1_Yf = f = f1_X$, so the parallel maps $gf, 1_X\colon X \to X$ become the same when composed with $f\colon X \to Y$. As f is a monomorphism we can conclude that $gf = 1_X$, which means that g is a two-sided inverse for f.
(b) This is dual to (a).
(c) Suppose that f has an inverse, say g, so $fg = 1_Y$ and $gf = 1_X$. Let p be any left inverse for f, so $pf = 1_X$. Then $p = p1_Y = pfg = 1_Xg = g$, which shows that g is the only left inverse for f. Dually, g is the only right inverse for f. A fortiori, it is the only two-sided inverse for f.

□

DEFINITION 36.24. [`defn-inverse`]
If $f\colon X \to Y$ is an isomorphism, we write $f^{-1}\colon Y \to X$ for the inverse of f (which is well-defined by the above proposition).

PROPOSITION 36.25. [`prop-isos`]
Let $X \xrightarrow{f} Y \xrightarrow{g} Z$ be morphisms in a category $\mathcal{C}$.

(a) *1_X is an isomorphism (with inverse 1_X).*
(b) *If $f\colon X \to Y$ is an isomorphism, then so is $f^{-1}\colon Y \to X$ (with inverse f).*
(c) *If $f\colon X \to Y$ and $g\colon Y \to Z$ are isomorphisms, then $gf\colon X \to Z$ is an isomorphism (with inverse $f^{-1}g^{-1}$).*

PROOF. Clear. □

COROLLARY 36.26. (a) *Any object is isomorphic to itself.*
(b) *X is isomorphic to Y if and only if Y is isomorphic to X.*
(c) *If X is isomorphic to Y and Y is isomorphic to Z then X is isomorphic to Z.*

PROOF. Immediate from the proposition. □

Definition 36.27. [defn-groupoid]
A *groupoid* is a category in which every morphism is an isomorphism.

One very important example is the fundamental groupoid of a space, which we discuss in in Section 28. Here we will give some more algebraic examples.

Example 36.28. [eg-fields-fixed-order]
Fix a number q that is a power of a prime, and let $\mathcal{F}(q)$ be the category of finite fields of order q. Every homomorphism of fields is injective, so if the source and target have the same finite order then it must be an isomorphism. It follows that $\mathcal{F}(q)$ is a groupoid.

Example 36.29. [eg-translation-groupoid]
Let G be a group, and let X be a set with an action of G. We can define a groupoid $\mathrm{Trans}(G,X)$ as follows: the objects are the elements of X, and the morphisms from x to y are the elements $g \in G$ for which $g.x = y$. Composition is given by the group operation in G.

36.3. Subcategories.

Definition 36.30. Let $\mathcal{C}$ be a category. A *subcategory* of $\mathcal{C}$ is a category $\mathcal{D}$ such that:

(a) Every object of $\mathcal{D}$ is an object of $\mathcal{C}$.
(b) For every pair of objects X, Y of $\mathcal{D}$, the set $\mathcal{D}(X,Y)$ is a subset of $\mathcal{C}(X,Y)$.
(c) For every object X of $\mathcal{D}$, the identity morphism $1_X \in \mathcal{C}(X,X)$ lies in $\mathcal{D}(X,X)$.
(d) The composition rule for $\mathcal{D}$ is just the restriction of the composition rule for $\mathcal{C}$. In particular, if $f \in \mathcal{D}(X,Y) \subseteq \mathcal{C}(X,Y)$ and $g \in \mathcal{D}(Y,Z) \subseteq \mathcal{C}(Y,Z)$ then the composite $gf \in \mathcal{C}(X,Z)$ actually lies in $\mathcal{D}(X,Z)$.

Definition 36.31. Let $\mathcal{C}$ be a category, and let $\mathcal{D}$ be a subcategory.

(a) We say that $\mathcal{D}$ is *full* if for all objects $X, Y \in \mathcal{D}$ we have $\mathcal{D}(X,Y) = \mathcal{C}(X,Y)$.
(b) We say that $\mathcal{D}$ is *replete* if it is full, and whenever X is isomorphic to an object of $\mathcal{D}$, it actually lies in $\mathcal{D}$.
(c) We say that $\mathcal{D}$ is *skeletal* (or is a *skeleton* of $\mathcal{C}$) if it is full, and contains precisely one object in each isomorphism class of objects of $\mathcal{C}$.
(d) We say that $\mathcal{D}$ is *wide* if every object of $\mathcal{C}$ is an object of $\mathcal{D}$ (so $\mathcal{C}$ and $\mathcal{D}$ have the same objects).

Example 36.32. [eg-subcat-groups]
In the category **Groups** we have a replete subcategory **FinGroups** of finite groups, and another replete subcategory **AbGroups** of abelian groups. We can also let $\mathcal{D}$ be the category whose objects are all the subgroups of $\mathbb{Z}$, and whose morphisms are all group homomorphisms between such subgroups. This is a subcategory of **Groups** that is full but not replete.

Example 36.33. [eg-subcat-vs]
Let $\mathcal{C}$ be the category of finite-dimensional vector spaces over $\mathbb{R}$, and let $\mathcal{D}$ be the full subcategory whose objects are the space $\mathbb{R}^n$ for $n \in \mathbb{N}$. Then $\mathcal{D}$ is a skeleton of $\mathcal{C}$.

Example 36.34. [eg-wide-subcat]
Let $\mathcal{D}$ be the category whose objects are all groups, and whose morphisms are all surjective group homomorphisms. This is legitimate because identity homomorphisms are surjective, and the composite of any two surjective homomorphisms is again surjective. This defines a wide subcategory $\mathcal{D}$ of **Groups**.

36.4. Functors.

Definition 36.35. [defn-functor]
Let $\mathcal{C}$ and $\mathcal{D}$ be categories. A *functor* from $\mathcal{C}$ to $\mathcal{D}$ is a rule that assigns to each object $X \in \mathcal{C}$ and object $F(X) \in \mathcal{D}$, and to each morphism $f\colon X \to Y$ in $\mathcal{C}$ a morphism $F(f)\colon F(X) \to F(Y)$ in $\mathcal{D}$, in such a way that

F0: For each $X \in \mathcal{C}$ we have $F(1_X) = 1_{F(X)}$.
F1: For each composable pair of morphisms $X \xrightarrow{f} Y \xrightarrow{g} Z$ in $\mathcal{C}$, we have $F(gf) = F(g)F(f)$ in $\mathcal{D}(F(X), F(Z))$.

Remark 36.36. [rem-star-notation]

In cases where the notation for the functor F is cumbersome, it is common to write f_* rather than $F(f)$ for the map $F(X) \to F(Y)$ induced by a map $f: X \to Y$. With this notation the axioms say $(1_X)_* = 1_{FX}$ and $(gf)_* = g_* f_*$.

Remark 36.37. [rem-cat-cat]

For any category $\mathcal{C}$, there is an identity functor $1_{\mathcal{C}}: \mathcal{C} \to \mathcal{C}$. Given functors $F: \mathcal{C} \to \mathcal{D}$ and $G: \mathcal{D} \to \mathcal{E}$ we can compose them in an evident way to get a functor $GF: \mathcal{C} \to \mathcal{E}$. This almost means that we have a category of categories and functors, but in fact there are some set-theoretic difficulties which we shall not explore here.

Example 36.38. [eg-involutions]

We can define a functor F: **Groups** $\to$ **Sets** as follows. On objects, we put $F(G) = \{g \in G \;:\; g^2 = 1\}$. Given a homomorphism $\alpha: G \to H$ and an element $g \in F(G)$ we note that $\alpha(g)^2 = \alpha(g^2) = \alpha(1) = 1$, so $\alpha(g) \in F(H)$. We can thus put $F(\alpha) = \alpha|_{F(\alpha)}$, which is a map of sets from $F(G)$ to $F(H)$. It is clear that $F(1_G) = 1_{F(G)}$ and $F(\beta\alpha) = F(\beta)F(\alpha)$, so we have a functor as claimed.

Example 36.39. [eg-group-algebra]

Let **Rings** be the category of (not necessarily commutative) rings. For any finite group G, we let $\mathbb{C}[G]$ denote the set of all functions $a: G \to \mathbb{C}$, considered as a ring using the convolution product $(a * b)(y) = \sum_{x \in G} a(x)b(x^{-1}y)$. Equivalently, if we let $[g]: G \to \mathbb{C}$ denote the function that sends g to 1 and all other elements to 0, then the convolution product is characterised by the fact that $[g] * [h] = [gh]$. For any homomorphism $\alpha: G \to H$ and any $a \in \mathbb{C}[G]$ we define $\alpha_*(a) \in \mathbb{C}[H]$ by $\alpha_*(a)(h) = \sum_{\alpha(g)=h} a(g)$. Equivalently, this is characterised by the fact that $\alpha_*[g] = [\alpha(g)]$ for all $g \in G$. We find that $\alpha_*(a * b) = \alpha_*(a) * \alpha_*(b)$, so α_* is a homomorphism of rings. If $\beta: H \to K$ is another homomorphism we find that $\beta_*(\alpha_*([g])) = \beta_*([\alpha(g)]) = [\beta(\alpha(g))] = (\beta\alpha)_*([g])$. As the elements $[g]$ form a basis for $\mathbb{C}[G]$ over $\mathbb{C}$, it follows that $\beta_* \circ \alpha_* = (\beta\alpha)_*: \mathbb{C}[G] \to \mathbb{C}[K]$. This means that we can define a functor R: **FinGroups** $\to$ **Rings** by $R(G) = \mathbb{C}[G]$ on objects, and $R(\alpha) = \alpha_*$ on morphisms.

Example 36.40. [eg-centre-functor]

For any group G we let $Z(G)$ denote the centre of G. If $\phi: G \to H$ is a homomorphism, and $g \in Z(G)$, then $f(g)$ need not lie in $Z(H)$. For an example, take $G = C_2 = \{1, g\}$ and let H be the symmetric group Σ_3 and define $\phi: G \to H$ by $\phi(g) = (1\ 2)$. Thus, there is no obvious way to make Z into a functor. In fact, there is no way at all. To see this, let G and H be as above, and note that the signature gives a homomorphism $\psi: H \to G$ with $\psi\phi = 1_G$. If Z were a functor we would have homomorphisms $ZG \xrightarrow{\phi_*} ZH \xrightarrow{\psi_*} ZG$ whose composite is the identity. Here $ZG = G$ and $ZH = 1$ so this is clearly impossible.

Now let $\mathcal{D}$ be the wide subcategory of groups and *surjective* homomorphisms. If $\phi: G \to H$ is surjective, it is easy to check that ϕ carries central elements to central elements. Thus, we can define a functor $Z: \mathcal{D} \to$ **AbGroups** by $Z(\phi) = \phi|_{Z(G)}$.

Example 36.41. [eg-group-action]

Suppose we have a set X and a group G that acts on X. Thus, for each $g \in G$ we have a map $\mu_g: X \to X$ given by $\mu_g(x) = gx$, and these maps satisfy $\mu_1 = 1_X$ and $\mu_g \circ \mu_h = \mu_{gh}$. We can then define a functor $F: bG \to$ **Sets** which sends the unique object of bG to the set X, and sends each morphism g of bG to the map μ_g.

Example 36.42. Suppose we have any category $\mathcal{C}$, and some objects X_0, X_1, X_2 and X_3 in $\mathcal{C}$, and morphisms

$$\begin{tikzcd} X_0 \arrow[r, "f"] \arrow[d, "g"'] & X_1 \arrow[d, "h"] \\ X_2 \arrow[r, "k"'] & X_3 \end{tikzcd}$$

such that the square commutes (ie $kg = hf$). Let $\mathcal{Q}$ be the category described in Example 36.10. Then there is a functor $F: \mathcal{Q} \to \mathcal{C}$ given on objects by by $F(i) = X_i$, and on morphisms by $F(p) = f$, $F(q) = g$,

$F(r) = h$ and $F(s) = k$. Indeed, commutative squares in $\mathcal{C}$ are essentially the same thing as functors from $\mathcal{Q}$ to $\mathcal{C}$.

Example 36.43. [eg-mat-functor]
Let $\mathbf{Mat}_{\mathbb{R}}$ be as in Example 36.11, and let $\mathbf{FinVect}_{\mathbb{R}}$ be the category of finite dimensional vector spaces over $\mathbb{R}$ and linear maps. We can define $F\colon \mathbf{Mat}_{\mathbb{R}} \to \mathbf{FinVect}_{\mathbb{R}}$ as follows. On objects, we put $F(n) = \mathbb{R}^n$ (regarded as the space of column vectors of length n). On morphisms, if A is an $m \times n$ matrix we let $F(A)\colon \mathbb{R}^n \to \mathbb{R}^m$ be the linear map given by $F(A)(u) = Au$.

Example 36.44. [eg-forgetful]
We can define a functor $U\colon$ **Groups** $\to$ **Sets** as follows. On objects, $U(G)$ is just the underlying set of elements of the group G. For a homomorphism $\alpha\colon G \to H$, the function $U(\alpha)$ is just α itself. This is called the *forgetful functor*. There are similar forgetful functors from rings to sets, or from topological spaces to sets. There is also a partially forgetful functor from rings to abelian groups, that remembers the additive group structure and forgets the multiplication.

Example 36.45. [eg-inclusion-functor]
If $\mathcal{D}$ is a subcategory of $\mathcal{C}$ then there is an inclusion functor $J\colon \mathcal{D} \to \mathcal{C}$ given by $J(X) = X$ and $J(f) = f$.

Definition 36.46. Let $F\colon \mathcal{C} \to \mathcal{D}$ be a functor.
(a) We say that F is *faithful* if the maps $F\colon \mathcal{C}(X,Y) \to \mathcal{C}(F(X),F(Y))$ are all injective.
(b) We say that F is *full* if the maps $F\colon \mathcal{C}(X,Y) \to \mathcal{C}(F(X),F(Y))$ are all surjective.
(c) We say that F is *essentially surjective* if for all objects $U \in \mathcal{D}$ there exists an object $X \in \mathcal{C}$ such that U is isomorphic to $F(X)$.
(d) We say that F is an *equivalence* if it is full, faithful, and essentially surjective.

Example 36.47. [eg-func-props]
(a) Forgetful functors are faithful but not full.
(b) Consider the forgetful functor $U\colon$ **FinGroups** $\to$ **FinSets**. This is not essentially surjective, because the empty set does not biject with the underlying set of any group. Now let $\mathbf{FinSets}'$ denote the category of finite nonempty sets, and let $U'\colon$ **FinGroups** $\to \mathbf{FinSets}'$ be the forgetful functor. If $T \in \mathbf{FinSets}'$ then the order $n = |T|$ is nonzero, so there is a finite cyclic group $\mathbb{Z}/n\mathbb{Z}$ of order n, and we can choose a bijection $U'(\mathbb{Z}/n\mathbb{Z}) \to T$. This shows that U' is essentially surjective.
(c) It is a standard fact of linear algebra that every linear map $\mathbb{R}^n \to \mathbb{R}^m$ comes from a unique $m \times n$ matrix, and that every finite-dimensional vector space over $\mathbb{R}$ is isomorphic to $\mathbb{R}^n$ for some n. From this we see that the functor $\mathbf{Mat}_{\mathbb{R}} \to \mathbf{FinVect}_{\mathbb{R}}$ in Example 36.43 is an equivalence.

Definition 36.48. [defn-nat-trans]
Let F and G be two functors from $\mathcal{C}$ to $\mathcal{D}$. A *natural transformation* α from F to G consists of morphisms $\alpha_X\colon F(X) \to G(X)$ in $\mathcal{D}$ (for all $X \in \mathcal{C}$) such that for every morphism $f\colon X \to Y$ in $\mathcal{C}$, the following diagram commutes:

$$\begin{tikzcd} F(X) \arrow[r, "\alpha_X"] \arrow[d, "F(f)"'] & G(X) \arrow[d, "G(f)"] \\ F(Y) \arrow[r, "\alpha_Y"'] & G(Y). \end{tikzcd}$$

(This diagram is called a *naturality square*.) We say that α is a *natural isomorphism* if, in addition, all the morphisms $\alpha_X\colon F(X) \to G(X)$ are isomorphisms.

Remark 36.49. If α is a natural isomorphism from F to G, it is straightforward to check that the inverse maps $\alpha_X^{-1}\colon G(X) \to F(X)$ comprise another natural isomorphism from G to F.

Example 36.50. [eg-nat-trans-i]
Define functors $F, G\colon$ **AbGroup** $\to$ **AbGroup** by

$$F(A) = \{a \in A \;:\; 4a = 0\} \qquad\qquad F(f) = f|_{F(A)}$$
$$G(A) = \{a \in A \;:\; 2a = 0\} \qquad\qquad G(f) = f|_{G(A)}.$$

Then define $\alpha_A\colon F(A) \to G(A)$ by $\alpha_A(a) = 2a$. We claim that this is a natural transformation. Indeed, the naturality condition just says that for any homomorphism $f\colon A \to B$ and any element $a \in F(A)$ we have $f(2a) = 2f(a)$, which is immediate from the definition of a homomorphism.

Example 36.51. [`eg-nat-trans-ii`]

Let $U\colon$ **Rings** $\to$ **Sets** be the forgetful functor. For any ring R, we can define a function $\alpha_R\colon R \to R$ by $\alpha_R(x) = x - x^2$. This is not a homomorphism, so it is better to think of it as a morphism $\alpha_R\colon U(R) \to U(R)$ in the category of sets. We claim that this defines a natural transformation $\alpha\colon U \to U$. Indeed, the naturality condition says that for all ring homomorphisms $f\colon R \to S$, the diagram

$$\begin{tikzcd} U(R) \arrow[r, "\alpha_R"] \arrow[d, "U(f)"'] & U(R) \arrow[d, "U(f)"] \\ U(S) \arrow[r, "\alpha_S"'] & U(S). \end{tikzcd}$$

In other words, for all $x \in R$ we should have $f(x - x^2) = f(x) - f(x)^2$. This is an easy consequence of the fact that f is a ring homomorphism.

Example 36.52. [`eg-frobenius-natural`]

Let p be a prime, and let $1\colon \mathbf{Fields}_p \to \mathbf{Fields}_p$ be the identity functor. As in Proposition 36.21, we define a homomorphism $\phi_K\colon K \to K$ by $\phi_K(a) = a^p$. We claim that this gives a natural map $\phi\colon 1 \to 1$. Equivalently, we claim that for any homomorphism $f\colon K \to L$, the following diagram commutes:

$$\begin{tikzcd} K \arrow[r, "\phi_K"] \arrow[d, "f"'] & K \arrow[d, "f"] \\ L \arrow[r, "\phi_L"'] & L. \end{tikzcd}$$

This just says that $f(a^p) = f(a)^p$, which is clear.

Definition 36.53. Suppose we have functors $F\colon \mathcal{C} \to \mathcal{D}$ and $G\colon \mathcal{D} \to \mathcal{C}$. Let $1_{\mathcal{C}}$ denote the identity functor from $\mathcal{C}$ to $\mathcal{C}$, and similarly for $1_{\mathcal{D}}$. We say that G is *inverse* to F (and *vice versa*) if there exist natural isomorphisms $1_{\mathcal{C}} \to GF$ and $1_{\mathcal{D}} \to FG$.

Proposition 36.54. [`prop-cat-equiv`]

If $F\colon \mathcal{C} \to \mathcal{D}$ has an inverse, then it is an equivalence. Conversely, if $F\colon \mathcal{C} \to \mathcal{D}$ is an equivalence, then we can construct an inverse provided that we ignore set-theoretic difficulties.

This will rely on a simple lemma:

Lemma 36.55. [`lem-PQRP`]

Let $P \xrightarrow{f} Q \xrightarrow{g} R \xrightarrow{h} P$ be functions between sets, such that f and g are injective and h is bijective and $hgf = 1_P$. Then f and g are also bijective.

Proof. First, as h is bijective we have an inverse map $h^{-1}\colon P \to R$. As $hgf = 1_P$ we have $h^{-1}hgf = h^{-1}1_P$, or in other words $gf = h^{-1}$. It then follows that $gfh = h^{-1}h = 1_R$. Now for any $r \in R$ we can put $q = fh(r)$ and we find that $r = g(q)$. This shows that g is surjective, but it was also assumed to be injective, so it is a bijection. Now the equation $gf = h^{-1}$ gives $f = g^{-1}h^{-1}$, with g and h bijections, so f is a bijection. □

Proof of Proposition 36.54. First suppose that F has an inverse. We choose an inverse $G\colon \mathcal{D} \to \mathcal{C}$, and natural isomorphisms $\alpha\colon 1_{\mathcal{C}} \to GF$ and $\beta\colon 1_{\mathcal{D}} \to FG$. For $X, Y \in \mathcal{C}$ we define

$$\phi_{XY}\colon \mathcal{C}(GF(X), GF(Y)) \to \mathcal{C}(X, Y)$$

by

$$\phi_{XY}(GF(X) \xrightarrow{w} GF(Y)) = (X \xrightarrow{\alpha_X} GF(X) \xrightarrow{w} GF(Y) \xrightarrow{\alpha_Y^{-1}} Y).$$

This is clearly a bijection, with $\phi_{XY}^{-1}(u) = \alpha_Y \circ u \circ \alpha_X^{-1}$. On the other hand, by inspecting the naturality square

$$\begin{tikzcd} X \arrow[r, "\alpha_X", "\simeq"'] \arrow[d, "u"'] & GF(X) \arrow[d, "GF(u)"] \\ Y \arrow[r, "\simeq", "\alpha_Y"'] & GF(Y) \end{tikzcd}$$

we see that the composite

$$\mathcal{C}(X,Y) \xrightarrow{F_{XY}} \mathcal{D}(F(X),F(Y)) \xrightarrow{G_{F(X),F(Y)}} \mathcal{C}(GF(X),GF(Y)) \xrightarrow{\phi_{XY}} \mathcal{C}(X,Y)$$

is the identity. It follows easily that F_{XY} is injective. A similar argument shows that $G_{UV}\colon \mathcal{D}(U,V) \to \mathcal{C}(G(U),G(V))$ is injective for all U and V. In particular, we can take $U = F(X)$ and $V = F(Y)$ to see that $G_{F(X),F(Y)}$ is injective. We can now apply Lemma 36.55 to the above composite to see that F_{XY} is a bijection, which means that F is full and faithful. Moreover, for any $Y \in \mathcal{D}$ we have $Y \simeq FG(Y)$ via β_Y, which shows that F is essentially surjective.

Now suppose we start instead from the assumption that F is full, faithful, and essentially surjective. As F is essentially surjective, for each $U \in \mathcal{D}$ we can choose an object $GU \in \mathcal{C}$ and an isomorphism $\beta_U\colon U \to FGU$. Often we will know that F is essentially surjective by some kind of constructive argument that provides a choice of GU and β_U. In other cases we may need to make an arbitrary choice, simultaneously for all objects $U \in \mathcal{D}$, and there may be a proper class of such objects. This is set-theoretically dubious, although there are various ways to make it respectable. This will not be a crucial point for us, so we will ignore the details. So far we have only defined G on objects. Suppose we have a morphism $s\colon U \to V$ in $\mathcal{D}$. Let $\psi_{UV}(s)$ be the composite

$$FG(U) \xrightarrow{\beta_U^{-1}} U \xrightarrow{s} V \xrightarrow{\beta_V} FG(V).$$

As F is assumed full and faithful, we see that the map $F_{G(U),G(V)}\colon \mathcal{C}(G(U),G(V)) \to \mathcal{C}(FG(U),FG(V))$ is bijective. There is thus a unique morphism $s'\colon G(U) \to G(V)$ such that $F(s') = \psi_{UV}(s)$. We define $G(s)$ to be this morphism s'. In other words, $G(s)$ is characterised uniquely by the fact that the square

$$\begin{tikzcd} U \arrow[r, "\beta_U", "\simeq"'] \arrow[d, "s"'] & FG(U) \arrow[d, "FG(s)"] \\ V \arrow[r, "\simeq", "\beta_V"'] & FG(V) \end{tikzcd}$$

is commutative. Now suppose we have another morphism $t\colon V \to W$. It is easy to see that $G(t) \circ G(s)$ has the defining property of $G(ts)$, so $G(ts) = G(t)G(s)$. It is also clear from the definition that $G(1_U) = 1_{G(U)}$, so we have defined a functor $G\colon \mathcal{D} \to \mathcal{C}$. We can now regard the above square as a naturality square, showing that β is a natural isomorphism $1_{\mathcal{D}} \to FG$. Next, consider the isomorphism $\beta_{FX}\colon FX \to FGFX$. As F is full and faithful, there is a unique morphism $\alpha_X\colon X \to GFX$ such that $F(\alpha_X) = \beta_{FX}$. Similarly, there is also a unique $\alpha'_X\colon GFX \to X$ with $F(\alpha'_X) = \beta_{FX}^{-1}$. This means that

$$F(\alpha'_X\alpha_X) = F(\alpha'_X)F(\alpha_X) = \beta_{FX}^{-1}\beta_{FX} = 1_{FX} = F(1_X).$$

As F is faithful, we can deduce that $\alpha'_X\alpha_X = 1_X$. A similar argument shows that $\alpha_X\alpha'_X = 1_{GF(X)}$, so α'_X is inverse to α_X, proving that α_X is an isomorphism. We claim that α is also natural, or in other words that the left hand square below commutes:

$$\begin{tikzcd} X \arrow[r, "\alpha_X"] \arrow[d, "u"'] & GF(X) \arrow[d, "GF(u)"] \\ Y \arrow[r, "\alpha_Y"'] & GF(Y) \end{tikzcd} \qquad \begin{tikzcd} F(X) \arrow[r, "F(\alpha_X)"] \arrow[d, "F(u)"'] & FGF(X) \arrow[d, "FGF(u)"] \\ F(Y) \arrow[r, "F(\alpha_Y)"'] & FGF(Y) \end{tikzcd} \qquad \begin{tikzcd} F(X) \arrow[r, "\beta_{FX}"] \arrow[d, "F(u)"'] & FGF(X) \arrow[d, "FGF(u)"] \\ F(Y) \arrow[r, "\beta_{FY}"'] & FGF(Y). \end{tikzcd}$$

This says that two maps $X \to GF(Y)$ are the same. As F is faithful, it will suffice to check that the resulting maps $F(X) \to FGF(Y)$ are the same, or equivalently that the middle square above commutes. By the definition of α, the middle square is the same as the right hand square, which is commutative because β is natural. We have thus constructed a natural isomorphism $\alpha\colon 1_{\mathcal{C}} \to GF$ as well as a natural isomorphism $\beta\colon 1_{\mathcal{D}} \to FG$, so G is inverse to F as claimed. □

Example 36.56. [`eg-choose-basis`]
As in Example 36.43 we define $F\colon \mathbf{Mat}_{\mathbb{R}} \to \mathbf{FinVect}_{\mathbb{R}}$ by $F(n) = \mathbb{R}^n$. We observed in Example 36.47(c) that this is full, faithful and essentially surjective. In principle this must therefore have an inverse. However, to construct one, we would need to choose a basis for every finite-dimensional real vector space simultaneously.

36.5. Adjoint functors.

Definition 36.57. [`defn-adjoint`]
Let $\mathcal{C}$ and $\mathcal{D}$ be categories. An *adjoint pair* consists of functors $F\colon \mathcal{C} \to \mathcal{D}$ (the *left adjoint*) and $G\colon \mathcal{D} \to \mathcal{C}$ (the *right adjoint*), together with natural transformations $\eta\colon 1_{\mathcal{C}} \to FG$ (the *unit*) and $\epsilon\colon GF \to 1_{\mathcal{D}}$ (the *counit*) such that the following diagrams commute for all $X \in \mathcal{C}$ and $U \in \mathcal{D}$:

$$\begin{tikzcd} G(X) \arrow[r, "G(\eta_X)"] \arrow[dr, "1_{G(X)}"'] & GFG(X) \arrow[d, "\epsilon_{G(X)}"] \\ & G(X) \end{tikzcd} \qquad \begin{tikzcd} F(X) \arrow[r, "\eta_{F(X)}"] \arrow[dr, "1_{F(X)}"'] & FGF(X) \arrow[d, "F(\epsilon_X)"] \\ & F(X). \end{tikzcd}$$

These are called the *triangular identities*.

Example 36.58. [`eg-abelianisation`]
In this example we will write the group law for a generic group using multiplicative notation, whether or not it is abelian.

Let $J\colon \mathbf{AbGroups} \to \mathbf{Groups}$ be the inclusion functor. For any group G, let G' be the subgroup generated by all elements of the form $xyx^{-1}y^{-1}$. It is standard (and not hard to check) that G' is a normal subgroup, that G/G' is an abelian group, and that G' is the smallest subgroup with these two properties. If $\alpha\colon G \to H$ is a homomorphism then $\alpha(xyx^{-1}y^{-1}) = \alpha(x)\alpha(y)\alpha(x)^{-1}\alpha(y)^{-1}$, so $\alpha(G') \leq H'$, so there is an induced homomorphism $\overline{\alpha}\colon G/G' \to H/H'$ given by $\overline{\alpha}(xG') = \alpha(x)H'$. We can thus define a functor $Q\colon \mathbf{Groups} \to \mathbf{AbGroups}$ by $Q(G) = G/G'$ and $Q(\alpha) = \overline{\alpha}$. Next, we can define $\eta_G\colon G \to JQ(G) = G/G'$ by $\eta(g) = gG'$. This is easily seen to be natural. Next, let A be an abelian group. Then $QJ(A) = A/A' = A/\{1\}$. Strictly speaking, this is not the same as A; the elements of $A/\{1\}$ are the cosets $\{a\}$ for $a \in A$, not the elements a themselves. We can define an isomorphism $\epsilon_A\colon QJ(A) \to A$ by $\epsilon_A(\{a\}) = a$, and this gives a natural isomorphism $\epsilon\colon QJ \to 1$. We claim that this gives an adjoint pair, with Q as the left adjoint and J as the right adjoint. Indeed, the triangular diagrams take the form

$$\begin{tikzcd} A \arrow[r, "\eta"] \arrow[dr, "1"'] & A/1 \arrow[d, "\epsilon"] \\ & A \end{tikzcd} \qquad \begin{tikzcd} G/G' \arrow[r, "\eta"] \arrow[dr, "1"'] & (G/G')/1 \arrow[d, "\epsilon"] \\ & G/G' \end{tikzcd}$$

and they are easily seen to commute.

Example 36.59. [`eg-topological-adjunctions`]
Consider the category **Spaces** of topological spaces, and the category **Sets** of sets. We have already seen the forgetful functor $U\colon \mathbf{Spaces} \to \mathbf{Sets}$. In the opposite direction, for any set T we let $D(T)$ denote T with the discrete topology, and we let $I(T)$ denote T with the indiscrete topology. For any function $f\colon T \to S$, we see that f is continuous when regarded as a map $D(T) \to D(S)$, or as a map $I(T) \to I(S)$. We can therefore make D and I into functors $\mathbf{Sets} \to \mathbf{Spaces}$ by putting $D(f) = f$ and $I(f) = f$. Next, for any set T we let $\eta_T\colon T \to UD(T)$ be the identity map. For any space X, we note that $DU(X)$ is the same set as X, but with the discrete topology. We let ϵ_X denote the identity function, considered as a map from $DU(X)$ to X. This is continuous, because every subset of $DU(X)$ is open, so there is nothing to check. We claim that this gives an adjunction, with D as the left adjoint and U as the right adjoint. Indeed, all that is left is

to check the triangular identities, which are trivial, because all maps involved are the identity maps of the underlying sets.

Next, for any space X we define ζ_X to be the identity map, considered as a map $X \to IU(X)$. This is continuous, because the only open sets in $IU(X)$ are $\emptyset$ and X, which are open in the original topology as well. We also let ξ_T denote the identity map $T \to T = UI(T)$. We find that this gives another adjunction, with U now being the left adjoint, I the right adjoint, ζ the unit and ξ the counit.

Proposition 36.60. [`prop-adjoint`]
Any adjunction (F, G, η, ϵ) gives a system of bijections

$$\rho_{XU} \colon \mathcal{C}(X, G(U)) \to \mathcal{D}(F(X), U)$$

given by

$$\rho_{XU}(X \xrightarrow{q} G(U)) = (F(X) \xrightarrow{F(q)} FG(U) \xrightarrow{\epsilon_U} U$$
$$\rho_{XU}^{-1}(F(X) \xrightarrow{p} U) = (X \xrightarrow{\eta_X} GF(X) \xrightarrow{G(p)} G(U)).$$

These are natural in the sense that for all

$$W \xrightarrow{f} X \qquad X \xrightarrow{p} GU \qquad U \xrightarrow{m} V$$

we have

$$\rho_{WV}(W \xrightarrow{f} X \xrightarrow{p} G(U) \xrightarrow{G(m)} G(V)) = (F(W) \xrightarrow{F(f)} F(X) \xrightarrow{\rho(p)} U \xrightarrow{m} V).$$

Conversely, if we have functors F and G and maps ρ_{XU} with this naturality property, then there is a unique adjunction (F, G, η, ϵ) giving rise to them.

Proof. We can certainly define maps

$$\begin{tikzcd} \mathcal{C}(X, G(U)) \arrow[r, shift left, "\rho_{XU}"] & \mathcal{D}(F(X), U) \arrow[l, shift left, "\lambda_{XU}"] \end{tikzcd}$$

by

$$\rho_{XU}(q) = \epsilon_U \circ F(q) \qquad \lambda_{XU}(p) = G(p) \circ \eta_X.$$

Now consider the left hand diagram below:

$$\begin{tikzcd} X \arrow[r, "q"] \arrow[d, "\eta_X"'] & G(U) \arrow[d, "\eta_{G(U)}"'] \arrow[dr, "1"] & \\ GF(X) \arrow[r, "GF(q)"'] & GFG(U) \arrow[r, "G(\epsilon_U)"'] & G(U) \end{tikzcd} \qquad \begin{tikzcd} F(X) \arrow[r, "F(\eta_X)"] \arrow[dr, "1"'] & FGF(X) \arrow[r, "FG(p)"] \arrow[d, "\epsilon_{F(X)}"] & FG(U) \arrow[d, "\epsilon_U"] \\ & F(X) \arrow[r, "p"'] & U \end{tikzcd}$$

The square commutes because η is natural, and the triangle commutes by the triangular identities. The composite along the bottom is $G(\epsilon_U \circ F(q)) = G(\rho(q))$, so the left edge followed by the bottom is $G(\rho(q)) \circ \eta_X = \lambda(\rho(q))$. As the diagram commutes, we have $\lambda(\rho(q)) = q$. Similarly, we can use the right hand diagram to show that $\rho(\lambda(p)) = p$ for all $p \colon F(X) \to U$, so ρ and λ are mutually inverse bijections, as claimed.

Now suppose we have maps

$$W \xrightarrow{f} X \qquad X \xrightarrow{p} GU \qquad U \xrightarrow{m} V$$

and consider the diagram

$$\begin{tikzcd} F(W) \arrow[r, "F(f)"] & F(X) \arrow[r, "F(p)"] \arrow[dr, "\rho(p)"'] & FG(U) \arrow[r, "FG(m)"] \arrow[d, "\epsilon_U"] & FG(V) \arrow[d, "\epsilon_V"] \\ & & U \arrow[r, "m"'] & V. \end{tikzcd}$$

The triangle commutes by the definition of $\rho(p)$, and the square commutes by the naturality of ϵ. The upper composite from $F(X)$ to V is by definition $\rho(G(m) \circ p \circ f)$, whereas the lower composite it $m \circ \rho(p) \circ F(f)$, and the diagram commutes so these are the same.

Now suppose instead that we start with a system of bijections

$$\rho_{XU}\colon \mathcal{C}(X,G(U)) \to \mathcal{D}(F(X),U)$$

satisfying

$$\rho_{WV}(W \xrightarrow{f} X \xrightarrow{p} G(U) \xrightarrow{G(m)} G(V)) = (F(W) \xrightarrow{F(f)} F(X) \xrightarrow{\rho_{XU}(p)} U \xrightarrow{m} V)$$

for all f, p and m as before. We put

$$\lambda_{XU} = \rho_{XU}^{-1}\colon \mathcal{D}(F(X),U) \to \mathcal{C}(X,G(U)).$$

Note that the naturality rule for ρ gives

$$\rho_{WV}(W \xrightarrow{f} X \xrightarrow{\lambda_{XU}(q)} G(U) \xrightarrow{G(m)} G(V)) = (F(W) \xrightarrow{F(f)} F(X) \xrightarrow{q} U \xrightarrow{m} V)$$

and we can apply λ_{WV} to this to get

$$\lambda_{WV}(F(W) \xrightarrow{F(f)} F(X) \xrightarrow{q} U \xrightarrow{m} V) = W \xrightarrow{f} X \xrightarrow{\lambda_{XU}(q)} G(U) \xrightarrow{G(m)} G(V).$$

Now note that we have an element $1_{G(U)} \in \mathcal{C}(G(U),G(U))$ and a bijection

$$\rho_{G(U),U}\colon \mathcal{C}(G(U),G(U)) \to \mathcal{D}(FG(U),U),$$

giving a morphism

$$\epsilon_U = \rho_{G(U),U}(1_{G(U)})\colon FG(U) \to U.$$

Now suppose we have a morphism $m\colon U \to V$ in $\mathcal{D}$. The naturality rule for the triple $(f,p,m) = (1_{GU},1_{GU},m)$ gives $\rho(G(m)) = m \circ \epsilon_U$, whereas the naturality rule for $(Gm,1_{GV},1_V)$ gives $\rho(G(m)) = \epsilon_V \circ FG(m)$. It follows that $m \circ \epsilon_U = \epsilon_V \circ FG(m)$, which means precisely that the maps ϵ_U give a natural transformation $FG \to 1_{\mathcal{D}}$. The naturality rule for $(q,1_{GU},1_U)$ gives $\rho(q) = \epsilon_U \circ F(q)$, so ϵ determines ρ just as before. By a dual argument, we see that the maps

$$\eta_X = \lambda_{X,F(X)}(1_{F(X)})\colon X \to GF(X)$$

give a natural transformation, and that $\lambda(p) = G(p) \circ \eta_X$ for all $p\colon F(X) \to U$. In particular, we can take $X = G(U)$ and $p = \epsilon_U$ to get $\lambda(\epsilon_U) = G(\epsilon_U) \circ \eta_{G(U)}$. On the other hand, we also have $\epsilon_U = \rho(1_{G(U)})$, so $\lambda(\epsilon_U) = 1_{G(U)}$. This gives $G(\epsilon_U) \circ \eta_{G(U)} = 1_{G(U)}$, which is one of the triangular identities. A similar argument gives the other triangular identity, so we have an adjunction as claimed. □

EXAMPLE 36.61. [`eg-group-ring-adjunction`]
Let **Rings** be the category of (not necessarily commutative) rings. For any ring R, we let $U(R)$ denote the set of invertible elements in R, which is a group under multiplication. If $f\colon R \to S$ is a ring homomorphism and $u \in U(R)$ then $f(u^{-1})$ is an inverse for $f(u)$, so $f(u) \in U(S)$. Thus, f restricts to give a group homomorphism $U(R) \to U(S)$, so we can regard U as a functor **Rings** $\to$ **Groups**. We will show that this has a left adjoint. The construction is essentially the same as in Example 36.39: for any group G we form a free abelian group $\mathbb{Z}[G]$ with one basis element $[g]$ for each $g \in G$, and we make this a ring by the rule $[g] * [h] = [gh]$. Given a group homomorphism $p\colon G \to U(R)$ we define a ring homomorphism $\rho_{GR}(p)\colon \mathbb{Z}[G] \to R$ by

$$\rho_{GR}(p)(\sum_i n_i[g_i]) = \sum_i n_i p(g_i).$$

This gives a bijection

$$\rho_{GR}\colon \mathbf{Groups}(G,U(R)) \to \mathbf{Rings}(\mathbb{Z}[G],R)$$

which has the naturality property described in Proposition 36.60 and so leads to an adjunction. The unit map $\eta\colon G \to U(\mathbb{Z}[G])$ is just $\eta(g) = [g]$, and the counit map $\epsilon\colon \mathbb{Z}[U(R)] \to R$ is $\epsilon(\sum_i n_i[u_i]) = \sum_i n_i u_i$.

36.6. Products and coproducts.

Definition 36.62. [`defn-terminal`]
Let $\mathcal{C}$ be a category. An object $A \in \mathcal{C}$ is *initial* if for each $X \in \mathcal{C}$, there is a unique morphism $A \to X$. An object $B \in \mathcal{C}$ is *terminal* if for each $X \in \mathcal{C}$, there is a unique morphism $X \to B$.

Proposition 36.63. [`prop-terminal-unique`]
If A and A' are both initial objects, then there is a unique isomorphism from A to A'. Similarly, if B and B' are both terminal objects, then there is a unique isomorphism from B to B'.

Proof. First note that there is a unique map from A to A, and the identity is one such map, so it must be the only one. Similarly, the identity is the only morphism from A' to A'. Next, as A is initial there is a unique morphism $f\colon A \to A'$, and as A' is initial there is a unique morphism $f'\colon A' \to A$. Now $f'f$ is a morphism from A to itself so we must have $f'f = 1_A$, and similarly $ff' = 1_{A'}$. Thus, f and f' are isomorphisms. The terminal case is proved dually. $\square$

Because of this proposition, it is usually harmless to talk about "the initial object" or "the terminal object", even though it may not strictly be unique.

Example 36.64.

(a) In **Sets**, the empty set is initial, and any set with precisely one element is terminal.
(b) The empty set has a unique topology, and with that topology it becomes an initial object in **Spaces**. Similarly, any singleton set has a unique topology, and with that topology it is terminal in **Spaces**.
(c) In **Groups**, the trivial group is both initial and terminal.
(d) In **Rings**, the ring $\mathbb{Z}$ is initial and the zero ring is terminal.
(e) In the category of fields of characteristic $p > 0$, the field $\mathbb{Z}/p$ is initial and there is no terminal object. Similarly, in the category of fields of characteristic zero, $\mathbb{Q}$ is initial and there is no terminal object. In the category of all fields there is neither an initial object nor a terminal object.
(f) If G is a nontrivial group then bG has neither an initial object nor a terminal object.

Definition 36.65. [`defn-categorical-product`]
Let $\mathcal{C}$ be a category, and let A and B be objects of $\mathcal{C}$. A *product diagram* for A and B is a diagram $A \xleftarrow{p} P \xrightarrow{q} B$ such that for any other diagram $A \xleftarrow{f} T \xrightarrow{g} B$ there is a unique morphism $m\colon T \to P$ with $pm = f$ and $qm = g$. Equivalently, for any object T there is a map

$$\lambda_T\colon \mathcal{C}(T,P) \to \mathcal{C}(T,A) \times \mathcal{C}(T,B)$$

given by $\lambda_T(m) = (pm, qm)$, and the diagram $A \xleftarrow{p} P \xrightarrow{q} B$ is a product diagram iff the maps λ_T is a bijection for all T.

Example 36.66. [`eg-product-set`]
In the category **Sets**, we can take $P = A \times B$ and $p(a,b) = a$ and $q(a,b) = b$. Given another diagram $A \xleftarrow{f} T \xrightarrow{g} B$ we can define $m\colon T \to A \times B$ by $m(t) = (f(t), g(t))$, and this is clearly the unique function with $pm = f$ and $qm = g$.

Example 36.67. [`eg-product-space`]
Let X and Y be topological spaces, and give the set $X \times Y$ the product topology as in Definition 5.14. As proved in Proposition 5.16, this makes the diagram $X \xleftarrow{p} X \times Y \xrightarrow{q}$ into a product diagram in the category **Spaces**.

Example 36.68. [`eg-product-group`]
Now consider instead the category **Groups**. We can again take $P = A \times B$, equipped with the usual group structure so that $(a,b).(a',b') = (aa', bb')$. Then the projections p and q are group homomorphisms, so we have a diagram $A \xleftarrow{p} P \xrightarrow{q} B$ in **Groups**. If we have another diagram $A \xleftarrow{f} T \xrightarrow{g} B$ in **Groups** we can again define $m\colon T \to A \times B$ by $m(t) = (f(t), g(t))$, and we find that this is also a group homomorphism, and it is clearly the unique one with $pm = f$ and $qm = g$. The same procedure works in the category of rings.

Example 36.69. [`eg-poset-product`]

Let X be a set. Let P the set of subsets of X, ordered by inclusion, and let sP be the corresponding category. Thus, for any pair of subsets $U, V \subseteq X$, there is a unique morphism $U \to V$ if $U \subseteq V$, and no morphisms otherwise. For any pair of subsets A and B, we have a diagram $A \xleftarrow{p} A \cap B \xrightarrow{q} B$, and we claim that this is a product diagram. Indeed, if we have a diagram $A \xleftarrow{f} T \xrightarrow{g} B$ then T must be a subset of A and also a subset of B, so $T \subseteq A \cap B$, so we have a morphism $m \colon T \to A \cap B$. Because parallel maps in sP are always equal, we see that m is unique and satisfies $pm = f$ and $qm = g$ as required.

Example 36.70. [`eg-matrix-product`]

Let d and e be natural numbers, and consider them as objects of the category $\mathbf{Mat}_R$. Put

$$\begin{aligned} p &= [I_d \mid 0_{de}] \colon d + e \to d \\ q &= [0_{ed} \mid I_e] \colon d + e \to e. \end{aligned}$$

Suppose we are given morphisms $d \xleftarrow{f} t \xrightarrow{g} e$, so f is a $d \times t$ matrix and g is an $e \times t$ matrix. We can stack them to form a $(d + e) \times t$ matrix $m = \left[\begin{array}{c} f \\ \hline g \end{array}\right]$, and we find that this is the unique matrix such that $pm = f$ and $qm = g$. Thus, the diagram $d \xleftarrow{p} (d + e) \xrightarrow{q} e$ is a product diagram in $\mathbf{Mat}_R$.

Example 36.71. [`eg-field-product`]

The theory of products in $\mathbf{Fields}_0$ is complex. Given fields K and L of characteristic zero, we can always form the product ring $K \times L$, but that will not be a field, so it will not give a product diagram in $\mathbf{Fields}_0$. In some cases, there will be a be a different diagram that does provide a product in $\mathbf{Fields}_0$. For one class of examples, recall that K has a unique subfield K_0 isomorphic to $\mathbb{Q}$, and similarly for L, so $K_0 \simeq L_0$. Suppose that there are no other cases where a subfield of K is isomorphic to a subfield of L. (When K and L are finite extensions of $\mathbb{Q}$, there are only finitely many subfields and this condition can often be checked explicitly by Galois theory.) If we have a diagram $K \xleftarrow{f} T \xrightarrow{g} L$ then $f(T) \simeq T \simeq g(T)$ so $f(T) = K_0$ and $g(T) = L_0$ and $T \simeq \mathbb{Q}$. Using this we see that the diagram $K \leftarrow \mathbb{Q} \to L$ is a product diagram. For example, this applies when $K = \mathbb{Q}(\sqrt{2})$ and $L = \mathbb{Q}(\sqrt{3})$. Using similar methods one can show that the diagram $\mathbb{Q}(\sqrt[3]{2}) \xleftarrow{1} \mathbb{Q}(\sqrt[3]{2}) \xrightarrow{1} \mathbb{Q}(\sqrt[3]{2})$ is also a product diagram, but that there is no product diagram of the form $\mathbb{Q}(\sqrt{2}) \leftarrow P \to \mathbb{Q}(\sqrt{2})$. (These two cases are different because $\mathrm{Aut}(\mathbb{Q}(\sqrt[3]{2})) = \{1\}$, but $\mathbb{Q}(\sqrt{2})$ has a nontrivial automorphism sending $\sqrt{2}$ to $-\sqrt{2}$.)

Remark 36.72. [`rem-products-terminal`]

Suppose we have a category $\mathcal{C}$ and objects $A, B \in \mathcal{C}$. We can define an auxiliary category $\mathcal{D}$ as follows: the objects are diagrams of type $(A \xleftarrow{f} T \xrightarrow{g} B)$ with $T \in \mathcal{C}$, and the morphisms from $(A \xleftarrow{f} T \xrightarrow{g} B)$ to $(A \xleftarrow{f'} T' \xrightarrow{g'} B)$ are the morphisms $m \colon T \to T'$ in $\mathcal{C}$ for which $f'm = f$ and $g'm = g$. Equivalently, they are the morphisms m that make the following diagram commute:

$$\begin{tikzcd} & T \arrow[dl, "f"'] \arrow[dr, "g"] \arrow[dd, "m"] & \\ A & & B \\ & T' \arrow[ul, "f'"] \arrow[ur, "g'"'] & \end{tikzcd}$$

We now see from the definitions that a product diagram for A and B is precisely the same as a terminal object in the category $\mathcal{D}$. Proposition 36.63 therefore tells us that any two product diagrams are connected by a unique morphism, and that that morphism is an isomorphism in $\mathcal{D}$. More explicitly, if $(A \xleftarrow{p} P \xrightarrow{q} B)$ and $(A \xleftarrow{p'} P' \xrightarrow{q'} B)$ are both product diagrams, then there is a unique morphism $m \colon P \to P'$ with $p'm = p$ and $q'm = q$, and that morphism is an isomorphism. Because of this, it is generally harmless to refer to P as "the product of A and B" and to denote it by $A \times B$, even though it may not strictly be unique.

We now consider the dual notion.

Definition 36.73. [defn-categorical-coproduct]

Let $\mathcal{C}$ be a category, and let A and B be objects of $\mathcal{C}$. A *coproduct diagram* for A and B is a diagram $A \xrightarrow{i} P \xleftarrow{j} B$ such that for any other diagram $A \xrightarrow{f} T \xleftarrow{g} B$ there is a unique morphism $m\colon P \to T$ with $mi = f$ and $mj = g$. Equivalently, for any object T there is a map

$$\mu_T\colon \mathcal{C}(P,T) \to \mathcal{C}(A,T) \times \mathcal{C}(B,T)$$

given by $\mu_T(m) = (mi, mj)$, and the diagram $A \xrightarrow{i} P \xleftarrow{j} B$ is a coproduct diagram iff the map μ_T is a bijection for all T.

Remark 36.74. [rem-coproducts-initial]

By the dual of Remark 36.72, coproducts are initial objects in a certain auxiliary category, so they are unique up to canonical isomorphism if they exist. It is therefore usually harmless to refer to P as "the coproduct of A and B" and to denote it by $A \amalg B$.

Example 36.75. [eg-coproduct-set]

Let A and B be sets. We then let $A \amalg B$ denote the set of pairs of the form $(0,a)$ (with $a \in A$) or $(1,b)$ (with $b \in B$), and define maps $A \xrightarrow{i} A \amalg B \xleftarrow{j} B$ by $i(a) = (0,a)$ and $j(b) = (1,b)$. If we are given another diagram $A \xrightarrow{f} T \xleftarrow{g} B$, we can define $m\colon A \amalg B \to T$ by $m(0,a) = f(a)$ and $m(1,b) = g(b)$; this is clearly the unique map with $mi = f$ and $mj = g$, so we have a coproduct diagram.

Alternatively, if A and B happen to come as disjoint subsets of some other set C, we can just define $P' = A \cup B$ and $i'(a) = a$ and $j'(b) = b$. This gives another coproduct diagram. As expected, there is a bijection $m\colon P \to P'$ given by $m(0,a) = a$ and $m(1,b) = b$.

Example 36.76. [eg-biproduct]

Let A and B be abelian groups; we will write the group operation as addition. In this context, it is traditional to write $A \oplus B$ instead of $A \times B$ for the product group. As well as the projection maps $A \xleftarrow{p} A \oplus B \xrightarrow{q} B$, we also have maps $A \xrightarrow{i} A \oplus B \xleftarrow{j} B$ given by $i(a) = (a,0)$ and $j(b) = (0,b)$. We claim that this gives a coproduct diagram in **AbGroups**. Indeed, suppose we have another abelian group T and homomorphisms $A \xrightarrow{f} T \xleftarrow{g} B$. We can then define $m\colon A \oplus B \to T$ by $m(a,b) = f(a) + g(b)$. Using the fact that $(a,b) = i(a) + j(b)$, we see that m is the unique homomorphism with $mi = f$ and $mj = g$, as required.

Example 36.77. [eg-free-product]

Let G and H be groups (now written with multiplicative notation) that need not be abelian. There is then a group $G*H$ (called the *free product* of G and H) and morphisms $G \xrightarrow{i} G*H \xleftarrow{j} H$ giving a coproduct diagram. To describe $G*H$, put $G' = G \setminus \{1\}$ and $H' = H \setminus \{1\}$. Then $G*H$ is the set of sequences $u = (u_1, \dots, u_r)$ such that either

(a) $u_i \in G'$ for all odd i, and $u_i \in H'$ for all even i; or
(b) $u_i \in H'$ for all odd i, and $u_i \in G'$ for all even i.

The empty sequence is permitted, and gives the identity element of $G*H$. The inverse of $(u_1, \dots, u_r)$ is $(u_r^{-1}, \dots, u_1^{-1})$. To define uv, we first join them together to get a longer sequence w that may or may not lie in $G*H$. If w contains adjacent terms that lie in the same group then we combine them by multiplying in that group. If the resulting sequence contains an identity element anywhere, then we discard it. After a finite number of steps of these two types, we obtain a element of $G*H$, which we take to be uv. Some work is required to prove that this is well-defined and associative, but we will not give details here. For $g \in G'$ we define $i(g)$ to be the sequence (g), and we also define $i(1)$ to be the empty sequence. This gives a homomorphism $i\colon G \to G*H$, and we define $j\colon H \to G*H$ in the same way. Given homomorphisms $G \xrightarrow{p} T \xleftarrow{q} H$ we define $m\colon G*H \to T$ by $m(u_1, \dots, u_r) = p(u_1)q(u_2)p(u_3)\cdots$ (in case (a)) or $m(u_1, \dots, u_r) = q(u_1)p(u_2)q(u_3)\cdots$ (in case (b)). One can check that this is a homomorphism. Given that, it is clearly the unique one for which $mi = p$ and $mj = q$. We thus have a coproduct diagram as claimed.

In some special cases we can be more explicit. For one example, define maps $r_k, s_k\colon \mathbb{Z} \to \mathbb{Z}$ by $r_k(n) = k - n$ and $s_k(n) = k + n$. The set

$$A = \{r_k \;:\; k \in \mathbb{Z}\} \cup \{s_k \;:\; k \in \mathbb{Z}\} = \{n \mapsto \epsilon n + k \;:\; \epsilon \in \{1,-1\}, k \in \mathbb{Z}\}$$

is then a group under composition (with $s_0 = 1$), and the sets $G = \{1, r_0\}$ and $H = \{1, r_1\}$ are subgroups. We have $r_0^2 = r_1^2 = 1$ and $s_k = (r_1 r_0)^k$ and $r_k = (r_1 r_0)^{k-1} r_1$, and using these we find that the evident homomorphism $G * H \to A$ is an isomorphism, or equivalently that the inclusion maps give a coproduct diagram $G \to A \leftarrow H$. Note here that G and H are abelian but $G * H$ is not, so the coproduct of G and H in **Groups** is different from the coproduct in **AbGroups**.

For another example, consider the group $SL_2(\mathbb{Z})$ of 2×2 integer matrices with determinant one, and the quotient group $P = SL_2(\mathbb{Z})/\{I, -I\}$. Put $g = \left[\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right]$ and $h = \left[\begin{smallmatrix} 0 & -1 \\ 1 & -1 \end{smallmatrix}\right]$ (considered as elements of P). These generate cyclic groups $G = \langle g \rangle$ and $H = \langle h \rangle$ of order 2 and 3 respectively. It is known that the diagram $G \to P \leftarrow H$ is a coproduct, but we will not discuss the proof here.

Example 36.78. [`eg-coproduct-ring`]
Readers who are familiar with tensor products can consider the following. Let **CRings** denote the category of commutative rings. If $A, B \in$ **CRings**, we can make $A \otimes B$ into a ring by defining

$$\left(\sum_{i \in I} a_i \otimes b_i\right)\left(\sum_{j \in J} c_j \otimes d_j\right) = \sum_{i \in I}\sum_{j \in J} (a_i c_j) \otimes (b_i d_j).$$

Equivalently, multiplication is defined by the distributivity rule together with the rule $(a \otimes b)(c \otimes d) = (ac) \otimes (bd)$. We then have homomorphisms $A \xrightarrow{i} A \otimes B \xleftarrow{j} B$ given by $i(a) = a \otimes 1$ and $j(b) = 1 \otimes b$. If we are given another diagram $A \xrightarrow{f} T \xleftarrow{g} B$, we can define a homomorphism $m \colon A \otimes B \to T$ by the rule

$$m\left(\sum_{i \in I} a_i \otimes b_i\right) = \sum_i f(a_i) g(b_i),$$

and we find that this is the unique homomorphism with $mi = f$ and $mj = g$. Thus $A \otimes B$ is the coproduct of A and B in **CRings**.

Example 36.79. [`eg-poset-coproduct`]
Let X be a set, and let P the poset of subsets of X. By arguments similar to those in Example 36.69, we see that $A \cup B$ is the coproduct of A and B in sP.

We now discuss (co)products of more than two factors.

Definition 36.80. [`defn-indexed-product`]
Let $\mathcal{C}$ be a category, and let $(A_i)_{i \in I}$ be a family of objects of $\mathcal{C}$. A *cone* for the family is an object P equipped with a family of morphisms $p_i \colon P \to A_i$ for all i. Such a cone is a *product diagram* if for every other cone $(T \xrightarrow{f_i} A_i)_{i \in I}$, there is a unique morphism $m \colon T \to P$ with $p_i m = f_i$ for all i. This means that the maps p_i induce a bijection

$$\mathcal{C}(T, P) \to \prod_{i \in I} \mathcal{C}(T, A_i)$$

for all T.

Dually, a *cocone* is an object Q with a family of morphisms $q_i \colon A_i \to Q$. Such a cocone is a *coproduct diagram* if for every other cocone $(A_i \xrightarrow{g_i} T)_{i \in I}$ there is a unique morphism $m \colon Q \to T$ such that $mq_i = g_i$ for all i. This means that the maps q_i induce a bijection

$$\mathcal{C}(Q, T) \to \prod_{i \in I} \mathcal{C}(A_i, T)$$

for all T.

Remark 36.81. [`rem-terminal-product-cones`]
There is a evident way to define a category of cones, and we find that a product diagram is simply a terminal object in that category. Similarly, a coproduct diagram is an initial object in the category of cocones. This means that products and coproducts are unique up to canonical isomorphism if they exist.

Example 36.82. [eg-infinite-products]

Let $(A_i)_{i\in I}$ be a family of sets. Recall that the cartesian product $\prod_{i\in I} A_i$ is the set of all families $a = (a_i)_{i\in I}$, where $a_i \in A_i$ for all i. We have maps $\pi_j \colon \prod_i A_i \to A_j$ given by $\pi_j((a_i)_{i\in I}) = a_j$, giving a cone. Now suppose we have another set T and a family of maps $f_i \colon T \to A_i$. We can then define $m\colon T \to \prod_i A_i$ by $m(t) = (f_i(t))_{i\in I}$, and this is clearly the unique map with $\pi_i m = f_i$ for all i. Thus, we have a product diagram in **Sets**. If each A_i is a group then we can make $\prod_i A_i$ into a group using the obvious rule

$$(a_i)_{i\in I}.(b_i)_{i\in I} = (a_i b_i)_{i\in I}.$$

We find that the maps π_i are group homomorphisms and that they give a product diagram in **Groups** as well as in **Sets**. Products in **Rings** can be constructed in the same way. Similarly, if each A_i is a topological space then we can equip $\prod_i A_i$ with the product topology and the maps π_i then give a product diagram in **Spaces**, as we see from Proposition 5.16.

Example 36.83. [eg-infinite-coproducts]

Again, let $(A_i)_{i\in I}$ be a family of sets. Recall that the disjoint union $\coprod_{i\in I} A_i$ is the set of all pairs (i, a) where $i \in I$ and $a \in A_i$. We have maps $\iota_j \colon A_j \to \coprod_i A_i$ given by $\iota_j(a) = (j, a)$. Given maps $g_i \colon A_i \to T$ for all i, we can define $m\colon \coprod_i A_i \to T$ by $m(i, a) = g_i(a)$, and this is the unique map with $m\iota_i = g_i$ for all i. Thus, the maps ι_i give a coproduct diagram in **Sets**. If the sets A_i have given topologies, then we can topologise $\coprod_i A_i$ as in Definition 5.40, and we see using Proposition 5.43 that this gives a coproduct diagram in **Spaces**.

Example 36.84. [eg-infinite-direct-sum]

Now let $(A_i)_{i\in I}$ be a family of abelian groups. We let $\bigoplus_i A_i$ denote the subgroup of $\prod_i A_i$ consisting of families $(a_i)_{i\in I}$ for which $a_i = 0$ for all but finitely many indices i. We have homomorphisms $\iota_j \colon A_j \to \bigoplus_i A_i$ given by

$$\iota_j(a)_i = \begin{cases} a & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$$

If we have a family of homomorphisms $g_i \colon A_i \to T$, we can define $m\colon \bigoplus_i A_i \to T$ by

$$m((a_i)_{i\in I}) = \sum_{i\in I} g_i(a_i)$$

(which is well-defined, because only finitely many of the terms are nonzero). We find that this is the unique homomorphism with $m\iota_i = g_i$ for all i. This shows that the homomorphisms ι_i give a coproduct diagram.

We next discuss the sense in which products and coproducts are functorial.

Definition 36.85. [defn-power-category]

Let $\mathcal{C}$ be a category, and let I be a set. We introduce a new category $[I, \mathcal{C}]$ as follows. An object of $[I, \mathcal{C}]$ is a family of objects $(A_i)_{i\in I}$ with $A_i \in \mathcal{C}$ for all i. A morphism in $[I, \mathcal{C}]$ from $(A_i)_{i\in I}$ to $(B_i)_{i\in I}$ is a system of morphisms $f_i \colon A_i \to B_i$ for each $i \in I$. Equivalently, we have

$$[I, \mathcal{C}]((A_i)_{i\in I}, (B_i)_{i\in I}) = \prod_{i\in I} \mathcal{C}(A_i, B_i).$$

If we have morphisms

$$(A_i)_{i\in I} \xrightarrow{(f_i)_{i\in I}} (B_i)_{i\in I} \xrightarrow{(g_i)_{i\in I}} (C_i)_{i\in I}$$

then the composite is defined to be the family $(h_i)_{i\in I}$, where $h_i = g_i f_i$.

Proposition 36.86. [prop-product-functor]

Suppose that we have a construction that produces a product diagram $(\prod_I A_i \xrightarrow{\pi_j} A_j)_{j\in I}$ for each family $(A_i)_{i\in I}$ in $[I, \mathcal{C}]$. Then for any morphism

$$f = (f_i)_{i\in I} \colon (A_i)_{i\in I} \to (B_i)_{i\in I}$$

there is a unique morphism $\prod_i f_i$ in $\mathcal{C}$ making the following diagram commute for all j:

$$\begin{tikzcd}
\prod_I A_i \arrow[r, "\prod_i f_i"] \arrow[d, "\pi_j"'] & \prod_I B_i \arrow[d, "\pi_j"] \\
A_j \arrow[r, "f_j"'] & B_j.
\end{tikzcd}$$

Moreover, this construction gives us a functor $\prod_I \colon [I,\mathcal{C}] \to \mathcal{C}$.

We will leave it to the reader to formulate the dual statement.

Proof. We have a family of morphisms $f_j\pi_j \colon \prod_I A_i \to B_j$. As the morphisms $\pi_j \colon \prod_I B_i \to B_j$ form a product diagram, there is a unique morphism $m \colon \prod_I A_i \to \prod_i B_i$ with $\pi_j m = f_j \pi_j$ for all j. We define $\prod_i f_i$ to be m; this clearly has the stated property. Now suppose we have another family of morphisms $g_i \colon B_i \to C_i$. Consider the diagram:

$$\begin{tikzcd}
\prod_I A_i \arrow[r, "\prod_I f_i"] \arrow[d, "\pi_j"'] & \prod_I B_i \arrow[r, "\prod_I g_i"] \arrow[d, "\pi_j"'] & \prod_I C_i \arrow[d, "\pi_j"] \\
A_j \arrow[r, "f_j"'] & B_j \arrow[r, "g_j"'] & C_j.
\end{tikzcd}$$

By inspecing this, we see that the composite $(\prod_I g_i) \circ (\prod_I f_i)$ has the property that defines $\prod_i(g_i f_i)$. A similar argument shows that the identity morphism $1_{\prod_I A_i}$ has the property that defines $\prod_I 1_{A_i}$. This shows that $\prod_I \colon [I,\mathcal{C}] \to \mathcal{C}$ is a functor, as claimed. □

Remark 36.87. [rem-diagonal-adjoint]
For each object $A \in \mathcal{C}$ we have a constant family $\Delta(A) \in [I,\mathcal{C}]$ defined by $\Delta(A)_i = A$ for all i. If we have another family $(B_i)_{i\in I}$, then a morphism from $\Delta(A)$ to B is just a family of morphisms $f_i \colon A \to B_i$, which gives rise to a morphism $f \colon A \to \prod_I B_i$ in $\mathcal{C}$. In other words, we have a bijection

$$\mathcal{C}(A, \prod_I B_i) \simeq [I,\mathcal{C}](\Delta(A), (B_i)_{i\in I}).$$

One can check that this has the naturality property described in Proposition 36.60, so we see that the functor $\prod_I \colon [I,\mathcal{C}] \to \mathcal{C}$ is right adjoint to Δ. By a dual argument, the functor $\coprod_I \colon [I,\mathcal{C}] \to \mathcal{C}$ is left adjoint to Δ.

36.7. Limits and colimits.

Definition 36.88. [defn-diagram-category]
We say that a category I is *small* if the collection of objects is a set (rather than a proper class). Given a small category I and an arbitrary category $\mathcal{C}$, we write $[I,\mathcal{C}]$ for the category of functors from I to $\mathcal{C}$. More precisely, the objects of $[I,\mathcal{C}]$ are functors, and the morphisms are the natural transformations. We will often refer to $[I,\mathcal{C}]$ as a *diagram category*, and to the objects as *I-shaped diagrams*. The examples below will make it clear why this is reasonable.

Example 36.89. [eg-discrete-diagrams]
Let I be a set. We then have a category dI with object set I and only identity morphisms (this is called a *discrete category*). We find that a functor $A \colon dI \to \mathcal{C}$ is determined by the family of objects $(A(i))_{i\in I}$, so $[dI,\mathcal{C}]$ is equivalent to the category $[I,\mathcal{C}]$ discussed in Section 36.6. We will usually not distinguish between the set I and the discrete category dI.

Example 36.90. [eg-equivariant]
Let G be a group, and let $\mathcal{C}$ be a category. A *G-object* in $\mathcal{C}$ is an object $A \in \mathcal{C}$ equipped with morphisms $\mu_g \colon A \to A$ for each $g \in G$ satisfying $\mu_1 = 1_A$ and $\mu_{gh} = \mu_g\mu_h$. For example, we can consider G-sets,

G-spaces, G-rings and so on. Suppose we have G-objects A and B, and a morphism $f\colon A \to B$. We say that f is *G-equivariant* if the following diagram commutes for all $g \in G$:

$$\begin{tikzcd} A \arrow[r, "f"] \arrow[d, "\mu_g"'] & B \arrow[d, "\mu_g"] \\ A \arrow[r, "f"'] & B. \end{tikzcd}$$

We write $G\mathcal{C}$ for the category of G-objects and equivariant morphisms. By a slight elaboration of Example 36.41, we see that $G\mathcal{C}$ is equivalent to the functor category $[bG, \mathcal{C}]$.

Example 36.91. [eg-square-functor]
Let I be the category with four objects that we discussed in Example 36.10:

$$\begin{tikzcd} \bullet \arrow[r] \arrow[d] & \bullet \arrow[d] \\ \bullet \arrow[r] & \bullet \end{tikzcd}$$

Then an object of $[I, \mathcal{C}]$ is just a commutative square in $\mathcal{C}$.

Example 36.92. [eg-sequence-functor]
Consider $\mathbb{N}$ as a poset in the obvious way, and consider the category $s\mathbb{N}$. For each $n \leq m$ there is a unique morphism $u_{n,m}\colon n \to m$ in $s\mathbb{N}$. If we put $v_n = u_{n,n+1}$, we see that $u_{n,m}$ is the composite

$$n \xrightarrow{v_n} (n+1) \xrightarrow{v_{n+1}} (n+2) \xrightarrow{v_{n+2}} \cdots \xrightarrow{v_{m-2}} (m-1) \xrightarrow{v_{m-1}} m.$$

Thus, for any functor $F\colon s\mathbb{N} \to \mathcal{C}$ we have

$$F(u_{n,m}) = (F(n) \xrightarrow{F(v_n)} \cdots \xrightarrow{F(v_{m-1})} F(m)).$$

Using this, we see that the objects of $[s\mathbb{N}, \mathcal{C}]$ are just the diagrams of type

$$A_0 \xrightarrow{f_0} A_1 \xrightarrow{f_2} A_2 \to \cdots.$$

A morphism between two such diagrams is given by a commutative ladder as follows:

$$\begin{tikzcd} A_0 \arrow[r, "f_0"] \arrow[d, "p_0"'] & A_1 \arrow[r, "f_1"] \arrow[d, "p_1"'] & A_2 \arrow[r, "f_2"] \arrow[d, "p_2"'] & A_3 \arrow[r, "f_3"] \arrow[d, "p_3"'] & A_4 \arrow[r, "f_4"] \arrow[d, "p_4"'] & \cdots \\ B_0 \arrow[r, "g_0"'] & B_1 \arrow[r, "g_1"'] & B_2 \arrow[r, "g_2"'] & B_3 \arrow[r, "g_3"'] & B_4 \arrow[r, "g_4"'] & {} \end{tikzcd}$$

Remark 36.93. [rem-diagram-notation]
Let X be an I-shaped diagram in $\mathcal{C}$, so $X\colon I \to \mathcal{C}$. We will often emphasise the analogy with Examples 36.89 and 36.92 by writing X_i instead of $X(i)$ for the objects in the diagram.

Definition 36.94. [defn-limit-cone]
Let X be an I-shaped diagram in $\mathcal{C}$. A *cone* for X is an object $P \in \mathcal{C}$ together with a system of morphisms $p_i\colon P \to X_i$ such that for all $u\colon i \to j$ in I the left-hand diagram below commutes:

$$\begin{tikzcd} & P \arrow[dl, "p_i"'] \arrow[dr, "p_j"] & \\ X_i \arrow[rr, "u_*"'] & & X_j. \end{tikzcd} \qquad \begin{tikzcd} P \arrow[rr, "m"] \arrow[dr, "p_i"'] & & Q \arrow[dl, "q_i"] \\ & X_i & \end{tikzcd}$$

Now suppose we have two cones, say $(P \xrightarrow{p_i} X_i)_{i \in I}$ and $(Q \xrightarrow{q_i} X_i)_{i \in I}$. A *morphism of cones* between them is a morphism $m\colon P \to Q$ in $\mathcal{C}$ such that for all $i \in I$, the right-hand diagram above commutes. This defines a category of cones. A *limit cone* is a terminal object in this category. If a limit cone exists then it is unique up to canonical isomorphism, and we denote it by $\varprojlim_I X_i$.

Example 36.95. [`eg-products-as-limits`]

If I is a set (regarded as a discrete category) then an I-shaped diagram in $\mathcal{C}$ is just an I-indexed family of objects of $\mathcal{C}$, and we have $\varprojlim_I X_i = \prod_I X_i$. More precisely: if a product for the family exists, then it is a limit for the diagram, and *vice versa*.

Example 36.96. [`eg-fixpoints-as-limit`]

Let X be a G-set, regarded as a functor $bG \to \mathbf{Sets}$. As bG has only a single object, a cone for this diagram just consists of a set P with a single map $p\colon P \to X$, and the cone condition says that $\mu_g \circ p = p$ for all $g \in G$. Put

$$X^G = \{x \in X \;:\; gx = x \text{ for all } g \in G\},$$

and let $i\colon X^G \to X$ be the inclusion map. It is now easy to see that $X^G \xrightarrow{i} X$ is a limit cone, or in other words that $\varprojlim_{bG} X = X^G$.

Example 36.97. [`eg-equaliser`]

Let I denote the following category:

$$\begin{tikzcd} \bullet \arrow[r, shift left] \arrow[r, shift right] & \bullet \end{tikzcd}$$

I-shaped diagrams are called *forks*; they have the form

$$\begin{tikzcd} X \arrow[r, shift left, "u"] \arrow[r, shift right, "v"'] & Y \end{tikzcd}$$

A limit for this fork is called an *equaliser* of u and v. A cone for the fork consists of an object P together with maps $p\colon P \to X$ and $q\colon Q \to Y$ such that $up = q = vp$. As p determines q we need not mention q explicitly. Thus, we can say that a cone is a morphism $p\colon P \to X$ with $up = vp$.

Now let $\mathcal{C}$ be the category of sets, and put $E = \{x \in X \;:\; u(x) = v(x)\}$, and let $i\colon E \to X$ be the inclusion. For any cone $P \xrightarrow{p} X$ we see that $p(P)$ must be a subset of E, so p gives a map $m\colon P \to E$ with $im = p$. Using this, we see that E is an equaliser. If X and Y are groups and u and v are homomorphisms then E is a subgroup of X and we see that we have a limit cone in **Groups**. By similar arguments, we see that this construction gives equalisers in **Rings**, **Fields**, $\mathbf{Vect}_{\mathbb{R}}$ and similar algebraic categories. In the case of abelian groups we can also describe E as $\ker(u - v)$.

Example 36.98. [`eg-pullback`]

Consider a commutative square

$$\begin{tikzcd} W \arrow[r, "p"] \arrow[d, "q"'] & X \arrow[d, "r"] \\ Y \arrow[r, "s"'] & Z \end{tikzcd}$$

(in some category $\mathcal{C}$). We say that this is a *pullback square* if for every object T and every pair of maps $(Y \xleftarrow{g} T \xrightarrow{f} X)$ with $rf = sg$, there is a unique map $m\colon T \to W$ with $pm = f$ and $qm = g$.

$$\begin{tikzcd} T \arrow[drr, bend left, "f"] \arrow[ddr, bend right, "g"'] \arrow[dr, dotted, "m"] & & \\ & W \arrow[r, "p"] \arrow[d, "q"'] & X \arrow[d, "r"] \\ & Y \arrow[r, "s"'] & Z \end{tikzcd}$$

This can be reinterpreted in terms of limits as follows. We can regard the diagram $(Y \xrightarrow{s} Z \xleftarrow{r} X)$ as a functor from the category $I = (\bullet \leftarrow \bullet \rightarrow \bullet)$ to $\mathcal{C}$. A cone for this is an object T together with morphisms f,

g and h making the following diagram commute:

$$\begin{tikzcd} T \arrow[r, "f"] \arrow[d, "g"'] \arrow[dr, dotted, "h"] & X \arrow[d, "r"] \\ Y \arrow[r, "s"'] & Z \end{tikzcd}$$

Note that $h = sg = rf$, so we need not mention h as a separate datum: we can just say that a cone is a diagram $(Y \xleftarrow{g} T \xrightarrow{f} X)$ with $sg = rf$, so that the square above commutes. Given this, it is clear that our original square is a pullback iff $(Y \xleftarrow{q} W \xrightarrow{p} X)$ is a limit cone for $(Y \xrightarrow{s} Z \xleftarrow{r} X)$.

In the category of sets, we can construct a pullback for $(Y \xrightarrow{s} Z \xleftarrow{r} X)$ by taking

$$W = \{(x, y) \in X \times Y \ : \ r(x) = s(y)\}$$

with $p(x, y) = x$ and $q(x, y) = y$. The same construction works in **Groups**, **Rings**, **Fields**, $\mathbf{Vect}_{\mathbb{R}}$ and so on. (The case of fields requires a little extra thought, because $X \times Y$ is then a ring but not a field. However, after recalling that r and s are necessarily injective we find that W is a field, and this turns out to be enough.)

Example 36.99. [`eg-tower`]

A *tower* in $\mathcal{C}$ is a functor $s\mathbb{N}^{\mathrm{op}} \to \mathcal{C}$, or equivalently a diagram of the form

$$X_0 \xleftarrow{u_0} X_1 \xleftarrow{u_1} X_2 \xleftarrow{u_2} \cdots$$

In the category of sets (or groups, or rings) we can construct a limit by the rule

$$\varprojlim_i X_i = \{x = (x_i)_{i=0}^{\infty} \ : \ f_i(x_{i+1}) = x_i \text{ for all } i\}.$$

Example 36.100. [`eg-limit-initial`]

Suppose that I has an initial object, say i_0. We claim that for any diagram $X \colon I \to \mathcal{C}$ we have $\varprojlim_I X_i = X_{i_0}$. For a more precise statement, let a_i be the unique morphism from i_0 to i in I, so for each $u \colon i \to j$ we must have $u \circ a_i = a_j$. We then have a family of maps $(a_i)_* \colon X_{i_0} \to X_i$, and using the fact that $ua_i = a_j$ we see that they form a cone. We claim that it is a limit cone. Indeed, if $(T \xrightarrow{f_i} X_i)$ is any other cone, we have a morphism $f_{i_0} \colon T \to X_{i_0}$. The cone condition for T says that whenever $u \colon i \to j$ we have $u_* \circ f_i = f_j$; by taking $u = a_i$, we see that $(a_i)_* \circ f_{i_0} = f_i$, which means that f_{i_0} is a morphism of cones. If $g \colon T \to X_{i_0}$ is any morphism of cones, then by definition we have $(a_i)_* \circ g = f_i$ for all i. We can take $i = i_0$ and note that $a_{i_0} = 1$ to see that $g = f_{i_0}$. Thus, we have a unique morphism of cones, proving that X_{i_0} is the limit as claimed.

Proposition 36.101. [`prop-product-equaliser`]

Suppose that $\mathcal{C}$ has a product for every family of objects, and an equaliser for every fork. Then $\mathcal{C}$ has a limit for every diagram.

Proof. Consider a diagram $X \colon I \to \mathcal{C}$. Let M be the set of all morphisms in I. For each morphism $(u \colon i \to j) \in M$, we put $s(u) = i$ (the *source* of u) and $t(u) = j$ (the *target*). We then let $(P \xrightarrow{p_i} X_i)_{i \in I}$ be a product diagram for the family $(X_i)_{i \in I}$, and we let $(Q \xrightarrow{q_u} X_{t(u)})_{u \in M}$ be a product cone for the family $(X_{t(u)})_{u \in M}$. We define maps $\lambda_u, \mu_u \colon P \to X_{t(u)}$ as follows: λ_u is just the projection $p_{t(u)}$, whereas μ_u is the composite

$$P \xrightarrow{p_s(u)} X_{s(u)} \xrightarrow{u_*} X_{t(u)}.$$

By the defining property of Q, there is a unique morphism $\lambda \colon P \to Q$ with $q_u \lambda = \lambda_u$ for all $u \in M$, and also a unique morphism $\mu \colon P \to Q$ with $q_u \mu = \mu_u$ for all $u \in M$. Let $L \xrightarrow{\phi} P$ be an equaliser for λ and μ, and put $r_i = p_i \phi$. Note that $\lambda \phi = \mu \phi$, so for each $(u \colon i \to j) \in M$ we have $\lambda_u \phi = q_u \lambda \phi = q_u \mu \phi = \mu_u \phi$. Filling in the definitions of λ_u and μ_u, this gives $p_j \phi = u_* p_i \phi$, or in other words $r_j = u_* r_i$. This means that the maps $(L \xrightarrow{r_i} X_i)_{i \in I}$ form a cone. Suppose we have another cone $(T_i \xrightarrow{f_i} X_i)_{i \in I}$. By the defining property of P, there is a unique map $g \colon T \to P$ with $p_i g = f_i$ for all i. We claim that $\lambda g = \mu g$. By the uniqueness clause in the defining property of Q, it will suffice to check that $q_u\lambda g = q_u\mu g$ for all $u\colon i \to j$. Here $q_u\lambda g = \lambda_u g = p_j g = f_j$ and $q_u\mu g = \mu_u g = u_* p_i g = u_* f_i$. As $(f_i)_{i\in I}$ is a cone, these are the same, as required. As $\phi\colon L \to P$ is an equaliser of λ and μ, we see that there is a unique $h\colon T \to L$ with $\phi h = g$. We now have $r_i h = p_i\phi h = p_i g = f_i$, so h is a morphism of cones from $(T \xrightarrow{f_i} X_i)_{i\in I}$ to $(L \xrightarrow{r_i} X_i)_{i\in I}$. By partially reversing the argument we see that it is the unique such morphism. We thus have a limit cone, as required. □

Corollary 36.102. *Each of the categories **Sets**, **Groups**, **AbGroups**, **Rings** and $\textbf{Vect}_{\mathbb{R}}$ has limits for all diagrams.* □

Remark 36.103. [rem-limit-too-big]
Recall that according to our definitions, a diagram in **Sets** is a functor from a *small* category to **Sets**. One can formulate the notion of a limit cone for a functor $I \to \mathbf{Sets}$ even if I is large, but no such cone need exist. For example, we can consider the class of all ordinals as a large discrete category, and consider the constant functor sending each ordinal to the set $\{0,1\}$; this has no limit. (Informally we can say that the limit should be $\prod_\alpha\{0,1\}$, and that this is too big to exist as a set.) On the other hand, we can consider the constant functor $\alpha \mapsto \{0\}$ and we find that this does have a limit, which is again the set $\{0\}$. Thus, some large diagrams do have limits.

Proposition 36.104. [prop-preserves-limits]
Suppose that $G\colon \mathcal{D} \to \mathcal{C}$ has a left adjoint. Then for any diagram $Y\colon I \to \mathcal{D}$ and any limit cone $(Q \xrightarrow{q_i} Y_i)_{i\in I}$, the system $(GQ \xrightarrow{G(q_i)} GY_i)_{i\in I}$ is a limit cone for the diagram $I \xrightarrow{Y} \mathcal{D} \xrightarrow{G} \mathcal{C}$. More loosely, we can say that

$$G(\varprojlim_I Y_i) = \varprojlim_I GY_i,$$

or that G preserves limits.

Proof. Let $F\colon \mathcal{C} \to \mathcal{D}$ be left adjoint to G, so we have bijections

$$\rho\colon \mathcal{D}(C, GD) \to \mathcal{C}(FC, D)$$

with naturality properties as in Proposition 36.60. Now consider a cone $(P \xrightarrow{p_i} GY_i)_{i\in I}$, so for $u\colon i \to j$ in I we have $G(u_*) \circ p_i = p_j\colon P \to GY_j$. We then have morphisms $\rho(p_i)\colon FP \to Y_i$. These satisfy $u_* \circ \rho(p_i) = \rho(G(u_*) \circ p_i) = \rho(p_j)$, so they form a cone. As $(Q \xrightarrow{q_i} Y_i)_{i\in I}$ is assumed to be a limit cone, there is a unique morphism $m\colon FP \to Q$ with $q_i m = \rho(p_i)$ for all i. If we put $n = \rho^{-1}(m)\colon P \to GQ$ we find that $\rho(G(q_i)n) = q_i\rho(n) = q_i m = \rho(p_i)$, but ρ is a bijection so $G(q_i)n = p_i$. Thus, n is a morphism of cones from $(P \xrightarrow{p_i} GY_i)$ to $(GQ \xrightarrow{G(q_i)} GY_i)$. By similar arguments, we see that it is the unique such morphism. □

Example 36.105. [eg-preserves-limits]
Define $F, G\colon \mathbf{AbGroups} \to \mathbf{AbGroups}$ by $F(A) = A/2A$ and $G(A) = A[2] = \{a \in A \;:\; 2a = 0\}$. There are then evident bijections

$$\mathbf{AbGroups}(F(A), B) \simeq \{f\colon A \to B \;:\; 2f = 0\} \simeq \mathbf{AbGroups}(A, G(B)),$$

so F is left adjoint to G. Proposition 36.104 therefore tells us that G preserves limits. For example, for any family of abelian groups B_i we have $(\prod_i B_i)[2] = \prod_i B_i[2]$. Similarly, if we have two homomorphisms $f, g\colon A \to B$ with equaliser E, then the equaliser of the restricted maps $A[2] \to B[2]$ is $E[2]$. Both of these facts can easily be seen directly.

Proposition 36.106. [prop-limit-as-adjoint]
Fix a small category I and an arbitrary category $\mathcal{C}$. Suppose we have a construction that gives, for each diagram $X\colon I \to \mathcal{C}$, a limit cone $(\varprojlim_I X_i \xrightarrow{p_j} X_j)_{j\in I}$. Then there is a canonical way to define an action on morphisms giving a functor $\varprojlim_I\colon [I, \mathcal{C}] \to \mathcal{C}$. Moreover, we also have a functor $\Delta\colon \mathcal{C} \to [I, \mathcal{C}]$ given by $\Delta(T)(i) = T$ for all I, and $\varprojlim_I$ is right adjoint to Δ.

Proof. This is just a slight elaboration of Proposition 36.86, which is the special case where I is discrete. Suppose we have diagrams X and Y, and a morphism $f\colon X \to Y$ in $[I,\mathcal{C}]$; we need to define a map $\varprojlim_I f_i\colon \varprojlim_I X_i \to \varprojlim_I Y_i$. Consider the diagram

$$\begin{tikzcd} \varprojlim_I X_i \arrow[r, "p_j"] \arrow[d, dotted, "\varprojlim_I f_i"'] & X_j \arrow[r, "u_*"] \arrow[d, "f_j"'] & X_k \arrow[d, "f_k"] \\ \varprojlim_I Y_i \arrow[r, "p_j"'] & Y_j \arrow[r, "u_*"'] & Y_k. \end{tikzcd}$$

For every j we have a morphism $f_j p_j\colon \varprojlim_I X_i \to Y_j$ as above. For every morphism $u\colon j \to k$ we have a square as shown on the right, which commutes because f is assumed to be a morphism of diagrams. We also have $u_* p_j = p_k$, and using this we see that the maps $f_j p_j$ form a cone for Y. By the defining property of $\varprojlim_I Y_i$, there is a unique map $\varprojlim_I X_i \to \varprojlim_I Y_i$ such that the left square commutes for all j. This defines an action on morphisms, and we leave it to the reader to check that this makes $\varprojlim_I$ into a functor. A morphism of diagrams from $\Delta(A)$ to X is just the same as a cone of the form $(A \xrightarrow{e_i} X_i)_{i\in I}$, and such cones biject with morphisms $A \to \varprojlim_I X_i$. This shows that $\varprojlim_I$ is left adjoint to Δ, as claimed. □

It is also useful to know that different types of limit constructions commute with each other, in a sense that we now explain.

Proposition 36.107. *Let I and J be small categories, let $\mathcal{C}$ be a category to which Proposition 36.106 applies, and let $X\colon I \times J \to \mathcal{C}$ be a functor. Then there are natural isomorphisms*

$$\varprojlim_I \varprojlim_J X_{ij} = \varprojlim_{I\times J} X_{ij} = \varprojlim_J \varprojlim_I X_{ij}.$$

Proof. Roughly speaking, the idea is as follows. For any test object T, morphisms from T to the iterated limit $\varprojlim_I \varprojlim_J X_{ij}$ biject with compatible families of morphisms $T \to \varprojlim_J X_{ij}$ for all i. Moreover, for fixed i, morphisms from T to $\varprojlim_J X_{ij}$ biject with compatible families of morphisms to X_{ij}. By combining these descriptions, we see that morphisms from T to $\varprojlim_I \varprojlim_J X_{ij}$ biject with compatible families of morphisms $T \to X_{ij}$ for all i and j, and this is the definining property of $\varprojlim_{I\times J} X_{ij}$.

We will now give a more detailed account.

We start by formulating the statement in a more careful way. By definition we have

$$(I \times J)((i,j),(i',j')) = I(i,i') \times J(j,j').$$

Thus, for each $i \in I$ we can define a functor $\phi_i\colon J \to I \times J$ by $\phi_i(j) = (i,j)$ on objects, and $\phi_i(v) = (1_i, v)$ on morphisms. We then have a functor $X \circ \phi_i\colon J \to \mathcal{C}$ and thus an inverse limit $\varprojlim_J (X \circ \phi_i)_j$, which we also denote more briefly by $\varprojlim_J X_{ij}$. Next, for any morphism $u\colon i \to i'$ in I we have a family of morphisms $(u, 1_j)\colon (i,j) \to (i',j)$ in $I \times J$, and thus morphisms

$$(u,1_j)_*\colon (X \circ \phi_i)_j = X_{ij} \to X_{i'j} = (X \circ \phi_{i'})_j$$

in $\mathcal{C}$. It is straightforward to check that these give a morphism of diagrams $\lambda_u\colon X \circ \phi_i \to X \circ \phi_{i'}$, and thus a morphism

$$\mu_u = \varprojlim_J \lambda_u\colon \varprojlim_J (X \circ \phi_i)_j \to \varprojlim_J (X \circ \phi_{i'})_j$$

in $\mathcal{C}$. The objects $\varprojlim_J (X \circ \phi_i)_j$ together with the morphisms μ_u give a diagram $I \to \mathcal{C}$, whose limit is naturally denoted by $\varprojlim_I \varprojlim_J X_{ij}$. It is this that we claim is isomorphic to $\varprojlim_{I\times J} X_{ij}$. In the proof, we will use the notation

$$
\begin{aligned}
p_{ij}\colon & \varprojlim_J X_{ij} \to X_{ij} \\
q_i\colon & \varprojlim_I \varprojlim_J X_{ij} \to \varprojlim_J X_{ij} \\
r_{ij}\colon & \varprojlim_{I\times J} X_{ij} \to X_{ij}
\end{aligned}
$$

for the canonical projections.

From the definitions, we see that the following two diagrams commute (for all $u\colon i \to i'$ in I and $v\colon j \to j'$ in J).

$$
\begin{tikzcd}
& X_{ij} \arrow[rr, "(u,1)_*"] \arrow[dd, "(1,v)_*"'] & & X_{i'j} \arrow[dd, "(1,v)_*"'] & \\
\varprojlim_J X_{ij} \arrow[ur, "p_{ij}"] \arrow[dr, "p_{ij'}"'] & & \varprojlim_{I\times J} X_{ij} \arrow[ul, "r_{ij}"] \arrow[dl, "r_{ij'}"'] \arrow[ur, "r_{i'j}"'] \arrow[dr, "r_{i'j'}"] & & \varprojlim_J X_{i'j} \arrow[ul, "p_{i'j}"'] \arrow[dl, "p_{i'j'}"] \\
& X_{ij'} \arrow[rr, "(u,1)_*"'] & & X_{i'j'} &
\end{tikzcd}
$$

$$
\begin{tikzcd}
& & \varprojlim_I \varprojlim_J X_{ij} \arrow[dl, "q_i"'] \arrow[dr, "q_{i'}"] & & \\
& \varprojlim_J X_{ij} \arrow[rr, "\mu_u"] \arrow[dl, "p_{ij'}"'] \arrow[d, "p_{ij}"] & & \varprojlim_J X_{i'j} \arrow[d, "p_{i'j}"'] \arrow[dr, "p_{i'j'}"] & \\
X_{ij'} & X_{ij} \arrow[l, "(1,v)_*"] \arrow[rr, "(u,1)_*"'] & & X_{i'j} \arrow[r, "(1,v)_*"'] & X_{i'j'}
\end{tikzcd}
$$

Unfinished □

We next discuss the dual theory briefly.

Definition 36.108. [`defn-colimit-cone`]
Let X be an I-shaped diagram in $\mathcal{C}$. A *cocone* for X is an object $Q \in \mathcal{C}$ together with a system of morphisms $q_i\colon X_i \to Q$ such that for all $u\colon i \to j$ in I the left-hand diagram below commutes:

$$
\begin{tikzcd}
X_i \arrow[rr, "u_*"] \arrow[dr, "q_i"'] & & X_j \arrow[dl, "q_j"] \\
& Q &
\end{tikzcd}
\qquad\qquad
\begin{tikzcd}
& X_i \arrow[dl, "q_i"'] \arrow[dr, "r_i"] & \\
Q \arrow[rr, "n"'] & & R.
\end{tikzcd}
$$

Now suppose we have two cocones, say $(X_i \xrightarrow{q_i} Q)_{i\in I}$ and $(X_i \xrightarrow{r_i} R)_{i\in I}$. A *morphism of cocones* between them is a morphism $n\colon Q \to R$ in $\mathcal{C}$ such that for all $i \in I$, the right-hand diagram above commutes. This defines a category of cocones. A *colimit cocone* is an initial object in this category. If a colimit cocone exists then it is unique up to canonical isomorphism, and we denote it by $\varinjlim_I X_i$.

Example 36.109. In the case where I is a discrete category, colimits are the same as coproducts. If X is a G-set considered as a functor $bG \to \mathbf{Sets}$, then the quotient map from X to the orbit set X/G is a colimit cocone.

Example 36.110. [`eg-coequaliser`]

The colimit of a fork is called a *coequaliser*. In the category of abelian groups, the coequaliser of a fork

$$\begin{tikzcd} A \arrow[r, shift left, "u"] \arrow[r, shift right, "v"'] & B \end{tikzcd}$$

is the quotient group $B/\operatorname{img}(u-v)$. Now consider instead a fork

$$\begin{tikzcd} X \arrow[r, shift left, "u"] \arrow[r, shift right, "v"'] & Y \end{tikzcd}$$

in the category of sets. Put

$$R_0 = \{(y,y) \;:\; y \in Y\} \cup \{(u(x),v(x)) \;:\; x \in X\} \cup \{(v(x),u(x)) \;:\; x \in X\}$$

then define R_n recursively for $n > 0$ by

$$R_{n+1} = \{(y,z) \in Y^2 \;:\; \text{ there exists } a \in Y \text{ with } (y,a) \in R_n \text{ and } (a,z) \in R_n\}.$$

Finally, we put $R_\infty = \bigcup_n R_n$. We identify subsets of Y^2 with relations on Y in the usual way, so ySz means that $(y,z) \in S$. With this convention, we see that each of the relations R_n is reflexive and symmetric, that $R_n \subseteq R_{n+1}$, and that R_∞ is also transitive. This means that R_∞ is an equivalence relation, and in fact it is the smallest equivalence relation containing $\{(u(x),v(x)) \;:\; x \in X\}$. We can therefore put $Q = Y/R_\infty$ (the set of equivalence classes) and let $q\colon Y \to Q$ be the quotient map (sending $y \in Y$ to the corresponding equivalence class $[y]$). As $(u(x),v(x)) \in R_0 \subseteq R_\infty$ we see that $[u(x)] = [v(x)]$, so $qu = qv\colon X \to Q$, so q is a cocone for the fork. We claim that it is a coequaliser. Indeed, if $r\colon Y \to R$ is another cocone, we can put $E = \{(y,z) \in Y^2 \;:\; r(y) = r(z)\}$. This is easily seen to be an equivalence relation containing R_0, so it contains R_n for all n by induction on n, so it contains R_∞. In other words, if $[y] = [z]$ we have $r(y) = r(z)$. We thus have a well-defined map $n\colon Q \to R$ given by $n([y]) = r(y)$, and this is the unique map with $nq = r$, as required.

EXAMPLE 36.111. [`eg-coproduct-coequaliser`]

Let $\mathcal{C}$ be a category (such as **Sets** or **AbGroups**) that has a coproduct for every family of objects, and a coequaliser for every fork. We can then construct a colimit for an arbitrary diagram $X\colon I \to \mathcal{C}$ by the dual of Proposition 36.101. With some slight abuse of language, we can say that $\varinjlim_I X_i$ is the coequaliser of the fork

$$\begin{tikzcd} \coprod_u X_{s(u)} \arrow[r, shift left, "\alpha"] \arrow[r, shift right, "\beta"'] & \coprod_i X_i \end{tikzcd}$$

where α maps $X_{s(u)}$ to $X_{s(u)}$ by the identity, and β maps $X_{s(u)}$ to $X_{t(u)}$ by u_*.

EXAMPLE 36.112. [`eg-pushout`]

Consider a commutative square

$$\begin{tikzcd} W \arrow[r, "p"] \arrow[d, "q"'] & X \arrow[d, "r"] \\ Y \arrow[r, "s"'] & Z \end{tikzcd}$$

(in some category $\mathcal{C}$). We say that this is a *pushout square* if for every object T and every pair of maps $(X \xrightarrow{f} T \xleftarrow{g} Y)$ with $fp = gq$, there is a unique map $m\colon Z \to T$ with $mr = f$ and $ms = g$.

$$\begin{tikzcd} W \arrow[r, "p"] \arrow[d, "q"'] & X \arrow[d, "r"] \arrow[rdd, bend left, "f"] & \\ Y \arrow[r, "s"'] \arrow[rrd, bend right, "g"'] & Z \arrow[rd, dotted, "m"] & \\ & & T \end{tikzcd}$$

It is equivalent to say that $(X \xrightarrow{r} Z \xleftarrow{s} Y)$ is a colimit cocone for the diagram $(X \xleftarrow{p} W \xrightarrow{q} Y)$

In the category of abelian groups, we can construct a pushout for $B \xleftarrow{p} A \xrightarrow{q} C$ as follows. We have a homomorphism $d\colon A \to B \oplus C$ given by $d(a) = (p(a), -q(a))$, and thus a quotient group $D = (B \oplus C)/d(A)$. We define maps $(B \xrightarrow{r} D \xleftarrow{s} C)$ by $r(b) = (b,0) + d(A)$ and $s(c) = (0,c) + d(A)$. We then find that the square

$$\begin{tikzcd} A \arrow[r, "p"] \arrow[d, "q"'] & B \arrow[d, "r"] \\ C \arrow[r, "s"'] & D \end{tikzcd}$$

is a pushout.

Proposition 36.113. [prop-preserves-colimits]
Suppose that $F\colon \mathcal{C} \to \mathcal{D}$ has a right adjoint. Then for any diagram $X\colon I \to \mathcal{C}$ and any colimit cocone $(X_i \xrightarrow{q_i} Q)_{i \in I}$, the system $(FX_i \xrightarrow{F(q_i)} Q)_{i \in I}$ is a colimit cocone for the diagram $I \xrightarrow{X} \mathcal{C} \xrightarrow{F} \mathcal{D}$. More loosely, we can say that

$$F(\varinjlim_I X_i) = \varinjlim_I FX_i,$$

or that F preserves colimits.

Proof. Dual to Proposition 36.104. ☐

36.8. Filtered colimits.

Definition 36.114. [defn-filtered]
Let I be a small category. We say that I is *filtered* if

FC0: There is at least one object in I.
FC2: For all $i, j \in I$ there exists an object $k \in I$ and morphisms $i \xrightarrow{u} k \xleftarrow{v} j$.
FC3: For any two parallel morphisms $u, v\colon i \to j$ in I, there is an object k and a morphism $w\colon j \to k$ with $wu = wv$.

A *filtered diagram* means a diagram indexed by a filtered category, and a *filtered colimit* means the colimit of such a diagram.

We will show that filtered colimits behave better than more general colimits in a variety of ways. First, however, we give some examples.

Definition 36.115. [defn-directed]
A *directed set* is a partially ordered set in which every finite subset has an upper bound.

Remark 36.116. [rem-directed]
Let D be a nonempty partially ordered set in which every pair of elements has an upper bound. Consider a finite subset $F \subseteq D$. If $F = \emptyset$ then any element is an upper bound, if $F = \{i\}$ then i is an upper bound, and if $F = \{i, j\}$ then by hypothesis there is an upper bound. If $|F| > 2$ then we can write $F = F' \amalg \{j\}$ say, and by induction there is an upper bound (say i) for F', and by hypothesis there is an upper bound (say k) for $\{i, j\}$, and we see that k is an upper bound for F. This means that the category sD is filtered. (Note that axiom FC3 is vacuous here, because parallel morphisms in sD are always equal.)

Example 36.117. [eg-Nr-directed]
For any $r \geq 0$, the poset $\mathbb{N}^r$ is directed. Indeed, it is clearly nonempty, and if $i, j \in \mathbb{N}^r$ we can put

$$k = (\max(i_1, j_1), \ldots, \max(i_r, j_r))$$

to get an element $k \in \mathbb{N}^r$ that is an upper bound for $\{i, j\}$.

Example 36.118. [eg-subsets-directed]
Let X be any set, and let D be the set of finite subsets of X, ordered by inclusion. Given any finite subset $F = \{A_1, \ldots, A_r\}$, the union $B = \bigcup_i A_i$ is an upper bound for F in D. It follows that D is directed.

Example 36.119. [eg-discrete-not-filtered]
Let I be a set, regarded as a discrete category. Then I can only be filtered if $|I| = 1$. Similarly, if G is a nontrivial group, then bG is not filtered.

Example 36.120. If I is any small category with a terminal object, then I is filtered.

36.9. The Yoneda lemma.

Definition 36.121. [defn-representable]
Let $\mathcal{C}$ be a category. For any object $A \in \mathcal{C}$, we define a functor $YA \colon \mathcal{C} \to \mathbf{Sets}$ by $(YA)(X) = \mathcal{C}(A, X)$. For a morphism $f \colon W \to X$ in $\mathcal{C}$, the induced map

$$f_* \colon (YA)(W) = \mathcal{C}(A, W) \to \mathcal{C}(A, X) = (YA)(X)$$

is defined by $f_*(u) = f \circ u$. We say that a functor $F \colon \mathcal{C} \to \mathbf{Sets}$ is *representable* if there is a natural isomorphism $YA \to F$ for some A (in which case we say that A *represents* F).

Example 36.122. [eg-representable-abgroups]
We can define functors $F, G, H \colon \mathbf{AbGroups} \to \mathbf{Sets}$ by

$$\begin{aligned} F(A) &= A/2A \\ G(A) &= \{a \in A \ : \ 2a = 0\} \\ H(A) &= A^2. \end{aligned}$$

It is easy to see that $\mathbf{AbGroups}(\mathbb{Z}/2, A) \simeq G(A)$ and $\mathbf{AbGroups}(\mathbb{Z} \oplus \mathbb{Z}, A) \simeq H(A)$ and that these isomorphisms are natural in A; so G and H are representable. On the other hand, the inclusion $i \colon \mathbb{Z} \to \mathbb{Q}$ is injective but the induced map $i_* \colon F(\mathbb{Z}) \to F(\mathbb{Q})$ is not; it follows easily from this that F is not representable.

Example 36.123. [eg-representable-crings]
Let $\mathbf{CRings}$ be the category of commutative rings. We can define functors $F, G \colon \mathbf{CRings} \to \mathbf{Sets}$ by

$$\begin{aligned} F(R) &= \{(a, b, c) \in R^3 \ : \ a^3 + b^3 + c^3 = 0\} \\ G(R) &= \{a \in R \ : \ a \text{ is invertible } \}. \end{aligned}$$

We claim that F is represented by the ring $A = \mathbb{Z}[x, y, z]/(x^3 + y^3 + z^3 - 1)$, and that G is represented by the ring $B = \mathbb{Z}[x, y]/(xy - 1)$. Indeed, suppose we have an element $(a, b, c) \in F(R)$. We can then define $f_0 \colon \mathbb{Z}[x, y, z] \to R$ by

$$f_0 \left(\sum_{i,j,k \geq 0} m_{ijk} x^i y^j z^k \right) = \sum_{i,j,k \geq 0} m_{ijk} a^i b^j c^k.$$

Now let I be the ideal in $\mathbb{Z}[x, y, z]$ generated by $x^3 + y^3 + z^3 - 1$. As $a^3 + b^3 + c^3 = 1$ we see that $f_0(x^3 + y^3 + z^3 - 1) = 0$ and so $f_0(I) = 0$, so we can define $f \colon A \to R$ by $f(p + I) = f_0(p)$. We define $\alpha_R(a, b, c)$ to be this homomorphism f, so we have a function $\alpha_R \colon F(R) \to \mathbf{CRings}(A, R)$. In the opposite direction, given a homomorphism $f \colon A \to R$ we put $\beta_R(f) = (f(x + I), f(y + I), f(z + I)) \in F(R)$. One can check that α and β are natural and mutually inverse, so A represents F. The proof that B represents G is similar.

Theorem 36.124 (The Yoneda Lemma). *Let $\mathcal{C}$ be a category, let $F \colon \mathcal{C} \to$ **Sets** be a functor, and let A be an object of $\mathcal{C}$. Then natural transformations from YA to F biject with elements of FA.*

A more precise and detailed statement is embedded in the proof.

Proof. Let $\alpha \colon YA \to F$ be a natural map. This means that for every object $P \in \mathcal{C}$ we have a function $\alpha_P \colon (YA)(P) = \mathcal{C}(A, P) \to FP$, and for every morphism $u \colon P \to Q$ the diagram

$$\begin{tikzcd} \mathcal{C}(A,P) \arrow[r, "\alpha_P"] \arrow[d, "u_*"'] & FP \arrow[d, "Fu"] \\ \mathcal{C}(A,Q) \arrow[r, "\alpha_Q"'] & FQ. \end{tikzcd}$$

commutes. In particular, we can take $P = A$ to get a function $\alpha_A\colon \mathcal{C}(A,A) \to F(A)$ and thus an element $x = \alpha_A(1_A) \in F(A)$. Now consider an element $u \in (YA)(Q)$, or in other words a morphism $u\colon A \to Q$. We can again take $P = A$ above and chase the element 1_A around the diagram to get $\alpha_Q(u_*(1_A)) = (Fu)(\alpha_A(1_A))$, or in other words $\alpha_Q(u) = (Fu)(x)$. This shows that α is determined by x.

Suppose instead that we start with an element $x \in FA$. We can then define $\phi_P\colon (YA)(P) = \mathcal{C}(A,P) \to F(P)$ by $\phi_P(u) = (Fu)(x)$. We claim that this gives a natural map $\phi\colon YA \to F$, or equivalently that for every morphism $u\colon P \to Q$ the diagram

$$\begin{tikzcd} \mathcal{C}(A,P) \arrow[r, "\phi_P"] \arrow[d, "u_*"'] & FP \arrow[d, "Fu"] \\ \mathcal{C}(A,Q) \arrow[r, "\phi_Q"'] & FQ. \end{tikzcd}$$

commutes, or that for every $v\colon A \to P$ we have $(Fu)(\phi_P(v)) = \phi_Q(uv)$. Note that $(Fu) \circ (Fv) = F(uv)$, so from the definitions we have

$$(Fu)(\phi_P(v)) = (Fu)((Fv)(x)) = F(uv)(x) = \phi_Q(uv)$$

as required. It is also clear that $\phi_A(1_A) = F(1_A)(x) = x$, so our two constructions are mutually inverse, as claimed. □

Definition 36.125. [`defn-universal-example`]

We say that an element $x \in F(A)$ is a *universal example* for F if the associated natural transformation $YA \to F$ is an isomorphism (so for each P and $a \in FP$ there is a unique morphism $u\colon A \to P$ with $u_*(x) = a$).

Example 36.126. In Example 36.122, the element $1 \in G(\mathbb{Z}/2)$ is a universal example for G, and the element $((1,0),(0,1)) \in H(\mathbb{Z}\oplus\mathbb{Z})$ is a universal example for H. In Example 36.123, we let $\overline{x}$, $\overline{y}$ and $\overline{z}$ denote the images of x, y and z in the quotient ring $A = \mathbb{Z}[x,y,z]/(x^3+y^3+z^3-1)$; we then find that the triple $(\overline{x},\overline{y},\overline{z}) \in F(A)$. Similarly, the image of x in $B = \mathbb{Z}[x,y]/(xy-1)$ gives a universal example for G.

Example 36.127. [`eg-yoneda-units`]

As in Example 36.123, we define a functor $G\colon$ **CRings** $\to$ **Sets** by

$$G(R) = \{a \in R \;:\; a \text{ is invertible }\}.$$

For each $n \in \mathbb{Z}$ we can define a natural map $\alpha_n\colon G \to G$ by $\alpha_n(a) = a^n$, and another natural map $\beta_n\colon G \to G$ by $\beta_n(a) = -a^n$. We claim that these are the only natural maps from G to G. To see this, recall that G is represented by the ring $B = \mathbb{Z}[x,y]/(xy-1)$, so natural maps from G to G biject with elements of $G(B)$. Now B can also be described as the ring of Laurent polynomials $f(x) = \sum_{k=-N}^{N} c_k x^k$ with integer coefficients. If $f(x)$ is invertible then there must be another Laurent polynomial $g(x) = \sum_{m=-M}^{M} d_m x^m$ with $f(x)g(x) = 1$. By considering the highest and lowest terms in $f(x)$, $g(x)$ and $f(x)g(x)$ we see that $f(x)$ must consist of a single term (say $f(x) = c_k t^k$) and $g(x)$ must also consist of a single term (say $g(x) = d_m x^m$). The relation $f(x)g(x) = 1$ then reduces to $m = -k$ and $c_k d_m = 1$ but $c_k, d_m \in \mathbb{Z}$ so $c_k = d_m = \pm 1$. We therefore see that $f(x)$ is either x^k (corresponding to α_k) or $-x^k$ (corresponding to β_k) and the claim follows.

Example 36.128. [`eg-yoneda-idempotents`]

Now define another functor $E\colon$ **CRings** $\to$ **Sets** by

$$E(R) = \{ \text{ idempotents in } R\} = \{e \in R \;:\; e^2 = e\}.$$

Note that if $e \in E(R)$ then $(1-2e)^2 = 1 - 4e + 4e^2 = 1$, so $1 - 2e \in G(R)$, and similarly $2e - 1 \in G(R)$. We thus have four natural maps $E \to G$ given by

$$\phi_0(e) = 1 \qquad \phi_1(e) = -1 \qquad \phi_2(e) = 1 - 2e \qquad \phi_3(e) = 2e - 1.$$

We claim that there are no others. To see this, note that E is represented by the ring $C = \mathbb{Z}[x]/(x^2 - x)$. There is a unique ring map $\rho\colon C \to \mathbb{Z}\times\mathbb{Z}$ given by $\rho(x) = (1,0)$, and one can check that this is an isomorphism, with $\rho^{-1}(n,m) = nx + m(1-x)$. It follows that natural transformations from E to G biject with the set

$$G(\mathbb{Z}\times\mathbb{Z}) = \{(1,1),(-1,-1),(-1,1),(1,-1)\}.$$

By chasing through the various identifications, we find that these four elements correspond to ϕ_0, ϕ_1, ϕ_2 and ϕ_3 respectively.

PROPOSITION 36.129. [`prop-auto-functor`]
Let $G\colon \mathcal{D}\to\mathcal{C}$ be a functor. Suppose we have an object $F(X)\in\mathcal{D}$ for each $X\in\mathcal{C}$, and a system of bijections $\rho_{XU}\colon \mathcal{C}(X,G(U))\to\mathcal{D}(F(X),U)$ for all $U\in\mathcal{D}$. Suppose that these are natural in U, in the sense that for all $p\colon X\to G(U)$ and $m\colon U\to V$ we have

$$\rho_{XV}(X\xrightarrow{p}G(U)\xrightarrow{G(m)}G(V)) = (F(X)\xrightarrow{\rho_{XU}(p)}U\xrightarrow{m}V).$$

Then there is a unique compatible way to define F on morphisms, so that F becomes a functor and the maps ρ_{UV} give an adjunction between F and G.

PROOF. **Prove this.** □

36.10. Regular monomorphisms and epimorphisms. Write this

36.11. Cartesian closure. Write this

36.12. Monoidal structures. Write this

Bibliography

[1] Peter T. Johnstone, *Stone spaces*, Cambridge Studies in Advanced Mathematics, vol. 3, Cambridge University Press, 1982.
[2] David Pincus and Robert M. Solovay, *Definability of measures and ultrafilters*, J. Symbolic Logic **42** (1977), no. 2, 179–190. MR0480028 (58 #227)
[3] Martin Väth, *Nonstandard analysis*, Birkhäuser Verlag, Basel, 2007. MR2265332 (2007g:03087)