Topology for the working mathematician
(Working title)

Michael Müger

11.07.2018, 11:58

Contents

0 Preface **11**

Part I: Fundamentals **14**

1 Introduction **15**

2 Basic notions of point-set topology **19**

2.1 Metric spaces: A reminder . . . 19
2.1.1 Pseudometrics. Metrics. Norms . . . 19
2.1.2 Convergence in metric spaces. Closure. Diameter . . . 21
2.1.3 Continuous functions between metric spaces . . . 22
2.2 From metrics to topologies . . . 25
2.2.1 The metric topology . . . 25
2.2.2 Equivalence of metrics . . . 27
2.3 Some standard topologies . . . 28
2.4 Closed and clopen subsets. Connectedness . . . 30
2.5 The separation axioms T_1 and T_2 . . . 31
2.6 Interior. Closure. Boundary . . . 33
2.7 Neighborhoods. Density . . . 37
2.7.1 Neighborhoods. Topologies from neighborhoods . . . 37
2.7.2 Dense subsets. Nowhere dense subsets . . . 39
2.7.3 ⋆ Accumulation points. Perfect sets. Scattered spaces . . . 40
2.8 Some more exotic types of spaces . . . 41
2.8.1 ⋆ Irreducible spaces . . . 41
2.8.2 T_0-spaces . . . 42
2.8.3 ⋆⋆ Alexandrov or Smallest Neighborhood Spaces . . . 43

3 Metric spaces: Completeness and its applications **47**

3.1 Completeness . . . 47
3.2 Completions . . . 50
3.3 Baire's theorem for complete metric spaces. G_δ-sets . . . 53
3.3.1 Baire's theorem . . . 53
3.3.2 Baire's theorem and the choice axioms . . . 54
3.3.3 G_δ and F_σ sets . . . 55
3.3.4 Application: Croft's Lemma . . . 56
3.3.5 Application: A dense G_δ-set of nowhere differentiable functions . . . 57
3.4 ⋆ Oscillation. Extending continuous functions. Complete metrizability . . . 58

3.4.1 Oscillation and sets of continuity 58
3.4.2 Extending continuous functions between metric spaces 59
3.4.3 More on complete metrizability 61

4 More basic topology **63**
4.1 Bases. Second countability. Separability 63
4.1.1 Bases 63
4.1.2 Second countability and separability 63
4.1.3 Spaces from bases 67
4.2 Subbases and order topologies 69
4.2.1 Subbases. Topologies generated by families of subsets 69
4.2.2 Order topologies 69
4.3 Neighborhood bases. First countability 70

5 Convergence and continuity **75**
5.1 Convergence in topological spaces: Sequences, nets, filters 75
5.1.1 Sequences 75
5.1.2 Nets 77
5.1.3 $\star$ Filters 81
5.2 Continuous, open, closed functions. Homeomorphisms 83
5.2.1 Continuity at a point 83
5.2.2 Continuous functions. The category $\mathcal{TOP}$ 85
5.2.3 Homeomorphisms. Open and closed functions 87
5.2.4 $\star$ Semicontinuous $\mathbb{R}$-valued functions 89

6 New spaces from old **91**
6.1 Initial and final topologies 91
6.1.1 The final topology 91
6.1.2 The initial topology 92
6.2 Subspaces 93
6.3 Direct sums 95
6.4 Quotient spaces 97
6.4.1 Quotient topologies. Quotient maps 97
6.4.2 Quotients by equivalence relations 98
6.4.3 A few geometric applications 101
6.5 Direct products 105
6.5.1 Basics 105
6.5.2 $\star\star$ More on separability and the Souslin property for products 109
6.5.3 Products of metric spaces 111
6.5.4 Joint versus separate continuity 112
6.6 $\star$ Pushouts and Pullback (fiber product) 114

Part II: Covering and Separation axioms (beyond T_2) **117**

7 Compactness and related notions **119**
7.1 Covers. Subcovers. Lindelöf and compact spaces 119
7.2 Compact spaces: Equivalent characterizations 121
7.3 Behavior of compactness and Lindelöf property under constructions 123

7.4 More on compactness . . . 125
7.4.1 More on compactness and subspaces . . . 125
7.4.2 More on compactness and continuity. Quotients and embeddings . . . 127
7.4.3 ⋆ Second countability for images under closed maps . . . 128
7.4.4 ⋆ Extending continuous maps into compact Hausdorff spaces . . . 128
7.5 Compactness of products. Tychonov's theorem . . . 130
7.5.1 The slice lemma. Compactness of finite products . . . 130
7.5.2 Tychonov's theorem . . . 131
7.5.3 Complements . . . 132
7.5.4 ⋆ Ultrafilters. Alternative proof of the subbase lemma . . . 133
7.5.5 ⋆⋆ Third proof of Tychonov, using nets . . . 134
7.5.6 ⋆⋆ Universal nets. Fourth proof of Tychonov . . . 135
7.6 ⋆ Compactness of ordered topological spaces. Supercompact spaces . . . 136
7.7 Compactness: Variations, metric spaces and subsets of $\mathbb{R}^n$ . . . 138
7.7.1 Countable compactness. Weak countable compactness . . . 138
7.7.2 Sequential compactness . . . 140
7.7.3 Compactness of metric spaces I: Equivalences . . . 141
7.7.4 Compactness of metric spaces II: Applications . . . 145
7.7.5 Subsets of $\mathbb{R}^n$ I: Compactness . . . 149
7.7.6 Subsets of $\mathbb{R}^n$ II: Convexity . . . 152
7.7.7 ⋆ Compactness in function spaces I: Ascoli-Arzelà theorems . . . 154
7.8 One-point compactification. Local compactness . . . 156
7.8.1 Compactifications: Definition and Examples . . . 156
7.8.2 ⋆ Compactifications: Some general theory . . . 157
7.8.3 The one-point compactification X_∞ . . . 159
7.8.4 Locally compact spaces . . . 162
7.8.5 Continuous extensions of $f : X \to Y$ to X_∞. Proper maps . . . 167
7.8.6 Functoriality and universal property of X_∞ . . . 169
7.8.7 Further applications of properness . . . 171
7.8.8 Proper metric spaces . . . 174
7.8.9 $C_0(X, \mathbb{F})$ . . . 176
7.9 Compact-open topology. Compactly generated spaces . . . 178
7.9.1 The compact-open topology . . . 178
7.9.2 The exponential law . . . 179
7.9.3 ⋆⋆ Compactly generated spaces = k-spaces . . . 182

8 Stronger separation axioms and their uses 185
8.1 T_3- and T_4-spaces . . . 185
8.1.1 Basics . . . 185
8.1.2 Normality of subspaces. Hereditary normality (T_5) . . . 189
8.1.3 Normality of finite products . . . 191
8.1.4 ⋆ Normality of infinite products . . . 192
8.1.5 Normality and shrinkings of covers . . . 194
8.2 Urysohn's "lemma" and its applications . . . 195
8.2.1 Urysohn's Lemma . . . 195
8.2.2 Perfect normality (T_6) . . . 197
8.2.3 The Tietze-Urysohn extension theorem . . . 199
8.2.4 Urysohn's metrization theorem . . . 201

8.2.5 Partitions of unity. Locally finite families . . . 204
8.3 Completely regular spaces. Stone-Čech compactification . . . 207
8.3.1 $T_{3.5}$: Completely regular spaces . . . 207
8.3.2 Embeddings into products . . . 209
8.3.3 The Stone-Čech compactification . . . 210
8.3.4 Topologies vs. families of pseudometrics . . . 212
8.3.5 Functoriality and universal property of βX . . . 213
8.3.6 Stone-Čech compactification via ultrafilters . . . 217
8.4 ⋆ Applications of the Stone-Čech compactification . . . 217
8.4.1 Čech-completeness . . . 217
8.4.2 Characterization of completely metrizable spaces . . . 219
8.4.3 From Hausdorff compactifications to Proximities . . . 221
8.4.4 From proximities to Hausdorff compactifications . . . 223
8.4.5 Freudenthal compactification of rimcompact spaces . . . 224
8.5 Paracompactness and some of its uses . . . 224
8.5.1 The basic facts . . . 224
8.5.2 Paracompactness and local compactness . . . 229
8.5.3 Paracompactness and normality of product spaces . . . 230
8.5.4 The Nagata-Smirnov metrization theorem . . . 233
8.5.5 Two applications of paracompactness . . . 234
8.5.6 ⋆⋆ A glimpse at generalized metric spaces . . . 235
8.6 Summary of the generalizations of compactness . . . 236

Part III: Connectedness. Steps towards algebraic topology **238**

9 Connectedness: Fundamentals **239**
9.1 Connected spaces and components . . . 239
9.1.1 Basic results . . . 239
9.1.2 Connected components and local connectedness . . . 241
9.1.3 ⋆ Quasi-components . . . 242
9.2 Connectedness of Euclidean spaces . . . 243
9.2.1 Basics . . . 243
9.2.2 Intermediate value theorem and applications . . . 245
9.2.3 n-th roots in $\mathbb{R}$ and $\mathbb{C}$ . . . 246

10 Higher-dimensional connectedness **249**
10.1 Introduction . . . 249
10.2 The cubical Sperner lemma. Proof of Theorem 10.1.2 . . . 249
10.3 The theorems of Poincaré-Miranda and Brouwer . . . 252
10.4 The dimensions of I^n and $\mathbb{R}^n$ . . . 254
10.5 Other notions of topological dimension . . . 257

11 Highly disconnected spaces. Peano curves **259**
11.1 Highly disconnected spaces . . . 259
11.1.1 Totally disconnected spaces. The connected component functor π_c . . . 259
11.1.2 Totally separated spaces . . . 261
11.1.3 Zero-dimensional spaces. Stone spaces . . . 262

11.1.4 Extremely disconnected spaces. Stonean spaces . . . 263
11.1.5 Infinite products of discrete spaces . . . 264
11.1.6 Zero-dimensional spaces: βX and embeddings $X \hookrightarrow K$ . . . 265
11.1.7 Embeddings $K \hookrightarrow \mathbb{R}$. The Cantor set . . . 267
11.1.8 K maps onto every compact metrizable space . . . 271
11.1.9 Projective limits of topological spaces . . . 272
11.1.10 Stone spaces = profinite spaces. Profinite groups . . . 272
11.1.11 Stone duality. Connections to βX . . . 274
11.2 Peano curves and the problem of dimension . . . 280
11.2.1 Peano curves using the Cantor set and Tietze extension . . . 281
11.2.2 Lebesgue's construction and the Devil's staircase . . . 282
11.2.3 Schöneberg's construction . . . 284
11.2.4 There are no differentiable Peano maps . . . 285
11.2.5 Digression: Sard's theorem and other uses of measure zero . . . 286

12 Paths in topological and metric spaces **289**
12.1 Paths. Path components. The π_0 functor . . . 289
12.2 Path-connectedness vs. connectedness . . . 291
12.3 The Jordan curve theorem . . . 294
12.4 Paths in metric spaces. Geodesic spaces. Length spaces . . . 297
12.4.1 Geodesic metric spaces. Menger's theorem . . . 297
12.4.2 Path lengths . . . 299
12.4.3 Length spaces. The Hopf-Rinow theorem . . . 303

13 Homotopy. The Fundamental Group(oid). Coverings **307**
13.1 Homotopy of maps and spaces. Contractibility . . . 307
13.2 Alternative proof of non-contractibility of S^1. Borsuk-Ulam for S^1, S^2 . . . 311
13.3 Path homotopy. Algebra of paths . . . 312
13.4 The fundamental groupoid functor $\Pi_1 : \mathrm{Top} \to \mathrm{Grpd}$ . . . 314
13.5 Homotopy invariance of π_1 and Π_1 . . . 319
13.5.1 Homotopy invariance of π_1 . . . 320
13.5.2 Homotopy invariance of Π_1 . . . 321
13.6 The Seifert-van Kampen theorem . . . 323
13.7 Covering spaces and applications . . . 323
13.7.1 Covering spaces. Lifting of paths and homotopies . . . 323
13.7.2 Computation of some fundamental groups . . . 326
13.7.3 Properly discontinuous actions of discrete groups . . . 328
13.7.4 The universal covering space . . . 329
13.7.5 Classification of covering spaces . . . 329
13.7.6 Seifert-van Kampen via coverings . . . 329

Appendices **329**

A Background on sets and categories **331**
A.1 Reminder of basic material . . . 331
A.1.1 Notation. Sets. Cartesian products . . . 331
A.1.2 More on Functions . . . 333

A.1.3 Relations . . . 334
A.2 Disjoint unions and direct products . . . 336
A.2.1 Disjoint unions . . . 336
A.2.2 Arbitrary direct products . . . 337
A.3 Choice axioms and their equivalents . . . 338
A.3.1 Three formulations of the Axiom of Choice . . . 338
A.3.2 Weak versions of the Axiom of Choice . . . 339
A.3.3 Zorn's Lemma . . . 340
A.3.4 Proof of AC $\Rightarrow$ Zorn . . . 342
A.3.5 Some other statements equivalent to AC and Zorn's Lemma . . . 343
A.3.6 Transfinite induction. The Δ-system lemma . . . 344
A.4 Lattices. Boolean algebras . . . 345
A.5 Basic definitions on categories . . . 346

B Spectra of commutative rings. Spectral spaces **351**

C More on Topological Groups **355**
C.1 Basics . . . 355
C.2 Separation axioms and metrizability for topological groups . . . 356

D The fixed point theorems of Banach and Caristi **359**
D.1 Banach's contraction principle and variations . . . 359
D.2 Caristi's fixed point theorem . . . 361
D.3 Application: Iterated function systems . . . 364

E The sequence spaces $\ell^p(S)$ **367**
E.1 Definition and completeness . . . 367
E.2 Separability . . . 368
E.3 Compactness in function spaces II: $\ell^p(S,\mathbb{F})$ and $L^p(\mathbb{R}^n)$ . . . 369

F Between topology and functional analysis: $C_0(X,\mathbb{F})$ **373**
F.1 The Stone-Weierstrass theorem . . . 373
F.1.1 The main result . . . 373
F.1.2 Generalizations . . . 374
F.1.3 Applications . . . 375
F.2 Weak Gelfand duality: Characters of $C_0(X,\mathbb{F})$ . . . 376

G Normed spaces: Applications of topology **379**
G.1 Linear maps: Boundedness and continuity . . . 379
G.2 Applications of Baire's theorem . . . 379
G.2.1 Uniform boundedness . . . 379
G.2.2 Open mappings and closed graphs . . . 381
G.2.3 Continuous functions with divergent Fourier series . . . 383
G.3 Alaoglu's theorem . . . 384
G.4 Schauder's fixed point theorem . . . 385
G.5 A quick look at locally convex vector spaces . . . 386

H Rough schedule of a lecture course (16 weeks) 389
References 391
Index 402

Chapter 0

Preface

THIS IS NOT YET A PREFACE! (IT IS MORE LIKE A SALES PITCH.)

Some distinctive features of our presentation are the following: We

- believe in the unity of mathematics. Therefore, connections to order theory (smallest neighborhood spaces vs. preorders, Stone and Priestley duality), algebra (pure and topological), analysis (real and functional) and (metric) geometry are emphasized rather than downplayed. The boundaries between (general) topology and analysis and metric geometry are impossible to define anyway.
- believe in lemma extraction (as clearly do some of the authors cited below): Where the same argument is used repeatedly, it is split off as a lemma. Example: Lemma 7.4.2 is deduced from Lemma 7.4.1 which also immediately gives that compact Hausdorff spaces are regular.
- did our best to let no single proof be longer than a page.
- avoid ordinal numbers and topological examples based on them.
- (re)state results in categorical language, where appropriate, hopefully without overdoing it.
- resist the temptation of including counterexamples for all non-implications. (E.g., we don't prove $T_3 \not\Rightarrow T_{3.5}$.) But we do provide counterexamples where they seem helpful for avoiding misconceptions, e.g. the Arens-Fort space proving that a topology on a countable set need not be first countable.
- give four proofs of Tychonov's and two of Alaoglu's theorem and discuss various ramifications (Kelley's converse, the ultrafilter lemma).
- give three approaches to constructing the Stone-Čech compactification: Embedding into cubes, ultrafilters, characters on $C_b(X)$.
- prove Ekeland's variational principle and Caristi's fixed point theorem.
- discuss the basics of geodesic and length spaces and prove the Hopf-Rinow-Cohn-Vossen theorem.
- give a more thorough discussion of the Lebesgue property of metric spaces than is usual. (The only exception seems to be the recent book [183] of Naimpally and Warrack.)
- define proximity spaces, but use them only for the classification of Hausdorff compactifications.

- discuss Stone spaces in relation to profinite spaces (and groups) and Stone duality, including connections to the Stone-Čech compactification.

- believe that defining only the fundamental group, but not the fundamental groupoid is quite outdated and misleading. After all, paths do not need to be loops in order to be composed. Other textbook authors increasingly seem to think the same, cf. e.g. [154, 37, 226].

- give two proofs of van Kampen's theorem: for the fundamental groupoid by manipulating paths, and via covering space theory (but only for the fundamental group version, to keep things simple).

- prove the basic results on separation separation axioms and metrizability for topological groups (rarely found in books) and topological vector spaces.

- define the Gromov-Hausdorff metric and study iterated function systems, complementing the discussion of the Cantor set.

- present, deviating from common practice (among the very few exceptions there are [25, 192]), the most basic results from topological algebra, concerning separation axioms and metrizability for topological groups and local convexity and normability of topological vector spaces. We also discuss the standard applications of topological ideas to topological algebra: the uses of Baire's theorem, weak compactness (Alaoglu), Schauder's fixed point theorem. However, this not being an introduction to abstract harmonic analysis or functional analysis, we do not include results, even fundamental ones, if they do not relate closely to point-set topology, e.g. the Hahn-Banach theorem.

- give two proofs of the Banach-Steinhaus theorem: A strong version via Baire's theorem and an 'elementary' proof.

- The biggest omission probably is the theory of uniform spaces. Since they have very few uses outside topology proper, the author is not convinced that they belong to the core that 'everyone' should know. Also, there are many good expositions of the subject to which we would have nothing to add. Cf. [237, 75, 130, 126], etc.

We include some (relatively) new approaches that we consider real gems:

- Grabiner's more conceptual treatment of the Tietze-Urysohn theorem, using an approximation lemma that also applies to the open mapping theorem.

- McMaster's very short construction (as a quotient of βX) of the Hausdorff compactification corresponding to a proximity. This leads to a quick construction of the Freudenthal compactification.

- Maehara's short proof of Jordan's curve theorem.

- Kulpa's proof of Brouwer's fixed point theorem via the Poincaré-Miranda theorem, using a cubical Sperner lemma. This must surely be the shortest self-contained proof in the literature. We emphasize the rôle of higher dimensional connectedness (Theorem 10.1.2) and include a short deduction of the invariance of dimension for the cubes due to van Mill.

- The beautiful approach of Hanche-Olsen and Holden for proving the theorems of Ascoli-Arzelà and Kolmogorov-Riesz-Fréchet (which we prove only for the sequence spaces $\ell^p(S)$).

- Palais' new proof of Banach's contraction principle.
- Penot's proof of Ekeland's variational principle and Ekeland's recent proof of Nash-Moser-Hamilton style results using the latter.
- A very short proof of Menger's theorem, improving on the already short one by Goebel and Kirk.
- The little known proof (found in [64, Chapter 3, §5]) of the fact (used in the standard proof of algebraic closedness of $\mathbb{C}$) that complex numbers have n-th roots.
- A slick topological proof [18] of the continuous dependence of the roots of a complex polynomial on the coefficients.

Acknowledgments. I thank Arnoud v. Rooij for several very attentive readings of the growing manuscript and many constructive comments, including several simplifications of proofs. Thanks are also due to Noud Aldenhoven, Bas Jordans, Nesta van der Schaaf, Carmen van Schoubroeck, Sohail Sheikh, Marco Stevens, Luuk Verhoeven and Julius Witte for spotting errors. The responsibility for the remaining mistakes is, of course, entirely mine.

Part I: Fundamentals

Chapter 1

Introduction

Virtually everyone writing about topology feels compelled to begin with the statement that “topology is geometry without distance” or “topology is rubber-sheet geometry”, i.e. the branch of geometry where there is no difference between a donut and a cup (in the sense that the two can be continuously deformed into each other without cutting or gluing). While there is some truth[1] to these ‘definitions’, they leave much to be desired: On the one hand, the study of metric spaces belongs to topology, even though they do have a notion of distance. On the other hand, the above definitions are purely negative and clearly insufficient as a foundation for a rigorous theory.

A preliminary positive definition might be as follows: Topology is concerned with the study of topological spaces, where a topological space is a set X equipped with some additional structure that allows to determine whether (i) a sequence (or something more general) with values in X **converges** and (ii) whether a function $f : X \to Y$ between two topological spaces X, Y is **continuous**.

The above actually defines ‘General Topology’, also called ‘set-theoretic topology’ or ‘point-set topology’, which provides the foundations for all branches of topology. It only uses some set theory and logic, yet proves some non-trivial theorems. Building upon general topology, one has several other branches:

- Algebraic Topology uses tools from algebra to study and (partially) classify topological spaces.
- Geometric and Differential Topology study spaces that ‘locally look like $\mathbb{R}^n$’, the difference roughly being that differential topology uses tools from analysis, whereas geometric topology doesn’t.
- Topological Algebra is concerned with algebraic structures that at the same time have a topology such that the algebraic operations are continuous. Example: $\mathbb{R}$ with the usual topology is a topological field. (Topological algebra is not considered a branch of topology. Nevertheless we will look a bit at topological groups and vector spaces.)

Figure 1.1 attempts to illustrate the position of the branches of topology (general, geometric, differential, algebraic) in the fabric of mathematics. (Arrows show dependencies, dotted lines weaker connections.) As one sees, even pure algebra uses notions of general topology, e.g. via the Krull topology in the theory of infinite Galois extensions or the Jacobson topology on the set of ideals of an associative algebra.

One may certainly say that (general) topology is the language of a very large part of mathematics. But it is more than a language since it has its share of non-trivial theorems, some of which we

[1] See http://commons.wikimedia.org/wiki/File:Mug_and_Torus_morph.gif

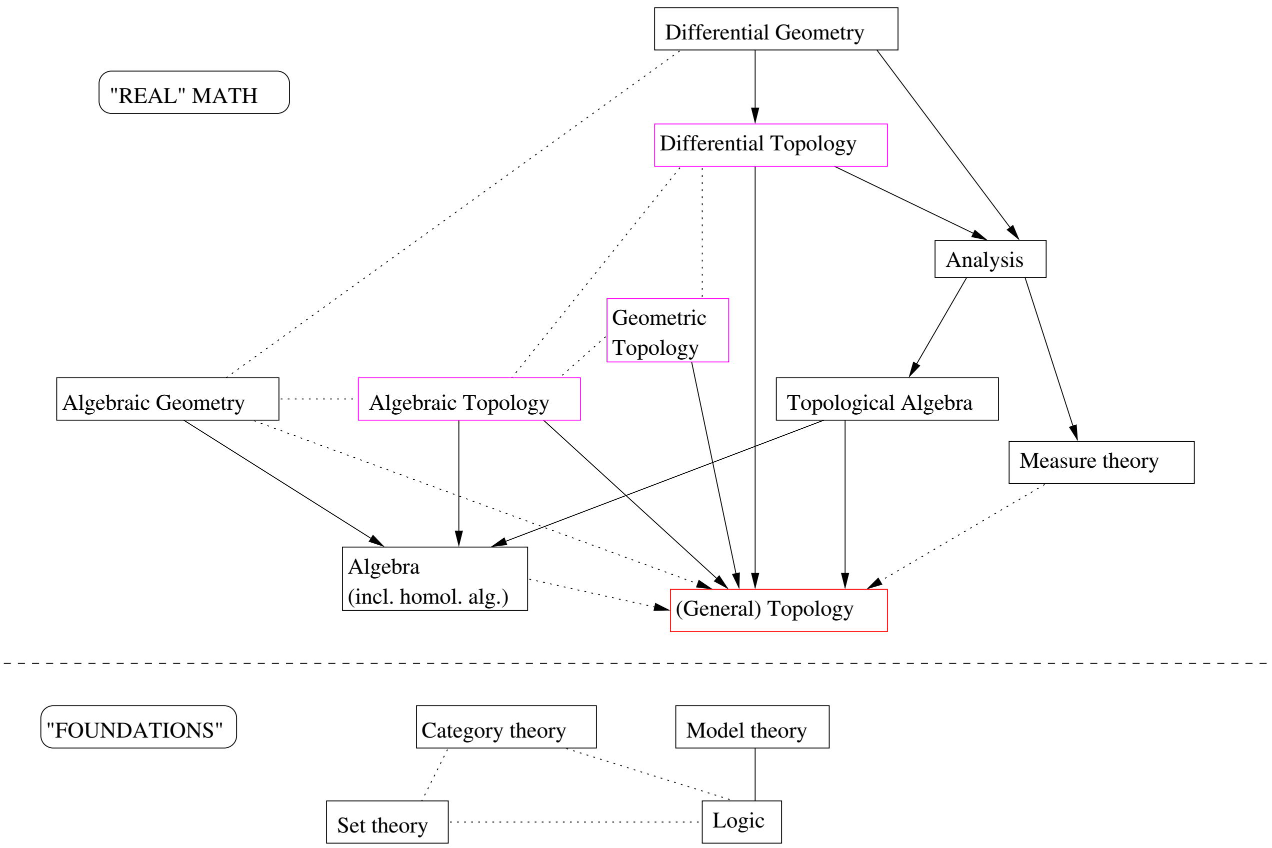


Figure 1.1: The branches of Topology in mathematics

will prove: Tychonov's theorem, the Nagata-Smirnov metrization theorem, Brouwer's fixed point theorem, etc.

General topology is a very young subject which started for real only in the 20th century with the work of Fréchet[2] and Hausdorff[3]. (Of course there were many precursors.) For more on its history see [173, 164, 132, 11].

By comparison, algebraic topology is much older. (While this may seem paradoxical, it parallels the development of analysis, whose foundations were only understood at a fairly late stage.) Its roots lie in work of L. Euler and C. F. Gauss[4], but it really took off with B. Riemann after 1850. In the beginning, the subject was called 'analysis situs'. The modern term 'topology' was coined by J. B. Listing[5] in 1847. For the history of algebraic topology (which was called combinatorial topology in the early days) cf. [191, 57].

A serious problem for the teaching of topology is that the division of topology in general and algebraic topology[6] has only become more pronounced since the early days, as a look at [152] and

[2]Maurice Fréchet (1878-1973), French mathematician. Introduced metric spaces in his 1906 PhD thesis [82].

[3]Felix Hausdorff (1868-1942), German mathematician. One of the founding fathers of general topology. His book [109] was extremely influential.

[4]Leonhard Euler (1707-1783), Carl Friedrich Gauss (1777-1855).

[5]Georg Friedrich Bernhard Riemann (1826-1866), Johann Benedict Listing (1808-1882).

[6]Alexandroff-Hopf (Topologie I, 1935): Die and und für sich schwierige Aufgabe, eine solche Darstellung eines immerhin jungen Zweiges der mathematischen Wissenschaft zu geben, wird im Falle der Topologie dadurch besonders erschwert, daß die Entwicklung der Topologie in zwei voneinander gänzlich getrennten Richtungen vor sich gegangen ist: In der algebraisch-kombinatorischen und in der mengentheoretischen – von denen jede in mehrere weitere Zweige zerfällt, welche nur lose miteinander zusammenhängen.

Alexandroff (Einfachste Grundbegriffe der Topologie, 1932): Die weitere Entwicklung der Topologie steht zunächst

[131] shows. Cf. also [133]. Some algebraic topologists consider a short appendix on general topology sufficient for most purposes (for an exception see [28]), but this does no justice to the needs of analysis, geometric topology, algebraic geometry and other fields. In the present introduction the focus therefore is on general topology, but in Part III we gradually switch to more algebraic-topological matters.

In a sense, General Topology simply is concerned with sets and certain families of subsets of them and functions between them. (In fact, for a short period no distinction was made between set theory and general topology, cf. [109], but this is no more the case.) This means that the only prerequisite is a reasonable knowledge of some basic (naive) set theory and elementary logic. At least in principle, the subject could therefore be taught and studied in the second semester of a math programme. But such an approach does not seem very reasonable, and the author is not aware of any institution where it is pursued. Usually the student encounters metric spaces during her study of calculus/analysis. Already functions of one real variable motivate the introduction of (norms and) metrics, namely in the guise of the uniform distance between bounded functions (and the L^p-norms). Analysis in $1 < d < \infty$ dimensions naturally involves various metrics since there is no really distinguished metric on $\mathbb{R}^d$. We will therefore also assume some basic familiarity with metric spaces, including the concepts of Cauchy sequence and completeness. The material in [201] or [224] is more than sufficient. Nevertheless, we prove the main results, in particular uniqueness and existence of completions. No prior exposure to the notion of topological space is assumed.

Sections marked with a star ($\star$) can be omitted on first reading, but their results will be used at some later point. Two stars ($\star\star$) are used to mark optional sections that are not referred to later.

Many exercises are spread throughout the text, and many results proven there are used freely afterwards.

im Zeichen einer scharfen Trennung der mengentheoretischen und der kombinatorischen Methoden: Die kombinatorische Topologie wollte sehr bald von keiner geometrischen Realität, außer der, die sie im kombinatorischen Schema selbst (und seinen Unterteilungen) zu haben glaubte, etwas wissen, während die mengentheoretische Richtung derselben Gefahr der vollen Isolation von der übrigen Mathematik auf dem Wege der Auftürmung von immer spezielleren Fragestellungen und immer komplizierteren Beispielen entgegenlief.

Chapter 2

Basic notions of point-set topology

2.1 Metric spaces: A reminder

2.1.1 Pseudometrics. Metrics. Norms

As mentioned in the introduction, some previous exposure to metric spaces is assumed. Here we briefly recall the most important facts, including proofs, in order to establish the terminology.

Definition 2.1.1 *If X is a set, a pseudometric on X is a map $d : X \times X \to \mathbb{R}_{\geq 0}$ such that*

(i) $d(x, y) = d(y, x)\ \forall x, y$. *(Symmetry)*

(ii) $d(x, z) \leq d(x, y) + d(y, z)\ \forall x, y, z$. *(Triangle inequality)*

(iii) $d(x, x) = 0\ \forall x$.

A metric is pseudometric d such that $x \neq y \Rightarrow d(x, y) \neq 0$.

Remark 2.1.2 Obviously every statement that holds for pseudometrics in particular holds for metrics. The converse is not at all true. (On the other hand, when we state a result only for metrics, this should not automatically be interpreted as saying that the generalization to pseudometrics is false.) □

Pseudometrics are easy to come by:

Exercise 2.1.3 Let $f : X \to \mathbb{R}$ be a function. Prove:

(i) $d(x, y) = |f(x) - f(y)|$ is a pseudometric.

(ii) d is a metric if and only if f is injective.

(Taking $f = \mathrm{id}_{\mathbb{R}}$, we recover the well-known fact that $d(x, y) = |x - y|$ is a metric on $\mathbb{R}$.)

Exercise 2.1.4 For a pseudometric d on X prove:

$$|d(x, z) - d(y, z)| \leq d(x, y) \quad \forall x, y, z \tag{2.1}$$

$$|d(x, y) - d(x_0, y_0)| \leq d(x, x_0) + d(y, y_0) \quad \forall x, y, x_0, y_0, \tag{2.2}$$

$$\sup_{z \in X} |d(x, z) - d(y, z)| = d(x, y) \quad \forall x, y. \tag{2.3}$$

Definition 2.1.5 *A (pseudo)metric space is a pair (X,d), where X is a set and d is a (pseudo) metric on X.*

Remark 2.1.6 A set X with $\#X \geq 2$ admits infinitely many different metrics. (Just consider $d' = \lambda d$, where $\lambda > 0$.) Therefore it is important to make clear which metric is being used. Nevertheless, we occasionally allow ourseleves to write 'Let X be a metric space' when there is no risk of confusion. □

Every pseudometric space has a metric quotient space:

Exercise 2.1.7 Let d be a pseudometric on a set X. Prove:

(i) $x \sim y \Leftrightarrow d(x,y) = 0$ defines an equivalence relation $\sim$ on X.

(ii) Let $p : X \to X/\sim$ be the quotient map arising from $\sim$. Show that there is a unique metric d' on $X/\sim$ such that $d(x,y) = d'(p(x),p(y))\ \forall x,y \in X$.

From now on we will mostly restrict our attention to metrics, but we will occasionally use pseudometrics as a tool. A basic, if rather trivial, example of a metric is given by this:

Example 2.1.8 On any set X, the following defines a metric, the standard discrete metric:

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

(This should not be confused with the weaker notion of 'discrete metric' encountered later.) □

Example 2.1.9 Let p be a prime number. For $0 \neq x \in \mathbb{Q}$ write $x = \frac{r}{s}\, p^{n_p(x)}$, where $n_p(x), r, s \in \mathbb{Z}$ and p divides neither r nor s. This uniquely defines $n_p(x)$. Now

$$\|x\|_p = \begin{cases} p^{-n_p(x)} & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$

is the p-adic norm on $\mathbb{Q}$. It is obvious that $\|x\|_p = 0 \Leftrightarrow x = 0$ and $\|xy\|_p = \|x\|_p\|y\|_p$. With a bit of work one shows $\|x+y\|_p \leq \|x\|_p + \|y\|_p\ \forall x,y$. This implies that $d_p(x,y) = \|x-y\|_p$ is a metric on $\mathbb{Q}$. ($\|\cdot\|_p$ is not to be confused with the norms $\|\cdot\|_p$ on $\mathbb{R}^n$ defined below. In fact, it is not even quite a norm in the sense of the following definition.) □

Definition 2.1.10 *Let V be a vector space over $\mathbb{F} \in \{\mathbb{R},\mathbb{C}\}$. A norm on V is a map $V \to [0,\infty),\ x \mapsto \|x\|$ satisfying*

(i) $\|x\| = 0 \ \Leftrightarrow\ x = 0$. (Faithfulness)

(ii) $\|\lambda x\| = |\lambda|\,\|x\| \quad \forall \lambda \in \mathbb{F},\ x \in V$. (Homogeneity)

(iii) $\|x+y\| \leq \|x\| + \|y\| \quad \forall x,y \in V$. (Triangle inequality or subadditivity)

A normed space is a pair $(V, \|\cdot\|)$, where V is vector space over $\mathbb{F} \in \{\mathbb{R},\mathbb{C}\}$ and $\|\cdot\|$ is a norm on V.

Remark 2.1.11 The generalization of a norm, where one drops the requirement $\|x\| = 0 \Rightarrow x = 0$, is universally called a seminorm. For the sake of consistency, one should thus speak of 'semimetrics' instead of pseudometrics, but only a minority of authors does this. □

Lemma 2.1.12 *If V is a vector space over $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and $\|\cdot\|$ is a norm on V then $d_{\|}(x,y) = \|x-y\|$ defines a metric on V. Thus every normed space is a metric space.*

Proof. We have $d(x,y) = \|x-y\| = \|-(x-y)\| = \|y-x\| = d(y,x)$, and $d(x,y) = 0$ holds if and only if $\|x-y\| = 0$, which is equivalent to $x = y$. Finally, $d(x,z) = \|x-z\| = \|(x-y)+(y-z)\| \leq \|x-y\| + \|y-z\| = d(x,y) + d(y,z)$. ∎

Example 2.1.13 Let $n \in \mathbb{N}$ and $p \in [1,\infty)$. The following are norms on $\mathbb{R}^n$ and $\mathbb{C}^n$:

$$\begin{aligned} \|x\|_\infty &= \max_{i \in \{1,\ldots,n\}} |x_i|, \\ \|x\|_p &= \left(\sum_{i=1}^n |x_i|^p \right)^{1/p}. \end{aligned}$$

For $n = 1$ and any $p \in [1,\infty]$, this reduces to $\|x\|_p = |x|$, but for $n \geq 2$ all these norms are different. That $\|\cdot\|_\infty$ and $\|\cdot\|_1$ are norms is easy to see. $\|\cdot\|_2$ is the well known Euclidean norm. For all $1 < p < \infty$, it is immediate that $\|\cdot\|_p$ satisfies requirements (i) and (ii), while (iii) is the inequality of Minkowski

$$\|x+y\|_p \leq \|x\|_p + \|y\|_p. \tag{2.4}$$

This is proven using the Hölder inequality

$$\left| \sum_{i=1}^n x_i y_i \right| \leq \|x\|_p \cdot \|y\|_q, \tag{2.5}$$

valid when $\frac{1}{p} + \frac{1}{q} = 1$. This is surely well known for $p = q = 2$, in which case (2.5) is known as the Cauchy-Schwarz inequality. For the general proof, cf. e.g. [203].

The infinite dimensional generalization $\ell^p(S)$ of $(\mathbb{R}^n, \|\cdot\|_p)$ is considered in Section E. □

2.1.2 Convergence in metric spaces. Closure. Diameter

An important reason for introducing metrics is to be able to define the notions of convergence and continuity:

Definition 2.1.14 *A sequence in a set X is a map $\mathbb{N} \to X$, $n \mapsto x_n$. We will usually denote the sequence by $\{x_n\}_{n \in \mathbb{N}}$ or just $\{x_n\}$.*

Definition 2.1.15 *A sequence $\{x_n\}$ in a metric space (X,d) converges to $z \in X$, also denoted $x_n \to z$, if for every $\varepsilon > 0$ there is $N \in \mathbb{N}$ such that $n \geq N \Rightarrow d(x_n, z) < \varepsilon$.*

If $\{x_n\}$ converges to z then z is the limit of $\{x_n\}$. We assume as known (but will later reprove in a more general setting) that a sequence in a metric space has at most one limit, justifying the use of 'the'.

Lemma 2.1.16 *Let (X,d) be a metric space and $Y \subseteq X$. Then for a point $x \in X$, the following are equivalent:*

(i) For every $\varepsilon > 0$ there is $y \in Y$ such that $d(x,y) < \varepsilon$.

(ii) There is a sequence $\{y_n\}$ in Y such that $y_n \to x$.

The set of points satisfying these (equivalent) conditions is called the closure $\overline{Y}$ *of* Y. *It satisfies* $Y \subseteq \overline{Y} = \overline{\overline{Y}}$. *A subset* $Y \subseteq X$ *is called closed if* $Y = \overline{Y}$.

Proof. (ii)$\Rightarrow$(i) This is obvious since $y_n \to x$ is the same as $d(y_n, x) \to 0$. (i)$\Rightarrow$(ii) For every $n \in \mathbb{N}$, use (i) to choose $y_n \in Y$ such that $d(y_n, x) < 1/n$. Clearly $y_n \to x$. (This of course uses the axiom of countable choice, cf. Section A.3.2.)

It is clear that $Y \subseteq \overline{Y}$. Finally, $x \in \overline{\overline{Y}}$ means that for every $\varepsilon > 0$ there is a point $y \in \overline{Y}$ with $d(x, y) < \varepsilon$. Since $y \in \overline{Y}$, there is a $z \in Y$ such that $d(y, z) < \varepsilon$. By the triangle inequality we have $d(x, z) < 2\varepsilon$, and since ε was arbitrary we have proven that $x \in \overline{Y}$. Thus $\overline{\overline{Y}} = \overline{Y}$. ■

Definition 2.1.17 *If* (X, d) *is a metric space then the diameter of a subset* $Y \subseteq X$ *is defined by* $\mathrm{diam}(Y) = \sup_{x,y \in Y} d(x, y) \in [0, \infty]$ *with the understanding that* $\mathrm{diam}(\emptyset) = 0$.

A subset Y *of a metric space* (X, d) *is called bounded if* $\mathrm{diam}(Y) < \infty$.

Exercise 2.1.18 For $Y \subseteq (X, d)$, prove $\mathrm{diam}(\overline{Y}) = \mathrm{diam}(Y)$.

Definition 2.1.19 *If* (X, d) *is a metric space,* $A, B \subseteq X$ *are non-empty and* $x \in X$, *define*

$$\begin{aligned} \mathrm{dist}(A, B) &= \inf_{\substack{a \in A \\ b \in B}} d(a, b), \\ \mathrm{dist}(x, A) &= \mathrm{dist}(\{x\}, A) = \inf_{a \in A} d(x, a). \end{aligned}$$

(If A *or* B *is empty, we leave the distance undefined.)*

Exercise 2.1.20 Let (X, d) be a metric space and $A, B \subseteq X$.

(i) Prove that $|\mathrm{dist}(x, A) - \mathrm{dist}(y, A)| \leq d(x, y)$.

(ii) Prove that $\mathrm{dist}(x, A) = 0$ if and only if $x \in \overline{A}$.

(iii) Prove that A is closed if and only if $\mathrm{dist}(x, A) = 0$ implies $x \in A$.

(iv) Prove that $\overline{A} \cap \overline{B} \neq \emptyset \;\Rightarrow\; \mathrm{dist}(A, B) = 0$.

(v) For $X = \mathbb{R}$ with $d(x, y) = |x - y|$, give examples of non-empty closed sets $A, B \subseteq X$ with $\mathrm{dist}(A, B) = 0$ but $A \cap B = \emptyset$. (Thus the converse of (iv) is not true in general!)

Remark: With Definition 2.1.21, (i) directly gives that $x \mapsto \mathrm{dist}(x, A)$ is continuous.

2.1.3 Continuous functions between metric spaces

Definition 2.1.21 *Let* $(X, d), (X', d')$ *be metric spaces and* $f : X \to X'$ *a function.*

- f *is called continuous at* $x \in X$ *if for every* $\varepsilon > 0$ *there is* $\delta > 0$ *such that* $d(x, y) < \delta \Rightarrow d'(f(x), f(y)) < \varepsilon$.
- f *is called continuous if it is continuous at each* $x \in X$.
- f *is called a homeomorphism if it is a bijection, continuous, and the inverse* $f^{-1} : X' \to X$ *is continuous.*
- f *is called an isometry if* $d'(f(x), f(y)) = d(x, y)$ $\forall x, y \in X$.

- *f is called <u>bounded</u> if $f(X) \subseteq Y$ is bounded w.r.t. d'. (Equivalently there is an $R \in [0, \infty)$ such that $d'(f(x), f(y)) \leq R \ \forall x, y \in X$.)*

  (This actually does not refer to d, thus it makes sense for every $f : X \to (X', d')$.)

Remark 2.1.22 1. Obviously, an isometry is both continuous and injective.

2. Since the inverse function of a bijective isometry again is an isometry, and thus continuous, we have

$$\text{isometric bijection} \Rightarrow \text{homeomorphism} \Rightarrow \text{continuous bijection}.$$

As the following examples show, the converse implications are not true!

3. If $d(x, y) = |x - y|$ then $(\mathbb{R}, d_{\text{disc}}) \to (\mathbb{R}, d)$, $x \mapsto x$ is a continuous bijection, but not a homeomorphism.

4. If (X, d) is a metric space and $d'(x, y) = 2d(x, y)$ then $(X, d) \to (X, d')$, $x \mapsto x$ is a homeomorphism, but not an isometry.

5. A less trivial example, used much later: Let $X = (-1, 1)$ and $d(x, y) = |x - y|$. Then $f : (\mathbb{R}, d) \to (X, d)$, $x \mapsto \frac{x}{1+|x|}$ is a continuous and has $g : y \mapsto \frac{y}{1-|y|}$ as continuous inverse map. Thus f, g are homeomorphisms, but clearly not isometries. □

The connection between the notions of continuity and convergence is provided by:

Lemma 2.1.23 *Let $(X, d), (X', d')$ be metric spaces and $f : X \to X'$ a function. Then the following are equivalent (t.f.a.e.):*

(i) f is continuous at $x \in X$.

(ii) For every sequence $\{x_n\}$ in X that converges to x, the sequence $\{f(x_n)\}$ in X' converges to $f(x)$. ('f is sequentially continuous'.)

Proof. (i)⇒(ii) Let $\{x_n\}$ be a sequence such that $x_n \to x$, and let $\varepsilon > 0$. Since f is continuous at x, there is a $\delta > 0$ such that $d(x, y) < \delta \Rightarrow d(f(x), f(y)) < \varepsilon$. Since $x_n \to x$, there is $N \in \mathbb{N}$ such that $n \geq N$ implies $d(x_n, x) < \delta$. But then $d(f(x_n), f(x)) < \varepsilon \ \forall n \geq N$. This proves $f(x_n) \to f(x)$.

(ii)⇒(i) Assume that f is not continuous at $x \in X$. Now, $\neg(\forall\varepsilon\exists\delta\forall y \cdots) = \exists\varepsilon\forall\delta\exists y \neg \cdots$. This means that there is $\varepsilon > 0$ such that for every $\delta > 0$ there is a $y \in X$ with $d(x, y) < \delta$ such that $d(f(x), f(y)) \geq \varepsilon$. Thus we can choose a sequence $\{x_n\}$ in X such that $d(x, x_n) < 1/n$ and $d(f(x), f(x_n)) \geq \varepsilon$ for all $n \in \mathbb{N}$. Now clearly $x_n \to x$, but $f(x_n)$ does not converge to $f(x)$. This contradicts the assumption that (ii) is true. ■

Definition 2.1.24 *The set of all bounded / continuous / bounded and continuous functions $f : (X, d) \to (X', d')$ are denoted $B((X, d), (X', d'))$ / $C((X, d), (X', d'))$ / $C_b((X, d), (X', d'))$, respectively. (In practice, we may write $B(X, X'), C(X, X'), C_b(X, X')$.)*

Proposition 2.1.25 (Spaces of bounded functions) *Let $(X, d), (Y, d')$ be metric spaces. Define*

$$D(f, g) = \sup_{x \in X} d'(f(x), g(x)). \tag{2.6}$$

(i) The equation (2.6) defines a metric D on $B(X, Y)$.

(ii) $C_b(X, Y) := C(X, Y) \cap B(X, Y) \subseteq B(X, Y)$ is closed w.r.t. D.

Proof. (i) Let $f, g \in B(X, Y)$. For any $x_0 \in X$, we have

$$\begin{aligned} D(f,g) &= \sup_{x\in X} d'(f(x),g(x)) \leq \sup_{x\in X}\,[d'(f(x),f(x_0)) + d'(f(x_0),g(x_0)) + d'(g(x_0),g(x))] \\ &\leq d'(f(x_0),g(x_0)) + \sup_{x\in X} d'(f(x),f(x_0)) + \sup_{x\in X} d'(g(x),g(x_0)) \\ &\leq d'(f(x_0),g(x_0)) + \operatorname{diam}(f(X)) + \operatorname{diam}(g(X)) < \infty, \end{aligned}$$

thus D is finite on $B(X, Y)$. It is clear that D is symmetric and that $D(f, g) = 0 \Leftrightarrow f = g$. Furthermore,

$$\begin{aligned} D(f,h) &= \sup_{x\in X} d'(f(x),h(x)) \leq \sup_{x\in X}\,[d'(f(x),g(x)) + d'(g(x),h(x))] \\ &\leq \sup_{x\in X} d'(f(x),g(x)) + \sup_{x\in X} d'(g(x),h(x)) = D(f,g) + D(g,h). \end{aligned}$$

Thus D satisfies the triangle inequality and is a metric on $B(X, Y)$.

(ii) Let $\{f_n\} \subseteq C_b(X, Y)$ and $g \in B(X, Y)$ such that $D(f_n, g) \to 0$. Let $x \in X$ and $\varepsilon > 0$. Choose N such that $n \geq N \Rightarrow D(f_n, g) < \varepsilon/3$. Since f_N is continuous, we can choose $\delta > 0$ such that $d(x, y) < \delta \Rightarrow d'(f_N(x), f_N(y)) < \varepsilon/3$. Now we have

$$d'(g(x),g(y)) \leq d'(g(x),f_N(x)) + d'(f_N(x),f_N(y)) + d'(f_N(y),g(y)) < \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon,$$

thus g is continuous at x. Since this works for every x, g is continuous. Since $g \in B(X, Y)$ by assymption, we thus have $g \in C_b(X, Y)$. By Lemma 2.1.16, the elements of $B(X, Y)$ that are limits w.r.t. D of elements of $C_b(X, Y)$ constitute the closure $\overline{C_b(X, Y)}$. We thus have shown $\overline{C_b(X, Y)} \subseteq C_b(X, Y)$ and therefore that $C_b(X, Y)) \subseteq B(X, Y)$ is closed. ■

Definition 2.1.26 *If $(X, d), (Y, d')$ are metric spaces and $\{f_n\}$ is a sequence in $B(X, Y)$ or (more often) in $C_b(X, Y)$ such that $D(f_n, g) \to 0$ then f_n converges uniformly to g. And D is called the metric of uniform convergence or simply the uniform metric.*

Remark 2.1.27 1. Statement (ii) of the proposition is just a shorter (and more conceptual) formulation of the fact that the limit of a uniformly convergent sequence of continuous functions is continuous (from which we obtained it). The reader probably knows that pointwise convergence (i.e. $f_n(x) \to g(x)$ for each x) of a sequence of continuous functions does not imply continuity of g. Example: $f_n(x) = \min(1, nx)$ is in $C([0, 1], [0, 1])$ for each $n \in \mathbb{N}$ and converges pointwise to the discontinuous function g, where $g(0) = 0$ and $g(x) = 1$ for all $x > 0$.

2. Part (ii) of the lemma shows that uniformity of the convergence $f_n \to g$ is sufficient for continuity of g. But note that is not necessary. In other words, continuity of g does not imply that the convergence $f_n \to g$ is uniform! Example: The function $f_n : [0, 1] \to [0, 1]$ defined by

$$f_n(x) = \max(0, 1 - n|x - 1/n|) = \begin{cases} nx & x \in [0, 1/n] \\ 1 - n(x - 1/n) & x \in [1/n, 2/n] \\ 0 & x \in [2/n, 1] \end{cases}$$

(draw this) is continuous for each $n \in \mathbb{N}$ and converges pointwise to $g \equiv 0$. But the convergence is not uniform since $D(f_n, g) = 1\ \forall n$.

3. However, if X is sufficiently nice (countably compact, for example a closed bounded subset of $\mathbb{R}^n$) and $\{f_n\} \subseteq C(X, \mathbb{R})$ converges pointwise monotonously, i.e. $f_{n+1}(x) \geq f_n(x)$ for all $x \in X$, $n \in \mathbb{N}$, to a continuous $g \in C(X, \mathbb{R})$ then the convergence is uniform! This is Dini's theorem, which we will prove in Section 7.7.4. □

2.2 From metrics to topologies

2.2.1 The metric topology

Why would anyone want to generalize metric spaces? Here are the most important reasons:

1. (Closure) The category of metric spaces is not closed w.r.t. certain constructions, like direct products (unless countable) and quotients (except under strong assumptions on the equivalence relation), nor does the space $C((X,d),(Y,d'))$ of (not necessarily bounded) functions have an obvious metric (unless X is compact).

2. (Clarity) Most properties of metric spaces (compactness, connectedness,...) can be defined in terms of the topology induced by the metric and therefore depend on the chosen metric only via its equivalence class. Eliminating irrelevant details from the theory actually simplifies it by clarifying the important concepts.

3. (Aesthetic) The definition of metric spaces involves the real numbers (which themselves are a metric space and a rather non-trivial one at that) and therefore is extrinsic. A purely set-theoretic definition seems preferable.

4. (A posteriori) As soon as one has defined a good generalization, usually many examples appear that one could not even have imagined beforehand.

In generalizing metric spaces one certainly still wants to be able to talk about convergence and continuity. Examining Definitions 2.1.15 and 2.1.21, one realizes the centrality of the following two concepts:

Definition 2.2.1 *Let (X,d) be a pseudometric space.*

(i) The <u>open ball</u> of radius r around x is defined by $B(x,r) = \{y \in X \mid d(x,y) < r\}$. (If necessary, we write $B^d(x,r)$ to make clear which metric is being used.)

(ii) We say that $Y \subseteq X$ is <u>open</u> if for every $x \in Y$ there is an $\varepsilon > 0$ such that $B(x,\varepsilon) \subseteq Y$.

The set of open subsets of X is denoted τ_d. (Clearly $\tau_d \subseteq P(X)$.)

Consistency of our language requires that open balls are open:

Exercise 2.2.2 Prove that every $B(x,\varepsilon)$ with $\varepsilon > 0$ is open.

Exercise 2.2.3 Prove that a subset $Y \subseteq (X,d)$ is bounded (in the sense of Definition 2.1.17) if and only if $Y \subseteq B(x,r)$ for some $x \in X$ and $r > 0$.

Lemma 2.2.4 *The open subsets of a pseudometric space (X,d) satisfy the following:*

(i) $\emptyset \in \tau_d,\ X \in \tau_d$.

(ii) If $U_i \in \tau_d$ for every $i \in I$ then $\bigcup_{i\in I} U_i \in \tau_d$.

(iii) If $U_1,\dots,U_n \in \tau_d$ then $\bigcap_{i=1}^n U_i \in \tau_d$.

In words: The empty and the full set are open, <u>arbitrary unions</u> and <u>finite intersections</u> of open sets are open.

Proof. (i) is obvious. (ii) Let $U_i \in \tau\ \forall i \in I$ and $U = \bigcup_i U_i$. Then any $x \in U$ is contained in some U_i. Now there is a $\varepsilon > 0$ such that $B(x,\varepsilon) \subseteq U_i \subseteq U$. Thus $U \in \tau$. (iii) Let $U_i \in \tau$ for all $i = 1,\ldots,n$ and $U = \bigcap_i U_i$. If $x \in U$ then $x \in U_i$ for every $i \in \{1,\ldots,n\}$. Thus there are $\varepsilon_1,\ldots,\varepsilon_n$ such that $B(x,\varepsilon_i) \subseteq U_i$ for all i. With $\varepsilon = \min(\varepsilon_1,\ldots,\varepsilon_n) > 0$ we have $B(x,\varepsilon) \subseteq U_i$ for all i, thus $B(x,\varepsilon) \subseteq U$, implying $U \in \tau$. ■

Remark 2.2.5 We do not require intersections of infinitely many open sets to be open, and in most topological spaces they are not! Consider, e.g., $X = \mathbb{R}$ with $d(x,y) = |x-y|$. Then $U_n = (-1/n, 1/n)$ is open for each $n \in \mathbb{N}$, but $\bigcap_{n=1}^\infty U_n = \{0\}$, which is not open. (See however Section 2.8.3.) □

We will take this as the starting point of the following generalization:

Definition 2.2.6 *If X is a set, a subset $\tau \subseteq P(X)$ is called a topology on X if it has the properties (i)-(iii) of Lemma 2.2.4 (with τ_d replaced by τ). A subset $U \subseteq X$ is called (τ-)open if $U \in \tau$. A topological space is a pair (X,τ), where X is a set and τ is a topology on X.*

Example 2.2.7 The empty space $\emptyset$ has the unique topology $\tau = \{\emptyset\}$. (The axioms only require $\{\emptyset, X\} \subseteq \tau$, but not $\emptyset \neq X$.) Every one-point space $\{x\}$ has the unique topology $\tau = \{\emptyset, \{x\}\}$. Already the two-point space $\{x,y\}$ allows several topologies: $\tau_1 = \{\emptyset, \{x,y\}\}, \tau_2 = \{\emptyset, \{x\}, \{x,y\}\}, \tau_3 = \{\emptyset, \{y\}, \{x,y\}\}, \tau_4 = \{\emptyset, \{x\}, \{y\}, \{x,y\}\}$. □

Definition 2.2.8 *(i) A topology τ_d arising from a metric is called metric topology.*
(ii) A topological space (X,τ) is called metrizable if $\tau = \tau_d$ for some metric d on X.

While the metric spaces are our main motivating example for Definition 2.2.6, there are others that have nothing to do (a priori) with metrics. In fact, we will soon see that not every topological space is metrizable!

Exercise 2.2.9 (Subspaces) (i) Let (X,d) be a metric space and $Y \subseteq X$. If d_Y is the restriction of d to Y, it is clear that (Y,d_Y) is a metric space. If τ and τ_Y denote the topologies on X and Y induced by d and d_Y, respectively, prove

$$\tau_Y = \{U \cap Y \mid U \in \tau\}. \tag{2.7}$$

(ii) Let (X,τ) be a topological space and $Y \subseteq X$. Define $\tau_Y \subseteq P(Y)$ by (2.7). Prove that τ_Y is a topology on Y.

(iii) If (X,τ) is a topological space and $Z \subseteq Y \subseteq X$ then $\tau_Z = (\tau_Y)_Z$.

The topology τ_Y is called the subspace topology (or induced topology, which we tend to avoid), and (Y,τ_Y) is a subspace of (X,τ). (Occasionally it is more convenient to write $\tau \upharpoonright Y$.)

We will have more to say about subspaces in Section 6.2.

Remark 2.2.10 Let (X,τ) be a topological space and $Y \subseteq X$ given the subspace topology. By definition, a set $Z \subseteq Y$ is open (in Y) if and only if it is of the form $U \cap Y$ with $U \in \tau$. Thus unless $Y \subseteq X$ is open, a subset $Z \subseteq Y$ can be open (in Y) without being open in X! Example: If $X = \mathbb{R},\ Y = [0,1], Z = [0,1)$ then Z is open in Y since $Z = Y \cap (-1,1)$, where $(-1,1)$ is open in X. □

A natural modification of Definition 2.2.1(i) leads to closed balls in a metric space:

Exercise 2.2.11 Let (X, d) be a metric space. For $x \in X,\ r > 0$ define closed balls by

$$\overline{B}(x, r) = \{y \in X \mid d(x, y) \leq r\}.$$

Prove:

(i) $\overline{B}(x, r)$ is closed (in the sense of Lemma 2.1.16.)

(ii) The inclusion $\overline{B(x, r)} \subseteq \overline{B}(x, r)$ always holds.

(iii) $\overline{B(x, r)} = \overline{B}(x, r)$ holds for all $x \in X,\ r > 0$ if and only if for all $x, y \in X$ with $x \neq y$ and $\varepsilon > 0$ there is $z \in X$ such that $d(x, z) < d(x, y)$ and $d(z, y) < \varepsilon$.

(iv) Give an example of a metric space where $\overline{B(x, r)} = \overline{B}(x, r)$ does not hold (for certain x, r).

2.2.2 Equivalence of metrics

Definition 2.2.12 *Two metrics d_1, d_2 on a set are called equivalent ($d_1 \simeq d_2$) if they give rise to the same topology, i.e. $\tau_{d_1} = \tau_{d_2}$.*

It is obvious that equivalence of metrics indeed is an equivalence relation.

Exercise 2.2.13 Let d_1, d_2 be metrics on X. Prove that the following are equivalent:

(i) d_1, d_2 are equivalent, i.e. $\tau_{d_1} = \tau_{d_2}$.

(ii) For every $x \in X$ and every $\varepsilon > 0$ there is a $\delta > 0$ such that

$$B^{d_2}(x, \delta) \subseteq B^{d_1}(x, \varepsilon), \quad \text{and} \quad B^{d_1}(x, \delta) \subseteq B^{d_2}(x, \varepsilon).$$

(iii) The map $(X, d_1) \to (X, d_2),\ x \mapsto x$ is a homeomorphism.

(iv) A sequence $\{x_n\}$ converges to x w.r.t. d_1 if and only if it converges to x w.r.t. d_2.

(v) A sequence $\{x_n\}$ converges w.r.t. d_1 if and only if it converges w.r.t. d_2.

Hint: For (v)$\Rightarrow$(iv), use the fact that (v) holds for all sequences to show that $\{x_n\}$ cannot have different limits w.r.t. d_1 and d_2.

Exercise 2.2.14 (i) Let (X, d) be a metric space and $f : [0, \infty) \to [0, \infty)$ a function satisfying

(α) $f(t) = 0 \ \Leftrightarrow \ t = 0$.

(β) $\lim_{t \to 0} f(t) = 0$.

(γ) f is non-decreasing, i.e. $s \leq t \Rightarrow f(s) \leq f(t)$.

(δ) f is subadditive, i.e. $f(s + t) \leq f(s) + f(t)\ \forall s, t \geq 0$.

Prove that $d'(x, y) = f(d(x, y))$ is a metric on X that is equivalent to d.

(ii) Use (i) to prove that

$$d_1(x, y) = \min(1, d(x, y)), \qquad d_2(x, y) = \frac{d(x, y)}{1 + d(x, y)}$$

are metrics equivalent to d.

Definition 2.2.15 *Two norms $\|\cdot\|_1, \|\cdot\|_2$ on a real or complex vector space V are called equivalent ($\|\cdot\|_1 \simeq \|\cdot\|_2$) if there are constants $c_2 \geq c_1 > 0$ such that $c_1\|x\|_1 \leq \|x\|_2 \leq c_2\|x\|_1$ for all $x \in V$.*

Exercise 2.2.16 (i) Prove that equivalence of norms is an equivalence relation.

(ii) Prove that for every $s \in [1, \infty)$ there is a constant $c_{s,n} > 0$ such that the norms on $\mathbb{R}^n$ defined in Example 2.1.13 satisfy

$$\|x\|_\infty \leq \|x\|_p \leq c_{p,n}\|x\|_\infty \quad \forall x \in \mathbb{R}^n,$$

giving the best (i.e. smallest possible) value for $c_{p,n}$.

(iii) Conclude that the norms $\|\cdot\|_p,\ p \in [1, \infty]$ are all equivalent.

(iv) Let $\|\cdot\|_1, \|\cdot\|_2$ be arbitrary norms on V, and define the metrics $d_i(x, y) := \|x - y\|_i, i = 1, 2$. Prove that $\|\cdot\|_1 \simeq \|\cdot\|_2 \ \Leftrightarrow\ d_1 \simeq d_2$. Hint: For $\Leftarrow$ use axiom (ii) of Definition 2.1.10.

Remark 2.2.17 1. In Section 7.7.5 we will prove that *all* norms on $\mathbb{R}^n$ ($n < \infty$) are equivalent.
2. If d_1, d_2 are metrics on X, it is clear that existence of constants $c_2 \geq c_1 > 0$ such that

$$c_1 d_1(x, y) \leq d_2(x, y) \leq c_2 d_1(x, y) \quad \forall x, y \in X \tag{2.8}$$

implies $d_1 \simeq d_2$. And if d_1, d_2 are obtained from norms $\|\cdot\|_i, i = 1, 2$ then by the preceding exercise we have $d_1 \simeq d_2 \Leftrightarrow \|\cdot\|_1 \simeq \|\cdot\|_2 \Leftrightarrow$ (2.8). But if at least one of the metrics d_1, d_2 does not come from a norm, equivalence $d_1 \simeq d_2$ **does not imply** (2.8): Consider $X = \mathbb{R}$ with $d_1(x, y) = |x - y|$ and $d_2(x, y) = \max(1, d_1(x, y))$. Then $d_1 \simeq d_2$ by Exercise 2.2.14, but (2.8) cannot hold since d_1 is unbounded and d_2 is bounded. □

Definition 2.2.18 *The topology on $\mathbb{R}^n$ (and $\mathbb{C}^n$) defined by the equivalent norms $\|\cdot\|_p,\ p \in [1, \infty]$ is called the usual or Euclidean topology.*

We see that passing from a metric space (X, d) to the topological space (X, τ_d), we may lose information. This actually is one of the main reasons for working with topological spaces, since even when all spaces in sight are metrizable, the actual choice of the metrics may be irrelevant and therefore distracting! Purely topological proofs tend to be cleaner than metric proofs.

2.3 Some standard topologies

It is time to see some topologies that do not come from a metric! Some standard topologies can actually be defined on any set X:

Definition/Proposition 2.3.1 *Let X be a set. Then the following are topologies on X:*

- *The discrete topology* $\tau_{\text{disc}} = P(X)$.
- *The indiscrete topology* $\tau_{\text{indisc}} = \{\emptyset, X\}$.
- *The cofinite topology* $\tau_{\text{cofin}} = \{X\backslash Y \mid Y \subseteq X \text{ finite}\} \cup \{\emptyset\}$.
- *The cocountable topology* $\tau_{\text{cocnt}} = \{X\backslash Y \mid Y \subseteq X \text{ countable}\} \cup \{\emptyset\}$.

A discrete topological space is a space equipped with the discrete topology, etc.

Proof. That $\tau_{\rm disc}, \tau_{\rm indisc}$ are topologies is obvious. By definition, $\tau_{\rm fin}, \tau_{\rm cofin}$ contain $\emptyset, X$. Let $U_i \in \tau_{\rm cofin}$ for each $i \in I$. The non-empty U_i are of the form $U_i = X\backslash F_i$ with each F_i finite. Now $\bigcup_i U_i = \bigcup_i X\backslash F_i = X\backslash \bigcap_i F_i$. Since an intersection of finite sets is finite, this is in $\tau_{\rm cofin}$. Let $U_1, U_2 \in \tau_{\rm cofin}$. If either of them is empty then $U_1 \cap U_2 = \emptyset \in \tau_{\rm cofin}$. Otherwise $U_i = X\backslash F_i$ with F_1, F_2 finite. Then $U_1 \cap U_2 = X\backslash(F_1 \cup F_2)$, which is in $\tau_{\rm cofin}$ since the union of two finite sets is finite. The same reasoning works for $\tau_{\rm cocnt}$. (Since a countable union of countable sets is countable, we actually find that $\tau_{\rm cocnt}$ is closed under countable intersections. With later language, for $\tau_{\rm cofin}$ all G_δ-sets are open.) ■

A one-point subset $\{x\} \subseteq X$ is often called a singleton. Nevertheless, we may occasionally allow ourselves to write 'points' when 'singletons' is meant.

Definition 2.3.2 *If (X, τ) is a topological space, a point $x \in X$ is called isolated if $\{x\} \in \tau$.*

Exercise 2.3.3 (i) Prove that (X, τ) is discrete if and only if every $x \in X$ is isolated.

(ii) If d is a metric on X, prove that τ_d is discrete if and only if for every $x \in X$ there is $\varepsilon_x > 0$ such that $d(x, y) \geq \varepsilon_x \; \forall y \neq x$.

Metrics satisfying the equivalent conditions in (ii) are called discrete. Clearly the standard discrete metric is discrete.

Exercise 2.3.4 Let X be arbitrary. Prove

(a) $\tau_{\rm indisc} \subseteq \tau_{\rm cofin} \subseteq \tau_{\rm cocnt} \subseteq \tau_{\rm disc}$.

(b) If $2 \leq \#X < \infty$ then $\tau_{\rm indisc} \subsetneq \tau_{\rm cofin} = \tau_{\rm cocnt} = \tau_{\rm disc}$.

(c) If X is countably infinite then $\tau_{\rm indisc} \subsetneq \tau_{\rm cofin} \subsetneq \tau_{\rm cocnt} = \tau_{\rm disc}$.

(d) If X is uncountable then $\tau_{\rm indisc} \subsetneq \tau_{\rm cofin} \subsetneq \tau_{\rm cocnt} \subsetneq \tau_{\rm disc}$.

The above exercise has provided examples of inclusion relations between different topologies on a set. This merits a definition:

Definition 2.3.5 *Let X be a set and τ_1, τ_2 topologies on X. If $\tau_1 \subseteq \tau_2$ then we say that τ_1 is coarser than τ_2 and that τ_2 is finer than τ_1. (Some authors use weaker/stronger instead of coarser/finer.)*

Exercise 2.3.6 Let X, I be sets and τ_i a topology on X for every $i \in I$. Prove that $\tau = \bigcap_{i \in I} \tau_i$ is a topology on X.

Clearly, for any set, the indiscrete topology is the coarsest topology and the discrete topology the finest. And $\bigcap_i \tau_i$ is coarser than each τ_i.

Definition 2.3.7 *A property P that a topological space may or may not have is called hereditary if every subspace of a space with property P automatically has property P.*

Exercise 2.3.8 Prove that the following properties are hereditary: (i) metrizability, (ii) discreteness, (iii) indiscreteness, (iv) cofiniteness and (v) cocountability.

In order to avoid misconceptions, we emphasize that the properties of discreteness, indiscreteness, cofiniteness and cocountability are quite exceptional in that they completely determine the topology. For other properties, like metrizability, this typically is not the case.

2.4 Closed and clopen subsets. Connectedness

Definition 2.4.1 *Let (X, τ) be a topological space. A set $Y \subseteq X$ is called closed if and only if $X \backslash Y$ is open.*

The following is obvious:

Lemma 2.4.2 *Let (X, τ) be a topological space. Then*

(i) $\emptyset$ and X are closed.

(ii) If C_i is closed for every $i \in I$ then $\bigcap_{i \in I} C_i$ is closed.

(iii) If $C_1, \ldots, C_n$ are closed then $\bigcup_{i=1}^n C_i$ is closed.

Thus the family of closed sets is closed under arbitrary intersections and finite unions.

It is clear that we can also specify a topology on X by giving a family of sets satisfying (i)-(iii) above and calling their complements open. In fact, the cofinite (resp. cocountable) topology on X is defined more naturally by declaring as closed X and all its finite (resp. countable) subsets. It is then obvious that (i)-(iii) in Lemma 2.4.2 are satisfied.

Example 2.4.3 Another example where it is more convenient to specify the closed sets is provided by the definition of the Zariski[1] topology on an algebraic variety. In the simplest situation this goes as follows: Let k be a field, $n \in \mathbb{N}$ and $X = k^n$. If $P \subseteq k[x_1, \ldots, x_n]$ is a (possibly infinite) set of polynomials in n variables $x_1, \ldots, x_n$, we define

$$Y_P = \{x \in k^n \mid p(x) = 0 \quad \forall p \in P\} \ \subseteq \ k^n.$$

(We say that Y_P is the zero-set of P.) A set $Y \subseteq k^n$ is algebraic if $Y = Y_P$ for some P as above. We have $Y_\emptyset = k^n$, thus $X = k^n$ is algebraic. Letting P contain two contradictory equations (e.g. $P = \{x_1, x_1 - 1\}$) we obtain $Y_P = \emptyset$, thus $\emptyset$ is algebraic. If I is any index set and $P_i \subseteq k[x_1, \ldots, x_n]$ $\forall i \in I$, it is easy to see that $\bigcap_{i \in I} Y_{P_i} = Y_Q$ for $Q = \bigcup_{i \in I} P_i$. Thus arbitrary intersections of algebraic sets are algebraic. Now let $P_1, P_2 \subseteq k[x_1, \ldots, x_n]$ and define $Q = \{p_1 p_2 \mid p_1 \in P_1, \ p_2 \in P_2\}$. It is not hard to check that $Y_Q = Y_{P_1} \cup Y_{P_2}$, and by induction we see that finite unions of algebraic sets are algebraic. We have thus proven that the family of algebraic subsets of $X = k^n$ satisfies (i)-(iii) of Lemma 2.4.2, so that they are the closed sets of a topology on X, the Zariski topology. (This can be generalized considerably, cf. Section B and books like [195, 196, 107].)

It should be noted that infinite unions of algebraic sets need not be algebraic. (In order to adapt the above argument to infinite unions, we would need to make sense of infinite products, which is difficult in a purely algebraic context.) Thus we have a non-trivial example for the 'arbitrary unions, finite intersection' situation that is completely different from the metric topologies. (In fact, Zariski topologies usually are not metrizable.) □

Exercise 2.4.4 (i) Prove that $Y_{P_1} \cup Y_{P_2} = Y_Q$ for $Q = \{p_1 p_2 \mid p_1 \in P_1, \ p_2 \in P_2\}$.

(ii) Prove that for $n = 1$, the Zariski topology is just the cofinite topology on k.

Unfortunately, the terminology open/closed is quite misleading: A set $Y \subseteq (X, \tau)$ can be neither open nor closed, e.g.: $(0, 1] \subseteq \mathbb{R}$. On the other hand, a set $Y \subseteq (X, \tau)$ can be open and closed at the same time!

[1] Oscar Zariski (1899-1986) was born in Ukraine (then part of Russia), emigrated first to Italy, then to the US. He was one of the pioneers of modern algebraic geometry.

Definition 2.4.5 *A subset of a topological space is called clopen if it is closed and open. The set of clopen subsets of X is called* $\mathrm{Clop}(X)$.

- In every topological space (X, τ), the subsets $\emptyset$ and X are clopen.
- If X is discrete, every $Y \subseteq X$ is clopen.
- If $C, D \subseteq X$ are clopen then $C \cup D$, $C \cap D$ and $\neg X := X \backslash C$ are clopen. It is easy to show (provided one knows Definition 11.1.66) that $(\mathrm{Clop}(X), \cup, \cap, \neg, \emptyset, X)$ is a Boolean algebra. This Boolean algebra has interesting applications, cf. Section 11.1.11.

Definition 2.4.6 *A topological space X is connected if $\emptyset$ and X are the only clopen subsets.*

We defer the detailed discussion of the notion of connectedness and its many ramifications (which touch upon algebraic topology) until Section 9. But we will encounter it every now and then and prove some small facts. For now, we only note:

- X is connected if and only if it cannot be written as $X = U \cup V$ with U and V both non-empty, disjoint and open (equivalently, both closed). (This is often taken as the definition of connectedness, but we prefer the above one for its conciseness.)
- All indiscrete spaces are connected.
- Discrete spaces with more than one point are not connected.

2.5 The separation axioms T_1 and T_2

We have seen that a space (X, τ) is discrete if and only if all singletons $\{x\}$ are open. When are the singletons closed?

Exercise 2.5.1 Prove that for a topological space (X, τ), the following are equivalent:

(i) For every $x \in X$, the singleton $\{x\} \subseteq X$ is closed.

(ii) For any $x, y \in X$ with $x \neq y$ there is an open set U such that $x \in U,\ y \notin U$.

(iii) For every $x \in X$, we have $\{x\} = \bigcap\{U \mid x \in U \in \tau\}$.

Definition 2.5.2 *A space satisfying the equivalent properties of Exercise 2.5.1 is called a T_1-space.*

Many topological spaces actually satisfy the following stronger axiom:

Definition 2.5.3 *A topological space (X, τ) is called Hausdorff space or T_2-space if for any $x, y \in X$ with $x \neq y$ we can find open U, V such that $x \in U,\ y \in V$ and $U \cap V = \emptyset$.*

One also says: The open sets separate the points of X.

Lemma 2.5.4 *If (X, d) is a metric space then the metric topology τ_d is T_2 (Hausdorff).*

Proof. If $x \neq y$ then $d := d(x, y) > 0$. Let $U = B(x, d/2), V = B(y, d/2)$. Then $x \in U \in \tau,\ y \in V \in \tau$. It remains to prove that $U \cap V = \emptyset$. Assume $z \in U \cap V$. Then $d(x, z) < d/2$ and $d(y, z) < d/2$. Thus $d = d(x, y) \leq d(x, z) + d(z, y) < d/2 + d/2 = d$, which is absurd. ■

Remark 2.5.5 1. The preceding result is false for pseudometric spaces that are not metric!

2. The T_1- and T_2-properties are called separation axioms. (The T in T_1, T_2 stands for 'Trennung', German for separation.) We will encounter quite a few more.

3. Hausdorff originally included the T_2-axiom in the definition of topological spaces, but this was dropped when it turned out that also non-T_2 spaces are important.

4. Discreteness can be understood as the strongest possible separation property: One easily checks that a space X is discrete if and only if for *any* two disjoint sets $A, B \subseteq X$ there are disjoint open sets U, V containing A and B, respectively. □

Exercise 2.5.6 Prove:

(i) The T_1 and T_2-properties are hereditary.

(ii) Let τ_1, τ_2 be topologies on X, where τ_2 is finer than τ_1. Prove that if τ_1 is T_1 (resp. T_2) then τ_2 is T_1 (resp. T_2).

Exercise 2.5.7 Prove the following:

(i) $T_2 \Rightarrow T_1$.

(ii) Every discrete space (X, τ_{disc}) is T_2.

(iii) If $\#X \geq 2$ then $(X, \tau_{\text{indisc}})$ is not T_1 (thus not T_2).

(iv) The Zariski topology on k^n is T_1 for all k and n.

(v) Every cofinite space (X, τ_{cofin}) is T_1.

(vi) If (X, τ) is T_1 then $\tau \supseteq \tau_{\text{cofin}}$. (Thus τ_{cofin} is the coarsest T_1 topology on X.)

(vii) Every finite T_1-space is discrete (and thus T_2).

(viii) If $\#X = \infty$ then (X, τ_{cofin}) is not T_2. Thus $T_1 \not\Rightarrow T_2$.

Corollary 2.5.8 $(X, \tau_{\text{indisc}})$ *with* $\#X \geq 2$ *and* (X, τ_{cofin}) *with* $\#X = \infty$ *are not metrizable.*

The existence of non-metrizable spaces is the second main reason for studying general topology: In various situations, topological spaces arise that are not metrizable and therefore simply could not be discussed in a theory of metric spaces. We will see that quotient spaces of metric spaces may fail to be metrizable. Other examples of non-metrizable topologies are the Zariski topology from Example 2.4.3, which is T_1, but rarely T_2 (it is discrete when k is finite), and the 'weak topologies' of functional analysis.

Exercise 2.5.9 Prove:

(i) Every finite subspace of a T_1-space is discrete.

(ii) Deduce that connectedness is not hereditary.

(iii) A connected T_1-space with more than one point has no isolated point.

2.6 Interior. Closure. Boundary

If (X, τ) is a topological space and $Y \subseteq X$, one may ask whether there is a largest open (or closed) subset of Y or a smallest open (or closed) subset containing Y. Since the union of any number of open sets is open, and the intersection of any number of closed sets is closed, two of these four questions have a positive answer. (The other two in general do not, but see Section 2.8.3.)

Definition 2.6.1 *Let $Y \subseteq (X, \tau)$. Then the interior Y^0 and closure $\overline{Y}$ of Y are defined by*

$$\begin{aligned} Y^0 &= \bigcup\{U \mid U \text{ open}, \ U \subseteq Y\}, \\ \overline{Y} &= \bigcap\{C \mid C \text{ closed}, \ C \supseteq Y\}. \end{aligned}$$

The points of Y^0 are called interior points of Y, those of $\overline{Y}$ adherent points of Y.

(Some authors write 'closure point', 'proximate point' or 'limit point'. But the latter term is also used for the different concept of 'accumulation point'!)

Exercise 2.6.2 Let (X, τ) be a topological space and $Y, Z \subseteq X$. Prove:

(i) Y^0 is open, $\overline{Y}$ is closed, and $Y^0 \subseteq Y \subseteq \overline{Y}$.

(ii) If Y is open then $Y^0 = Y$. If Y is closed then $\overline{Y} = Y$.

(iii) $\emptyset^0 = \overline{\emptyset} = \emptyset$ and $X^0 = \overline{X} = X$.

(iv) $Y^{00} = Y^0$ and $\overline{\overline{Y}} = \overline{Y}$. (Idempotency)

(v) If $Y \subseteq Z$ then $Y^0 \subseteq Z^0$ and $\overline{Y} \subseteq \overline{Z}$. (Monotonicity)

The following simple fact has many uses:

Lemma 2.6.3 *If $U \cap V = \emptyset$ and U is open then $U \cap \overline{V} = \emptyset$.*

Proof. Since U is open, $X \backslash U$ is closed. And $U \cap V = \emptyset$ is equivalent to $V \subseteq X \backslash U$. Thus $X \backslash U$ appears in the family over which the intersection is taken in the definition of $\overline{V}$. Thus $\overline{V} \subseteq X \backslash U$, which is equivalent to $\overline{V} \cap U = \emptyset$. ■

The connection between the interior and closure operations is provided by complements:

Lemma 2.6.4 *For every $Y \subseteq X$ we have*

$$X \backslash Y^0 = \overline{X \backslash Y}, \qquad (X \backslash Y)^0 = X \backslash \overline{Y}.$$

Proof. We only prove the first identity:

$$\begin{aligned} X \backslash Y^0 &= X \backslash \bigcup\{U \text{ open} \mid U \subseteq Y\} = \bigcap\{X \backslash U \mid U \text{ open}, U \subseteq Y\} \\ &= \bigcap\{C \text{ closed} \mid X \backslash C \subseteq Y\} = \bigcap\{C \text{ closed} \mid X \backslash Y \subseteq C\} = \overline{X \backslash Y}. \end{aligned}$$

The first equality is just the definition of Y^0, the second is de Morgan, the third results by replacing the open set U by $X \backslash C$ with C closed, and the last results from the equivalence between $X \backslash C \subseteq Y$ and $X \backslash Y \subseteq C$. ■

Exercise 2.6.5 (i) Prove the equivalence used in the last sentence of the proof.

(ii) Prove the second identity of Lemma 2.6.4.

How the interior and closure operations interact with union and intersection is less obvious:

Lemma 2.6.6 *Let (X, τ) be a topological space and $A, B \subseteq X$. Then*

(i) $\overline{A \cup B} = \overline{A} \cup \overline{B}$.

(ii) $(A \cap B)^0 = A^0 \cap B^0$.

(iii) $\overline{A \cap B} \subseteq \overline{A} \cap \overline{B}$.

(iv) $A^0 \cup B^0 \subseteq (A \cup B)^0$.

Proof. Combining the trivial inclusions

$$A \subseteq A \cup B, \quad B \subseteq A \cup B, \quad A \cap B \subseteq A, \quad A \cap B \subseteq B$$

with monotonicity we obtain

$$\overline{A} \subseteq \overline{A \cup B}, \quad \overline{B} \subseteq \overline{A \cup B}, \quad \overline{A \cap B} \subseteq \overline{A}, \quad \overline{A \cap B} \subseteq \overline{B},$$

$$A^0 \subseteq (A \cup B)^0, \quad B^0 \subseteq (A \cup B)^0, \quad (A \cap B)^0 \subseteq A^0, \quad (A \cap B)^0 \subseteq B^0,$$

from which we obtain

$$\overline{A} \cup \overline{B} \subseteq \overline{A \cup B}, \quad \overline{A \cap B} \subseteq \overline{A} \cap \overline{B}, \quad A^0 \cup B^0 \subseteq (A \cup B)^0, \quad (A \cap B)^0 \subseteq A^0 \cap B^0.$$

We thus have proven (iii) and (iv) and 'half of' (i),(ii). Now, $\overline{A} \cup \overline{B}$ is a closed subset containing $A \cup B$, so that it appears in the family defining $\overline{A \cup B}$. Thus $\overline{A \cup B} \subseteq \overline{A} \cup \overline{B}$. Since the converse inclusion was proven before, we have (i). Similarly, the fact that $A^0 \cap B^0$ is an open subset of $A \cap B$ implies $A^0 \cap B^0 \subseteq (A \cap B)^0$, thus (ii). ∎

Remark 2.6.7 1. It is very important to understand that equality need not hold in (iii) and (iv)! Thus $\overline{A} \cap \overline{B}$ may be strictly smaller than $\overline{A \cap B}$. (For $X = \mathbb{R}$, $A = (0,1)$, $B = (1,2)$ we have $\overline{A} \cap \overline{B} = \{1\}$, but $\overline{A \cap B} = \emptyset$.) Similarly, $A^0 \cap B^0$ may be strictly smaller than $(A \cap B)^0$. (For $X = \mathbb{R}$, $A = [0,1]$, $B = [1,2]$ we have $(A \cup B)^0 = (0,2)$, but $A^0 \cup B^0 = (0,1) \cup (1,2)$.)

2. Induction over n gives the generalization $\overline{\bigcup_{i=1}^n Y_i} = \bigcup_{i=1}^n \overline{Y_i}$ of (i), which we will often use. Similarly, the interior of a finite intersection equals the intersection of the interiors. But these statements may very well be false for infinite unions/intersections! Example:

$$\bigcup_{x \in \mathbb{Q}} \overline{\{x\}} = \bigcup_{x \in \mathbb{Q}} \{x\} = \mathbb{Q} \neq \mathbb{R} = \overline{\mathbb{Q}} = \overline{\bigcup_{x \in \mathbb{Q}} \{x\}}.$$

(This is closely related to the fact that an infinite union of closed sets need not be closed.) □

Exercise 2.6.8 (Topology from closure operation (Kuratowski 1922)) [2] Let X be a set and cl: $P(X) \to P(X)$ a map satisfying the properties

[2] Kazimierz Kuratowski (1896-1980). Polish mathematician and logician.

(α) $\mathrm{cl}(\emptyset) = \emptyset$.

(β) $\mathrm{cl}(Y) \supseteq Y$ for every $Y \subseteq X$.

(γ) $\mathrm{cl}(A \cup B) = \mathrm{cl}(A) \cup \mathrm{cl}(B)$.

Prove:

(i) $A \subseteq B \Rightarrow \mathrm{cl}(A) \subseteq \mathrm{cl}(B)$.

(ii) There is a unique topology τ on X such that $Y \subset X$ is τ-closed if and only if $\mathrm{cl}(Y) = Y$.

(iii) Assuming in addition (δ) $\mathrm{cl}(\mathrm{cl}(Y)) = \mathrm{cl}(Y)$, we have $\mathrm{cl}(Y) = \overline{Y}^\tau$ for every $Y \subseteq X$.

Remark 2.6.9 1. Combining the preceding exercise with Exercise 2.6.2 and Lemma 2.6.6, we see that specifying a topology on a set X is equivalent to giving a closure operator satisfying (α)-(δ).

2. Since the interior operation $Y \mapsto Y^0$ has properties dual to the closure, one could also obtain a topology from an interior operation having properties dual to (α)-(δ). □

Definition 2.6.10 *If $Y \subseteq (X, \tau)$ then the boundary ∂Y of Y is*

$$\partial Y = \overline{Y} \backslash Y^0.$$

(Some authors write 'frontier' instead of 'boundary', in symbols $\mathrm{Fr}\, Y$*.)*

Lemma 2.6.11 *Let $Y \subseteq (X, \tau)$. Then*

*(i) $\partial Y = \overline{Y} \cap (X \backslash Y^0) = \overline{Y} \cap \overline{X \backslash Y} = \partial(X \backslash Y)$.
Thus a subset and its complement have the same boundary.*

(ii) ∂Y is closed.

(iii) $\overline{Y} = Y \cup \partial Y$ and $Y^0 = Y \backslash \partial Y$.

(iv) $\partial Y = \emptyset \;\Leftrightarrow\; \overline{Y} = Y^0 \;\Leftrightarrow\; Y$ is clopen.

(v) $\partial Y = X \;\Leftrightarrow\; \overline{Y} = X$ and $Y^0 = \emptyset$.

Proof. (i) The first identity is just Lemma 2.6.4. It is clear that $\overline{Y} \cap \overline{X \backslash Y}$ is unchanged under the replacement $Y \rightsquigarrow X \backslash Y$. (ii) Obvious from $\partial Y = \overline{Y} \cap \overline{X \backslash Y}$. (iii) We have $Y \cup \partial Y = Y \cup (\overline{Y} \backslash Y^0) = Y \cup \overline{Y} = \overline{Y}$ and $Y \backslash \partial Y = Y \backslash (\overline{Y} \backslash Y^0) = Y^0$. Both computations use $Y^0 \subseteq Y \subseteq \overline{Y}$. (iv) We first note that $\partial Y = \emptyset$ is equivalent to (*) $Y^0 = \overline{Y}$. If Y is clopen then $Y^0 = Y = \overline{Y}$, thus $(*)$ holds. Conversely, (*) together with $Y^0 \subseteq Y \subseteq \overline{Y}$ implies $Y^0 = Y = \overline{Y}$, thus Y is open and closed. (v) In view of $Y^0 \subseteq \overline{Y}$ it is clear that $\overline{Y} \backslash Y^0 = X$ holds if and only if $\overline{Y} = X$ and $Y^0 = \emptyset$. ■

Exercise 2.6.12 Let X be a topological space. Prove or disprove (by counterexample) the following statements:

(i) $(\partial Y)^0 = \emptyset$ for every $Y \subseteq X$. (I.e. boundaries have empty interior.)

(ii) $(\partial Y)^0 = \emptyset$ holds whenever $Y \subseteq X$ is closed.

Exercise 2.6.13 Let (X, τ) be a topological space, $(Y, \tau_Y) \subseteq X$ a subspace, and $Z \subseteq Y$. Prove:

(i) We have $\tau_Y \subseteq \tau$ (interpreting subsets of Y as subsets of X) if and only if $Y \subseteq X$ is open.

(ii) The interior of Z in Y, denoted $\mathrm{Int}_Y(Z)$, contains Z^0 (the interior of Z in X).

(iii) Assume Y is open. Then the interiors of Z in X and in Y coincide (thus $\mathrm{Int}_Y(Z) = Z^0$), and Z is open in Y if and only if it is open in X.

(iv) Z is closed in Y if and only if $Z = Y \cap C$ for some closed $C \subseteq X$.

(v) The closure of Z in (Y, τ_Y), denoted $\mathrm{Cl}_Y(Z)$, equals $\overline{Z} \cap Y$. ($\overline{Z}$ is the closure of Z in X).

(vi) Assume Y is closed. Then the closures of Z in X and Y coincide (thus $\mathrm{Cl}_Y(Z) = \overline{Z}$), and Z is closed in Y if and only if it is closed in X.

Exercise 2.6.14 Let (X, τ) be a topological space. An open set $U \subseteq X$ is called regular open if $U = \overline{U}^0$. A closed set C is called regular closed if $C = \overline{C^0}$.

(i) Prove that the complement of a regular open set is regular closed, and vice versa.

(ii) Prove that the intersection of two regular open sets is regular open.

(iii) Give examples of open sets in $\mathbb{R}$ that are (a) regular, (b) not regular.

(iv) Give two regular open sets in $\mathbb{R}$ whose union is not regular open.

(v) Prove that $U \subseteq \overline{U}^0$ for every open U and $\overline{C^0} \subseteq C$ for every closed C.

(vi) Prove that each $\overline{Y}^0$ is regular open, i.e. $\overline{\overline{Y}^0}^0 = \overline{Y}^0$.

(vii) Show that in every non-discrete T_1 space there is a non-regular open set.

Remark 2.6.15 Let U, V be regular open. Then $U \wedge V := U \cap V$ is regular open by (ii), and $U \vee V := \overline{U \cup V}^0$ and $U^\perp := (X \backslash U)^0 = \overline{X \backslash U}^0$ are regular open by (v). Now one easily shows that $(A, \vee, \wedge, \perp, \emptyset, X)$, where A is the set of regular open sets, is a Boolean algebra. A clopen set obviously is regular open. Indeed, one easily checks that the Boolean algebra $\mathrm{Clop}(X)$ of clopen subsets is a Boolean subalgebra of the Boolean algebra of regular open sets. □

The following is not more than an amusing curiosity:

Exercise 2.6.16 (Kuratowski's closure-complement theorem) Let (X, τ) be a topological space. To every subset $Y \subseteq X$ we can associate to new subsets $Y^c \equiv X \backslash Y$ and $\overline{Y}$.

(i) Use Exercise 2.6.14(vi) to prove $\overline{\overline{\overline{\overline{Y}^c}^c}^c} = \overline{\overline{Y}^c}$.

(ii) Use this to show that beginning from a single subset $Y \subseteq X$ and applying the operations closure and complement, one can produce at most 14 different subsets of X.

2.7 Neighborhoods. Density

2.7.1 Neighborhoods. Topologies from neighborhoods

Definition 2.7.1 *Let (X,τ) be a topological space and $x \in X$.*

(i) An open neighborhood of x is a $U \in \tau$ such that $x \in U$. The set of open neighborhoods of x is denoted $\mathcal{U}_x$.

(ii) A neighborhood of x is a set $N \subseteq X$ that contains an open neighborhood of x. The set of all neighborhoods of x is denoted $\mathcal{N}_x$.

Lemma 2.7.2 *Let (X,τ) be a topological space.*

(i) $\mathcal{U}_x$ is closed w.r.t. finite intersections.

(ii) $N \subseteq X$ is a neighborhood of $x \in X$ if and only if $x \in N^0$.

(iii) $N \subseteq X$ is open if and only if $y \in N \Rightarrow N \in \mathcal{N}_y$.

(iv) $\mathcal{N}_x$ has the following properties:

- *$\mathcal{N}_x \neq \emptyset$ and $\emptyset \notin \mathcal{N}_x$.*
- *If $N \in \mathcal{N}_x$ and $M \supseteq N$ then $M \in \mathcal{N}_x$.*
- *If $N, M \in \mathcal{N}_x$ then $N \cap M \in \mathcal{N}_x$.*

A non-empty family of non-empty sets with these two properties is called a filter. In particular, $\mathcal{N}_x$ is the neighborhood filter of x. For more on filters see Sections 5.1.3 and 7.5.4.)

Proof. (i) Obvious. (ii) $N^0 \subseteq N$ is open, thus if $x \in N^0$ then N is a neighborhood of x. If $N \in \mathcal{N}_x$ then there is an open U with $x \in U \subseteq N$, thus $x \in U \subseteq N^0 \subseteq N$. (iii) Every open U clearly is a neighborhood for each $y \in U$. If N is a neighborhood of each $y \in N$ then there are open U_y with $y \in U_y \subseteq N$. But then $N = \bigcup_{y \in N} U_y$, thus N is open. (iv) Obvious. ■

Lemma 2.7.3 *Let $Y \subseteq (X,\tau)$. Then $x \in \overline{Y}$ if and only if $N \cap Y \neq \emptyset$ for every (open) neighborhood N of x.*

Proof. By Lemma 2.6.4, $\overline{Y} = X \backslash (X \backslash Y)^0$. Thus $x \in \overline{Y}$ is equivalent to $x \notin (X\backslash Y)^0$, which is equivalent to: There is no open set U such that $x \in U \subseteq X\backslash Y$. But this in turn is equivalent to: every open set U with $x \in U$ satisfies $U \cap Y \neq \emptyset$. This proves the claim for open neighborhoods. The claim for arbitrary neighborhoods follows from the facts that (a) open neighborhoods are neighborhoods and (b) every neighborhood contains an open neighborhood. ■

Remark 2.7.4 1. Lemma 2.7.3 gives an alternative proof of Lemma 2.6.3: If $x \in U$, then U is an open neighborhood of x disjoint from V. Thus Lemma 2.7.3 gives $x \notin \overline{V}$, so that $U \cap \overline{V} = \emptyset$.

2. If $x \in X$ is an isolated point then $\{x\}$ is an open neighborhood of x, thus $x \notin Y \subseteq X \;\Rightarrow\; x \notin \overline{Y}$ by Lemma 2.7.3, whence the term 'isolated'.

3. If (X,d) is a metric space and $Y \subseteq X$, Lemma 2.7.3 implies that the closures of Y in the metric (Lemma 2.1.16) and the topological sense (Definition 2.4.1) coincide.

4. If $Y \subseteq (X,\tau)$ then $x \in \partial Y$ if and only if every (open) neighborhood of x contains points of Y and of $X\backslash Y$. (This follows from $\partial Y = \overline{Y} \cap \overline{X\backslash Y}$.)

5. Every $X \subseteq \mathbb{R}$ that is bounded above has a supremum $\sup(X) \in \mathbb{R}$. If $\sup(X) \notin X$ then the definition of the supremum implies that $(\sup(X) - \varepsilon, \sup(X)) \cap X \neq \emptyset$ for every $\varepsilon > 0$. Thus we always have $\sup(X) \in \overline{X}$, and similarly for the infimum. □

The following shows that a topology τ can also be defined in terms of axioms for neighborhoods:

Exercise 2.7.5 (Topology from neighborhood axioms) Let X be a set, and let for every $x \in X$ a non-empty family $\mathcal{M}_x \subseteq P(X)$ be given, such that the following holds:

(α) $x \in N \ \forall N \in \mathcal{M}_x$.

(β) If $N \in \mathcal{M}_x$ and $M \supseteq N$ then $M \in \mathcal{M}_x$.

(γ) If $N, M \in \mathcal{M}_x$ then $N \cap M \in \mathcal{M}_x$.

(δ) For every $N \in \mathcal{M}_x$ there is a $U \in \mathcal{M}_x$ such that $U \subseteq N$ and $U \in \mathcal{M}_y$ for each $y \in U$.

Prove:

(i) $\tau = \{U \subseteq X \mid \forall x \in U : U \in \mathcal{M}_x\}$ is a topology on X.

(ii) $\mathcal{M}_x$ equals $\mathcal{N}_x$, the set of τ-neighborhoods of x.

(iii) τ is the unique topology τ on X for which (b) holds.

Remark 2.7.6 1. In view of Lemma 2.7.2 and Exercise 2.7.5, specifying a topology on a set X is equivalent to specifying a neighborhood system $\{\mathcal{M}_x\}_{x\in X}$ satisfying (α)-(δ).

2. ⋆⋆ Combining this with the bijection between topologies and closure operators satisfying (α)-(δ) in Exercise 2.6.8, we clearly also have a bijection between closure operators and neighborhood systems. Interestingly, this bijection remains intact if one omits the respective axioms (δ) from the definitions of closure operators and neighborhood systems. This leads to a generalization of topological spaces, called 'pre-topological spaces'. □

Exercise 2.7.7 Let τ be the standard topology on $\mathbb{R}$. Let $\mathbb{N} = \{1, 2, 3, \ldots\}$.

(i) Prove that there is a topology $\widetilde{\tau}$ on $\mathbb{R}$ such that:

- the $\widetilde{\tau}$-neighborhoods of $x \neq 0$ are the same as the τ-neighborhoods of x.
- The $\widetilde{\tau}$-neighborhoods of $x = 0$ are the sets that contain

$$(-\varepsilon, \varepsilon) \backslash \left\{ \frac{1}{n} \mid n \in \mathbb{N} \right\}$$

for some $\varepsilon > 0$.

(ii) Prove that $\widetilde{\tau}$ is finer than τ.

(iii) Prove that $\widetilde{\tau}$ is Hausdorff.

(iv) Prove that $C = \left\{ \frac{1}{n} \mid n \in \mathbb{N} \right\}$ is closed w.r.t. $\widetilde{\tau}$, but not w.r.t. τ.

(v) Prove that there are no $U, V \in \widetilde{\tau}$ with $U \cap V = \emptyset$ and $0 \in U, C \subseteq V$.
(Later we will say: $(\mathbb{R}, \widetilde{\tau})$ is not regular (T_3).)

2.7.2 Dense subsets. Nowhere dense subsets

Definition 2.7.8 *A set $Y \subseteq (X,\tau)$ is called dense (in X) if $\overline{Y} = X$.*

(Some authors write 'everywhere dense', but the 'everywhere' is redundant.)

Lemma 2.7.9 *Let (X,τ) be a topological space with $X \neq \emptyset$. Then $Y \subseteq X$ is dense if and only if $Y \cap W \neq \emptyset$ whenever $\emptyset \neq W \in \tau$.*

Proof. By definition, Y is dense if and only if $x \in \overline{Y}$ for every $x \in X$. By Lemma 2.7.3, this is equivalent to $Y \cap U \neq \emptyset$ whenever $x \in U \in \tau$. Since the only rôle of $x \in U$ here is to guarantee that $U \neq \emptyset$, this condition is equivalent to the one in the Lemma.

Alternative argument: If there is $\emptyset \neq W \in \tau$ with $Y \cap W = \emptyset$, then $\overline{Y} \cap W = \emptyset$ by Lemma 2.6.3, thus $\overline{Y} \neq X$. If no such W exists, we have $(X \backslash Y)^0 = \emptyset$, thus $X \backslash \overline{Y} = \emptyset$ by Lemma 2.6.4 and therefore $\overline{Y} = X$. ■

Notice that the intersection of two dense sets need not be dense. It can even be empty, as in the following example: $U_1 = \mathbb{Q}$ and $U_2 = \mathbb{Q} + \sqrt{2}$ are both dense in $\mathbb{R}$, but $U_1 \cap U_2 = \emptyset$. (Otherwise we could deduce that $\sqrt{2}$ is rational.) But we have:

Lemma 2.7.10 *(i) If $Y \subseteq X$ is dense and $V \subseteq X$ is open then $V \subseteq \overline{V \cap Y}$.*

(ii) If $V, Y \subseteq X$ are both dense and V is open then $V \cap Y$ is dense.

Proof. (i) Let $x \in V$. We want to show $x \in \overline{V \cap Y}$. In view of Lemma 2.7.3 this amounts to showing $W \cap (V \cap Y) \neq \emptyset$ for every open $W \ni x$. But $W \cap (V \cap Y) = (W \cap V) \cap Y$. Now, $W \cap V$ is open and non-empty (since it contains x), thus density of Y and Lemma 2.7.9 give $(W \cap V) \cap Y \neq \emptyset$, and we are done.

(ii) By (i), $V \subseteq \overline{V \cap Y}$. Now density of V gives $X = \overline{V} \subseteq \overline{V \cap Y}$, thus $\overline{V \cap Y} = X$. ■

Corollary 2.7.11 *Any finite intersection of dense open sets is dense.*

Remark 2.7.12 For infinitely many dense open sets this is not necessarily true: Consider (X, τ_{cofin}) for X countably infinite. For every $x \in X$, the set $X \backslash \{x\}$ is open (by definition of τ_{cofin}) and dense (since it is not closed and X is the only set that is strictly bigger). But the countable intersection $\bigcap_x (X \backslash \{x\})$ is empty and thus certainly not dense.

In Section 3.3 we will see that this cannot happen for complete metric spaces. □

With Lemma 2.6.4, we have that $Y \subseteq X$ is dense $\Leftrightarrow$ $X \backslash Y$ has empty interior. The following property is stronger than having empty interior:

Definition 2.7.13 *If (X,τ) is a topological space, $Y \subseteq X$ is nowhere dense if $\overline{Y}^0 = \emptyset$.*

Remark 2.7.14 (i) Every closed set with empty interior is nowhere dense, e.g. $\mathbb{Z} \subseteq \mathbb{R}$.

(ii) A non-closed example of a nowhere dense set is given by $\{1/n \mid n \in \mathbb{N}\} \subseteq \mathbb{R}$.

(iii) While a set can be dense and have empty interior, e.g. $\mathbb{Q} \subseteq \mathbb{R}$, a dense set clearly cannot be nowhere dense (unless $X = \emptyset$). □

Exercise 2.7.15 Let X be a topological space. Prove that $Y \subseteq X$ is nowhere dense if and only if for every non-empty open $U \subseteq X$ there is a non-empty open $V \subseteq U$ such that $V \cap Y = \emptyset$.

2.7.3 ⋆ Accumulation points. Perfect sets. Scattered spaces

In this short section we discuss some notions that are related to the closure of a subset and to isolated points. This material is of lesser importance and may safely be ignored until it is needed.

Definition 2.7.16 *A topological space is called dense-in-itself if it has no isolated points. A subset of a topological space is called perfect if it is closed and dense-in-itself (as a subspace).*

By Exercise 2.5.9, a connected T_1 space with more than one point is dense-in-itself.

Definition 2.7.17 *If X is a topological space and $Y \subseteq X$, a point $x \in X$ is called an accumulation point of Y if every neighborhood of x contains a point of Y different from x. The set of accumulation points of Y is called the derived set Y'.*

Remark 2.7.18 1. Using Lemma 2.7.3 one sees immediately that $x \in Y' \Leftrightarrow x \in \overline{Y\backslash\{x\}}$.

2. A point of Y may or may not be in Y'. In fact, if (X, τ) is discrete then each $Y\backslash\{x\}$ is closed for any Y, x, so that 1. implies that $Y' = \emptyset$ for every $Y \subseteq X$.

3. We have $x \in X'$ if and only if every neighborhood of X contains a point other than x, which is true if and only if $\{x\}$ is not open. Thus the derived set X' of the total space is the complement of the set of isolated points, and X is dense-in-itself if $X' = X$.

4. The notion of accumulation point played a central rôle in the early development of set theory and point set topology. But the simpler notions of open and closed sets have turned out to be more fundamental. Accumulation points continue to be relevant for certain specialized matters, like the discussion of (weak) countable compactness. □

Exercise 2.7.19 Let X be a topological space X and $Y \subseteq X$. Prove:

(i) $\overline{Y}\backslash Y \subseteq Y' \subseteq \overline{Y}$.

(ii) $Y' \neq \emptyset \;\Leftrightarrow\; Y$ is non-closed or not discrete (as a subspace).

(iii) $\overline{Y} = Y \cup Y'$. Thus Y is closed if and only if $Y' \subseteq Y$.

(iv) Y is dense in itself if and only if $Y \subseteq Y'$.

(v) Y is perfect if and only if $Y = Y'$.

Definition 2.7.20 *A topological space X is scattered if every subspace has an isolated point. Equivalently, no subset of X is dense-in-itself.*

Example 2.7.21 1. Obviously every discrete space is scattered.

2. The set $\{1/n \mid n \in \mathbb{N}\} \cup \{0\} \subseteq \mathbb{R}$ with the topology inherited from $\mathbb{R}$ is scattered, but not discrete (since 0 is not isolated). □

Exercise 2.7.22 Let X be a topological space. Prove:

(i) If $Y \subseteq X$ is dense-in-itself then the same holds for $\overline{Y}$.

(ii) If $Y_i \subseteq X$ is dense-in-itself for all $i \in I$ then the same holds for $\bigcup_i Y_i$.

(iii) There are a perfect subset Y and a scattered subset Z such that $X = Y \cup Z$ and $Y \cap Z = \emptyset$.

The following notions are used only incidentally, cf. e.g. Exercises 4.1.19, 7.2.7 and Proposition 7.7.6:

Definition 2.7.23 *Let X be a topological space and $Y \subseteq X$ a subset. A point $x \in X$ is called*

- *ω-accumulation point of Y if $U \cap Y$ is infinite for every open neighborhood U of x,*
- *condensation point of Y if $U \cap Y$ is uncountable for every open neighborhood U of x.*
- *complete accumulation point of Y if $\#(U \cap Y) = \#Y$ (i.e. Y and $U \cap Y$ have the same cardinalities) for every open neighborhood U of x.*

We denote by Y^ω (Y^{cd}, Y^{cpl}) the sets of ω-accumulation (condensation, complete accumulation) points of Y.

We obviously have $Y^{\mathrm{cd}} \subseteq Y^\omega \subseteq Y' \subseteq \overline{Y}$, and for uncountable Y we have $Y^{\mathrm{cpl}} \subseteq Y^{\mathrm{cd}}$.

Exercise 2.7.24 Prove: If X is a T_1-space then every accumulation point of $Y \subseteq X$ is an ω-accumulation point, i.e. $Y^\omega = Y'$.

Exercise 2.7.25 Let X be a topological space and $A, B \subseteq X$. Prove:

(i) $(A \cup B)^\omega = A^\omega \cup B^\omega$ and $(A \cup B)^{\mathrm{cd}} = A^{\mathrm{cd}} \cup B^{\mathrm{cd}}$.

(ii) A^ω and A^{cd} are closed.

Remark 2.7.26 Let X be the space from Example 2.7.21.2. Then $X^\omega = \{0\}$ and $X^{\mathrm{cd}} = \emptyset$. Thus $(X^\omega)^\omega = (X^{\mathrm{cd}})^{\mathrm{cd}} = \emptyset$, showing that $(X^\omega)^\omega = X^\omega$ need not hold. Here $(X^{\mathrm{cd}})^{\mathrm{cd}} = X^{\mathrm{cd}}$ does hold, if trivially. For more on this, cf. Exercise 4.1.19. $\square$

2.8 Some more exotic types of spaces

2.8.1 ⋆ Irreducible spaces

Exercise 2.8.1 Let X be a topological space. Show that the following are equivalent:

(i) If $C, D \subseteq X$ are closed and $X = C \cup D$ then $C = X$ or $D = X$.

(ii) If $U, V \subseteq X$ are non-empty open sets then $U \cap V \neq \emptyset$.

(iii) Every non-empty open $U \subseteq X$ is dense.

Definition 2.8.2 *A space satisfying these equivalent conditions is called irreducible. Otherwise, i.e. if there are two disjoint non-empty open sets, it is called reducible.*

Exercise 2.8.3 Prove:

(i) Every irreducible space is connected.

(ii) An irreducible space with more than one point is never Hausdorff.

(iii) If $\#X = \infty$ then $(X, \tau_{\mathrm{cofin}})$ is irreducible.

(iv) If X is uncountable then (X, τ_{cocnt}) is irreducible.

Remark 2.8.4 1. Irreducible spaces play an important rôle in modern algebraic geometry. (In [107], they appear on page 3. The Zariski topologies are irreducible.)

2. In view of (i), irreducible spaces are also called hyperconnected.

3. $\mathbb{R}$ (with the usual topology) is T_2 and, as we will prove later, connected. Thus: connected $\not\Rightarrow$ irreducible.

4. (iii),(iv) show that an irreducible space can be T_1. □

2.8.2 T_0-spaces

In Section 2.5 we have defined T_1-spaces as spaces in which every point (more precisely: every singleton) is closed. So far, the only non-T_1-spaces that we have met were the – rather uninteresting – indiscrete spaces. But there actually are spaces 'in nature' (for example in algebraic geometry) that are not T_1:

Example 2.8.5 Consider $X = \{x, y\}$, $\tau = \{\emptyset, \{x\}, X\}$. X is irreducible, thus connected. Since $\{x\}$ is open, $\{y\} = X\backslash\{x\}$ is closed, but $\{x\}$ is not closed. In fact $\overline{\{x\}} = X$, so that X is not T_1. (Thus Exercise 2.5.9 does not apply. Indeed x is isolated.) Yet we have $\overline{\{x\}} \neq \overline{\{y\}}$, thus the points $x \neq y$ are distinguished by their closures. □

Exercise 2.8.6 Let (X, τ) be a topological space. Prove that the following are equivalent:

(i) Given $x, y \in X$, $x \neq y$, there is a $U \in \tau$ containing precisely one of the two points. (I.e. all points are distinguished by τ.)

(ii) If $x \neq y$ then $\overline{\{x\}} \neq \overline{\{y\}}$.

Definition 2.8.7 *A topological space is called T_0-space if it satisfies the equivalent characterizations in Exercise 2.8.6.*

Obviously, $T_1 \Rightarrow T_0$. The space in Example 2.8.5 is T_0, but not T_1. A very important class of non-trivial T_0-spaces is discussed in Appendix B. The point x in Example 2.8.5 is an example for the following:

Exercise 2.8.8 Let (X, τ) be a topological space and $x \in X$. Prove that $\overline{\{x\}} = X$ holds if and only if x is contained in every non-empty open set.

A point with these equivalent properties is called generic point.

Exercise 2.8.9 Given a topological space (X, τ), define a relation on X by $x \leq_\tau y \Leftrightarrow x \in \overline{\{y\}}$. Prove:

(i) $\leq_\tau$ is a reflexive and transitive (thus a preorder, called the specialization preorder).

(ii) τ is T_0 if and only if $\leq_\tau$ is also antisymmetric (thus a partial order).

(iii) τ is T_1 if and only if $\leq_\tau$ is trivial in the sense of $x \leq_\tau y \Leftrightarrow x = y$.

(iv) τ is indiscrete if and only if $x \leq_\tau y$ for all x, y.

Spaces that are not even T_0 can arise from pseudometrics:

Exercise 2.8.10 Let d be a pseudometric on a set X. Prove:

(i) τ_d is the indiscrete topology if and only if $d \equiv 0$.

(ii) τ_d is T_0 if and only if d is a metric.

(Thus d is a metric $\Rightarrow \tau_d$ is $T_2 \Rightarrow T_1 \Rightarrow T_0 \Rightarrow d$ is a metric.)

2.8.3 ⋆⋆ Alexandrov or Smallest Neighborhood Spaces

As remarked earlier, intersections of infinitely many open subsets need not be open. But there clearly are spaces where this is true, to wit the discrete spaces (for the trivial reason that every subset is open). But there are more interesting examples, and this subsection is devoted to a quick look at them.

Exercise 2.8.11 Let (X, τ) be a topological space. Prove that the following are equivalent:

(i) Every union of closed subsets is closed.

(ii) Every intersection of open subsets is open.

(iii) Every $x \in X$ has a smallest open neighborhood, i.e. an open neighborhood U_x contained in every open set that contains x.

Definition 2.8.12 *Topological spaces with the equivalent properties from Exercise 2.8.11 are called smallest neighborhood spaces or Alexandrov[3] spaces[4].*

Lemma 2.8.13 *(i) Every discrete space is a smallest neighborhood space.*

(ii) Every smallest neighborhood T_1-space is discrete.

(iii) Every finite topological space is a smallest neighborhood space.

Proof. (i) Follows from Exercise 2.8.11(i) since every subset is closed.

(ii) In a T_1 space every singleton is closed, thus with Exercise 2.8.11(i) every subset is closed.

(iii) If a set X is a finite then so is every topology τ on it, thus every intersection of open sets is open. ■

Proposition 2.8.14 *Let X be a set.*

(i) For a preorder $\leq$ on X, define

$$\tau_{\leq} = \{U \subseteq X \mid x \in U,\ y \in X,\ x \leq y \ \Rightarrow \ y \in U\}.$$

(I.e. $\tau_{\leq}$ is the set of subsets of X that are upward-closed.) Then $\tau_{\leq}$ is a topology on X with the smallest neighborhood property.

(ii) For $x \in X$, the set $\overline{M}(x) = \{y \in X \mid y \geq x\}$ of (non-strict) majorants of x is in $\tau_{\leq}$. For every $U \in \tau_{\leq}$ we have $U = \bigcup_{x \in U} \overline{M}(x)$.

---

[3]Pavel Sergeevich Alexandrov (1896-1982), Russian mathematician. (Also transliterated as Aleksandrov or Alexandroff.) We will also encounter the Alexandrov compactification.

[4]Not to be confused with the Alexandrov spaces in metric geometry (cf. e.g. [39]) introduced by Aleksandr Danilovich Aleksandrov (1912-1999). For this reason, we prefer to write 'smallest neighborhood space'.

(iii) The specialization preorder (Exercise 2.8.9) $\leq_{\tau_\leq}$ on X arising from $\tau_\leq$ coincides with $\leq$.

(iv) If (X,τ) is a smallest neighborhood space and $\leq_\tau$ is the specialization preorder arising from τ then $\tau_{\leq_\tau} = \tau$.

(v) For every X there is a bijection between preorders and smallest neighborhood topologies on X. For corresponding $\tau, \leq$ we have that $\leq$ is trivial (i.e. $\{(x,x)\}$) if and only if τ is T_1 if and only if τ is discrete. And τ is T_0 if and only if $\leq$ is partial order.

Proof. (i) Clearly $\emptyset, X \in \tau_\leq$. Let $U = \bigcup_{i\in I} U_i$, where $U_i \in \tau_\leq\ \forall i$. If $x \in U, y \in X, x \leq y$ then $x \in U_i$ for some $i \in I$. But then $y \in U_i$, thus $y \in U$. Let $U = \bigcap_{i\in I} U_i$, where $U_i \in \tau_\leq\ \forall i$. Assume $x \in U, y \in X, x \leq y$. Then $x \in U_i\ \forall i$, thus also $y \in U_i\ \forall i$ and thus $y \in U$. Since this is true for *all* intersections, $\tau_\leq$ is a smallest neighborhood topology.

(ii) For the first claim it suffices to observe that $\overline{M}(x)$ is upward-closed. The second claim follows from the fact that $x \in U \in \tau_\leq$ implies $x \in \overline{M}(x) \subseteq U$.

(iii) By Lemma 2.7.3, we have $x \in \overline{\{y\}}$ if and only $y \in U$ holds for every $\tau_\leq$-open set $U \ni x$. Since $\overline{M}(x)$ is among the latter, $x \in \overline{\{y\}}$ implies $x \leq_{\tau_\leq} y$. Conversely, $x \in U \in \tau_\leq$ implies $\overline{M}(x) \subseteq U$, thus y is contained in every open set containing x. This proves that the specialization preorder $\leq_{\tau_\leq}$ of $\tau_\leq$ coincides with $\leq$.

(iv) Recall that since (X,τ) is a smallest neighborhood space, each x has a smallest open neighborhood U_x, and for each $U \in \tau$ we have $U = \bigcup_{x\in U} U_x$. As for (iii) we use that $x \in \overline{\{y\}}$ if and only if $x \in U \in \tau$ implies $y \in U$. This gives that $x \in \overline{\{y\}}$ is equivalent to $y \in U_x$. Combining this with the definition of $\leq_\tau$, we have

$$x \leq_\tau y \quad \Leftrightarrow \quad x \in \overline{\{y\}} \quad \leq \quad y \in U_x.$$

This implies that $U_x = \overline{M}(x)$, the majorant set (w.r.t. $\leq_\tau$) of x. Now $\tau = \tau_{\leq_\tau}$ follows from

$$U \in \tau \quad \Leftrightarrow \quad U = \bigcup_{x\in X} U_x \quad \Leftrightarrow \quad U = \bigcup_{x\in X} \overline{M}(x) \quad \Leftrightarrow \quad U \in \tau_{\leq_\tau},$$

where the last identity is due to (ii).

(v) This is an immediate consequence of (i)+(iii)+(iv), Exercise 2.8.9(ii)-(iii) and our earlier observation that a smallest neighborhood space is T_1 if and only if it is discrete. ■

It is natural to ask what continuity of a map between smallest neighborhood spaces means in terms of the corresponding preorders. In this discussion we use the notion of a base (Definition 4.1.1) and of a continuous function (Definition 5.2.7).

Proposition 2.8.15 *Let $(X,\tau), (Y,\tau')$ be smallest neighborhood spaces and $\leq, \leq'$ the associated (specialization) preorders. Then a function $f: X \to Y$ is continuous if and only if it is order-preserving, i.e. $x \leq x' \Rightarrow f(x) \leq' f(x')$.*

Proof. Continuity of f means $U \in \tau' \Rightarrow f^{-1}(U) \in \tau$. Since the majorant sets $\overline{M}(y)$ form a base for τ', we only need to check whether $f^{-1}(\overline{M}(y)) \in \tau$ for each $y \in Y$, cf. Exercise 5.2.8(iii). Now, $f^{-1}(\overline{M}(y)) = \{x \in X \mid f(x) \geq' y\}$. This set is in τ if and only if it is upward closed. Thus continuity of f is equivalent to $x, x' \in X, y \in Y, f(x) \geq' y, x' \geq x \Rightarrow f(x') \geq' y$. If this is true then taking $y = f(x)$ we obtain the implication $x, x' \in X, x' \geq x \Rightarrow f(x') \geq' f(x)$. Conversely, if the latter true then we also have $x' \geq x, f(x) \geq' y \Rightarrow f(x') \geq' f(x) \geq' y$. Thus continuity of $f: (X,\tau) \to (Y,\tau')$ is equivalent to $f: (X,\leq) \to (Y,\leq')$ being order-preserving. ■

Corollary 2.8.16 *The category of smallest neighborhood spaces and continuous maps is isomorphic to the category of preordered sets and order-preserving maps. This restricts to an isomorphism between the categories of smallest neighborhood T_0-spaces and partially ordered sets.*

Thus the study of smallest neighborhood spaces reduces to a branch of order theory! For more on smallest neighborhood spaces see [6], where also several other characterizations of these spaces are given, and [220] for the special case of finite spaces. For an important area of application see [113].

Chapter 3

Metric spaces: Completeness and its applications

3.1 Completeness

Definition 3.1.1 *A sequence $\{x_n\}$ in a metric space (X, d) is a Cauchy*[1] *sequence if for every $\varepsilon > 0$ there is an $N \in \mathbb{N}$ such that $n, m \geq N \Rightarrow d(x_n, x_m) < \varepsilon$.*

Lemma 3.1.2 *Every convergent sequence in a metric space is a Cauchy sequence.*

Proof. Assume $x_n \to z$. If $\varepsilon > 0$ then there is $N \in \mathbb{N}$ such that $n \geq N \Rightarrow d(x_n, z) < \varepsilon/2$. If now $n, m \geq N$ then $d(x_n, x_m) \leq d(x_n, z) + d(z, x_m) < \varepsilon/2 + \varepsilon/2 = \varepsilon$. ■

The converse is not true: If $X = (0, 1] \cap \mathbb{Q}$ with $d(x, y) = |x - y|$ then $\{x_n = 1/n\}$ is a Cauchy sequence, but it does not converge in X. Less trivially, $\mathbb{Q}$ is not complete w.r.t. the same metric d as above (nor w.r.t. the p-adic metrics $d_p(x, y) = \|x - y\|_p$). This motivates:

Definition 3.1.3 *If d is a metric on X such that every Cauchy sequence converges, both the metric and the metric space (X, d) are called complete.*

Definition 3.1.4 *A Banach space*[2] *is a normed space $(V, \|\cdot\|)$ such that the metric space $(X, d_{\|})$ is complete.*

We assume as known from a course on calculus/analysis that $(\mathbb{R}, d)$, where $d(x, y) = |x - y|$ is complete.

Lemma 3.1.5 *Let $p \in [1, \infty]$ and $d_p(x, y) = \|x - y\|_p$. Then $(\mathbb{R}^d, d_p)$ is complete for every $d \in \mathbb{N}$.*

Proof. Let $p \in [1, \infty]$. From the definition of $\|\cdot\|_p$ it is clear that for $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$ and $1 \leq i \leq d$ we have $|x_i| \leq \|x\|_p$. Thus if a sequence $\{x^n\}$ is Cauchy we have $|x_i^n - x_i^m| \leq \|x^n - x^m\|_p$, so that the sequence $\{x_i^n\}_n$ in $\mathbb{R}$ is Cauchy. Since $\mathbb{R}$ is complete, we have $x_i^n \stackrel{n\to\infty}{\longrightarrow} y_i \in \mathbb{R}$. Since $\|x\|_p = \|(x_1, \ldots, x_d)\|_p$ depends continuously on $x_1, \ldots, x_d$, this implies $\|x^n - y\|_p \to 0$. ■

(For a generalization to infinite dimensions, cf. Section E.)

The main reason for the importance of completeness is the following: Given a sequence $\{x_n\}$ in a metric space, it is usually much easier to show that it is Cauchy and then invoking completeness

[1] Augustin-Louis Cauchy (1789-1857). French mathematician and pioneer of rigorous analysis.
[2] Stefan Banach (1892-1945). Polish mathematician and pioneer of functional analysis. Also known for B. algebras, B.'s contraction principle, the B.-Tarski paradox and the Hahn-B. and B.-Steinhaus theorems, etc.

than proving convergence directly, which requires already knowing the limit. An example for this strategy is the proof of Banach's contraction principle (Theorem D.1.2), probably known from a course in analysis. But often completeness is used indirectly via its consequences that don't involve Cauchy sequences in their statements, like Cantor's intersection theorem (see the exercise below) and Baire's theorem (cf. Theorem 3.3.1). In this book, Cantor's theorem will be used for the extension of continuous functions between metric spaces (Section 3.4.2) and for the results related to Ekeland's Variational Principle and Caristi's Fixed Point Theorem D.2.2, which in turn are used to prove Menger's Theorem 12.4.8. Some applications of Baire's theorem will be discussed in Section 3.3, others in Appendix G.2.

Exercise 3.1.6 (Cantor's Intersection Theorem) [3] Let (X,d) be a metric space. Prove:

(i) If $\{C_n\}_{n\in\mathbb{N}}$ are sets satisfying $X \supseteq C_1 \supseteq C_2 \supseteq \cdots$ and $\lim_{n\to\infty} \mathrm{diam}(C_n) = 0$ then $\bigcap_n C_n$ contains at most one point.

(ii) If (X,d) is complete and $\{C_n\}_{n\in\mathbb{N}}$ are non-empty closed sets satisfying $X \supseteq C_1 \supseteq C_2 \supseteq \cdots$ and $\lim_{n\to\infty} \mathrm{diam}(C_n) = 0$ then $\bigcap_n C_n$ is non-empty (thus a singleton by (i)).

(iii) Assume that $\bigcap_n C_n \neq \emptyset$ for every family $\{C_n\}$ as in (ii). Then (X,d) is complete.

(For $C_n = (0, 1/n) \subseteq \mathbb{R}$ we have $\bigcap_n C_n = \emptyset$. Thus closedness of the C_n cannot be omitted.)

Lemma 3.1.7 *Let (X,d) be a metric space.*

(i) If (X,d) is complete and $Y \subseteq X$ is closed then (Y,d) is complete.

(ii) If $Y \subseteq X$ is such that (Y,d) is complete then $Y \subseteq X$ is closed.

Proof. (i) Let $Y \subseteq X$ be closed. A Cauchy sequence $\{y_n\}$ in Y clearly is a Cauchy sequence in X. By completeness of X we have $y_n \to x$ for some $x \in X$. The closedness of Y implies $x \in Y$. Thus Y is complete.

(ii) Let $x \in \overline{Y}$. By definition of the closure, there is a sequence $\{y_n\}$ in Y that converges to x. A convergent sequence is Cauchy, and since $\{y_n\} \subseteq Y$ and (Y,d) is complete, there is a $y \in Y$ such that $y_n \to y$. By uniqueness of the limit, $x = y$. Thus $\overline{Y} \subseteq Y$, and Y is closed. ■

In view of the above it is clear that completeness is not hereditary. (One could say it is 'closed-hereditary'.) (While non-closed subsets of complete metric spaces are not complete, we will see that open subsets are 'completely metrizable', cf. Proposition 3.4.18.)

Proposition 3.1.8 *Let $(X,d),(Y,d')$ be metric spaces. Let D be as in (2.6). Then: (Y,d') is complete $\Leftrightarrow$ $(B(X,Y),D)$ is complete $\Leftrightarrow$ $(C_b(X,Y),D)$ is complete.*

Proof. Assume that (Y,d') is complete, and let $\{f_n \in B(X,Y)\}$ be a Cauchy sequence w.r.t. the metric D. The definition of D implies that $d'(f(x),g(x)) \leq D(f,g)$ for every $x \in X$. Thus $\{f_n(x)\}$ is a Cauchy sequence in Y for every $x \in X$. By completeness of Y, $\lim_{n\to\infty} f_n(x)$ exists for every x, and we define $g(x) = \lim_n f_n(x)$. By assumption, f_n is Cauchy uniformly in x: For every $\varepsilon > 0$ there is $N \in \mathbb{N}$ such that $n,m \geq N \Rightarrow d'(f_n(x), f_m(x)) < \varepsilon$ for all $x \in X$. Letting $m \to \infty$ (for fixed x) and using continuity of d' (Exercise 2.1.4), we obtain $n \geq N \Rightarrow d'(f_n(x), g(x)) \leq \varepsilon$ for all x. This proves both $g \in B(X,Y)$ and $D(f_n, g) \to 0$. Thus $(B(X,Y),D)$ is complete.

Assume that $(B(X,Y),D)$ is complete. By Proposition 2.1.25(ii), $C_b(X,Y) \subseteq B(X,Y)$ is closed, and therefore complete by Lemma 3.1.7(i).

[3] Georg Ferdinand Ludwig Philipp Cantor (1845-1918). German mathematician. Founder of modern set theory.

Finally, assume that $(C_b(X,Y),D)$ is complete, and let $\{y_n\}$ be a Cauchy sequence in Y. Consider the constant functions $f_n : X \to Y,\ x \mapsto y_n$ for all x. Then trivially $f_n \in C_b(X,Y)$ and $D(f_n,f_m) = d'(y_n,y_m)$, thus $\{f_n\}$ is a Cauchy sequence. By completeness of $C_b(X,Y)$, there is $g \in C_b(X,Y)$ such that $D(f_n,g) \to 0$. For $x,y \in X$ we have $d'(g(x),g(y)) \le d'(g(x),f_n(x)) + d'(f_n(x),f_n(y)) + d'(f_n(y),g(y)) \le 2D(f_n,g)$ since f_n is constant. Since $D(f_n,g) \to 0$, it follows that g is constant, thus there is $y \in Y$ such that $g(x) = y$ for all $x \in X$. Now $d'(y_n,y) = D(f_n,g) \to 0$, thus $y_n \to y$, so that (Y,d') is complete. ■

Lemma 3.1.9 *(i) If $\alpha : (X,d) \to (X',d')$ is a bijective isometry then (X,d) is complete if and only if (X',d') is complete.*

(ii) The conclusion of (i) is not not true if α is not surjective or only a homeomorphism.

Proof. (i) Obvious. (ii) If (X,d) is complete and $Y \subseteq X$ is non-closed then (Y,d) is non-complete by Lemma 3.1.7(ii). The inclusion map $Y \hookrightarrow X$ is a (non-surjective) isometry, showing the first claim. By Remark 2.1.22.5 there is a homeomorphism between $(\mathbb{R},d)$ and $((-1,1),d)$, where $d(x,y) = |x-y|$. Since $(\mathbb{R},d)$ is complete but $((-1,1),d)$ is not (since $(-1,1) \subseteq \mathbb{R}$ is non-closed), we see that completeness of metric spaces is not preserved under homeomorphisms. ■

So far we have discussed completeness of a given metric d on a set X. But as we know, different metrics d_1,d_2 can give rise to the same topology, in which case they are called equivalent. This raises the question whether completeness of metrics is preserved under equivalence. In some cases, this is true:

Exercise 3.1.10 Assume that d is a complete metric on X. Prove that any equivalent metric $d' \simeq d$ constructed as in Exercise 2.2.14 is complete.

The following example shows that two metrics d_1,d_2 on a set X can be equivalent even though d_1 is complete and d_2 is not!

Example 3.1.11 By Remark 2.1.22.5, $\alpha : (\mathbb{R},d) \to ((-1,1),d),\ x \mapsto \frac{x}{1+|x|}$, where d is the distance metric, is a homeomorphism. It follows that $d'(x,y) = d(\alpha(x),\alpha(y))$ is a metric on $\mathbb{R}$ that is equivalent to d. (Cf. Exercise 2.2.13.) By construction, $\alpha : (\mathbb{R},d') \to ((-1,1),d)$ is an isometry. Since the metric d on $(-1,1)$ is not complete, it follows that the metric d' on $\mathbb{R}$ is not complete, even though d' is equivalent to the complete metric d. □

The example motivates the following definition:

Definition 3.1.12 *(i) A topological space (X,τ) is called completely metrizable if there is a complete metric d on X such that $\tau = \tau_d$.*

(ii) A metric space (X,d) is called completely metrizable if (X,τ_d) is completely metrizable. (I.e. there is a complete metric d' on X satisfying $\tau_{d'} = \tau_d$, i.e. equivalent to d.)

Given a topological space (X,τ), it is clear that a priori there are four possibilities:

1. (X,τ) is not metrizable.

2. (X,τ) is metrizable, admitting both complete and non-complete metrics.

3. (X,τ) is metrizable, but not completely metrizable.

4. (X, τ) is metrizable and *every* metric d compatible with τ is complete.

We have seen examples for the first case (e.g. any non-Hausdorff space) and for the second, cf. Example 3.1.11. Examples for the other two cases will be found in Proposition 3.3.7(iv) and in Section 7.7.3, where we will show that every compact metrizable space is in the fourth class.

For further results on complete metrizability see Sections 3.4.3 and 8.4.2.

Remark 3.1.13 Completeness is a property that a metric space has or has not. But it is not a topological notion, i.e. it makes no sense to ask whether a topological space is complete since in a topological space we have no way of defining a Cauchy sequence. There is, however, the topological notion of Čech-completeness, with which one proves that a topological space is completely metrizable if and only if it is metrizable and Čech-complete, cf. Section 8.4.2. □

3.2 Completions

Since completeness is a very desirable property of metric spaces, it is natural to ask whether a metric space can be 'made complete'. The precise formulation of this is:

Definition 3.2.1 *Let (X, d) be a metric space. A completion of (X, d) is a metric space $(\widehat{X}, \widehat{d})$ together with a map $\iota : X \to \widehat{X}$ such that:*

(i) $(\widehat{X}, \widehat{d})$ is complete.

(ii) ι is an isometry.

(iii) $\iota(X) \subseteq \widehat{X}$ is dense, thus $\overline{\iota(X)} = \widehat{X}$.

We will soon prove that every metric space has a completion. But first we show that completions are unique, in the sense that between any two completions there is an isometric bijection, in fact a unique one:

Proposition 3.2.2 (Uniqueness of completions) *If $((\widehat{X}_1, \widehat{d}_1), \iota_1)$, $((\widehat{X}_2, \widehat{d}_2), \iota_2)$ are completions of (X, d), then there is a unique isometric bijection $f : \widehat{X}_1 \to \widehat{X}_2$ such that $\iota_2 = f \circ \iota_1$.*

Proof. It is clear that we must define f on $\iota_1(X)$ by $f(\iota_1(x)) = \iota_2(x)$. We want f to be an isometry, thus continuous. Thus if $y \in \widehat{X}_1 \backslash \iota_1(X)$ and $\{x_n\}$ is a sequence in X such that $\iota_1(x_n) \to y$, we must have $f(y) = f(\lim_n \iota_1(x_n)) = \lim_n f(\iota_1(x_n)) = \lim_n \iota_2(x_n)$. We show that the limit on the r.h.s. exists: Since the sequence $\{\iota_1(x_n)\}$ converges, it is a Cauchy-sequence, and since ι_1 is an isometry, $\{x_n\}$ is a Cauchy-sequence in X. Since ι_2 is an isometry, $\{\iota_2(x_n)\}$ is a Cauchy-sequence in $\widehat{X}_2$, and by completeness it converges to some $z \in \widehat{X}_2$. In order to define $f(y) = z$, one must show that z depends only on y, but not on the choice of the sequence $\{x_n\}$. So let $\{x'_n\}$ be another sequence in X such that $\iota_1(x'_n) \to y$. But this means that $\widehat{d}_1(\iota_1(x_n), \iota_1(x'_n)) \to 0$, and therefore $\widehat{d}_2(\iota_2(x_n), \iota_2(x'_n)) \to 0$, since ι_1, ι_2 are isometries. This implies $\lim_n \iota_2(x'_n) = \lim_n \iota_2(x_n) = z$ and thus f is well-defined. The above reasoning also shows that this $f : \widehat{X}_1 \to \widehat{X}_2$ is uniquely determined by the requirements of continuity and $f \circ \iota_1 = \iota_2$. Now let $y, y' \in \widehat{X}_1$ and let $\{x_n\}, \{x'_n\}$ be sequences in X such that $\iota_1(x_n) \to y, \iota_1(x'_n) \to y'$. By the definition of f, we have $f(y) = \lim_n \iota_2(x_n)$, $f(y') = \lim_n \iota_2(x'_n)$ and

$$\widehat{d}_1(y, z) = \lim_n \widehat{d}_1(\iota_1(x_n), \iota_1(x'_n)) = \lim_n d(x_n, x'_n) = \lim_n \widehat{d}_2(\iota_2(x_n), \iota_2(x'_n)) = \widehat{d}_2(f(y), f(z)),$$

proving that $f : \widehat{X}_1 \to \widehat{X}_2$ is an isometry. It remains to show that f is surjective. This can be done in two ways: Reversing the rôles of $\widehat{X}_1, \widehat{X}_2$, the above gives an isometry $g : \widehat{X}_2 \to \widehat{X}_1$ such that $g \circ \iota_2 = \iota_1$. Now the function $f \circ g : \widehat{X}_2 \to \widehat{X}_2$ is the identity map on $\iota_2(X) \subseteq \widehat{X}_2$, and continuity of f, g and density of $\iota_2(X) \subseteq \widehat{X}_2$ imply that $f \circ g = \mathrm{id}_{\widehat{X}_2}$, so that f is surjective. (In fact, f and g are inverse functions of each other.)

Alternatively, observe that $f : \widehat{X}_1 \to \widehat{X}_2$ is an isometry, thus injective, so that $f : \widehat{X}_1 \to f(\widehat{X}_1)$ is an isometric bijection. This implies that the metric subspace $(f(\widehat{X}_1), \widehat{d}_2)$ of $(\widehat{X}_2, \widehat{d}_2)$ is complete. Thus by Lemma 3.1.7(ii), $f(\widehat{X}_1) \subseteq \widehat{X}_2$ is closed. Since $f(\widehat{X}_1)$ contains $\iota_2(X)$, which is dense in $\widehat{X}_2$, we have $f(\widehat{X}_1) = \overline{f(\widehat{X}_1)} \supseteq \overline{\iota_2(X)} = \widehat{X}_2$, so that f is surjective. ■

Remark 3.2.3 1. Since every metric space can be isometrically embedded into a bigger one, uniqueness of completions clearly wouldn't hold without requirement (iii) in Definition 3.2.1.

2. If we drop the requirement $\iota_2 = f \circ \iota_1$ in Proposition 3.2.2, there may be many different isometries between different completions of a given metric space (X, d).

3. The identity $\iota_2 = f \circ \iota_1$ in the proposition can be stated by saying that "the diagram

$$\begin{tikzcd} X \arrow[r, "\iota_1"] \arrow[dr, "\iota_2"'] & \widehat{X}_1 \arrow[d, "f"] \\ & \widehat{X}_2 \end{tikzcd}$$

commutes". This may seem a tremendous waste of space, but in more complicated situations, involving more sets and functions, commutative diagrams often greatly clarify what is going on. (See e.g. (6.8).) □

Corollary 3.2.4 *Let (X, d) be a complete metric space and $Y \subseteq X$. Then the (unique) completion of (Y, d) is given by $(\overline{Y}, d)$.*

Proof. By Lemma 3.1.7, (Y, d) is complete if and only if Y is closed. Thus $(\overline{Y}, d)$ is complete, and by uniqueness of completions, this is *the* completion of (Y, d). ■

E.g., if $d(x, y) = |x - y|$ then both $((0, 1), d)$ and $((0, 1], d)$ have $([0, 1], d)$ as completion.

Corollary 3.2.4 clearly only helps if we have an isometric embedding of (Y, d) into a complete metric space. To prove that every metric space has a completion, one needs to work harder. Usually this is done using the set of Cauchy sequences in X, generalizing the classical completion of $\mathbb{R}$. Since this construction can be found in any number of references, we follow a different (and perhaps more elegant) route. (Cf. e.g. [180, 204].)

Theorem 3.2.5 *Every metric space has a completion.*

Proof. Given a metric space (X, d), the idea is to find a complete metric space (Y, D) and an isometry $\iota : (X, d) \to (Y, D)$. Then $((\overline{\iota(X)}, D), \iota)$ is a completion of (X, d) by Corollary 3.2.4. Since $\mathbb{R}$ with the standard metric is complete, $(C_b(X, \mathbb{R}), D)$ is complete by Proposition 3.1.8. It remains to construct an isometry $\iota : X \to C_b(X, \mathbb{R})$.

Pick $x_0 \in X$ once and for all. For $x \in X$, the function $f_x : X \to \mathbb{R},\ z \mapsto d(z, x) - d(z, x_0)$ is continuous and bounded by $d(x, x_0)$, cf. (2.1). Thus $f_x \in C_b(X, \mathbb{R})$. This allows us to define

$\iota : X \to C_b(X, \mathbb{R}),\ x \mapsto f_x$. With the metric D on $C_b(X, \mathbb{R})$ defined in (2.6) we have

$$D(\iota(x), \iota(y)) = \sup_{z \in X} |(d(z, x) - d(z, x_0)) - (d(z, y) - d(z, x_0))| = \sup_{z \in X} |d(z, x) - d(z, y)| = d(x, y),$$

where the final identity is (2.3). Thus $\iota : X \to C_b(X, \mathbb{R})$ is an isometry. ■

Remark 3.2.6 1. We briefly sketch another (and somewhat more common) method of constructing a (thus the) completion: Let $\widetilde{X} \subseteq \mathrm{Fun}(\mathbb{N}, X)$ be the set of all Cauchy sequences in (X, d). If $\{x_i\}, \{y_i\} \in \widetilde{X}$, then the inequality $|d(x_i, y_i) - d(x_j, y_j)| \le d(x_i, x_j) + d(y_i, y_j)$ implies that $\{d(x_i, y_i)\}$ is a Cauchy sequence in $\mathbb{R}$. The latter converges since $\mathbb{R}$ is complete, and we define $\widetilde{d}(\{x_i\}, \{y_i\}) = \lim_i d(x_i, y_i)$. One checks that this is a pseudometric. As in Exercise 2.1.7, defining $\widehat{X} = \widetilde{X}/\sim$ one obtains a (true) metric $\widehat{d}$ on $\widehat{X}$. Now a diagonal argument shows that $(\widehat{X}, \widehat{d})$ is complete. (A slight modification of the above reasoning also produces $\mathbb{R}$ as the completion of $X = \mathbb{Q}, d(x, y) = |x - y|$.)

2. The construction of $(\widehat{X}, \widehat{d})$ via $C_b(X, \mathbb{R})$ surely is more elegant, but has its own drawbacks: It is less economic in that it begins with $\mathrm{Fun}(X, \mathbb{R})$ instead of $\mathrm{Fun}(\mathbb{N}, X)$. More importantly, since it assumes the metric space $(\mathbb{R}, d)$ as given, it clearly cannot be used to construct $\mathbb{R}$ as the completion of $\mathbb{Q}$. But, at least in the author's view, the construction of $\mathbb{R}$ in terms of Dedekind sections is preferable anyway. Cf. e.g. [201, App. to Chap. 1].

3. If $(V, \|\cdot\|)$ is a normed space, it is easy to show that the completion of the metric space $(V, d_{\|\|})$ again comes from a normed space $\overline{(V, \|\cdot\|)}$, which then is a Banach space. □

Now we can give an interesting and useful characterization of complete metric spaces:

Proposition 3.2.7 *A metric space (X, d) is complete if and only if $\iota(X) \subseteq Y$ is closed whenever (Y, d) is a metric space and $\iota : X \to Y$ is an isometry. (One says (X, d) is universally closed.)*

Proof. Assume (X, d) is complete. If $\iota : (X, d) \to (Y, d')$ is an isometry then $(\iota(X), d') \subseteq (Y, d')$ is isometric to (X, d). Thus $(\iota(X), d')$ is complete, thus $\iota(X) \subseteq Y$ is closed by Lemma 3.1.7(ii).

Assume (X, d) is not complete, and let $((\widehat{X}, \widehat{d}), \iota)$ be a completion. Then (X, d) is isometric to the subspace $\iota(X) \subseteq \widehat{X}$. If $\iota(X) \subseteq \widehat{X}$ was closed, $(\iota(X), \widehat{d}) \cong (X, d)$ would be complete by Lemma 3.1.7(i), contradicting the assumption. Thus $\iota(X) \subseteq \widehat{X}$ is not closed. ■

Remark 3.2.8 There is a class of spaces sitting properly between metric and topological spaces, called uniform spaces. A uniform space is a pair $(X, \mathcal{U})$ where X is a set and the uniform structure (or uniformity) $\mathcal{U} \subseteq P(X \times X)$ (as opposed to $\tau \subseteq P(X)$ for a topology) satisfies certain axioms. Now, every metric d on X gives rise to a uniformity $\mathcal{U}_d$ on X, and every uniformity $\mathcal{U}$ defines a topology $\tau_{\mathcal{U}}$. (Of course $\tau_{\mathcal{U}_d} = \tau_d$.) In uniform spaces one can define a notion of Cauchy sequence and therefore also the property of completeness. Every uniform space has a completion. For more on uniform spaces see e.g. [130, 237, 75].

This being said, it seems that the applications of uniform spaces outside topology proper are quite few, the most important being to topological groups: Every topological group has two canonical uniform structures, so one can consider their completeness (in the senses of Weyl and Raikov) and completions. Cf. [8] and the papers by Comfort in [152] and Tkachenko in [11, Vol.3]. □

3.3 Baire's theorem for complete metric spaces. G_δ-sets

3.3.1 Baire's theorem

Recall that a finite intersection of dense open sets is dense (Corollary 2.7.11) in every topological space, but that this need not be true for infinitely many dense open sets (Remark 2.7.12).

Theorem 3.3.1 [4] *Let (X, d) be a complete metric space and $\{U_n\}_{n\in\mathbb{N}}$ a countable family of dense open subsets. Then $\bigcap_{n=1}^\infty U_n$ is dense.*

Proof. Let $\emptyset \neq W \in \tau$. Since U_1 is dense, $W \cap U_1 \neq \emptyset$ by Lemma 2.7.9, so we can pick $x_1 \in W \cap U_1$. Since $W \cap U_1$ is open, we can choose $\varepsilon_1 > 0$ such that $\overline{B(x_1, \varepsilon_1)} \subseteq W \cap U_1$. We may also assume $\varepsilon_1 < 1$. Since U_2 is dense, $U_2 \cap B(x_1, \varepsilon_1) \neq \emptyset$ and we pick $x_2 \in U_2 \cap B(x_1, \varepsilon_1)$. By openness, we can pick $\varepsilon_2 \in (0, 1/2)$ such that $\overline{B(x_2, \varepsilon_2)} \subseteq U_2 \cap B(x_1, \varepsilon_1)$. Continuing this iteratively, we find points x_n and $\varepsilon_n \in (0, 1/n)$ such that $\overline{B(x_n, \varepsilon_n)} \subseteq U_n \cap B(x_{n-1}, \varepsilon_{n-1})$ $\forall n$. If $i > n$ and $j > n$ we have by construction that $x_i, x_j \in B(x_n, \varepsilon_n)$ and thus $d(x_i, x_j) \leq 2/n$. Thus $\{x_n\}$ is a Cauchy sequence, and by completeness it converges to some $z \in X$. Since $n > k \Rightarrow x_n \in \overline{B(x_k, \varepsilon_k)}$, the limit z is contained in $\overline{B(x_k, \varepsilon_k)}$ for each k, thus

$$z \in \bigcap_n \overline{B(x_n, \varepsilon_n)} \subseteq W \cap \bigcap_n U_n,$$

thus $W \cap \bigcap_n U_n$ is non-empty. Since W was an arbitrary non-empty open set, Lemma 2.7.9 gives that $\bigcap_n U_n$ is dense. ■

Remark 3.3.2 1. Baire's Theorem 3.3.1 and its reformulation in Proposition 3.3.5 have countlessly many applications, cf. the overviews [136, 238]. Some of the best known are in functional analysis: The Open Mapping Theorem, the Closed Graph Theorem and the Uniform Boundedness Theorem of Banach and Steinhaus. (Proofs of these are given in Appendix G.2, where we also use them to construct a dense set of periodic continuous functions whose Fourier series diverges at a given point.) Here we consider three applications, to metric topology (Propositions 3.3.7 and 3.3.13), to 'Croft's lemma' in classical analysis, cf. Subsection 3.3.4, and to constructing a dense set of continuous functions that are nowhere differentiable (Theorem 3.3.18).

2. Notice the similarity to the proof of Cantor's Intersection Theorem (Exercise 3.1.6(ii): We produce a Cauchy sequence and use completeness to conclude that the latter converges, which immediately proves that a certain set (namely $\bigcap_i C_i$ and $W \cap \bigcap_n U_n$, respectively) is non-empty. The details are a bit more involved in the case of Baire's theorem, but the only conceptual difference is that for Cantor's theorem we need the axiom of countable choice, whereas Baire's theorem requires a bit more, cf. Section 3.3.2. In view of this observation, the reputation of Baire's theorem (and its applications) of being difficult or deep seems exaggerated.

3. The conclusion of Baire's theorem makes no reference to a metric and therefore makes sense for general topological spaces. This motivates the next definition. □

Definition 3.3.3 *A topological space is a Baire space if every countable intersection of dense open sets is dense.*

Corollary 3.3.4 *Every completely metrizable space is a Baire space.*

[4] René-Louis Baire (1874-1932), French mathematician, proved this for $\mathbb{R}^n$ in his 1899 doctoral thesis. The generalization is due to Hausdorff (1914).

In Section 8.4.1 we will encounter a large class of not necessarily metrizable Baire spaces, the Čech-complete spaces, which contains all (locally) compact Hausdorff spaces.

The following equivalent formulations of the Baire property are often used:

Proposition 3.3.5 *For a topological space X, the following are equivalent:*

(i) X is a Baire space.

(ii) If $C_n \subseteq X$ is closed with empty interior for each $n \in \mathbb{N}$ then $\bigcup_{n=1}^{\infty} C_n$ has empty interior.

(iii) Every countable union of nowhere dense subsets of X has empty interior.

Proof. The equivalence (i)⇔(ii) is seen by taking complements. Since closed sets with empty interior are nowhere dense, we have (iii)⇒(ii). Now assume (ii) and let $\{C_n\}$ be nowhere dense sets. Then the sets $D_n = \overline{C_n}$ are closed and have empty interior, thus (ii) implies that $\bigcup_n D_n$ has empty interior. But then clearly also $\bigcup_n C_n \subseteq \bigcup_n D_n$ has empty interior. ■

Remark 3.3.6 There is a considerable amount of additional terminology around Baire's theorem. E.g., $Y \subseteq X$ is called <u>meager</u> if Y is a countable union of nowhere dense sets. The Baire property then amounts to the statement that a meager set has empty interior. Meager sets are also called sets 'of first category', all other sets being called 'of second category' (whence the name 'Baire category theorem'). In the author's opinion, the latter pair of terms should be avoided since it is unhelpful and potentially confusing in view of other uses of 'category' in mathematics. □

Here is a first application of Theorem 3.3.1 and, more generally, the Baire property:

Proposition 3.3.7 *(i) If (X, τ) is non-empty, has no isolated points, is T_1 and Baire then X is uncountable.*

(ii) A non-empty complete metric space without isolated points is uncountable.

(iii) A space that is countably infinite without isolated points is not completely metrizable.

(iv) $\mathbb{Q}^n$ with the Euclidean topology inherited from $\mathbb{R}^n$ is not completely metrizable.

Proof. (i) Since X is T_1, every $\{x\}$ is closed, thus $U_x = X\backslash\{x\}$ is open. Since X has no isolated points, $\{x\}$ is not open, thus U_x is non-closed. The only subset of X properly containing U_x is X, thus U_x is dense. Assuming that X is countable, the Baire property implies denseness of $\bigcap_{x\in X} U_x = \bigcap_{x\in X} X\backslash\{x\} = \emptyset$. Thus $X = \emptyset$, contradicting the assumption.

(ii) Follows from (i) since complete metric spaces are T_1 and Baire.

(iii) If d is a complete metric such that $\tau = \tau_d$ then we have a contradiction with (ii).

(iv) $\mathbb{Q}^n$ has no isolated points since a Euclidean ball $B(x, \varepsilon)$ contains infinitely many points of $\mathbb{Q}^n$. Now apply (iii). ■

3.3.2 Baire's theorem and the choice axioms

It is clear that we used the Axiom of Choice in the proof of Theorem 3.3.1. Since we made only countably many choices, one might think that we only need the Axiom of Countable Choice (ACC). However, this is not true since the choice of x_2 must take the preceding choice for x_1 into account, the choice for x_3 depended on x_2, and so on! What is really needed is the Axiom of Countable Dependent Choice (ACDC), cf. Definition A.3.7.

Exercise 3.3.8 Rewrite the proof of Theorem 3.3.1 so as to make clear that the Axiom of Countable Dependent Choice suffices.

The following surprising result, proven in [20, 94], shows that ACDC is actually equivalent to Baire's theorem:

Theorem 3.3.9 *The axioms of set theory (without any choice axiom) together with the Baire property of complete metric spaces imply ACDC.*

3.3.3 G_δ and F_σ sets

In many applications of Baire's theorem, e.g. Theorems 3.3.18 and G.2.2 below, the individual dense open sets U_n carry little interest, but the fact that the dense set obtained at the end is a countable intersection of open sets does. This motivates the following definition:

Definition 3.3.10 *A countable intersection of open sets in a topological space is called a G_δ-set. A countable union of closed sets is a F_σ-set.*

We will encounter G_δ-sets quite often, e.g. in the guise of closed sets in metric spaces (Exercise 3.3.12), sets of continuity of functions (Proposition 3.4.6), in the characterization of completely metrizable spaces (Theorem 3.4.20) and also in non-metric contexts. One reason is that some results that are true for open sets generalize to G_δ-sets.

Exercise 3.3.11 Prove:

(i) If $Y \subseteq X$ is G_δ and $Z \subseteq Y$ is G_δ in Y, then Z is G_δ in X.

(ii) If $U \subseteq X$ is G_δ and $Y \subseteq X$ then $U \cap Y$ is G_δ in Y.

(iii) If $U \subseteq Y$ is G_δ, where $Y \subseteq X$, then there is a G_δ-set $V \subseteq X$ such that $U = V \cap Y$.

Obviously open sets are G_δ. The subsets $(a, b] = \bigcap_{n=1}^\infty (a, b + 1/n) \subseteq \mathbb{R}$ are G_δ, but not open. In metric spaces all closed sets are G_δ:

Exercise 3.3.12 Let (X, d) be a metric space. For $A \subseteq X$ and $\varepsilon > 0$ define

$$A_\varepsilon = \bigcup_{x \in A} B(x, \varepsilon). \tag{3.1}$$

Since this is open, $B = \bigcap_{n=1}^\infty A_{1/n}$ clearly is G_δ. Prove that $B = \overline{A}$ and deduce that closed subsets of metric spaces are G_δ.

Generalizing the proof of Proposition 3.3.7, we can prove that certain sets are not G_δ:

Proposition 3.3.13 *Let (X, τ) be non-empty, T_1 and Baire. (E.g. a non-empty complete metric space.) If $Y \subseteq X$ is countable, dense and such that no $y \in Y$ is isolated in X then Y is not G_δ (but $X \backslash Y$ is dense G_δ).*

Proof. Since X is T_1, every $\{y\}$ is closed, thus $X \backslash \{y\}$ open. Since $y \in Y$ is not isolated in X, $\{y\}$ is not open, $X \backslash \{y\}$ is non-closed, thus dense ($X \backslash \{y\} \subsetneq \overline{X \backslash \{y\}} \subseteq X$ implies $\overline{X \backslash \{y\}} = X$). Since Y is countable, $X \backslash Y = \bigcap_{y \in Y} X \backslash \{y\}$ is G_δ, and by the Baire property it is dense.

Assuming that Y is G_δ, we have $Y = \bigcap_{n=1}^\infty U_n$, where each U_n is open. Then

$$\bigcap_{n\in\mathbb{N}} U_n \;\cap\; \bigcap_{y\in Y} X\backslash\{y\} = Y \cap (X\backslash Y) = \emptyset. \tag{3.2}$$

Each U_n is dense since it contains Y, which is dense. Thus $\{U_n \mid n \in \mathbb{N}\} \cup \{X\backslash\{y\} \mid y \in Y\}$ is a countable family of dense open subsets of X. By the Baire property, the intersection of these sets is dense, thus non-empty since $X \neq \emptyset$, which contradicts (3.2). Thus Y is not G_δ. ■

Corollary 3.3.14 *No countable dense subset of* $\mathbb{R}^n$ *is* G_δ. *E.g.,* $\mathbb{Q} \subseteq \mathbb{R}$ *is non-*G_δ.

Proof. This follows from the Proposition since $\mathbb{R}^n$ with the Euclidean metric is complete and has no isolated point. ■

3.3.4 Application: Croft's Lemma

The following result, often called "Croft's Lemma", is proven in [54]. It is a nice application of Baire's theorem in classical analysis.

Theorem 3.3.15 *Let* $f : (0,\infty) \to \mathbb{R}$ *be continuous and satisfying* $\lim_{n\to\infty} f(nx) = 0$ *for every* $x > 0$. *Then* $\lim_{x\to\infty} f(x) = 0$.

Proof. Let $\varepsilon > 0$. For $m \in \mathbb{N}$, define

$$C_m = \{x \geq 1 \mid |f(nx)| \leq \varepsilon \ \forall n \geq m\} = \bigcap_{n\geq m} n^{-1} f^{-1}([-\varepsilon, \varepsilon]).$$

For every $x \geq 1$ we have $f(nx) \to 0$, thus there is m_x such that $n \geq m_x$ implies $|f(nx)| \leq \varepsilon$. This means that $x \in C_{m_x}$, so that we have proven $\bigcup_{m\in\mathbb{N}} C_m = [1,\infty)$. Continuity of f implies that $C_m \subseteq [1,\infty)$ is closed for each m. Since $\mathbb{R}$ with the standard metric $d(x,y) = |x-y|$ is complete and $[1,\infty) \subseteq \mathbb{R}$ is closed, $([1,\infty), d)$ is a complete metric space. Now Baire's theorem implies that one of the sets C_m must have non-empty interior and therefore contain an open interval (a,b) (since otherwise $\bigcup_m C_m = [1,\infty)$ would have empty interior, which is absurd). Now $(a,b) \subseteq C_m$ means that $|f(x)| \leq \varepsilon$ for every $x \in \bigcup_{n\geq m}(na, nb)$. Since $(n+1)/n \to 1 < b/a$, there is n_0 such that $n \geq n_0 \Rightarrow (n+1)/n < b/a$. Thus for $n \geq n_0$ we have $na < (n+1)a < nb < (n+1)b$, so that the intervals (na, nb) and $((n+1)a, (n+1)b)$ overlap. Together with $nb \to +\infty$ this implies $\bigcup_{n\geq n_0}(na, nb) = (n_0 a, \infty)$. Thus with $x_0 = n_0 a$ we have $x > x_0 \Rightarrow x \in \bigcup_{n\geq n_0}(na, nb) \Rightarrow |f(x)| \leq \varepsilon$. Since ε was arbitrary, we are done. ■

Remark 3.3.16 There are proofs of Croft's Lemma that avoid using Baire's theorem, cf. e.g [5, pp. 17, 149] or [193, p. 174], but they are much less transparent. All these proofs also use the Axiom of Countable Dependent Choice, just as the proof of Baire's theorem, so that they are not preferable from a foundational point of view. □

3.3.5 Application: A dense G_δ-set of nowhere differentiable functions

It is well known that there are continuous functions $f \in C(I, \mathbb{R})$, where $I \subseteq \mathbb{R}$ is an interval or all of $\mathbb{R}$, that are nowhere differentiable. It is not hard to write down candidates:

$$\begin{aligned} f_1(x) &= \sum_{n=1}^{\infty} \frac{1}{n^2} \sin(n^2 x), \\ f_2(x) &= \sum_{n=1}^{\infty} \frac{1}{2^n} \sin(2^n x). \end{aligned}$$

Both functions are obviously continuous, but the problem is to prove nowhere differentiability. In fact, despite having (supposedly) been discussed by Riemann, f_1 turned out to be differentiable precisely at countably many points! (This was completely clarified only more than 100 years after Riemann's death, cf. [89]. The function f_1 is extremely interesting, cf. e.g. [62].) The function f_2 indeed is nowhere differentiable, as was shown by Weierstrass and Hardy. (Cf. e.g. [217].) Here we'll follow a different route to nowhere differentiable functions, namely using Baire's theorem to prove existence of infinitely many such functions.

Let $I = [0, 1]$ and consider $(C_b(I, \mathbb{R}), D)$, where D is as in (2.6). (It is proven in Analysis courses (and again in Section 7) that $C_b(X, I) = C(X, I)$.) The metric space $(C_b(I, \mathbb{R}), D)$ is complete by Proposition 3.1.8. For $n \in \mathbb{N}$, define

$$U_n = \left\{ f \in C_b(I, \mathbb{R}) \mid \forall x \in I \ \exists y \in I : y \neq x, \ \left| \frac{f(y) - f(x)}{y - x} \right| > n \right\}.$$

Exercise 3.3.17 Prove that, for each n, U_n is (a) open and (b) dense in $(C_b(I, \mathbb{R}), D)$.

Theorem 3.3.18 *The topological space $(C_b([0, 1], \mathbb{R}), \tau_D)$ contains a dense G_δ-set of nowhere differentiable functions.*

Proof. Let $G = \bigcap_n U_n$. This obviously is a G_δ-set, and assuming the results of the exercise, it is dense by Baire's theorem. We claim that any $f \in G$ is nowhere differentiable. To prove this, assume that to the contrary f is differentiable at some $x \in I$ with $f'(x) = c$. By definition of differentiability,

$$\lim_{y \to x} \frac{f(y) - f(x)}{y - x} = c.$$

Thus picking some $\varepsilon > 0$, there is $\delta > 0$ such that $0 < |x - y| < \delta$ implies $\left| \frac{f(y)-f(x)}{y-x} - c \right| < \varepsilon$. Thus $\left| \frac{f(y)-f(x)}{y-x} \right|$ is bounded (by $|c| + \varepsilon$) if y satisfies $0 < |x - y| < \delta$. And if $|x - y| \geq \delta$ we have $\left| \frac{f(y)-f(x)}{y-x} \right| \leq \frac{2C}{\delta}$, where C is such that $|f(x)| \leq C$. (Recall $f \in C_b(I, \mathbb{R})$.) Thus the set $\left\{ \frac{f(y)-f(x)}{y-x} \mid y \neq x \right\}$ (where x is fixed) is bounded. On the other hand, $f \in G = \bigcap_n U_n$ means that for every x there is a $y \neq x$ such that $\left| \frac{f(y)-f(x)}{y-x} \right|$ is as large as desired. This contradiction proves that f is nowhere differentiable. ■

Remark 3.3.19 1. We emphasize that the proof is non-constructive: It does not give us any idea how to obtain such an f. On the other hand it gives us more than the explicit formulas above, namely a dense G_δ-set of nowhere differentiable functions.

2. Baire's theorem can also be used to prove that the space $\{f \in C([0,1],\mathbb{C}) \mid f(0) = f(1)\}$ of continuous periodic functions contains a dense (w.r.t. τ_D) G_δ-set of functions whose Fourier series all diverge at some fixed point, cf. Appendix G.2.3. The original way to prove this goes under the colorful names of the 'gliding hump method' or 'condensation of singularities'. Later this was streamlined using Baire's theorem and the 'principle of uniform boundedness' (Theorem G.2.2) that follows from it. (There are constructions of a continuous function with divergent Fourier series that avoid Baire's theorem and its corollaries, cf. e.g. [144, Section I.18], but they all use the Axiom of Countable Dependent Choice in one way or other.) □

3.4 ⋆ Oscillation. Extending continuous functions. Complete metrizability

3.4.1 Oscillation and sets of continuity

Definition 3.4.1 *Let (X, τ) be a topological space and (Y, d) a metric space. Let $A \subseteq X$ and $f : A \to Y$ a function. For $x \in \overline{A}$, define the oscillation of f at x by*

$$\operatorname{osc}(f, x) = \inf_{\substack{U \in \tau \\ x \in U}} \operatorname{diam}(f(A \cap U)).$$

Remark 3.4.2 1. We required $x \in \overline{A}$ since this is equivalent to having $A \cap U \neq \emptyset$ for every neighborhood U of x by Lemma 2.7.3.

2. Note that the oscillation is defined for every point in $\overline{A}$. If A is non-closed, this includes some points where f itself is not defined! This will be put to good use in the next section. □

Lemma 3.4.3 *Let $(X, \tau), (Y, d), A, f$ as in Definition 3.4.1. Then*

(i) For $x \in A$ we have $\operatorname{osc}(f, x) = 0$ if and only if for every $\varepsilon > 0$ there is an open neighborhood U of x such that $y \in A \cap U \ \Rightarrow\ d(f(x), f(y)) < \varepsilon$.

(ii) The set $B_f = \{x \in \overline{A} \mid \operatorname{osc}(f, x) = 0\}$ is a G_δ-set in $\overline{A}$.

Proof. (i) The statement $\operatorname{osc}(f, x) = 0$ clearly is equivalent to saying that for every $\varepsilon > 0$ we can find an open neighborhood U of x such that $\operatorname{diam}(f(A \cap U)) < \varepsilon$. Since $x \in A$, this is equivalent to $d(f(x), f(y)) < \varepsilon$ for every $y \in A \cap U$.

(ii) Defining $B_\varepsilon = \{x \in \overline{A} \mid \operatorname{osc}(f, x) < \varepsilon\}$, where $\varepsilon > 0$, we have $B_f = \bigcap_{n=1}^\infty B_{1/n}$. Now $\operatorname{osc}(f, x) < \varepsilon$ is equivalent to existence of an open neighborhood U of x with $\operatorname{diam}(f(A \cap U)) < \varepsilon$. Thus

$$\begin{aligned} B_\varepsilon &= \{x \in \overline{A} \mid \exists U \in \tau : x \in U,\ \operatorname{diam}(f(A \cap U)) < \varepsilon\} \\ &= \overline{A} \cap \{x \in X \mid \exists U \in \tau : x \in U,\ \operatorname{diam}(f(A \cap U)) < \varepsilon\} \\ &= \overline{A} \cap \bigcup \{U \in \tau \mid \operatorname{diam}(f(A \cap U)) < \varepsilon\}, \end{aligned}$$

which is open in $\overline{A}$. (In the second and third line one may worry about U's such that $U \cap A = \emptyset$, but they don't matter since they also satisfy $U \cap \overline{A} = \emptyset$.) Thus B_f is G_δ in $\overline{A}$. ■

Remark 3.4.4 We cannot hope to prove that B_f always is G_δ in X. If this was true, then taking $A \subseteq X$ closed and a continuous $f : A \to Y$ (e.g. f constant), we would conclude that $A = B_f$ is G_δ. But in general topological spaces it is not true that every closed set is G_δ. □

Definition 3.4.5 *If the equivalent conditions in (i) are satisfied, we say f is continuous at x, and x is called a point of continuity of f. (This notion will later be generalized to the case where also Y is a topological space, cf. Exercise 5.2.1.)*

Proposition 3.4.6 *Let (X, τ) be a topological space, (Y, d) a metric space and $f : X \to Y$ a function. Then the set of points of continuity of f is a G_δ-set.*

Proof. Considering Lemma 3.4.3 with $A = X$, we have $B_f \subseteq \overline{A} = A$, so that by (i) B_f coincides with the set of continuity points of f, and by (ii) B_f is G_δ in $\overline{A} = X$. ■

Combining this with Corollary 3.3.14, we see that there is no function $f : \mathbb{R} \to \mathbb{R}$ that has $\mathbb{Q}$ as its set of continuity points. But $\emptyset$ and $\mathbb{R}\backslash\mathbb{Q}$ are G_δ, and indeed:

Exercise 3.4.7 (i) Let $f_1 : \mathbb{R} \to \mathbb{R}$ be given by $f_1 = \chi_\mathbb{Q}$ (i.e. $f(x) = 1$ for $x \in \mathbb{Q}$ and $= 0$ otherwise). Prove that f_1 is nowhere continuous.

(ii) Define $f_2 : \mathbb{R} \to \mathbb{R}$ by $f_2(x) = 0$ for $x \in \mathbb{R}\backslash\mathbb{Q}$. For $x \in \mathbb{Q}$, put $f_2(x) = 1/n$, where $x = m/n$ with $n \in \mathbb{N}$ and m, n relatively prime. Prove that the set of continuity points of f_2 is $\mathbb{R}\backslash\mathbb{Q}$.

One may ask whether the converse of Proposition 3.4.6 is true, in the sense that for every G_δ-set $Z \subseteq X$ there is a function $f : X \to Y$ having Z as its set of continuity points. In this generality this is not true: If $\{z\} \subseteq X$ is clopen then every $f : X \to Y$ that is continuous on $X\backslash\{z\}$ actually is continuous on X. Thus $X\backslash\{z\}$ is not the continuity set of any function.

Proposition 3.4.8 *If X is a topological space containing two disjoint dense subsets S, T then for every G_δ-set $Z \subseteq X$ there is a function $f : X \to \mathbb{R}$ whose set of continuity equals Z.*

Proof. Let $Z \subseteq X$ be G_δ, thus $Z = \bigcap_{n=1}^\infty U_n$ with U_i open. For $x \in Z$, put $f(x) = 0$. If $x \notin Z$, let $n(x) = \min\{n \mid x \notin U_n\}$ and define $f(x) = 1/n(x)$ if $x \in S$ and $f(x) = -1/n(x)$ if $x \notin S$. Now it is not hard to check that f is continuous precisely on Z. ■

3.4.2 Extending continuous functions between metric spaces

Definition 3.4.9 *Let X, Y be metric spaces, $A \subseteq X$ and $f : A \to Y$ a continuous function. If*

$$A \subsetneq B \subseteq X, \quad \widehat{f} \in C(B, Y) \quad \text{and} \quad \widehat{f}\restriction A = f$$

then $\widehat{f}$ is called an extension of f.

(Once we define continuous functions between topological spaces, the above definition will obviously generalize.) In this section we construct extensions of $f : X \supseteq A \to Y$ under the assumption that X, Y are metric spaces with Y complete.

Proposition 3.4.10 *Let $(X, d), (Y, d')$ be metric spaces with (Y, d') complete, $A \subseteq X$ and $f : A \to Y$ continuous. Then*

(i) B_f defined as in Lemma 3.4.3(ii) satisfies $A \subseteq B_f \subseteq \overline{A}$ and is G_δ in X.

(ii) For $x \in B_f$ the set $\bigcap_{\delta>0} \overline{f(A \cap B(x, \delta))} \subseteq Y$ contains exactly one point. This defines a function $\widehat{f}(x) : B_f \to Y$.

(iii) $\widehat{f}\restriction A = f$.

(iv) $\widehat{f}(x) : B_f \to Y$ *is continuous.*

Proof. (i) $B_f \subseteq \overline{A}$ holds by definition, and $A \subseteq B_f$ follows from continuity of f on A and Lemma 3.4.3(i). By Lemma 3.4.3(ii) B_f is G_δ in $\overline{A}$, thus also in X since $\overline{A}$ is G_δ by Exercise 3.3.12.

(ii) Let $x \in B_f$. The sets $C_n = \overline{f(A \cap B(x, 1/n))} \subseteq Y$ are non-empty, closed and decreasing ($Y \supseteq C_1 \supseteq C_2 \supseteq \cdots$). The assumption $x \in B_f$, i.e. $\mathrm{osc}(f, x) = 0$, implies $\mathrm{diam}(C_n) \to 0$. Since (Y, d') is complete, Cantor's Intersection Theorem (Exercise 3.1.6) applies.

(iii) If $x \in A$ then $f(x) \in \bigcap_{\delta>0} \overline{f(A \cap B(x, \delta))}$, implying $\widehat{f}(x) = f(x)$.

(iv) Now let $x \in B_f$ and $\varepsilon > 0$. As noted before, there is a $\delta > 0$ such that $\mathrm{diam}(\overline{f(A \cap B(x, \delta))}) < \varepsilon$. For $y \in B(x, \delta)$, let $\delta' = \delta - d(x, y) > 0$. Then $B(x, \delta') \cup B(y, \delta') \subseteq B(x, \delta)$, implying $\overline{f(A \cap B(x, \delta'))} \cup \overline{f(A \cap B(y, \delta'))} \subseteq \overline{f(A \cap B(x, \delta))}$. By definition of $\widehat{f}$, $\widehat{f}(x)$ and $\widehat{f}(y)$ are contained in the two sets on the left hand side. Thus $d'(\widehat{f}(x), \widehat{f}(y)) \leq \mathrm{diam}(\overline{f(A \cap B(x, \delta))}) < \varepsilon$, proving continuity of $\widehat{f}$. ■

Remark 3.4.11 If A is closed then $B_f = A$, so that the Proposition becomes empty. But if $A \subseteq X$ is not G_δ (and thus also not closed by Exercise 3.3.12) we necessarily have $B_f \supsetneq A$, implying that every $f \in C(A, Y)$ has a proper extension! Proposition 3.3.13 provides such sets, e.g. any countable dense subset of $\mathbb{R}^n$. □

Definition 3.4.12 *If* $(X, d), (Y, d')$ *are metric spaces, then* $f : X \to Y$ *is* <u>*uniformly continuous*</u> *if for every* $\varepsilon > 0$ *there is* $\delta > 0$ *such that* $d(x, y) < \delta \Rightarrow d'(f(x), f(y)) < \varepsilon$.

The following corollary (which could also be proven directly) will have many uses:

Corollary 3.4.13 *Let* $(X, d), (Y, d')$ *be metric spaces with* Y *complete, let* $A \subseteq X$ *and* $f : (A, d) \to (Y, d')$ *uniformly continuous. Then* f *has a uniformly continuous extension* $\widehat{f} : \overline{A} \to Y$.

Proof. Uniform continuity of f implies $\mathrm{osc}(f, x) = 0$ for all $x \in \overline{A}$, thus $B_f = \overline{A}$. Now apply Proposition 3.4.10. Reviewing the proof of continuity given there, we also see that also $\widehat{f} : \overline{A} \to Y$ is uniformly continuous. ■

Remark 3.4.14 In particular, every uniformly continuous $f : (X, d) \to (Y, d')$, where (Y, d') is complete, has a unique extension $\widehat{f}$ to the completion $(\widehat{X}, \widehat{d})$ of (X, d). (One can show that the completion is characterized by this fact.) □

Theorem 3.4.15 (Lavrentiev) [5] *Let* $(X, d), (X', d')$ *be complete metric spaces,* $A \subseteq X$, $A' \subseteq X'$ *and* $f : A \to A'$ *a homeomorphism. Then there are* G_δ*-sets* B, B' *such that* $A \subseteq B \subseteq \overline{A}$, $A' \subseteq B' \subseteq \overline{A'}$ *and a homeomorphism* $\widehat{f} : B \to B'$ *extending* f.

Exercise 3.4.16 Prove Lavrentiev's theorem. Hint: Use Proposition 3.4.10 to obtain G_δ-sets C, C' such that $A \subseteq C \subseteq \overline{A}$, $A' \subseteq C' \subseteq \overline{A'}$ and maps $\widehat{f} : C \to Y$, $\widehat{f^{-1}} : C' \to X$ extending f, f^{-1}. Then put $B = C \cap \widehat{f}^{-1}(C')$, $B' = C' \cap \widehat{f^{-1}}^{-1}(C)$ and prove that $\widehat{f} \restriction B$ and $\widehat{f^{-1}} \restriction B'$ are mutually inverse.

[5] Mikhail Alekseevich Lavrent(i)ev (1900-1980), Russian mathematician.

3.4.3 More on complete metrizability

The following proposition is usually derived from Lavrentiev's Theorem 3.4.15, but doing so obscures the simplicity of the underlying idea. We prefer to give a direct proof.

Proposition 3.4.17 *If (X, d) is a metric space and $Y \subseteq X$ admits a complete metric d' such that $d' \simeq d \restriction Y$ (i.e. (Y, d) is completely metrizable) then $Y \subseteq X$ is G_δ.*

Proof. Let d' be a complete metric on Y such that $d' \simeq d \restriction Y$. By Exercise 2.2.13(i)⇔(iii), the maps $f = \mathrm{id}_Y : (Y, d) \to (Y, d')$ and $g = \mathrm{id}_Y : (Y, d') \to (Y, d)$ are continuous and mutually inverse, in particular $g \circ f = \mathrm{id}_Y$. Since (Y, d') is complete, applying Proposition 3.4.10 to the situation $f : (X, d) \supseteq (Y, d) \xrightarrow{f} (Y, d')$ gives us a G_δ-set B such that $Y \subseteq B \subseteq \overline{Y}$ (where the closure is in X w.r.t. the metric d) and a continuous extension $\widehat{f} : (B, d) \to (Y, d')$ of f. The composite map $(B, d) \xrightarrow{\widehat{f}} (Y, d') \xrightarrow{g} (Y, d)$ is continuous and its restriction to $Y \subseteq B$ is the identity map of Y. Since Y is d-dense in B, it follows by continuity that $g \circ \widehat{f}$ is the identity map of B. This implies that $\widehat{f} : B \to Y$ is injective, but since $f : Y \to Y$ already was surjective and $\widehat{f}$ extends f to $B \supseteq Y$, it follows that $B = Y$. By construction, $B \subseteq X$ is G_δ, thus the claim follows. ■

Proposition 3.4.18 *If (X, d) is a complete metric space and $Y \subseteq X$ is G_δ, then (Y, d) is completely metrizable.*

Proof. Let $\{U_s\}_{s\in\mathbb{N}}$ be open sets such that $Y = \bigcap_s U_s$. In view of Exercise 2.1.20, each of the functions $y \mapsto \mathrm{dist}(y, X\backslash U_s)$ is continuous and vanishes if and only if y is in the closed set $X\backslash U_s$, which is impossible for $y \in Y$. For $y_1, y_2 \in Y = \bigcap_s U_s$ define

$$d'(y_1, y_2) = d(y_1, y_2) + \sum_{s=1}^{\infty} 2^{-s} \min\left(1, \left|\frac{1}{\mathrm{dist}(y_1, X\backslash U_s)} - \frac{1}{\mathrm{dist}(y_2, X\backslash U_s)}\right|\right).$$

To see that this is a metric on Y it suffices to note that d is a metric, that each summand on the right is a pseudometric and that the sum converges for all $y_1, y_2 \in Y$. If $\{y_n\}$ is a sequence in Y and $d(y_n, y) \to 0$ where $y \in Y$, then each of the functions $(\mathrm{dist}(y_1, X\backslash U_s))^{-1}$ converges to $(\mathrm{dist}(y, X\backslash U_s))^{-1}$, thus $d'(y_n, y) \to 0$. Now assume that the sequence $\{y_n\}$ in Y is Cauchy w.r.t. d'. In view of $d \le d'$ it is Cauchy with respect to d, which is complete, thus $\{y_n\}$ converges to some $x \in X$. If we can show that $x \in Y$, we have proven that the metric d' on Y is complete and equivalent to $d \restriction Y$.

Let $\varepsilon > 0$ and $s \in \mathbb{N}$. Then there is $N \in \mathbb{N}$ such that $i, j \ge N$ implies $d'(x_i, x_j) < 2^{-s}\varepsilon$ and thus

$$\left|\frac{1}{\mathrm{dist}(y_i, X\backslash U_s)} - \frac{1}{\mathrm{dist}(y_j, X\backslash U_s)}\right| < \varepsilon.$$

Thus in particular for $i \ge N$ we have

$$\mathrm{dist}(y_i, X\backslash U_s)^{-1} \in [\mathrm{dist}(y_N, X\backslash U_s)^{-1} - \varepsilon, \mathrm{dist}(y_N, X\backslash U_s)^{-1} + \varepsilon].$$

If we choose ε small enough, this implies

$$\mathrm{dist}(y_i, X\backslash U_s) \ge (\mathrm{dist}(y_N, X\backslash U_s)^{-1} - \varepsilon)^{-1} > 0.$$

With $\delta = \mathrm{dist}(y_N, X\backslash U_s)^{-1} - \varepsilon > 0$ we find $y_i \in C_s = \{y \in X \mid \mathrm{dist}(x, X\backslash U_s) \ge \delta\} \subseteq U_s$. Note that $C_s \subseteq X$ is closed. Thus for each s we find a closed $C_s \subseteq U_s$ such that the sequence $\{x_i\}$ eventually lives in C_s. Together with the fact that $\{y_n\}$ converges to $x \in X$, this implies $x \in C_s$ for all s. Thus $x \in \bigcap_s C_s \subseteq \bigcap_s U_s = Y$, and we are done. ■

Remark 3.4.19 Alternatively, one can prove this result for open $Y \subseteq X$, so that the index s disappears, and then deduce the result for G_δ-sets using Exercise 6.5.23 and Corollary 6.5.36. However, we prefer the above proof since it is technically simpler. □

Theorem 3.4.20 *For a metric space (X, d) the following are equivalent:*

(i) $\iota(X) \subseteq X'$ is G_δ for every isometry $\iota : (X, d) \to (X', d')$.

(ii) $\iota(X) \subseteq X'$ is G_δ for every isometry $\iota : (X, d) \to (X', d')$ with (X', d') complete.

(iii) X is a G_δ-subset in its completion $(\widehat{X}, \widehat{d})$.

(iv) $\iota(X) \subseteq X'$ is G_δ for some isometry $\iota : (X, d) \to (X', d')$ with (X', d') complete.

(v) (X, d) is completely metrizable.

Proof. (i)⇒(ii)⇒(iii)⇒(iv) is trivial. (iv)⇒(v) is Proposition 3.4.18. (v)⇒(i) is Proposition 3.4.17, cf. also Remark 3.1.13.1. ■

Compare this with Proposition 3.2.7: A metric space (X, d) is complete (resp. completely metrizable) if and only if it is universally closed (resp. universally G_δ), i.e. (X, d) is closed (resp. G_δ) in every metric space isometrically containing it. (Recall that every closed set in a metric space is G_δ.) We will return to this subject in Section 8.4.2.

Chapter 4

More basic topology

4.1 Bases. Second countability. Separability

4.1.1 Bases

For (X, d) a metric space, it is easy to see that $Y \subseteq X$ is open if and only if Y is a union of open balls $B(x, r)$. This motivates the following:

Definition 4.1.1 *Let (X, τ) be a topological space. A subset $\mathcal{B} \subseteq \tau$ is called a* <u>*base for τ*</u> *if every $U \in \tau$ can be expressed as a union of elements of $\mathcal{B}$.*

Equivalently, $\mathcal{B}$ is a base if whenever $x \in U \in \tau$, there is $V \in \mathcal{B}$ such that $x \in V \subseteq U$.

Example 4.1.2 Thus if (X, d) is metric then $\mathcal{B} = \{B(x, r) \mid x \in X, r > 0\}$ is a base for τ_d.

Specializing to $(\mathbb{R}, \tau_d)$, where $d(x, y) = |x - y|$, we find that the open intervals

$$\mathcal{B} = \{(a, b) \mid a < b\}$$

form a base for the standard topology τ_d of $\mathbb{R}$. $\square$

Example 4.1.3 If (X, τ) is discrete, $\mathcal{B} = \{\{x\} \mid x \in X\}$ is a base for τ. Actually this is the unique smallest base for the discrete topology.

Warning: It is very rare for a topology to have a unique smallest base! But if (X, τ) is a smallest neighborhood space, cf. Section 2.8.3, and each $x \in X$ has smallest neighborhood U_x, then $\{U_x \mid x \in X\}$ clearly is the unique smallest base for τ. $\square$

If (X, d) is metric, one can find proper subsets of $\mathcal{B} = \{B(x, r) \mid x \in X, r > 0\}$ that are still bases, e.g. $\mathcal{B}' = \{B(x, 1/n) \mid x \in X, n \in \mathbb{N}\}$. (Here we use the Archimedian property of $\mathbb{R}$: For every $r > 0$ there exists $n \in \mathbb{N}$ such that $\frac{1}{n} < r$.)

This raises the question how small, in terms of cardinality, a base can be. It is clear that a finite base exists if and only if τ is finite. The example of the indiscrete topology shows that this can happen even if the underlying space is infinite. But if X is infinite and τ is T_1 then $\{X \backslash \{x\} \mid x \in X\} \subseteq \tau$, thus τ is infinite and therefore also any base for τ.

4.1.2 Second countability and separability

We will now focus on the countable case, but see Remark 4.1.6.2.

Definition 4.1.4 *A space (X, τ) satisfies the* <u>*second axiom of countability*</u> *(= is* <u>*second countable*</u>*) if it admits a countable base.*

Example 4.1.5 Let τ be the usual topology on $\mathbb{R}$. Then also

$$\mathcal{B}' = \{(a, b) \mid a, b \in \mathbb{Q},\ a < b\}$$

is a base for τ since for real numbers $a < x < b$ we can find $a', b', \in \mathbb{Q}$ such that $a < a' < x < b' < b$, thus $x \in (a', b')$. We conclude that $(\mathbb{R}, \tau)$ is second countable, despite the uncountability of $\mathbb{R}$. That the the same holds for $(\mathbb{R}^n, \tau_{\text{eucl}})$ could be proven directly, but it is convenient to introduce some more formalism. □

Remark 4.1.6 1. Later we will prove that if X is second countable, i.e. admits a countable base, and $\mathcal{V}$ is any base, there is a countable base $\mathcal{V}_0 \subseteq \mathcal{V}$. Cf. Proposition 7.1.10.

2. In the serious literature on general topology, e.g. [75], second countability is generalized as follows: If χ is a cardinal number, a topological space (X, τ) is said to have <u>weight</u> $w(X) \leq \chi$ if there is a base for τ having cardinality $\leq \chi$. One can then study how the weight behaves under various constructions. We will almost exclusively consider the case $\chi = \aleph_0 = \#\mathbb{N}$, which is sufficient for many applications (like Proposition 8.1.15 and Theorems 8.2.33). □

Exercise 4.1.7 (i) Prove that second countability is hereditary.

(ii) Prove that a discrete subset of a second countable space is (at most) countable.

Exercise 4.1.8 Let (X, τ) be a T_1-space and $\mathcal{B}$ a base for τ. Prove:

(i) $\#X \leq 2^{\#\mathcal{B}} \equiv \#P(\mathcal{B})$.

(ii) If X is second countable then $\#X \leq \mathfrak{c}$ (where $\mathfrak{c} = 2^{\aleph_0} = \#P(\mathbb{N}) = \#\mathbb{R}$).

(iii) (BONUS) Reprove (i) assuming only T_0.

Definition 4.1.9 *A space is called* <u>*separable*</u> *if it has a countable dense subset.*

Unfortunately, some authors (mainly in topological group theory) write 'separable' when they mean second countable, which can create confusion.

Lemma 4.1.10 *Let (X, τ) be a topological space.*

(i) If $\mathcal{B}$ is a base for τ then there is a dense subset $S \subseteq X$ such that $\#S \leq \#\mathcal{B}$.

(ii) If (X, τ) is second countable then it is separable.

Proof. (i) Using the axiom of choice, we choose a point $x_U \in U$ for each $U \in \mathcal{B}$, $U \neq \emptyset$ and define $S = \{x_U \mid \emptyset \neq U \in \mathcal{B}\}$. Let $W \subseteq X$ be open and non-empty. By definition of a base, there is a non-empty $U \in \mathcal{B}$ such that $U \subseteq W$. Now $x_U \in U \subseteq W$, and $x_U \in S$, thus $W \cap S \neq \emptyset$. Now Lemma 2.7.9 gives that $S \subseteq X$ is dense. By construction we have $\#S \leq \#\mathcal{B}$.

(ii) Second countable means that there is a countable base $\mathcal{B}$. Now by (i) there is a countable dense subset, thus X is separable. ■

The converse of (ii) is not true in general: For uncountable X, the space (X, τ_{cofin}) is separable, but not second countable, cf. Exercise 4.1.16(iv),(vi).

Lemma 4.1.11 *(i) If a metric space (X, d) is separable, it is second countable.*

(ii) Thus for metric spaces, separability $\Leftrightarrow$ second countability.

Proof. (i) Let $Y \subseteq X$ be countable and dense. Let

$$\mathcal{B} = \{B(y, 1/n) \mid y \in Y,\ n \in \mathbb{N}\}.$$

Now $\#\mathcal{B} = \#(Y \times \mathbb{N}) \leq \#(\mathbb{N} \times \mathbb{N}) = \#\mathbb{N}$, thus $\mathcal{B}$ is countable. It remains to prove that $\mathcal{B}$ is a base for τ_d. So let $U \in \tau$ be non-empty and $x \in U$. Since U is open, there is an $n \in \mathbb{N}$ such that $B(x, 1/n) \subseteq U$. Since Y is dense, there is $y \in Y$ such that $d(x, y) < 1/2n$ ($\Leftrightarrow x \in B(y, 1/2n)$). If now $z \in B(y, 1/2n)$ then $d(x, z) \leq d(x, y) + d(y, z) < \frac{1}{2n} + \frac{1}{2n} = \frac{1}{n}$. This proves that $V := B(y, 1/2n)$, which clearly is an element of $\mathcal{B}$, is contained in $B(x, 1/n)$ and therefore in U. Thus $x \in V \subseteq U$, proving that $\mathcal{B}$ is a base for τ_d. (ii) is now obvious. ■

Corollary 4.1.12 *$\mathbb{R}^n$ with the usual topology is second countable for any $n \in \mathbb{N}$.*

Proof. $\mathbb{Q}^n \subseteq \mathbb{R}^n$ is dense w.r.t. τ_{eucl}, thus $\mathbb{R}^n$ is separable. Now apply Lemma 4.1.11. ■

Many properties considered so far have been hereditary: discreteness, indiscreteness, cofiniteness, cocountability, metrizability, T_1, T_2 and second countability. But connectedness and (the metric property of) completeness are not hereditary, and the same holds for separability:

Exercise 4.1.13 (i) Given any topological space (X, τ), put $X' = X \cup \{p\}$ (where $p \notin X$). Give a topology τ' on X' such that $\tau' \upharpoonright X = \tau$ and $\{p\} \subseteq X'$ is dense.

(ii) Conclude that separability is not hereditary.

The spaces produced by the above construction are not very nice (irreducible and non-T_1). A Hausdorff example will be given in Lemma 6.5.16.

The existence of non-hereditary properties motivates the following:

Definition 4.1.14 *If a property P is not hereditary, a space is called hereditarily P if it and all its subspaces have property P. (Obviously, the property of being hereditarily P is hereditary.)*

Exercise 4.1.15 Prove:

(i) Open subspaces of separable spaces are separable. (I.e. separability is 'open-hereditary'.)

(ii) Second countable spaces are hereditarily separable.

(iii) Separable metric spaces are hereditarily separable.

Exercise 4.1.16 Prove:

(i) Every countable topological space is separable.

(ii) Every countable metric space is second countable.

(iii) If X is countable then (X, τ_{cofin}) and (X, τ_{cocnt}) are separable.

(iv) If X is uncountable then (X, τ_{cofin}) is separable, but (X, τ_{cocnt}) is not.

(v) If X is countable then (X, τ_{cofin}) is second countable.

(vi) If X is uncountable then $(X, \tau_{\mathrm{cofin}})$ and $(X, \tau_{\mathrm{cocnt}})$ are not second countable.

Hint: Let X be uncountable and $\mathcal{F}$ a family of finite subsets of X such that every finite subset of X is contained in some $F \in \mathcal{F}$. Prove that $\mathcal{F}$ is uncountable.

Warning: There are topologies (necessarily non-metrizable) on countable sets that are not second countable! Cf. Exercise 4.3.14.

Besides second countability and separability, there is a third countability property (somewhat less important):

Definition 4.1.17 *A topological space X has the Souslin property*[1] *or the countable chain condition (c.c.c.) if every family of mutually disjoint non-empty open subsets of X is countable.*

Exercise 4.1.18 Prove:

(i) An uncountable discrete space is not Souslin (thus neither separable nor second countable.)

(ii) If $\mathcal{U}$ is a family of mutually disjoint non-empty open subsets of X and $S \subseteq X$ is dense then $\#\mathcal{U} \le \#S$. In particular, every separable space has the Souslin property.

(Thus: second countable $\Rightarrow$ separable $\Rightarrow$ Souslin.)

(iii) Every irreducible space is Souslin.

(iv) The cocountable topology on any set has the Souslin property. (Thus Souslin $\not\Rightarrow$ separable, since on an uncountable set τ_{cocnt} is not separable (Exercise 4.1.16(iv)).)

(v) A metrizable space with the Souslin property is separable (thus also second countable).

We summarize the most important properties of the cofinite and cocountable topologies (in the non-discrete cases):

Space	2nd cnt.	separable	Souslin
τ_{cofin} for countably infinite X	Yes	Yes	Yes
τ_{cofin} for uncountable X	No	Yes	Yes
τ_{cocnt} for uncountable X	No	No	Yes

The example of a non-separable space with the Souslin property provided by (iv) is unsatisfactory since it is irreducible, thus not Hausdorff. See Example 6.5.28 for a better result.

The following exercise continues Exercise 2.7.25:

Exercise 4.1.19 (Cantor-Bendixson Theorem) Let X be a topological space and $Y \subseteq X$. Recalling that Y^{cd} denotes the condensation points of Y, prove:

(i) If X is second countable then $Y \backslash Y^{\mathrm{cd}}$ is countable, $(Y^{\mathrm{cd}})^{\mathrm{cd}} = Y^{\mathrm{cd}}$, and Y^{cd} is perfect.

(ii) Every second countable space has a perfect subspace whose complement is countable.

(iii) A second countable scattered space is countable.

[1] Mikhail Yakovlevich Souslin (1894-1919), Russian mathematician. Made important contributions to general topology and descriptive set theory, then died from typhus at age 24. (The French transliteration is due to the fact that S.'s few publications were in French.)

4.1.3 Spaces from bases

So far, we used bases to study given topological spaces. But bases can also be used to construct topologies: Given a set X and $\mathcal{B} \subseteq P(X)$, thus a family of subsets of X, the question arises whether there is a topology τ on X such that $\mathcal{B}$ is a base for τ. Since τ must contain $\mathcal{B}$, be closed under arbitrary unions and we want every $U \in \tau$ to be a union of elements of $\mathcal{B}$, we must put

$$\tau = \left\{ \bigcup_{i \in I} U_i \mid I \text{ a set}, \ U_i \in \mathcal{B} \ \ \forall i \in I \right\}. \tag{4.1}$$

In words: $\tau \subseteq P(X)$ consists of all unions of sets in $\mathcal{B}$[2]. Now one has:

Proposition 4.1.20 *Let X be a set, $\mathcal{B} \subseteq P(X)$ and τ defined as in (4.1). Then τ is a topology if and only if*

(a) $\bigcup \mathcal{B} = X$. (I.e. every $x \in X$ is contained in some $U \in \mathcal{B}$.)

(b) For every $U, V \in \mathcal{B}$ we have

$$U \cap V = \bigcup \{W \in \mathcal{B} \mid W \subseteq U \cap V\}. \tag{4.2}$$

(Equivalently, if $U, V \in \mathcal{B}$ and $x \in U \cap V$ then there is a $W \in \mathcal{B}$ such that $x \in W \subseteq U \cap V$.)

If this is the case then $\mathcal{B}$ is a base for τ.

Proof. It is clear that $\emptyset \in \tau$ and that τ, defined by (4.1), is closed under arbitrary unions. Furthermore, $X \in \tau$ is equivalent to (a). Thus assume (a) holds. It remains to show that (b) is equivalent to τ being a topology. Assume the latter is the case, and let $U, V \in \mathcal{B}$. Then $U, V \in \tau$, thus $U \cap V \in \tau$, thus $U \cap V$ is a union of elements of $\mathcal{B}$, thus also (4.2) holds.

As to the converse, assume (4.2) holds for all $U, V \in \mathcal{B}$. Assume $U, V \in \tau$, thus $U = \bigcup_i U_i, V = \bigcup_j V_j$ with certain $U_i, V_j \in \mathcal{B}$. Now

$$U \cap V = (\bigcup_i U_i) \cap (\bigcup_j V_j) = \bigcup_{i,j} (U_i \cap V_j) = \bigcup_{i,j} \bigcup \{W \in \mathcal{B} \mid W \subseteq U_i \cap V_j\},$$

thus $U \cap V$ is a union of elements of $\mathcal{B}$, and therefore in τ. Thus $U, V \in \tau \Rightarrow U \cap V \in \tau$, proving that τ is a topology. ■

The following is a first example of the use of the lemma to construct exotic topologies. For others, see the exercises below.

Example 4.1.21 Let $X = \mathbb{R}$ and $\mathcal{B} = \{[a, b) \mid a < b\}$. (Notice that this is NOT a base for the metric topology τ_d since $[a, b)$ is not in τ_d.) It is clear that the union over $\mathcal{B}$ equals $\mathbb{R}$. Now, if $e = \max(a, c)$ and $f = \min(b, d)$ then

$$[a, b) \cap [c, d) = \left\{ \begin{array}{cc} \emptyset & \text{if } e \geq f \\ [e, f) & \text{if } e < f \end{array} \right\},$$

which either empty or in $\mathcal{B}$. Thus $\mathcal{B}$ is the base of a topology τ_S, called the Sorgenfrey topology[3], whose elements clearly are of the form $\bigcup_i [a_i, b_i)$. Now $(\mathbb{R}, \tau_S)$ is called the Sorgenfrey line. □

[2] A much more efficient way of stating this is $\tau = \{\bigcup \mathcal{F} \mid \mathcal{F} \subseteq \mathcal{B}\}$, cf. Defininition/Proposition A.1.1, but for unclear reasons this notation seems to be unpopular.

[3] Robert Sorgenfrey (1915-1995), American mathematician.

Exercise 4.1.22 Prove that the Sorgenfrey topology τ_S is finer than the Euclidean topology $\tau_{\rm eucl}$ on $\mathbb{R}$ and deduce that τ_S is Hausdorff.

Example 4.1.23 Let (X, τ) be a topological space. Let $\mathcal{B}_r \subseteq \tau$ be the family of regular open sets, cf. Exercise 2.6.14. Since X is regular open (being clopen) and the intersection of two regular open sets is regular, by Proposition 4.1.20 there is a topology τ_r having $\mathcal{B}_r$ as a base. It is clear that $\tau_r \subseteq \tau$, thus τ_r is coarser than τ. The space (X, τ_r) is called the semiregularization of (X, τ). And (X, τ) is called semiregular if $\tau_r = \tau$, i.e. every open set is a union of regular open sets. $\square$

Exercise 4.1.24 (Infinitude of primes) Let $\mathbb{N} = \{1, 2, 3, \ldots\}$ (thus $0 \not\in \mathbb{N}$!!). For $a \in \mathbb{Z}$, $b \in \mathbb{N}$ we define

$$N_{a,b} = a + b\mathbb{Z} = \{a + bn \mid n \in \mathbb{Z}\} \subseteq \mathbb{Z}.$$

(i) Prove: $N_{a,b} = \{c \in \mathbb{Z} \mid c - a \equiv 0 \pmod b\} = \{c \in \mathbb{Z} \mid b \text{ divides } c - a\}$.

(ii) Prove: If $c \in N_{a,b}$ then $N_{c,b} = N_{a,b}$.

(iii) Prove that $\mathcal{B} = \{N_{a,b} \mid a \in \mathbb{Z},\, b \in \mathbb{N}\}$ is a base for a topology τ on $\mathbb{Z}$.

(iv) Prove that each $U \in \tau$ with $U \neq \emptyset$ is infinite.

(v) Prove that each $N_{a,b}$ is clopen.

(vi) Let $P = \{2, 3, 5, \ldots\}$ be the set of prime numbers. Prove that

$$\mathbb{Z}\backslash\{1, -1\} = \bigcup_{p \in P} N_{0,p}.$$

(vii) Prove that P is infinite, using only the facts proven above.[4]

The topology constructed in the preceding exercise is not Hausdorff. A slight modification gives a Hausdorff topology:

Exercise 4.1.25 Let $\mathbb{N} = \{1, 2, 3, \ldots\}$ and $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$. Let $\mathcal{B}$ the family of subsets of $\mathbb{N}$ of the form $U_{a,b} = \{a + nb \mid n \in \mathbb{N}_0\}$, where $a, b \in \mathbb{N}$ are relatively prime (thus $(a, b) = gcd(a, b) = 1$). Prove:

(i) $\mathcal{B}$ is a base for a topology τ on $\mathbb{N}$.

(ii) τ is Hausdorff.

(iii) τ is connected.

Remark: We thus have a countably infinite space whose topology is Hausdorff and connected! (Recall that finite T_1 spaces are discrete, thus non-connected.)

[4]While this exercise really gives a proof of the infinitude of P (due to H. Fürstenberg, 1955), it seems rather mystifying. See [45, 168] for topology-free elucidations of what 'really' is behind the proof.

4.2 Subbases and order topologies

4.2.1 Subbases. Topologies generated by families of subsets

Condition (b) in Proposition 4.1.20 is not always convenient. This motivates the following

Definition 4.2.1 *Let (X, τ) be a topological space. A subset $\mathcal{S} \subseteq \tau$ is called a subbase for τ if taking all finite intersections of elements of $\mathcal{S}$ one obtains a base $\mathcal{B}$ for τ. (The intersection $\bigcap \emptyset$ over an empty family of subsets of X is interpreted as the ambient space X.)*[5]

Lemma 4.2.2 *Let X be a set and $\mathcal{S} \subseteq P(X)$. Then there is a unique topology τ such that $\mathcal{S}$ is a subbase for τ.*

Proof. Let $\mathcal{B} \subseteq P(X)$ be the set of all finite intersections of elements of $\mathcal{S}$. Since the empty intersection $\bigcap \emptyset$ is interpreted as X, we have $X \in \mathcal{B}$. Clearly $\mathcal{B}$ is closed w.r.t. finite intersections. Thus $\mathcal{B}$ satisfies (a) and (b) in Proposition 4.1.20. ■

Here is a different perspective at the topology obtained from $\mathcal{S}$:

Definition 4.2.3 *Let X be a set and $\mathcal{U} \subseteq P(X)$ arbitrary. The topology $\tau_{\mathcal{U}}$ on X generated by $\mathcal{U}$ is the intersection of all topologies on X that contain $\mathcal{U}$, i.e. $\tau_{\mathcal{U}} = \bigcap \{\tau \supseteq \mathcal{U} \text{ a topology on } X\}$.*

Remark 4.2.4 1. By Exercise 2.3.6, any intersection of topologies on X is a topology on X. Clearly, the above $\tau_{\mathcal{U}}$ is the smallest topology on X that contains $\mathcal{U}$. (But one should not take this as the definition of $\tau_{\mathcal{U}}$ without proving that a smallest topology containing $\mathcal{U}$ exists, which is what we have done above.) Furthermore, the topology generated by $\mathcal{U} \subseteq P(X)$ precisely consists of all unions of finite intersections of elements of $\mathcal{U}$. Thus every $\mathcal{S} \subseteq P(X)$ is a subbase for the topology on X generated by $\mathcal{S}$.

2. If X is a set and $\mathcal{F}$ is an infinite family of subsets of X (thus X is infinite) then the family $\mathcal{F}'$ of all finite intersections of elements of $\mathcal{F}$ has the same cardinality as $\mathcal{F}$. Thus the spaces admitting a countable subbase are precisely the second countable ones. □

4.2.2 Order topologies

Definition 4.2.5 *Let $(X, \leq)$ be a totally ordered set. For $x \in X$, let*

$$L_x = \{y \in X \mid y < x\}, \qquad R_x = \{y \in X \mid y > x\}$$

and

$$\mathcal{S} = \{L_x \mid x \in X\} \cup \{R_x \mid x \in X\}. \tag{4.3}$$

$\mathcal{S}$ is a subbase for a topology $\tau_{\leq}$ on X, called the order topology, and $(X, \tau_{\leq})$ is called a (totally) ordered topological space.

Remark 4.2.6 1. Clearly this is interesting only if $\#X \geq 2$, in which case $\mathcal{S} \neq \emptyset$.

2. Since $L_x \cap L_y = L_{\min(x,y)}$ and similarly for R_x, every finite intersection of elements of $\mathcal{S}$ is of one of the following forms: $\emptyset, L_x, R_x, L_b \cap R_a$. The last one is also denoted (a, b).

[5] If $\mathcal{F}$ is the empty family of sets then $\bigcap \mathcal{F}$ should in principle denote the 'all-set' that contains "everything". Such a set does not exist since its existence, together with the Axiom of Separation, would lead to Russel's paradox. Thus $\bigcap \emptyset$ makes sense only in the context of some ambient set. Speaking formally, for every set X there is a map $I_X : P(P(X)) \to P(X),\ \mathcal{F} \mapsto \{x \in X \mid S \in \mathcal{F} \Rightarrow x \in S\}$. Now we have $I_X(\emptyset) = X$, and all is well.

3. For $\mathbb{R}$ with the usual ordering, $\tau_{\leq}$ is the usual (metric) topology since the latter has the above $\mathcal{S}$ as a subbase. □

Exercise 4.2.7 Let $(X, \leq)$ be a totally ordered set with $\#X \geq 2$. Prove that the order topology $\tau_{\leq}$, cf. Definition 4.2.5, is Hausdorff (T_2).

Hint: For $x < y$, proceed differently according to $\{z \mid x < z < y\} = \emptyset$ or $\neq \emptyset$.

Definition 4.2.8 *Let $(X, \leq), (Y, \leq)$ be totally ordered sets. Then the lexicographic ordering on $X \times Y$ is given by $(x, y) \leq (x', y') :\Leftrightarrow x < x' \vee (x = x' \wedge y \leq y')$.*

Exercise 4.2.9 Consider the lexicographic order on $X = \mathbb{Z} \times [0, 1)$ (where $\mathbb{Z}$ and $[0, 1)$ have the natural orders). Prove that the map $f : X \to \mathbb{R}, (n, r) \mapsto n + r$ is an order-preserving bijection. Deduce that the order topology $\tau_{\leq}$ on X and the usual topology τ on $\mathbb{R}$ satisfy $\tau = \{f(U) \mid U \in \tau_{\leq}\}$. ('An order isomorphism induces a homeomorphism.')

Remark 4.2.10 Let $(X, \leq)$ be totally ordered. Defining $\overline{R}_x = \{y \in X \mid y \geq x\}$ and

$$\widetilde{\mathcal{S}} = \{L_x \mid x \in X\} \cup \{\overline{R}_x \mid x \in X\},$$

$\widetilde{\mathcal{S}}$ is a subbase for a topology $\widetilde{\tau}$, which for $X = \mathbb{R}$ just is the Sorgenfrey topology τ_S. The latter clearly is an example for the following definition. □

Definition 4.2.11 *Let $(X, \leq)$ be totally ordered. A subset $Y \subseteq X$ is called convex if $x, y \in Y$ implies $z \in Y$ whenever $x < z < y$. If a topology τ on X has a base consisting of convex sets then $(X, \leq, \tau)$ is called a generalized ordered space.*

Remark 4.2.12 Let $(X, \leq)$ be a totally ordered space and let $\tau_{\leq}$ be the order topology on X. If $Y \subseteq X$ then on the one hand $(Y, \leq_{|Y})$ is a totally ordered space and thus has an order topology $\tau_{\leq_{|Y}}$. On the other hand, Y has the subspace (or induced) topology $(\tau_{\leq})_Y$ induced from $\tau_{\leq}$. These topologies do not always coincide!! But a subspace of an ordered space always is a generalized ordered space. Conversely, every generalized ordered space can always be embedded (topologically and as an ordered space) into an ordered space. □

4.3 Neighborhood bases. First countability

Definition 4.3.1 *Let (X, τ) be a topological space and $x \in X$. A family $\mathcal{N} \subseteq P(X)$ is called a (open) neighborhood base of x if*

(i) Every $N \in \mathcal{N}$ is a (open) neighborhood of x. (I.e. $\mathcal{N} \subseteq \mathcal{N}_x$ resp. $\mathcal{N} \subseteq \mathcal{U}_x$.)

(ii) For every neighborhood $M \in \mathcal{N}_x$ of x there is an $N \in \mathcal{N}$ such that $N \subseteq M$.

Exercise 4.3.2 Let (X, τ) be a topological space. For every $x \in X$, let $\mathcal{V}_x$ be an open neighborhood base for x. Prove that $\mathcal{B} = \bigcup_x \mathcal{V}_x$ is a base for τ.

Lemma 4.3.3 *For a topological space (X, τ), the following are equivalent:*

(i) Every $x \in X$ has a countable neighborhood base $\mathcal{N}$.

(ii) Every $x \in X$ has a countable open neighborhood base $\mathcal{U}$.

(iii) Every $x \in X$ has an open neighborhood base $\mathcal{V} = \{V_i\}_{i \in \mathbb{N}}$ such that $V_1 \supseteq V_2 \supseteq \cdots$.

Proof. It is obvious that (iii)⇒(ii)⇒(i).

(i)⇒(ii) For every $N \in \mathcal{N}$ there exists, by definition of a neighborhood, an open U_N such that $x \in U_N \subseteq N$. Now $\mathcal{U} = \{U_N \mid N \in \mathcal{N}\}$ is a countable open neighborhood base for x.

(ii)⇒(iii) Let $\mathcal{U}$ be a countable open neighborhood base for x. Choose a bijection $\mathbb{N} \to \mathcal{U}$, $i \mapsto U_i$ and define $V_i = \bigcap_{k=1}^{i} U_k$. Then clearly $V_1 \supseteq V_2 \supseteq \cdots$, and $V_i \subseteq U_i \; \forall i$ implies that $\mathcal{V} = \{V_i\}_{i \in \mathbb{N}}$ is an open neighborhood base for x. (It should be clear how to modify this should $\mathcal{U}$ happen to be finite.) ■

Definition 4.3.4 *A topological space satisfying the equivalent statements in Lemma 4.3.3 is first countable or satisfies the first axiom of countability.*

Exercise 4.3.5 Prove that first countability is hereditary.

Lemma 4.3.6 *Every metric space has the first countability property.*

Proof. If U is open and $x \in X$, there is an $n \in \mathbb{N}$ such that $x \in B(x, 1/n) \subseteq U$. Thus for every $x \in X$,

$$\mathcal{N} = \{B(x, 1/n) \mid n \in \mathbb{N}\}$$

is a countable open neighborhood base of x. ■

Many results that are true for metric spaces actually only use the first countability. We will soon see this in Propositions 5.1.7 and 5.1.13. But see Remark 5.2.28.

Lemma 4.3.7 *The second countability property implies the first.*

Proof. Let $\mathcal{B}$ be a countable base for the topology and let $x \in X$. We claim that

$$\mathcal{N} = \{U \in \mathcal{B} \mid x \in U\},$$

which clearly is countable, is a neighborhood base for x. To see this, let $x \in V \in \tau$. By definition of a base, there is a $U \in \mathcal{B}$ such that $x \in U \subseteq V$. With the definition of $\mathcal{N}$, we have $U \in \mathcal{N}$, thus $\mathcal{N}$ is an open neighborhood base for x. ■

For Hausdorff spaces, the following is an improvement of Exercise 4.1.8 (since second countability implies both first countability and separability, but not conversely):

Exercise 4.3.8 Let X be a separable and first countable Hausdorff space. Prove that $\#X \leq \mathfrak{c}$.

Remark 4.3.9 Exercises 4.1.8 and 4.3.8 are just the beginning of the theory of 'cardinal functions', on which much research has been done in recent decades. Cf. [106, Sections a03, a04] and [152, Chapters 1,2]. □

Remark 4.3.10 Figure 4.1 summarizes the implications proven so far.

It is straightforward to check that these implications rule out 9 of the $2^4 = 16$ conceivable combinations of the first and second countability properties, separability and metrizability. All others are actually possible, as the following table shows:

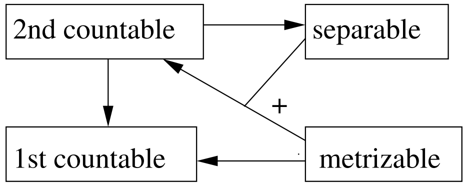


Figure 4.1: Implications between countability axioms

1st cnt.	2nd cnt.	separable	metrizable	Example
0	0	0	0	$(\{0,1\}, \tau_{\text{disc}})^N$ with $\#N > \mathfrak{c} = \#\mathbb{R}$
0	0	1	0	(X, τ_{cofin}) with X uncountable
1	0	0	0	$[0,1]^2$ with lexicogr. order topology
1	0	0	1	(X, τ_{disc}) or $\ell^2(X)$ for X uncountable
1	0	1	0	Sorgenfrey line
1	1	1	0	$(X, \tau_{\text{indisc}})$ with $\#X \geq 2$
1	1	1	1	finite discrete space

Most of these examples will be discussed below. (The first example uses the product topology, in particular Exercises 6.5.15, 6.5.24 and Corollary 6.5.36. The proof of non-metrizability of $[0,1]^2$ with the lexicographical order topology will only be given in Remark 7.7.28.) $\square$

Exercise 4.3.11 Consider $X = [0,1] \times [0,1]$ equipped with the order topology coming from the lexicographic order. Prove that X is first countable, but does not have the Souslin property (and thus is neither second countable nor separable).

Exercise 4.3.12 Prove that the Sorgenfrey line $(\mathbb{R}, \tau_S)$ is

(i) separable,

(ii) first countable,

(iii) not second countable,

(iv) not metrizable.

Exercise 4.3.13 Prove that the cofinite and cocountable topologies on an uncountable set are not first countable.

In the preceding exercise, the underlying space was uncountable. But there are also topologies on countable sets that not first countable, thus also not second countable! An example is provided in the next exercise:

Exercise 4.3.14 (Arens-Fort space) [6] Let $X = \mathbb{N}_0 \times \mathbb{N}_0$, where $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$. Define $\tau \subseteq P(X)$ as follows: $U \subseteq X$ is in τ if and only if U satisfies one of the following conditions:

- $(0,0) \notin U$.
- $(0,0) \in U$ and $U_m := \{n \in \mathbb{N}_0 \mid (m,n) \notin U\}$ is infinite for at most finitely many $m \in \mathbb{N}_0$. (Thus: At most finitely many columns of U lack infinitely many elements of X.)

[6] Richard Friederich Arens (1919-2000): German-American topologist and functional analyst. Marion Kirkland Fort, Jr. (1921-1964): American general topologist.

(i) Prove that τ is a topology.

(ii) Prove that τ is Hausdorff (T_2).

(iii) Prove that $(0,0) \in X$ cannot have a countable neighborhood base. Hint: Given open neighborhoods $\{U_i\}_{i\in\mathbb{N}}$ of $(0,0)$, construct an open $V \ni (0,0)$ that contains none of the U_i.

(iv) Conclude that (X,τ) has neither the first nor the second countability property.

(v) Conclude that countability of X does not imply that (X,τ) is first countable, nor vice versa.

Exercise 4.3.15 (The Niemytzki plane) [7] Let $B_r(x,y) \subseteq \mathbb{R}^2$ be the open ball around (x,y) with radius r. Define

$$\begin{aligned} L &= \{(x,0) \mid x \in \mathbb{R}\}, \\ Y &= \{(x,y) \in \mathbb{R}^2 \mid y > 0\}, \\ X &= Y \cup L = \{(x,y) \in \mathbb{R}^2 \mid y \geq 0\}. \end{aligned}$$

For $(x,y) \in X,\ r > 0$ define

$$U_{(x,y),r} = \begin{cases} B_r(x,y) \cap Y & \text{if} \quad y > 0 \\ B_r(x,r) \cup \{(x,0)\} & \text{if} \quad y = 0 \end{cases} \tag{4.4}$$

(i) Make a drawing where L, Y, X and $U_{(x,y),r}$ (with $y > 0$ and $y = 0$) can be understood.

(ii) Show that $B = \{U_{(x,y),r} \mid (x,y) \in X,\ r > 0\}$ is a base for a topology $\widetilde{\tau}$ on X.

(iii) Show that $\widetilde{\tau}\restriction Y$ is the standard topology, whereas the subspace $(L, \widetilde{\tau}\restriction L)$ is discrete.

(iv) Show that $(X,\widetilde{\tau})$ is first countable, but not second countable.

(v) With (iii) it follows that $A = \{(x,0) \mid x \in \mathbb{R}\backslash\mathbb{Q}\}$ and $B = \{(x,0) \mid x \in \mathbb{Q}\}$ are disjoint closed subsets of $(X,\widetilde{\tau})$. Show that there are no $U, V \in \widetilde{\tau}$ such that $A \subseteq U,\ B \subseteq V,\ U \cap V = \emptyset$. (Later we will say: $(X,\widetilde{\tau})$ is not normal (T_4).) Hint: Baire's theorem.

[7] Viktor Vladimirovich Nemytskii (1900-1967). Russian mathematician.

Chapter 5

Convergence and continuity

In general topology it is easy to forget that the initial aim of (general) topology was to provide a general framework for the study of convergence and continuity. The author is aware of a book in which these subjects appear on pages 254 and 175, respectively! (Admittedly, there are situations, mostly in algebra, where one studies certain topologies without being interested in convergence or continuity: The Krull topology on Galois groups, the Zariski topology on algebraic varieties, etc.) We therefore now turn to the first of these subjects, convergence.

5.1 Convergence in topological spaces: Sequences, nets, filters

5.1.1 Sequences

The Definition 2.1.14 of sequences obviously applies to topological spaces. But we need a new notion of convergence:

Definition 5.1.1 *Let (X, τ) be a topological space, $\{x_n\}_{n\in\mathbb{N}}$ a sequence in X and $z \in X$. We say that $\underline{x_n \text{ converges to } z}$ or $\underline{z \text{ is a limit of } x_n}$ if for every (open) neighborhood U of z there is an $N \in \mathbb{N}$ such that $n \geq N \;\Rightarrow\; x_n \in U$. In this case we write $x_n \to z$.*

Remark 5.1.2 1. $x_n \to z$ is equivalent to: For every neighborhood U of z, $x_n \in U$ for all but finitely many n.

2. It does not matter whether we include 'open' in the definition.

3. In order to show $x_n \to z$ it suffices to verify the condition in Definition 5.1.1 for the elements of any neighborhood base for z.

4. If (X, d) is a metric space and $\tau = \tau_d$ then convergence $x_n \to z$ in the metric (Definition 2.1.15) and topological (Definition 5.1.1) senses are equivalent since $\{B(z, \varepsilon) \mid \varepsilon > 0\}$ is a neighborhood base for z w.r.t. τ_d.

5. The notation $z = \lim_{n\to\infty} x_n$ suggests that there is a unique limit, which however is not always true! We will therefore only use this notation when uniqueness is known to hold, cf. Proposition 5.1.4, and write $x_n \to z$ otherwise. □

Exercise 5.1.3 Prove:

(i) If (X, τ) is indiscrete then *any* sequence $\{x_n\}$ in X converges to *any* $z \in X$.

(ii) If (X, τ) is discrete then a sequence $\{x_n\}$ in X converges to $z \in X$ if and only if there is $N \in \mathbb{N}$ such that $n \geq N \Rightarrow x_n = z$. (I.e. $\{x_n\}$ is 'eventually constant'.)

Proposition 5.1.4 *If (X, τ) is a Hausdorff space then every sequence has at most one limit. (I.e., if $x_n \to y$ and $x_n \to z$ then $y = z$.)*

Proof. Assume $x_n \to y$ and $x_n \to z$, where $y \neq z$. Since X is T_2 we can find $U, V \in \tau$ such that $y \in U, z \in V$ and $U \cap V = \emptyset$. Now, there are $N, M \in \mathbb{N}$ such that $n \geq N \Rightarrow x_n \in U$ and $n \geq M \Rightarrow x_n \in V$. Thus if $n \geq \max(N, M)$ then $x_n \in U \cap V$, but this contradicts $U \cap V = \emptyset$. ■

Lemma 5.1.5 *If $\{y_n\}_{n\in\mathbb{N}} \subseteq Y \subseteq X$ such that $y_n \to x$ then $x \in \overline{Y}$.*

Proof. As noted above, $y_n \to x$ means that every neighborhood U of x contains y_n for infinitely many n. This obviously implies $U \cap Y \neq \emptyset$ for every neighborhood U of x, and by Lemma 2.7.3 this is equivalent to $x \in \overline{Y}$. ■

Question: Is every $z \in \overline{Y}$ a limit of a sequence $\{y_n\}$ in Y? If X is metrizable, the answer is yes by Lemma 2.1.16 and Remark 2.7.4.2. But in general it is NO!

Example 5.1.6 Let X be uncountable and τ the cocountable topology on X. Let $Y \subsetneq X$ be uncountable but $Y \neq X$. By definition of the cocountable topology, we have $\overline{Y} = X$ (since the only closed uncountable subset is X). But if $\{y_i\}$ is a sequence in Y then $\{y_1, y_2, \ldots\} \subseteq X$ is countable, thus closed. Thus if $y_i \to z$ then $z \in \overline{\{y_1, y_2, \ldots\}} = \{y_1, y_2, \ldots\} \subseteq Y$. Thus no point $z \in X \backslash Y$ can be obtained as limit of a sequence in Y! As the next result shows, this phenomenon is closely related to the lack of first countability proven in Exercise 4.3.13. (But cf. Remark 5.1.9.2.) □

Proposition 5.1.7 *Let (X, τ) satisfy the first countability axiom. Then:*

(i) If $V_1 \supseteq V_2 \supseteq \cdots$ is a shrinking neighborhood base for $z \in X$ and $x_i \in V_i \ \forall i$ then $x_i \to z$.

(ii) The closure of any $Y \subseteq X$ coincides with the set of limits of sequences taking values in Y.

Proof. (i) Let W be any neighborhood of z. Since $\{V_i\}$ is a neighborhood base for x, there is $i \in \mathbb{N}$ such that $V_i \subseteq W$, thus also $V_j \subseteq W \ \forall j \geq i$ due to the shrinking character of the $V_i's$. By our choice of the sequence $\{x_i\}$ we then have $j \geq i \Rightarrow x_j \in V_j \subseteq W$. Thus $x_i \to z$.

(ii) Let $z \in \overline{Y}$. By Lemma 4.3.3 there is a shrinking open neighborhood base $V_1 \supseteq V_2 \supseteq \cdots$ for z. Lemma 2.7.3 implies $V_i \cap Y \neq \emptyset$ for all $i \in \mathbb{N}$. Thus for each i, we can choose $x_i \in V_i \cap Y$ (by countable choice). Thus $x_i \in V_i \ \forall i$, so that (i) gives $x_n \to z$. ■

Corollary 5.1.8 *A subset $Y \subseteq X$ of a first countable space X is closed if and only if $\{x_n\}_{n\in\mathbb{N}} \subseteq Y$ and $x_n \to y$ implies $y \in Y$.*

Remark 5.1.9 1. Since metric spaces are first countable, the proposition shows that Definition 2.4.1 is consistent with the definition of closedness in Lemma 2.1.16.

2. In the literature, e.g. [75], one finds notions of topological spaces that are slightly more general than the first countable ones: Spaces satisfying the conclusion of Proposition 5.1.7(ii) are called 'Fréchet spaces' (unrelated to the Fréchet spaces of functional analysis), and spaces with the property in Corollary 5.1.8) are 'sequential spaces'. Clearly: metrizable ⇒ first countable ⇒ Fréchet ⇒ sequential. Some of the results that we will prove for first countable spaces hold more generally for Fréchet or even sequential spaces. But it turns out that even the largest of these classes, that of sequential spaces, is still very close to metric spaces, cf. Remark 6.4.6. □

Definition 5.1.10 *If $\{x_n\}$ is a sequence in a topological space (X,τ), then $z \in X$ is called accumulation point of $\{x_n\}$ if for every open neighborhood U of z there are infinitely many $n \in \mathbb{N}$ such that $x_n \in U$. (Equivalently, there are arbitrarily large n such that $x_n \in U$.)*

Every limit of a sequence is an accumulation point, but the converse need not hold: The sequence $\{x_n = (-1)^n\}$ has ± 1 as accumulation points, but no limit.

Definition 5.1.11 *A subsequence of a sequence $\{x_n\}$ is a sequence of the form $\{x_{n_m}\}_{m\in\mathbb{N}}$, where $1 \le n_1 < n_2 < \cdots$ is a strictly increasing sequence in $\mathbb{N}$.*

Lemma 5.1.12 *Let $\{x_n\}$ be a sequence in the topological space (X,τ). If it has a subsequence $\{x_{n_m}\}$ converging to $z \in X$ then z is an accumulation point of $\{x_n\}$.*

Proof. Let U be an open neighborhood of z. Since the subsequence $\{x_{n_m}\}$ converges to z, there is an $N \in \mathbb{N}$ such that $m \ge N \Rightarrow x_{n_m} \in U$. But this implies that there are infinitely many n such that $x_n \in U$, thus z is an accumulation point. ■

Without further assumptions, it is not true that every accumulation point of a sequence is the limit of a subsequence! (Cf. Example 7.7.13.) In metric spaces there is no problem, but again first countability is sufficient:

Proposition 5.1.13 *If (X,τ) is first countable then for every accumulation point z of a sequence $\{x_n\}$ in X there is a subsequence $\{x_{n_m}\}$ converging to z.*

Proof. Let z be an accumulation point of the sequence. Let $V_1 \supseteq V_2 \supseteq \cdots$ be a shrinking open neighborhood base of z as in Lemma 4.3.3(iii). Since z is an accumulation point of $\{x_n\}$, there clearly is an n_1 such that $x_{n_1} \in V_1$. Since there are infinitely many n such that $x_n \in V_2$, we can find $n_2 > n_1$ such that $x_{n_2} \in V_2$. Continuing like this, we can obtain $n_1 < n_2 < n_3 < \cdots$ such that $x_{n_m} \in V_m \ \forall m$. Now Proposition 5.1.7(i) gives that $x_{n_m} \stackrel{m\to\infty}{\longrightarrow} z$. ■

5.1.2 Nets

Is there a way to repair the failure of Proposition 5.1.7 in the non-first countable situation? The most straightforward (but not always the best) solution is provided by nets. The idea is to replace the index set $\mathbb{N}$ by some other set that is allowed to depend on the situation. We also need a substitute for the ordering of $\mathbb{N}$. It turns out that instead of a total order, the following requirement is sufficient:

Definition 5.1.14 *A directed set is a pair $(I,\le)$, where I is a set and $\le$ is a binary relation on I that is reflexive and transitive (i.e. a preorder, cf. Definition A.1.8) and satisfies directedness, i.e. for any $\iota_1,\iota_2 \in I$ there is a $\iota_3 \in I$ such that $\iota_3 \ge \iota_1,\ \iota_3 \ge \iota_2$.*

Definition 5.1.15 *A net[1] in a space X consists of a directed set $(I,\le)$ and a map $I \to X, \iota \mapsto x_\iota$. (Usually we will denote the net as $\{x_\iota\}_{\iota\in I}$ or just $\{x_\iota\}$.)*

Remark 5.1.16 1. We follow the common practice of using lower case Greek letters for the elements of a directed set, but there are exceptions (as in the proof of Proposition 5.1.19).

2. In many situations, the relation $\le$ of a directed set will also satisfy antisymmetry and thus be a partial order. But not being part of the definition, this property will never be used in proofs.

[1] Nets were introduced in 1922 by the American mathematicians Eliakim H. Moore (1862-1932) and Herman L. Smith (1892-1950). For this reason, in the older literature one finds the term 'Moore-Smith convergence', but this name has now gone out of fashion.

3. Every totally ordered set $(I, \leq)$ is directed: Given $\iota_1, \iota_2 \in I$, put $\iota_3 = \max(\iota_1, \iota_2)$. In particular $(\mathbb{N}, \leq)$ is a directed set, thus every sequence is a net (with $I = \mathbb{N}$).

4. But not every partially ordered set is directed: If X is a set with $\#X \geq 2$, take $I = P(X)\backslash\{X\}$ and $\leq$ the inclusion order on I.) □

Definition 5.1.17 *(i) If $\{x_\iota\}_{\iota\in I}$ in a net in X and $Y \subseteq X$, we say x_ι is eventually in Y if there exists a $\iota_0 \in I$ such that $\iota \geq \iota_0 \Rightarrow x_\iota \in Y$.*

(ii) If X is a topological space and $y \in X$, a net $\{x_\iota\}_{\iota\in I}$ converges to $y \in X$ if it eventually is in every (open) neighborhood of y. We then write $x_\iota \to y$.

Remark 5.1.18 1. There is no point in writing $\iota \to \infty$: The notation $x_\iota \to y$ is unambiguous, the direction being built into the definition of a directed set.

2. If $\tau \subseteq \tau'$ are topologies on the same set X (thus τ' is finer than τ) then a net in X that converges w.r.t. τ' converges w.r.t. τ, but not vice versa. Thus finer topologies have fewer convergent nets. □

The next two results shows that nets not only do not share the defect of sequences encountered in Example 5.1.6, they are also sufficient to test whether a space is Hausdorff:

Proposition 5.1.19 *Let $Y \subseteq (X, \tau)$. Then the closure $\overline{Y}$ coincides with the set of limits of nets taking values in Y.*

Proof. If $\{y_\iota\}_{\iota\in I}$ is a net with values in $Y \subseteq X$ and $y_\iota \to x$, one proves exactly as for sequences that $x \in \overline{Y}$. Now let $x \in \overline{Y}$. We want to construct a net $\{y_\iota\}$ in Y such that $y_\iota \to x$. Let $I = \mathcal{U}_x = \{U \in \tau \mid x \in U\}$ (the set of open neighborhoods of x). We define a partial order on I by 'reverse inclusion', i.e. $U \leq V :\Leftrightarrow U \supseteq V$. (Thus the 'larger' elements of I w.r.t. $\leq$ are the smaller neighborhoods of x!) This clearly is a partial order, and it is directed: If $U, V \in I$ then $W = U \cap V$ is an open neighborhood of x, thus in I and $W \subseteq U, W \subseteq V$, thus $W \geq U, W \geq V$.

Since $x \in \overline{Y}$, every neighborhood $U \ni x$ satisfies $U \cap Y \neq \emptyset$ by Lemma 2.7.3. Thus we can define a map $I \to Y$ by assigning to each $U \in I$ a point $y_U \in U \cap Y$. This defines a net in Y indexed by $(I, \leq)$. In order to prove that it converges to x, let V be any (open) neighborhood of x. Thus $V \in I$, and for every $U \in I$ such that $U \geq V$, we have $U \subseteq V$ and thus $x_U \in U \subseteq V$. Thus $y_U \to x$. ■

Proposition 5.1.20 *A topological space (X, τ) is Hausdorff if and only if no net in X has two different limits.*

Proof. The proof of the implication $\Rightarrow$ is essentially the same as for sequences: If $x \neq y$ are limits with disjoint open neighborhoods U, V provided by Hausdorffness, let ι_1, ι_2 such that $\iota \geq \iota_1 \Rightarrow x_\iota \in U$ and $\iota \geq \iota_2 \Rightarrow x_\iota \in V$. Using directedness we find ι_3 such that $\iota_3 \geq \iota_1,\ \iota_3 \geq \iota_2$. Now $\iota \geq \iota_3$ implies $x_\iota \in U \cap V = \emptyset$, which is absurd.

Now assume that (X, τ) is not Hausdorff. Thus there are points $x \neq y$ such that whenever $U, V \in \tau$ with $x \in U, y \in V$ we have $U \cap V \neq \emptyset$. We will construct a net $\{x_\iota\}$ such that $x_\iota \to x$ and $x_\iota \to y$. Let $I = \mathcal{N}_x \times \mathcal{N}_y$. If $(P, Q), (R, S) \in I$, we say $(P, Q) \geq (R, S)$ if $P \subseteq R$ and $Q \subseteq S$. It is easy to see that this defines a directed partial order. Now define a map $I \to X$ as follows. To every $(U, V) \in I$ associate an arbitrary point $x_{(U,V)} \in U \cap V$. This can be done since any neighborhoods $U \in \mathcal{N}_x, V \in \mathcal{N}_y$ satisfy $U \cap V \neq \emptyset$. Now we claim that the net $\{x_{(U,V)}\}$ converges to x and to y. Namely, let A, B be neighborhoods of x and y, respectively. Thus $A \in \mathcal{N}_x, B \in \mathcal{N}_y$, and therefore $(A, B) \in I$. Now whenever $(A', B') \geq (A, B)$ we have $A' \subseteq A$ and $B' \subseteq B$ and thus $x_{(A',B')} \in A' \cap B' \subseteq A \cap B$. Thus the net converges to both x and y. ■

Remark 5.1.21 1. Thus in a Hausdorff space, the notation $\lim_\iota x_\iota = z$ is justified. In non-Hausdorff spaces it should be avoided since it misleadingly suggests uniqueness of limits.

2. If (X, τ) is non-Hausdorff, but first countable, one can combine the ideas of the proofs of Propositions 5.1.7 and 5.1.20 and construct a sequence that has two different limits.

3. The above proof shows clearly why the condition that $(I, \leq)$ be directed was imposed. This property is essential for most proofs involving nets. (Proposition 5.1.19 was an exception).

4. Nets are a straightforward generalization of sequences and easy to use (after getting used to them). For this reason they are popular with many analysts. Proofs involving convergence of nets have a somewhat 'dynamic' flair if one interprets the index $\iota \in I$ as 'time', which is consistent with the terms 'eventually' and 'frequently'. But some proofs involving nets look quite tautological, compare Propositions 5.1.19, 5.1.20, 5.2.5. In many situations one might prefer proofs that are more 'static' or set-theoretic and therefore avoid nets. One way to do this is using the notion of filters, briefly touched in the next section. (Most set-theoretic topologists seem to prefer filters over nets.) But often one can find proofs that avoid both filters and nets, as in our first proof of Tychonov's theorem in Section 7.5.2. (But we we will also give a proof using filters and two that use nets!) □

Example 5.1.22 (Unordered sums) Let $f : S \to \mathbb{C}$ be a function where S is an arbitrary set. We want to make sense of $\sum_{s \in S} f(s)$. Let $\mathcal{F}$ be the family of finite subsets of S. Partially ordered by (ordinary!) inclusion of sets, $\mathcal{F}$ clearly is a directed set. Thus the map $\mathcal{F} \to \mathbb{C}$ defined by $F \mapsto \sum_{s \in F} f(s)$ is a net. If $\lim_F \sum_{s \in F} f(s) = A$ exists (in the sense of Definition 5.1.17) we write $\sum_{s \in S} f(s) = A$. One can prove the following:

1. If $f(s) \geq 0 \ \forall s$ then $\sup_{F \in \mathcal{F}} \sum_{s \in F} f(s)$ always exists in $[0, \infty]$, and $\sum_{s \in S} f(s)$ exists if and only if the supremum is finite, in which case sum and supremum coincide.

2. If $f(s) \geq 0 \ \forall s$ and $\sup_{F \in \mathcal{F}} \sum_{s \in F} f(s) < \infty$ then $\{s \in S \mid f(s) \neq 0\}$ is at most countable by Lemma E.2.2.

3. If $\sum_{s \in S} |f(s)| < \infty$ then $\sum_{s \in S} f(s)$ exists and $|\sum_{s \in S} f(s)| \leq \sum_{s \in S} |f(s)|$.

4. Surprisingly (at first sight) the converse also holds: If $\sum_{s \in S} f(s)$ exists then $\sum_{s \in S} |f(s)| < \infty$! (For a proof see [17, Theorem 5.2.3].) If this seems to contradict what you know about convergence of series (indexed by $\mathbb{N}$), consider (i) that no order is given on the set S and (ii) Riemann's theorem saying that the sum of a series is invariant under permutation of the terms if and only if the series is absolutely convergent! □

Example 5.1.23 (Riemann integral) Let $f : [a, b] \to \mathbb{C}$. A partition of $[a, b]$ is a finite set $P \subseteq [a, b]$ such that $\{a, b\} \subseteq P$. The set $\mathcal{P}$ of partitions of $[a, b]$, ordered by inclusion, clearly is a directed set. Given a partition P, we can order its elements like $a = x_0 < x_1 < \cdots < x_N = b$, where $P = \{x_0, \ldots, x_N\}$, and define

$$\begin{aligned} U(f, P) &= \sum_{i=1}^{N} (x_i - x_{i-1}) \sup_{x \in [x_{i-1}, x_i]} f(x), \\ L(f, P) &= \sum_{i=1}^{N} (x_i - x_{i-1}) \inf_{x \in [x_{i-1}, x_i]} f(x). \end{aligned}$$

Now, $P \mapsto U(f, P)$, $P \mapsto L(f, P)$ are nets in $\mathbb{C}$ indexed by $\mathcal{P}$, and by Darboux' characterization of Riemann integrability (which in many books is taken as the definition of Riemann integrability, cf.

[201, 223]), f is Riemann integrable on $[a,b]$ if and only if $\lim_P(U(f,P)-L(f,P))=0$. In that case we have $\int_a^b f(x)dx := \lim_P U(f,P) = \lim_P L(f,P)$.

It is easy to prove that every $f \in C([a,b],\mathbb{C})$ is Riemann integrable. With more effort one proves that $f:[a,b]\to\mathbb{C}$ is Riemann integrable if and only if it is bounded and almost everywhere continuous (i.e. the set of points where f is not continuous has measure zero in the sense of Definition 11.2.11). □

Remark 5.1.24 On the basis of the foregoing examples, a topological imperialist would have a point in claiming that much of classical analysis, at least as long as no measure theory is involved, can be construed as a part of topology. (This actually seems to be the view of various analysts, e.g. [17].) At least it shows that any attempt at drawing a dividing line between topology and analysis is futile. □

Definition 5.1.25 *(i) A net $\{x_\iota\}_{\iota\in I}$ in X is frequently in $Y\subseteq X$ if for every $\iota_0\in I$ there exists $\iota\geq\iota_0$ such that $x_\iota\in Y$.*

(ii) A point $x\in X$ is an accumulation point of $\{x_\iota\}$ if it is frequently in every (open) neighborhood $U\ni x$.

Remark 5.1.26 1. Clearly every limit of a net is an accumulation point, but not conversely.

2. For a net defined on $I=\mathbb{N}$ with the usual order $\leq$, the above definition of accumulation point is equivalent to the one in Definition 5.1.10. □

Exercise 5.1.27 Prove that a net is frequently in $Y\subseteq X$ if and only if it is not eventually in $X\backslash Y$.

We would like to have a notion of subnets of a net $\{x_\iota\}$, such that $z\in X$ is a limit of a subnet if and only if z is an accumulation point (without assuming first countability of course). Such a notion exists, but it is not entirely obvious. (There even is some controversy as to what the 'right' definition is, cf. Remarks 5.1.32 and 7.5.27.)

Definition 5.1.28 *Let $(I,\leq),(J,\leq)$ be directed sets. A map $\alpha:J\to I$ is called cofinal if for every $\iota_0\in I$ there is a $\lambda_0\in J$ such that $\lambda\geq\lambda_0 \Rightarrow \alpha(\lambda)\geq\iota_0$.*

Definition 5.1.29 *Let $(I,\leq)$ be a directed set and $\{x_\iota\}_{\iota\in I}$ a net in X. If $(J,\leq)$ is another directed set, a net $J\to X$, $\lambda\mapsto y_\lambda$ is called a subnet of $\{x_\iota\}$ if there is a cofinal map $\alpha:J\to I$ such that $y_\lambda=x_{\alpha(\lambda)}$.*

Proposition 5.1.30 *A point $z\in X$ is an accumulation point of the net $\{x_\iota\}$ if and only if $\{x_\iota\}$ has a subnet that converges to z.*

Proof. Assume that $\{x_\iota\}$ has a subnet converging to z. I.e., there is a directed set $(J,\leq)$ and a cofinal map $\alpha:J\to I$ such that $x_{\alpha(\lambda)}\to z$. Let U be an open neighborhood of z and let $\iota_0\in I$. Since α is cofinal, there is $\lambda_0\in J$ such that $\lambda\geq\lambda_0 \Rightarrow \alpha(\lambda)\geq\iota_0$, and since $\lambda\mapsto x_{\alpha(\lambda)}$ converges to z, there is $\lambda_1\in J$ such that $\lambda\geq\lambda_1 \Rightarrow x_{\alpha(\lambda)}\in U$. Thus if $\lambda\geq\lambda_0$ and $\lambda\geq\lambda_1$ then $\iota=\alpha(\lambda)\geq\iota_0$ and $x_\iota\in U$. Thus x_ι frequently is in U, i.e. z is an accumulation point.

If, conversely, z is an accumulation point of $\{x_\iota\}$ then applying the lemma below to $\mathcal{F}=\mathcal{N}_x$ (or $\mathcal{U}_x$), we obtain a subnet that eventually is in every $N\in\mathcal{N}_x$, i.e. converges to z. ■

Lemma 5.1.31 ([189]) *Let $\{x_\iota\}$ be a net in X and $\mathcal{F} \subseteq P(X)$ a family of subsets such that (i) x_ι is frequently in every $F \in \mathcal{F}$ and (ii) given $F_1, F_2 \in \mathcal{F}$, there is $F_3 \in \mathcal{F}$ with $F_3 \subseteq F_1 \cap F_2$. Then there is a subnet $J \to X,\ \lambda \mapsto x_{\alpha(\lambda)}$ that is eventually in every $F \in \mathcal{F}$.*

Proof. Let $J = \{(\iota, F) \in I \times \mathcal{F} \mid x_\iota \in F\}$ and define

$$(\iota_1, F_1) \geq (\iota_2, F_2) \quad \Leftrightarrow \quad \iota_1 \geq \iota_2 \ \text{ and } \ F_1 \subseteq F_2.$$

Let $(\iota_1, F_1), (\iota_2, F_2) \in J$. Pick $F_3 \in \mathcal{F}$ with $F_3 \subseteq F_1 \cap F_2$. Since x_ι frequently is in F_3, there is ι_3 such that $\iota_3 \geq \iota_1, \iota_3 \geq \iota_2$ and $x_{\iota_3} \in F_3$. Thus $(\iota_3, F_3) \in J$ and $(\iota_3, F_3) \geq (\iota_k, F_k)$ for $k = 1, 2$. Thus $(J, \leq)$ is directed. Define $\alpha : J \to I$ by $(\iota, F) \mapsto \iota$. This map is cofinal: Given $\iota_0 \in I$, take $\lambda_0 = (\iota_0, X)$, which clearly is in J. Now $\lambda = (\iota, F) \geq \lambda_0$ means $\iota \geq \iota_0$, thus $\alpha(\iota, F) = \iota \geq \iota_0$.

It remains to show that $\lambda \mapsto x_{\alpha(\lambda)}$ satisfies the last claim. So let $F \in \mathcal{F}$. Since x_ι frequently is in F, we can find $\iota_0 \in I$ such that $x_{\iota_0} \in F$. Then $\lambda_0 := (\iota_0, F)$ is in J. Now it is clear that $\lambda = (\iota, F') \geq \lambda_0$ implies $F' \subseteq F$ and $x_{\alpha(\iota, F')} = x_\iota \in F' \subseteq F$. Thus $\lambda \mapsto x_{\alpha(\lambda)}$ is eventually in F. ■

Remark 5.1.32 1. Since every sequence is a net, we see that an accumulation point of a sequence always is the limit of a sub**net**, whether or not the space is first countable.

2. As mentioned above, there are variants of the definition of subnets. Some authors, e.g. [237], require the map $\alpha : J \to I$ to be increasing, i.e. $\lambda_2 \geq \lambda_1 \ \Rightarrow \ \alpha(\lambda_2) \geq \alpha(\lambda_1)$. (In the presence of this requirement, the cofinality can be simplified to: $\forall \iota \in I\ \exists \lambda \in J : \alpha(\lambda) \geq \iota$.) This definition is more restrictive than Definition 5.1.29. But since the subnets produced in the proof of Lemma 5.1.31 clearly satisfies the stronger condition, the Lemma and the Proposition are true for both definitions of subnets. □

Exercise 5.1.33 Let $\{x_\iota\}_{\iota \in I}$ be a net in the space X. Prove:

(i) If $x_\iota \to x$ then every subnet of $\{x_\iota\}$ converges to x.

(ii) If $x \in X$ and every subnet of $\{x_\iota\}$ has a subnet converging to x then $x_\iota \to x$.

Note: We do not require that every subnet of $\{x_\iota\}$ converge to x !

5.1.3 ⋆ Filters

In this section we briefly look at the notion of filters, which are an alternative to nets. At first encounter, filters may be less intuitive, but they have some advantages, like less redundancy in the proofs and fewer invocations of the axiom of choice. Here we limit ourselves to the basics. For more on filters, cf. Section 7.5.4. Cf. also Remark 7.5.27 for some opinions.

Definition 5.1.34 *If X is a set, a filter[2] on X is a family $\mathcal{F} \subseteq P(X)$ of subsets satisfying:*

(i) If $F, G \in \mathcal{F}$ then $F \cap G \in \mathcal{F}$.

(ii) If $F \in \mathcal{F}$ and $G \supseteq F$ then $G \in \mathcal{F}$.

(iii) $\emptyset \notin \mathcal{F}$.

---

[2] Filters were invented in 1937 by the French mathematician Henri Cartan (1904-2008), an important member of the Bourbaki group. Unsurprisingly, the best reference on filters is [25].

Preference for nets or filters is sometimes put as a question of American vs. European (in particular French) tastes, but this is simplistic. Most contemporary research in general topology is actually done in terms of filters, not nets.

(iv) $\mathcal{F} \neq \emptyset$.

In view of Lemma 2.7.2, the family $\mathcal{N}_x$ of not-necessarily-open neighborhoods of a point is a filter, the neighborhood filter of x.

Definition 5.1.35 *A filter $\mathcal{F}$ on a topological space X is said to converge to $x \in X$ if it contains the neighborhood filter of x, i.e. $\mathcal{N}_x \subseteq \mathcal{F}$.*

Exercise 5.1.36 Prove that a space is Hausdorff if and only if every filter in it converges to at most one point.

We briefly consider the connection between filters and nets.

Exercise 5.1.37 Let $\{x_\iota\}$ be a net in (X, τ). Prove:

(i) The family $\mathcal{F} = \{F \subseteq X \mid x_\iota \text{ is eventually in } F\}$ is a filter.

(ii) The filter $\mathcal{F}$ converges to x if and only if the net $\{x_\iota\}$ does so.

Remark 5.1.38 An alternative proof for the 'if' ($\Leftarrow$) direction in Exercise 5.1.36 can be given using nets: If X is not Hausdorff then by Proposition 5.1.20 there is a net $\{x_\iota\}$ in X that converges to two different points. But then the filter $\mathcal{F}$ associated with $\{x_\iota\}$, cf. Exercise 5.1.37 converges to these different points. □

Exercise 5.1.39 Let $\mathcal{F}$ be a filter in (X, τ).

(i) Prove that $(\mathcal{F}, \leq)$, where $\leq = \supseteq$ (reverse inclusion), is a directed set.

In view of filter axiom (iii) there are nets $\{x_N\}_{N \in \mathcal{F}}$ satisfying $x_N \in N$ for all $N \in \mathcal{F}$.

(ii) Prove: If $\mathcal{F}$ converges to $z \in X$ then every net $\{x_N\}_{N \in \mathcal{F}}$ satisfying $x_N \in N$ converges to z.

(iii) Prove that if $\mathcal{F}$ does not converge to z then there is a net $\{x_N\}_{N \in \mathcal{F}}$ satisfying $x_N \in N\ \forall N$ that does not converge to z.

There is a notion of a base for a filter, somewhat analogous to that of a base for a topology:

Definition 5.1.40 *Let $\mathcal{F} \subseteq P(X)$ be a filter in X. A subset $\mathcal{B} \subseteq \mathcal{F}$ is called filter base for $\mathcal{F}$ if every $F \in \mathcal{F}$ contains some $B \in \mathcal{B}$. Equivalently, $\mathcal{F} = \{F \subseteq X \mid \exists B \in \mathcal{B} : B \subseteq \mathcal{F}\}$.*

In analogy to bases for a topology we can ask which subsets of $P(X)$ can be filter bases:

Lemma 5.1.41 *Let X be a set and $\mathcal{B} \subseteq P(X)$. Then $\mathcal{F} = \{F \subseteq X \mid \exists B \in \mathcal{B} : B \subseteq \mathcal{F}\}$ is a filter, clearly having $\mathcal{B}$ as filter base, if and only if*

- $\mathcal{B} \neq \emptyset$.
- $\emptyset \notin \mathcal{B}$.
- *If $B_1, B_2 \in \mathcal{B}$ then there is $B_3 \in \mathcal{B}$ such that $B_3 \subseteq B_1 \cap B_2$.*

Proof. $\Rightarrow$ Assume $\mathcal{F}$ is a filter. $\mathcal{B} = \emptyset$ implies $\mathcal{F} = \emptyset$ and $\emptyset \in \mathcal{B}$ implies $\emptyset \in \mathcal{F}$, both of which are forbidden. Thus $\mathcal{B} \neq \emptyset \not\in \mathcal{B}$. If $B_1, B_2 \in \mathcal{B}$ then $B_1, B_2 \in \mathcal{F}$, thus $B_1 \cap B_2 \in \mathcal{F}$, so that by definition of $\mathcal{F}$ there is a $B_3 \in \mathcal{B}$ with $B_3 \subseteq B_1 \cap_2$.

$\Leftarrow$ Obviously $\mathcal{B} \neq \emptyset$ implies $\mathcal{F} \neq \emptyset$. If $F \in \mathcal{F}$ then $F \supseteq B$ for some $B \in \mathcal{B}$. By assumption $B \neq \emptyset$, thus $F \neq \emptyset$. It is evident from the definition that $\mathcal{F}$ is closed under supersets. If $F_1, F_2 \in \mathcal{F}$, there are $B_1, B_2 \in \mathcal{B}$ with $B_i \subseteq F_i$. Then $F_1 \cap F_2 \supseteq B_1 \cap B_2 \supseteq B_3$ for some $B_3 \in \mathcal{B}$. Thus $F_1 \cap F_2 \in \mathcal{F}$. This proves that $\mathcal{F}$ is a filter. ■

Anticipating the result proven next, we can give a characterization of continuity of a function at a point in terms of the neighborhood filters:

Lemma 5.1.42 *Let X, Y be topological spaces and $f : X \to Y$. Then f is continuous at $x \in X$ if and only if $\{f(N) \mid N \in \mathcal{N}_x\}$ is a filter base for the filter $\mathcal{N}_{f(x)}$ in Y.*

Proof. The above statement is equivalent to saying that for every $M \in \mathcal{N}_{f(x)}$ there is an $N \in \mathcal{N}_x$ such that $f(N) \subseteq M$. This clearly is equivalent to Exercise 5.2.1(iii), and therefore to continuity of f at x. ■

5.2 Continuous, open, closed functions. Homeomorphisms

Having discussed the basics of the notion of convergence, we now turn to continuity. In (general) topology, the notion of continuous functions is much more important than convergence, since it provides the morphisms in the category of topological spaces, cf. Definition 5.2.14.

5.2.1 Continuity at a point

Exercise 5.2.1 Let $(X, \tau), (Y, \sigma)$ be topological spaces, $f : X \to Y$ a function and $x \in X$. Prove that the following are equivalent:

(i) For every open neighborhood V of $f(x)$ there exists an open neighborhood U of x such that $f(U) \subseteq V$.

(ii) For every open neighborhood V of $f(x)$ we have that $f^{-1}(V) \subseteq X$ is a neighborhood of x (not necessarily open). (I.e.: $U \in \mathcal{U}_{f(x)} \Rightarrow f^{-1}(U) \in \mathcal{N}_x$.)

(iii) For every $N \in \mathcal{N}_{f(x)}$ we have $f^{-1}(N) \in \mathcal{N}_x$.

Remark 5.2.2 1. Note that in (i) we did *not* require that $f^{-1}(U) \subseteq X$ be open for every open neighborhood U of $f(x)$, which potentially is a stronger statement!

2. Continuity of f at x implies $\bigcap_{U \in \mathcal{U}_x} f(U) \subseteq \bigcap_{V \in \mathcal{U}_{f(x)}} V \equiv \bigcap \mathcal{U}_{f(x)}$. If Y is T_1 then the r.h.s. equals $\{f(x)\}$ (Exercise 2.5.1), thus $\bigcap_{U \in \mathcal{U}_x} f(U) = \{f(x)\}$. (Deducing continuity of f at x from $\bigcap_{U \in \mathcal{U}_x} f(U) = \{f(x)\}$ is much harder and works only under restrictive additional assumptions. □

Definition 5.2.3 *Let $(X, \tau), (Y, \sigma)$ be topological spaces, $f : X \to Y$ a function and $x \in X$. If the equivalent statements in Exercise 5.2.1 hold, then f is called <u>continuous at x</u> and x is a <u>continuity point</u> of f.*

Exercise 5.2.4 Let $(X, d), (Y, d')$ be metric spaces and equip them with the metric topologies. For a function $f : X \to Y$, prove that the following are equivalent:

(i) f is continuous at $x \in X$ in the above (topological) sense.

(ii) f is continuous at $x \in X$ in the (metric) sense of Definition 2.1.21.

(iii) $f(x_n) \to f(x)$ for every sequence $\{x_n\}$ such that $x_n \to x$.

As recalled in Exercise 5.2.4, for metric spaces $(X,d),(Y,d')$ we have continuity of $f : X \to Y$ at $x \in X$ if and only if $f(x_n) \to f(x)$ for every sequence $\{x_n\}$ such that $x_n \to x$. As we have seen, for general topological spaces, sequences are not good enough to probe closedness, and the same is true for continuity. Nets do not have this defect, and indeed they can be used to characterize continuity of functions:

Proposition 5.2.5 *Let $(X,\tau),(Y,\sigma)$ be topological spaces and $f : X \to Y$ a function. Then f is continuous at $x \in X$ if and only if $f(x_\iota) \to f(x)$ for every net $\{x_\iota\}_{\iota\in I}$ in X such that $x_\iota \to x$.*

Proof. ($\Rightarrow$) Let f be continuous at x and let $x_\iota \to x$. We want to prove that $f(x_\iota) \to f(x)$. Let V be an open neighborhood of $f(x)$. By continuity of f at x, there is an open neighborhood U of x such that $f(U) \subseteq V$, cf. Exercise 5.2.1. Since $x_\iota \to x$, there is $\iota_0 \in I$ such that $\iota \ge \iota_0 \;\Rightarrow\; x_\iota \in U$. But by construction, $f(U) \subseteq V$, thus $\iota \ge \iota_0 \;\Rightarrow\; f(x_\iota) \in V$. Thus $f(x_\iota) \to f(x)$.

($\Leftarrow$) Assume that f is not continuous at x. Then by (ii) of Exercise 5.2.1, there exists an open neighborhood V of $f(x)$ such that $f(U) \not\subseteq V$ for every open neighborhood U of x. Thus, for every $U \in \mathcal{U}_x$ we can choose a point $x_U \in U$ such that $f(x_U) \not\in V$. Taking $I = \mathcal{U}_x$ and $\le = \supseteq$ as in the proof of Proposition 5.1.19, $(I,\le)$ is a directed set and the net $\{x_U\}_{U\in I}$ converges to x. But since for every U we have $f(x_U) \not\in V$, the net $\{f(x_U)\}_{U\in I}$ is not eventually in the neighborhood V of $f(x)$, thus does not converge to $f(x)$. ■

Remark 5.2.6 1. In order to conclude that f is continuous at $x \in X$ it is enough to prove that $f(x_\iota) \to f(x)$ for every net $\{x_\iota\}$ in $X\backslash\{x\}$ such that $x_\iota \to x$. This is because the net constructed in the proof of $\Leftarrow$ automatically satisfies $x_U \neq x$ for all U.

2. Let X, Y be topological spaces, where X is first countable, and $f : X \to Y$ is not continuous at $x \in X$. Let V be as in the proof of Proposition 5.2.5 $\Leftarrow$, and let $U_1 \supseteq U_2 \supseteq \cdots$ a decreasing countable open neighborhood base at x. Then for every $n \in \mathbb{N}$ we can choose $x_n \in U_n$ such that $f(x_n) \not\in V \;\forall n$. Then $\{x_n\}$ is a *sequence* such that $x_n \to x$ and $f(x_n) \not\to f(x)$.

3. If $f : X \to Y$ is function and $x_0 \in X,\; y_0 \in Y$, one shows (similarly to Proposition 5.2.5) that the following are equivalent:

(i) For every net $\{x_\iota\}$ in $X\backslash\{x_0\}$ that converges to x_0, the net $\{f(x_\iota)\}$ converges to y_0.

(ii) For every open neighborhood V of y_0, there is an open neighborhood U of x_0 such that $f(U\backslash\{x_0\}) \subseteq V$.

(iii) The function $\widetilde{f}$ defined by $\widetilde{f}(x_0) = y_0$ and $\widetilde{f}(x) = f(x)$ for $x \neq x_0$ is continuous at x_0. (Thus f at worst has a discontinuity at x_0 that can be removed by changing $f(x_0)$.)

If these (equivalent) conditions (for which the value $f(x_0)$ is irrelevant!) are satisfied one says $\underline{f(x) \text{ converges to } y_0 \text{ as } x \to x_0}$, in symbols $\lim_{x\to x_0} f(x) = y_0$. In particular, f is continuous at x_0 if and only if $\lim_{x\to x_0} f(x) = f(x_0)$. In this sense, the notion of convergence at a point generalizes that of continuity at a point. It should be noted that this concept is more popular with analysts than with topologists. □

5.2.2 Continuous functions. The category $\mathcal{TOP}$

Definition 5.2.7 *Let $(X, \tau), (Y, \sigma)$ be topological spaces and $f : X \to Y$. Then f is called continuous if $f^{-1}(U) \in \tau$ for every $U \in \sigma$. The set of continuous functions $X \to Y$ is usually denoted $C(X, Y)$, suppressing the topologies. (We may occasionally write $C(X)$ instead of $C(X, \mathbb{R})$.)*

Exercise 5.2.8 Let $(X, \tau), (Y, \sigma)$ be topological spaces and $f : X \to Y$ a function. Prove that the following are equivalent:

(i) f is continuous (in the sense of Definition 5.2.7).

(ii) $f^{-1}(C) \subseteq X$ is closed for every closed $C \subseteq Y$.

(iii) $f^{-1}(U)$ is open for every U in a base (or subbase) of σ.

(iv) f is continuous at every $x \in X$.

(v) $f(\overline{Z}) \subseteq \overline{f(Z)}$ for every $Z \subseteq X$.

Hint: The implications involving (v) are somewhat more difficult, but very instructive. For (i)⇒(v) assume $x \in \overline{Z}$ and use Lemma 2.7.3 twice to prove $f(x) \in \overline{f(Z)}$. To obtain (v)⇒(ii), assume $C \subseteq Y$ is closed and apply (v) to $Z = f^{-1}(C)$ to prove that $f^{-1}(C)$ is closed.

Exercise 5.2.9 Prove the equivalence (i)⇔(v) of Exercise 5.2.8 using nets.

Exercise 5.2.10 Let $(X, \tau), (Y, \sigma)$ be topological spaces. Prove:

(i) If X is discrete or Y is indiscrete then every function $f : X \to Y$ is continuous.

(ii) If X is connected and Y is discrete then every continuous $f : X \to Y$ is constant. (This will be generalized later.)

(iii) If X is irreducible and Y Hausdorff, then every continuous $f : X \to Y$ is constant.

Exercise 5.2.11 Prove:

(i) If $f : X \to Y$ is continuous and surjective and $A \subseteq X$ is dense then $f(A) \subseteq Y$ is dense.

(ii) If X is separable and $f : X \to Y$ is continuous and surjective then Y is separable.

Remark 5.2.12 If X, Y are metric spaces, (2.6) defines a metric D on the space $B(X, Y)$ of bounded functions from X to Y, and $C_b(X, Y) = C(X, Y) \cap B(X, Y) \subseteq B(X, Y)$ is a closed subspace, cf. Proposition 2.1.25. Now that we have the notion of continuity between topological spaces, we can generalize this to the case where (X, τ) only is a topological space. Reexamining the proofs of Propositions 2.1.25 and 3.1.8, one finds that they easily generalize. (Essentially the only thing one needs to change is to replace the $\delta > 0$ in the proof of Proposition 2.1.25(ii) by an open neighborhood $U \ni x$ such that $y \in U \Rightarrow d'(f_N(x), f_N(y)) < \varepsilon/3$.) □

Exercise 5.2.13 Let $f : X \to Y$ and $g : Y \to Z$ be continuous. Prove that $g \circ f$ is continuous.

Given the preceding result, it is natural to start using some categorical language, cf. Appendix A.5:

Definition 5.2.14 *Topological spaces and continuous maps form a* category $\mathcal{TOP}$. *Its objects are topological spaces, and for topological spaces X, Y we have* $\mathrm{Hom}_{\mathcal{TOP}}(X, Y) = C(X, Y)$. *Composition of morphisms is given by composition of maps (cf. the preceding exercise), and the identity morphism* id_X *of every space X just is the identity map $x \mapsto x$.*

The full subcategory consisting of Hausdorff spaces is denoted $\mathcal{TOP}_{T_2}$.

Exercise 5.2.15 Prove: If Y is T_i, where $i \in \{1, 2\}$, and $f : X \to Y$ is continuous and injective then X is T_i.

Exercise 5.2.16 For Y Hausdorff and $f, g : X \to Y$ continuous, prove:

(i) The coincidence set $C = \{x \in X \mid f(x) = g(x)\} \subseteq X$ of f and g is closed. Hint: Prove that $X \backslash C$ is open.

(ii) If f and g coincide on a dense subset of X then $f = g$.

(iii) If $A \subseteq B \subseteq \overline{A} \subseteq X$ and $f \in C(A, Y)$, then f has at most one extension $\widehat{f}$ to B. (I.e. a function $\widehat{f} \in C(B, Y)$ such that $\widehat{f} \restriction A = f$.)

Remark 5.2.17 1. The uniqueness result in (iii) will be extremely useful. Assume that $f \in C(X, Y)$ has dense image and that $g, h \in C(Y, Z)$, where Z is Hausdorff, satisfy $g \circ f = h \circ f$. Then f and g coincide on the dense subset $f(X) \subseteq Y$, so that Exercise 5.2.16(ii) implies $g = h$. In categorical language this means: 'In the category $\mathcal{TOP}_{T_2}$ of Hausdorff spaces, every continuous map with dense image is an epimorphism', cf. Definition A.5.5. The converse is also true: Every epimorphism in the category $\mathcal{TOP}_{T_2}$ has dense image, cf. Remark 6.6.6. (But: In the category $\mathcal{TOP}$ of *all* topological spaces, epimorphisms coincide with surjective maps.)

2. Proving existence of extensions is more difficult and requires further assumptions. Given a continuous function $f : X \supseteq A \to Y$, where Y is Hausdorff, one can try to extend it to $\overline{A}$ as follows: For $x \in \overline{A}$ and a net $\{x_\iota\}$ in A converging to x, the net $\{f(x_\iota)\}$ in Y has at most one limit z. If this limit exists and is independent of the chosen net $\{x_\iota\}$, this defines $\widehat{f}(x)$. In practice, one prefers a more set (in fact filter) theoretic approach over the use of nets. Cf. Proposition 3.4.10 in a metric setting (using completeness) and Theorem 7.4.20 for topological spaces (using compactness).

3. If $A \subseteq X$ is closed the above strategy for extending $f : A \to Y$ to some $B \supsetneq A$ does not work. Yet, there are some existence results, cf. Theorem 8.2.20 and 8.5.36, but little on uniqueness. □

By Proposition 5.2.5, continuity of a function $f : X \to Y$ (at a point $x \in X$) can be interpreted in terms of convergence of nets (or sequences, in favorable cases). But from a categorical point of view one can argue that the concept of continuity (of functions) is more fundamental than that of convergence (of sequences/nets), since continuous functions are the morphisms in the category $\mathcal{TOP}$, whereas the conceptual meaning of (convergent) sequences/nets is less clear. It therefore is interesting that convergence can be considered as a special case of continuity:

Exercise 5.2.18 (i) Write $\mathbb{N}_\infty = \mathbb{N} \cup \{\infty\}$ and define

$$\tau_\infty = P(\mathbb{N}) \cup \{\mathbb{N}_\infty \backslash F \mid F \subseteq \mathbb{N} \text{ finite}\}.$$

Prove that τ_∞ is a topology on $\mathbb{N}_\infty$.

(ii) If (X, τ) is a topological space, $\{x_n\}_{n \in \mathbb{N}}$ a sequence in X, and $z \in X$, define $f : \mathbb{N}_\infty \to X$ by $f(n) = x_n$ $\forall n \in \mathbb{N}$ and $f(\infty) = z$. Prove that the following are equivalent:

(α) $f : (\mathbb{N}_\infty, \tau_\infty) \to (X, \tau)$ is continuous.

(β) f is continuous at ∞.

(γ) $\lim_{n\to\infty} x_n = z$.

(The space $(\mathbb{N}_\infty, \tau)$ has natural interpretations, cf. Exercise 5.2.22 and Remark 7.8.16.3.)

(iii) (Bonus) Can you generalize (i), (ii) to nets?

5.2.3 Homeomorphisms. Open and closed functions

Definition 5.2.19 *A function $f : (X, \tau) \to (Y, \sigma)$ is called a homeomorphism if it is bijective, continuous, and the inverse function $f^{-1} : Y \to X$ is continuous.*

Two spaces $(X, \tau), (Y, \sigma)$ are called homeomorphic ($X \cong Y$) if and only if there exists a homeomorphism $f : X \to Y$.

Exercise 5.2.20 Prove:

(i) All open intervals $(a, b) \subseteq \mathbb{R}$ are mutually homeomorphic, and the same holds with 'open' replaced by 'closed' or by 'half-open', meaning $(a, b]$ and $[a, b)$.

Remark: Later we will prove that $(a, b), [a, b], [a, b)$ are mutually non-homeomorphic.

(ii) The functions from $(-1, 1)$ to $\mathbb{R}$ given by

$$f_1 : x \mapsto \frac{x}{1 - |x|}, \qquad f_2 : x \mapsto \frac{x}{\sqrt{1 - x^2}}, \qquad f_3 : x \mapsto \tan \frac{x\pi}{2}$$

are homeomorphisms. Give the inverse functions.

(iii) Prove $[0, 1) \cong [0, \infty)$.

(iv) Prove $\mathbb{R} \cong (0, \infty)$ in two ways: Using (i)-(iii), and using $x \mapsto e^x$.

Remark 5.2.21 1. A continuous function $f : (X, \tau) \to (Y, \sigma)$ is a homeomorphism if and only if there is a continuous function $g : (Y, \sigma) \to (X, \tau)$ such that $g \circ f = \mathrm{id}_X$ and $f \circ g = \mathrm{id}_Y$. Thus homeomorphisms are the isomorphisms in the category $\mathcal{TOP}$.

2. Just as groups (rings, fields, vector spaces) that are isomorphic in their respective categories are the "same" for all purposes of algebra, homeomorphic topological spaces are the "indistinguishable" for the purposes of topology. The best one could hope for in topology would be classification of topological spaces up to homeomorphism. (This is still completely hopeless.)

3. A property P that a topological space may have or not (like T_1, T_2, metrizability, first or second countability) is called topological if a space Y homeomorphic to X has property P if and only if X has it. Some authors make a big point out of pointing out for every property that they define that it is topological (or giving the proofs as exercises). This seems rather pointless, since it is utterly obvious for all properties defined purely in terms of the topology τ. (After all, every bijection $f : X \to Y$ gives rise to a canonical bijection $f' : P(X) \to P(Y)$, and $f : (X, \tau) \to (Y, \sigma)$ is a homeomorphism if and only if f' restricts to a bijection $\tau \to \sigma$.) The property "$42 \in X$" clearly is not topological, and the author is not aware of less artificial examples. (Completeness of metric spaces is not preserved under homeomorphisms, thus it is not a topological property. But it is not even a property of topological spaces!)

4. Recall the Bernstein-Schröder theorem from set theory: If there are injective maps $X \to Y$ and $Y \to X$ then there is a bijection $X \stackrel{\cong}{\to} Y$. In topology this is not true! On can find topological spaces $(X,\tau),(Y,\sigma)$ and continuous bijections $X \to Y$ and $Y \to X$ such that X and Y are non-homeomorphic! Cf. [60, p.112].

5. Isometric bijections between metric spaces are the isomorphisms in the category of metric spaces and isometric maps. □

Exercise 5.2.22 Prove that the space $(\mathbb{N}_\infty, \tau)$ from Exercise 5.2.18 is homeomorphic to the subspace $X = \{1/n \mid n \in \mathbb{N}\} \cup \{0\}$ of $(\mathbb{R}, \tau_d)$ (τ_d is the Euclidean topology).

Definition 5.2.23 *A function $f : (X,\tau) \to (Y,\sigma)$ is called open (resp. closed) if $f(Z) \subseteq Y$ is open (resp. closed) whenever $Z \subseteq X$ is open (resp. closed).*

Remark 5.2.24 If $f : (X,\tau) \to (Y,\sigma)$, where $\mathcal{B}$ is a base for τ, then f is open if and only $f(U) \in \sigma$ for each $U \in \mathcal{B}$. (The analogous statement for subbases need not be true!) □

The following is an immediate consequence of the definitions:

Lemma 5.2.25 *Let X be a set and τ_1, τ_2 topologies on X. Let $\mathrm{id}_X : x \mapsto x$ be the identical function on X. Then:*

(i) $\mathrm{id}_X : (X,\tau_1) \to (X,\tau_2)$ *is open* $\Leftrightarrow$ $\tau_1 \subseteq \tau_2$.

(ii) $\mathrm{id}_X : (X,\tau_1) \to (X,\tau_2)$ *is continuous* $\Leftrightarrow$ $\tau_2 \subseteq \tau_1$.

(iii) $\mathrm{id}_X : (X,\tau_1) \to (X,\tau_2)$ *is a homeomorphism* $\Leftrightarrow$ $\tau_1 = \tau_2$.

Generalizing from identity maps to bijections, we have:

Lemma 5.2.26 *Let $(X,\tau),(Y,\sigma)$ be topological spaces.*

(i) If $f : X \to Y$ is a bijection with inverse $g : Y \to X$ then g is continuous $\Leftrightarrow$ f is open $\Leftrightarrow$ f is closed.

(ii) A function $f : X \to Y$ is a homeomorphism $\Leftrightarrow$ f is bijective, continuous and open $\Leftrightarrow$ f is bijective, continuous and closed.

Proof. (i) If $f : X \to Y$ and $g : Y \to X$ are mutually inverse functions then for every $Z \subseteq X$ we have $g^{-1}(Z) = f(Z)$, which gives that g is continuous if and only if f is open. Since f is a bijection, it satisfies $f(X \backslash Z) = Y \backslash f(Z)$ for every $Z \subseteq X$. From this it is immediate that openness and closedness of f are equivalent (for bijections!).

(ii) is an obvious consequence of (i). ■

We will have many occasions to see that openness and closedness are useful properties even for functions that are not-bijective. Here is a first example:

Exercise 5.2.27 If (X,τ) is first (respectively second) countable and $f : X \to Y$ is continuous, open and surjective, prove that (Y,σ) is first (respectively second) countable.

Remark 5.2.28 1. The analogous statement for separability (Exercise 5.2.11(ii)) was true without the openness assumption, but for second countability of $f(X)$ this is not the case! Cf. Proposition 7.4.17 for a result on second countability of images under closed maps.

2. For every topological space Y there is a continuous open surjection $f : X \to Y$ with X Hausdorff, cf. [237, 13H].

3. Every metric space is first countable. Thus if M is metric and $f : M \to X$ is continuous, open and surjective then X is first countable by Exercise 5.2.27. Actually every first countable space arises in this way! For T_1-spaces the proof is not difficult, cf. e.g. [225, p. 179-180]. Now the result under 2. can be used to remove the T_1 assumption.) □

Exercise 5.2.29 Given a function $f : X \to Y$, prove that the following are equivalent:

(i) f is closed.

(ii) $\overline{f(A)} \subseteq f(\overline{A})$ holds for every $A \subseteq X$.

(iii) For every $y \in Y$ and every open $U \subseteq X$ such that $f^{-1}(y) \subseteq U$ there is an open $V \subseteq Y$ such that $y \in V$ and $f^{-1}(V) \subseteq U$.

Corollary 5.2.30 *The identity $\overline{f(A)} = f(\overline{A})$ holds for all $A \subseteq X$ if and only if f is continuous and closed.*

5.2.4 ⋆ Semicontinuous $\mathbb{R}$-valued functions

Recall the usual topology σ of $\mathbb{R}$ has a subbase $\mathcal{S} = \{(a,\infty) \mid a \in \mathbb{R}\} \cup \{(-\infty, a) \mid a \in \mathbb{R}\}$. By Exercise 5.2.8(iii), this implies that $f : (X,\tau) \to (\mathbb{R},\sigma)$ is continuous if and only if $f^{-1}((a,\infty))$ and $f^{-1}((-\infty,a))$ are in τ for every $a \in \mathbb{R}$. This motivates the following definition:

Definition 5.2.31 *Let (X,τ) be a topological space. A function $f : X \to \mathbb{R}$ is called*

- *lower semicontinuous (resp. upper semicontinuous) at $x \in X$ if for every $\varepsilon > 0$ there is an open neighborhood $U \ni x$ such that $f(U) \subseteq (f(x) - \varepsilon, \infty)$ (resp. $f(U) \subseteq (-\infty, f(x) + \varepsilon)$).*
- *lower semicontinuous (resp. upper semicontinuous) if $f^{-1}((a,\infty))$ (resp. $f^{-1}((-\infty,a))$) is open for every $a \in \mathbb{R}$.*

Exercise 5.2.32 Let $f : (X,\tau) \to \mathbb{R}$ be a function. Prove:

(i) f is lower (resp. upper) semicontinuous at $x \in X \Leftrightarrow$ for every $\varepsilon > 0$ and every net $\{x_\iota\}_{\iota \in I}$ such that $x_\iota \to x$, there is a $\iota_0 \in I$ such that $\iota \geq \iota_0 \Rightarrow f(x_\iota) > f(x) - \varepsilon$ (resp. $f(x_\iota) < f(x) + \varepsilon$).

(ii) f is upper (lower) semicontinuous if and only if f is upper (lower) semicontinuous at every $x \in X$.

(iii) f is continuous (at x) if and only if it is upper and lower semicontinuous (at x).

(iv) f is upper semicontinuous if and only if $-f$ is lower semicontinuous.

(v) If $\mathcal{F}$ is a family of lower semicontinuous functions then $g(x) = \sup\{f(x) \mid f \in \mathcal{F}\}$ is lower semicontinuous.

(vi) A finite sum of lower semicontinuous functions is lower semicontinuous.

(vii) If $\mathcal{F}$ is a family of non-negative lower semicontinuous functions such that for all $x \in X$, $g(x) = \sum_{f\in\mathcal{F}} f(x) < \infty$ (in the sense of unordered summation, cf. Example 5.1.22) then g is lower semicontinuous.

(viii) If $\mathcal{F}$ is as in (vii) with g continuous, then $x \mapsto \sum_{f\in\mathcal{F}'} f(x)$ is continuous for every $\mathcal{F}' \subseteq \mathcal{F}$. In particular, every $f \in \mathcal{F}$ is continuous.

Another way of testing semicontinuity is using the lower and upper limits (lim inf and lim sup) which can be defined as for sequences. For a net $\{y_\iota\}$ in $\mathbb{R}$ indexed by the directed set $(I, \leq)$ and $\iota_0 \in I$, define

$$U_{\iota_0} = \sup\{y_\iota \mid \iota \geq \iota_0\}, \qquad L_{\iota_0} = \inf\{y_\iota \mid \iota \geq \iota_0\},$$

both taking values in the extended reals $\widetilde{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$. It is clear that $\iota_1 \geq \iota_2$ implies $U_{\iota_1} \leq U_{\iota_2}$ and $L_{\iota_1} \geq L_{\iota_2}$, thus the limits

$$\limsup y_\iota := \lim_\iota U_\iota, \qquad \liminf y_\iota := \lim_\iota L_\iota$$

always exist in $\widetilde{\mathbb{R}}$.

Exercise 5.2.33 Consider a function $f : X \to \mathbb{R}$ and $x \in X$. Prove that f is ...

(i) lower semicontinuous at x if and only if $f(x) \leq \liminf f(x_\iota)$ for every net $\{x_\iota\}$ with $x_\iota \to x$,

(ii) upper semicontinuous at x if and only if $f(x) \geq \limsup f(x_\iota)$ for every net $\{x_\iota\}$ with $x_\iota \to x$.

For results involving semicontinuous functions see Theorem 8.5.33 and Theorem D.2.2.

Chapter 6

New spaces from old

6.1 Initial and final topologies

It often happens that a family of maps f_i from a set to certain topological spaces is given (or the other way round) and we want to find the "best" topology on the set making all f_i continuous. What "best" means depends on whether we consider maps $f_i : X \to (Y_i, \sigma_i)$ or $f_i : (X_i, \tau_i) \to Y$.

6.1.1 The final topology

We begin with the (slightly nicer) case, where maps $f_i : (X_i, \tau_i) \to Y$ are given. This is relevant for direct sums and quotients of topological spaces.

Definition 6.1.1 *Let (X_i, τ_i) be a topological space for each $i \in I$, and let Y be a set. Let functions $f_i : X_i \to Y,\ i \in I$ be given. The final topology on Y induced by the maps f_i is the finest topology σ_{fin} such that all maps $f_i : (X_i, \tau_i) \to (Y, \sigma_{\text{fin}}),\ i \in I$ are continuous.*

Lemma 6.1.2 *The final topology σ_{fin} always exists, is unique and is given by*

$$\sigma_{\text{fin}} = \{U \subseteq Y \mid f_i^{-1}(U) \in \tau_i \quad \forall i \in I\}. \tag{6.1}$$

Proof. For the purpose of this proof, denote the r.h.s. of (6.1) by σ'. We claim that σ' is a topology. It should be obvious that $\emptyset, Y \in \sigma'$. And if each $U_k \subseteq Y$ satisfies $f_i^{-1}(U_k) \in \tau_i$ for all $i \in I$ then by basic set theory (cf. Appendix A.1) we have $f_i^{-1}(\bigcup_k U_k) = \bigcup_k f_i^{-1}(U_k) \in \tau_i$ and (for finitely many k) $f_i^{-1}(\bigcap_k U_k) = \bigcap_k f_i^{-1}(U_k) \in \tau_i$ for all i. Thus σ' is a topology.

Now, the very definition of continuity implies that every topology σ on Y for which all maps $f_i : (X_i, \tau_i) \to (Y, \sigma)$ are continuous is contained in σ'. Thus σ' is the finest (=largest) topology on Y making all $f_i : X_i \to Y$ continuous, thus $\sigma_{\text{fin}} = \sigma'$. ■

Exercise 6.1.3 Prove that a subset $Z \subseteq Y$ is σ_{fin}-closed if and only if $f_i^{-1}(Z) \subseteq X_i$ is closed for all $i \in I$.

Proposition 6.1.4 *Let $(X_i, \tau_i), Y, f_i$ be as in Definition 6.1.1, and let σ_{fin} be the corresponding final topology on Y. Then a map $h : (Y, \sigma_{\text{fin}}) \to (Z, \gamma)$ is continuous if and only if the composition $h \circ f_i : (X, \tau_i) \to (Z, \gamma)$ is continuous for every $i \in I$.*

Proof. By definition of the final topology σ_{fin}, all $f_i : (X_i, \tau_i) \to (Y, \sigma_{\text{fin}})$ are continuous. Thus if $h : (Y, \sigma_{\text{fin}}) \to (Z, \gamma)$ is continuous, so are $h \circ f_i$.

Now assume that $h \circ f_i : (X_i, \tau_i) \to (Z, \gamma)$ is continuous for every $i \in I$. I.e., for every $U \in \gamma$ we have $f_i^{-1}(h^{-1}(U)) = (h \circ f_i)^{-1}(U) \in \tau_i$. Thus if we write $V = h^{-1}(U) \subseteq Y$, we have $f_i^{-1}(V) \in \tau_i \ \forall i$. But in view of Lemma 6.1.2, this is equivalent to $V \in \sigma_{\text{fin}}$. Thus h is continuous. ■

6.1.2 The initial topology

We now turn to the case, where maps $f_i : X \to (Y_i, \sigma_i)$ are given. This applies to subspaces and products of topological spaces.

Definition 6.1.5 *Let X be a set and let (Y_i, σ_i) a topological space for each $i \in I$. Let functions $f_i : X \to Y_i,\ i \in I$ be given. The initial topology induced by the maps f_i is the coarsest topology τ_{ini} on X such that all maps $f_i : (X, \tau_{\text{ini}}) \to (Y_i, \sigma_i),\ i \in I$ are continuous.*

Lemma 6.1.6 *(i) The initial topology always exists and is unique.*

(ii) A subbase for the initial topology τ_{ini} is given by

$$\mathcal{S} = \{f_i^{-1}(U) \mid i \in I,\ U \in \sigma_i\}. \tag{6.2}$$

Proof. Define $\mathcal{S} \subseteq P(X)$ as in (6.2). If τ is a topology on X then continuity of all maps $f_i : (X, \tau) \to (Y_i, \sigma_i)$ is equivalent to $\mathcal{S} \subseteq \tau$. We know from Lemma 4.2.2 that there is a unique weakest topology τ on X containing $\mathcal{S}$, obtained either as the intersection of all topologies that contain $\mathcal{S}$ or as the family of all sets that can be written as arbitrary unions of finite intersections of elements of $\mathcal{S}$. (These two descriptions are different only at first sight.) This topology clearly is τ_{ini}, and it has $\mathcal{S}$ as subbase by construction. ■

Remark 6.1.7 The conclusion of Lemma 6.1.6 is not quite as nice as that of Lemma 6.1.2: While we could write down τ_{fin} explicitly, τ_{ini} could only be defined by writing down a subbase. This is due to the fact that inverse images of functions have better algebraic properties than images, cf. Lemma A.1.7. This is also the reason why there is no nice analogue of Exercise 6.1.3 for the initial topology. □

The following is entirely analogous to Proposition 6.1.4:

Proposition 6.1.8 *Let $X, (Y_i, \sigma_i), f_i$ be as in Definition 6.1.5, and let τ_{ini} be the corresponding initial topology on X. Then a map $h : (Z, \gamma) \to (X, \tau_{\text{ini}})$ is continuous if and only if the composition $f_i \circ h : (Z, \gamma) \to (Y, \sigma_i)$ is continuous for every $i \in I$.*

Proof. By definition of the initial topology τ_{ini}, all $f_i : (X, \tau_{\text{ini}}) \to (Y_i, \sigma_i)$ are continuous. Thus if $h : (Z, \gamma) \to (X, \tau_{\text{ini}})$ is continuous, so are the composites $f_i \circ h$.

Now assume that $f_i \circ h : (Z, \gamma) \to (Y, \sigma_i)$ is continuous for every $i \in I$. Thus for every $i \in I$ and $U \in \sigma_i$, we have $(f_i \circ h)^{-1}(U) \in \gamma$. But this is the same as $h^{-1}(f_i^{-1}(U)) \in \gamma$. By Lemma 6.1.6(ii), $\{f_i^{-1}(U) \mid i \in I,\ U \in \sigma_i\}$ is a subbase $\mathcal{S}$ for τ_{ini}. Thus $h^{-1}(V) \in \gamma$ for every $V \in \mathcal{S}$, and continuity of h follows from Exercise 5.2.8(iii). ■

The following has no analogue for the final topology:

Proposition 6.1.9 *Let $X, (Y_i, \sigma_i), f_i$ be as in Definition 6.1.5, and let τ_{ini} be the corresponding initial topology on X. Then a net $\{x_\iota\}$ in X converges to $z \in (X, \tau_{\text{ini}})$ if and only if $f_i(x_\iota)$ converges to $f_i(z)$ in (Y_i, σ_i) for each $i \in I$.*

Proof. If x_ι converges in (X, τ_{ini}) then by Proposition 5.2.5, $f_i(x_\iota)$ converges in (X_i, σ_i) for each $i \in I$, since f_i is continuous by definition of τ_{ini}. Now assume that $f_i(x_\iota)$ converges to $z_i \in X_i$ for each i, let $z \in X$ be such that $f_i(z) = z_i \ \forall i$, and let $z \in U \in \tau_{\text{ini}}$. The $\mathcal{S}$ of Lemma 6.1.6 is a subbase for τ_{ini}, thus there are $i_1, \ldots, i_n \in I$ and $U_k \in \sigma_{i_k}$ such that

$$z \in V = f_{i_1}^{-1}(U_1) \cap \cdots \cap f_{i_k}^{-1}(U_k) \subseteq U.$$

Since all nets $f_i(x_\iota)$ converge, we can find $\iota_k \in I$ such that $\iota \geq \iota_k \Rightarrow f_{i_k}(x_\iota) \in U_k$. Since $(I, \leq)$ is directed, we can find $\iota_0 \in I$ such that $\iota_0 \geq \iota_k \ \forall k = 1, \ldots, n$. Now, if $\iota \geq \iota_0$ we have $f_{i_k}(x_\iota) \in U_k \ \forall k = 1, \ldots, n$, and therefore $x_\iota \in V$. This proves $x_\iota \to z$. ■

Remark 6.1.10 1. Notice that we do NOT assert that $f_i(x_\iota) \to z_i \in X_i \ \forall i \in I$ implies the existence of $z \in X$ such that $f_i(z) = z_i$ and $x_\iota \to z$. The existence of $z \in X$ must be given!

2. If $Z = \{z \in X \mid f_i(z) = z_i \ \forall i \in I\}$ and $f_i(x_\iota) \to z_i \ \forall i$, then the above shows that x_ι converges to every $z \in Z$! Thus by Proposition 5.1.20, the initial topology τ_{ini} will fail to be Hausdorff if the map $f : X \to \prod_k Y_k$ defined by $f(x)_i = f_i(x)$ is not injective. (If f is injective, we say that the family $f_i : X \to Y_i$ separates the points of X.) □

6.2 Subspaces

We have encountered subspaces very early (Exercise 2.2.9). Now we study those aspects of subspaces that involve continuous functions.

Lemma 6.2.1 *Let (X, τ) be a topological space and $Y \subseteq X$. Then the subspace topology τ_Y, cf. (2.7), coincides with the initial topology τ_{ini} on Y induced by the inclusion map $\iota : Y \hookrightarrow X$.*

Proof. By Lemma 6.1.6(ii), $\mathcal{S} = \{\iota^{-1}(U) \mid U \in \tau\}$ is a subbase for the initial topology τ_{ini}. But $\iota^{-1}(U) = U \cap Y$, from which it is clear that $\mathcal{S}$ is already closed under unions and finite intersections, so that $\tau_{\text{ini}} = \mathcal{S} = \tau_Y$. ■

Corollary 6.2.2 *Let $Y \subseteq (X, \tau)$ and let τ_Y be the subspace topology. Then $f : (Z, \eta) \to (Y, \tau_Y)$ is continuous if and only if f is continuous as a map $(Z, \eta) \to (X, \tau)$. (Strictly speaking, we should write $\iota \circ f : Z \to X$, where $\iota : Y \hookrightarrow X$ is the inclusion map.)*

Proof. In view of Lemma 6.2.1, this is immediate by Proposition 6.1.8.

Also the direct proof is very simple. Assume $f : Z \to Y$ is continuous as a map to X. Let $V \in \tau_Y$. Then there is a $U \in \tau$ such that $V = Y \cap U$. Now $f^{-1}(V) = f^{-1}(U)$, which is open by continuity of $f : Z \to X$. The converse is clear since ι is continuous. ■

Remark 6.2.3 1. If this corollary was not true, then clearly something would be wrong with our definition of subspaces.

2. If this was the only application of initial topologies, it would hardly justify introducing the notion. The initial topology will come into its own in the discussion of the product topology, cf. Section 6.5, where it really provides the right perspective. (The corollary can, of course, be proven without the formalism of initial topologies, but that proof would just be a restatement of that of Proposition 6.1.8 in the special case at hand.)

3. In the discussion of subspaces, we consider maps $f : Y \to (X, \sigma)$ that are injective. Without this assumption, the initial topology τ_{ini} on Y induced by f can be quite badly behaved. (Cf. also

Remark 6.1.10.) E.g. in the extreme case where $f = \text{const} = x \in X$, we find that $f^{-1}(U)$ equals either Y if $x \in U \subseteq X$ or $\emptyset$ if $x \notin U$. Thus τ_{ini} is the indiscrete topology. □

Exercise 6.2.4 Let $Y \subseteq (X, \tau)$ be a subspace. Then the inclusion map $\iota : Y \hookrightarrow X$ is open (resp. closed) if and only if $Y \subseteq X$ is open (resp. closed).

Exercise 6.2.5 Let $(X, \tau), (Y, \sigma)$ be topological spaces and $f : X \to Y$ a function. Prove:

(i) If f is continuous and $A \subseteq X$ then $f \restriction A : (A, \tau_A) \to (Y, \sigma)$ is continuous.

(ii) Continuity of $f \restriction A$ for each $A \in \mathcal{A}$ together with $\bigcup \mathcal{A} = X$ does not imply continuity of f.

(iii) If $\mathcal{U}$ is a family of open subsets of X such that $f \restriction U$ is continuous for every $U \in \mathcal{U}$ and $\bigcup_{U \in \mathcal{U}} = X$ then f is continuous.

(iv) A statement analogous to (ii) holds for finite families of closed subsets. Explain why this does not generalize to infinite families.

The following immediate consequence is used very often:

Corollary 6.2.6 (Gluing of functions) *Let X, Y be topological spaces. Let $\mathcal{A}$ be a family of subsets of X such that $\bigcup_{A \in \mathcal{A}} A = X$. Assume that all elements of $\mathcal{A}$ are open, or $\mathcal{A}$ is finite and all elements are closed. Let $\{f_A : A \to Y\}_{A \in \mathcal{A}}$ be continuous functions such that $f_A \restriction A \cap A' = f_{A'} \restriction A \cap A'$ whenever $A, A' \in \mathcal{A}$ and $A \cap A' \neq \emptyset$. Then the function $f : X \to Y$ defined by $f(x) = f_A(x)$, where we choose any $A \in \mathcal{A}$ with $x \in A$, is continuous.*

The following notion involving subspaces will have many uses:

Definition 6.2.7 *A map $f : (X, \tau) \to (Y, \sigma)$ is called an embedding if $f : X \to f(X)$ is a homeomorphism w.r.t. the subspace (=induced) topology on $f(X) \subseteq Y$.*

Example 6.2.8 Let (X, τ) be a topological space and $Y \subseteq X$. Then the inclusion map $(Y, \tau_Y) \to (X, \tau)$ is an embedding, quite trivially. □

Given $f : X \to Y$, it is clear that $f : X \to f(X)$ is automatically surjective, and injective if and only if $f : X \to Y$ is injective. Giving $f(X) \subseteq Y$ the subspace topology, continuity of $f : (X, \tau) \to (f(X), \sigma \restriction f(X))$ is equivalent to continuity of $f : (X, \tau) \to (Y, \sigma)$ by Corollary 6.2.2. Thus in order for $f : X \to Y$ to be an embedding, it must be continuous and injective.

Lemma 6.2.9 *Let $f : (X, \tau) \to (Y, \sigma)$ be continuous and injective.*

(i) The following are equivalent:

(α) f is an embedding

(β) f is open as a map $X \to f(X)$.

(γ) f is closed as a map $X \to f(X)$.

(δ) $f(x) \notin \overline{f(C)}$ whenever $C \subseteq X$ is closed and $x \in X \backslash C$.

(ii) If $f : X \to Y$ is open or closed then $f : X \to f(X)$ is open (closed), thus an embedding.

Proof. (i) By Lemma 5.2.26, f is an embedding if and only if $f : X \to f(X)$ is injective, surjective, continuous and open (or, equivalently, closed). Since our assumptions imply that $f : X \to f(X)$ is a continuous bijection, we thus have $(\alpha) \Leftrightarrow (\beta) \Leftrightarrow (\gamma)$.

To conclude, we prove $(\gamma) \Leftrightarrow (\delta)$. Let $C \subseteq X$ be closed. By Exercise 2.6.13(iii), closedness of $f(C)$ is equivalent to $f(C) = \mathrm{Cl}_{f(X)}(f(C)) = \overline{f(C)} \cap f(X)$. The inclusion $f(C) \subseteq \overline{f(C)} \cap f(X)$ is trivially true. The converse inclusion $\overline{f(C)} \cap f(X) \subseteq f(C)$ is just the statement that if $y = f(x) \in \overline{f(C)}$ for some $x \in X$, then $y \in f(C)$. Since f is injective, this implies $x \in C$. Thus f is an embedding if and only if ($C \subseteq X$ closed, $x \in X$, $f(x) \in \overline{f(C)} \Rightarrow x \in C$). This is just the contraposition of (δ).

(ii) Assume that $f : X \to Y$ is open (resp. closed). If now $Z \subseteq X$ is open (resp. closed) then $f(Z)$ is open (resp. closed) in Y. But then $f(Z) = f(Z) \cap f(X)$ is open (resp. closed) in $f(X)$ with its subspace topology. Now apply (i). ■

Later we will use (local) compactness to prove that certain continuous injections automatically are embeddings, cf. Propositions 7.4.11(iii) and 7.8.63.

6.3 Direct sums

The direct sum (or coproduct) operation on topological spaces is not terribly interesting in itself (and therefore omitted by some authors). But it plays an important rôle in algebraic topology, where it is combined with the quotient operation to 'attach' a space to another, cf. Definition 6.6.7. On the other hand it helps to better understand the notion of connectedness, cf. Proposition 6.3.7.

The direct sum of topological spaces is defined in terms of the disjoint union of sets, which is defined and studied in Section A.2.

Definition 6.3.1 *Let (X_i, τ_i) be a topological space for each $i \in I$. Let $\bigoplus_k X_k$ be the disjoint union. The direct sum topology on $\bigoplus_k X_k$ is the final topology τ induced by the inclusion maps $\iota_i : X_i \to \bigoplus_k X_k,\ x \mapsto (i, x)$.*

The topological space $(\bigoplus_k X_k, \tau)$ is denoted by $\bigoplus_k (X_k, \tau_k)$, the direct sum of the (X_i, τ_i).

Remark 6.3.2 1. Some authors call the direct sum the coproduct and write $\coprod_k (X_k, \tau_k)$ instead. From a categorical perspective, this perfectly justified since the coproduct (direct sum) behaves dual to the (direct) product. But we find the symbols $\oplus, \bigoplus$ more immediately recognizable and less prone to confusion with $\prod$.

2. If I is finite we will usually take $I = \{1, \ldots, n\}$ and also denote $\bigoplus_i X_i$ by $X_1 \oplus \cdots \oplus X_n$. But notice that this raises some issues: The total ordering of written text forces us to put a total order on the index set I, which is absent from Definition A.2.6. This however is spurious, and we should either read something like $(x, y) \in X \times Y$ as a function from a two-element set to $X \cup Y$ or 'identify' $(x, y) \in X \times Y$ with $(y, x) \in Y \times X$. One way to do this is to consider topological spaces with direct sum operation as a 'symmetric monoidal category'. To further complicate the matter we note that, according to many authors, (co)products should not be defined in terms of an explicit construction but rather in terms of their universal property, but doing so makes them uniquely defined only up to isomorphism... □

Applying Lemma 6.1.2 to the present situation, we immediately have:

Lemma 6.3.3 *The direct sum topology on $\bigoplus_k X_k$ is given by*

$$\tau = \left\{ U \subseteq \bigoplus_k X_k \mid \iota_i^{-1}(U) \in \tau_i \quad \forall i \in I \right\}.$$

Remark 6.3.4 If $X_i \cap X_j = \emptyset$ whenever $i \neq j$, then the map $\bigoplus_k X_k \ni (x,i) \mapsto x \in \bigcup_k X_k$ clearly is a bijection. In that case, we can identify each X_i with the corresponding subset of $\bigcup_k X_k$ and write $U \cap X_i$ instead of $\iota_i^{-1}(U)$. □

The universal property of the disjoint union, cf. Proposition A.2.5 has a topological version:

Proposition 6.3.5 *Let (X_i, τ_i), $i \in I$ and (Y, σ) be topological spaces. Then there is a bijection between continuous maps $f : \bigoplus_k (X_k, \tau_k) \to (Y, \sigma)$ and families of continuous maps $\{f_i : (X_i, \tau_i) \to (Y, \sigma)\}_{i \in I}$.*

Proof. By Proposition A.2.5, there is a bijection between maps $f : \bigoplus_k (X_k, \tau_k) \to (Y, \sigma)$ and families of maps $\{f_i : (X_i, \tau_i) \to (Y, \sigma)\}_{i \in I}$. Now Proposition 6.1.4 implies that f is continuous if and only if all f_i are continuous. ■

The inclusion maps $\iota_i : X_i \to \bigoplus_k X_k$ are continuous by definition of the direct sum topology as final topology. But we have more:

Lemma 6.3.6 *Let (X_i, τ_i) be a topological space for each $i \in I$. Then*

(i) $Y \subseteq \bigoplus_i X_i$ is closed if and only if $\iota_i^{-1}(Y) \subseteq X_i$ is closed for every $i \in I$.

(ii) The maps ι_i are open and closed.

(iii) Each $\iota_i(X_i)$ is a clopen subset of $\bigoplus_k X_k$.

(iv) Each $\iota_i : (X_i, \tau_i) \to \bigoplus_k (X_k, \tau_k)$ is an embedding.

Proof. (i) This is a special case of Exercise 6.1.3.

(ii) Let $U \in \tau_i$. By Lemma 6.3.3, we must check that $\iota_j^{-1}(\iota_i(U)) \in \tau_j$ for all j. But this is the empty set if $j \neq i$ and U otherwise. This gives openness, and the proof of closedness is analogous, using the result of (i).

(iii) Each X_i is open and closed as subset of itself. Now the claim follows from (ii).

(iv) The maps ι_i are injective, continuous and open, thus the claim follows from Lemma 6.2.9(ii). (Alternatively, just observe that restricting the direct sum topology to $\iota_i(X_i) \subseteq \bigoplus_k X_k$ gives the topology τ_i back.) ■

Proposition 6.3.7 *A topological space (X, τ) is connected if and only if it is not homeomorphic to a direct sum $(X_1, \tau_1) \oplus (X_2, \tau_2)$ with $X_1 \neq \emptyset \neq X_2$.*

Proof. If $X_1 \neq \emptyset \neq X_2$ then $X = (X_1, \tau_1) \oplus (X_2, \tau_2)$ is not connected, since $\iota_1(X_1) \subseteq X$ is clopen and neither $\emptyset$ not X. Thus also a space that is homeomorphic to a non-trivial direct sum is not connected.

Now assume that (X, τ) is non-connected, i.e. there is a clopen $Y \subseteq X$ with $\emptyset \neq Y \neq X$. Since Y is clopen, we have $Y, X \backslash Y \in \tau$. Thus $\tau_Y = \{U \cap Y \mid U \in \tau\} \subseteq \tau$, and also $\tau_{X \backslash Y} \subseteq \tau$. We claim that (X, τ) is homeomorphic to the direct sum of the subspaces (Y, τ_Y) and $(X \backslash Y, \tau_{X \backslash Y})$. It is clear that (as sets) we can identify X with the disjoint union $Y \oplus (X \backslash Y)$. Proving that this is a homeomorphism amounts to showing for every $U \subseteq X$ that $U \in \tau$ if and only if $U \cap Y \in \tau_Y$ and $U \cap (X \backslash Y) \in \tau_{X \backslash Y}$. The direction $\Rightarrow$ is obvious by definition of τ_Y and $\tau_{X \backslash Y}$. For the converse, assume $U \cap Y \in \tau_Y$ and $U \cap (X \backslash Y) \in \tau_{X \backslash Y}$. In view of $\tau_Y \subseteq \tau$, $\tau_{X \backslash Y} \subseteq \tau$, this implies $U \cap Y \in \tau$ and $U \cap (X \backslash Y) \in \tau$, thus also $U = (U \cap Y) \cup (U \cap (X \backslash Y)) \in \tau$. ■

Remark 6.3.8 Since non-connected spaces decompose into non-trivial direct sums, one might think that by iterating this decomposition one will arrive at $(X, \tau) \cong \bigoplus_i (X_i, \tau_i)$ with all (X_i, τ_i) connected. In Section 9.1, we will see that this is false in general, even for compact spaces! (This is related to the fact that there are spaces that are non-discrete, but in which singletons are the only connected subspaces.) $\square$

Exercise 6.3.9 Let $\{X_i\}_{i \in I}$ be topological spaces. Prove that the direct sum $X = \bigoplus_i X_i$ is

(i) Hausdorff if and only if each X_i is Hausdorff.

(ii) metrizable if and only each X_i is metrizable.

(iii) first countable if and only if each X_i is first countable.

(iv) second countable if and only if each X_i is second countable and $X_i \neq \emptyset$ for at most countably many $i \in I$.

6.4 Quotient spaces

6.4.1 Quotient topologies. Quotient maps

In this section, we study the following situation: We are given a topological space (X, τ) and a surjective map $f : X \to Y$, and want to put a natural topology on Y.

Definition 6.4.1 *Let (X, τ) be a topological space, Y a set and $p : X \to Y$ a surjective map. Then the quotient topology on Y is the final topology induced by the map p.*

On the other hand, a continuous surjection $f : (X, \tau) \to (Y, \sigma)$ is called a (topological) quotient map if σ coincides with the quotient topology induced by τ and f.

Thus by Lemma 6.1.2, the quotient topology is given by

$$\sigma = \{V \subseteq Y \mid f^{-1}(V) \in \tau\}. \tag{6.3}$$

In analogy to Exercise 2.2.9(iii), we have a result on iterated quotient constructions:

Exercise 6.4.2 Let $(X, \tau) \xrightarrow{f} Y \xrightarrow{g} Z$ be surjective maps. Let σ be the quotient topology on Y. Then the quotient topologies on Z arising from the quotients $g : (Y, \sigma) \to Z$ and $g \circ f : (X, \tau) \to Z$ coincide.

What can be said about the final topology when $f : X \to Y$ is not surjective?

Exercise 6.4.3 Let $f : (X, \tau) \to Y$ be arbitrary and σ the final topology on Y induced by f. Prove that $(Y, \sigma) \cong (f(X), \sigma') \oplus (Y \backslash f(X), \tau_{\text{disc}})$, where σ' is the final topology induced by the (surjective) map $f : X \to f(X)$.

Remark 6.4.4 Just as not every continuous injective map is an embedding, not every continuous surjective map $f : (X, \tau) \to (Y, \sigma)$ is a quotient map! [If $\widetilde{\sigma} = \{U \subseteq Y \mid f^{-1}(U) \in \tau\}$ is the final topology on Y induced by f, the continuity of f w.r.t. σ means that $\sigma \subseteq \widetilde{\sigma}$. This in turn means that f factorizes as $(X, \tau) \to (Y, \widetilde{\sigma}) \to (Y, \sigma)$, where the first map *is* a quotient map and the second is the identity map of Y, equipped with two a priori different topologies.] It therefore is useful to have criteria implying that a continuous surjective map is a quotient map. $\square$

Lemma 6.4.5 *Let $f : (X,\tau) \to (Y,\sigma)$ be continuous and surjective.*

(i) f is a quotient map (i.e. σ is the final topology on Y induced by f) if and only if $V \subseteq Y$ and $f^{-1}(V) \in \tau$ imply $V \in \sigma$.

(ii) If f is open or closed then it is a quotient map. (This is sufficient, but not necessary!)

Proof. (i) Continuity gives $V \in \sigma \Rightarrow f^{-1}(V) \in \tau$. If the converse implication also holds, then (6.3) is true, thus σ is the quotient topology.

(ii) Assume f is open. Let $V \subseteq Y$. If $f^{-1}(V) \in \tau$ then $f(f^{-1}(V)) \in \sigma$ by openness of f. Since f is surjective, we have $f(f^{-1}(V)) = V$, cf. Lemma A.1.6, thus V is open. Now assume that f is closed, $D \subseteq Y$ and $f^{-1}(D) \subseteq X$ is closed. By the same reasoning as above, but using closedness of f, we obtain that D is closed. If now $f^{-1}(V) \in \tau$ then $f^{-1}(Y\backslash V) = X\backslash f^{-1}(V)$ is closed, thus $Y\backslash V$ is closed, so that V is open. Thus in both cases we have $f^{-1}(V) \in \tau \Rightarrow V \in \sigma$, so that σ is the quotient topology. ■

Remark 6.4.6 Later, we will use the above lemma to prove that a continuous surjection $f : X \to Y$ automatically is a quotient map when X is compact and Y Hausdorff, cf. Proposition 7.4.11(iii).

2. In Remark 5.2.28 we have noted that the first countable spaces are precisely the images of metric spaces under continuous surjective open maps. By Lemma 6.4.5, every such map is a quotient map, but not conversely. Thus the class of quotient spaces of metric spaces is potentially larger than that of first countable spaces, and indeed one finds that it precisely is the class of sequential spaces (cf. Remark 5.1.9)! Between first countable spaces and sequential spaces one has the Fréchet-spaces. It turns out that the latter precisely are the images of metric spaces under maps $f : M \to X$ that are 'hereditarily quotient', i.e. $f^{-1}(Z) \to Z$ is a quotient map for every $Z \subseteq X$. This is equivalent to f being continuous, surjective and 'pseudo-open' in the sense that for any $x \in X$ and any open $U \subseteq M$ such that $f^{-1}(x) \subseteq U$ we have $x \in f(U)^0$.

These results (for proofs cf. e.g. [225, p. 180-183]) show that first countable, Fréchet and sequential spaces, while being more general than metric spaces, still are quite close to metric spaces. This shows that in topology one cannot hope to get very far using only sequences. □

6.4.2 Quotients by equivalence relations

Surjective maps $f : X \to Y$ most often arise as quotient maps $X \to X/\sim$, where $\sim$ is an equivalence relation on X. (We assume as known the basics of equivalence relations and related constructions, cf. Section A.1.3.)

We therefore now focus on this situation. Thus whenever (X,τ) is a topological space and $\sim$ is an equivalence relation on X, the quotient $X/\sim$ is understood to be equipped with the quotient topology coming from the quotient map $p : X/ \to X/\sim$. We need a further definition:

Definition 6.4.7 *Let $\sim$ be an equivalence relation on a set X. A subset $Y \subseteq X$ is ($\sim$-)<u>saturated</u> if $x \sim y \in Y$ implies $x \in Y$. Equivalently, Y is a union of $\sim$-equivalence classes. The <u>saturation</u> of $Y \subseteq X$ is given by*

$$Y^{\sim} = \{x \in X \mid \exists y \in Y : x \sim y\}.$$

The importance of this definition is due to the following obvious facts:

- If $Z \subseteq X/\sim$ then $p^{-1}(Z)$ is $\sim$-saturated.
- If $Y \subseteq X$ then $p^{-1}(p(Y)) = Y^{\sim}$.

Elevating Lemma A.1.11 to a topological statement, we obtain the universal property of the quotient space construction:

Proposition 6.4.8 *Let X, Y be topological spaces and $\sim$ an equivalence relation on X.*

(i) There is a bijection between continuous maps $g : X/\!\sim \to Y$ and continuous maps $f : X \to Y$ that are constant on equivalence classes such that $f = g \circ p$.

(ii) The function $g : X/\!\sim \to Y$ corresponding to $f : X \to Y$ is open if and only if $f(U) \subseteq Y$ is open for every $\sim$-saturated open $U \subseteq X$. In particular, this holds if f is open.

(iii) $g : X/\!\sim \to Y$ is a homeomorphism if and only if f is surjective, $f(x) = f(y) \;\Rightarrow\; x \sim y$, and $f(U) \subseteq Y$ is open for every $\sim$-saturated open $U \subseteq X$.

Proof. (i) The bijection $f \leftrightarrow g$ between functions (disregarding continuity) was shown in Lemma A.1.11. Thus every $f : X \to Y$ constant on equivalence classes is of the form $f = g \circ p$ for a unique map $g : X/\!\sim \to Y$. Since the quotient topology on $X/\!\sim$ is a final topology for the map $p : X \to X/\!\sim$, Proposition 6.1.4 immediately gives that $f : X \to Y$ is continuous if and only if $g : X/\sim \to Y$ is continuous.

(ii) By definition, the map g is open if $g(U) \subseteq Y$ is open for every open $U \subseteq X/\!\sim$. By definition of the quotient topology, $U \subseteq X/\!\sim$ is open if and only if $V = p^{-1}(U) \subseteq X$ (which is $\sim$-saturated) is open. Now the first claim follows from $g(U) = f(V)$. The last claim follows, since openness of $f : X \to Y$ means that $f(U) \subseteq Y$ is open for every open $U \subseteq X$, whether $\sim$-saturated or not.

(iii) By Lemma 5.2.26, g is a homeomorphism if and only if it is injective, surjective, continuous, and open. Continuity is automatic by (i). The equivalence of the three remaining conditions to those stated under (iii) follows from (ii) and from (ii),(iii) of Lemma A.1.11. ■

Remark 6.4.9 1. The existence in (i) of a continuous map $g : X/\!\sim \to Y$ such that $f = g \circ p$ is expressed as 'f factors through the quotient map $p : X \to X/\!\sim$'.

2. The above result is the main reason why we are interested in quotient spaces. It will be used quite often in the sequel, beginning in Section 6.4.3. □

Equipping $X/\!\sim$ with the quotient topology, the quotient map $p : X \to X/\!\sim$ is automatically continuous. But is it open? closed?

Lemma 6.4.10 *The map $p : X \to X/\!\sim$ is open (resp. closed) if and only if $U^\sim$ is open (resp. closed) for every open (resp. closed) $U \subseteq X$.*

Proof. By definition, p is open (closed) if and only if $p(U)$ is open (closed) for every open (closed) $U \subseteq X$. But by definition of the quotient topology, $p(U) \subseteq X/\!\sim$ is open (closed) if and only if $p^{-1}(p(U))$ is open (closed). As mentioned above, $p^{-1}(p(U)) = U^\sim$. ■

Definition 6.4.11 *An equivalence relation $\sim$ on X is called open (resp. closed) if $p : X \to X/\!\sim$ is open (resp. closed). (Cf. the equivalent conditions above.)*

As far as separation axioms are concerned, the quotient space construction can be very badly behaved. In view of the definition of the quotient topology, one obviously has:

Lemma 6.4.12 *Let X be a topological space and $\sim$ an equivalence relation on X. Then the quotient space $X/\!\sim$ is ...*

(i) *indiscrete if and only if the only $\sim$-saturated open subsets of X are $\emptyset$ and X;*

(ii) *discrete if and only if each equivalence class $[x] \subseteq X$ is open;*

(iii) *T_1 if and only if each equivalence class $[x] \subseteq X$ is closed.*

Proof. (i) A set $Z \subseteq X/\sim$ is open if and only if its preimage $p^{-1}(Z)$, which is $\sim$-saturated, is open. Thus $X/\sim$ is indiscrete if and only if the only $\sim$-saturated open sets in X are $\emptyset, X$.

(ii) $X/\sim$ is discrete if and only if each singleton in $X/\sim$ is open. This is equivalent to each $\sim$-equivalence class in X being open.

(iii) $X/\sim$ is T_1 if and only if each singleton in $X/\sim$ is closed. By Exercise 6.1.3, $\{x\} \in X/\sim$ is closed if and only if $p^{-1}(x) \subseteq X$ is closed. (Recall that the quotient topology is a final topology.) Since $p^{-1}(x)$ is a $\sim$-equivalence class, the above is equivalent to each $\sim$-equivalence class in X being closed. ■

Exercise 6.4.13 For $X = \mathbb{R}$ with the usual topology, define $\sim$ by $x \sim y \Leftrightarrow x - y \in \mathbb{Q}$. Show that $X/\sim$ is indiscrete.

Lemma 6.4.14 *Let (X, τ) be a topological space and $\sim$ an equivalence relation on X. Then the quotient topology on $X/\sim$ is Hausdorff if and only if given two different equivalence classes $[x] \neq [y]$ there are $U, V \in \tau$ such that $[x] \subseteq U, [y] \subseteq V$, $U \cap V = \emptyset$ <u>and</u> U, V are $\sim$-saturated.*

Proof. $X/\sim$ is Hausdorff if and only if for any $[x], [y] \in X/\sim$, $[x] \neq [y]$ there are disjoint open neighborhoods $U', V' \subseteq X/\sim$. If this is true then $U = p^{-1}(U'), V = p^{-1}(V')$ are disjoint saturated open sets in X containing $[x]$ and $[y]$, respectively. Clearly this is necessary and sufficient. ■

Definition 6.4.15 *For a topological space (X, τ) and $Y \subseteq X$, let $\sim_Y$ be the smallest equivalence relation on X that identifies all points of Y with each other. We then write $X/Y := X/\sim_Y$.*

On the positive side, we have the following:

Exercise 6.4.16 Prove:

(i) If $Y \subseteq X$ is open (resp. closed) the equivalence relation $\sim_Y$ from Definition 6.4.15 is open (resp. closed).

(ii) If (X, τ) is Hausdorff and $Y \subseteq X$ is finite, then X/Y is Hausdorff. (Using Lemma 6.4.14 saves work!)

For infinite $Y \subseteq X$, however, X/Y may fail to be Hausdorff! (But see Exercise 8.1.19.)

Example 6.4.17 Let $X = \mathbb{R} \times \{0, 1\} \subseteq \mathbb{R}^2$ (union of two parallel lines). Let $\sim$ be the equivalence relation that (besides containing the diagonal) identifies $(x, 0)$ and $(x, 1)$ whenever $x \neq 0$, but $(0, 0) \not\sim (0, 1)$. The equivalence classes are closed, thus $X/\sim$ is T_1. But if U, V are open neighborhoods of $(0, 0)$ and $(0, 1)$, respectively, there is $\varepsilon > 0$ such that $U \supseteq (-\varepsilon, +\varepsilon) \times \{0\}$ and $V \supseteq (-\varepsilon, +\varepsilon) \times \{1\}$. Now it is clear that $U^\sim \cap V^\sim \neq \emptyset$. Thus the requirement of the lemma cannot be satisfied, and $X/\sim$ is not Hausdorff. □

Exercise 6.4.18 Show that the equivalence relation in Example 6.4.17 is open, but not closed.

Example 6.4.19 In Example 6.4.17 we saw a nice space (even metrizable) and an open equivalence relation, where the quotient fails to be Hausdorff. We will see later that images under closed maps, in particular quotients by closed equivalence relations, tend to be better behaved. Nevertheless, even if X is Hausdorff and $f : X \to Y$ is continuous, closed and surjective, it does not follow that $Y = f(X)$ is Hausdorff! One can find Hausdorff spaces (even $T_{3.5}$-spaces) containing disjoint closed sets A, B that cannot be separated by disjoint open neighborhoods. (For an example, cf. Proposition 8.1.37.) Given such a space, define an equivalence relation $\sim$ on X by

$$x \sim y \;:\Leftrightarrow\; x = y \;\vee\; \{x, y\} \subseteq A \;\vee\; \{x, y\} \subseteq B.$$

I.e., $\sim$ identifies all points of A with each other and all points of B with each other (but not with those of A). The $\sim$-saturation of $C \subseteq X$ can be $C, C \cup A, C \cup B, C \cup A \cup B$, each of which is closed for closed C. Thus the equivalence relation $\sim$ and the quotient map p are closed. But $X/\!\sim$ cannot be Hausdorff: The images $a, b \in X/\!\sim$ of A and B, respectively, have no disjoint open neighborhoods U, V since otherwise $p^{-1}(U) \supseteq p^{-1}(a) = A$, $p^{-1}(V) \supseteq p^{-1}(b) = B$ would be disjoint open sets, contradicting our assumption on X. □

Remark 6.4.20 In Proposition 7.8.65 we will see that if X is Hausdorff and $f : X \to Y$ is surjective, continuous, closed and *proper* then Y is Hausdorff.

We will also encounter a separation axiom stronger than T_2 (normality=T_4) that is preserved under quotients by closed equivalence relations and therefore gives T_2 quotients. (The space in Example 6.4.17 actually is normal, but this does not help since $\sim$ is not closed.) □

In view of Exercise 2.8.10, the following is a generalization to topological spaces of Exercise 2.1.7:

Exercise 6.4.21 (The Kolmogorov quotient) Let (X, τ) be a topological space. For $x, y \in X$ define $x \sim y$ by '$\forall U \in \tau : x \in U \Leftrightarrow y \in U$'. Prove:

(i) $\sim$ is an equivalence relation.

(ii) Prove that (X, τ) is T_0 if and only if $x \sim y \Rightarrow x = y$.

(iii) The quotient space $X/\!\sim$ is T_0.

Exercise 6.4.22 Let X be a topological space, $\sim$ an equivalence relation on X, and $Y \subseteq X$. Let $\sim_Y$ be the restriction of $\sim$ to Y (i.e. $\sim_Y = \sim \cap (Y \times Y)$). Let $p : X \to X/\!\sim$ be the quotient map.

(i) Prove that there is a natural continuous bijection $\alpha : Y/\!\sim_Y \to p(Y) \subseteq X/\!\sim$.

(ii) Prove that α is a homeomorphism if $Y \subseteq X$ is $\sim$-saturated and open (or closed).

(iii) Give an example where α is not a homeomorphism.

6.4.3 A few geometric applications

According to the popular slogans mentioned in the introduction, topology is "rubber-sheet geometry" or "what remains of geometry when we forget about coordinates and distances". It therefore is high time that we make some contact with geometric intuition. Indeed, the quotient space construction is a method for making rigorous certain intuitive constructions of non-trivial spaces from simpler ones. It also plays an important rôle in algebraic topology.

When studying a quotient space $X/\!\sim$, we often have a suspicion that $X/\!\sim$ is homeomorphic to some given space Y. The examples below will illustrate that Proposition 6.4.8 provides a systematic way of proving this.

Definition 6.4.23 • *The (closed) unit n-disk is* $D^n = \{x \in \mathbb{R}^n \mid \|x\|_2 \leq 1\} \subseteq \mathbb{R}^n$.

- *The unit n-sphere is* $S^n = \{x \in \mathbb{R}^{n+1} \mid \|x\|_2 = 1\}$. *Thus* $\partial D^n = S^{n-1}$. *For* $n = 1$ *we also use the representation* $S^1 = \{z \in \mathbb{C} \mid |z| = 1\}$ *which results from the identification* $\mathbb{C} \cong \mathbb{R}^2$.

Lemma 6.4.24 *The n-sphere can be considered as quotient of* $\mathbb{R}^{n+1}\backslash\{0\}$:

$$S^n \cong (\mathbb{R}^{n+1}\backslash\{0\})/(x \sim \lambda x\ \forall \lambda > 0).$$

Proof. The map $f : \mathbb{R}^{n+1}\backslash\{0\} \to \mathbb{R}^{n+1},\ x \mapsto \frac{x}{\|x\|_2}$ takes values in S^n and as a map $\mathbb{R}^{n+1}\backslash\{0\} \to S^n$ it is surjective. It is obviously continuous and easily seen to be open. Finally, $f(x) = f(y)$ holds if and only if $\frac{x}{\|x\|_2} = \frac{y}{\|y\|_2}$, which is the case if and only $x = \lambda y$ with $\lambda > 0$. Thus $f(x) = f(y) \Leftrightarrow x \sim y$. Now Proposition 6.4.8(iii) implies that $g : \mathbb{R}^{n+1}\backslash\{0\}/\!\sim\to S^n$ is a homeomorphism. ■

The following result is the precise formulation of the statement "identifying the two ends on an interval, we obtain a circle":

Lemma 6.4.25 *Let* $\sim$ *be the equivalence relation on* $I = [0,1]$ *which only identifies* 0 *and* 1. *(Formally:* $\sim = \{(x,x) \mid x \in I\} \cup \{(0,1),(1,0)\}$.*) Then the quotient space* $I/\!\sim$ *(which we may simply denote* $[0,1]/(0 \sim 1)$*) is homeomorphic to* S^1.

Proof. The map $f : I \to S^1,\ x \mapsto e^{2\pi i x}$ is continuous and surjective. The only way for $f(x) = f(y)$ to happen with $x \neq y$ is $x = 0, y = 1$ or $\leftrightarrow$. This is exactly the equivalence relation defined in the statement. Now, let $U \subseteq [0,1]$ be open and $\sim$-saturated. This means that if U contains 0 then it must contain 1 and vice versa. If an open $U \subseteq [0,1]$ contains neither 0 nor 1 then it is $\sim$-saturated and it is easy to see that $f(U) \subseteq S^1$ is open. The other alternative is that $\{0,1\} \subseteq U$. But then there is an $\varepsilon > 0$ such that $[0,\varepsilon) \cup (1-\varepsilon,1] \subseteq U$, and using the fact that f is periodic with period 1 we have $f(U) = f(U \cap (0,1)) \cup f((-\varepsilon,\varepsilon))$, which again is open. Now apply Proposition 6.4.8(iii). ■

Remark 6.4.26 1. Note that while f satisfies condition (ii) of Proposition 6.4.8, it is not an open map since, e.g., $[0,1/2)$ is open in $[0,1]$ but $f([0,1/2)) \subseteq S^1$ is not open!

2. The preceding result has the following generalization to all dimensions: For each $n \in \mathbb{N}$, there is a homeomorphism $S^n \cong I^n/\!\sim$, where $\sim$ identifies all points of ∂I^n with each other and leaves the interior of I^n alone. (This is very useful in higher homotopy theory, cf. Exercise 13.4.23.) We defer the proof (Exercise 7.8.21) until we have the tools to do it easily. We now focus on some other quotients of $I \times I$. □

In the same way one shows (with $I^2 = I \times I$.[1])

$$\begin{aligned} C = I^2/((x,0) \sim (x,1)) &\cong I \times S^1 \quad \text{(hollow cylinder)} \\ T = I^2/((x,0) \sim (x,1), (0,y) \sim (1,y)) &\cong S^1 \times S^1 \quad (2-\text{torus}). \end{aligned}$$

More interestingly, $M = I^2/((0,y) \sim (1,1-y))$ is the well-known Möbius strip[2], obtained by applying a 180^0 twist to a ribbon before glueing two opposite ends. (Notice that while the boundary

[1] We have not yet defined the product space $I \times I$. This is no problem since we can consider it as the subspace $\{(x,y) \mid x,y \in [0,1]\}$ of $\mathbb{R}^2$ equipped with the usual (Euclidean) topology.

[2] August Ferdinand Möbius (1790-1868). German mathematician and theoretical astronomer.

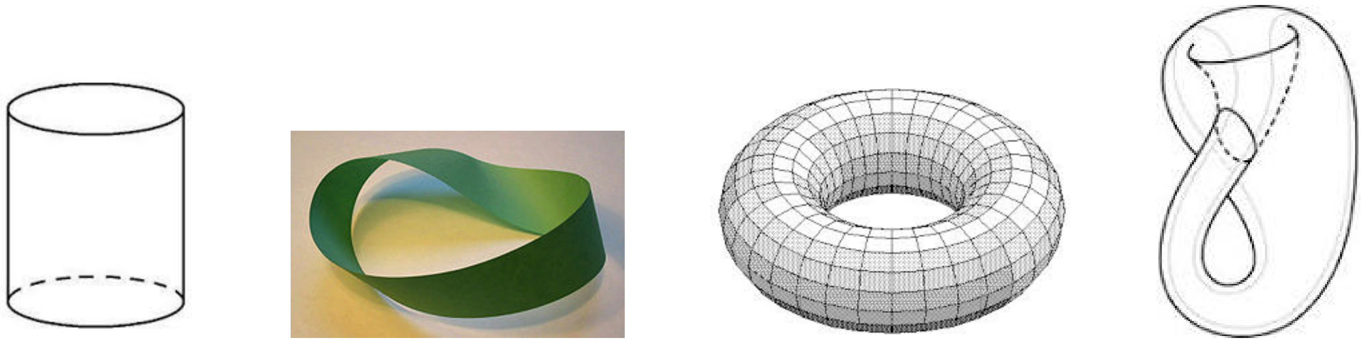


Figure 6.1: Cylinder, Möbius band, torus, Klein bottle

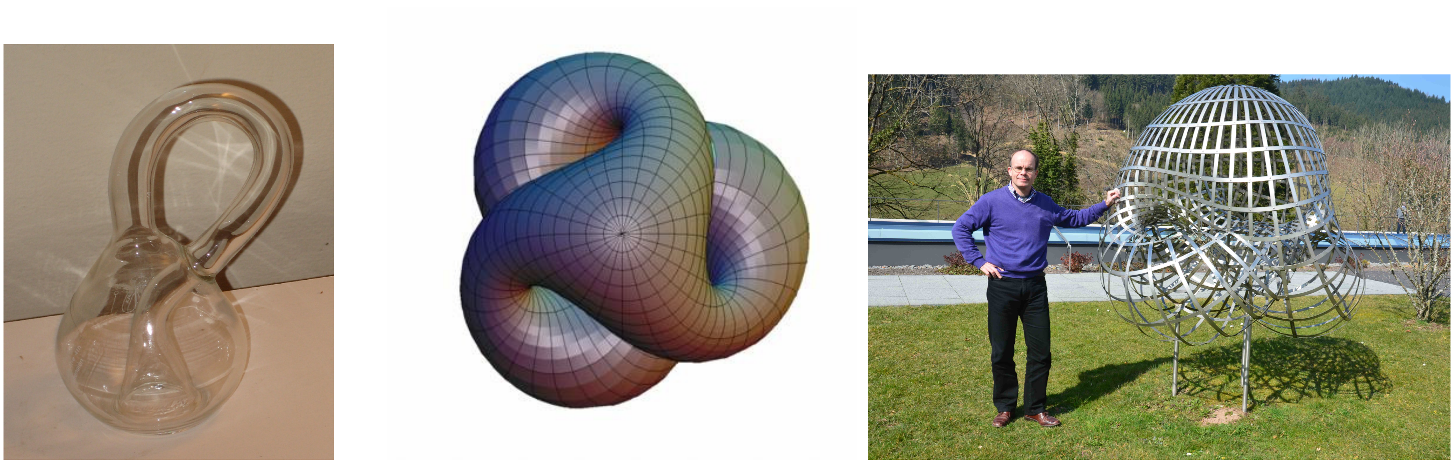


Figure 6.2: Another Klein bottle and two projective planes (Boy surface)

of the cylinder consists of two circles, the boundary of the Möbius strip is a single circle!) The 2-torus T can be obtained from the square in two steps

$$I^2 \rightsquigarrow C = I^2/((x,0) \sim (x,1)) \rightsquigarrow I^2/((x,0) \sim (x,1), (0,y) \sim (1,y)),$$

the second of which consists in gluing the cylinder $I \times S^1$ along its two boundary circles in the natural way. But the cylinder has another quotient

$$KB = C/((0,y) \sim (1,1-y)) = I^2/((x,0) \sim (x,1), (0,y) \sim (1,1-y)),$$

which is known as the Klein bottle[3], cf. Figure 6.2. Reversing the order of the quotient operations, we obtain the Klein bottle as a quotient of the Möbius strip: $KB = M/((x,0) \sim (x,1))$. (Since M has only one boundary circle ∂M, this second quotient operation consists in pairwise identification of points of ∂M, which is harder to visualize.) Like the cylinder, also the Möbius strip has two natural quotients, the second being

$$M/((x,0) \sim (1-x,1))) = I^2/((0,y) \sim (1,1-y), (x,0) \sim (1-x,1)). \tag{6.4}$$

Before we try to understand the nature of this space, we show in Figure 6.3 a very convenient way of representing the identifications made in the above quotient space constructions. In each of the diagrams, two parallel lines marked with a arrows are identified, where the orientation is preserved if the arrows point in the same direction or reversed otherwise. (In more complicated situations, one should mark the arrows in some way to make clear which sides are identified.)

[3] Christian Felix Klein (1849-1925), German mathematician.

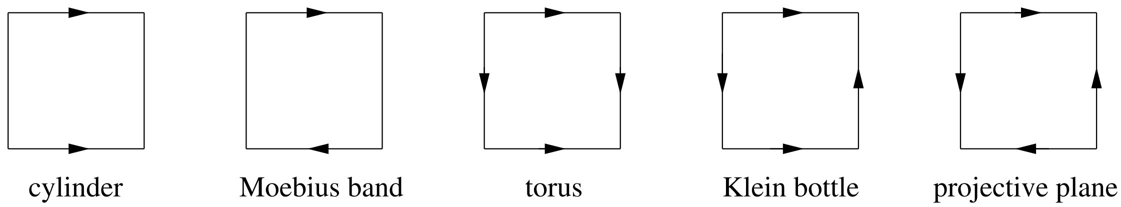


Figure 6.3: Quotient spaces of a square

Returning to the quotient space defined by (6.4), we notice that under the homeomorphism $[0,1]^2 \to [-1,1]^2, (x,y) \mapsto (2x-1, 2y-1)$, the identifications $((x,0) \sim (1-x,1), (0,y) \sim (1,1-y))$ correspond to the identification $x \sim -x$ of antipodal points in $\partial[-1,1]^2$. For a subset $X \subseteq \mathbb{R}^2$ such that ∂X is stable under $x \mapsto -x$, we denote by $\sim_\partial$ the equivalence relation that identifies antipodal points (x and $-x$) of ∂X and does nothing else. Thus we see that (6.4) is homeomorphic to $[-1,1]^2/\sim_\partial$, which is more convenient to work with. This quotient (and its higher dimensional analogues) has another, somewhat simpler interpretation as a quotient space (unfortunately equally challenging to visualize), known as the (real) projective plane $\mathbb{RP}^2$. The latter just is the first (interesting) of the following infinite family of spaces, studied in projective geometry, a subject with a long and venerable history:

Definition 6.4.27 *The real projective space $\mathbb{RP}^n$ is the quotient space $S^n/(x \sim -x)$ obtained by identifying antipodal pairs $\{x, -x\}$ of points of S^n.*

In view of Lemma 6.4.24 and Exercise 6.4.2, this is equivalent to

$$\mathbb{RP}^n \cong (\mathbb{R}^{n+1}\backslash\{0\})/(x \sim \lambda x \ \forall \lambda \neq 0),$$

which has the advantage of working (purely algebraically) for any field k instead of $\mathbb{R}$. $\mathbb{RP}^n$ can be interpreted as the "space of lines through 0 in $\mathbb{R}^{n+1}$", suitably topologized.

Lemma 6.4.28 *(i) Let $\sim_\partial$ be the equivalence relation on $[-1,1]^n$ that identifies each $x \in \partial[-1,1]^n$ with $-x$. For all $n \in \mathbb{N}$, we have homeomorphisms*

$$\frac{[-1,1]^n}{\sim_\partial} \cong \frac{D^n}{\sim_\partial} \cong \frac{S^n}{(x \sim -x)} = \mathbb{RP}^n. \tag{6.5}$$

(ii) $\mathbb{RP}^1 \cong S^1$. (Remark: We will later prove $\mathbb{RP}^n \not\cong S^n$ for $n \geq 2$.)

Proof. (i) $K = [-1,1]^n \subseteq \mathbb{R}^n$ is compact convex with non-empty interior and $-K = K$. Proposition 7.7.61, proven later, provides a homeomorphism $g : D^n \to [-1,1]^n$ that satisfies $g(-x) = -g(x) \ \forall x$. From this one deduces a homeomorphism $[-1,1]^n/\sim_\partial \to D^n/\sim_\partial$ between the quotient spaces, which is the first half of (6.5). We now wish to relate this to the projective space $\mathbb{RP}^n = S^n/(x \sim -x)$. Defining $S^n_\pm = \{x \in S^n \mid \pm x_{n+1} \geq 0\}$, we have $S^n = S^n_+ \cup S^n_-$. Now the maps $D^n \to S^n_\pm$, $x = (x_1, \ldots, x_n) \mapsto (x_1, \ldots, x_n, \pm\sqrt{1 - \|x\|^2})$ are continuous bijections with continuous inverses $(x_1, \ldots, x_{n+1}) \mapsto (x_1, \ldots, x_n)$, thus homeomorphisms. Thus we can interpret S^n as two discs D^n glued together at their boundaries. The map $x \mapsto -x$ clearly is a homeomorphism $S^n_+ \to S^n_-$. Thus if $p : S^n \to S^n/(x \sim -x)$ is the quotient map, we have $p(S^n_+) = S^n/(x \sim -x)$. Restricted to the interior of S^n_+, the map p is injective, but it identifies $(x_1, \ldots, x_n, 0)$ with $(-x_1, \ldots, -x_n, 0)$. Thus set-theoretically we have $S^n/(x \sim -x) \cong D^n/\sim_\partial$, and as in the proof of Lemma 6.4.25 one applies Proposition 6.4.8 to prove that this is a homeomorphism. (Or use Proposition 7.4.11(iv).) This is the second half of (6.5).

(ii) This follows from the instance $\mathbb{RP}^1 \cong D^1/\sim_\partial$ of (i), together with $D^1 = [-1,1]$ and Lemma 6.4.25. ■

Remark 6.4.29 The above constructions of quotient spaces are 'purely topological' in that they do not take place in some ambient space $\mathbb{R}^n$ into which everything is embedded. Cylinder and 2-torus can be embedded into $\mathbb{R}^3$, but the Klein bottle cannot. It can be 'immersed' into $\mathbb{R}^3$, but only at the price of self-intersections. It is easy to see that it can be embedded into $\mathbb{R}^4$. For the projective plane $\mathbb{RP}^2$, the same is true. A nice immersion into $\mathbb{R}^3$ was found by W. Boy in 1901, having been instructed by his supervisor D. Hilbert to prove that such an immersion does not exist! Cf. [44, Section 2.2] for a thorough explanation. □

Exercise 6.4.30 Use Lemma 6.4.14 to prove that the quotient spaces C, T, KB, M, $\mathbb{R}P^2$ considered in this subsection are Hausdorff. (Do not use the homeomorphisms with known spaces proven here!)

6.5 Direct products

6.5.1 Basics

Given any family $\{X_i\}_{i\in I}$ of sets, we can define the direct product $\prod_{i\in I} X_i$, cf. Appendix A.2.2. In this section we will study topologies on $\prod_i X_i$, assuming that each X_i is a topological space.

Definition 6.5.1 *Let $(X_i, \tau_i), i \in I$ be topological spaces. The product topology τ_Π on $\prod_k X_k$ is the initial topology defined by the projection maps $p_i : \prod_k X_k \to X_i,\ f \mapsto f(i)$. The topological space $(\prod_k X_k, \tau_\Pi)$ is also denoted by $\prod_k (X_k, \tau_k)$.*

As for sums, if I is finite we will usually take $I = \{1, \ldots, n\}$ and also denote $\prod_i X_i$ by $X_1 \times \cdots \times X_n$. (Remark 6.3.2 applies also here.)

What does Definition 6.5.1 mean in more concrete terms? By Lemma 6.1.6(ii), a subbase for τ_Π is given by
$$\mathcal{S}_\Pi = \{p_i^{-1}(U) \mid i \in I,\ U \in \tau_i\}. \tag{6.6}$$
Thus a base $\mathcal{B}$ for τ_Π is obtained by considering all finite intersections of elements of $\mathcal{S}_\Pi$:
$$\mathcal{B}_\Pi = \{p_{i_1}^{-1}(U_{i_1}) \cap \cdots \cap p_{i_n}^{-1}(U_{i_n}) \mid i_1, \ldots, i_n \in I,\ U_{i_k} \in \tau_{i_k}\ \forall k = 1, \ldots, n\}. \tag{6.7}$$

Since we defined the product topology on $\prod_i (X_i, \tau_i)$ as an initial topology, the general results from Section 6.1 apply. Our first result, which is just the specialization of Proposition 6.1.8 to the product, shows that the universal property of the product of sets, cf. Proposition A.2.7, lifts to topological spaces and continuous maps:

Proposition 6.5.2 *Given topological spaces (X, τ) and $(Y_i, \sigma_i), i \in I$, there is a bijection between continuous maps $f : X \to \prod_k Y_k$ (with the product topology) and families of continuous maps $\{f_i : X \to Y_i\}_{i\in I}$, given by $f \mapsto \{p_i \circ f\}_{i\in I}$.*

Lemma 6.5.3 *A net $\{x_\iota\}$ in a product space $\prod_i (X_i, \tau_i)$ converges if and only if the net $\{p_i(x_\iota)\}$ in X_i converges for each $i \in I$. ("The net converges coordinatewise".)*

Proof. This is just an application of Proposition 6.1.9, modulo one observation: If $p_i(x) \to x_i \in X_i$ for all $i \in I$, there is a unique point x in $\prod_k X_k$ such that $p_k(x) = x_k\ \forall k$. (This is not true in the generality of Proposition 6.1.9.) ■

Remark 6.5.4 If $I = \bigcup_{k\in K} I_k$ is a partition of the index set I, i.e. $I_i \cap I_j = \emptyset$ for $i \neq j$, then there is a canonical homeomorphism

$$\prod_{i\in I} X_i \cong \prod_{k\in K} \left(\prod_{j\in I_k} X_j \right).$$

We omit the trivial but tedious details. □

Remark 6.5.5 If (X,τ) is a topological space and I is a set, we write $(X,\tau)^I$ for $\prod_{i\in I}(X,\tau)$. In view of Lemma 6.5.3, this is just the set of all functions $f : I \to X$, equipped with the topology of pointwise convergence. (When there is no risk of confusion about the topology τ, we may simply write X^I.) Since X^I depends only on the cardinality of I, this also defines $(X,\tau)^N$ for a cardinal number N. Notice that when X is a topological space, Y^X often denotes the set $C(X,Y)$ of *continuous* functions (which can also be topologized). When X is discrete, $C(X,Y)$ coincides with Y^X as just defined, which hopefully will limit the risk of confusion. □

Proposition 6.5.2 and Lemma 6.5.3 clearly show that our definition of the product topology is 'the right one', even though it is not very intuitive. In order to obtain better insight, and to preempt misconceptions about the product topology, it is very instructive to consider another, probably more intuitive, topology on $\prod_i X_i$:

Exercise 6.5.6 Let (X_i,τ_i) be a topological space for each $i \in I$, and let $X = \prod_i X_i$.

(i) Prove that

$$\mathcal{B}'_\Pi = \left\{ \prod_{i\in I} U_i \mid U_i \in \tau_i \ \ \forall i \in I \ \text{ and } \ \#\{i \in I \mid U_i \neq X_i\} < \infty \right\}$$

satisfies the conditions of Proposition 4.1.20 and therefore is a base for a topology τ'_Π on X.

(ii) Prove that $\mathcal{B}'_\Pi$ equals $\mathcal{B}_\Pi$ from (6.7), thus τ'_Π equals the product topology τ_Π.

(iii) If $\emptyset \neq U \in \tau_\Pi$, prove that $\{i \in I \mid p_i(U) \neq X_i\}$ is finite.

(iv) Defining

$$\mathcal{B}_\Box = \left\{ \prod_{i\in I} U_i \mid U_i \in \tau_i \ \ \forall i \in I \right\},$$

prove that $\mathcal{B}_\Box$ is the base for a topology $\tau_\Box$ on $\prod_k X_k$, the box topology.

(v) Prove $\tau_\Box \supseteq \tau_\Pi$. (Thus $\tau_\Box$ is finer than the product topology τ_Π.)

(vi) Prove that $\tau_\Box = \tau_\Pi$ when $\#I < \infty$.

(vii) Prove that $\tau_\Box \neq \tau_\Pi$ when $\{i \in I \mid \tau_i \neq \tau_{\text{indisc}}\}$ is infinite. (I.e. infinitely many of the X_i's have an open subset that is different from $\emptyset$ and X_i.)

Remark 6.5.7 By (vi) above, there is no harm in thinking in terms of the more intuitive box topology $\tau_\Box$ when dealing with **finite** products. But (except certain trivial cases), the box topology on an infinite product of non-trivial spaces differs from the product topology. This clearly means that it cannot have the same properties as the latter. In fact: If we replace the product topology τ_Π by the box topology $\tau_\Box$, Proposition 6.5.2 and Lemma 6.5.3 are false for infinite products! (Since the base $\mathcal{B}_\Box$ of $\tau_\Box$ can be written as $\{\prod_i U_i = \bigcap_i p_i^{-1}(U_i),\ U_i \in \tau_i\ \forall i\}$, adapting the proof of Lemma 6.5.3,

we would then need to find a ι_0 bigger than an infinite number of ι_i's, and there is no guarantee that this exists.) Also the facts that arbitrary products of connected (or compact) spaces are connected (respectively compact), which we will prove later, are false for the box topology! □

We know by construction that the projections $p_i : \prod_k X_k \to X_i$ are continuous. Furthermore:

Proposition 6.5.8 *Let $(X_i, \tau_i),\ i \in I$ be topological spaces. Then for each $i \in I$ the projection $p_i : \prod_k X_k \to X_i$ is open, and therefore a quotient map.*

Proof. If one of the X_k is empty, so is the product, and the claim is true. Thus we may assume $X_k \neq \emptyset\ \forall k$. In view of Remark 5.2.24 it suffices to show that $p_i(U) \subseteq X_i$ is open for the elements U of the base $\mathcal{B}_\Pi$ in (6.7). In view of Exercise 6.5.6(i), every $U \in \mathcal{B}$ is of the form $\prod_{i\in I} U_i$ where $U_i \in \tau_i$ and all but finitely many U_i are equal to X_i. From this it is obvious that $p_j(U) = U_j$, which is open. Thus the map p_j is open and thus a quotient map by Lemma 6.4.5. ■

It will follow from Exercise 6.5.35 below that the product topology on the n-fold product $\mathbb{R}\times\cdots\times\mathbb{R}$ coincides with the standard topology on $\mathbb{R}^n$ arising from any of the norms discussed in Example 2.1.13 (which are all equivalent, cf. Exercise 2.2.16(iii) and Theorem 7.7.51). For the purpose of the next exercise, this may be assumed.

While in certain situations one can prove projection maps to be closed, cf. Exercise 7.5.5, in general they are not:

Exercise 6.5.9 Prove that the projection $p_1 : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is not closed.

If $\emptyset \neq U_i \in \tau_i\ \forall i \in I$ then $\prod_k U_k$ is in the box topology $\tau_\Box$, but it is in τ_Π only if all but finitely many U_i are X_i. On the other hand, products of closed subsets behave nicely:

Exercise 6.5.10 Let $(X_i, \tau_i),\ i \in I$ be topological spaces and $C_i \subseteq X_i,\ i \in I$ closed subsets. Prove that $\prod_i C_i \subseteq \prod_i X_i$ is closed.

Exercise 6.5.11 For $A \subseteq X, B \subseteq Y$, compute $(A \times B)^0$, $\overline{A \times B}$ and $\partial(A \times B)$ in terms of $A^0, \overline{A}, B^0, \overline{B}$.

Exercise 6.5.12 Let X_i be a topological space for each $i \in I$.

(i) Let $A_i \subseteq X_i$ for all $i \in I$. Prove $\overline{\prod_i A_i} = \prod_i \overline{A_i}$.

(ii) If $A_i \subseteq X_i$ is dense for all $i \in I$, prove that $\prod_i A_i$ is dense in $\prod_i X_i$.

(iii) If $X_1, \ldots, X_n$ are separable, prove that $\prod_{i=1}^n X_i$ is separable.

(iv) Redo (iii), now for countable I. (You need a new approach, but may of course use (iii).)

Lemma 6.5.13 *Let $(X_i, \tau_i),\ i \in I$ be topological spaces. Let $z \in \prod_k X_k$. (Thus in particular $X_i \neq \emptyset\ \forall i$.) For each $i \in I$, define a map $\iota_{z,i} : X_i \to \prod_k X_k$ by saying that $\iota_{z,i}(x)$ is the unique point y defined by*

$$p_k(y) = \begin{cases} z_k & \text{if } \ k \neq i \\ x & \text{if } \ k = i \end{cases}$$

Then $\iota_{z,i}$ is an embedding, thus each (X_i, τ_i) is homeomorphic to a subspace of $\prod_k (X_k, \tau_k)$. This subspace is closed if all X_i are T_1.

Proof. Injectivity of $\iota_{z,i}$ is clear, thus $\iota_{z,i} : X_i \to \iota_{z,i}(X_i)$ is a bijection. The image of $\iota_{z,i}$ is $\bigcap_{j\neq i} p_j^{-1}(z_j) \subseteq \prod_k X_k$, and the subspace topology on this coming from the product topology on $\prod_k X_k$ is is exactly τ_i. Thus $\iota_{z,i}$ is an embedding. If all X_i are T_1 then $\{z_i\} \subseteq X_i$ is closed for each i, thus $X_i \times \prod_{j\neq i}\{z_j\} \subseteq \prod_i X_i$ is closed by Exercise 6.5.10. ■

Corollary 6.5.14 *Let P be a hereditary property of topological spaces. If $X_i \neq \emptyset\ \forall i \in I$ and $\prod_i X_i$ has property P then every X_i has property P.*

Recall that first and second countability, the separation axioms T_1 and T_2 and metrizability are hereditary. This already gives part of the following:

Exercise 6.5.15 Let (X_i, τ_i) be topological spaces with $X_i \neq \emptyset\ \forall i \in I$. Prove that

(i) $\prod_i(X_i, \tau_i)$ is T_1 if and only if and only if each (X_i, τ_i) is T_1.

(ii) $\prod_i(X_i, \tau_i)$ is T_2 if and only if and only if each (X_i, τ_i) is T_2.

(iii) $\prod_i(X_i, \tau_i)$ is first (resp. second) countable if and only if each (X_i, τ_i) is first (resp. second) countable and at most countably many X_i are not indiscrete.

Using products, we can improve on Exercise 4.1.13:

Lemma 6.5.16 *(i) The Sorgenfrey plane $(\mathbb{R}, \tau_S)^2$ has a closed discrete subspace that has the cardinality $\mathfrak{c}$ of the continuum.*

(ii) There exists a Hausdorff space that is separable, but not hereditarily separable.

Proof. (i) Let $X = (\mathbb{R}, \tau_S) \times (\mathbb{R}, \tau_S)$ and $Y = \{(x, -x) \mid x \in \mathbb{R}\} \subseteq X$. Now Y is closed w.r.t. the usual topology and therefore closed w.r.t. the topology $(\tau_S)^2$ since τ_S is finer than the standard topology on $\mathbb{R}$. Now $[x, \infty) \times [-x, \infty) \in \tau_S^2$ and it intersects Y only in $(x, -x)$. Thus $\{(x, -x)\} \in Y$ is open, so that the subspace topology on Y is discrete. Since there is an obvious bijection $Y \cong \mathbb{R}$, Y has cardinality $\mathfrak{c}$.

(ii) By Exercises 4.1.22 and 6.5.15(ii), the Sorgenfrey plane is Hausdorff. By Exercise 4.3.12, the Sorgenfrey line is separable, thus also the Sorgenfrey plane is separable by Exercise 6.5.12. But by (i), $(\mathbb{R}, \tau_S)^2$ has a closed discrete subspace that has cardinality $\mathfrak{c}$ and thus is not separable. ■

Exercise 6.5.17 Prove that $\mathbb{R}^n\backslash\{0\} \cong S^{n-1} \times (0, \infty)$ by giving mutually inverse continuous maps both ways.

The following exercises show that Hausdorffness of spaces and continuity of functions can be characterized in terms of closedness of certain subsets in direct products:

Exercise 6.5.18 Prove the following statements:

(i) A space X is Hausdorff if and only if the <u>diagonal</u> $\Delta_X = \{(x, x) \mid x \in X\} \subseteq X \times X$ is closed.

(ii) Use (i) to give a new proof of Exercise 5.2.16(ii).

Definition 6.5.19 *Let $f : X \to Y$ a function. Then the <u>graph</u> of f is the subset $G(f) = \{(x, f(x)) \mid x \in X\} \subseteq X \times Y$.*

Remark 6.5.20 1. For $\mathrm{id}_X : X \to X$ we find $G(\mathrm{id}_X) = \Delta_X$.

2. Since a function $f : X \to Y$ is defined as a relation $R \subseteq X \times Y$ satisfying some assumptions, $G(f)$ in principle just **equals** f, but the perspective is somewhat different. □

Exercise 6.5.21 (i) For $f : X \to Y$, prove $G(f) = (\mathrm{id}_X \times f)(\Delta_X) = (f \times \mathrm{id}_Y)^{-1}(\Delta_Y)$.

(ii) Prove: If Y is Hausdorff and $f : X \to Y$ is continuous then $G(f) \subseteq X \times Y$ is closed.

Remark 6.5.22 For the converse implication of (ii), we need stronger assumptions. Cf. Exercise 7.5.7. □

Exercise 6.5.23 Let X be a topological space and $X_i \subseteq X$ $\forall i \in I$. Define $Y = \bigcap_i X_i$ and

$$f : Y \to \prod_{i \in I} X_i, \quad y \mapsto \prod_{i \in I} y,$$

which makes sense since $Y \subseteq X_i$ $\forall i$.

(i) Prove that $f : Y \to \prod_i X_i$ is an embedding.

(ii) If X is Hausdorff, prove that $f(Y) \subseteq \prod_i X_i$ is closed.

6.5.2 ⋆⋆ More on separability and the Souslin property for products

This section is somewhat more difficult than what we have done so far and may be skipped since it is not essential for what follows.

Remarkably, one can do better than Exercise 6.5.12(iv):

Exercise 6.5.24 Prove:

(i) If $\#I \leq \mathfrak{c} = \#\mathbb{R}$ and X_i is separable for each $i \in I$ then $\prod_i X_i$ is separable.

(ii) If $\prod_{i \in I} X_i$ is separable and each X_i reducible then $\#I \leq \#\mathbb{R}$.

(For reducibility see Definition 2.8.2. Every Hausdorff space with ≥ 2 points is reducible.)

Remark 6.5.25 1. There also is a necessary and sufficient condition for separability of $\prod_i X_i$.

2. More general (but not stronger in the case $\#I = \mathfrak{c}$) is the Hewitt-Marczewski-Pondiczery[4] theorem: Let χ be an infinite cardinal number, X_i a space admitting a dense subset of cardinality $\leq \chi$ for each $i \in I$, where $\#I \leq 2^\chi$. Then $\prod_i X_i$ has a dense subset of cardinality $\leq \chi$. Cf. e.g. [75, Theorem 2.3.15]. □

Exercise 6.5.15(iii) and Exercise 6.5.24 show that separability, which is weaker than second countability, is better behaved under products than the latter.

This leads to the question how the Souslin property, which is even weaker than separability, behaves under products. It turns out that the innocent-looking **question whether the product of any two spaces with the Souslin propery is Souslin cannot be answered based on the usual ZFC axioms of set theory** (including the axiom of choice)!

The problem turns out to be related to the Continuum Hypothesis (CH), which is the statement that every uncountable subset of $\mathbb{R}$ has the same cardinality as $\mathbb{R}$. It was shown by Gödel[5] (1940) and

[4]Edwin Hewitt (1920-1999), American mathematician who mostly worked in topology and analysis. Edward Marczewski (1907-1976), Polish mathematician (until 1940 Edward Szpilrajn). E. S. Pondiczery actually was a pseudonym of the American mathematician Ralph P. Boas, Jr. (1912-1992).

[5]Kurt Friedrich Gödel (1906-1978). Austrian (later American) mathematician and logician.

Cohen[6] (1963/4), respectively, that there are models of ZFC set theory where CH is true, respectively false. One can prove that the the product of two Souslin spaces always is Souslin when Martin's[7] axiom $MA(\aleph_1)$ holds. CH implies that $MA(\aleph_1)$ is false, but there are models of set theory in which $MA(\aleph_1) \wedge \neg CH$ holds, and in such a set theory, products of Souslin spaces are Souslin. On the other hand, under the set theoretic 'diamond axiom' $\diamond$ (which is somewhat stronger than CH), one can construct Souslin spaces X, Y such that $X \times Y$ is not Souslin. For more on this see [151, 51].

This is perhaps the simplest instance of a question in general topology that can be answered definitely only making set-theoretic assumptions beyond ZFC. There are many others, and a considerable part of research in general topology since Cohen has been predicated on stronger set-theoretic hypotheses.[8]

However, one can prove the following:

Proposition 6.5.26 *(i) If $f : X \to Y$ is continuous and X has the Souslin property then so does $f(X)$.*

(ii) Let $X_i \neq \emptyset\ \forall i$. Then $\prod_{i\in I} X_i$ has the Souslin property if and only if $\prod_{j\in J} X_j$ has the Souslin property for each finite subset $J \subseteq I$.

Since separability implies the Souslin property and is preserved under finite products, one has the following remarkable consequence:

Corollary 6.5.27 *If X_i is separable for every $i \in I$ then $X = \prod_{i\in I} X_i$ has the Souslin property.*

Example 6.5.28 The cube $[0,1]^\chi$ is Hausdorff and has the Souslin property for all χ, but for $\#\chi > \mathfrak{c} = \#\mathbb{R}$ it is not separable. □

The proof of Proposition 6.5.26 requires the following lemma from infinitary combinatorics, proven in Section A.3.6 using transfinite recursion:

Lemma 6.5.29 (Δ-system lemma) *If $\mathcal{A}$ is an uncountable family of finite sets then there exist an uncountable subfamily $\mathcal{A}_0 \subseteq \mathcal{A}$ and a finite set A such that $X \cap Y = A$ for all $X, Y \in \mathcal{A}_0$ with $X \neq Y$.*

Proof of Proposition 6.5.26. (i) Let $f : X \to Y$ be continuous and surjective, where X has the Souslin property. Let $\mathcal{U}$ be a family of mutually disjoint non-empty open sets in Y. Then $\{f^{-1}(U) \mid U \in \mathcal{U}\}$ is a family of disjoint open sets in X that are non-empty due to the surjectivity of f. Since X is Souslin, $\mathcal{U}$ must be countable. Thus the continuous image of a space with the Souslin property also has that property.

(ii) ⇒ If $X_i \neq \emptyset\ \forall i$ and $X = \prod_i X_i$ has the Souslin property, then $X_i = p_i(X)$ has the Souslin property by (i).

[6] Paul Joseph Cohen (1934-2007). American mathematician. He received the Fields medal for this work.

[7] Donald A. Martin (1940-). American set theorist and philosopher of mathematics

[8] The reader should not think that this surprising state of affairs arises only in topology. See e.g. [74] for a similar situation in algebra involving just discrete abelian groups (Whitehead's problem). [An abelian group A is called a Whitehead group if for every surjective homomorphism $\alpha : B \to A$, where B is an abelian group and $\ker(\alpha) \cong \mathbb{Z}$, there is a homomorphism $\beta : A \to B$ such that $\alpha \circ \beta = \mathrm{id}_A$. (Thus $B \cong A \oplus \mathbb{Z}$.) It is easy to show that every free abelian group is Whitehead, and with more effort one proves that every countable Whitehead group is free. Whitehead's problem is to prove that uncountable Whitehead groups are free. This was done by S. Shelah (Israeli mathematician and logician, born 1945) assuming Gödel's axiom $V = L$ of constructibility (which implies $\diamond$ and CH). On the other hand, assuming $MA(\aleph_1) \wedge \neg CH$ one can construct Whitehead groups that are not free. (Shelah later showed that existence of non-free Whitehead groups is possible even assuming GCH.) Thus $MA(\aleph_1) \wedge \neg CH$ implies a simple state of affairs in topology but not so in algebra, while $V = L$ implies a nice statement in algebra, but not in topology!]

$\Leftarrow$ Assume $\{U_j\}_{j\in J}$ is an uncountable family of mutually disjoint non-empty open sets in $X = \prod_{i\in I} X_i$. Each U_j contains a (non-empty) basic open set of the form $V_j = \prod_{i\in I} W_{k,i}$, where each $W_{k,i} \subseteq X_i$ is open and $A_j = \{i \in I \mid W_{j,i} \neq X_i\}$ is finite. For $j, j' \in J$, $j \neq j'$ we cannot have $A_j \cap A_{j'} = \emptyset$ since that would imply $V_j \cap V_{j'} \neq \emptyset$, contradicting the assumption that the U_j are mutually disjoint. Now, $\mathcal{A} = \{A_j \mid j \in J\}$ is an uncountable family of finite sets, and the Δ-system lemma provides an uncountable subfamily $\mathcal{A}_0 = \{A_j \mid j \in J_0\} \subseteq \mathcal{A}$ and a finite $A \subseteq I$ such that $j, j' \in J_0,\ j \neq j' \Rightarrow A_j \cap A_{j'} = A$. Since any two A_j have non-empty intersection, we have $A \neq \emptyset$. The projection π from $X = \prod_{i\in I} X_i$ to the finite product $\prod_{i\in A} X_i$ is an open map. [It suffices to prove that $\pi(U)$ is open for basic open sets in $\prod_i X_i$. This is quite obvious since the latter are of the form $\prod_i V_i$, where each V_i is open in X_i and $U_i = X_i$ for all but finitely many i.] If $j, j' \in J_0,\ j \neq j'$ then $A_j \cap A_{j'} = A \neq \emptyset$, which implies $\pi(U_j) \cap \pi(U_{j'}) = \emptyset$ since $V_j \cap V_{j'} = \emptyset$, this disjointness being due to the coordinates in $A = A_j \cap A_{j'}$. Thus $\{\pi(U_j)\}_{j\in J}$ is an uncountable family of mutually disjoint non-empty open sets in the finite product $\prod_{i\in A} X_i$, contradicting the assumption that the latter has the Souslin property. ■

An alternative (but related) proof of Corollary 6.5.27 uses the Knaster property:

Definition 6.5.30 *A topological space X has the Knaster property if every uncountable family $\mathcal{U}$ of non-empty open subsets has an uncountable subfamily $\mathcal{U}' \subseteq \mathcal{U}$ such that $U, V \in \mathcal{U}' \Rightarrow U \cap V \neq \emptyset$.*

Exercise 6.5.31 Prove:

(i) The following implications hold: Separability $\Rightarrow$ Knaster property $\Rightarrow$ Souslin property.

(ii) If X_i has the Knaster property for each $i \in I$ then $\prod_i X_i$ has the Knaster property.

Hint: (i) is easy. The proof of (ii) has much in common with that of Proposition 6.5.26(ii).

6.5.3 Products of metric spaces

Definition 6.5.32 *Let (X_i, d_i) be a metric space for each $i \in I$. A product metric on the product $X = \prod_{i\in I} X_i$ is a metric D such that the metric topology τ_D equals the product $\tau_\Pi = \prod_{i\in I} \tau_{d_i}$ of the metric topologies.*

Exercise 6.5.33 Let (X_i, d_i) be metric spaces for $i = 1, \ldots, n$. Defining

$$D_\infty(x,y) = \max_{i\in\{1,\ldots,n\}} d_i(x_i, y_i),$$

$$D_s(x,y) = \left(\sum_{i=1}^n d_i(x_i,y_i)^s\right)^{1/s} \qquad (1 \le s < \infty),$$

prove:

(i) D_s is a metric on X for each $s \in [1, \infty]$. Hint: For $s < \infty$, use Minkowski's inequality (2.4).

(ii) A sequence $\{x^k\}$ in $\prod_i X_i$ converges w.r.t. D if and only if it converges coordinatewise, i.e. $\{x_i^k\}$ converges in (X_i, d_i) for each i.

(iii) D is a product metric.

Corollary 6.5.34 *For all $n \in \mathbb{N}$, $\mathbb{R}^n$ equipped with the Euclidean topology is homeomorphic to the direct product of n copies of $\mathbb{R}$ (with the Euclidean topology).*

If I is uncountable and (X_i, d_i) satisfies $\#X_i \geq 2$ for each i then $\prod_i (X_i, \tau_{d_i})$ is not first countable by Exercise 6.5.15 and therefore is not metrizable. On the other hand:

Exercise 6.5.35 Let $(X_n, d_n), n \in \mathbb{N}$ be non-empty metric spaces such that all d_n are bounded by 1. For $x, y \in X = \prod_n X_n$ (i.e. $x = (x_1, x_2, \ldots)$) we define

$$D(x,y) = \sum_{n=1}^{\infty} 2^{-n} d_n(x_n, y_n).$$

(i) Prove that D is a metric on $\prod_n X_n$.

(ii) Prove that a sequence $\{x^k\}$ in X converges to $z \in X$ w.r.t. D if and only if it converges coordinatewise, i.e. $\lim_k x_n^k = z_n$ for each n.

(iii) Prove that (X, D) is complete if and only if each (X_n, d_n) is complete.

(iv) Give a direct proof of $\tau_D = \tau_\Pi := \prod_n \tau_{d_n}$. (A less direct proof follows from (ii) and the proof of Exercise 6.5.33(iii).)

Corollary 6.5.36 *Countable products of (completely) metrizable spaces are (completely) metrizable.*

Proof. Choose (complete) metrics on the spaces inducing the given topologies. By Exercises 2.2.14 and 3.1.10 we may assume that the metrics are bounded by 1. Now apply Exercise 6.5.35. ■

Remark 6.5.37 For a product of metric spaces there is a third topology besides τ_Π and $\tau_\Box$:
Let $\{(X_i, d_i)\}_{i \in I}$ be any number of metric spaces. For $x, y \in X = \prod_i X_i$, define

$$D(x,y) = \sup_{i \in I} d_i'(x_i, y_i) \quad \text{where} \quad d_i'(x_i, y_i) = \min(1, d_i(x_i, y_i)).$$

This clearly is a metric on X, the metric of uniform convergence. Convergence w.r.t. τ_D implies pointwise convergence ($x_i \to y_i \ \forall i \in I$), thus τ_D is finer than the product topology τ_Π. Since τ_D is a metric topology, it has $\mathcal{B}_D = \{B^D(x,r) = \prod_i B^{d_i'}(x_i, r) \mid x \in X, \ r > 0\}$ as base. Thus $\tau_\Pi \subseteq \tau_D \subseteq \tau_\Box$. (The three topologies coincide for finite I.) □

6.5.4 Joint versus separate continuity

The universal property of the product, as given by Proposition 6.5.2, tells us that a map into a product of topological spaces is well-behaved if and only if it is well-behaved component-wise. It is essential to understand that this is not at all true for maps out of a product-space! The slogan is: **Separate continuity does not imply joint continuity!**

Definition 6.5.38 *Given topological spaces (X_i, τ_i), $i \in I$ and a function $f : \prod_k X_k \to (Y, \sigma)$, we say that f is separately continuous if the map $f \circ \iota_{z,i} : X_i \to Y$ is continuous for each $z \in \prod_k X_k$ and each $i \in I$ (where $\iota_{z,i}$ is as in Lemma 6.5.13). Less formally: f is continuous w.r.t. $x_i \in X_i$ when the other $x_j, j \neq i$ are kept fixed.*

Occasionally, continuity of $f : \prod_k X_k \to Y$ in the usual sense, namely w.r.t. the topologies τ_Π and σ, is called joint continuity in order to distinguish it from separate continuity.

That separate continuity of $f : \prod_k X_k \to Y$ does not imply continuity of f, not even in the case $[0,1] \times [0,1] \to \mathbb{R}$ is shown by the following simple example which should be known from Analysis courses. (Cf. e.g. [224, Exercise 13.2.11].) Unfortunately, experience tells that it isn't (or has already been forgotten).

Consider $f : [0,1] \times [0,1] \to \mathbb{R}$ defined by

$$f(x,y) = \begin{cases} \frac{xy}{x^2+y^2} & \text{if } (x,y) \neq (0,0) \\ 0 & \text{if } (x,y) = (0,0) \end{cases}$$

It is easy to check that the maps $[0,1] \to \mathbb{R}$ given by $x \mapsto f(x,y_0)$ and $y \mapsto f(x_0,y)$ are continuous for all $x_0, y_0 \in [0,1]$. But f is NOT jointly continuous: f vanishes on the axes $x = 0$ and $y = 0$ (and therefore is continuous restricted to them), but $f(x,x) = 1/2$ for $x \neq 0$. Thus f is not continuous at $(0,0)$.

Another way to look at f is using polar coordinates $(x,y) = r(\cos\varphi, \sin\varphi)$. Then $f(r,\varphi) = \cos\varphi \sin\varphi = \frac{\sin 2\varphi}{2}$. This function is independent of r but not of φ. It assumes all values in $[0,1/2]$ in any neighborhood of $(x,y) = (0,0)$ and therefore clearly is not continuous.

The problem with f is that its limit as we approach $(0,0)$ depends on the direction of approach, a fact that we do not see if we move only horizontally and vertically in the plane. In view of Proposition 5.2.5, we must (at least a priori) consider all ways of approaching a point in the product space in order to test whether f is continuous there.

Under certain assumptions on the spaces, there are simpler ways to check joint continuity:

Theorem 6.5.39 *If $f : X \times Y \to Z$ is jointly continuous and $h : X \to Y$ is continuous then $f_h : X \to Z,\ x \mapsto f(x,h(x))$ is continuous.*

If $X = Y = [0,1]$ and $Z = \mathbb{R}$, then continuity of f_h for all $h \in C(X,Y)$ implies joint continuity of f.

More generally, this holds under the following assumptions: X, Y, Z are first countable, X is $T_{3.5}$ and dense-in-itself, Y is locally path-connected, and Z is T_3. (Those properties that have not yet been defined soon will be. First countability, T_3 and $T_{3.5}$ follow from metrizability.)

Proof. The map $X \to X \times Y,\ x \mapsto (x,h(x))$ is continuous (by Proposition 6.5.2!), so if f is (jointly!) continuous then f_h is continuous. The special case $X = Y = [0,1]$ and $Z = \mathbb{R}$ of the converse was proven by Luzin in 1948, and the generalization can be found in [55]. ■

An often more convenient approach to proving joint continuity is the following:

Exercise 6.5.40 (i) Recall that if A, B are sets we write $\mathrm{Fun}(A,B)$ for the set of all functions $f : A \to B$. State the obvious map

$$\Lambda : \mathrm{Fun}(X \times Y, Z) \to \mathrm{Fun}(X, \mathrm{Fun}(Y,Z)),$$

and prove that it is a bijection by giving the inverse map.

(ii) Let X, Y be topological spaces and (Z,d) a metric space. Prove: If $f \in \mathrm{Fun}(X \times Y, Z)$ is such that $\Lambda(f) \in C(X, C_b(Y,Z))$, where $C_b(Y,Z)$ has the topology coming the metric $D(g,h) = \sup_y d(g(y),h(y))$, then $f \in C(X \times Y, Z)$.

Thus: If f is continuous and bounded in y for every x and <u>continuous in x uniformly in y</u> then f is jointly continuous. (For a converse result, cf. Exercise 7.7.45(i).)

6.6 $\star$ Pushouts and Pullback (fiber product)

The four constructions in the preceding sections – subspaces, quotient spaces, direct sums and direct products – can be combined in many ways to produce more complicated spaces. Here we briefly look at two of the more important ones, even if their main use lies in algebraic topology.

Definition 6.6.1 *Let X_0, X_1, X_2 be sets and $f_i : X_0 \to X_i$, $i = 1, 2$ functions. Consider the canonical inclusion maps $\iota_i : X_i \to X_1 \oplus X_2$. Let $\sim$ be the smallest equivalence relation on $X_1 \oplus X_2$ identifying $\iota_1(f_1(x))$ with $\iota_2(f_2(x))$ for each $x \in X_0$. Then $X_1 \oplus_{X_0} X_2 := (X_1 \oplus X_2)/\!\sim$ is called the pushout for the 'diagram' $X_1 \overset{f_1}{\leftarrow} X_0 \overset{f_2}{\rightarrow} X_2$.*

The pushout comes with canonical maps $g_i : X_i \to X_1 \oplus_{X_0} X_2$, $i = 1, 2$ (defined as $g_i = p \circ \iota_i$, where $p : X_1 \oplus X_2 \to X_1 \oplus_{X_0} X_2$ is the quotient map). There also is a map $f : X_0 \to X_1 \oplus_{X_0} X_2$, defined as $f = g_i \circ f_i$ (which is independent of $i \in \{1, 2\}$ by construction).

If X_1, X_2 are topological spaces, $X_1 \oplus_{X_0} X_2$ is given the quotient topology coming from the direct sum topology on $X_1 \oplus X_2$.

Also this construction has a universal property:

Proposition 6.6.2 *Let $X_1 \overset{f_1}{\leftarrow} X_0 \overset{f_2}{\rightarrow} X_2$ be as in Definition 6.6.1. Let Y be a set and $h_i : X_i \to Y$, $i = 1, 2$ functions such that $h_1 \circ f_1 = h_2 \circ f_2$.*

$$\begin{tikzcd}
 & X_1 \arrow[dr, "\iota_1"] \arrow[drr, "g_1"] \arrow[drrr, "h_1"] & & & \\
X_0 \arrow[ur, "f_1"] \arrow[dr, "f_2"'] & & X_1 \oplus X_2 \arrow[r, "p"] & X_1 \oplus_{X_0} X_2 \arrow[r, dashed, "h"] & Y \\
 & X_2 \arrow[ur, "\iota_2"] \arrow[urr, "g_2"] \arrow[urrr, "h_2"'] & & &
\end{tikzcd} \tag{6.8}$$

Then there is a unique function $h : X_1 \oplus_{X_0} X_2 \to Y$ such that $h \circ g_i = h_i$, $i = 1, 2$.

If Y, X_1, X_2 have topologies and h_1, h_2 are continuous then $h : X_1 \oplus_{X_0} X_2 \to Y$ is continuous. If in addition the $f_i : X_0 \to X_i$ are continuous then so is $f : X_0 \to X_1 \oplus_{X_0} X_2$.

Proof. By the universal property of direct sums of sets (Proposition A.2.5) there is a unique function $h_0 : X_1 \oplus X_2 \to Y$ such that $h_0 \circ \iota_i = h_i$, $i = 1, 2$. The only non-trivial equivalences w.r.t. $\sim$ are $\iota_1(f_1(x)) \sim \iota_2(f_2(x))$ for each $x \in X_0$. The assumption $h_1 \circ f_1 = h_2 \circ f_2$ implies that $h_0 : X_1 \oplus X_2 \to Y$ is constant on the $\sim$-equivalence classes, so that the universal property of the quotient operation (Lemma A.1.11) implies the existence of a unique map $h : X_1 \oplus_{X_0} X_2 \equiv (X_1 \oplus X_2)/\!\sim \to Y$ such that $h \circ p = h_0$. If Y, X_1, X_2 are topological spaces and the h_i continuous, Proposition 6.3.5 gives continuity of h_0, and Proposition 6.4.8 gives the continuity of h. If furthermore $f_i : X_0 \to X_i$ for $i \in \{1, 2\}$ is continuous then also $f = p \circ \iota_i \circ f_i : X_0 \to X_1 \oplus_{X_0} X_2$ is continuous. ■

Remark 6.6.3 1. If X_0 is a fixed topological space, a space below X_0 is a topological space X equipped with a continuous map $\iota_X : X_0 \to X$. If $(X, \iota_X), (Y, \iota_Y)$ are spaces below X_0, a map $f : X \to Y$ is a map of spaces below X_0 if $f \circ \iota_X = \iota_Y$. Spaces and maps below X_0 form a category $\mathcal{TOP}^{X_0}$. If $(X_1, f_1), (X_2, f_2)$ are objects in $\mathcal{TOP}^{X_0}$ then $(X_1 \oplus_{X_0} X_2, f)$, as defined above, is an

object in $\mathcal{TOP}^{X_0}$. If now $(Y,k) \in \mathcal{TOP}^{X_0}$ and $h_i : X_i \to Y$ are maps of spaces below X_0 then it follows that $k = h_i \circ f_i$ for $i = 1,2$. This implies $f_1 \circ h_1 = f_2 \circ h_2$, so that by the proposition there is a unique $h : X_1 \oplus_{X_0} X_2 \to Y$. Now also $h \circ f = k$, thus h is a map of spaces below X_0. This proves that $(X_1 \oplus_{X_0} X_2, f)$ is the coproduct of $(X_1, f_1), (X_2, f_2)$ in the category $\mathcal{TOP}^{X_0}$.

2. The construction of the pushout, as well as its universal property, generalize easily to any family $\{f_i : X_0 \to X_i\}_{i\in I}$ of (continuous) maps, giving rise to $\bigoplus_{i\in I} {}_{X_0} X_i$. □

We now consider some applications of the pushout:

Exercise 6.6.4 Let X be a topological space and $A, B \subseteq X$ open subsets such that $X = A\cup B$. Give A, B the subspace topologies. Prove that the pushout $A \oplus_{A\cap B} B$ for the diagram $A \leftarrow A \cap B \to B$, where the arrows are the inclusion maps, is homeomorphic to X.

Exercise 6.6.5 Let X be a Hausdorff space and $A \subseteq X$ a closed subset. Prove that the pushout $X \oplus_A X$ for the diagram $X \leftarrow A \to X$, where the arrows are the obvious inclusion maps, is Hausdorff.

Remark 6.6.6 Let X, Y be Hausdorff spaces and $f \in C(X,Y)$. Assume that f is 'an epimorphism in the category of Hausdorff spaces', meaning that if also Z is Hausdorff and $g, h \in C(Y,Z)$ satisfy $g \circ f = h \circ f$ then $g = h$. Let $A = \overline{f(X)} \subseteq Y$ and define $Z = Y \oplus_A Y$, which is T_2 by Exercise 6.6.5. If $p : Y \oplus Y \to Y \oplus_A Y$ is the quotient map, let $\widetilde{\iota}_{1/2} = p \circ \iota_{1/2} : Y \to Y \oplus_A Y$. By construction, $\widetilde{\iota}_1$ and $\widetilde{\iota}_2$ coincide on A, thus $\widetilde{\iota}_1 \circ f = \widetilde{\iota}_2 \circ f$. Since f is epi, we have $\widetilde{\iota}_1 = \widetilde{\iota}_2$. This clearly implies $A = Y$, to wit $\overline{f(X)} = Y$.

Thus epimorphisms in the category of Hausdorff spaces have dense image, which is the promised converse of Remark 5.2.17(ii). □

Definition 6.6.7 *Let X, Z be topological spaces, $Y \subseteq X$ a subspace and $f : Y \to Z$ a continuous map. Then we write $X \oplus_f Z$ (instead of $X \oplus_Y Z$) for the pushout for the diagram $X \hookleftarrow Y \stackrel{f}{\to} Z$, where the first arrow is just the inclusion map. This construction is called 'attaching X to Z along $f : Y \to Z$'.*

Remark 6.6.8 Note that this construction identifies each $y \in Y \subseteq X$ with $f(y) \in Z$. Thus if f is not injective, certain points of Y become identified in $X \oplus_f Z$. The attachment construction plays an important rôle in algebraic topology, where it is often used to 'attach an n-cell'. In this case $X = D^n = \{x \in \mathbb{R}^n \mid \|x\|_2 \leq 1\}$ and $Y = \partial D^n = S^{n-1}$. □

Dualizing the construction of the pushout as a quotient of a direct sum, one obtains the 'pullback' as a subspace of the direct product.

Definition 6.6.9 *Let X_0, X_1, X_2 be sets and $f_i : X_i \to X_0$, $i = 1,2$ functions. The pullback or fiber product $X_1 \times_{X_0} X_2$ is defined as*

$$X_1 \times_{X_0} X_2 = \{(x_1, x_2) \in X_1 \times X_2 \mid f_1(x_1) = f_2(x_2)\}.$$

It comes with canonical maps $q_i : X_1 \times_{X_0} X_2 \to X_i$, $q_i : (x_1, x_2) \mapsto x_i$, $i = 1,2$. There also is a function $p : X_1 \times_{X_0} X_2 \to X_0$, given by $p = f_i \circ q_i$, which is independent of $i \in \{1,2\}$.

If X_1, X_2 are topological spaces, also $X_1 \times_{X_0} X_2$ is given the subspace topology arising from the product topology on $X_1 \times X_2$.

Lemma 6.6.10 *Let* $X_1 \xleftarrow{f_1} X_0 \xrightarrow{f_2} X_2$ *as in Definition 6.6.9. Let* $h_i : Y \to X_i,\ i = 1, 2$ *be functions such that* $f_1 \circ h_1 = f_2 \circ h_2$.

$$\begin{tikzcd}
 & & & X_1 \arrow[dr, "f_1"] & \\
Y \arrow[r, dashed, "h"] \arrow[urrr, "h_1"] \arrow[drrr, "h_2"'] & X_1 \times_{X_0} X_2 \arrow[r, "\iota"] \arrow[urr, "q_1"] \arrow[drr, "q_2"'] & X_1 \times X_2 \arrow[ur, "p_1"] \arrow[dr, "p_2"'] & & X_0 \\
 & & & X_2 \arrow[ur, "f_2"'] &
\end{tikzcd} \tag{6.9}$$

Then there is a unique function $h : Y \to X_1 \times_{X_0} X_2$ *such that* $q_i \circ h = h_1,\ i = 1, 2$.

If Y, X_1, X_2 *have topologies and* $h_1, h_2 : Y \to X_i$ *are continuous then* $h : Y \to X_1 \times_{X_0} X_2$ *is continuous. If in addition* $f_i : X_i \to X_0$ *is continuous for* $i \in \{i, 2\}$ *then also* $p = f_i \circ q_i$ *is continuous (since* q_i *is continuous by construction).*

Proof. Since q_i is just the restriction of p_i to $X_1 \times_{X_0} X_2 \subseteq X_1 \times X_2$, a map $h : Y \to X_1 \times_{X_0} X_2$ satisfies $q_i \circ h = h_i$ for $i = 1, 2$ if and only if it satisfies $p_i \circ h = h_i$ for $i = 1, 2$. But this forces the definition $h : Y \to X_1 \times X_2,\ y \mapsto (h_1(y), h_2(y))$. (Compare Proposition A.2.7.) In view of $f_1(h_1(y)) = f_2(h_2(y))$, we indeed have $h(y) \in X_1 \times_{X_0} X_2$.

If the h_i are continuous then $h : Y \to X_1 \times X_2, y \mapsto (h_1(y), h_2(y))$ is continuous by Proposition 6.5.2. Now Corollary 6.2.2 gives that h is continuous as a map $Y \to X_1 \times_{X_0} X_2$. ■

Remark 6.6.11 1. If X_0 is a fixed topological space, a space above X_0 is a topological space X equipped with a continuous map $p_X : X \to X_0$. If $(X, p_X), (Y, p_Y)$ are spaces above X_0, a map $f : X \to Y$ is a map of spaces above X_0 if $p_Y \circ f = p_X$. Spaces and maps above X_0 form a category $\mathcal{TOP}_{X_0}$. Arguing as for the pushout, one finds that the pullback $(X_1 \times_{X_0} X_2, p)$ is the direct product of $(X_1, p_1), (X_2, p_2)$ in the category $\mathcal{TOP}_{X_0}$.

2. Given any number of spaces X_i equipped with maps $f_i : X_i \to X_0$, the above construction generalizes straightforwardly to a fiber product $\prod_{X_0} X_i$. □

Part II:

Covering and Separation axioms (beyond T_2) [9]

[9] This author agrees with Ioan James [130]: "Most accounts of the theory go on to discuss separation axioms, [...]. But in my view compactness should come first, because it is of fundamental importance."

Chapter 7

Compactness and related notions

By well-known classical results from basic analysis, every continuous function $f : [a, b] \to \mathbb{R}$ is uniformly continuous (Heine) and bounded, and it assumes its infimum and supremum. In introductory textbooks like [88, 223], these results are deduced from the Bolzano-Weierstrass theorem according to which every bounded sequence in $\mathbb{R}$ has a convergent subsequence. Later we will say '$[a, b]$ is sequentially compact'. As we know, sequences are insuffiicent for studying general topological spaces, which is why sequential compactness is not the best notion and we will focus on the 'better' notion of compactness. Compactness actually is the most important property that a space can have. It has many applications to analysis and functional analysis, cf. [115]. But also many topologies arising in purely algebraic contexts are compact, like the Krull topology on Galois groups, the Zariski topologies (cf. Appendix B) and the topology on the Stone dual (Section 11.1.11) of a Boolean algebra. For this reason, compactness and its many relatives and generalizations merit a very thorough study, which is why this is the longest chapter in this text.[1]

7.1 Covers. Subcovers. Lindelöf and compact spaces

Definition 7.1.1 *Let (X, τ) be a topological space.*

- *A <u>cover</u> of X is a family $\mathcal{U} \subseteq P(X)$ of subsets of X such that $\bigcup \mathcal{U} = \bigcup_{U \in \mathcal{U}} U = X$.*
- *A cover is called open (resp. closed) if every $U \in \mathcal{U}$ is open (resp. closed).*
- *A <u>subcover</u> of a cover $\mathcal{U}$ is a subfamily $\mathcal{V} \subseteq \mathcal{U}$ such that still $\bigcup \mathcal{V} = X$.*

Remark 7.1.2 More explicitly (and tediously) a cover is a family $\mathcal{U} = \{U_i \mid i \in I\}$ such that $\bigcup_{i \in I} U_i = X$, and if $J \subseteq I$ is such that $\bigcup_{i \in J} U_i = X$, then $\mathcal{V} = \{U_i \mid i \in J\}$ is a subcover. (Indexing by a set I gives the additional liberty of having $U_i = U_j$ for some $i \neq j$, but this is never needed.) □

Definition 7.1.3 *A topological space is called*

- *<u>compact</u> if every open cover has a finite subcover.*
- *<u>Lindelöf</u>[2] if every open cover has a countable subcover.*

---

[1]The impatient reader might want to look at Proposition 7.4.11 for an important application of compactness and at Theorems 7.7.51, 7.7.63 or Exercise 7.8.71 for examples of results whose statements do not involve compactness but whose proofs do.

[2]Ernst Leonard Lindelöf (1870-1946), Finnish mathematician. He actually proved that $\mathbb{R}$ is 'Lindelöf'.

Remark 7.1.4 1. Obviously, every compact space is Lindelöf.

2. Compactness does, of course, not just mean that (X, τ) admits a finite open cover. This would be trivial since every space (X, τ) has the finite open cover $\{X\}$.

3. Like second countability and separability, the Lindelöf property can be generalized by replacing the cardinal number $\aleph_0 = \#\mathbb{N}$ by any other (infinite) cardinal $\aleph$, but we stick to $\aleph_0$.

4. Some authors include the Hausdorff property in the definition of compact spaces and call our compact spaces 'quasi-compact'. Other authors, in particular in the older literature, call our compact Hausdorff spaces 'bicompact' (and use 'compact' for our 'countably compact'). Similarly, some authors include the T_2 or T_3-axiom in the definition of Lindelöf spaces. □

We begin with some very easy results and examples of compact and Lindelöf spaces:

Exercise 7.1.5 Let (X, τ) be a topological space and $\mathcal{U}$ an open cover of X. Prove:

(i) There is a subcover $\mathcal{V} \subseteq \mathcal{U}$ such that $\#\mathcal{V} \leq \#X$.

(ii) There is a subcover $\mathcal{V} \subseteq \mathcal{U}$ such that $\#\mathcal{V} \leq \#\mathcal{B}$, where $\mathcal{B}$ is any base for τ.

Exercise 7.1.6 Let (X, τ) be a topological space. Prove:

(i) Every indiscrete space is compact.

(ii) If τ is finite (resp. countable) then it is compact (resp. Lindelöf).

(iii) If X is finite (resp. countable) then τ is compact (resp. Lindelöf).

(iv) A second countable space is hereditarily Lindelöf.

(v) A discrete space is compact (resp. Lindelöf) if and only if it is finite (resp. countable).

(vi) The cofinite (resp. cocountable) topology on any set is compact (resp. Lindelöf).

(vii) The Euclidean and Sorgenfrey topologies on $\mathbb{R}$ are non-compact.

Remark 7.1.7 1. The discrete topology on a countably infinite set provides an example of space that is Lindelöf, but not compact.

2. By Exercise 7.1.6(iv), the Lindelöf property is weaker than second countability, and much of its usefulness derives from the fact that many (but not all!) results that hold for second countable spaces generalize to Lindelöf spaces. It will also play a rôle in the discussion of compactness of metric spaces.

3. There are many spaces that are Lindelöf, but not second countable! Examples are provided by: (a) The cocountable topology on an uncountable set is Lindelöf by Exercise 7.1.6(vi), but not second countable by Exercise 4.1.16(vi). (b) Countable spaces are Lindelöf by Exercise 7.1.6(ii)) but can fail to be second countable, cf. Exercise 4.3.14. (c) Cubes $[0,1]^\chi$, where χ is an uncountable cardinal number, are not second countable by Exercise 6.5.15(iii). But as a consequence of Tychonov's theorem, proven later, such a space is even compact, thus a fortiori Lindelöf. (d) The Sorgenfrey line is another example, cf. Exercise 7.1.8.

4. When $\chi > \mathfrak{c}$, the space in the above example (c) is also non-separable, proving Lindelöf $\not\Rightarrow$ separable. That separable $\not\Rightarrow$ Lindelöf is demonstrated by the Sorgenfrey plane $(\mathbb{R}, \tau_S)^2$, which clearly is separable but not Lindelöf by Corollary 8.1.39. □

Exercise 7.1.8 Goal: The Sorgenfrey line $(\mathbb{R}, \tau_S)$ is Lindelöf.

(i) Show that for proving the Lindelöf property it is sufficient to consider covers by open sets of the form $[a, b)$, where $a < b$.

(ii) Given an open cover $\mathcal{U} = \{[a_i, b_i)\}_{i \in I}$, let $Y = \bigcup_i (a_i, b_i)$. Prove that $\mathbb{R} \backslash Y$ is at most countable.

(iii) Show that every subset of $\mathbb{R}$ with the usual topology is Lindelöf.

(iv) Combine this to obtain a countable subcover of $\mathcal{U}$. (Thus $(\mathbb{R}, \tau_S)$ is Lindelöf.)

As we have seen above and in Section 4.1.2, we have the implications

$$\text{compact} \quad \Rightarrow \quad \text{Lindelöf} \quad \Leftarrow \quad \text{second countable} \quad \Rightarrow \quad \text{separable} \quad \Rightarrow \quad \text{Souslin},$$

none of which is invertible in general. For metrizable spaces, however, the situation simplifies since the last four properties become equivalent (but compactness remains stronger):

Exercise 7.1.9 Prove that a metrizable Lindelöf space is separable and second countable.

The following result beautifully applies the Lindelöf property without even having it in its statement:

Proposition 7.1.10 *If a space (X, τ) is second countable, i.e. τ admits a countable base $\mathcal{B}$, then* **every** *base $\mathcal{V}$ for τ has a countable subfamily $\mathcal{V}_0 \subseteq \mathcal{V}$ that still is a base.*

Proof. Let $\mathcal{U} = \{U_1, U_2, \ldots\}$ be a countable base and $\mathcal{V}$ any base. For $i \in \mathbb{N}$, define $\mathcal{B}_i = \{V \in \mathcal{V} \mid V \subseteq U_i\} \subseteq \mathcal{V}$. Since $\mathcal{V}$ is a base, we have $\bigcup \mathcal{B}_i = U_i$. Now, the subspaces $U_i \subseteq X$ are second countable, thus Lindelöf. Thus the open cover $\mathcal{B}_i$ of U_i has a countable subcover $\mathcal{B}_i^0 \subseteq \mathcal{B}_i$. Now define $\mathcal{V}_0 = \bigcup_{i=1}^{\infty} \mathcal{B}_i^0$. As a countable union of countable sets, this is a countable subfamily of $\mathcal{V}$. If $W \subseteq X$ is open, it is a union of the U_i contained in W, and each such U_i is a union over a subfamily of $\mathcal{B}_i^0 \subseteq \mathcal{V}_0$. Thus $\mathcal{V}_0 \subseteq \mathcal{V}$ is a base for τ. ∎

7.2 Compact spaces: Equivalent characterizations

The Lindelöf property essentially just is another countability property like second countability, separability and the Souslin property (all of which are equivalent for metric spaces, as we have seen). We will see that compactness has a very different character.

The property of compactness has a long and complicated history, cf. [194] or the historical notes in [237, 75]. Before the 'right' (i.e. most useful and best behaved) definition was arrived at, mathematicians studied various other notions, like 'sequential compactness' or '(weak) countable compactness'. We will have a quick look at these alternative notions in Section 7.7, where in particular we will see that they are all equivalent to compactness for metric spaces. But for more general spaces, this is not the case, and compactness clearly is the most important notion. (Later we will meet quite a few generalizations of compactness: local compactness, compact generation, σ-compactness, hemicompactness, paracompactness, ...)

In this subsection, we consider two important equivalent characterizations of compactness in terms of closed sets and nets.

Definition 7.2.1 *A family $\mathcal{F}$ of subsets of a set X has the finite intersection property if any intersection of finitely many elements of $\mathcal{F}$ is non-empty.*

Lemma 7.2.2 *For a topological space (X, τ), the following are equivalent:*

(i) (X, τ) *is compact.*

(ii) Whenever $\mathcal{F} \subseteq P(X)$ *is a family of closed subsets of* X *such that* $\bigcap \mathcal{F} = \emptyset$ *then there are* $C_1, \ldots, C_n \in \mathcal{F}$ *such that* $C_1 \cap \cdots \cap C_n = \emptyset$.

(iii) Whenever $\mathcal{F} \subseteq P(X)$ *consists of closed subsets of* X *and has the finite intersection property then* $\bigcap \mathcal{F} \neq \emptyset$.

Proof. (i) and (ii) are dualizations of each other, using de Morgan's formulas, and (iii) is the contraposition of (ii). ■

(Similar equivalences can be given for the Lindelöf property, but we will not need them.)

The reader probably knows from analysis courses that a metric space (X, d) is compact if and only if every sequence in X has an accumulation point or, equivalently, a convergent subsequence. A similar result holds for topological spaces, but we must replace sequences by nets:

Proposition 7.2.3 *A topological space* (X, τ) *is compact if and only if every net in* X *has an accumulation point (equivalently, a convergent subnet).*

Proof. $\Leftarrow$: Assume X is not compact, thus there is an open cover $\mathcal{U} = \{U_i\}_{i \in I}$ admitting no finite subcover. The set $\mathcal{J} = \{J \subseteq I \mid J \text{ finite}\}$ is partially ordered by inclusion and upward directed. By assumption, $X_J := X \backslash \bigcup_{j \in J} U_j \neq \emptyset \ \forall J \in \mathcal{J}$, thus (invoking AC) we can choose points $x_J \in X_J$. Now $(\mathcal{J}, \leq) \to X, \ J \mapsto x_J$ is a net. By assumption, the net $\{x_J\}$ has an accumulation point x. Since $\mathcal{U}$ is a cover, there is $i \in I$ such that $x \in U_i$. Clearly $\{i\} \in \mathcal{J}$, and if $J \geq \{i\}$ then $x_J \notin U_i$ by construction. Thus, whenever $K \geq \{i\}$, we have $x_L \notin X_L \ \forall L \geq K$. Thus x is not frequently in U_i, contradicting the fact that x is an accumulation point.

$\Rightarrow$: If $\{x_\iota\}_{\iota \in I}$ has no accumulation point, then every $x \in X$ has a neighborhood U_x such that x_ι is not frequently in U_x, thus by Exercise 5.1.27 x_ι eventually is in $X \backslash U_x$. Thus there is ι_x such that $\iota \geq \iota_x \ \Rightarrow \ x \notin U_x$. Now $\{U_x\}_{x \in X}$ is an open cover of X, and by compactness there is a finite subcover $\{U_{x_1}, \ldots, U_{x_n}\}$. Using the directedness axiom, we can find $\lambda \in I$ such that $\lambda \geq \iota_{x_k}$ for $k = 1, \ldots, n$. Now $\iota \geq \lambda$ implies $x_\iota \notin \bigcup_{k=1}^n U_{x_k} = X$, which is absurd. ■

Whether a space is compact can be checked by looking only at covers consisting of base elements:

Exercise 7.2.4 Let (X, τ) be a topological space and $\mathcal{B} \subseteq \tau$ a base for the topology. Prove that X is compact if and only if every cover of X by elements of $\mathcal{B}$ has a finite subcover.

The result of the above exercise is not very useful, but the following result definitely is:

Lemma 7.2.5 (Alexander's Subbase Lemma) [3] *Let* (X, τ) *be a topological space and* $\mathcal{S} \subseteq \tau$ *a subbase for* τ. *If every cover of* X *by elements of* $\mathcal{S}$ *has a finite subcover, then* X *is compact.*

Proof. Assuming that X is non-compact, the family $\mathfrak{F}$ of open covers $\mathcal{V}$ that have no finite subcover is non-empty. Partially ordering $\mathfrak{F}$ by inclusion, let $\mathfrak{G}$ be a totally ordered subset of $\mathfrak{F}$. Then $\mathcal{W} = \bigcup \mathfrak{G} = \bigcup_{\mathcal{V} \in \mathfrak{G}} \mathcal{V}$ clearly is an open cover of X. If $U_1, \ldots, U_n \in \mathcal{W}$ then each U_i comes from some $\mathcal{V}_i \in \mathfrak{G}$. Since $\mathfrak{G}$ is totally ordered, we have $\{U_1, \ldots, U_n\} \subseteq \max(\mathcal{V}_1, \ldots, V_n) = \overline{\mathcal{V}} \in \mathfrak{G}$. Since $\overline{\mathcal{V}}$ has no finite subcover, the same holds for $\mathcal{W}$. Thus $\mathcal{W} \in \mathfrak{F}$, and it is an upper bound for the chain $\mathfrak{G}$. Thus the assumptions of Zorn's Lemma are satisfied, so that $\mathfrak{F}$ has a maximal element $\mathcal{M}$. Thus $\mathcal{M}$ is an open cover without finite subcover, and maximality of $\mathcal{M}$ implies that $\mathcal{M} \cup \{V\}$ does have a finite subcover for every non-empty open $V \notin \mathcal{M}$.

[3] James Waddell Alexander II (1888-1971), American topologist (general and algebraic).

Defining $\mathcal{S}' = \mathcal{M} \cap \mathcal{S}$, no finite subfamily of $\mathcal{S}'$ covers X, since this is true for $\mathcal{M} \in \mathfrak{G}$. Thus if we prove that $\mathcal{S}'$ covers X, we have arrived at a contradiction with the hypothesis.

To this end, assume $x \in X \backslash \bigcup \mathcal{S}'$. Since $\mathcal{M}$ is a cover, there is a $U \in \mathcal{M}$ such that $x \in U$. Since $\mathcal{S}$ is a subbase, there are $V_1, \ldots, V_n \in \mathcal{S}$ such that $x \in \bigcap_{i=1}^n V_i \subseteq U$. In view of $x \notin \bigcup \mathcal{S}'$, we have $V_i \notin \mathcal{S}'$ for each $i = 1, \ldots, n$. Thus for each i the open cover $\mathcal{M} \cup \{V_i\}$ has a finite subcover, so that $X = Y_i \cup V_i$, where Y_i is a finite union of elements of $\mathcal{M}$. In view of $\bigcap_i V_i \subseteq U$ we have

$$X = Y_1 \cup \cdots \cup Y_n \cup \bigcap_{i=1}^n V_i \subseteq Y_1 \cup \cdots \cup Y_n \cup U.$$

Since $U \in \mathcal{M}$ and each Y_i is a finite union of elements of $\mathcal{M}$, we find that X is a finite union of elements of $\mathcal{M}$. But this contradicts the fact that $\mathcal{M}$ by construction has no finite subcover. This contradiction proves $\bigcup \mathcal{S}' = X$, and therefore the lemma. ■

Remark 7.2.6 The subbase lemma is often useful when topologies are defined in terms of subbases, as is the case with initial topologies, cf. Section 6.1, in particular Lemma 6.1.6. The most important classes of examples are given by product spaces and spaces with the order topology, discussed in Section 7.6. Lemma 7.2.5 will be used to prove Tychonov's Theorem 7.5.9 on the compactness of product spaces and a criterion for compactness of spaces with order topology, cf. Theorem 7.6.2. □

Exercise 7.2.7 Prove: A topological space (X, τ) is compact if and only if every infinite subset $Y \subseteq X$ has a complete accumulation point. (The $\Rightarrow$ direction requires some easy cardinal number arithmetic. The $\Leftarrow$ direction is harder and usually done using ordinal numbers.)

In the next section we will study in the behavior of compactness and the Lindelöf property under various constructions to the extent that the two properties behave similarly, before we turn to results for compact spaces that have no analogue for Lindelöf spaces. On the way, we will encounter many useful applications of compactness. The usefulness of compactness raises the question whether a non-compact space can be compactified, in analogy to completion of a metric space. They can, though not uniquely in contrast to completions (Proposition 3.2.2). Compactifications will be considered in Sections 7.8 and 8.3.3.

7.3 Behavior of compactness and Lindelöf property under constructions

We now study the behavior of Lindelöf property and compactness under passage to subspaces, direct sums and under continuous functions. (Products will be considered later.)

The following construction is similar to that of Exercise 4.1.13 (but different!):

Exercise 7.3.1 Let (X, τ) be any topological space. Let $X' = X \cup \{p\}$ (where $p \notin X$) and $\tau' = \tau \cup \{X'\}$. (Thus the open sets of (X', τ') are the open $U \subseteq X$, considered as subsets of X', and the total space X'.) Prove:

(i) τ' is a topology, and it is compact (thus Lindelöf).

(ii) $\tau' \upharpoonright X = \tau$.

(iii) Conclude that neither compactness nor the Lindelöf property are hereditary.

(As to a Hausdorff example, we will see that $[0,1]$ is compact, but $(0,1) \cong \mathbb{R}$ is not.)

Given a subset $Y \subseteq (X, \tau)$, equipped with the subspace topology τ_Y, the question whether (Y, τ_Y) is compact or Lindelöf can be formulated without reference to τ_Y:

Lemma 7.3.2 *Let (X, τ) be a topological space and $Y \subseteq X$. Then the subspace $(Y, \tau_Y) \subseteq (X, \tau)$ is compact (resp. Lindelöf) if and only if every family $\{U_i \in \tau\}_{i \in I}$ such that $\bigcup_i U_i \supseteq Y$ has a finite (resp. countable) subfamily whose union still contains Y.*

Proof. If $\{U_i \in \tau\}_{i \in I}$ is a family of open sets such that $\bigcup_i U_i \supseteq Y$, then $\{V_i = U_i \cap Y \in \tau_Y\}_{i \in I}$ is an open cover of Y. Conversely, every open cover $\{V_i\}$ of Y arises in this way (usually non-uniquely). If $U_i \subseteq X$ and $V_i \subseteq Y$ are related as above and $J \subseteq I$ then it is clear that $\bigcup_{j \in J} U_j \supseteq Y$ holds if and only if $\{V_j\}_{j \in J}$ is a subcover of $\{V_i\}$. ■

Remark 7.3.3 This result has a dual version which is occasionally useful:

If $U \subseteq X$ is open such that $X \backslash U$ is compact, and $\{F_i\}_{i \in I}$ is a family of closed sets such that $\bigcap_{i \in I} F_i \subseteq U$, then there is a finite subset $J \subseteq I$ such that $\bigcap_{j \in J} F_j \subseteq U$. □

Lemma 7.3.4 *If (X, τ) is compact (resp. Lindelöf) and $Y \subseteq X$ is closed then (Y, τ_Y) is compact (resp. Lindelöf).*

Proof. Let $\{U_i \in \tau\}_{i \in I}$ be such that $\bigcup_i U_i \supseteq Y$. Since Y is closed, $U_0 = X \backslash Y$ is open, and $\{U_0\} \cup \{U_i \mid i \in I\}$ is an open cover of X. Since X is compact (resp. Lindelöf), there is a finite (resp. countable) subset $J \subseteq I$ such that $\{U_0\} \cup \{U_j \mid j \in J\}$ still covers X. But this means that $\bigcup_{j \in J} U_j \supseteq Y$, thus Y is compact (resp. Lindelöf) by Lemma 7.3.2. ■

Exercise 7.3.5 Prove that a direct sum $\bigoplus_{i \in I}(X_i, \tau_i)$ of topological spaces is

(i) compact if and only if each X_i is compact and the set $\{i \in I \mid X_i \neq \emptyset\}$ is finite.

(ii) Lindelöf if and only if each X_i is Lindelöf and the set $\{i \in I \mid X_i \neq \emptyset\}$ is countable.

Note: The two proofs are very similar, thus it suffices to write down one and indicate how the other differs.

Lemma 7.3.6 *If (X, τ) is compact (resp. Lindelöf) and $f : (X, \tau) \to (Y, \sigma)$ is continuous then the subspace $f(X) \subseteq Y$ is compact (resp. Lindelöf).*

Proof. Thus let $\{U_i\}_{i \in I}$ be open sets in Y such that $\bigcup_i U_i \supseteq f(X)$. By continuity, each $f^{-1}(U_i) \subseteq X$ is open. Since the U_i cover $f(X)$, the family $\{f^{-1}(U_i)\}$ is an open cover of X. By compactness (resp. the Lindelöf property) of X there is a finite (resp. countable) set $J \subseteq I$ such that $\{f^{-1}(U_j)\}_{j \in J}$ still covers X. This is equivalent to $\bigcup_{j \in J} U_j \supseteq f(X)$, thus the subspace $f(X) \subseteq Y$ is compact (resp. Lindelöf) by Lemma 7.3.2. ■

Corollary 7.3.7 *(i) If X is compact (resp. Lindelöf) and $f : X \to Y$ is continuous and surjective then Y is compact (resp. Lindelöf).*

(ii) Quotient spaces of compact (resp. Lindelöf) spaces are compact (resp. Lindelöf).

Actually, surjective continuous images of compact spaces have a tendency to be quotient spaces! Cf. Proposition 7.4.11(iii).

In Section 7.5 we will see that products of arbitrarily many compact spaces are compact.

On the other hand, in Corollary 8.1.39(iii) we will find that the square $(\mathbb{R}, \tau_S)^2$ of the Sorgenfrey line is not Lindelöf, thus not even products of two Lindelöf spaces need to be Lindelöf.

Exercise 7.3.8 Let $\{x_n\}_{n\in\mathbb{N}}$ be a sequence in a topological space X such that $x_n \to x \in X$. Prove that $Y = \{x, x_1, x_2, \ldots\} \subseteq X$ is compact.

7.4 More on compactness

In the rest of this Section we focus on results that really require compactness (not just Lindelöf).

7.4.1 More on compactness and subspaces

Lemma 7.4.1 *If (X, τ) is Hausdorff, $Y \subseteq X$ is a compact subspace and $x \in X\backslash Y$ then there are open U, V such that $Y \subseteq U$, $x \in V$ and $U \cap V = \emptyset$.*

Proof. For every $y \in Y$ we have $y \neq x$, thus using the Hausdorff property we can find $U_y, V_y \in \tau$ such that $y \in U_y$, $x \in V_y$ and $U_y \cap V_y = \emptyset$. In view of $y \in U_y$, we have $\bigcup_y U_y \supseteq Y$. By compactness of Y and Lemma 7.3.2 there is a finite subset $\{y_1, \ldots, y_n\} \subseteq Y$ such that $U := U_{y_1} \cup \cdots \cup U_{y_n} \supseteq Y$. Now $V := V_{y_1} \cap \cdots \cap V_{y_n}$ is an open neighborhood of x, and

$$V \cap U = (V_{y_1} \cap \cdots \cap V_{y_n}) \cap (U_{y_1} \cup \cdots \cup U_{y_n}) = \bigcup_{k=1}^{n} (V_{y_1} \cap \cdots \cap V_{y_n} \cap U_{y_k}) = \emptyset$$

due to $U_{y_k} \cap V_{y_k} = \emptyset$. ■

We have already proven that closed subspaces of compact spaces are compact (and similarly for Lindelöf spaces). For Hausdorff spaces and compactness, there is a converse:

Lemma 7.4.2 *If (X, τ) is Hausdorff and $Y \subseteq X$ is compact (with the relative topology) then $Y \subseteq X$ is closed.*

Proof. By Lemma 7.4.1, every $x \in X\backslash Y$ has an open neighborhood V contained in $X\backslash Y$. Thus $X\backslash Y$ is open, thus Y is closed. ■

Corollary 7.4.3 *A subspace of a compact Hausdorff space is compact if and only if it is closed.*

Remark 7.4.4 1. If X has the indiscrete topology, then every subspace $Y \subseteq X$ is indiscrete, thus compact, but $\overline{Y} = X$ whenever $Y \neq \emptyset$. Thus the T_2 assumption in Lemma 7.4.2 is necessary.

2. Lemmas 7.3.4 and 7.4.2 should be compared to Lemma 3.1.7 concerning completeness of subspaces. For a (less perfect) analogue of Proposition 3.2.7 see Corollary 8.3.22.

3. A subset of a topological space is called <u>relatively compact</u> if its closure is compact.

4. Lemma 7.4.2 and the needs of algebraic topology motivate the definition of several closely related notions of <u>weak Hausdorff</u> spaces:

(WH1) Every compact subspace $Y \subseteq X$ is closed.

(WH1') If Y is compact and $f : Y \to X$ is continuous then $f(Y) \subseteq X$ is closed.

(WH2) If Y is compact Hausdorff and $f : Y \to X$ is continuous then $f(Y) \subseteq X$ is closed.

We then have the following implications: $T_2 \Rightarrow$ WH1 $\Leftrightarrow$ WH1' $\Rightarrow$ WH2 $\Rightarrow T_1$. (The first two $\Rightarrow$ follow from Lemmas 7.4.2 and 7.3.6, respectively. For WH1'$\Rightarrow$WH1, consider the embedding map $f : Y \hookrightarrow X$.) For more on these notions and their relevance in algebraic topology see [166, 121]. □

Exercise 7.4.5 Prove that disjoint compact subsets C, D in a Hausdorff space have disjoint open neighborhoods $U \supseteq C,\ V \supseteq D$. Hint: Combine the result of Lemma 7.4.1 with the method of its proof.

Remark 7.4.6 Later we will call a T_1-space X regular or T_3 if for every closed $C \subseteq X$ and $x \in X \backslash C$ then there are disjoint open $U, V \subseteq X$ such that $Y \subseteq U,\ x \in V$. And a T_1-space will be called normal or T_4 if given disjoint closed C, D there are disjoint open U, V such that $C \subseteq U,\ D \subseteq V$. Since closed subsets of a compact space are compact by Lemma 7.3.4, Lemma 7.4.1 and Exercise 7.4.5 have as corollaries that every compact Hausdorff space is regular and normal. In Section 8 we will study such stronger separation axioms quite extensively. □

Exercise 7.4.7 Let X be a topological space, not necessarily compact.

(i) Prove that any finite union of compact subsets of X is compact.

(ii) Let $K, C \subseteq X$ where K is compact and C is closed. Prove that $K \cap C$ is compact.

(iii) Prove that in a Hausdorff space, every intersection of compact sets is compact.

(iv) Give an example of two compact sets in a non-Hausdorff space whose intersection is not compact.

In view of the meta-Definition 2.3.7, a space X is called hereditarily compact if *all* subspaces $Y \subseteq X$ are compact. Obviously every indiscrete space is hereditarily compact. There are not many hereditarily compact Hausdorff spaces, but more hereditarily compact T_1 spaces:

Exercise 7.4.8 (i) Prove that a Hausdorff space (X, τ) is hereditarily compact if and only if X is finite and τ is the discrete topology.

(ii) Prove that the property of being cofinite is hereditary. Deduce that cofinite spaces (X, τ_{cofin}) are hereditarily compact.

Exercise 7.4.9 Let (X, τ) be a topological space. Prove that the following are equivalent:

(i) X is hereditarily compact.

(ii) For every family $\{U_i \in \tau\}_{i \in I}$ there is a finite subset $J \subseteq I$ such that $\bigcup_{j \in J} U_j = \bigcup_{i \in I} U_i$.

(iii) Every strictly increasing chain $U_1 \subsetneq U_2 \subsetneq \cdots$ of open sets in X is finite.

(iv) Every strictly decreasing chain $C_1 \supsetneq C_2 \supsetneq \cdots$ of closed sets in X is finite.

Remark 7.4.10 Because of the statements (iii) and (iv), hereditarily compact spaces are also called Noetherian. In Section 2.8 we will see that this is more than an analogy with Noetherian rings. For more on hereditarily compact spaces see [219]. □

7.4.2 More on compactness and continuity. Quotients and embeddings

We know that continuous images of compact spaces are compact, and similarly for Lindelöf spaces. For compact spaces, we have more, thanks to Lemma 7.4.2. The following has countless applications:

Proposition 7.4.11 *Let (X, τ) be compact, (Y, σ) Hausdorff and $f : X \to Y$ continuous. Then:*

(i) f is closed. (In particular $f(X) \subseteq Y$ is closed.)

(ii) If f is a bijection then it is a homeomorphism.

(iii) If f is injective then it is an embedding.

(iv) If f is surjective then it is a quotient map. (σ is the final topology on Y induced by f.)

Note that (ii) and (iii) imply that (X, τ) is Hausdorff, which was not assumed!

Proof. (i) If $C \subseteq X$ is closed then it is compact by Lemma 7.3.4. Thus by Lemma 7.3.6, $f(C) \subseteq Y$ is compact, and thus closed by Lemma 7.4.2.

(ii) If f is bijective then (i) with Lemma 5.2.26 implies that it is a homeomorphism.

(iii) It is clear that $f : X \to f(X)$ is a continuous bijection. Since $f(X) \subseteq Y$ inherits the Hausdorff property, (ii) implies that $f : X \to f(X)$ is a homeomorphism.

(iv) f is continuous, surjective, and closed by (i). Now Lemma 6.4.5 gives that σ is the quotient topology, thus f is a quotient map. (For another proof, independent of Lemma 6.4.5, cf. Remark 7.4.15. ■

Corollary 7.4.12 *If X is a compact space, and $\sim$ is an equivalence relation on X such that $X/{\sim}$ is Hausdorff then $\sim$ is closed.*

Remark 7.4.13 In Section 8.1 we will prove a converse of this: If X is compact Hausdorff and $\sim$ is a closed equivalence relation, then the quotient space $X/{\sim}$ is Hausdorff. Thus: If X is compact Hausdorff then $X/{\sim}$ is Hausdorff if and only if $\sim$ is closed! □

Corollary 7.4.14 *Let X be a set and τ_1, τ_2 topologies on X. If τ_1 is compact, τ_2 is Hausdorff and $\tau_1 \supseteq \tau_2$ (i.e. τ_1 is finer than τ_2) then $\tau_1 = \tau_2$.*

In particular: If two compact Hausdorff topologies on the same set are comparable (i.e. $\tau_1 \subseteq \tau_2$ or $\tau_2 \subseteq \tau_1$) then they coincide!

Proof. By Lemma 5.2.25, the map $\mathrm{id}_X : (X, \tau_1) \to (X, \tau_2)$ is continuous. Thus by Proposition 7.4.11(ii) it is a homeomorphism, thus $\tau_1 = \tau_2$. The second statement obviously follows. ■

Remark 7.4.15 The statements (ii)-(iv) of Proposition 7.4.11 can also be interpreted in terms of Corollary 7.4.14. E.g. for (iv) proceed like this: Consider the the quotient (=final) topology $\widetilde{\sigma} = \{U \subseteq Y \mid f^{-1}(U) \in \tau\}$ on Y induced by f. Continuity of f implies $\widetilde{\sigma} \supseteq \sigma$. Since (X, τ) is compact, $\widetilde{\sigma}$ is compact by Lemma 7.3.6. Since σ is Hausdorff, Corollary 7.4.14 now gives $\widetilde{\sigma} = \sigma$. Thus σ is the quotient topology and f is a quotient map. □

Exercise 7.4.16 Let (X, τ) be compact Hausdorff and τ' another topology on X. Prove:

(i) If $\tau' \supsetneq \tau$ then τ' is Hausdorff, but not compact.

(ii) If $\tau' \subsetneq \tau$ then τ' is compact, but not Hausdorff.

The next two subsections give two more advanced applications of compactness and may be skipped until needed.

7.4.3 ⋆ Second countability for images under closed maps

In Exercise 5.2.27 we have seen that the image of a second countable space under a continuous open map is second countable. For non-open maps, such results are much harder to come by. Here is a result for closed maps of compact spaces:

Proposition 7.4.17 *If (X,τ) is compact and second countable, (Y,σ) is T_1 and $f : X \to Y$ is continuous, closed and surjective, then σ is second countable.*

Proof. Let $\mathcal{U}$ be a countable base for τ. Let $\mathcal{S}$ be the family of finite subsets of $\mathcal{U}$. Defining for each $S \in \mathcal{S}$,

$$W_S = Y \backslash f(X \backslash \bigcup_{s \in S} U_s),$$

we have $W_S \in \sigma$ since f is closed. Since $\mathcal{S}$ is countable, our claim will follow once we prove that $\{W_S,\ S \in \mathcal{S}\}$ is a base for σ.

Let $y \in W \in \sigma$. Then $f^{-1}(y) \subseteq f^{-1}(W) \subseteq X$. Since Y is T_1, $\{y\}$ is closed, and continuity of f gives that $f^{-1}(W)$ is open and $f^{-1}(y)$ is closed, thus compact by Lemma 7.3.4. Since $\mathcal{U}$ is a base for τ, we have $f^{-1}(W) = \bigcup\{U \in \mathcal{U} \mid U \subseteq f^{-1}(W)\}$. The U's appearing on the right hand side cover $f^{-1}(x)$, and by compactness of the latter and Lemma 7.3.2, there is a finite family $S \in \mathcal{S}$ such that

$$f^{-1}(y) \subseteq \bigcup_{s \in S} U_s \subseteq f^{-1}(W).$$

This implies

$$f(X \backslash f^{-1}(W)) \subseteq f(X \backslash \bigcup_{s \in S} U_s) \subseteq f(X \backslash f^{-1}(y)),$$

and using the surjectivity of f and the definition of W_S this becomes $Y \backslash W \subseteq Y \backslash W_S \subseteq Y \backslash \{y\}$, or just $y \in W_S \subseteq W$. This means that $\{W_S,\ S \in \mathcal{S}\}$ is a base for σ, and we are done. ■

Remark 7.4.18 1. The result remains true if we replace compactness of X by compactness of $f^{-1}(y)$ for each $y \in Y$, i.e. properness, which will be defined in Section 7.8.5. Clearly, if f is the quotient map arising from an equivalence relation $\sim$, this holds if and only if each equivalence class $[x] \subseteq X$ is compact.

2. Since the family of finite subsets of some infinite set X has the same cardinality as X, essentially the same proof gives the following more general result: If X is compact, Y is T_1 and $f : X \to Y$ is continuous, closed and surjective then $w(Y) \leq w(X)$, where $w(X)$ is the weight of X mentioned in Remark 4.1.6.

3. Preview: We will show later that if (X,τ) is compact Hausdorff then it is 'normal', which is better than T_2, and that normality is preserved under quotients by closed equivalence relations (Proposition 8.1.17). If (X,τ) is also second countable, the above result applies and gives second countability of $X/{\sim}$. We will also show (Corollary 8.2.36) that a compact Hausdorff space is metrizable if and only if it is second countable. Putting all this together we obtain: If X is compact metrizable and $\sim$ is a closed equivalence relation on X, then $X/{\sim}$ is compact metrizable! □

7.4.4 ⋆ Extending continuous maps into compact Hausdorff spaces

Given a continuous function $f : A \to Y$, where $A \subseteq X$ is dense and Y is Hausdorff, we proved in Exercise 6.5.18(iii) that f has at most one continuous extension to X. Concerning the more difficult question of existence, so far we only have Proposition 3.4.10 in the context of metric spaces. But a complete answer exists when the target space Y is compact Hausdorff.

Lemma 7.4.19 *If (X, τ) is compact Hausdorff then:*

(i) If $x \in U \in \tau$ there is an open V such that $x \in V \subseteq \overline{V} \subseteq U$.

(ii) If $x, y \in X$, $x \neq y$ then there are open $U \ni x$, $V \ni y$ such that $\overline{U} \cap \overline{V} = \emptyset$.

Proof. (i) Let $x \in U' \in \tau$. Then $Y = X \backslash U'$ is closed, thus compact. Now by Lemma 7.4.1 there are disjoint open sets U, V such that $x \in U$ and $Y' \subseteq V$. Now $\overline{U} \cap V = \emptyset$, thus $\overline{U} \subseteq U'$.

(ii) Since X is Hausdorff, there are disjoint open sets $U' \ni x$, $V' \ni y$. Now use the preceding argument to find open U, V such that $x \in U \subseteq \overline{U} \subseteq U'$ and $y \in V \subseteq \overline{V} \subseteq V'$. ■

Theorem 7.4.20 *Let $A \subseteq X$ be dense and $f : A \to Y$ continuous, with Y compact Hausdorff. Then there is a continuous extension $\widehat{f} : X \to Y$ (i.e. $\widehat{f} \upharpoonright A = f$) if and only if*

$$C, D \subseteq Y \text{ closed}, \ C \cap D = \emptyset \quad \Rightarrow \quad \overline{f^{-1}(C)} \cap \overline{f^{-1}(D)} = \emptyset. \tag{7.1}$$

(In X, not in A!) If an extension $\widehat{f}$ exists, it is unique.

Proof. Uniqueness of $\widehat{f}$ follows from Exercise 6.5.18(iii). Assume that $\widehat{f}$ exists, and let $C, D \subseteq Y$ be disjoint closed sets. Then $\widehat{f}^{-1}(C), \widehat{f}^{-1}(D) \subseteq X$ are closed and disjoint. Now,

$$\overline{f^{-1}(C)} = \overline{\widehat{f}^{-1}(C) \cap A} \subseteq \widehat{f}^{-1}(C),$$

by closedness of $\widehat{f}^{-1}(C)$. Similarly, $\overline{f^{-1}(D)} \subseteq \widehat{f}^{-1}(D)$, and therefore (7.1) holds.

Now assume (7.1). For $x \in X$ define $\mathcal{F}(x) = \{\overline{f(N \cap A)} \mid N \in \mathcal{N}_x\} \subseteq P(Y)$. If $N_1, \ldots, N_n \in \mathcal{N}_x$ then

$$\overline{f(N_1 \cap A)} \cap \cdots \cap \overline{f(N_n \cap A)} \supseteq \overline{f(N_1 \cap A) \cap \cdots \cap f(N_n \cap A)} \supseteq \overline{f(N_1 \cap \cdots \cap N_n \cap A)} \neq \emptyset, \tag{7.2}$$

since $N = N_1 \cap \cdots \cap N_n \in \mathcal{N}_x$, thus $N \cap A \neq \emptyset$ since $A \subseteq X$ is dense. Thus the family $\mathcal{F}(x)$ has the finite intersection property, and by compactness of Y and Lemma 7.2.2, we have $\bigcap \mathcal{F}(x) = \bigcap_{N \in \mathcal{N}_x} \overline{f(N \cap A)} \neq \emptyset$.

We claim that this intersection contains precisely one point. Assume $y_1 \neq y_2$, $\{y_1, y_2\} \subseteq \bigcap \mathcal{F}(x) \subseteq Y$. By Lemma 7.4.19 we can find open neighborhoods V_1, V_2 of y_1, y_2, respectively, such that $\overline{V_1} \cap \overline{V_2} = \emptyset$. By (7.1), we have $\overline{f^{-1}(V_1)} \cap \overline{f^{-1}(V_2)} \subseteq \overline{f^{-1}(\overline{V_1})} \cap \overline{f^{-1}(\overline{V_2})} = \emptyset$, thus obviously $x \notin \overline{f^{-1}(V_1)} \cap \overline{f^{-1}(V_2)}$. Thus there is $k \in \{1, 2\}$ such that $y \notin \overline{f^{-1}(V_k)}$. This means that $X \backslash \overline{f^{-1}(V_k)}$ is an open neighborhood of x. By definition of $\mathcal{F}(x)$ this implies $\{y_1, y_2\} \subseteq \mathcal{F}(x) \subseteq \overline{f(A \backslash f^{-1}(V_k))} \subseteq Y \backslash V_k$, which contradicts $y_k \in V_k$. Thus $y_1 = y_2$, and $\bigcap \mathcal{F}(x)$ is a singleton for every $x \in X$. This allows us to define $\widehat{f}(x)$ by $\{\widehat{f}(x)\} = \bigcap \mathcal{F}(x)$. For $x \in A$ we obviously have $f(x) \in \bigcap \mathcal{F}(x)$, implying $\widehat{f} \upharpoonright A = f$.

It remains to show that $\widehat{f}$ is continuous. Let $U \subseteq Y$ be an open neighborhood of $\widehat{f}(x)$. Since $\bigcap_{N \in \mathcal{N}_x} \overline{f(N \cap A)} = \{\widehat{f}(x)\} \subseteq U$ and $X \backslash U$ is closed, thus compact, Remark 7.3.3 implies that there are $N_1, \ldots, N_n \in \mathcal{N}_x$ such that $\overline{f(N_1 \cap A)} \cap \cdots \cap \overline{f(N_n \cap A)} \subseteq U$. We may and will assume that the N_i are open. Thus $N = N_1 \cap \cdots \cap N_n$ is an open neighborhood of x, and by (7.2) we have $\overline{f(N \cap A)} \subseteq \overline{f(N_1 \cap A)} \cap \cdots \cap \overline{f(N_n \cap A)} \subseteq U$. Now, for every $x' \in N$, we have $N \in \mathcal{N}_{x'}$, thus $\widehat{f}(x') \in \overline{f(N \cap A)} \subseteq U$. Thus $\widehat{f}(N) \subseteq U$, and $\widehat{f}$ is continuous. ■

Remark 7.4.21 1. Notice the conceptual similarity of this proof to that of Theorem 3.4.10. The main difference is that we replace completeness by compactness as the main tool.

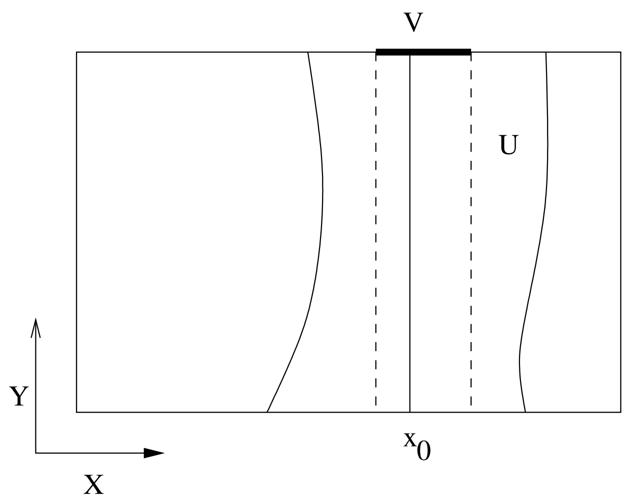
2. For an important application of the above result, see Proposition 7.8.9. □

7.5 Compactness of products. Tychonov's theorem

7.5.1 The slice lemma. Compactness of finite products

The following lemma is extraordinarily important since it is behind most results involving products of two spaces at least one of which is compact (Exercises 7.5.3(ii), 7.5.5 and 7.7.45, Lemma 7.9.5, Proposition 8.5.23).

Lemma 7.5.1 (Slice lemma) *For X arbitrary and Y compact, let $x_0 \in X$ and $U \subseteq X \times Y$ open such that $\{x_0\} \times Y \subseteq U$. Then there is an open $V \subseteq X$ such that $x_0 \in V$ and $V \times Y \subseteq U$.*



Remark 7.5.2 For non-compact Y, the conclusion of the lemma fails: For $X = Y = \mathbb{R}$, consider $U = \{(x, y) \mid |xy| < 1\}$. Then $U \subseteq \mathbb{R}^2$ is open and contains $\{0\} \times \mathbb{R}$. But since $(\varepsilon/2, 4/\varepsilon)$ has arbitrarily small x-coordinate but is not in U, there is no $\varepsilon > 0$ such that $(-\varepsilon, \varepsilon) \times \mathbb{R} \subseteq U$. □

Exercise 7.5.3 (i) Prove Lemma 7.5.1. Hint: Use the proof of Lemma 7.4.1 as inspiration.

(ii) Use (i) to prove: If X and Y are compact then $X \times Y$ is compact.

Corollary 7.5.4 *Any finite direct product of compact spaces is compact.*

We will soon prove that this result is also true without the finiteness assumption.

Another application of the slice lemma concerns projection maps from a product space to its direct factors, which as we saw in Exercise 6.5.9 are not always closed:

Exercise 7.5.5 Let X be arbitrary and Y compact. Prove that the projection map $p_1 : X \times Y \to X$ is closed. Hint: Use Lemma 7.5.1.

Remark 7.5.6 The converse is also true: If $p_1 : X \times Y \to X$ is closed for every X then Y is compact. (In fact, it suffices if this holds for all normal X.) Cf. [75, Theorem 3.1.16]. □

Now we can prove a converse of the implication in Exercise 6.5.21:

Exercise 7.5.7 Let X, Y be topological spaces with Y compact and $f : X \to Y$ such that the graph $G(f) \subseteq X \times Y$ is closed. Prove that f is continuous.

Hint: This can be proven using Exercise 7.5.5 or, alternatively, using nets, Exercise 5.1.33 and Propositions 5.2.5, 7.2.3.

Exercise 7.5.8 Let X, Y be topological spaces, $A \subseteq X,\ B \subseteq Y$ compact subspaces and $U \subseteq X \times Y$ open such that $A \times B \subseteq U$. Prove that there are open $V \subseteq X,\ W \subseteq Y$ such that $A \times B \subseteq V \times W \subseteq U$.

7.5.2 Tychonov's theorem

Corollary 7.5.4 result can be improved considerably:

Theorem 7.5.9 (Tychonov 1929) [4] *Let $X_i \neq \emptyset\ \forall i \in I$. Then $\prod_i (X_i, \tau_i)$ is compact if and only if (X_i, τ_i) is compact for every $i \in I$.*

Proof. $\Rightarrow$ If $X_i \neq \emptyset\ \forall i \in I$ and $X = \prod_k X_k$ is compact then compactness of each X_i follows from $X_i = p_i(X)$ and Corollary 7.3.7.

$\Leftarrow$ Let (X_i, τ_i) be compact for each I. As we know, $\mathcal{S} = \{p_i^{-1}(V) \mid i \in I,\ V \in \tau_i\}$ is a subbase for the product topology τ on $X = \prod_i X_i$. By Alexander's Subbase Lemma 7.2.5, X is compact if (and only if) every cover $\mathcal{U} \subseteq \mathcal{S}$ of X by subbasic sets has a finite subcover. To prove this by contradiction, assume that $\mathcal{U} \subseteq \mathcal{S}$ covers X, but *no* finite $\mathcal{U}' \subseteq \mathcal{U}$ does. For each $i \in I$, define

$$\mathcal{V}_i = \{V \in \tau_i \mid p_i^{-1}(V) \in \mathcal{U}\} \subseteq \tau_i. \tag{7.3}$$

For every $U \in \mathcal{U}$ there are $i \in I, V \in \tau_i$ such that $U = p_i^{-1}(V)$. But then (7.3) gives $V \in \mathcal{V}_i$, and in view of $\bigcup \mathcal{U} = X$ we have:

$$\text{For every } \ x \in X \ \text{ there are } \ i \in I,\ V \in \mathcal{V}_i \ \text{ such that } \ p_i(x) \in V. \tag{7.4}$$

We now claim that $\mathcal{V}_i$ cannot be an open cover of X_i for any $i \in I$. Otherwise compactness of X_i would allow us to find a finite subcover $\{V_1, \ldots, V_n\} \subseteq \mathcal{V}_i$. But then $\mathcal{U}' = \{p_i^{-1}(V_1), \ldots, p_i^{-1}(V_n)\}$ (which is contained in $\mathcal{U}$ by the definition (7.3) of $\mathcal{V}_i$) would be a finite subcover of $\mathcal{U}$, contradicting our assumption. This means that for each $i \in I$ we can find an $x_i \in X_i \backslash \bigcup \mathcal{V}_i$. Now the point $x \in X$ determined by $p_i(x) = x_i\ \forall i$ violates (7.4). This is the desired contradiction. ■

Remark 7.5.10 Tychonov's theorem is one of the most important results of general topology and has countless applications, from abstract algebra and logic to functional analysis, several of which we will meet later. For this reason, we will give another proof of the subbase lemma, and two alternative proofs of Tychonov's theorem using nets[5]. Of course, it is not necessary to know them all, but each of them introduces new tools that have other applications: While the first proof of the subbase lemma was pedestrian, using only Zorn's lemma, the second one makes the use of (ultra)filters explicit (which were implicit in the first proof). The third proof of Tychonov's theorem uses the characterization of compactness in terms of convergence of subnets, and the fourth uses universal nets, the analogue for nets of ultrafilters. □

[4] Andrey Nikolayevich Tikhonov (1906-1993). Born in Russia before the revolution and died after the demise of the Soviet Union. It has been said that "Tychonov's theorem is due to Čech, while Tychonov discovered the Stone-Čech compactification". The real truth is even more complicated, cf. [80].

[5] Whoever finds this excessive should look at Titchmarsh's marvelous book *The theory of the Riemann zeta function*, where seven proofs of the functional equation of the ζ-function are given, plus several variants. Sir Michael Atiyah (Fields medal): "If you have only one proof of a theorem then you cannot say that you understand it very well."

As we have seen in Exercise 6.5.6(vi), the box topology on an infinite product of spaces differs from the product topology (if we exclude trivial exceptions). The next result is the main reason why the product topology is 'better':

Exercise 7.5.11 Consider $[0,1]$ with its standard (Euclidean) topology and prove that the box topology on $[0,1]^{\mathbb{N}}$ is non-compact.

7.5.3 Complements

The proof of Tychonov's theorem given above used Zorn's lemma, which is equivalent to the axiom of choice, and the same is true of the other proofs given below. The following observation shows that this cannot be avoided:

Theorem 7.5.12 (Kelley 1950) [6] *The statement that all products of compact T_1-spaces are compact implies the Axiom of Choice.*

Proof. Let $X_i \neq \emptyset\ \forall i \in I$. We equip each X_i with the cofinite topology, which is T_1 by Exercise 2.5.7(iv) and compact by Exercise 7.1.6(vi). Obviously, the one-point space $(\{\infty\}, \tau)$, where τ is the unique topology, is compact and T_1. Thus the direct sums $(Y_i, \tau_i) = (X_i, \tau_{\text{cofin}}) \oplus (\{\infty\}, \tau)$ are compact and T_1. Now our assumption gives that $Y = \prod_i (Y_i, \tau_i)$ is compact. For each $i \in I$ let $C_i = \{y \in Y \mid y_i \in X_i\} = p_i^{-1}(X_i)$, which is closed since $X_i \subseteq Y_i$ is closed. Now let $J \subseteq I$ be finite and consider the finite intersection $\bigcap_{j \in J} C_j$ of closed sets. This is the set $\{y \in Y \mid p_j(y) \in X_j\ \forall j \in J\}$. Since each X_i is non-empty we can choose $x_j \in X_j$ for every $j \in J$. (Since J is finite, this requires only the finite axiom of choice!) Now let $y \in Y$ be the point whose j-th coordinate is x_j if $j \in J$ and ∞ otherwise. Clearly $y \in \bigcap_{j \in J} C_j$, so that we have proven that every finite intersection of the closed sets C_i is non-empty. Since Y is compact, Lemma 7.2.2 implies $\bigcap_{i \in I} C_i \neq \emptyset$. But by definition of C_i we have $\bigcap_{i \in I} C_i = \prod_{i \in I} X_i$, so that we have proven $\prod_{i \in I} X_i \neq \emptyset$, which is the Axiom of Choice! ■

Remark 7.5.13 1. By Kelley's result above, Tychonov's theorem implies the Axiom of Choice, even when assumed only for T_1-spaces. However, the restriction of Tychonov's theorem to T_2-spaces is strictly weaker than AC. The following statements actually turn out to be equivalent:

1. Tychonov's theorem restricted to T_2-spaces.
2. $[0,1]^N$ is compact for every cardinal number N.
3. $\{0,1\}^N$ is compact for every cardinal number N.
4. Alexander's Subbase Lemma 7.2.5.
5. The ultrafilter lemma: Every filter embeds into an ultrafilter. Cf. Section 7.5.4.
6. The ultranet lemma: Every net has a universal subnet. Cf. Section 7.5.6.
7. Alaoglu's theorem in functional analysis. Cf. Appendix G.3.
8. Stone duality between Stone spaces and Boolean algebras. Cf. Section 11.1.11.
9. Every commutative ring has a prime ideal. (Maximal ideals are prime, but not vice versa. Thus this is weaker than the existence of maximal ideals, which is equivalent to AC.)

[6] John L. Kelley (1916-1999), American topologist and author of the classic textbook [139].

This class of statements is provably weaker than the class containing the Axiom of Choice: There exists a model of ZF-set theory in which the above statements hold, but not the Axiom of Choice!

2. The above statement that the subbase lemma is strictly weaker than Tychonov's theorem and the (equivalent) axiom of choice does not contradict our deduction of Tychonov's theorem from the subbase lemma since that proof also contains a direct application of the axiom of choice. The same remark applies to the fourth proof, where Tychonov's theorem is proven using the ultranet lemma.

3. The fact that $[0,1]^{\mathbb{N}}$ is compact can be proven within Zermelo-Frenkel set theory without any choice axiom, cf. [114, Theorem 3.13]. □

7.5.4 ⋆ Ultrafilters. Alternative proof of the subbase lemma

The proof of the subbase lemma given earlier can be made more conceptual by formulating it explicitly in therms of filters, cf. Section 5.1.3, and ultrafilters. The terminology introduced for this purpose will have other applications later.

We notice that filters on X are (partially) ordered by inclusion (as subsets of $P(X)$).

Definition 7.5.14 *A maximal filter or ultrafilter is a filter that is not properly contained in another filter.*

Remark 7.5.15 Much later, we will see that the ultrafilters on a set X are closely related to Stone-Čech compactification βX of X. Cf. Theorem 11.1.82. □

The following result is crucial:

Lemma 7.5.16 (Ultrafilter Lemma) *Every filter is contained in an ultrafilter.*

Proof. Let X be a set and $\mathcal{F}$ a filter on X. The family $\mathfrak{F}$ of all filters on X that contain $\mathcal{F}$ is a partially ordered set w.r.t. inclusion. If $\mathcal{C} \subseteq \mathfrak{F}$ is a totally ordered subset of $\mathfrak{F}$, we claim that the union $\bigcup \mathcal{C}$ of all elements of $\mathcal{C}$ is a filter (obviously containing $\mathcal{F}$). A union of filters clearly has the properties (ii) and (iii) in Definition 5.1.34. Let $F_1, F_2 \in \bigcup \mathcal{C}$. By the total order of $\mathcal{C}$, there is a $\widetilde{\mathcal{F}} \in \mathcal{C}$ such that $F_1, F_2 \in \widetilde{\mathcal{F}}$ and thus $F_1 \cap F_2 \in \widetilde{\mathcal{F}} \subseteq \bigcup \mathcal{C}$. This proves requirement (i), thus $\bigcup \mathcal{C}$ is in $\mathfrak{F}$ and is an upper bound for the chain $\mathcal{C}$. Therefore Zorn's lemma applies and gives a maximal filter $\widehat{\mathcal{F}}$ containing $\mathcal{F}$. ■

Ultrafilters are characterized by a quite remarkable property:

Lemma 7.5.17 *A filter $\mathcal{F}$ on X is an ultrafilter if and only if for every $Y \subseteq X$ exactly one of the alternatives $Y \in \mathcal{F},\ X \backslash Y \in \mathcal{F}$ holds.*

Proof. We begin by noting that we cannot have both $Y \in \mathcal{F}$ and $X \backslash Y \in \mathcal{F}$ since (i) would imply $\emptyset = Y \cap (X \backslash Y) \in \mathcal{F}$, which is forbidden by (iii). Assume $\mathcal{F}$ contains Y or $X \backslash Y$ for every $Y \subseteq X$. This means that $\mathcal{F}$ cannot be enlarged by adding $Y \subseteq X$ since either already $Y \in \mathcal{F}$ or else $X \backslash Y \in \mathcal{F}$, which excludes $Y \in \mathcal{F}$. Thus $\mathcal{F}$ is an ultrafilter.

Now assume that $\mathcal{F}$ is an ultrafilter and $Y \subseteq X$. If there is an $F \in \mathcal{F}$ such that $F \cap Y = \emptyset$ then $F \subseteq X \backslash Y$, and property (ii) implies $X \backslash Y \in \mathcal{F}$. If, on the other hand, $Y \cap F \neq \emptyset\ \forall F \in \mathcal{F}$ then there is a filter $\widetilde{\mathcal{F}}$ containing $\mathcal{F}$ and Y. Since $\mathcal{F}$ is maximal, we must have $Y \in \mathcal{F}$. ■

Lemma 7.5.18 *A topological space on which every ultrafilter converges is compact.*

Proof. Let $\mathcal{U} = \{U_i\}_{i \in I}$ be an open cover admitting no finite subcover. Then for every finite $J \subseteq I$, the set $Y_J = X \backslash \bigcup_{j \in J} U_j$ is non-empty. In particular, $Y_J \cap Y_{J'} = Y_{J \cup J'} \neq \emptyset$. Therefore, the set $\mathcal{F}$ of all sets containing a Y_J is a filter. Let $\widehat{\mathcal{F}}$ be an ultrafilter containing $\mathcal{F}$. By assumption, $\widehat{\mathcal{F}}$ converges to some $x \in X$. Clearly $x \in U_i$ for some $i \in I$. But then $U_i \in \mathcal{N}_x$, and the convergence of $\widehat{\mathcal{F}}$ to x gives $U_i \in \widehat{\mathcal{F}}$. On the other hand, we have $X \backslash U_i \in \mathcal{F} \subseteq \widehat{\mathcal{F}}$ by construction of $\mathcal{F}$. In view of this contradiction, $\mathcal{U}$ has must have a finite subcover, and X is compact. ■

Lemma 7.5.19 *Let (X, τ) be a topological space. If there is a non-convergent ultrafilter on X then for every subbase $\mathcal{S} \subseteq \tau$ there is a cover of X by elements of $\mathcal{S}$ that has no finite subcover.*

Proof. Let $\mathcal{F}$ be an ultrafilter on X that does not converge. Then every $x \in X$ has an open neighborhood U_x such that $U_x \notin \mathcal{F}$. Since $\mathcal{S}$ is a subbase, there are $S_{x,1}, \ldots, S_{x,n_x} \in \mathcal{S}$ such that $x \in S_{x,1} \cap \ldots \cap S_{x,n_x} \subseteq U_x$. If all $S_{x,1}, \ldots, S_{x,n_x}$ were in $\mathcal{F}$, so would be their intersection and thus U_x. Thus for every $x \in X$ there is an $S_x \in \mathcal{S}$ such that $x \in S_x \notin \mathcal{F}$. Thus $\mathcal{U} = \{S_x \mid x \in X\}$ is an open cover by elements of $\mathcal{S}$. Assume there is a finite subcover $\{S_{x_1}, \ldots, S_{x_n}\}$. Since the S_{x_i} are not in $\mathcal{F}$, by Lemma 7.5.17 we must have $X \backslash S_{x_1}, \ldots, X \backslash S_{x_n} \in \mathcal{F}$. But then also $\emptyset = \bigcap_{i=1}^n X \backslash S_{x_i} \in \mathcal{F}$, which is impossible. Thus the cover $\mathcal{U} \subseteq \mathcal{S}$ has no finite subcover. ■

Corollary 7.5.20 *X is compact if and only if every ultrafilter on X converges.*

Proof. The 'if' direction is Lemma 7.5.18. If there is a non-convergent ultrafilter on X, Lemma 7.5.19 applied to $\mathcal{S} = \tau$ gives an open cover that has no finite subcover. Thus X is non-compact. ■

Proof of the Subbase Lemma. If every cover $\mathcal{U} \subseteq \mathcal{S}$ has a finite subcover, then Lemma 7.5.19 implies that every ultrafilter on X is convergent. Then X is compact by Lemma 7.5.18. ■

Remark 7.5.21 This proof clearly is not completely independent of the preceding one. Being closed w.r.t. finite intersections and not containing $\emptyset$, a filter $\mathcal{F} \subseteq P(X)$ has the finite intersection property. On the other hand, if $\mathcal{F}$ has the finite intersection property, then $\widetilde{\mathcal{F}}$ defined as the family of all supersets of finite intersections of elements of $\mathcal{F}$ is a filter. (In Bourbaki's terminology, $\mathcal{F}$ is a filter base for the filter $\widetilde{\mathcal{F}}$, cf. [25].) Furthermore, the maximal family $\widehat{\mathcal{F}}$ with finite intersection property considered in the first proof clearly is an ultrafilter. □

7.5.5 ⋆⋆ Third proof of Tychonov, using nets

Our third proof of Tychonov's theorem uses the characterization of compactness in terms of existence of accumulation points of all nets, cf. Proposition 7.2.3, and Lemma 6.5.3 on convergence of nets in product spaces.

Third Proof of Tychonov's Theorem. We begin with some terminology: We write $X = \prod_{i \in I} X_i$ and $X_J = \prod_{j \in J} X_j$ if $J \subseteq I$. A partial point in X is a pair (J, x), where $J \subseteq I$ and $x \in X_J$. If (J, x) is a partial point and $K \subseteq J$, then $x \restriction K \in X_K$ is the obvious restriction. A partial point (J, x) is called partial accumulation point of a net $\{x_\iota\}$ in X if $x \in X_J$ is an accumulation point of $x_\iota \restriction J$. The set $\mathcal{P}$ of partial accumulation points of x_ι is partially ordered by

$$(J, x) \leq (K, y) \quad \Leftrightarrow \quad J \subseteq K \;\text{ and }\; y \restriction J = x. \tag{7.5}$$

$\mathcal{P}$ is not empty since the nets $p_i(x_\iota)$ in X_i have accumulation points $x_i \in X_i$ by the compactness of X_i, thus $(\{i\}, x_i) \in \mathcal{P}\ \forall i \in I$. Now let $\mathcal{C} \subseteq \mathcal{P}$ be a totally ordered subset. Define $M = \bigcup\{J \mid (J,y) \in \mathcal{C}\} \subseteq I$ and define $x \in X_M$ by $x_i = y_i$, where $i \in M$ and $(J,y) \in \mathcal{C}$ with $J \ni i$. This is well-defined since $\mathcal{C}$ is totally ordered w.r.t. (7.5). We must prove that (M,x) is a partial accumulation point of x_ι. If $U \subseteq X_M$ is an open neighborhood of $x \in X_M$, then U contains a neighborhood V of the form

$$V = p_{i_1}^{-1}(U_1) \cap \cdots \cap p_{i_n}^{-1}(U_n),$$

where $i_1, \ldots, i_n \in M$ and $U_k \in \tau_{i_k}$. Now each i_k is contained in some J_k, where $(J_k, x_k) \in \mathcal{C}$. Since $\mathcal{C}$ is totally ordered, we can find $(J,x) \in \mathcal{C}$ such that $J \supseteq \{i_1, \ldots, i_n\}$. Since $(J, x \upharpoonright J) \in \mathcal{C}$ is a partial accumulation point, $x_\iota \upharpoonright M$ is frequently in V. Thus (M,x) is a partial accumulation point of x_ι, and (M,x) is an upper bound for $\mathcal{C}$.

Now Zorn's lemma applies and provides a maximal element $(J,x) \in \mathcal{P}$. If we can show that $J = I$, we have that $x \in X$ is an accumulation point for x_ι and we are done. So assume $J \subsetneq I$. Since $x \in X_J$ is an accumulation point of $x_\iota \upharpoonright J$, by Proposition 5.1.30 there is a subnet of x'_α of x_ι such that $x'_\alpha \upharpoonright J$ converges to $x \in X_J$. If now $i' \in I \backslash J$, compactness of $X_{i'}$ and Proposition 7.2.3 imply that there is a subnet $\{x''_\beta\}$ of x'_α such that $p_{i'}(x''_\beta)$ converges to $z \in X_{i'}$. But this means that taking $K = J \cup \{i'\}$ and defining $y \in X_K$ by $y \upharpoonright J = x$ and $y_{i'} = z$, (K,y) is a partial accumulation point and $(K,y) > (J,x)$, contradicting maximality of (J,x). Thus $J = I$. ■

Remark 7.5.22 This proof is due to P. R. Chernoff, 1992 [48]. Cf. also [204]. □

7.5.6 ⋆⋆ Universal nets. Fourth proof of Tychonov

The definition of universal nets is inspired by Lemma 7.5.17:

Definition 7.5.23 *A net $\{x_\iota\}_{\iota \in I}$ in a set X is universal if for every $Y \subseteq X$ the net eventually lives in Y or in $X \backslash Y$.*

Lemma 7.5.24 *(i) If $\{x_\iota\}$ is a universal net in X and $f : X \to Y$ a function then $\{f(x_\iota)\}$ is a universal net in Y.*

(ii) If a universal net in a topological space has an accumulation point, it converges to that point.

Proof. (i) If $D \subseteq Y$ then the net eventually lives in $f^{-1}(D)$ or in $X \backslash f^{-1}(D)$, thus $f(x_\iota)$ is eventually in D or in $Y \backslash D$.

(ii) Let $x \in X$ be an accumulation point of x_ι and U an open neighborhood of x. Since the net frequently is in U, it cannot be eventually in $X \backslash U$, so it is eventually in U. Thus it converges to x. ■

Every constant net is universal, but it may be hard to believe that non-constant universal nets can exist. In fact, in analogy to the Ultrafilter Lemma, we have:

Lemma 7.5.25 (Ultranet lemma) *Every net has a universal subnet.*

Proof. Given a net $\{x_\iota\}$ in X, one uses Zorn's lemma in the same way as done several times before to obtain a maximal family $\mathcal{F}$ of sets having the properties that (i) x_ι frequently is in each $F \in \mathcal{F}$ and (ii) $\mathcal{F}$ is directed w.r.t. reverse inclusion, i.e. given $F_1, F_2 \in \mathcal{F}$, there is $F_3 \in \mathcal{F}$ with $F_3 \subseteq F_1 \cap F_2$. (Note that at least one family satisfying (i) and (ii) exists, namely $\mathcal{F} = \{X\}$.) By maximality, $\mathcal{F}$ is an ultrafilter, and by Lemma 7.5.17, for every $Y \subseteq X$ it contains exactly one of Y and $X \backslash Y$. Now

we can apply Lemma 5.1.31 to x_ι and $\mathcal{F}$ and obtain a subnet that, for each $Y \subseteq X$, eventually lives in Y or $X \backslash Y$ (depending on which of the two is in $\mathcal{F}$). Thus the subnet is a universal net. ■

The following is the universal net analogue of Corollary 7.5.20:

Proposition 7.5.26 *A space X is compact if and only if every universal net in X converges.*

Proof. By Lemma 7.5.25, every net has a universal subnet. So if every universal net converges then every net has a convergent subnet and thus X is compact by Proposition 7.2.3. Conversely, let X be compact and x_ι a universal net. By Proposition 7.2.3, x_ι has a convergent subnet, i.e. an accumulation point. Then by Lemma 7.5.24(ii), the whole net $\{x_\iota\}$ converges to that point. Thus every universal net in X converges. ■

Fourth Proof of Tychonov's Theorem. Let $\{x_\iota\}$ be a universal net in $\prod_k X_k$. By Lemma 7.5.24(i), $p_i(x_\iota)$ is a universal net in X_i for each $i \in I$, and therefore convergent to some point x_i by compactness of X_i and Proposition 7.5.26. Let $x \in \prod_k X_k$ be the point for which $p_i(x) = x_i \ \forall i$. Since $p_i(x_\iota) \to p_i(x) = x_i \ \forall i$, Lemma 6.5.3 implies $x_\iota \to x$. Thus every universal net in $\prod_k X_k$ converges, and therefore $\prod_k X_k$ is compact by Proposition 7.5.26. ■

Remark 7.5.27 1. This approach is due to Kelley (1950, [138]). While the actual proof of Tychonov's theorem is short and pretty, the reader will have noticed that the fourth proof is not independent of the earlier ones since it uses some of the material introduced for the second and third proofs. Rather more seriously, the whole approach using universal nets is marred by its reliance on the theory of (ultra)filters. The proof of the ultranet lemma by passing from the net to a family of sets, essentially a filter, enlarging that to an ultrafilter, and using the latter to extract a universal subnet, is very unsatisfactory. Furthermore, the resulting ultranet proof essentially **is** the ultrafilter proof in Section 7.5.4 with a lot of added tautological overhead (particularly evident in Lemma 5.1.31).

2. In [1], a more appealing proof of the ultranet lemma is given, which proceeds by applying Zorn's lemma directly to the family of subnets of $\{x_\iota\}$, obtaining a universal subnet. However, [1] adopts a definition of subnets different from the one used above (due to Moore). A nice aspect of this definition is that a net is universal in the sense of Definition 7.5.23 if and only if it has no proper subnet in the sense of [1]. (For Moore-subnets, there is no meaningful notion of properness.) But the [1]-definition of subnets involves the filters associated to the nets in question! It seems that there is no way to work with nets that is self-contained in that it does not invoke filters. □

7.6 ⋆ Compactness of ordered topological spaces. Supercompact spaces

In this section, we give an application of Alexander's subbase lemma to the question when an ordered topological space, cf. Remark 4.2.5, is compact. This discussion is almost literally lifted from [27].

Definition 7.6.1 *Let $(X, \leq)$ be a totally ordered set and $Y \subseteq X$. A <u>supremum</u> (or <u>least upper bound</u>) for Y is an $x \in X$ such that (i) x is an upper bound for Y and (ii) no $z < x$ is an upper bound for Y.*

Theorem 7.6.2 *An ordered topological space is compact if and only if every subset (including $\emptyset$!) has a supremum.*

Proof. Assume every $Y \subseteq X$ has a supremum, and let $\mathcal{U}$ be an open cover of X by elements of the subbase (4.3). In particular X has a supremum, which is an upper bound M for X. Being maximal, M is not contained in any L_x, thus $\mathcal{U}$ must contain R_x for some $x \in X$. Clearly every $x \in X$ is an upper bound for the empty set $\emptyset$, so that existence of a supremum for $\emptyset$ amounts to existence of a smallest element m of X. Being minimal, m is not contained in any R_x, so $\mathcal{U}$ must contain L_x for some $x \in X$. Thus $Y = \{x \in X \mid L_x \in \mathcal{U}\}$ is not empty, and by assumption there exists $\sup(Y) =: y$. This y is not contained in any $L_x \in \mathcal{U}$ (otherwise $x > y$, contradicting the definition of y), thus $y \in R_x$ for some $R_x \in \mathcal{U}$, meaning $x < y$. Since y is a least upper bound for Y, x cannot be an upper bound for Y. We therefore must have $y > x$ for some $y \in Y$. By definition of y, we have $L_y \in \mathcal{U}$. Summing up, we have found $x < y$ such that $\{L_y, R_x\} \subseteq \mathcal{U}$. Since $L_y \cup R_x = X$, we have found a finite subcover of $\mathcal{U}$. Now Alexander's Subbase Lemma 7.2.5 implies that X is compact.

Now assume that the order topology on X is compact, and suppose that there is a $Y \subseteq X$ without a supremum. If $Y = \emptyset$, this means that X has no minimum. In that case, $\mathcal{U} = \{R_x \mid x \in X\}$ is an open cover (since $\forall x \in X\, \exists y \in X$ with $y < x$) admitting no finite subcover. (If there was a finite subcover $\{R_{x_1}, \dots, R_{x_n}\}$ then, defining $z = \min(x_1, \dots, x_n)$ we would have $X = R_z$. But this is impossible since $z \in X$, but $z \notin R_z$.)

This leaves us with $Y \neq \emptyset$. Begin by assuming that Y has no upper bound, thus Y contains arbitrarily large elements. Arguing as before, $\mathcal{U} = \{L_y \mid y \in Y\}$ covers X, so that by compactness there is a finite subcover $\{L_{y_1}, \dots, L_{y_n}\}$. Defining $z = \max(y_1, \dots, y_n)$ we find that $X = L_z$, which is impossible since $z \notin L_z$.

Thus Y must have an upper bound, but no least one. Let Z be the set of upper bounds of Y. Now $\{L_y \mid y \in Y\} \cup \{R_z \mid z \in Z\}$ covers X: If $x \in X$ is not in $\bigcup\{L_y \mid y \in Y\}$ then $x \geq y\ \forall y \in Y$, thus x is an upper bound for Y. Since Y admits no least upper bound, there is an upper bound $z \in Z$ such that $z < x$. But then $x \in \bigcup\{R_z \mid z \in Z\}$. Therefore, by compactness, there is a finite subcover $\{L_{y_1}, \dots, L_{y_n}, R_{z_1}, \dots, R_{z_m}\} \subseteq \mathcal{U}$. Defining $a = \max(y_1, \dots, y_n) \in Y$, $b = \min(z_1, \dots, z_m) \in Z$, we have $X = L_a \cup R_b$. Since b is an upper bound for Y and $a \in Y$, we have $a \leq b$. Actually $a < b$ since $a = b$ would be a least upper bound for Y. Now, since $a \notin L_a$, we must have $a \in R_b$, but this is impossible since because of $a < b$. $\lightning$ ■

Remark 7.6.3 The set $X = (0, 1) \subseteq \mathbb{R}$ has supremum 1 in the ambient space $\mathbb{R}$, but none in X, consistent with the non-compactness. $X = (0, 1]$ has a supremum (in X), but the empty subset $\emptyset \subseteq X$ has no supremum, thus again X is non-compact. □

Exercise 7.6.4 Prove that the order topology arising from the lexicographic order on $[0, 1] \times [0, 1]$ is compact.

The proof of Theorem 7.6.2 motivates some further developments:

Definition 7.6.5 *A topological space (X, τ) is supercompact if admits a subbase $S \subseteq \tau$ such that every open cover by elements of S has a subcover with at most two elements.*

Corollary 7.6.6 *Compact ordered ⇒ supercompact ⇒ compact.*

Proof. The first implication is contained in the first half of the proof of Theorem 7.6.2, and the second is immediate by Alexander's subbase lemma. ■

Remark 7.6.7 While ordered topological spaces are the most 'natural' examples of supercompact spaces, there are others: If X is a compact Hausdorff space, sufficient conditions for X being supercompact are: (i) X is second countable (=metrizable), (ii) X is a topological group. There

are examples of compact Hausdorff spaces that are not supercompact. It is not hard to show that arbitrary products of supercompact spaces are supercompact. We leave the subject here. □

7.7 Compactness: Variations, metric spaces and subsets of $\mathbb{R}^n$

Before the notion of compactness was established as the 'right one', mathematicians experimented with various related definitions. We will have a quick look at countable and sequential compactness and then turn to metric spaces. (For much more information, cf. [75].)

7.7.1 Countable compactness. Weak countable compactness

The definition of compactness (**every** open cover $\mathcal{U}$ has a **finite** subcover $\mathcal{V}$) can be weakened by limiting the cardinality of the cover $\mathcal{U}$ or by allowing certain infinite subcovers $\mathcal{V}$. The latter leads to the Lindelöf property, which we have already studied. On the other hand:

Definition 7.7.1 *A topological space is called countably compact if every countable open cover has a finite subcover.*

The following should be obvious:

Exercise 7.7.2 (i) Compact $\Leftrightarrow$ (countably compact & Lindelöf).

(ii) Countably compact & second countable $\Rightarrow$ compact.

In analogy to compactness and the Lindelöf property, one finds:

Exercise 7.7.3 For countably compact X, prove:

(i) If $Y \subseteq X$ is closed then Y is countably compact.

(ii) If $f : X \to Y$ is continuous then $f(X)$ is countably compact.

Remark 7.7.4 1. Exercise 7.3.1 also shows that countable compactness is not hereditary. A Hausdorff example: The countable open cover $\{(n, n+2) \mid n \in \mathbb{Z}\}$ of $\mathbb{R}$ has no countable subcover. Thus $\mathbb{R}$ is not countably compact, and the same holds for $(0,1) \cong \mathbb{R}$. But $[0,1]$ is compact, thus countably compact.

2. With respect to products, countable compactness behaves equally bad as the Lindelöf property: One can find countably compact spaces X, Y such that $X \times Y$ is not countably compact. (But if X, Y are countably compact and Lindelöf then the same holds for $X \times Y$.) □

Exercise 7.7.5 Prove that for a topological space X, the following are equivalent:

(i) X is countably compact.

(ii) $\bigcap \mathcal{F} \neq \emptyset$ for every countable family $\mathcal{F}$ of closed sets having the finite intersection property.

(iii) $\bigcap_n C_n \neq \emptyset$ for every sequence $\{C_n\}_{n \in \mathbb{N}}$ of non-empty closed sets such that $C_{n+1} \subseteq C_n$ $\forall n$.

In the rest of this section we will need the notion of ω-accumulation points (Definition 2.7.23). It is easy to see that every infinite subset Y of a compact space X has an ω-accumulation point: If this was false then every $x \in X$ would have an open neighborhood U_x such that $U_x \cap Y$ is finite. The U_x cover X, and by compactness there is a finite subcover. Now a finite union of sets U_x has finite intersection with Y, producing a contradiction. But we can do better, cf. the implication (i)⇒(ii) in the following:

Proposition 7.7.6 *For a topological space (X, τ), the following are equivalent:*

(i) X is countably compact.

(ii) Every infinite subset $Y \subseteq X$ has an ω-accumulation point. (Compare Exercise 7.2.7.)

(iii) Every sequence in X has an accumulation point.

Proof. (i)⇒(ii) It clearly is enough to prove this for countably infinite subsets $Y \subseteq X$. Thus let $Y \subseteq X$ be a countable subset without ω-accumulation point. Then every $x \in X$ has an open neighborhood U_x such that $U_x \cap Y$ is finite. Now for every finite subset $F \subseteq Y$, we define $U_F = \bigcup\{U_x \mid U_x \cap Y = F\}$ and note that we either have $U_F \cap Y = \emptyset$ (if there is no x such that $U_x \cap Y = F$) or $U_F \cap Y = F$. Thus in any case, $U_F \cap Y \subseteq F$. Since Y is countable, the family $\mathcal{F} = \{F \subseteq Y \mid F \text{ finite}\}$ is countable, and since every U_x is contained in some U_F (namely for $F = U_x \cap Y$), we see that $\{U_F\}_{F \in \mathcal{F}}$ is a countable open cover of X. By countable compactness, there is a finite subcover $\{U_{F_1}, \ldots, U_{F_n}\}$. Thus

$$Y = (U_{F_1} \cup \cdots \cup U_{F_n}) \cap Y = (U_{F_1} \cap Y) \cup \cdots \cup (U_{F_n} \cap Y) \subseteq F_1 \cup \cdots \cup F_n,$$

which is finite, contradicting the assumption that Y is infinite.

(ii)⇒(iii) Let $Y = \{x_1, x_2, \ldots\}$ be the set of values of the sequence. If Y is finite, there must be a $y \in Y$ such that $x_n = y$ for infinitely many $n \in \mathbb{N}$. This y obviously is an accumulation point of the sequence $\{x_n\}$. If, on the other hand, Y is infinite, then by (ii) there exists an $x \in X$ such that $Y \cap U$ is infinite for every neighborhood U of x. But this precisely means that x is an accumulation point of the sequence $\{x_n\}$.

(iii)⇒(i). Let $\{U_n\}_{n \in \mathbb{N}}$ be a countable open cover of X that does not admit a finite subcover. This means that $\bigcup_{k=1}^n U_k \neq X$ for all n, thus we can choose $x_n \in X \backslash \bigcup_{k=1}^n U_k$ for all $n \in \mathbb{N}$. Our construction of $\{x_n\}$ implies that $x_k \notin U_m$ if $k \geq m$. But this means that no point of U_m can be an accumulation point of $\{x_n\}$. Since $\bigcup_m U_m = X$, the sequence has no accumulation point. This contradiction proves (i). ■

The next result will not be needed later, but it nicely complements Proposition 7.2.3 and the discussion of sequential compactness in the next subsection:

Corollary 7.7.7 *A topological space (X, τ) is countably compact if and only if every sequence in X has a convergent sub<u>net</u>.*

Proof. By Proposition 7.7.6, X is countably compact if and only if every sequence $\{x_n\}$ in X has an accumulation point. By Proposition 5.1.30, this is equivalent to every sequence having a convergent subnet. ■

The point of this result is that not every accumulation point of a sequence is the limit of a subsequence. Cf. the next subsection.

The next exercise explores what happens if we drop the 'ω-' in Proposition 7.7.6(ii):

Exercise 7.7.8 Let X be a topological space and $Y \subseteq X$. Consider the statements

(i) X is countably compact.

(ii) Every infinite $Y \subseteq X$ has an accumulation point.

(iii) Every closed discrete subspace of X is finite.

Prove that (i)⇒(ii)⇔(iii), and if X is T_1 then (ii)⇒(i).

Definition 7.7.9 *A space with the equivalent properties (ii) and (iii) in Exercise 7.7.8 is called weakly countably compact (or limit point compact or Fréchet compact).*

There are non-T_1 spaces that are weakly countably compact but not countably compact!

7.7.2 Sequential compactness

Definition 7.7.10 *A topological space is called sequentially compact if every sequence in it has a convergent subsequence.*

Proposition 7.7.11 *Sequentially compact ⇒ countably compact.*

Proof. By Lemma 5.1.12, limits of convergent subsequences are accumulation points. Thus sequential compactness of X implies (iii) in Proposition 7.7.6, and therefore countable compactness. ■

Remark 7.7.12 While the properties of compactness and countable compactness were defined in terms of covers, i.e. 'statically', we have proven 'dynamical' characterizations: X is compact (resp. countably compact) if and only if every net (resp. sequence) has a convergent subnet. Together with the above definition of sequential compactness, this gives another perspective of the implications compact ⇒ countably compact ⇐ sequentially compact. □

For all other implications there are counterexamples. There are even compact Hausdorff spaces that are not sequentially compact:

Example 7.7.13 Let $I = [0, 1]$ and consider $X = I^I \equiv \prod_{x \in I}[0, 1]$. As a product of compact Hausdorff spaces, X is compact Hausdorff by Exercise 6.5.15(ii) and Tychonov's theorem. An element of X is a function $f : I \to I$. For $n \in \mathbb{N}$ and $x \in I$, let $f_n(x)$ be the n-th digit of x in its binary expansion. (The latter is unique if we forbid infinite chains of 1s.) Assume that the sequence $\{f_n\}$ has a convergent subsequence $m \mapsto f_{n_m}$. Choose $x \in [0, 1]$ such that its n_m-th digit is 0 or 1 according to whether m is even or odd. Now the sequence $m \mapsto f_{n_m}(x)$ is $\{0, 1, 0, 1, \ldots\}$, which does not converge as $n \to \infty$. This contradicts the convergence of f_{n_m} in I^I, which would have to be pointwise for all $x \in [0, 1]$.

X it is compact, thus countably compact, so that by Proposition 7.7.6 every sequence in X – in particular the one constructed above – has an accumulation point. Therefore we have an example of a sequence with an accumulation point that is not the limit of a subsequence, as promised before Proposition 5.1.13. □

Example 7.7.13 of a space that is compact, thus countable compact, but not sequentially compact shows that the converse of Proposition 7.7.11 is not true unconditionally. But in view of Proposition 5.1.13 we have:

Proposition 7.7.14 *(i) Countably compact & first countable ⇒ sequentially compact.*

(ii) If X is first countable then: countably compact $\Leftrightarrow$ sequentially compact.

(iii) If X is second countable then: compact $\Leftrightarrow$ countably compact $\Leftrightarrow$ sequentially compact.

Proof. (i) The implication (i)⇒(iii) of Proposition 7.7.6 implies that every sequence in X has an accumulation point and by Proposition 5.1.13 the latter is the limit of a subsequence. Thus X is sequentially compact.

(ii) This follows from (i) and Proposition 7.7.11.

(iii) Second countability implies first countability (Lemma 4.3.7) and the Lindelöf property (Exercise 7.1.5(iii)). The latter makes compactness and countable compactness equivalent (Exercise 7.7.2), so that the claim follows from (ii). ■

Exercise 7.7.15 Prove:

(i) Closed subspaces of sequentially compact spaces are sequentially compact.

(ii) Products of finitely or countably many sequentially compact spaces are sequentially compact. Hint: Diagonal argument.

Remark 7.7.16 1. In Section 5.1.1 we have seen that sequences tend to be quite defective in spaces without the first countablility property. But for first countable spaces, sequential and countable compactness are equivalent.

2. There actually are spaces that are first countable and countably compact (thus sequentially compact), but not compact, e.g. space #42 in [216]!

3. Exercise 7.7.15(ii) shows that sequential compactness is better behaved w.r.t. products than countable compactness (cf. Remark 7.7.4.2), but not as well as compactness. (Given the inherently countable nature of sequences, it would be unreasonable to expect more.) The fact that compactness leads to the strongest result on products (to wit Tychonov's theorem) is one of the main reasons why compactness won out over countable and sequential compactness. □

7.7.3 Compactness of metric spaces I: Equivalences

We now turn to metric spaces, which besides the topological properties discussed above also admit metric notions like completeness and some others to be introduced now.

Combining Lemma 2.5.4 and Lemma 7.4.2 we find that compact subsets of metric spaces are closed. But there is a simple direct proof:

Exercise 7.7.17 Prove directly, using only the definition of countable compactness:

(i) Countably compact metric spaces are bounded.

(ii) Countably compact subsets of metric spaces are closed.

The following property is stronger than boundedness:

Definition 7.7.18 *A metric space (X, d) is called totally bounded if for every $r > 0$ there are finitely many points $x_1, \ldots, x_n$ such that $B(x_1, r) \cup \cdots \cup B(x_n, r) = X$.*

Subsets (not necessarily closed) and closures of bounded sets are bounded. Similarly:

Exercise 7.7.19 Let (X, d) be a metric space. Prove:

(i) If (X,d) is totally bounded and $Y \subseteq X$ then (Y,d) is totally bounded.

(ii) If (Y,d) is totally bounded and $Y \subseteq X$ is dense then (X,d) is totally bounded.

Lemma 7.7.20 *A metric space is totally bounded if and only if every sequence has a Cauchy subsequence.*

Proof. $\Leftarrow$ If X is not totally bounded, there is an $\varepsilon > 0$ so that X cannot be covered by finitely many ε-balls. Then we can find a sequence $\{x_n\}$ such that $x_n \in X \backslash \bigcup_{i=1}^{n-1} B(x_i, \varepsilon)$ for all n, implying $d(x_i, x_j) \geq \varepsilon$ for all $i > j$. But this gives $i \neq j \Rightarrow d(x_i, x_j) \geq \varepsilon > 0$, so that no subsequence of $\{x_n\}$ can be Cauchy.

$\Rightarrow$ Let $\{x_n\}$ be a sequence in X. Since X is totally bounded, it can be covered by a finite number of balls of radius 1. One of these, call it B_1, must contain x_n for infinitely many n. Now B_1 can be covered by finitely many balls of radius 1/2. Again, one of those balls, call it B_2, has the property that $B_1 \cap B_2$ contains x_n for infinitely many n. Going on in this way, we find a sequence B_i of open balls of radius $1/i$ such that $B_1 \cap \cdots \cap B_k$ contains x_n for infinitely many n, for any k. Thus we can choose a subsequence $\{x_{n_i}\}$ such that $x_{n_i} \in B_1 \cap \cdots \cap B_i$ for each i. Now if $j \geq i$, both x_{n_i} and x_{n_j} are contained in B_i, thus $d(x_{n_i}, x_{n_j}) < 2/i$. This implies that $i \mapsto x_{n_i}$ is a Cauchy sequence. ■

Definition 7.7.21 *Let (X,d) be a metric space.*

(i) If $\mathcal{U}$ is an open cover of X, a real number $\lambda > 0$ is called a Lebesgue number[7] for the cover $\mathcal{U}$ if for every $Y \subseteq X$ with $\mathrm{diam}(Y) < \lambda$ there is a $U \in \mathcal{U}$ such that $Y \subseteq U$.

(ii) (X,d) has the Lebesgue property if every open cover admits a positive Lebesgue number.

Proposition 7.7.22 *In the following diagram, solid (respectively, dashed) arrows indicate implications that are true for all topological (respectively metric) metric spaces. A '+' indicates that a combination of two statements implies the third. Dotted arrows indicate implications that hold under additional assumptions that are weaker than metrizability (T_1, T_4 or first countability).*

Proof. We begin with the known implications: 1. Exercise 7.7.2(i). 2.+3. Exercise 7.7.8 (one direction needs T_1, which holds for metric spaces). Pseudocompactness will only be discussed in Section 7.7.4. 4. Proposition 7.7.11. 5. Proposition 7.7.14 (needs first countability, which holds for metric spaces by Lemma 4.3.6). 6. Exercise 7.1.5(iii). 7. Lemma 4.1.10. 8. Lemma 4.1.11. 9. Exercise 7.1.9. Now we prove the remaining implications.

10. If $\{x_n\}$ is a Cauchy sequence, sequential compactness gives us a convergent subsequence. But a Cauchy sequence with a convergent subsequence is convergent. Thus (X,d) is complete.

11. For every $r > 0$, the family $\{B(x,r)\}_{x \in X}$ is an open cover of X. Now compactness gives existence of a finite subcover, thus total boundedness.

12. If (X,d) is totally bounded, there are $x_1, \ldots, x_n$ such that $X = \bigcup_i B(x_i, 1)$. With $D = \max_{i,j} d(x_i, x_j)$, the triangle inequality gives $\mathrm{diam}(X) \leq D + 2 < \infty$, thus X is bounded.

13. By total boundedness, for every $m \in \mathbb{N}$, there are finitely many points $x_{m,1}, \ldots, x_{m,n_m}$ such that $\bigcup_{k=1}^{n_m} B(x_{m,k}, 1/m) = X$. Then $S = \{x_{m,k} \mid m \in \mathbb{N},\ k = 1, \ldots, n_m\}$ is countable. Given $x \in X$ and $\varepsilon > 0$, pick $\varepsilon^{-1} < m \in \mathbb{N}$. Then there is a k such that $d(x, x_{m,k}) < 1/m < \varepsilon$. Thus S is dense and (X, τ_d) is separable. Now second countability follows from Lemma 4.1.11, but one sees more directly that $\mathcal{B} = \{B(x_{m,k}, 1/m) \mid m \in \mathbb{N},\ k = 1, \ldots, n_m\}$ is a base for τ_d.

14. By sequential compactness, every sequence has a convergent subsequence, thus a Cauchy subsequence. Now Lemma 7.7.20 gives total boundedness.

[7] Henri Leon Lebesgue (1875-1941). Particularly known for his integration theory.

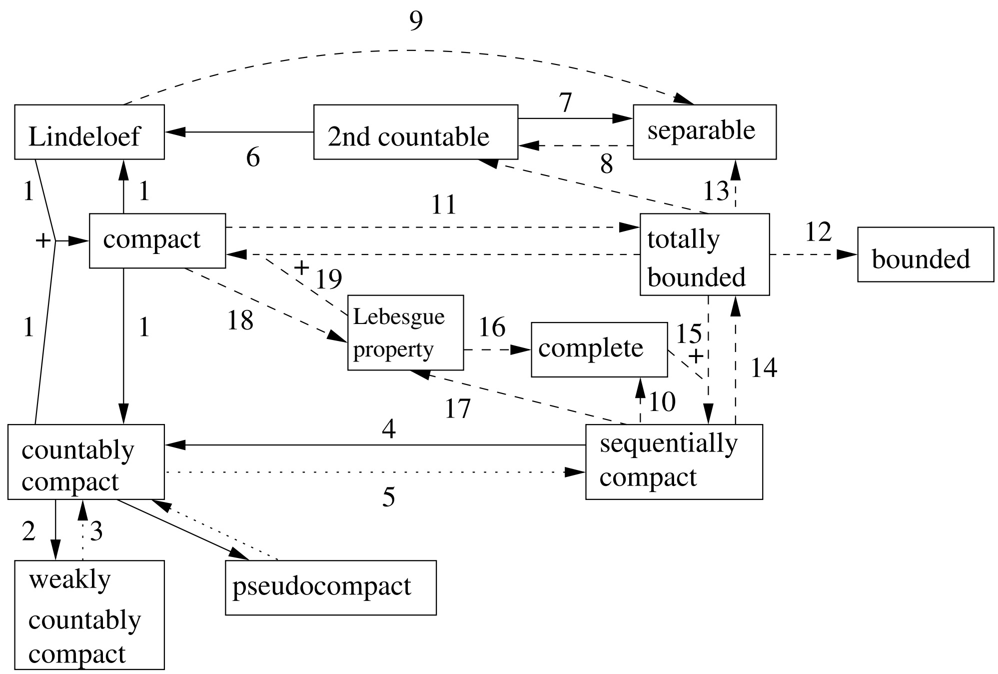


Figure 7.1: Implications of properties of metric spaces

15. By Lemma 7.7.20 every sequence has a Cauchy subsequence. By completeness, the latter converges. Thus the space is sequentially compact.

16. Let $\{x_i\}$ be a Cauchy sequence in X that does not converge. Then for every $x \in X$ there is an $\varepsilon_x > 0$ such that x_i frequently is not in $B(x, \varepsilon_x)$. Now $\mathcal{U} = \{B(x, \varepsilon_x)_{x \in X}$ is an open cover of X, which by assumption has a Lebesgue number $\lambda > 0$. Thus for every $y \in X$ there is an $x \in X$ such that $B(y, \lambda) \subseteq B(x, \varepsilon_x)$. This implies that x_i frequently is not in $B(y, \lambda)$. On the other hand, by Cauchyness there is $N \in \mathbb{N}$ such that $i, j \geq N$ implies $d(x_i, x_j) < \lambda$. In particular, $j \geq N \Rightarrow x_j \in B(x_N, \lambda)$, contradicting the fact that x_i frequently is not in $B(y, \lambda)$, for every y.

17. Assume there is an open cover $\mathcal{U}$ not admitting a Lebesgue number $\lambda > 0$. Then there are open sets of arbitrarily small diameter that are contained in no $U \in \mathcal{U}$. In particular, for each $n \in \mathbb{N}$ there is an x_n such that $B(x_n, 1/n) \not\subseteq U$ for all $U \in \mathcal{U}$. By sequential compactness, the sequence $\{x_n\}$ has an accumulation point x. Since $\mathcal{U}$ is a cover, we have $x \in U$ for some a $U \in \mathcal{U}$. Since U is open, there is $r > 0$ such that $B(x, r) \subseteq U$. Since x is an accumulation point of $\{x_n\}$, we can choose n such that $d(x, x_n) < r/2$ and $1/n < r/4$. If now $y \in B(x_n, 1/n)$ then $d(y, x) \leq d(y, x_n) + d(x_n, x) < \frac{1}{n} + \frac{r}{2} < \frac{r}{4} + \frac{r}{2} = \frac{3r}{4} < r$. Thus $B(x_n, 1/n) \subseteq U$, which is a contradiction with the choice of $\{x_n\}$.

18. For every $x \in X$, choose a $U_x \in \mathcal{U}$ such that $x \in U_x$. Since U_x is open, there is $\lambda_x > 0$ such that $B(x, 2\lambda_x) \subseteq U_x$. Now $\{B(x, \lambda_x)\}_{x \in X}$ is an open cover of X, so that by compactness there are $x_1, \ldots, x_n \in X$ such that $\bigcup_{k=1}^n B(x_k, \lambda_{x_k}) = X$. Let $\lambda := \min(\lambda_{x_1}, \ldots, \lambda_{x_n}) > 0$. Assuming now that $Y \subseteq X$ satisfies $\mathrm{diam}(Y) < \lambda$, pick any $y \in Y$ and a $k \in \{1, \ldots, n\}$ such that $y \in B(x_k, \lambda_{x_k})$. We claim the following inclusions

$$Y \subseteq B(y, \lambda) \subseteq B(x_k, 2\lambda_{x_k}) \subseteq U_{x_k}.$$

The first inclusion simply follows from $y \in Y$ and $\mathrm{diam}(Y) < \lambda$ and the third holds by the choice of λ_x. For the middle inclusion, note that $z \in B(y, \lambda)$ implies

$$d(z, x_k) \leq d(z, y) + d(y, x_k) < \lambda + \lambda_{x_k} \leq \lambda_{x_k} + \lambda_{x_k} = 2\lambda_{x_k}.$$

Since $U_{x_k} \in \mathcal{U}$, we are done.

19. Let $\mathcal{U}$ be an open cover of X. By assumption it has a Lebesgue number $\lambda > 0$. By total boundedness, the open cover $\{B(x, \lambda/3)\}_{x \in X}$ has a finite subcover $\{B(x_1, \lambda/3), \ldots, B(x_n, \lambda/3)\}$. In view of $\mathrm{diam}(B(x, \lambda/3)) \leq \frac{2}{3}\lambda$ and the definition of a Lebesgue number there is, for each $i = 1, \ldots, n$, a $U_i \in \mathcal{U}$ such that $B(x_i, \lambda/3) \subseteq U_i$. Now $\{U_i\}_{i=1}^n \subseteq \mathcal{U}$ is a finite subcover, thus X is compact. ■

In view of the implications proven above, the following is immediate:

Theorem 7.7.23 *For a metric space (X, d), the following are equivalent:*

(i) (X, τ_d) *is compact.*

(ii) (X, τ_d) *is countably compact.*

(iib) (X, τ_d) *is weakly countably compact.*

(iii) (X, τ_d) *is sequentially compact.*

(iv) (X, d) *is totally bounded and complete.*

(v) (X, d) *is totally bounded and has the Lebesgue property.*

Remark 7.7.24 1. Note that we have two ways of deducing compactness from the other statements, namely either from the combination of the Lebesgue property and total boundedness or from the combination of countable compactness and the Lindelöf property (deduced via separability from total boundedness). The first path is shorter, but apparently less well known.

2. In the above, the metric d was fixed. But as we know compactness is a topological notion, whereas a topology τ can be induced by different metrics (which then are called equivalent). In the light of this, the above means: If (X, τ) is a compact space then *every* compatible metric d is complete, totally bounded and Lebesgue. And if there is *some* compatible metric d that is totally bounded and complete then (X, τ) is compact.

3. The following table gives examples for spaces with all those combinatinations of properties that are not ruled out by the above results:

compact	tot.bdd.	Lebesgue	complete	2nd cnt.	Example
1	1	1	1	1	$[0,1]$ or finite discrete
0	1	0	0	1	$[0,1] \cap \mathbb{Q}$
0	0	1	1	1	$(\mathbb{N}, d_{\mathrm{disc}})$
0	0	1	1	0	X uncountable, d_{disc}
0	0	0	1	1	$\mathbb{R}$ with Eucl. metr.
0	0	0	1	0	$\ell^2(X), X$ uncountable
0	0	0	0	1	$\mathbb{Q}$ with Eucl. metr.
0	0	0	0	0	$\ell^2(X, \mathbb{Q}),\ X$ uncountable

(We will have more to say about the Lebesgue property in Section 7.7.4. For the Lebesgue property of (X, d_{disc}) cf. Remark 7.7.40.3.) □

The theorem has a number of corollaries that are immediate, but worth stating nevertheless:

Corollary 7.7.25 *Let (X, d) be a complete metric space.*

(i) If $Y \subseteq X$ then (Y, d) is compact if and only if it is closed and totally bounded.

(ii) $Y \subseteq X$ is relatively compact if and only it is totally bounded.

Proof. (i) If $Y \subseteq X$ is compact then it is closed (since metric spaces are Hausdorff) and totally bounded (by Theorem 7.7.23). Assume Y is closed and totally bounded. By Lemma 3.1.7(i), Y is complete. Thus it is compact by Theorem 7.7.23.

(ii) Since $Y \subseteq \overline{Y}$ is dense, Exercise 7.7.19 gives that $\overline{Y}$ is totally bounded if and only if Y is totally bounded. Now apply (i). ■

Corollary 7.7.26 *A metric space (X, d) is totally bounded if and only if its completion $(\widehat{X}, \widehat{d})$ is compact.*

Proof. By definition, $\widehat{X}$ is complete and $X \subseteq \widehat{X}$ is dense. If $\widehat{X}$ is compact, it is totally bounded, and so is X. If X is totally bounded then Corollary 7.7.25(ii) gives that $\widehat{X} = \overline{X}$ is compact. ■

Note that in view of the above, completion of a totally bounded metric space also is a compactification! For the purpose of later reference, we record the following fact used above:

Corollary 7.7.27 *Compact metrizable spaces are second countable, separable and Lindelöf.*

Remark 7.7.28 1. Now we can prove that $[0, 1]^2$ with the lexicographic order topology is not metrizable: By Exercise 7.6.4 this space is compact, but by Exercise 4.3.11 it is not second countable.

2. Later we will prove the following converse of Corollary 7.7.27: If X is compact, Hausdorff and second countable then X is metrizable. □

We note that completeness can be characterized without invoking Cauchy sequences, in a way that is very similar to the characterization of countable compactness in Exercise 7.7.5:

Lemma 7.7.29 *A metric space (X, d) is complete if and only if $\bigcap \mathcal{F} \neq \emptyset$ for every countable family $\mathcal{F}$ of closed sets having the finite intersection property and satisfying $\inf_{C \in \mathcal{F}} \operatorname{diam}(C) = 0$.*

Proof. Given a countable family $\mathcal{F}$, we choose a bijection $\mathbb{N} \to \mathcal{F}$, $n \mapsto C_n$. Then the sets $D_n = \bigcap_{k=1}^n C_n$ are closed, decreasing, non-empty (by the finite intersection property) and satisfy $\operatorname{diam}(D_n) \to 0$. The rest follows from Exercise 3.1.6. ■

7.7.4 Compactness of metric spaces II: Applications

In this section, we consider several applications of compactness in the context of metric spaces. Recall from Remark 5.2.12 that the results of Propositions 2.1.25, 3.1.8 adapt to the case where only (Y, d) is metric, but (X, τ) is a topological space. Thus we still have a natural topology τ_D on the set $C_b(X, Y)$, induced by the complete metric D of uniform convergence.

The first is result is straight-forward:

Corollary 7.7.30 *Let X be a compact topological space. Then:*

(i) Every continuous function $f : X \to Y$ with Y metric is bounded. Thus $C(X, Y) = C_b(X, Y)$, and $(C(X, Y), D)$ is a metric space.

(ii) Every continuous function $f : X \to \mathbb{R}$ is bounded and assumes its infimum and supremum.

Proof. (i) By Lemma 7.3.6, $f(X) \subseteq Y$ is compact, thus closed and bounded by Corollary 7.7.25. The rest was already proven in Proposition 3.1.8.

(ii) For a bounded $\mathbb{R}$-valued function, we have $\inf f, \sup f \in \overline{f(X)}$. Now the claim follows from the closedness of $f(X)$ given by (i). ■

Definition 7.7.31 *A topological space X is called pseudocompact if every $f \in C(X, \mathbb{R})$ is bounded.*

Thus compact ⇒ pseudocompact. In fact, there is a better result:

Exercise 7.7.32 (i) Prove that every continuous function from a countably compact space to a metric space is bounded.

(ii) Deduce that countably compact ⇒ pseudocompact.

Proposition 7.7.33 *For a topological space X, consider the following statements:*

(i) For every countable family $\mathcal{F}$ of open sets with the finite intersection property, $\bigcap\{\overline{U} \mid U \in \mathcal{F}\} \neq \emptyset$ holds.

(ii) For every sequence $\{U_n\}_{n\in\mathbb{N}}$ of non-empty open sets satisfying $U_{n+1} \subseteq U_n$ $\forall n$ we have $\bigcap_n \overline{U_n} \neq \emptyset$.

(iii) X is pseudocompact.

Then (i)⇔(ii)⇒(iii). For completely regular spaces, also (iii)⇒(ii).

Exercise 7.7.34 Prove the implications (i)⇔(ii)⇒(iii) in Proposition 7.7.33.

Remark 7.7.35 1. The implication (iii)⇒(ii) in Proposition 7.7.33 will will be proven in Proposition 8.2.52, together with a few more equivalent statements. Since the proof proceeds by by contradiction, one must produce an unbounded continuous function from the given information. This requires the theory concerning the existence of continuous real-valued functions discussed in Sections 8.2 and 8.3.1. In view of this, one could argue that complete regularity should be included in the definition of pseudocompactness, but I don't like such combined definitions.

2. We will see that every pseudocompact T_4-space is countably compact, cf. Lemma 8.2.28. Since metric spaces are $T_{3.5}$ and T_4, we can add pseudocompactness with its various equivalent characterizations to the list of equivalent properties of metric spaces given in Theorem 7.7.23. □

Corollary 7.7.30 has a generalization to semicontinuous functions:

Exercise 7.7.36 Let X be countably compact. Prove:

(i) If $f : X \to \mathbb{R}$ is lower semicontinuous then it is bounded below and assumes its infimum.

(ii) If $f : X \to \mathbb{R}$ is upper semicontinuous then it is bounded above and assumes its supremum.

Other applications of compactness include proofs of uniform continuity in various guises and uniform convergence. The following generalizes a classical result of Dini[8] for functions on a bounded interval:

[8] Ulisse Dini (1845-1918), Italian mathematician. Also known for a criterion for the convergence of Fourier series.

Proposition 7.7.37 (Dini's theorem) *Let X be countably compact, $\{f_n\}_{n\in\mathbb{N}} \subseteq C(X,\mathbb{R})$ and $g \in C(X,\mathbb{R})$ such that $f_n(x) \nearrow g(x)\ \forall x$ (pointwise monotone convergence). Then the convergence $f_n \to g$ is uniform, i.e. $\|f_n - g\| \equiv \sup_{x\in X} |f_n(x) - g(x)| \to 0$.*

Proof. For $\varepsilon > 0$ and $n \in \mathbb{N}$, define $U_n(\varepsilon) = \{x \in X \mid g(x) - f_n(x)| < \varepsilon\}$, which is open by continuity of $g - f_n$. Let $\varepsilon > 0$. Pointwise convergence implies that every $x \in X$ is contained in some $U_n(\varepsilon)$, thus $\bigcup_{n\in\mathbb{N}} U_n(\varepsilon) = X$. Countable compactness implies the existence of a finite subcover $\{U_{n_1}(\varepsilon), \ldots, U_{n_m}(\varepsilon)\}$. Since the convergence $f_n \to g$ is monotone, the $U_n(\varepsilon)$ are increasing with n. Thus with $N = \max(n_1, \ldots, n_m)$ we have $U_N(\varepsilon) = X$, which is equivalent to $\|g - f_N\| < \varepsilon$. Since $\varepsilon > 0$ was arbitrary, we have uniform convergence. ■

It should be known that a continuous function $f : [a,b] \to \mathbb{R}$ is uniformly continuous (Definition 3.4.12), cf. e.g. [223, Theorem 9.9.16]. (This was first proven in 1852 by Peter Gustav Lejeune Dirichlet (1805-1859).) Using the Lebesgue property, this can be generalized considerably:

Proposition 7.7.38 *Let $(X,d), (Y,d')$ be metric spaces, where (X,d) has the Lebesgue property (e.g. due to compactness). Then every continuous $f : X \to Y$ is uniformly continuous.*

Proof. Let $\varepsilon > 0$. Since f is continuous, for every $x \in X$ there is $\delta_x > 0$ such that $d(x,y) < \delta_x$ implies $d'(f(x), f(y)) < \varepsilon/2$. Now $\mathcal{U} = \{B(x,\delta_x)\}_{x\in X}$ is an open cover of X. Let $\lambda > 0$ be a Lebesgue number of this cover. If now $d(y,z) < \lambda$ then $\mathrm{diam}(\{y,z\}) < \lambda$, so that $\{y,z\} \subseteq B(x,\delta_x)$ for some $x \in X$. With the triangle inequality we have

$$d'(f(y), f(z)) \le d'(f(y), f(x)) + d'(f(x), f(z)) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,$$

where we used the fact that y and z are in $B(x,\delta_x)$, and the definition of δ_x. ■

The converse of Proposition 7.7.38 is also true, thus if every continuous $f : (X,d) \to (Y,d')$ is uniformly continuous then (X,d) has the Lebesgue property. Remarkably, these two equivalent properties are equivalent to many others. We need the following definition, which should be compared to that of discreteness of a metric (Exercise 2.3.3):

Definition 7.7.39 *A metric d on a set X is called uniformly discrete if there is $\varepsilon > 0$ such that $d(x,y) \ge \varepsilon$ whenever $x \ne y$.*

Remark 7.7.40 1. The standard discrete metric of Example 2.1.8 is uniformly discrete.

2. There are discrete metrics that are not uniformly discrete. E.g. $d(n,m) = \left|\frac{1}{n} - \frac{1}{m}\right|$ on $X = \mathbb{N}$.

3. If d is a uniformly discrete metric on X, then $\lambda = \varepsilon$ is a Lebesgue number for every open cover $\mathcal{U}$ since $\mathrm{diam}(Y) < \varepsilon$ implies that Y is a singleton. Thus (X,d) has the Lebesgue property. One also sees directly that *every* function $f : (X,d) \to (Y,d')$ is uniformly continuous for the trivial reason that $d(x,y) < \varepsilon$ implies $x = y$, thus $d'(f(x), f(y)) = 0$. □

Theorem 7.7.41 *For a metric space (X,d), the following are equivalent:*

(i) X has the Lebesgue property.

(ii) Every continuous $f : X \to (Y,d')$ with (Y,d') metric is uniformly continuous.

(iii) Every continuous $f : X \to \mathbb{R}$ is uniformly continuous (w.r.t. the standard metric on $\mathbb{R}$).

(iv) Every cover of X by two open sets has positive Lebesgue number.

(v) If $A \subseteq U \subseteq X$ with A closed and U open, there is $\varepsilon > 0$ such that $A_\varepsilon \subseteq U$.

(vi) If $A, B \subseteq X$ are non-empty closed subsets with $A \cap B = \emptyset$ then $\text{dist}(A, B) > 0$.

(vii) Every closed discrete subset $Y \subseteq X$ is uniformly discrete.

(viii) The derived set X' (i.e. set of non-isolated points) is compact and $X \backslash (X')_\varepsilon$ is uniformly discrete for every $\varepsilon > 0$.

Metric spaces satisfying these conditions are complete and have been called Lebesgue-spaces, UC-spaces or Atsuji-spaces.

Exercise 7.7.42 Prove the equivalence of statements (iv), (v), (vi) in Theorem 7.7.41.

Remark 7.7.43 1. Statement (vi) shows that the metric spaces with the Lebesgue property are precisely those where we cannot have the phenomenon encountered in Exercise 2.1.20(v), namely non-empty closed subsets A, B such that $\text{dist}(A, B) = 0$ and $A \cap B = \emptyset$.

2. That the Lebesgue property implies completeness was 16. in Proposition 7.7.22. We only give the simpler of the remaining proofs.

The implicaton (i)⇒(ii) was Proposition 7.7.38, and (ii)⇒(iii) is obvious, as is (i)⇒(iv).

(iii)⇒(vi) Define $f : X \to [0, 1]$ by $f(x) = \text{dist}(x, A)/(\text{dist}(x, A) + \text{dist}(x, B))$. One checks easily that f is continuous and $f \restriction A = 0$, $f \restriction B = 1$. By (iii), f is uniformly continuous, thus there is $\varepsilon > 0$ such that $d(x, y) < \varepsilon$ implies $|f(x) - f(y)| < 1/2$. Combined with $f \restriction A = 0$, this implies that $f(x) \leq 1/2 \ \forall x \in A_\varepsilon$, which together with $f \restriction B = 1$ gives $\text{dist}(A, B) \geq \varepsilon$.

For proofs of the remaining implications (vi)⇒(vii) ⇒(viii) ⇒(i) (and many others) see [150], where about 30 equivalent conditions are discussed! □

Corollary 7.7.44 *A metric space having at most finitely many isolated points has the (equivalent) properties listed in Theorem 7.7.41 if and only if it is compact.*

Proof. Every compact metric space has the Lebesgue property. Conversely, the Lebesgue property implies compactness of X'. If the set Y of isolated points is finite then it is not only open but also closed. Thus $X \cong X' \oplus Y$, implying compactness of X. ■

We now consider a converse of the result in Exercise 6.5.40:

Exercise 7.7.45 Let X be arbitrary, Y compact and (Z, d) metric. Topologize $C(Y, Z)$ using the metric $D(f, g) = \sup_y d(f(y), g(y))$.

(i) For $f \in C(X \times Y, Z)$, let $F = \Lambda(f) \in \text{Fun}(X, \text{Fun}(Y, Z))$. Prove $F \in C(X, C(Y, Z))$. (Thus if $x \to x_0$ then $\sup_y d(f(x, y), f(x_0, y)) \to 0$, i.e. $f(x, y)$ is continuous in x uniformly in y.) Hint: Use the Slice Lemma 7.5.1.

(ii) Combining (i) with Exercise 6.5.40(ii), conclude that Λ gives a bijection $C(X \times Y, Z) \to C(X, C(Y, Z))$.

(iii) Assuming in addition that X is compact with metric d_X, combine (ii) with Proposition 7.7.38 to show for every $f \in C(X \times Y, Z)$ that

$$\forall \varepsilon > 0 \ \exists \delta > 0 : \ d_X(x, x') < \delta \Rightarrow \forall y \in Y : d(f(x, y), f(x', y)) < \varepsilon.$$

(iv) Assuming again that X is compact and equipping $C(X \times Y, Z)$ and $C(X, C(Y, Z))$ with the natural metrics, prove that Λ is an isometry. (This is independent of (iii).)

The following will be useful later:

Lemma 7.7.46 *Let (X, d) be a metric space and $\{K_i \subseteq X\}_{i \in I}$ compact subsets. Let $\{S_k \subseteq X\}_{k \in \mathbb{N}}$ such that $\operatorname{diam}(S_k) \stackrel{k\to\infty}{\longrightarrow} 0$ and for all $k \in \mathbb{N},\ i \in I$ one has $S_k \cap K_i \neq \emptyset$. Then $\bigcap_i K_i \neq \emptyset$.*

Proof. It is sufficient to prove the claim in the case where I is finite. In the general case, this then implies that the family $\{K_i\}_{i \in I}$ has the finite intersection property and another invocation of compactness gives $\bigcap_i K_i \neq \emptyset$. Thus let $\{K_1, \ldots, K_n\}$ be given and consider $K = \prod_i K_i$ equipped with the metric $d_K(x, y) = \sum_i d(x_i, y_i)$. For every $k \in \mathbb{N}$ and $i \in \{1, \ldots, n\}$, choose an $x_{k,i} \in S_k \cap K_i$ and define $x_k = (x_{k,1}, \ldots, x_{k,n}) \in K$. By compactness of K there exists a point $z = (z_1, \ldots, z_n) \in K$ every neighborhood of which contains x_k for infinitely many k. Now, $d(z_i, z_j) \leq d(z_i, x_{k,i}) + d(x_{k,i}, x_{k,j}) + d(x_{k,j}, z_j) \leq 2d_K(z, x_k) + \mathrm{diam(S_k)}$. Since by construction every neighborhood of z contains points x_k with arbitrarily large k, we can make both terms on the r.h.s. arbitrarily small and conclude that $z = (x, \ldots, x)$ for some $x \in X$. Since $z_i \in K_i$ for all i, we have $x \in \bigcap_i K_i$, and are done. ■

7.7.5 Subsets of $\mathbb{R}^n$ I: Compactness

By Corollary 7.7.25, for a complete metric space (X, d) and $Y \subseteq X$ we have: Y totally bounded $\Rightarrow$ $\overline{Y}$ compact $\Rightarrow$ Y totally bounded $\Rightarrow$ Y bounded. For $X = \mathbb{R}^n$ we have more:

Lemma 7.7.47 *Let $n \in \mathbb{N}$. Every bounded subset $X \subseteq \mathbb{R}^n$ is totally bounded.*

Proof. If $X \subseteq \mathbb{R}^n$ is bounded, it is contained in some cube $C = [-a, a]^n$. Now, for every $\varepsilon > 0$, the ball $B(x, \varepsilon)$ contains a closed cube of some edge $M > 0$. Now it is clear that C can be covered by $\lceil 2a/M \rceil^n$ such balls. Thus C is totally bounded, thus also X by Exercise 7.7.19(i). (Alternatively, use this argument to show that $[-a, a]$ is totally bounded, thus compact and deduce compactness of C from Tychonov's theorem.) ■

Theorem 7.7.48 (Heine-Borel) [9] *A subspace of $K \subseteq \mathbb{R}^n$ is compact if and only if it is closed and bounded.*

Proof. Taking Lemma 7.7.47 into account, this follows from Corollary 7.7.25. ■

Remark 7.7.49 It is important to understand that the Heine-Borel theorem does not generalize to most metric spaces! To see this it suffices to take any non-compact metric space (X, d) like $\mathbb{R}^n$ and replace the metric d by an equivalent bounded metric d'. Now X is obviously closed and bounded with respect to d', but it still is non-compact. Metric spaces to which the Heine-Borel result does generalize are considered in Section 7.8.8. □

Exercise 7.7.50 Use Theorem 7.7.48 to prove that the sphere S^n is compact for every $n \in \mathbb{N}$.

In Exercise 2.2.16, we proved that the norms $\|\cdot\|_s$ on $\mathbb{R}^n$ defined in Example 2.1.13 are equivalent for all $s \in [1, \infty]$. Using the Heine-Borel theorem we can improve this considerably:

[9] Heinrich Eduard Heine (1821-1881), Émile Borel (1871-1956). This was proven by Borel for countable covers and by Lebesgue in generality. Heine had little to do with it! The equivalent (since $\mathbb{R}^n$ is metric) statement that one obtains replacing compactness by sequential compactness is known as the Bolzano-Weierstrass theorem.

Theorem 7.7.51 *All norms on a finite dimensional vector space over $\mathbb{R}$ or $\mathbb{C}$ are equivalent.*

Proof. Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. Let $B = \{e_1, \ldots, e_d\}$ be a basis for V, and define the Euclidean norm $\|\cdot\|_2$ of $x = \sum_i c_i e_i$ by $\|x\|_2 = (\sum_i |c_i|^2)^{1/2}$. It clearly is sufficient to show that any norm $\|\cdot\|$ is equivalent to $\|\cdot\|_2$. Using $|c_i| \le \|x\|_2 \ \forall i$ and the properties of any norm, we have

$$\|x\| = \left\| \sum_{i=1}^d c_i e_i \right\| \le \sum_{i=1}^d |c_i| \, \|e_i\| \le \left(\sum_{i=1}^d \|e_i\| \right) \|x\|_2. \tag{7.6}$$

This implies that $x \mapsto \|x\|$ is continuous w.r.t. the topology on V defined by $\|\cdot\|_2$. Since the sphere $S = \{x \in \mathbb{F}^n \mid \|x\|_2 = 1\}$ is compact by Exercise 7.7.50, Corollary 7.7.30 implies that there is $z \in S$ such that $\lambda := \inf_{x \in S} \|x\| = \|z\|$. Since $z \in S$ implies $z \ne 0$ and $\|\cdot\|$ is a norm, we have $\lambda = \|z\| > 0$. Now, for $x \ne 0$ we have $\frac{x}{\|x\|_2} \in S$, and thus

$$\|x\| = \|x\|_2 \left\| \frac{x}{\|x\|_2} \right\| \ge \|x\|_2 \lambda. \tag{7.7}$$

Combining (7.6, 7.7), we have $c_1 \|x\|_2 \le \|x\| \le c_2 \|x\|_2$ with $0 < c_1 = \inf_{x \in S} \|x\| \le \sum_i \|e_i\| = c_2$. (Note that $e_i \in S \ \forall i$, so that $c_2 \ge d c_1$, showing again that V must be finite dimensional.) ■

Corollary 7.7.52 *Let $(V, \|\cdot\|)$ be a normed vector space over $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and $V' \subseteq V$ a finite dimensional subspace. Then the restriction of $\|\cdot\|$ to V' is equivalent to any other norm on V' and is complete, and $V' \subseteq V$ is closed.*

Proof. The first claim is immediate by Theorem 7.7.51. The second follows, since every norm on a finite dimensional vector space is equivalent to the Euclidean one, thus complete. The last claim results from the fact that complete subspaces of metric spaces are closed. ■

Remark 7.7.53 On a purely topological level, one can prove that a finite dimensional vector space over $\mathbb{R}$ or $\mathbb{C}$ has precisely one topology making it a topological vector space (meaning that the abelian group structure of $(V, +, 0)$ and the action of the ground field are continuous). □

Lemma 7.7.54 (F. Riesz) [10] *Let $(V, \|\cdot\|)$ be a normed space and $W \subsetneq V$ a closed proper subspace. Then for each $\delta \in (0,1)$ there is an $x_\delta \in V$ such that $\|x_\delta\| = 1$ and $\mathrm{dist}(x_\delta, W) \ge \delta$, i.e. $\|x_\delta - x\| \ge \delta \ \forall x \in W$.*

Proof. If $x_0 \in V \backslash W$ then $\lambda = \mathrm{dist}(x_0, W) > 0$ by Exercise 2.1.20(iii). In view of $\delta \in (0,1)$, we have $\frac{\lambda}{\delta} > \lambda$. Thus we can choose $y_0 \in W$ with $\|x_0 - y_0\| < \frac{\lambda}{\delta}$. Putting

$$x_\delta = \frac{y_0 - x_0}{\|y_0 - x_0\|},$$

we have $\|x_\delta\| = 1$. If $x \in W$ then

$$\|x - x_\delta\| = \left\| x - \frac{y_0 - x_0}{\|y_0 - x_0\|} \right\| = \frac{\| \|y_0 - x_0\| x - y_0 + x_0 \|}{\|y_0 - x_0\|} \ge \frac{\mathrm{dist}(x_0, W)}{\|y_0 - x_0\|} \ge \frac{\lambda}{\frac{\lambda}{\delta}} = \delta,$$

where the first $\ge$ is due to $\|y_0 - x_0\| x - y_0 \in W$ and the second $\ge$ is due to $\|x_0 - y_0\| < \frac{\lambda}{\delta}$. Since $x \in W$ was arbitrary, we are done. ■

[10] Frigyes Riesz (1880-1956), Hungarian mathematician.

Theorem 7.7.55 *If $(V, \|\cdot\|)$ is an infinite dimensional normed space then:*

(i) Each closed ball $\overline{B}(x,r)$ (with $r>0$) is non-compact.

(ii) Every subset $Y \subseteq V$ with non-empty interior Y^0 is non-compact.

Proof. (i) Choose $x_1 \in V$ with $\|x_1\| = 1$. Then $\mathbb{C}x_1$ is a closed proper subspace, thus there exists $x_2 \in V$ with $\|x_2\| = 1$ and $\|x_1 - x_2\| \geq \frac{1}{2}$. Since V is infinite dimensional, $V_2 = \text{span}\{x_1, x_2\}$ is a closed proper subspace, thus there exists $x_3 \in V$ with $\text{dist}(x_3, V_2) \geq \frac{1}{2}$, thus in particular $\|x_3 - x_i\| \geq \frac{1}{2}$ for $i = 1, 2$. Continuing in this way we can construct a sequence of $x_i \in V$ with $\|x_i\| = 1$ and $\|x_i - x_j\| \geq \frac{1}{2}$ $\forall i \neq j$. The sequence $\{x_i\}$ clearly cannot have a convergent subsequence, thus the closed unit ball $\overline{B}(0,1)$ is non-compact. Since $x \mapsto \lambda x + x_0$ is a homeomorphism, all closed balls are non-compact.

(ii) If $Y \subseteq V$ and $Y^0 \neq \emptyset$ then Y contains some open ball $B(x,r)$, thus also $\overline{B}(x, r/2)$, which is non-compact. Thus neither Y nor $\overline{Y}$ are compact. ■

Remark 7.7.56 1. With the notion of local compactness that we will encounter in Section 7.8, the above results can be simply stated thus: A normed vector space is locally compact if and only if it is finite dimensional.

2. The above negative result means that compact sets in infinite dimensional normed spaces must have empty interior. In Sections 7.7.7-E.3 we will characterize the compact subsets of certain function spaces.

3. While closed balls in infinite dimensional normed spaces are non-compact w.r.t. the norm topology, they actually are compact w.r.t. the weak topology. This is the content of Alaoglu's theorem, cf. Section G.3. □

The following is probably the simplest proof of the algebraic closedness of the field of complex numbers. The proof will take for granted that every complex number w has an n-th root $w^{1/n}$ for every n. (Thus equations of the form $z^n - w = 0$ always have solutions, which is a special case of the result to be proven here. This fact will be proven in Corollary 9.2.20, using the intermediate value theorem.)

Theorem 7.7.57 (Fundamental Theorem of Algebra) [11] *Let $P \in \mathbb{C}[z]$ be complex polynomial of degree $n \geq 1$, i.e. $P(z) = a_n z^n + \cdots + a_1 z + a_0$, where $a_n \neq 0$. Then there is a $z \in \mathbb{C}$ such that $P(z) = 0$.*

Proof. We may assume that $a_n = 1$. Then $P(z) = z^n(1 + a_{n-1}z^{-1} + \cdots + a_0 z^{-n})$ implies that $|P(z)| \to \infty$ as $|z| \to \infty$. Thus there is a $C > 0$ such that $|z| > C$ implies $P(z) \neq 0$. Since $z \mapsto |P(z)|$ is continuous and $\{z \in \mathbb{C} \mid |z| \leq C\}$ is compact by Theorem 7.7.48, Corollary 7.7.30 implies that there is a $z_0 \in \mathbb{C}$ where $|P|$ assumes its infimum. If the latter is zero, we are done. Thus assume $|P(z_0)| = \inf |P(z)| > 0$. Putting $Q(z) = P(z+z_0)/P(z_0)$, we have $Q(z) = 1 + b_p z^p + \cdots + b_n z^n$, where $p = \min\{m \geq 1 \mid a_m \neq 0\}$. Since $b_p = a_p / P(z_0) \neq 0$, by existence of n-th roots there is an $z \in \mathbb{C}$ such that $z^p = -1/b_p$. With $0 < r < 1$ we have

$$\begin{aligned} Q(rz) &= 1 + r^p z^p b_p + r^{p+1} z^{p+1} b_{p+1} + \cdots + r^n z^n b_n \\ &= (1 - r^p) + r^{p+1} z^{p+1} b_{p+1} + \cdots + r^n z^n b_n, \end{aligned}$$

[11] This is a misnomer in two respects: On the one hand, the complex numbers do not have a very central place in modern algebra. On the other hand, all proofs make essential use of some continuity considerations and therefore are not purely algebraic.

thus
$$|Q(rz)| \leq 1 - r^p + r^{p+1}|z^{p+1}b_{p+1} + rz^{p+2}b_{p+1} + \cdots + r^{n-p-1}z^n b_n|.$$
For $r > 0$ small enough, the term with r^{p+1} is smaller than r^p, so that $|Q(rz)| < 1$. But this contradicts the fact that by construction $|Q|$ has its absolute minimum 1 at $z = 0$. ■

For more on roots of complex polynomials see Exercise 7.8.71.

7.7.6 Subsets of $\mathbb{R}^n$ II: Convexity

Definition 7.7.58 *Let V be an $\mathbb{R}$-vector space. A subset $X \subseteq V$ is called convex if $x, y \in X$, $\lambda \in [0,1] \Rightarrow \lambda x + (1-\lambda)y \in X$. (I.e., the straight line segment $\overline{xy}$ is contained in X.)*

Clearly the Euclidean balls $B^n, D^n \subseteq \mathbb{R}^n$ are convex. This can be generalized:

Lemma 7.7.59 *If $\|\cdot\|$ is any norm on a vector space V then the balls $B(x,r) = \{y \in V \mid \|x-y\| < r\}$ and $\overline{B}(x,r) = \{y \in V \mid \|x-y\| \leq r\}$ are convex.*

Proof. Since the metric $d(x,y) = \|x-y\|$ is translation invariant, we have $B(x,r) = x + B(0,r)$, thus it suffices to prove the claim for $x = 0$. If $x, y \in B(0,r)$ and $t \in [0,1]$ then
$$\|tx + (1-t)y\| \leq \|tx\| + \|(1-t)y\| = t\|x\| + (1-t)\|y\| < tr + (1-t)r = r,$$
thus $tx + (1-t)y \in B(0,r)$ $\forall t \in [0,1]$. Similarly for $\overline{B}(x,r)$. ■

Lemma 7.7.60 *Let $K \subseteq \mathbb{R}^n$ be compact and convex with non-empty interior K^0. Assume $0 \in K^0$ (as can be achieved by translation, if necessary). Then the map*
$$f : \partial K \to S^{n-1}, \quad x \mapsto \frac{x}{\|x\|},$$
where $\|\cdot\|$ is the Euclidean norm, is a homeomorphism.

Proof. The map $\mathbb{R}^n \backslash \{0\} \to S^{n-1}$, $x \mapsto \frac{x}{\|x\|}$ is continuous, thus its restriction f to $\partial K \subseteq \mathbb{R}^n \backslash \{0\}$ is continuous. We claim that $f : \partial K \to S^{n-1}$ is a bijection. Assuming this for a moment, Proposition 7.4.11 implies that f is a homeomorphism since ∂K is compact (as a closed subset of the compact space K) and S^{n-1} is Hausdorff.

To prove surjectivity, let $z \in S^{n-1}$. Since K is compact, thus bounded, we have $\mu = \sup\{\lambda \geq 0 \mid \lambda z \in K\} < \infty$. Now $0 \in K^0$ implies $\mu > 0$, closedness of K implies $\mu z \in K$, and by construction every neighborhood of μz contains points in $\mathbb{R}^n \backslash K$. Thus $\mu z \in \partial K$. Since $f(\mu z) = z$, we have surjectivity of $f : \partial K \to S^{n-1}$.

It remains to prove injectivity of f, which clearly is equivalent to the statement that every ray beginning at 0 intersects ∂K in at most one point. Let R be such a ray and let $p, q \in K \cap R$ with $0 < \|p\| < \|q\|$. Since $0 \in K^0$, there is some closed ball B with center 0 contained in K^0. Consider the union of all line segments from q to a point in B. Since K is convex, this set is contained in K and it contains p in its interior. Thus $p \in K^0$ and therefore $p \notin \partial K$. Thus R intersects ∂K in at most one point, which is equivalent to injectivity of $f : \partial K \to S^{n-1}$. ■

It seems likely that one could also show the openness of f using the information about the geometry of K, but invoking compactness surely is less painful.

The following result was already used in the discussion of the real projective spaces. In stating it, we appeal to the Heine-Borel Theorem 7.7.48 to write 'closed bounded' instead of 'compact', in order to make the geometric assumptions more explicit:

Proposition 7.7.61 *Let $K \subseteq \mathbb{R}^n$ be convex, closed and bounded with non-empty interior. Then:*

(i) There is a homeomorphism $g : D^n \to K$ restricting to $S^{n-1} = \partial D^n \xrightarrow{\cong} \partial K$.

(ii) If $x_0 \in K^0$ then the homeomorphism $g : D^n \to K$ can be chosen such that $g(0) = x_0$.

(ii) If $-K = K$ then we can choose g such that $g(-x) = -g(x)\ \forall x$, i.e. g is $\underline{\mathbb{Z}_2\text{-equivariant}}$.

Proof. (i) We may assume $0 \in K^0$. Let $f : \partial K \to S^{n-1}$ be the homeomorphism constructed in the lemma. Defining

$$g : D^n \to K, \quad x \mapsto \begin{cases} \|x\|\, f^{-1}\left(\frac{x}{\|x\|}\right) & x \neq 0 \\ 0 & x = 0 \end{cases}$$

it should be clear that g is a bijection from D^n to K. At $x \neq 0$, g clearly is continuous. Since K is compact, there is $M > 0$ such that $\|x\| \leq M\ \forall x \in K$. Thus $\|f^{-1}(y)\| \leq M\ \forall y \in S^{n-1}$ and $\|g(x)\| \leq M\|x\|$. This tends to zero as $x \to 0$, so g is continuous at zero. As a continuous bijection between compact Hausdorff spaces, g is a homeomorphism by Proposition 7.4.11.

(ii) If $x_0 \in K^0$ then defining $K' = K - x$ we have $0 \in K'^0$. Thus (i) gives us a homeomorphism $g_0 : D^n \to K'$ sending $0 \in D^n$ to $0 \in K'$. Now $g := g_0 + x_0$ is a homeomorphism $D^n \to K$ sending 0 to x_0.

(iii) The assumption $K = -K$ together with convexity implies $0 \in K$, so that we do not need to shift K. It is obvious that the map $f : x \mapsto x/\|x\|$ in Lemma 7.7.60 is equivariant, and then the same holds for g. ■

This result is very useful since it automatically provides us with homeomorphisms between all compact convex subsets of $\mathbb{R}^n$ that have non-empty interior, e.g. $D^n \cong I^n \cong P$, where $P \subseteq \mathbb{R}^n$ is any (full) convex polyhedron. E.g., a tetrahedron $T \subseteq \mathbb{R}^3$ is homeomorphic to the cube I^3. Constructing such homeomorphisms directly would be quite painful.

What about dropping the assumption on the interior?

Proposition 7.7.62 *For a compact convex subset $K \subseteq \mathbb{R}^n$, the following are equivalent:*

(i) K has non-empty interior.

(ii) K is not contained in a proper hyperplane (i.e. a set $x_0 + V$ of K, where $V \subseteq \mathbb{R}^n$ is a subspace of dimension $< n$).

(iii) $\mathrm{span}_{\mathbb{R}}\{x - y \mid x, y \in K\} = \mathbb{R}^n$.

Proof. The equivalence (ii)$\Leftrightarrow$(iii) is obvious. Assume (i). If $x_0 \in K^0$ and $\varepsilon > 0$ is such that $\overline{B_\varepsilon(x_0))} \subseteq K$, it is clear that the vectors $x - x_0$ where $\|x - x_0\| = \varepsilon$ span $\mathbb{R}^n$. Thus (iii) holds. Now assume (ii) and choose $x_0 \in K$. Now choose $x_1 \in K \backslash \{x_0\}$, and then $x_2 \in K$ but not in the line determined by x_0 and x_1. Going on like this we obtain $x_0, x_1, \ldots$ in K such that no point lies in the hyperplane defined by the others. In view of (ii), this process continues until we have found x_n. By convexity of K it is clear that the n-simplex

$$S = \left\{ \sum_{i=0}^{n} \lambda_i x_i \mid \lambda_i \geq 0,\ \sum_i \lambda_i = 1 \right\}$$

is contained in K. It is easy to see that

$$S^0 = \left\{ \sum_{i=0}^{n} \lambda_i x_i \mid \lambda_i > 0,\ \sum_i \lambda_i = 1 \right\},$$

and since this is contained in K^0, K has non-empty interior. ■

Theorem 7.7.63 *Let $K \subseteq \mathbb{R}^n$ be non-empty, convex, closed and bounded. Then there is $m \leq n$ such that K is homeomorphic to D^m.*

Proof. We may assume $0 \in K$, replacing K by $K - x_0$, where $x_0 \in K$, if necessary. Define $Y = \mathrm{span}_{\mathbb{R}}(K)$. Now K is a convex subset of the subspace $Y \subseteq \mathbb{R}^n$ and by construction it is not contained in a proper hyperplane of Y. By Proposition 7.7.62, it has non-empty interior, considered as a subspace of Y. Now Proposition 7.7.61 applies, and with $m = \dim Y$ we have $K \cong D^m$. ■

Remark 7.7.64 Later (Corollary 10.4.7) we will prove that D^r and D^s are non-homeomorphic when $r \neq s$. Thus the m in the theorem is uniquely determined by K. In particular, $m = n$ holds if and only if K has non-empty interior. □

7.7.7 ⋆ Compactness in function spaces I: Ascoli-Arzelà theorems

If (X, τ) is a topological space and (Y, d) metric, the set $C_b(X, Y)$ is topologized by the metric D from (2.6), cf. Remark 5.2.12. It is therefore natural to ask whether the (relative) compactness of a set $\mathcal{F} \subseteq C_b(X, Y)$ can be characterized in terms of the elements of $\mathcal{F}$, which after all are functions $f : X \to Y$. This will be the subject of this section, but we will restrict ourselves to compact X, for which $C(X, Y) = C_b(X, Y)$ by Corollary 7.7.30.

Definition 7.7.65 *Let (X, τ) be a topological space and (Y, d) a metric space. A family $\mathcal{F} \subseteq \mathrm{Fun}(X, Y)$ is called* <u>equicontinuous</u> *if for every $x \in X$ and $\varepsilon > 0$ there is an open neighborhood $U \ni x$ such that $\overline{f \in \mathcal{F},\ x' \in U} \Rightarrow d(f(x), f(x')) < \varepsilon$.*

This clearly implies $\mathcal{F} \subseteq C(X, Y)$, but the point is that the choice of U depends only on x and ε, but works for every $f \in \mathcal{F}$.

The following lemma is from [105]. Note that the $\varepsilon - \delta$ condition appearing there could be called 'uniform metric properness' of f since it says that preimages of bounded sets are bounded in a uniform way.

Lemma 7.7.66 *Let (X, d) be a metric space. Assume that for each $\varepsilon > 0$ there are a $\delta > 0$, a metric space (Y, d') and a continuous $h : X \to Y$ such that $(h(X), d')$ is totally bounded and such that $d'(h(x), h(x')) < \delta$ implies $d(x, x') < \varepsilon$. Then (X, d) is totally bounded.*

Proof. For $\varepsilon > 0$, pick $\delta, (Y, d'), h$ as asserted. Since $h(X)$ is totally bounded, there are $y_1, \ldots, y_n \in h(X)$ such that $h(X) \subseteq \bigcup_i B(y_i, \delta) \subseteq Y$. Then $X = \bigcup_i h^{-1}(B(y_i, \delta))$. Choose x_i such that $h(x_i) = y_i$. Now $x \in h^{-1}(B(y_i, \delta)) \Rightarrow d'(h(x), y_i) < \delta \Rightarrow d(x, x_i) < \varepsilon$, so that $h^{-1}(B(y_i, \delta)) \subseteq B(x_i, \varepsilon)$. Thus $X = \bigcup_i B(x_i, \varepsilon)$, and (X, d) is totally bounded. ■

Theorem 7.7.67 (Ascoli-Arzelà) [12] *Let (X, τ) be a compact topological space and (Y, d) a complete metric space. Then $\mathcal{F} \subseteq C(X, Y)$ is (relatively) compact (w.r.t. the uniform topology τ_D) if and only if*

- *$\{f(x) \mid f \in \mathcal{F}\} \subseteq Y$ is (relatively) compact for every $x \in X$,*

[12] Giulio Ascoli (1843-1896), Cesare Arzelà (1847-1912), Italian mathematicians. They proved special cases of this result, of which there also exist more general versions.

- *$\mathcal{F}$ is equicontinuous.*

Proof. $\Rightarrow$ If $f, g \in C(X, Y)$ then $d(f(x), g(x)) \leq D(f, g)$ for every $x \in X$. This implies that the evaluation map $e_x : C(X, Y) \to Y,\ f \mapsto f(x)$ is continuous for every x. Thus compactness of $\overline{\mathcal{F}}$ implies that $e_x(\overline{\mathcal{F}}) = \{f(x) \mid f \in \overline{\mathcal{F}}\}$ is compact, thus closed. Since $e_x(\overline{\mathcal{F}})$ contains $e_x(\mathcal{F})$, also $\overline{e_x(\mathcal{F})} \subseteq e_x(\overline{\mathcal{F}})$ is compact.

To prove equicontinuity, let $x \in X$ and $\varepsilon > 0$. Since $\mathcal{F}$ is totally bounded, there are $g_1, \ldots, g_n \in \mathcal{F}$ such that $\mathcal{F} \subseteq \bigcup_i B^D(g_i, \varepsilon)$. By continuity of the g_i, there are open $U_i \ni x,\ i = 1, \ldots, n$, such that $x' \in U_i \Rightarrow d'(g_i(x), g_i(x')) < \varepsilon$. Put $U = \bigcap_i U_i$. If now $f \in \mathcal{F}$, there is an i such that $f \in B^D(g_i, \varepsilon)$, to wit $D(f, g_i) < \varepsilon$. Now for $x' \in U \subseteq U_i$ we have

$$d'(f(x), f(x')) \leq d'(f(x), g_i(x)) + d'(g_i(x), g_i(x')) + d'(g_i(x'), f(x')) < 3\varepsilon,$$

proving equicontinuity of $\mathcal{F}$ (at x, but x was arbitrary).

$\Leftarrow$ Let $\varepsilon > 0$. Since $\mathcal{F}$ is equicontinuous, for every $x \in X$ there is an open neighborhood U_x such that $f \in \mathcal{F}, x' \in U \Rightarrow d'(f(x), f(x')) < \varepsilon$. Since X is compact, there are $x_1, \ldots, x_n \in X$ such that $X = \bigcup_{i=1}^n U_{x_i}$. Now define $h : \mathcal{F} \to Y^{\times n} : f \mapsto (f(x_1), \ldots, f(x_n))$. Now $\widetilde{d}((y_1, \ldots, y_n), (y_1', \ldots, y_n')) = \sum_i d'(y_i, y_i')$ is a product metric on $Y^{\times n}$. By assumption $\overline{\{f(x) \mid f \in \mathcal{F}\}}$ is compact for each $x \in X$, thus $\overline{h(\mathcal{F})} \subseteq \prod_i \overline{\{f(x_i) \mid f \in \mathcal{F}\}} \subseteq Y^{\times n}$ is compact, thus $(h(\mathcal{F}), \widetilde{d})$ is totally bounded. If now $f, g \in \mathcal{F}$ satisfy $\widetilde{d}(h(f), h(g)) < \varepsilon$ then $d'(f(x_i), g(x_i)) < \varepsilon\ \forall i$ by definition of $\widetilde{d}$. With $x \in U_{x_i}$, we have

$$d'(f(x), g(x)) \leq d'(f(x), f(x_i)) + d'(f(x_i), g(x_i)) + d'(g(x_i), g(x)) < 3\varepsilon.$$

Since the U_{x_i} cover X, this implies $D(f, g) < 3\varepsilon$. Thus the assumptions of Lemma 7.7.66 are satisfied, and we obtain total boundedness of $\mathcal{F}$. ■

If $Y = \mathbb{R}^n$ then in view of Theorem 7.7.48 the requirement of compactness of $\overline{\{f(x) \mid f \in \mathcal{F}\}}$ for each x reduces to that of boundedness of $\{f(x) \mid f \in \mathcal{F}\}$ for each x, i.e. <u>pointwise boundedness</u> of $\mathcal{F}$. With the equivalence of compactness and sequential compactness for the metric space $(C(X, \mathbb{R}^n), D)$ the following is equivalent to Theorem 7.7.67 (for $Y = \mathbb{R}^n$):

Corollary 7.7.68 *If (X, τ) is compact and $\{f_n\}_{n \in \mathbb{N}} \subseteq C(X, \mathbb{R}^n)$ is pointwise bounded and equicontinuous then the sequence $\{f_n\}$ has a uniformly convergent subsequence.*

Theorem 7.7.67 and its corollary will be used for the proof of the Hopf-Rinow Theorem 12.4.27.

Ascoli-Arzelà type theorems are often stated with the (superfluous) additional assumption that X is metric. Given two metric spaces $(X, d), (Y, d')$, one can formulate the notion of uniform equicontinuity, which often (imprecisely) is called equicontinuity:

Definition 7.7.69 *Let $(X, d), (Y, d')$ be metric spaces. A family $\mathcal{F} \subseteq \mathrm{Fun}(X, Y)$ is called <u>uniformly equicontinuous</u> if for every $\varepsilon > 0$ there is a $\delta > 0$ such that*

$$f \in \mathcal{F}, \quad x, y \in X, \quad d(x, y) < \delta \quad \Rightarrow \quad d'(f(x), f(y)) < \varepsilon.$$

(This clearly implies equicontinuity of $\mathcal{F}$ and uniform continuity of every $f \in \mathcal{F}$.)

We know that continuous functions from a compact metric space to any metric space are uniformly continuous. This generalizes to equicontinuous families, making it pointless (but also harmless) to require uniform equicontinuity instead of equicontinuity:

Lemma 7.7.70 *If $(X, d), (Y, d')$ are metric spaces with X compact then every equicontinuous family $\mathcal{F} \subseteq \mathrm{Fun}(X, Y)$ is uniformly equicontinuous.*

Proof. Given $\varepsilon > 0$, equicontinuity allows us to find for every $x \in X$ a $\delta_x > 0$ such that $f \in \mathcal{F}$, $d(x,y) < \delta_x$ implies $d'(f(x), f(y)) < \varepsilon$. The rest of the proof is identical to that of Proposition 7.7.38 except of course that now f is any element of $\mathcal{F}$. ■

7.8 One-point compactification. Local compactness

7.8.1 Compactifications: Definition and Examples

Since compact spaces have very nice properties, in particular in combination with the Hausdorff property, it is natural to ask whether a non-compact space can be 'compactified' by embedding it into a compact space. In analogy to completions of metric spaces (Definition 3.2.1) we define:

Definition 7.8.1 *A (Hausdorff) compactification of a topological space (X,τ) is a space $(\widehat{X},\widehat{\tau})$ together with a continuous map $\iota : X \to \widehat{X}$ such that*

- *$(\widehat{X},\widehat{\tau})$ is compact (Hausdorff).*
- *ι is an embedding. (I.e. $\iota : X \to \iota(X) \subseteq \widehat{X}$ is a homeomorphism.)*
- *$\iota(X)$ is dense in $\widehat{X}$.*

The points in $\widehat{X}\backslash\iota(X)$ are the infinite points of $\widehat{X}$.

Remark 7.8.2 We will occasionally suppress the embedding map ι from the notation, considering X as a subset of $\widehat{X}$ (but not when we consider different compactifications). □

Exercise 7.8.3 Let (X,τ) be a topological space, and let $X_\infty = X \cup \{\infty\}$. (Here and in the sequel, it is understood that $\infty \not\in X$.) Define τ' as in Exercise 7.3.1. (We now write ∞ instead of p.) Assuming $X \neq \emptyset$, prove:

(i) X is dense in X', but τ' is not Hausdorff.

(ii) Conclude that (X',τ') is a (non-T_2) compactification of (X,τ).

(iii) How do the above statements change if $X = \emptyset$?

The next example is better since it provides Hausdorff compactifications of $\mathbb{R}^n$ and shows that compactifications in general are not unique:

Example 7.8.4 1. $(0,1)$ has $[0,1]$ and S^1 as compactifications (among many others!).

2. More generally, there is a homeomorphism ι_n from $\mathbb{R}^n$ to the open unit ball in $\mathbb{R}^n$. (Why?) In view of $D^n = \overline{\iota_n(\mathbb{R}^n)}$ it is immediate that (D^n, ι_n) is a Hausdorff compactification of $\mathbb{R}^n$. If $\sim$ is any equivalence relation on $S^{n-1} = \partial D^n$, extend it to D^n in the minimal way, i.e. $x \sim y \Rightarrow x = y \vee \{x,y\} \subseteq S^{n-1}$. One checks that this equivalence relation on D^n is closed if and only if it is closed on S^{n-1}. As we will see later, this is equivalent to $D^n/\sim$ being Hausdorff. Now $\mathbb{R}^n_\sim = D^n/\sim$ is compact. If $p : D^n \to \mathbb{R}^n_\sim$ is the quotient map and $\iota_\sim = p \circ \iota$, one finds that $(\mathbb{R}^n_\sim, \iota_\sim)$ is a compactification of $\mathbb{R}^n$. Thus every closed equivalence relation on S^{n-1} gives rise to a Hausdorff compactification of $\mathbb{R}^n$! Examples: (i) the trivial equivalence relation, i.e. $\mathbb{R}^n_\sim \cong D^n$, (ii) the radical equivalence relation that identifies all points. We will see that it leads to $\mathbb{R}^n_\sim \cong S^n$. And (iii) the equivalence relation that identifies only pairs $(x,-x)$ of antipodal points, which leads to $\mathbb{R}^n_\sim \cong \mathbb{RP}^n$, as shown in Lemma 6.4.28. Thus $\mathbb{R}^n$ has $D^n, S^n, \mathbb{RP}^n$ as Hausdorff compactifications, but clearly there are many others.
□

Remark 7.8.5 1. Exercise 7.8.3 shows that every space admits a compactification, in analogy to the existence of completions for all metric spaces. However, the compactification given there is not very useful since it is never Hausdorff.

2. A non-Hausdorff space admits no Hausdorff compactification (since T_2 is hereditary).

3. If X is compact then every Hausdorff compactification $\widehat{X}$ is homeomorphic to X. (Since X is compact, also $\iota(X)$ is compact, thus $\iota(X) \subseteq \widehat{X}$ is closed by the Hausdorffness of $\widehat{X}$. We thus have $X \cong \iota(X) = \overline{\iota(X)} = \widehat{X}$, by density of $\iota(X)$.)

4. Hausdorffness of X is not sufficient for X to have a Hausdorff compactification! Later (Theorem 8.3.20) we will identify a necessary and sufficient condition, but for the time being we have no need for that. (And in Section 7.8.3 we construct a Hausdorff compactification whose existence follows from a simpler but stronger condition.)

5. In contradistinction to completions, there is no uniqueness for compactifications. But if a space has a Hausdorff compactification at all, it has a unique 'largest' one, cf. Corollary 8.3.32. Under the stronger assumption mentioned above, it also has a unique 'smallest' compactifications, cf. Corollary 7.8.59. □

7.8.2 ⋆ Compactifications: Some general theory

The fact that there are multiple compactifications of a given space X makes it natural to study *all* compactifications of X and the maps between them. Categorical language is best suited for this:

Definition 7.8.6 *If X is a Hausdorff space, the category $\mathcal{C}(X)$ of Hausdorff compactifications of X is defined as follows:* $\mathrm{Obj}\,\mathcal{C}(X) = \{(\widehat{X}, \iota)\}$*, where $\widehat{X}$ is compact Hausdorff and $\iota : X \to \widehat{X}$ is an embedding with dense image. If $(\widehat{X}_1, \iota_1), (\widehat{X}_2, \iota_2) \in \mathcal{C}(X)$ then*

$$\mathrm{Hom}_{\mathcal{C}(X)}((\widehat{X}_1, \iota_1), (\widehat{X}_2, \iota_2)) = \{f \in C(\widehat{X}_1, \widehat{X}_2) \mid f \circ \iota_1 = \iota_2\}.$$

Since compactifications are not unique, we have no complete analogue of Proposition 3.2.2. The next result is the next best we can hope for:

Proposition 7.8.7 *Let X be non-compact.*

(i) If $(\widehat{X}_1, \iota_1), (\widehat{X}_2, \iota_2)$ are objects in $\mathcal{C}(X)$ then $\mathrm{Hom}_{\mathcal{C}(X)}((\widehat{X}_1, \iota_1), (\widehat{X}_2, \iota_2))$ contains at most element.

(ii) If $\mathrm{Hom}_{\mathcal{C}(X)}((\widehat{X}_1, \iota_1), (\widehat{X}_2, \iota_2)) \neq \emptyset \neq \mathrm{Hom}_{\mathcal{C}(X)}((\widehat{X}_2, \iota_2), (\widehat{X}_1, \iota_1))$ then $(\widehat{X}_1, \iota_1) \cong (\widehat{X}_2, \iota_2)$.

(iii) If the $\mathrm{Hom}_{\mathcal{C}(X)}((\widehat{X}_1, \iota_1), (\widehat{X}_2, \iota_2))$ is non-empty empty then its unique element f is surjective, a quotient map and satisfies

$$f(\widehat{X}_1 \backslash \iota_1(X)) = \widehat{X}_2 \backslash \iota_2(X).$$

Thus f maps the infinite points of $\widehat{X}_1$ to (in fact onto) the infinite points of $\widehat{X}_2$.

(iv) The isomorphism classes of compactifications of X form a partially ordered set $\mathcal{C}(X)/\cong$, where $[(\widehat{X}_1, \iota_1)] \geq [(\widehat{X}_2, \iota_2)]$ if and only if $\mathrm{Hom}_{\mathcal{C}(X)}((\widehat{X}_1, \iota_1), (\widehat{X}_2, \iota_2)) \neq \emptyset$.

(v) A compactification $(\widehat{X}, \iota) \in \mathcal{C}(X)$ is an initial (resp. terminal) object if and only if $[(\widehat{X}, \iota)]$ is a greatest (resp. smallest) element of the partially ordered set $(\mathcal{C}(X)/\cong, \leq)$.

Proof. (i) Any two morphisms $(\widehat{X}_1,\iota_1) \xrightarrow{f,g} (\widehat{X}_2,\iota_2)$ coincide on the dense subset $\iota_1(X)$. Since $\widehat{X}_2$ is Hausdorff, Exercise 5.2.16(ii) gives $f = g$.

(ii) Let $f : (\widehat{X}_1,\iota_1) \to (\widehat{X}_2,\iota_2)$ and $g : (\widehat{X}_2,\iota_2) \to (\widehat{X}_1,\iota_1)$. Then $g \circ f$ coincides with $\mathrm{id}_{\widehat{X}_1}$ on the dense subset $\iota_1(X)$, thus $g \circ f = \mathrm{id}_{\widehat{X}_1}$. Similarly $f \circ g = \mathrm{id}_{\widehat{X}_2}$.

(iii) If $f \in \mathrm{Hom}_{\mathcal{C}(X)}((\widehat{X}_1,\iota_1),(\widehat{X}_2,\iota_2))$ then $f \circ \iota_1 = \iota_2$ implies that $f(\widehat{X}_1)$ contains $\iota_2(X)$, which is dense since $\widehat{X}_2$ is a compactification. Since $\widehat{X}_1$ is compact, so is $f(\widehat{X}_1)$, and since $\widehat{X}_2$ is Hausdorff $f(\widehat{X}_1)$ is closed, thus equal to $\widehat{X}_2$. Now f is a quotient map by Proposition 7.4.11(iv). The remaining claim $f(\widehat{X}_1 \backslash \iota_1(X)) = \widehat{X}_2 \backslash \iota_2(X)$ follows from Lemma 7.8.8 below since $\iota_1(X) \subseteq \widehat{X}_1$ is dense and $f \restriction \iota_1(X) = \iota_2 \circ \iota_1^{-1}$ is a homeomorphism with image $\iota_2(X)$.

(iv) The above results suggest that the isomorphism classes of compactifications of a space X form a partially ordered set. Indeed, defining $\leq$ as stated, reflexivity and transity of $\leq$ are trivial, whereas antisymmetry follows from (ii). The problem is that the compactifications of X definitely do not form a set, but a proper class. So the question is whether at least the isomorphism classes of compactifications of X form a set. This is indeed the case and can be proven in different ways. We will later see that whenever X has a Hausdorff compactification, it has compactification βX such that for any other compactification $\widehat{X}$ there is a morphism $\beta X \to \widehat{X}$ in $\mathcal{C}(X)$ (thus βX is an initial object in $\mathcal{C}(X)$), thus $\widehat{X}$ is (isomorphic to) a quotient of βX. But the isomorphism classes of quotient spaces of a fixed topological space X are in bijective correspondence to the equivalence relations on X, and the latter form a set since an equivalence relation is a subset of $X \times X$. (Alternatively, one can use the fact that if $\widehat{X}$ is a Hausdorff compactification of X, the cardinality of $\widehat{X}$ is bounded by $\#\widehat{X} \leq \#\mathbb{R}^{(2^{\#X})}$. Also this is proven using βX.)

(v) By (i), all hom-sets in $\mathcal{C}(X)$ contain at most one element. Thus an object $(\widehat{X},\iota)$ in $\mathcal{C}(X)$ is initial if and only if $\mathrm{Hom}((\widehat{X},\iota),(\widehat{X}',\iota')) \neq \emptyset$ for all $(\widehat{X}',\iota')$. Thus if and only if $[(\widehat{X},\iota)] \geq [(\widehat{X}',\iota')]$ for all $[(\widehat{X}',\iota')]$, which is the definition of a greatest element. Similarly, $(\widehat{X},\iota)$ is terminal if and only if $[(\widehat{X},\iota)]$ is smallest. ■

Lemma 7.8.8 *Let $f : X \to Y$ be continuous, where X is Hausdorff. Let $A \subseteq X$ be a dense subset such that $f \restriction A$ is a homeomorphism $A \to f(A)$. Then $f(X \backslash A) \subseteq Y \backslash f(A)$.*

Proof. Assume the claim is false. Then there are $a \in A,\ b \in X \backslash A$ such that $f(a) = f(b)$. Since X is Hausdorff, there are disjoint open $U \ni a, V \ni b$. Then $U \cap A$ is an open neighborhood in A of a. Since $f \restriction A$ is a homeomorphism, $f(U \cap A) \subseteq f(A)$ is an open neighborhood in $f(A)$ of $f(a)$. Thus there is an open $W \subseteq Y$ such that $f(U \cap A) = W \cap f(A)$. Clearly W is an open neighborhood in Y of $f(a) = f(b)$. Let $V' \subseteq X$ be open such that $b \in V' \subseteq V$. Since $A \subseteq X$ is dense, $V' \cap A \neq \emptyset$. In view of $V' \subseteq V \subseteq X \backslash U$ and the injectivity of $f \restriction A$, the non-empty set $f(V' \cap A)$ is disjoint from $f(U \cap A) = W \cap f(A)$. Thus $f(V' \cap A) \subseteq W$ cannot hold, clearly implying $f(V') \not\subseteq W$. Since this is the case whenever $b \in V' \subseteq V$, we have a contradiction with the continuity of f at b. ■

The next result provides a criterion for determining when two Hausdorff compactifications of a space X (assuming there are any) are isomorphic in $\mathcal{C}(X)$:

Proposition 7.8.9 *Let $(\widehat{X}_1,\iota_1),(\widehat{X}_2,\iota_2)$ be Hausdorff compactifications of the space X. Then $(\widehat{X}_1,\iota_1)$ and $(\widehat{X}_2,\iota_2)$ are isomorphic as objects of $\mathcal{C}(X)$ if and only if for any two disjoint closed sets $A, B \subseteq X$ we have*

$$\overline{\iota_1(A)} \cap \overline{\iota_1(B)} = \emptyset \quad \Leftrightarrow \quad \overline{\iota_2(A)} \cap \overline{\iota_2(B)} = \emptyset. \tag{7.8}$$

Proof. We have continuous maps $f_1 : \iota_2 \circ \iota_1^{-1} : \iota_1(X) \to \widehat{X}_2$ and $f_2 : \iota_1 \circ \iota_2^{-1} : \iota_2(X) \to \widehat{X}_1$ defined on the dense subspaces $\iota_1(X) \subseteq \widehat{X}_1,\ \iota_2(X) \subseteq \widehat{X}_2$, respectively. In order to obtain an isomorphism

$\widehat{X}_1 \to \widehat{X}_2$ in $\mathcal{C}(X)$, we must construct continuous extensions $\widehat{f}_1, \widehat{f}_2$ of these maps to $\widehat{X}_1, \widehat{X}_2$. The extensions will then automatically be mutually inverse since $\widehat{f}_2 \circ \widehat{f}_1 : \widehat{X}_1 \to \widehat{X}_2$ will be the identity on the dense subset $\iota_1(X_1) \subseteq \widehat{X}_1$, and similarly for $\widehat{f}_1 \circ \widehat{f}_2$.

By Theorem 7.4.20, f_1 extends if and only if given disjoint closed sets $A, B \in \widehat{X}_2$, we have $\overline{\iota_2(\iota_1^{-1}(A))} \cap \overline{\iota_2(\iota_1^{-1}(B))} = \emptyset$, and similarly for f_2. But every closed $A \in \widehat{X}_1$ is of the form $\overline{\iota_1(A_0)}$ for some closed $A_0 \subseteq X$, namely $A_0 = \iota_1^{-1}(A)$, and similarly for B. Thus f_1 extends continuously if and only if given closed $A_0, B_0 \subseteq X$ such that $\overline{\iota_1(A)} \cap \overline{\iota_1(B)} = \emptyset$, we have $\overline{\iota_2(A)} \cap \overline{\iota_2(B)} = \emptyset$. In the condition for extendability of f_2, ι_1 and ι_2 are exchanged. Thus the condition (7.8) implies $\widehat{X}_1 \cong \widehat{X}_2$ in $\mathcal{C}(X)$. Conversely, this isomorphism implies that f_1 and f_2 continuously extend (to mutually inverse maps), so that (7.8) is satisfied. ■

In Section 8.4.3 this result will be used to give a fairly intrinsic classification of Hausdorff compactifications.

7.8.3 The one-point compactification X_∞

In Exercise 7.8.3 we have seen that for every topological space (X, τ) there is a topology τ' on $X_\infty = X \cup \{\infty\}$ making (X_∞, τ') a compactification of (X, τ). This is conceptually interesting, but τ' was not Hausdorff, whereas we are mostly interested in Hausdorff compactifications. The following mainly serves to motivate Theorem 7.8.14 (but also provides a converse):

Lemma 7.8.10 *Let (X, τ) be a topological space, $X_\infty = X \cup \{\infty\}$, and assume that τ' is a topology on X_∞ such that (X_∞, τ') is a Hausdorff compactification of (X, τ). Then*

(i) X is Hausdorff and non-compact.

(ii) Every $x \in X$ has a compact neighborhood (i.e. $x \in U \subseteq K$ for some open U, compact K).

(iii) $\tau' = \tau \cup \{X_\infty \backslash K \mid K \subseteq X \text{ compact}\}$. (We interpret subsets of X as subsets of X_∞ in the obvious way.)

Proof. (i) Hausdorffness of (X, τ) follows from hereditarity of this property and the fact that $\tau' \upharpoonright X = \tau$. If X was compact then $X \subseteq X_\infty$ would be closed since X_∞ is Hausdorff, contradicting the requirement $\overline{X} = X_\infty$.

(ii) If $x \in X$ then by the Hausdorff property of τ' there are $U, V \in \tau'$ such that $U \cap V = \emptyset$, $x \in U, \infty \in V$. With $K = X_\infty \backslash V$ we have $x \in U \subseteq K \subseteq X$, where K is closed in X_∞ and thus compact. Thus K is a compact neighborhood of x.

(iii) By assumption τ' is T_2, thus T_1, thus $\{\infty\}$ is τ'-closed, so that $X = X_\infty \backslash \{\infty\} \in \tau'$. By assumption, $(X, \tau) \hookrightarrow (X_\infty, \tau')$ is an embedding, i.e. $\tau = \tau' \upharpoonright X = \{U \cap X \mid U \in \tau'\}$. This means for every $U \in \tau$ that $U \in \tau'$ or $U \cup \{\infty\} \in \tau'$. In the latter case, $X \in \tau'$ implies $U = (U \cup \{\infty\}) \cap X \in \tau'$. Thus $\tau \subseteq \tau'$. It is obvious that if $U \in \tau'$ and $U \subseteq X$ then $U = U \cap X$, thus $U \in \tau$. This shows that the τ'-open sets not containing ∞ are precisely the elements of τ.

If $\infty \in U \in \tau'$ then $X_\infty \backslash U \subseteq X$ is closed in X_∞, thus compact. Conversely, if $K \subseteq X$ is compact then K is closed as a subset of X_∞ since the latter is Hausdorff, thus $X_\infty \backslash K \in \tau'$. This proves that the τ'-open subsets containing ∞ are precisely the complements (in X_∞) of the compact subsets of X. ■

Since we are mainly interested in Hausdorff compactifications, the condition in (ii) merits a name. Spaces satisfying it are called locally compact.

Since we will meet many similar conditions later on, we consider the generalization right away:

Definition 7.8.11 *Let P be a property that a topological space can either have or not have, and let (X, τ) be a topological space. Then we say that*

- *X is weakly locally P if every $x \in X$ has a neighborhood that has property P.*
- *X is strongly locally P if given $x \in U \in \tau$ there is a neighborhood N of x such that N has property P and $N \subseteq U$. (Thus every $x \in X$ has a neighborhood base consisting of sets with property P.)*

In these notes, 'locally P' always means 'weakly locally P' unless specified otherwise.

Remark 7.8.12 We obviously have the implication 'X is strongly locally P' $\Rightarrow$ 'X is weakly locally P', but also 'X is P' $\Rightarrow$ 'X is weakly locally P' (since X is a neighborhood of every $x \in X$). In general, no other implications hold. (But for local compactness see Lemma 7.8.24.) □

Example 7.8.13 Here are some examples of locally compact spaces:

1. Compact spaces. (Trivial.)
2. Discrete spaces. (A compact neighborhood for $x \in X$ is $\{x\}$.)
3. $\mathbb{R}^n$ for any $n \in \mathbb{N}$. Finite dimensional normed spaces. (By Theorems 7.7.48 and 7.7.51.)
4. In view of 3., topological spaces where every point has a neighborhood homeomorphic to $\mathbb{R}^n$, are locally compact. Such spaces are called locally Euclidean. (Much of modern mathematics revolves around manifolds, which are locally Euclidean spaces satisfying some additional axioms.) □

Lemma 7.8.10 does not imply that taking the τ' appearing in (iii) as the definition of a topology will work, but this surely is not unreasonable to hope. While we are mainly interested in the case where X is Hausdorff, we do not assume this to begin with. This forces us to include the requirement 'closed' in the following definition. It can be omitted when X is Hausdorff.

Theorem 7.8.14 *Let (X, τ) be a topological space. Put $X_\infty = X \cup \{\infty\}$ (where $\infty \notin X$) and define $\tau_\infty \subseteq P(X_\infty)$ by*

$$\tau_\infty = \tau \cup \{X_\infty \backslash K \mid K \subseteq X \text{ closed and compact}\}. \tag{7.9}$$

Then

(i) (X_∞, τ_∞) is a topological space.

(ii) $\tau_\infty \upharpoonright X = \tau$, thus $X \hookrightarrow X_\infty$ is an embedding.

(iii) (X_∞, τ_∞) is compact.

(iv) $X \subseteq X_\infty$ is open and $\{\infty\} \subseteq X_\infty$ is closed. Furthermore,
$X \subseteq X_\infty$ is closed $\Leftrightarrow$ $\{\infty\} \subseteq X_\infty$ is open $\Leftrightarrow$ $X_\infty = X \oplus \{\infty\}$ $\Leftrightarrow$ X is compact.

(v) X_∞ is a compactification of X if and only if X is non-compact.

(vi) τ_∞ is Hausdorff if and only if τ is Hausdorff and locally compact.

Proof. (i) We have $\emptyset \in \tau \subseteq \tau_\infty$. And $K = \emptyset$ is closed and compact, thus $X_\infty \in \tau_\infty$. Any family $\mathcal{U} \subseteq \tau_\infty$ is of the form $\mathcal{U}_1 \cup \mathcal{U}_2$, where $\mathcal{U}_1 \subseteq \tau$ and $\mathcal{U}_2 = \{X_\infty \backslash K_i\}_{i \in I}$ with $K_i \subseteq X$ closed and compact. Clearly $\bigcup \mathcal{U}_1 \in \tau$ and $\bigcup \mathcal{U}_2 = X_\infty \backslash \bigcap_i K_i$. The intersection of any number of closed compact sets is closed and compact by Exercise 7.4.7, thus $\bigcup \mathcal{U}_2 \in \tau_\infty$. If $U_1 \in \tau$ and $U_2 = X_\infty \backslash K$ with $K \subseteq X$ closed compact, then $U_1 \cup U_2 = X_\infty \backslash K'$, where $K' = K \cap (X \backslash U_1)$. Since the intersection of the closed compact set K with the closed $X \backslash U_1$ is closed compact, we have $U_1 \cup U_2 \in \tau_\infty$.

Finally, consider $U_1 \cap U_2$. The case $U_1, U_2 \in \tau$ is clear. Furthermore, $(X_\infty \backslash K_1) \cap (X_\infty \backslash K_2) = X_\infty \backslash (K_1 \cup K_2)$, which is in τ_∞ since the union of two closed compact sets is closed and compact (Exercise 7.4.7). Finally, if $U \in \tau$ and $K \subseteq X$ is closed and compact then $U \cap (X_\infty \backslash K) = U \cap (X \backslash K)$, which is in $\tau \subseteq \tau_\infty$ since $X \backslash K \in \tau$.

(ii) By construction, $\tau \subseteq \tau_\infty$, thus $\tau_\infty \upharpoonright X \supseteq \tau$. Let $K \subseteq X$ be closed and compact. Then $X \cap (X_\infty \backslash K) = X \backslash K$, which is in τ. Thus intersected with X, the 'new' open sets $X_\infty \backslash K$ in τ_∞ become 'old' ones, so that $\tau_\infty \upharpoonright X = \tau$.

(iii) Let $\mathcal{U} \subseteq \tau_\infty$ be an open cover of X_∞. Since $\mathcal{U}$ must cover ∞, it contains at least one set of the form $U_0 = X_\infty \backslash K$ with $K \subseteq X$ closed and compact. Now by (ii), $\{X \cap U \mid U \in \mathcal{U}\}$ is a cover of K by elements of τ, and by compactness of K there is a finite subfamily $\mathcal{U}_0 \subseteq \mathcal{U}$ still covering K. Now $\mathcal{U}_0 \cup \{U_0\}$ is a finite subcover of $\mathcal{U}$, thus τ_∞ is compact.

(iv) Since $X \in \tau$, we have $X \in \tau_\infty$, thus $\{\infty\} = X_\infty \backslash X$ is closed. By definition of τ_∞, we have $\{\infty\} \in \tau_\infty$ if and only if X is compact. This in turn is equivalent to $\{\infty\}$ being clopen and thus a direct summand, cf. Proposition 6.3.7.

(v) Statement (ii) means that the inclusion map $\iota_\infty : X \hookrightarrow X_\infty$ is an embedding, and by (iv) $X \subseteq X_\infty$ is non-closed, and thus dense, if and only if X is non-compact.

(vi) Assume X_∞ is Hausdorff. Then the subspace $X \subseteq X_\infty$ is Hausdorff since this property is hereditary. Furthermore ∞ can be separated from any $x \in X$ by open sets. By definition of τ_∞ this means that there are $U \in \tau$ containing x and $K \subseteq X$ closed and compact such that $U \cap (X_\infty \backslash K) = \emptyset$. This is equivalent to $x \in U \subseteq K$, thus K is a compact neighborhood of x. Since the argument also works the other way round, if X is Hausdorff and every $x \in X$ has a compact neighborhood then X_∞ is Hausdorff. ■

Definition 7.8.15 (X_∞, τ_∞) *is called the* <u>*one-point*</u> *or* <u>*Alexandrov compactification*</u> *of* (X, τ).

Remark 7.8.16 1. Strictly speaking, it is incorrect to call (X_∞, τ_∞) the one-point compactification when X is already compact since then ∞ is an isolated point of X_∞ and $\overline{X} = X \neq X_\infty$. This slight inconsistency will cause no harm.

2. In the context of the one-point compactification X_∞, we will often suppress the embedding map ι_∞ from the notation and identify X with its image in X_∞.

3. It should be clear that the space $(\mathbb{N}_\infty, \tau)$ considered in Exercise 5.2.18 (cf. also Exercise 5.2.22) is nothing but the one-point compactification of $(\mathbb{N}, \tau_{\text{disc}})$. □

Combining Lemma 7.8.10 and Theorem 7.8.14, we have:

Corollary 7.8.17 *Given a topological space* (X, τ), *there exists a topology* τ' *on* $X_\infty = X \cup \{\infty\}$ *making* (X_∞, τ') *a Hausdorff compactification of* (X, τ) *if and only if* (X, τ) *is Hausdorff, locally compact and non-compact. In this case,* $\tau' = \tau_\infty$ *is the unique such topology.*

Exercise 7.8.18 Let (X_∞, τ_∞) be the one-point compactification of (X, τ). Prove:

(i) If $A \subseteq X$ then A is closed in X_∞ if and only if $A \subseteq X$ is closed and compact.

(ii) If $A \subseteq X$ is closed then $A \cup \{\infty\} \subseteq X_\infty$ is closed.

(iii) The closure $\overline{A} \subseteq X_\infty$ of $A \subseteq X$ is given by

$$\overline{A} = \begin{cases} \mathrm{Cl}_X(A) & \text{if } \mathrm{Cl}_X(A) \text{ is compact} \\ \mathrm{Cl}_X(A) \cup \{\infty\} & \text{if } \mathrm{Cl}_X(A) \text{ is non} - \text{compact} \end{cases}$$

Exercise 7.8.19 Let X be compact Hausdorff, $x \in X$, and let $Y = X \backslash \{x\}$. Prove that Y is locally compact Hausdorff and $Y_\infty \cong X$.

Exercise 7.8.20 (i) Prove that $S^n \backslash \{x\} \cong \mathbb{R}^n$. (Use stereographic projection.)

(ii) Prove that $(\mathbb{R}^n)_\infty \cong S^n$ for all $n \geq 1$.

(iii) Describe (by proving homeomorphisms to known spaces) the 1-point compactifications of $(0,1), [0,1), [0,1]$.

Exercise 7.8.21 For $I = [0,1]$, $n \in \mathbb{N}$, let $\partial I^n = \{x \in I^n \mid \exists i : x_i \in \{0,1\}\}$ be the boundary of I^n, and let $\sim$ be the equivalence relation on I^n defined by $x \sim y$ if $x = y$ or $\{x,y\} \subseteq \partial I^n$. Call $z \in I^n/{\sim}$ the image of ∂I^n under the quotient map $p : I^n \to I^n/{\sim}$. Prove:

(i) $I^n/{\sim}$ is compact Hausdorff.

(ii) $(I^n/{\sim}) \backslash \{z\} \cong (0,1)^n \cong \mathbb{R}^n$.

(iii) $I^n/{\sim} \cong (\mathbb{R}^n)_\infty \cong S^n$. (Use Exercise 7.8.19.)

Exercise 7.8.22 (i) Given a topological space $X \neq \emptyset$, prove that X_∞ is connected if and only if X has no compact direct summand.

(ii) Give an example of a non-connected space $X \neq \emptyset$ such that X_∞ is connected.

(iii) Give an example of a non-connected space $X \neq \emptyset$ such that X_∞ is non-connected.

7.8.4 Locally compact spaces

In this section we will devote some attention to the property of local compactness. It should be emphasized that the existence of a Hausdorff one-point compactification is neither the only nor the most important reason to study locally compact spaces. Actually more important are topologized algebraic structures:

Definition 7.8.23 • *A topological group is a group G equipped with a topology such that the algebraic operations $(g,h) \mapsto gh$ and $g \mapsto g^{-1}$ are continuous.*

- *Topological fields are defined analogously.*

- *If $\mathbb{F}$ is a topological field, a topological vector space over $\mathbb{F}$ is a vector space V equipped with a topology, such that V is a topological abelian group and the action of $\mathbb{F}$ on V is continuous (as a map $\mathbb{F} \times V \to V$).*

Local compactness of the above structures has many uses:

- Every locally compact group G carries a canonical positive measure μ, allowing to integrate reasonable functions defined on G. (G is compact if and only if $\mu(G) < \infty$.) This area is called abstract harmonic analysis. Cf. e.g. [116].

- Every locally compact abelian group A has a dual group $\widehat{A}$, which is again locally compact abelian. $\widehat{A}$ is compact (discrete) if and only if A is discrete (compact). For every A, one has an isomorphism $\widehat{\widehat{A}} \cong A$ of topological groups, called Pontrjagin duality, cf. [116, Vol.1]. The examples $\widehat{(\mathbb{R},+)} = (\mathbb{R},+)$, $\widehat{(\mathbb{Z},+)} \cong (S^1,\cdot)$ are the basis of classical Fourier analysis.

- The fields $\mathbb{R}$ and $\mathbb{C}$ are locally compact, but not compact. More generally, a local field is a topological field whose topology is locally compact and non-discrete. Local fields are quite well understood, cf. [234]. Those of characteristic zero (i.e. containing $\mathbb{Q}$) are precisely the finite field extensions of $\mathbb{R}$ and of the p-adic fields $\mathbb{Q}_p$ for p prime. (While the only finite field extension of $\mathbb{R}$ is $\mathbb{C}$, the $\mathbb{Q}_p$'s have many.) The local fields of prime characteristic p are the finite extensions of $\mathbb{F}_p((x))$, the field of formal Laurent series with coefficients in the prime field $\mathbb{F}_p$. ($\mathbb{F}_p((x))$ is the quotient field of the formal power series ring $\mathbb{F}_p[[x]]$.)

- Associated with every commutative Banach algebra B comes a certain locally compact Hausdorff space, the Gelfand spectrum of B. (The Gelfand spectrum of B is compact if and only if the algebra B has a unit.) Cf. e.g. [181].

Some authors define local compactness by the strong form. Clearly this implies our definition. On the other hand:

Lemma 7.8.24 *If (X,τ) is (weakly) locally compact and Hausdorff then every point has a neighborhood base of compact sets. (I.e., X is strongly locally compact.)*

Proof. Since (X,τ) is locally compact Hausdorff, (X_∞,τ_∞) is compact Hausdorff. Now let $x \in W \in \tau$. In view of $\tau \subseteq \tau_\infty$, $C := X_\infty \backslash W$ is closed, thus compact in X_∞. Applying Lemma 7.4.1 to X_∞, x and C, we obtain disjoint $U,V \in \tau_\infty$ containing C and x, respectively. Since $\infty \in C \subseteq U$, we have $\infty \notin V$, thus $V \in \tau$. On the other hand, $\infty \in U \in \tau_\infty$, thus $K := X_\infty \backslash U \subseteq X$ is compact. Thus $x \in V \subseteq K \subseteq W$, so that K is a compact neighborhood K of x contained in W. ■

In Exercise 7.4.5 we saw that disjoint compact subsets of a Hausdorff space X have disjoint open neighborhoods. Statement (ii) below shows that the assumptions on the subsets can be weakened when X is locally compact Hausdorff:

Proposition 7.8.25 *Let X be locally compact Hausdorff.*

(i) If $K \subseteq U \subseteq X$ with K compact and U open then there is an open V such that $\overline{V}$ is compact and $K \subseteq V \subseteq \overline{V} \subseteq U$.

(ii) If $K \subseteq X$ is compact, $C \subseteq X$ is closed and $C \cap K = \emptyset$ then there are open U,V such that $K \subseteq U$, $C \subseteq V$ and $U \cap V = \emptyset$.

Proof. (i) By Lemma 7.8.24, for every $x \in K$, we can find an open V_x such that $\overline{V_x}$ is compact and $x \in V_x \subseteq \overline{V_x} \subseteq U$. Now $\{V_x\}_{x \in K}$ covers K, thus by compactness we find $x_1,\ldots,x_n \in K$ such that $K \subseteq V_{x_1} \cup \cdots \cup V_{x_n} =: V$. Now, $\overline{V} = \overline{V_{x_1} \cup \cdots \cup V_{x_n}} = \overline{V_{x_1}} \cup \cdots \cup \overline{V_{x_n}}$, which is compact (as a finite union of compacts) and contained in U.

(ii) is equivalent to (i), as is seen taking $C = X \backslash U$. ■

How does local compactness behave w.r.t. the four ways of constructing new spaces out of old defined in Section 6? We begin with subspaces.

Definition 7.8.26 *A subset $Y \subseteq (X, \tau)$ is called* <u>*locally closed*</u> *if it is of the form $Y = U \cap C$ where U is open and C is closed.*

Remark 7.8.27 The definition of local closedness may seem a strange, but it has its rôles outside topology: In algebraic geometry, the subsets of projective space that are locally closed w.r.t. the Zariski topology are precisely the quasi-projective varieties. □

Exercise 7.8.28 Prove that $Y \subseteq X$ is locally closed if and only if Y is an open subset of $\overline{Y}$.

Exercise 7.8.29 Let X be locally compact Hausdorff. Prove:

(i) If $Y \subseteq X$ is closed then Y is locally compact. (This does not need Hausdorffness of X.)

(ii) If $Y \subseteq X$ is open then Y is locally compact.

(iii) If $Y \subseteq X$ is locally closed then Y is locally compact.

Corollary 7.8.30 *A topological space is locally compact Hausdorff if and only if it is homeomorphic to an open subspace of a compact Hausdorff space.*

Proof. Open subspaces of compact Hausdorff spaces are locally compact by Exercise 7.8.29(ii). Conversely, by Theorem 7.8.14(iv), every locally compact Hausdorff space X is (homeomorphic to) an open subspace of its one-point compactification X_∞, which is compact Hausdorff. ■

Now we have nice analogues of Lemma 7.4.2 and Corollary 7.4.3:

Proposition 7.8.31 *If X is Hausdorff and $Y \subseteq X$ is locally compact then $Y \subseteq X$ is locally closed.*

Proof. (i) By Exercise 7.8.28, $Y \subseteq X$ being locally closed is equivalent to Y being open in $\overline{Y}$. Replacing X by $\overline{Y}$, we thus reduce the problem to proving that a dense locally compact subspace $Y \subseteq X$ of a Hausdorff space is open. Let $y \in Y$. Since (Y, τ_Y) is locally compact, there is an open neighborhood $U \in \tau_Y$ of y whose closure $\mathrm{Cl}_Y(U)$ (in Y) is compact. In view of the definition of τ_Y there is an open $V \subseteq X$ such that $U = Y \cap V$. Using Exercise 2.6.13(v), we have

$$\mathrm{Cl}_Y(U) = \overline{U} \cap Y = \overline{Y \cap V} \cap Y.$$

Since $\mathrm{Cl}_Y(U)$ is compact and X Hausdorff, it follows that $\mathrm{Cl}_Y(U) = \overline{Y \cap V} \cap Y$ is closed in X. This closedness together with the trivial inclusion $Y \cap V \subseteq \overline{Y \cap V} \cap Y$ implies $\overline{Y \cap V} \subseteq \overline{Y \cap V} \cap Y$ and therefore $\overline{Y \cap V} \subseteq Y$. Now Lemma 2.7.10(i) gives $V \subseteq \overline{Y \cap V}$, so that we have $V \subseteq Y$. Since $V \subseteq X$ is open, this means that y is an interior point of Y. Since $y \in Y$ was arbitrary, we conclude that Y is open. ■

Combining this with Exercise 7.8.29(iii) we have:

Corollary 7.8.32 *If (X, τ) is locally compact Hausdorff then a subspace $Y \subseteq X$ is locally compact if and only if $Y \subseteq X$ is locally closed.*

Theorem 7.8.33 *A locally compact metric space (X, d) is completely metrizable.*

Proof. Embed X into its completion $\widehat{X}$. By Proposition 7.8.31, X is locally closed, thus open in its closure $\overline{X} = \widehat{X}$ by Exercise 7.8.28. Now Proposition 3.4.18 gives that X is completely metrizable. ■

Remark 7.8.34 1. Of course this does not mean that every locally compact metric space (X, d) is already complete: The space $(0, 1)$ with Euclidean metric $d(x, y) = |x - y|$ is locally compact, but the metric d is not complete.

2. The converse of Theorem 7.8.33 is false: By Theorem 7.7.55 no infinite dimensional normed space is locally compact, but there are many complete ones, e.g. $\ell^2(S)$ for infinite S. □

Exercise 7.8.35 A direct sum $\bigoplus_i X_i$ is locally compact if and only if each X_i is locally compact.

The question whether local compactness is preserved by continuous maps is complicated in general, but for open continuous maps it is staightforward:

Lemma 7.8.36 *If X is locally compact and $f : X \to Y$ is continuous and open then $f(X) \subseteq Y$ is locally compact.*

Proof. If $x \in U \subseteq K \subseteq X$ with U open and K compact then $f(x) \in f(U) \subseteq f(K)$. Now $f(U)$ is open (since f is open) and $f(K)$ is compact (by continuity of f and Lemma 7.3.6). Thus $f(K)$ is a compact neighborhood of $f(x)$, and $f(X)$ is locally compact. ■

Proposition 7.8.37 *Let $X_i \neq \emptyset\ \forall i$. Then $\prod_i (X_i, \tau_i)$ is locally compact if and only if each (X_i, τ_i) is locally compact and at most finitely many X_i are non-compact.*

Proof. Let X, Y be locally compact and $x \in X,\ y \in Y$. Let $N \subseteq X,\ M \subseteq Y$ be compact neighborhoods of x, y respectively. Then $N \times M$ is a compact neighborhood of $x \times y$. Thus a finite product of locally compact spaces is locally compact. In particular, the product of a locally compact space with a compact space is locally compact, and since the product of any number of compact spaces is compact by Tychonov's theorem, the 'if' direction is proven.

Now assume $X = \prod_i X_i$ is locally compact. The projections $p_i : \prod_k X_k \to X_i$ are continuous and open (Proposition 6.5.8). Thus each X_i is locally compact by Lemma 7.8.36. Now let $x \in U \subseteq K \subseteq X$ with U open and K compact. By definition of the product topology, we have $x \in V \subseteq U$, where $V = p_{i_1}^{-1}(U_1) \cap \cdots \cap p_{i_n}^{-1}(U_n)$ with $U_k \in \tau_{i_k}$. Thus

$$K \supseteq p_{i_1}^{-1}(U_1) \cap \cdots \cap p_{i_n}^{-1}(U_n) = \prod_{j \notin \{i_1, \ldots, i_n\}} X_j \ \times\ \prod_{k=1}^{n} U_k.$$

Thus for $j \notin \{i_1, \ldots, i_n\}$ we have $p_j(K) \supseteq p_j(U) = X_j$ and thus $p_j(K) = X_j$. Since K is compact and p_j continuous, X_j is compact. Thus with the (possible) exception of $X_{i_1}, \ldots, X_{i_n}$ all factors are compact. ■

Remark 7.8.38 Thus with $I = [0, 1]$, the space $I^n \times \mathbb{R}^m$ is locally compact if and only if $m < \infty$, whereas n can be any cardinal number n. It is therefore not quite correct to think of locally compact spaces as 'finite dimensional'. At best they are 'finite dimensional modulo compact factors', but this is difficult to make precise. (But a Banach space is locally compact if and only if it is finite dimensional. Cf. e.g. [204, Appendix B].) □

Remark 7.8.39 1. We now have clarified completely the behavior of local compactness w.r.t. sums, products, subspaces (the latter only in the Hausdorff case) and under open maps. In particular, quotients of locally compact spaces by open equivalence relations are locally compact. When f is not open, $f(X)$ need not be locally compact. In particular, images under closed maps need not be locally compact. But in Theorem 7.8.66, that the image $f(X)$ of a locally compact Hausdorff space

X is locally compact Hausdorff when f is continuous, closed and proper. (Properness will be defined very soon.)

2. The above implies that quotients of locally compact spaces by arbitrary equivalence relations are a proper generalization of locally compact spaces, called 'k-spaces' or 'compactly generated spaces'. Cf. Section 7.9, in particular Proposition 7.9.14. □

Exercise 7.8.40 Let X, Y be Hausdorff spaces, where X is compact and Y locally compact. Let $\sim$ be the equivalence relation on $X \times Y_\infty$ that identifies all points (x, ∞) with each other, doing nothing else.

(i) Give examples for X and Y with these properties such that $(X \times Y)_\infty$, the 1-point compactification of $X \times Y$ is not homeomorphic to $X \times Y_\infty$.

(ii) Prove that $(X \times Y_\infty)/\sim$ is homeomorphic to $(X \times Y)_\infty$.

(iii) Prove that $\mathbb{R}^2\backslash\{(0,0)\}$ is homeomorphic to $S^1 \times \mathbb{R}$.

(iv) Use (i) and (ii) to give a description of $(\mathbb{R}^2\backslash\{(0,0)\})_\infty$

(v) Make an instructive drawing of the situation in (v).

Exercise 7.8.41 Let X, Y be locally compact Hausdorff spaces. Prove that $(X \times Y)_\infty$ is homeomorphic to a quotient space of $X_\infty \times Y_\infty$. Hint: Use the obvious surjective map

In the context of metrization, we will be interested in whether (X_∞, τ_∞) is second countable. In order to answer this in Exercise 7.8.44, we need some preparations, which will also be crucial in the discussion of proper metric spaces in Section 7.8.8.

Definition 7.8.42 *A topological space X is called ...*

- *<u>σ-compact</u> if it admits a countable compact cover, i.e. a countable family $\{K_i\}_{i\in\mathbb{N}}$ of compact subsets such that $\bigcup_i K_i = X$.*
- *<u>hemicompact</u> if there are compact sets $\{K_i\}_{i\in\mathbb{N}}$ satisfying $\bigcup_{i\in\mathbb{N}} K_i = X$ and such that for every compact $K \subseteq X$ there is an $n \in \mathbb{N}$ such that $K \subseteq K_n$.*

Exercise 7.8.43 Let X be a topological space. Consider the following statements:

(α) There are open sets $\{U_i\}_{i\in\mathbb{N}}$ such that each $\overline{U_i}$ is compact, $\overline{U_i} \subseteq U_{i+1}$ $\forall i$, and $\bigcup_{i\in\mathbb{N}} U_i = X$.

(β) X is hemicompact.

(γ) X is σ-compact.

(δ) X is Lindelöf.

Prove:

(i) (α) $\Rightarrow$ locally compact (weakly).

(ii) (α) $\Rightarrow$ (β) $\Rightarrow$ (γ) $\Rightarrow$ (δ).

(iii) For X locally compact, (δ) $\Rightarrow$ (γ).

(iv) For X locally compact Hausdorff, $(\delta) \Rightarrow (\alpha)$.

Thus for X locally compact Hausdorff, the four properties are equivalent.

Exercise 7.8.44 Let X be any space and X_∞ its one-point compactification.

(i) Prove that X_∞ is second countable if and only if X is second countable and $\infty \in X_\infty$ has a countable open neighborhood base.

(ii) Use (i) and Exercise 7.8.43 to prove that if X is locally compact Hausdorff and second countable then X_∞ is second countable.

The following will not be used in the sequel, but it nicely complements Exercise 7.8.43, and the diagonal argument used in the proof is interesting:

Proposition 7.8.45 *Every first countable hemicompact Hausdorff space is locally compact. (Thus in particular hemicompact metrizable spaces are locally compact!)*

Proof. By hemicompactness we have compact sets $\{K_n\}_{n\in\mathbb{N}}$ such that $\bigcup_n K_n = X$ and every compact K is contained in some K_n. Replacing K_n by $\bigcup_{k=1}^n K_k$ if necessary we may assume $K_1 \subseteq K_2 \subseteq \cdots$. Let $x \in X$ and $U_1 \supseteq U_2 \supseteq \cdots$ an open neighborhood base for x. Suppose that $\overline{U_i}$ is non-compact for all $i \in \mathbb{N}$. Since X is Hausdorff, each K_i is closed. Thereforewe have $U_i \not\subseteq K_i\ \forall i$ (since otherwise $\overline{U_i} \subseteq \overline{K_i} = K_i$ would be compact). Thus we can choose an $x_i \in U_i \backslash K_i$ for each $i \in \mathbb{N}$. Clearly $x_i \to x$. This implies that the set $K = \{x, x_1, x_2, \ldots\}$ is compact. (Cf. Exercise 7.3.8.) Thus $K \subseteq K_n$ for some n, and in particular $x_n \in K_n$. This contradicts the choice of $x_n \in U_n \backslash K_n$. Thus some $\overline{U_i}$ is compact and therefore a compact neighborhood of x. ■

7.8.5 Continuous extensions of $f : X \to Y$ to X_∞. Proper maps

The developments of this section are motivated by the following question: Given a continuous function $f : X \to Y$, is there a continuous extension $\widehat{f} : X_\infty \to Y_\infty$? And if so, is it unique? We notice the following:

- If X is compact then any extension $\widehat{f} : X_\infty \to Y$ is continuous. (This follows from the fact, cf. Theorem 7.8.14(iv), that $X_\infty = X \oplus \{\infty\}$, thus ∞ is an isolated point.)

- If X is non-compact and Y is Hausdorff then there is at most one continuous extension $\widehat{f} : X_\infty \to Y$ of f. (This follows from density of X in X_∞ and Exercise 6.5.18(iii).)

- If $f : X \to Y$ is continuous, then so is $f : X \to Y \hookrightarrow Y_\infty$. Thus if X is non-compact and Y is locally compact Hausdorff, there is at most one continuous extension $\widehat{f} : X_\infty \to Y_\infty$.

Concerning the difficult problem of existence of continuous extensions, we have the powerful result Theorem 7.4.20. But the present situation being simpler since $X_\infty \backslash X$ consists only of one point, we prefer an elementary direct approach.

Lemma 7.8.46 *Let X be non-compact, $f : X \to Y$ continuous and $y_0 \in Y$. Define the extension $\widehat{f} : X_\infty \to Y$ by $\widehat{f}(\infty) = y_0$. Then the following are equivalent:*

(i) $\widehat{f}$ is continuous.

(ii) $X \backslash f^{-1}(U)$ is compact whenever $y_0 \in U \in \tau_Y$.

(iii) For every open $U \subseteq Y$ containing y_0 there is a compact $K \subseteq X$ such that $f(X\backslash K) \subseteq U$.

Proof. $\widehat{f}$ is continuous if and only if $\widehat{f}^{-1}(U) \subseteq X_\infty$ is open for every open $U \subseteq Y$. If $y_0 \notin U$ then $\widehat{f}^{-1}(U) = f^{-1}(U) \subseteq X$, which is open by continuity of f. If $y_0 \in U$ then $\widehat{f}^{-1}(U) = f^{-1}(U) \cup \{\infty\}$. By definition of τ_∞, this is open if and only if $X_\infty\backslash\widehat{f}^{-1}(U) = K$, where $K \subseteq X$ is closed and compact. Since $X_\infty\backslash\widehat{f}^{-1}(U) = X\backslash f^{-1}(U)$, we find that $\widehat{f}$ is continuous if and only if $K = X\backslash f^{-1}(U)$ is closed and compact whenever $y_0 \in U \in \tau$. Closedness being automatic by continuity of f, we have (i)⇔(ii). Now, if $K = X\backslash f^{-1}(U)$ is compact then $f^{-1}(U) = X\backslash K$, which implies $f(X\backslash K) \subseteq f(f^{-1}(U)) \subseteq U$, i.e. (iii). If, conversely, there is a compact K' such that $f(X\backslash K') \subseteq U$ then $f^{-1}(U) \supseteq X\backslash K'$, thus $X\backslash f^{-1}(U) \subseteq K'$, and as a closed subset of the compact K', this is compact. Thus (ii) holds. ■

Lemma 7.8.47 *Let $f : X \to Y$ be continuous. Then the extension $\widehat{f} : X_\infty \to Y_\infty$ satisfying $\widehat{f}(\infty_X) = \infty_Y$ is continuous if and only if $f^{-1}(K) \subseteq X$ is compact for every closed compact $K \subseteq Y$.*

Proof. Let f' be the composite $X \xrightarrow{f} Y \xrightarrow{\iota} Y_\infty$. f' clearly is continuous. Applying Lemma 7.8.47 to $f' : X \to Y_\infty$ and $y_0 = \infty_Y$ we find that $\widehat{f} = \widehat{f'} : X_\infty \to Y_\infty$ is continuous if and only if $X\backslash f'^{-1}(U)$ is compact whenever $\infty_Y \in U \in \tau_{Y_\infty}$. Such a U is of the form $Y_\infty\backslash K$, where $K \subseteq Y$ is compact. Now $X\backslash f'^{-1}(Y_\infty\backslash K) = X\backslash(X\backslash f^{-1}(K)) = f^{-1}(K)$. We thus find that $\widehat{f}$ is continuous if and only if $f^{-1}(K) \subseteq X$ is compact for every closed and compact $K \subseteq Y$. ■

Remark 7.8.48 If X is compact, the assumptions in both lemmas are trivially satisfied. This again shows that $\widehat{f}(\infty_X)$ is arbitrary when X is compact. □

The above results motivate the following definitions:

Definition 7.8.49 *Let $f : X \to Y$ be a function.*

(i) We say that f <u>tends to $y_0 \in Y$ at infinity</u> if for every open $U \subseteq Y$ containing y_0 there is a compact $K \subseteq X$ such that $f(X\backslash K) \subseteq U$.

(ii) f is called <u>proper</u> if $f^{-1}(K) \subseteq X$ is compact for every compact $K \subseteq Y$.

Remark 7.8.50 1. Note that we wrote 'compact' instead of 'closed compact'.

2. Some authors include continuity and/or closedness in the definition of proper functions, but we don't. Others write 'perfect' instead of 'proper' or require only compactness of $f^{-1}(y)$ for every $y \in Y$. Cf. Section 7.8.7 for some implications between the various definitions.

3. If $f : X \to Y$ and $g : Y \to Z$ are proper, it is obvious that $g \circ f : X \to Z$ is proper.

4. It is clear that homeomorphisms are proper.

5. A strange feature of the definition is that for compact X, every $f : X \to Y$ tends to $y_0 \in Y$ at infinity irrespective of y_0. □

Exercise 7.8.51 Let $f : X \to Y$ be a function (not a priori continuous) between topological spaces. Prove the following statements:

(i) If X is Hausdorff, Y is compact and f is proper then f is continuous.

(ii) If X is compact, Y is Hausdorff and f is continuous then f is proper.

(iii) If X, Y are compact Hausdorff then f is continuous if and only if it is proper.

Using the terminology of Definition 7.8.49, we can reformulate our results as follows:

Corollary 7.8.52 *Let $f : X \to Y$ a function. Then*

(i) For the extension $\widehat{f} : X_\infty \to Y_\infty$ satisfying $\widehat{f}(\infty_X) = \infty_Y$ we have the implications

$$f \text{ continuous and proper} \Rightarrow \ \widehat{f} \text{ continuous} \Rightarrow f \text{ continuous.}$$

If Y is Hausdorff then $\widehat{f}$ is continuous $\Leftrightarrow$ f is continuous and proper.

(ii) If X, Y are Hausdorff and $f : X \to Y$ is continuous, a continuous extension $\widehat{f} : X_\infty \to Y_\infty$ of f exists if and only if f is proper or tends to $y_0 \in Y$ at infinity. In these cases we must define $\widehat{f}(\infty_X) = \infty_Y$ or $\widehat{f}(\infty_X) = y_0$, respectively.

Proof. All this follows immediately from Lemmas 7.8.46, 7.8.47 once we note that for a subset of a Hausdorff space 'compact' and 'closed compact' are equivalent. ■

Exercise 7.8.53 Let X be non-compact, Y locally compact Hausdorff and $g \in C(X, Y)$. Give an alternative proof of Corollary 7.8.52(ii) by using Theorem 7.4.20.

Definition 7.8.54 *A net $\{x_\iota\}_{\iota \in I}$, in a topological space 'tends to infinity' or 'leaves every compact set', '$x_\iota \to \infty$' if for every compact $K \subseteq X$ there is $\iota_0 \in I$ such that $\iota \geq \iota_0 \Rightarrow x_\iota \notin K$.*

(This terminology does not imply that there is an 'infinite point', as in X_∞!) Nets provide a useful perspective on the notions of properness and tending to y_0 at infinity:

Exercise 7.8.55 Let X, Y be topological spaces, $f : X \to Y$ a function and $\{x_\iota\}$ a net in X. Prove:

(i) If $x_\iota \to \infty$ (in the sense of Definition 7.8.54) then $\iota_\infty(x_\iota)$ converges to the point $\infty \in X_\infty$.

(i') The converse of (i) holds when X is Hausdorff.

(ii) Let $f : X \to Y$ be continuous. If f tends to $y_0 \in Y$ at infinity (in the sense of Definition 7.8.49(i)) then $f(x_\iota) \to y_0$ whenever $x_\iota \to \infty$ (in the sense of Definition 7.8.54).

(ii') If X is Hausdorff then the converse of (ii) holds. Hint: Use $\widehat{f} : X_\infty \to Y$ with $\widehat{f}(\infty) = y_0$.

(iii) If $f : X \to Y$ is proper then $f(x_\iota) \to \infty_Y$ whenever $x_\iota \to \infty_X$.

(iii') If f is continuous and X, Y are Hausdorff then the converse of (iii) holds.
Hint: Use (i,i') to prove continuity of the extension $\widehat{f} : X_\infty \to Y_\infty$ with $\widehat{f}(\infty_X) = \infty_Y$.

7.8.6 Functoriality and universal property of X_∞

Definition 7.8.56 *By $\mathcal{TOP}$ we denote the category consisting of topological spaces and continuous maps, by $\mathcal{TOP}_c, \mathcal{TOP}_{lc}, \mathcal{TOP}_{cH}, \mathcal{TOP}_{lcH}$ the full subcategories of compact, respectively locally compact, (Hausdorff) spaces. A superscript p (as in $\mathcal{TOP}^p$) means that we consider the (non-full) subcategory having only proper maps.*

Corollary 7.8.57 *One-point compactification $X \mapsto X_\infty$ of spaces and extension $f \mapsto \widehat{f}$ by $\widehat{f}(\infty) = \infty$ of proper maps define a functor $F : \mathcal{TOP}^p \to \mathcal{TOP}_c$. The latter restricts to a functor $\mathcal{TOP}^p_{lcH} \to \mathcal{TOP}_{cH}$.*

Proof. All that remains to be verified is that $F(\mathrm{id}_X) = \mathrm{id}_{X_\infty}$ and $F(g \circ f) = F(g) \circ F(f)$ for proper continuous maps $X \stackrel{f}{\to} Y \stackrel{g}{\to} Z$, but both statements are entirely obvious. ■

As we saw in Example 7.8.4, a space may have many different compactifications. It is intuitively is clear that the 1-point compactification X_∞ is minimal in that it adds only one point. The following results make this precise:

Theorem 7.8.58 *Let (X, τ) be locally compact Hausdorff and non-compact. Let $(\widehat{X}, \widehat{\tau})$ be any Hausdorff compactification of X with embedding $\iota : X \hookrightarrow \widehat{X}$. Define a function $f : \widehat{X} \to X_\infty$ to the one-point compactification (X_∞, τ_∞) by*

$$f(z) = \begin{cases} \iota_\infty(x) & \text{if } z = \iota(x),\ x \in X \\ \infty & \text{if } z \in \widehat{X} \backslash \iota(X) \end{cases}$$

Then f is continuous, and the diagram

$$\begin{tikzcd} X \arrow[r, "\iota"] \arrow[dr, "\iota_\infty"'] & \widehat{X} \arrow[d, "f"] \\ & X_\infty \end{tikzcd}$$

commutes. Furthermore, f is the only function $\widehat{X} \to X_\infty$ with these two properties.

Proof. Commutativity of the diagram means $f(\iota(x)) = \iota_\infty(x)\ \forall x \in X$, which is true by the very definition of f.

We may and will identify X with its images $\iota(X) \subseteq \widehat{X}$ and $\iota_\infty(X) \subseteq X_\infty$. Since ι is an embedding, $X \subseteq \widehat{X}$ is locally compact. Since $\widehat{X}$ is Hausdorff, Proposition 7.8.31 implies that $X \subseteq \widehat{X}$ is locally closed, which by Exercise 7.8.28 is equivalent to X being open in $\overline{X}$. Together with $\overline{X} = \widehat{X}$, which holds since $\widehat{X}$ is a compactification of X, this gives $X \in \widehat{\tau}$. As in the proof of Lemma 7.8.10(i) this implies $\tau = \widehat{\tau} \upharpoonright X = \{U \cap X \mid U \in \widehat{\tau}\} \subseteq \widehat{\tau}$.

Now if $U \in \tau \subseteq \tau_\infty$ then $f^{-1}(U) = U \subseteq \widehat{X}$, since the 'infinite' points of $\widehat{X}$, i.e. the set $\widehat{X} \backslash X$, are mapped to $\infty \in X_\infty$. In view of $\tau \subseteq \widehat{\tau}$, we have $f^{-1}(U) \in \widehat{\tau}$. On the other hand, if $U \in \tau_\infty$ is of the form $X_\infty \backslash K$ with $K \subseteq X$ compact, then

$$f^{-1}(U) = \widehat{X} \backslash f^{-1}(K) = \widehat{X} \backslash K$$

is open since $K \subseteq \widehat{X}$ is compact and thus closed, $\widehat{X}$ being Hausdorff. Thus f is continuous.

It remains to prove the uniqueness claim. So let $g \in C(\widehat{X}, X_\infty)$ satisfying $g \circ \iota = \iota_\infty$. Then g coincides with the function f defined above on X, which is dense a dense subset of $\widehat{X}$ by definition of a compactification. Since X is locally compact Hausdorff, X_∞ is Hausdorff. Now Exercise 6.5.18(iii) implies $f = g$. ■

Theorem 7.8.58 is called the universal property of X_∞. The latter has a concise formulation in categorical language, and also a converse:

Corollary 7.8.59 *(i) If X is non-compact locally compact Hausdorff, the one-point compactification X_∞ is a terminal object in the category $\mathcal{C}(X)$ of Hausdorff compactifications.*

(ii) If the category $\mathcal{C}(X)$ of Hausdorff compactifications of X has a terminal object $(\widehat{X},\iota)$ then either X is compact and $\widehat{X}\cong X$, or X is non-compact locally compact Hausdorff and $(\widehat{X},\iota)\cong(X_\infty,\iota_\infty)$.

Proof. (i) By Theorem 7.8.58, there is a unique $f\in \mathrm{Hom}_{\mathcal{C}(X)}((Y,\iota),(X_\infty,\iota_\infty))$ for every $(Y,\iota)\in\mathcal{C}(X)$.

(ii) If X is compact then all objects of $\mathcal{C}(X)$ are isomorphic to X, which then is also terminal. Thus assume X to be non-compact. Let $(\widehat{X},\iota)\in\mathcal{C}(X)$. If $\widehat{X}\backslash\iota(X)$ has more than one element, choose two $x,y\in\widehat{X}\backslash\iota(X)$ and define $\widehat{X}'=\widehat{X}/\{x,y\}$ with quotient map $p:\widehat{X}\to\widehat{X}'$. Then $\widehat{X}'$ is compact and Hausdorff (by Exercise 6.4.16), so that $(\widehat{X}',p\circ\iota)$ is a compactification of X and strictly smaller than $(\widehat{X},\iota)$ w.r.t. the order $\leq$ on $(\mathcal{C}(X)/\cong,\leq)$. Thus $[(\widehat{X},\iota)]$ is not a smallest element of $(\mathcal{C}(X)/\cong,\leq)$, thus $(\widehat{X},\iota)$ not terminal in $\mathcal{C}(X)$ by Proposition 7.8.7(v). Thus a terminal object $(\widehat{X},\iota)$ has only one infinite point. Since $\widehat{X}$ by assumption is Hausdorff, Lemma 7.8.10 gives that X is locally compact Hausdorff and $(\widehat{X},\iota)\cong(X_\infty,\iota_\infty)$. ■

Remark 7.8.60 1. Theorem 7.8.58 generalizes the uniqueness result of Theorem 7.8.14(vii) and implies the latter easily: If τ' is a topology on $X_\infty=X\cup\{\infty\}$ such that $X\hookrightarrow(X_\infty,\tau')$ is a Hausdorff compactification, then the map $f:(X_\infty,\tau')\to(X_\infty,\tau_\infty)$ from Theorem 7.8.58 is a continuous bijection between compact Hausdorff spaces and therefore a homeomorphism. Thus $\tau'=\tau_\infty$.

2. By Corollary 7.8.59, the locally compact Hausdorff spaces are characterized (among the non-compact Hausdorff spaces) by the property of having a smallest (in the sense of being terminal) compactification. We will see later that *every* space X that admits a Hausdorff compactification has a compactification βX that is 'biggest' in the sense of being an initial object in $\mathcal{C}(X)$. In view of Proposition 7.8.7(ii), every compactification of X is a quotient of the maximal one, which makes βX much more important than X_∞.

3. Assume that X is non-compact and admits exactly one Hausdorff compactification $\widehat{X}$ (up to isomorphism in $\mathcal{C}(X)$). Then the unique compactification $\widehat{X}$ must be the one-point compactification. (Otherwise the argument in the proof of the corollary shows that we can obtain a different compactification $\widehat{X}'$ by identifying two points in $\widehat{X}\backslash X$, contradicting the uniqueness of $\widehat{X}$.) Thus X must be locally compact, but since this does not imply uniqueness of compactifications, X must satisfy an additional condition. This will be studied later, cf. Corollary 8.3.35. □

7.8.7 Further applications of properness

We now consider generalizations of several results from Section 7.4.2, beginning with a locally compact version of Proposition 7.4.11:

Proposition 7.8.61 *If X is an arbitrary space, Y is locally compact Hausdorff and $f:X\to Y$ is continuous and proper then f is closed.*

Proof. By Corollary 7.8.52, the extension $\widehat{f}:X_\infty\to Y_\infty$ with $\widehat{f}(\infty_X)=\infty_Y$ is continuous. Since X_∞ is compact and Y_∞ is Hausdorff, $\widehat{f}$ is closed by Proposition 7.4.11. If $C\subseteq X$ is closed then $C\cup\{\infty\}\subseteq X_\infty$ is closed by Exercise 7.8.18, so that $\widehat{f}(C\cup\{\infty\})\subseteq Y_\infty$ is closed. Now $Y_\infty\backslash\widehat{f}(C\cup\{\infty\})=Y\backslash f(C)$ is open in Y_∞ and thus (since it does not contain ∞_Y) in Y, thus $f(C)\subseteq Y$ is closed. ■

Corollary 7.8.62 *If X is an arbitrary space, Y is locally compact Hausdorff and $f:X\to Y$ is bijective, continuous and proper then f is a homeomorphism.*

Proof. By Proposition 7.8.61, f is closed. Now apply Lemma 5.2.26. ■

Cf. Exercises 7.8.70 and 7.8.71 for another proof and an application.

The following locally compact version of Proposition 7.4.11(iii) is an important ingredient in the proof of the Whitney embedding theorem for non-compact manifolds:

Proposition 7.8.63 *Let X be arbitrary, Y locally compact Hausdorff and $f : X \to Y$ continuous and injective. Then X is Hausdorff, and the following are equivalent:*

(i) f is proper.

(ii) f is closed.

(iii) $f(X)$ is closed and $f : X \to Y$ is an embedding.

If one (thus all) of these statements holds then X is locally compact.

Proof. Hausdorffness of X is just Exercise 5.2.15.

(i)⇒(ii): This is Proposition 7.8.61.

(ii)⇒(iii): f being closed, $f(X) \subseteq Y$ is closed. That f is an embedding follows from Lemma 6.2.9(ii).

(iii)⇒(i): Let $K \subseteq Y$ be compact. Since $f(X)$ is closed, $K \cap f(X)$ is compact by Exercise 7.4.7(ii), thus $f^{-1}(K) = f^{-1}(K \cap f(X))$ is compact since $f : X \to f(X)$ is a homeomorphism.

As to the last claim, assume (iii) holds. If $x \in X$, let $f(x) \in U \subseteq K \subseteq Y$ with U open and K compact. Then $x \in f^{-1}(U) \subseteq f^{-1}(K)$, where $f^{-1}(U)$ is open by continuity of f and $f^{-1}(K)$ is compact by properness. Thus X is locally compact. ■

Remark 7.8.64 The above proposition provides a necessary and sufficient condition for a continuous injection of locally compact Hausdorff spaces to be a *closed* embedding, i.e. an embedding with closed image. There clearly are non-closed embeddings, e.g. $X \hookrightarrow X_\infty$ for non-compact X. □

In Propositions 7.8.61 and 7.8.63, Y was assumed to be Hausdorff. If we want to *prove* that $f(X)$ is Hausdorff, we cannot use Proposition 7.8.61 and must *assume* closedness of f instead. Recall, however, from Example 6.4.19 that continuous images of Hausdorff spaces can fail to be Hausdorff even if the map is closed. This cannot happen if the map is also proper:

Proposition 7.8.65 *Let X be Hausdorff and $f : X \to Y$ surjective, closed and proper. Then Y is Hausdorff. (Continuity is not needed here.)*

Proof. Let $x, y \in Y$ with $x \neq y$. Since f is proper and singletons are compact, the disjoint subsets $C = f^{-1}(x)$, $D = f^{-1}(y)$ of X are compact, thus closed since X is Hausdorff. Now by Exercise 7.4.5, there are disjoint open $U, V \subseteq X$ such that $C \subseteq U$, $D \subseteq V$. Since $X\backslash U$, $X\backslash V$ are closed, $f(X\backslash U)$, $f(X\backslash V) \subseteq Y$ are closed by closedness of f. Consequentially, $Y\backslash f(X\backslash U)$, $Y\backslash f(X\backslash V) \subseteq Y$ are open. Furthermore, $f^{-1}(x) \subseteq U$, thus $x \notin f(X\backslash U)$, and therefore $x \in Y\backslash f(X\backslash U) =: U'$. In the same way, one obtains $y \in Y\backslash f(X\backslash V) =: V'$. Finally,

$$\begin{aligned}
[Y\backslash f(X\backslash U)] \cap [Y\backslash f(X\backslash V)] &= Y\backslash[f(X\backslash U) \cup f(X\backslash V)] = Y\backslash[f((X\backslash U) \cup (X\backslash V))] \\
&= Y\backslash f(X\backslash(U \cap V)) = Y\backslash f(X) = Y\backslash Y = \emptyset,
\end{aligned}$$

where we used surjectivity, i.e. $f(X) = Y$. Thus $U', V' \subseteq Y$ are disjoint open neighborhoods of x and y, respectively, and Y is Hausdorff. ■

Theorem 7.8.66 *Let X be locally compact Hausdorff and $f : X \to Y$ surjective, continuous closed and proper. Then Y is locally compact Hausdorff (and f is a quotient map).*

Proof. The claim in brackets just is Lemma 6.4.5. By Proposition 7.8.65, Y is Hausdorff. If $y \in Y$, properness of f implies that $K = f^{-1}(y) \subseteq X$ is compact. Applying Proposition 7.8.25(i) with $U = X$, we obtain an open $V \subseteq X$ such that $K \subseteq V$ and $\overline{V}$ is compact. As in the proof of Proposition 7.8.65 we see that $W = Y \backslash f(X \backslash V)$ is an open neighborhood of y. Furthermore, $W = Y \backslash f(X \backslash V) \subseteq Y \backslash f(X \backslash \overline{V}) \subseteq f(\overline{V})$, where the first inclusion is obvious and the second is due to surjectivity of f. Since $f(\overline{V})$ is compact as continuous image of the compact set $\overline{V}$ we see that $\overline{W}$ is compact, and thus a compact neighborhood of $y \in Y$. ■

Notice that the two preceding results use properness only via compactness of $f^{-1}(y)$. Together with closedness, this weaker assumption actually implies properness:

Proposition 7.8.67 *If $f : X \to Y$ is closed and $f^{-1}(y) \subseteq X$ is compact for every $y \in Y$, then f is proper.*

Proof. Let $K \subseteq Y$ be compact, and let $\{U_i\}_{i \in I}$ be a family of open sets in X such that $\bigcup_i U_i \supseteq f^{-1}(K)$. Let $\mathcal{J}$ be the family of finite subsets of I, and for $J \in \mathcal{J}$ let $U_J = \bigcup_{j \in J} U_j$. In view of Lemma 7.3.2, we need to find $J \in \mathcal{J}$ such that $f^{-1}(K) \subseteq U_J$. Since $\{U_i\}$ is an open cover of $f^{-1}(y)$, which is compact by assumption, for each $y \in Y$, there is $J_y \in \mathcal{J}$ such $f^{-1}(y) \subseteq U_{J_y}$. Since $X \backslash U_J$ and f are closed, $f(X \backslash U_J)$ is closed, thus $V_J = Y \backslash f(X \backslash U_J)$ is open. By definition, $y \notin f(X \backslash U_{J_y})$, thus $y \in V_{J_y}$. Therefore, $\{V_J\}_{J \in \mathcal{J}}$ is an open cover of K, and by compactness of K, there are $J_1, \ldots, J_n \in \mathcal{J}$ such that $K \subseteq V_{J_1} \cup \cdots \cup V_{J_n}$. Defining $S = J_1 \cup \cdots \cup J_n$, we have $S \in \mathcal{J}$ (as a finite union of finite sets). Now,

$$\begin{aligned} f^{-1}(K) &\subseteq \bigcup_{k=1}^{n} f^{-1}(V_{J_k}) = \bigcup_{k=1}^{n} f^{-1}(Y \backslash f(X \backslash U_{J_k})) = \bigcup_{k=1}^{n} (X \backslash f^{-1}(f(X \backslash U_{J_k}))) \\ &\subseteq \bigcup_{k=1}^{n} (X \backslash (X \backslash U_{J_k})) = \bigcup_{k=1}^{n} U_{J_k} = U_S, \end{aligned}$$

and we are done. (We have used that $f^{-1}(f(X \backslash U_{J_k})) \supseteq X \backslash U_{J_k}$.) ■

Here is a negative result concerning the existence of continuous extensions:

Lemma 7.8.68 *Let X be Hausdorff, $A \subsetneqq X$ dense and $f : A \to Y$ continuous, closed and proper. Then f admits no continuous extension $\widehat{f} : X \to Y$.*

Proof. Assume that a continuous extension $\widehat{f} : X \to Y$ does exist. Pick $x \in X \backslash A$ and replace X by $A \cup \{x\}$. Then $\overline{A} = X$ still holds. Now $f^{-1}(\widehat{f}(x)) \subseteq A$ is compact by properness of f and does not contain x. Then by Lemma 7.4.1 there are disjoint open subsets U, V of X such that $x \in U$ and $f^{-1}(\widehat{f}(x)) \subseteq V$. In view of $x \notin V$, we have $\widehat{f}^{-1}(f(A \backslash V)) = f^{-1}(f(A \backslash V)) \subseteq A$. Since $A \backslash V = A \cap (X \backslash V)$ is closed in A, we have that $f(A \backslash V) \subseteq Y$ is closed (by closedness of f), thus $\widehat{f}^{-1}(f(A \backslash V)) \subseteq X$ is closed (by continuity of $\widehat{f}$). Together with the trivial $A \backslash V \subseteq \widehat{f}^{-1}(f(A \backslash V))$, this implies $\overline{A \backslash V} \subseteq \widehat{f}^{-1}(f(A \backslash V)) \subseteq A$ (closure in X). But now we have

$$x \in U \subseteq X \backslash V = \overline{A} \cap \overline{X \backslash V} \subseteq \overline{A \cap (X \backslash V)} = \overline{A \backslash V} \subseteq A,$$

thus $x \in A$, contradicting our choice of $x \in X \backslash A$. ■

Remark 7.8.69 At first sight, this lemma contradicts Lemma 7.8.47 in the locally compact situation, according to which an extension of $f : X \to Y$ to X_∞ does exist when f is proper. But if X is non-compact, Y locally compact Hausdorff and $f : X \to Y$ continuous and proper, Lemma 7.8.47 and the discussion about uniqess preceding it say that the extension $\widehat{f} : X_\infty \to Y_\infty$ with $\widehat{f}(\infty) = \infty$ is the only continuous extension. Thus in particular, there is no extension that takes values in Y, consistent with Lemma 7.8.68 asserts. □

Exercise 7.8.70 There is a short direct proof of Corollary 7.8.62, not using the one-point compactification. Find it.

Exercise 7.8.71 (Roots of complex polynomials depend continuously on coefficients) If $z_1, \ldots, z_n$ are complex numbers (not necessarily different) then

$$P : z \mapsto \prod_{i=1}^{n}(z - z_i) = z^n + a_{n-1}z^{n-1} + \cdots + a_0 =: P_a(z)$$

is a monic polynomial of degree n. This defines a map $\sigma_0 : \mathbb{C}^n \to \mathbb{C}^n : (z_1, \ldots, z_n) \mapsto (a_0, \ldots, a_{n-1})$. Let S_n act on $\mathbb{C}^n$ by permutations: $\pi(z_1, \ldots, z_n) = (z_{\pi(1)}, \ldots, z_{\pi(n)})$. Then clearly $\sigma_0(\pi(z)) = \sigma_0(z)$ for all $z = (z_1, \ldots, z_n),\ \pi \in S_n$. Prove:

(i) σ_0 is continuous and gives rise to a continuous map $\sigma : \mathbb{C}^n/S_n \to \mathbb{C}^n$.

(ii) σ has an inverse map $\tau : \mathbb{C}^n \to \mathbb{C}^n/S_n$, defined using the Fundamental Theorem of Algebra (Theorem 7.7.57).

(iii) If $z \in \mathbb{C}$ satisfies $|z| > 1$ and $z^n + a_{n-1}z^{n-1} + \cdots + a_0 = 0$ then $|z| \le |a_0| + \cdots + |a_{n-1}|$.

(iv) $\sigma : \mathbb{C}^n/S_n \to \mathbb{C}^n$ is proper.

(v) $\sigma : \mathbb{C}^n/S_n \to \mathbb{C}^n$ is a homeomorphism, thus τ is continuous.

Remark: Continuity of $\tau : \mathbb{C}^n \to \mathbb{C}^n/S_n, (a_0, \ldots, a_{n-1}) \mapsto [(z_1, \ldots, z_n)]$ means that the solutions of a monic polynomial P of degree n depend continuously on the coefficients of P. (The standard proofs use Rouché's theorem from complex analysis, cf. e.g. [236, Appendix V, Theorem 4A].)

Remark 7.8.72 1. From (v) one can deduce the following: If $a = (a_0, \ldots, a_{n-1}) \in \mathbb{C}^n$ and $z \in \mathbb{C}$ is a single root of P_a then there are an open neighborhood $U \subseteq \mathbb{C}^n$ of a and a continuous map $f : U \to \mathbb{C}$ such that $P_{a'}(f(a')) = 0$ for all $a' \in U$.
2. Using an appropriate implicit function theorem one can prove that f is analytic. □

7.8.8 Proper metric spaces

In this section we discuss an application of properness to metric spaces.

Exercise 7.8.73 Let (X, d) be a metric space. Prove that the following are equivalent:

(i) (X, d) is locally compact.

(ii) For every $x \in X$ there is $R > 0$ such that $\overline{B}(x, R)$ is compact.

Example 7.8.74 If $X = (0,1)$ with $d(x,y) = |x-y|$, then $\overline{B}(x,R)$ is compact if and only if $R < \min(x, 1-x)$. □

Exercise 7.8.75 Let (X,d) be a metric space. Prove that the following are equivalent:

(ia) The map $X \to [0,\infty),\ x \mapsto d(x,x_0)$ is proper for every $x_0 \in X$.

(ib) The map $X \to [0,\infty),\ x \mapsto d(x,x_0)$ is proper for some $x_0 \in X$.

(iia) The closed balls $\overline{B}(x,r) = \{y \in X \mid d(y,x) \le r\}$ are compact for all $x \in X,\ r > 0$. (Keep in mind Exercise 2.2.11.)

(iib) The closed balls $\overline{B}(x,r)$ are compact for some $x \in X$ and all $r > 0$.

(iii) Every closed bounded $Y \subseteq X$ is compact.

Definition 7.8.76 *A metric space (X,d) is called ...*

- *proper if it satisfies the equivalent conditions in Exercise 7.8.75.*
- *properly metrizable if there exists a proper metric d' on X that is equivalent to d.*

Remark 7.8.77 1. It is very easy to see (but quite useless) that a metric space is compact if and only if it is proper and bounded.

2. The interest of proper metric spaces is that they behave almost like $\mathbb{R}^n$ in that a Heine-Borel style result holds. In fact, they share more properties with $\mathbb{R}^n$: □

Lemma 7.8.78 *Every proper metric space (X,d) is complete, locally compact, second countable, separable, Lindelöf, σ-compact, hemicompact.*

Proof. If $x \in X,\ r > 0$ then $x \in B(x,r) \subseteq \overline{B}(x,r)$, where $\overline{B}(x,r)$ is compact. Thus $\overline{B}(x,r)$ is a compact neighborhood of x, so that X is locally compact. If $\{x_n\}$ is a Cauchy sequence in X then there is $N \in \mathbb{N}$ such that $n,m \ge N$ implies $d(x_n,x_m) < 1$. Thus $\{x_n\}_{n\ge N}$ lives in the compact, thus complete, subspace $\overline{B}(x_N,1)$. Thus the Cauchy sequence converges. The subsets $K_n = \overline{B}(x,n)$ are compact and satisfy $K_n \subseteq K_{n+1}\ \forall n$ and $\bigcup_n K_n = X$. Thus τ_d is hemicompact. By Exercise 7.8.43 this quite trivially implies σ-compactness and the Lindelöf property. For metric spaces, the latter is equivalent to separability and second countability. ■

Theorem 7.8.79 *For a metric space (X,d) the following are equivalent:*

(i) (X,d) is properly metrizable.

(ii) There is a proper function $g : X \to \mathbb{R}$.

(iii) There are open sets $\{U_i\}_{i\in\mathbb{N}}$ such that each $\overline{U_i}$ is compact, $\overline{U_i} \subseteq U_{i+1}\ \forall i$, and $\bigcup_{i\in\mathbb{N}} U_i = X$.

(iv) (X,τ_d) is locally compact and any of the following: second countable, separable, Lindelöf, σ-compact, hemicompact.

Proof. (i)⇒(ii) By Exercise 7.8.75, $g(x) = d'(x_0, x)$ does the job if d' is a proper metric equivalent to d.

(ii)⇒(i) Let $g : X \to \mathbb{R}$ be proper. We may assume that $g \geq 0$. Define $d'(x, y) = d(x, y) + |g(x) - g(y)|$. This is a metric, and in view of $d \leq d'$ we have $d'(x_i, y) \to 0 \Rightarrow d(x_i, y) \to 0$. And if $d(x_i, y) \to 0$ then $d'(x_i, y) \to 0$ by continuity of g. Thus $d' \simeq d$. Now pick $x_0 \in X$ and $r > 0$. If $y \in \overline{B}^{d'}(x_0, r)$ then $d'(x_0, r) \leq r$, thus $|g(x_0) - g(y)| \leq r$, implying $\overline{B}^{d'}(x_0, r) \subseteq g^{-1}([0, g(x_0) + r])$. Since g is proper, $g^{-1}([0, f(x_0) + r])$ is compact, so that $\overline{B}^{d'}(x_0, r)$ is compact. Thus d' is proper.

(ii)⇒(iii) Defining $U_n = g^{-1}((-n, n))$, we have $\overline{U_n} \subseteq g^{-1}([-n, n])$, which is compact. The rest is obvious.

(iii)⇒(ii) We may assume $U_1 \neq \emptyset$. The sets $\overline{U_n}$ and $X \backslash U_{n+1}$ are closed and disjoint, thus for each $n \in \mathbb{N}$

$$f_n(x) = n + \frac{\operatorname{dist}(x, \overline{U_n})}{\operatorname{dist}(x, \overline{U_n}) + \operatorname{dist}(x, X \backslash U_{n+1})}$$

defines an element of $C(X, [n, n+1])$ satisfying $f_n \upharpoonright \overline{U_n} = n$ and $f_n \upharpoonright (X \backslash U_{n+1}) = n + 1$. (The denominator vanishes only on $\overline{U_n} \cap X \backslash U_{n+1} = \emptyset$.) Define $C_0 = \overline{U_1}$ and $C_n = \overline{U_{n+1}} \backslash U_n$ for $n \in \mathbb{N}$. Then C_n is closed for each $n \in \mathbb{N}_0$ and $\bigcup_{n=0}^{\infty} C_n = X$. Furthermore, we have

$$C_n \cap C_{n+1} = (\overline{U_{n+1}} \backslash U_n) \cap (\overline{U_{n+2}} \backslash U_{n+1}) = \overline{U_{n+1}} \backslash U_{n+1} = \partial U_{n+1}$$

and $C_n \cap C_m = \emptyset$ whenever $|n - m| \geq 2$. Define $g_0 \equiv 1$ on $C_0 = \overline{U_1}$. Let $g_n = f_n \upharpoonright C_n$ for $n \in \mathbb{N}$. Then clearly $g_n \in C(C_n, [0, \infty))$. For $x \in C_n \cap C_{n+1} = \partial U_{n+1} = \overline{U_{n+1}} \backslash U_{n+1}$ we have $g_n(x) = n+1 = g_{n+1}(x)$, thus g_n and g_{n+1} coincide on the intersection of their domains. This means that defining $g(x)$ to be $g_n(x)$ for any n such that $x \in C_n$, we obtain a continuous function $g : X \to \mathbb{R}$. (Cf. Corollary 6.2.6.) For $r > 0$ we have $g^{-1}([0, r]) \subseteq g^{-1}([0, \lceil r \rceil]) = \overline{U_{\lceil r \rceil}}$, which is compact by assumption. Thus $g : X \to [0, \infty)$ is continuous and proper.

(iii)⇔(iv) Property (iii) coincides with (α) in Exercise 7.8.43, and the claim follows from the implications proven there, keeping in mind that for metric spaces we have the result of Exercise 7.1.9. ■

Remark 7.8.80 1. For another result concerning proper metric spaces, cf. Theorem 12.4.27.

2. We will later prove that every second countable locally compact Hausdorff space is metrizable, cf. Corollary 8.2.40. Combining this with the above result, we arrive at the following very satisfactory statement (in which we cannot replace second countability by Lindelöf, since there are locally compact Hausdorff spaces that are Lindelöf but not second countable, e.g. $[0, 1]^S$ with S uncountable): □

Corollary 7.8.81 *A topological space (X, τ) is properly metrizable (i.e. admits a proper metric d such that $\tau = \tau_d$) if and only if it is locally compact Hausdorff and second countable.*

7.8.9 $C_0(X, \mathbb{F})$

Definition 7.8.82 *Let X be locally compact Hausdorff. Then*

$$C_0(X, \mathbb{C}) = \{f \in C(X, \mathbb{C}) \mid f \text{ tends to zero at infinity}\}.$$

It is clear that f tends to zero at infinity (in the sense of Definition 7.8.49) if and only if for every $\varepsilon > 0$ there is a compact $K \subseteq X$ such that $x \in X \backslash K \Rightarrow |f(x)| < \varepsilon$. (This is the more common definition.)

Exercise 7.8.83 Let X be locally compact Hausdorff. Writing $C_0(X) \equiv C_0(X, \mathbb{R})$, prove:

(i) The following are equivalent: (α) X is compact. (β) $C(X) = C_0(X)$. (γ) The non-zero constant functions are contained in $C_0(X)$.

(ii) If $g \in C(X_\infty)$ and $g(\infty) = 0$ then $g \restriction X \in C_0(X)$.

(iii) If $f : X \to \mathbb{R}$ is any function and $\widehat{f} : X_\infty \to \mathbb{R}$ is defined by $\widehat{f} \restriction X = f$ and $\widehat{f}(\infty) = 0$ then $\widehat{f} \in C(X_\infty)$ if and only if $f \in C_0(X)$. (Use Lemma 7.8.46.)

(iv) The maps

$$\begin{array}{rrcllcl} \alpha: & C_0(X) \oplus \mathbb{R} & \to & C(X_\infty), & (f,c) & \mapsto & \widehat{f} + c\mathbf{1} \\ \beta: & C(X_\infty) & \to & C_0(X) \oplus \mathbb{R}, & g & \mapsto & (g - g(\infty)\mathbf{1}, g(\infty)) \end{array}$$

are mutually inverse homomorphisms of $\mathbb{R}$-algebras.

Remark 7.8.84 Given an algebra A over some field k, define $A_\mathbf{1} = A \oplus k$. Turn $A_\mathbf{1}$ into a vector space by coordinatewise addition and multiplication by elements of k. Then $(a, \lambda)(b, \mu) = (ab, \lambda b + \mu a, \lambda\mu)$ is an associative multiplication on $A_\mathbf{1}$ that distributes over $+$. Thus $A_\mathbf{1}$ is a K-algebra, and it has $(0, 1)$ as unit. $A_\mathbf{1}$ is called the (minimal) unitization of A. Now the above exercise has the following interpretation:

When X is locally compact Hausdorff, $C(X_\infty)$ is (isomorphic to) the unitization of $C_0(X)$.

Note that unitization $A \mapsto A_\mathbf{1}$ is a functor: If $\delta : A \to B$ is an algebra homomorphism then $\delta_\mathbf{1} : (a, \lambda) \mapsto (\delta(a), \lambda)$ is a unital algebra homomorphism extending δ. $\square$

Exercise 7.8.85 (i) Let X, Y be locally compact Hausdorff and $f : X \to Y$ continuous and proper. Prove that $g \mapsto g \circ f$ defines a map $C_0(Y, \mathbb{C}) \to C_0(X, \mathbb{C})$ that is linear and an algebra homomorphism.

(ii) Prove that the assignments $F(X) = C_0(X)$ and $F(f) : g \mapsto g \circ f$ define a contravariant functor $F : \mathcal{TOP}^p_{lcH} \to \mathrm{CAlg}_\mathbb{C}$, where $\mathrm{CAlg}_\mathbb{C}$ is the category of commutative algebras over $\mathbb{C}$ and algebra homomorphisms.

(iii) If $U : A \to A_\mathbf{1}$ is the unitization functor from Remark 7.8.84, prove that the diagram

$$\begin{tikzcd} \mathcal{TOP}^p_{lcH} \arrow[r, "U"] \arrow[dr, "F"'] & \mathcal{TOP}_{cH} \arrow[d, "F"] \\ & \mathrm{CommAlg}_\mathbb{C} \end{tikzcd}$$

of functors commutes.

7.9 Compact-open topology. Compactly generated spaces

7.9.1 The compact-open topology

If X, Y are topological spaces, one would like to put a reasonable topology on the set $C(X,Y)$ of continuous functions. One possibility is the subspace topology inherited from $\mathrm{Fun}(X,Y)$, interpreted as product space $\prod_{x\in X} Y_x$, where $Y_x = Y$, or short Y^X. The topology on $\prod_x Y$ is the initial topology, for which a subbase is given by (6.6). In the present situation, where $f \in Y^X$, we write evaluation at $x \in X$ instead of the projection maps p_x. Thus the subbase can be written as $\mathcal{S} = \{F(x,U)\}$, where $x \in X,\ U \subseteq Y$ is open and $F(x,U) = \{f : X \to Y \mid f(x) \in U\}$. In view of the form of the subbase elements $F(x,U)$, this topology is called the point-open topology. We know from our discussion of the product topology that a net $f_\iota \in Y^X$ converges in this topology if and only if $f_\iota(x) \in Y$ converges for each $x \in X$. Thus the point-open topology is the topology of *pointwise* convergence.

However, pointwise convergence is not the best notion of convergence for (continuous) functions. In Section 2.1 (and Remark 5.2.12) we have seen that every metric on Y gives rise to a metric D on $C_b(X,Y)$. The topology defined by D is the topology of *uniform* convergence.

The aim of this section is to define a topology on $C(X,Y)$ that is closer to the uniform than to the pointwise topology, without assuming metrizability of Y.

Lemma 7.9.1 *Let $(X,\tau), (Y,\sigma)$ be topological spaces.*

(i) The sets

$$F(K,U) = \{f \in C(X,Y) \mid f(K) \subseteq U\} \quad \text{with } K \subseteq X \text{ compact, } U \subseteq Y \text{ open,}$$

form a subbase for a topology on $C(X,Y)$, the compact-open topology[13] τ_{co}.

(ii) If (Y,σ) is Hausdorff, so is τ_{co}.

Proof. 1. If $f \in C(X,Y)$, take $K = \{x\}$ with $x \in X$ and $U = Y$. Then $f \in F(K,U)$. This proves that the union of the $F(K,U)$ equals $C(X,Y)$, so that $\mathcal{S} = \{F(K,U)\}$ is the subbase of a topology τ_{co}.

2. If Y is Hausdorff, the point-open topology on $\mathrm{Fun}(X,Y)$ is Hausdorff since it equals the product topology on Y^X. It is obvious that the compact-open topology is finer than the point-open topology, thus it is Hausdorff. ■

The next result shows that, in certain cases, the compact-open topology is the topology of uniform convergence:

Proposition 7.9.2 *If X is compact and Y is metric, then the compact-open topology on $C(X,Y)$ coincides with the topology arising from the metric D in (2.6).*

Proof. Let $K \subseteq X$ be compact and $U \subseteq Y$ open. Let $f \in F(K,U) \subseteq C(X,Y)$. Then $L = f(K) \subseteq U \subseteq Y$ is compact. The function $L \to [0,\infty), y \mapsto \mathrm{dist}(y, Y\backslash U)$ is continuous, thus assumes its infimum due to compactness of L. This infimum is non-zero since L and $Y\backslash U$ are disjoint. Thus there is a $\varepsilon > 0$ such that $L_\varepsilon \subseteq U$. If now $g \in B^D(f,\varepsilon)$ then $g(K) \subseteq L_\varepsilon \subseteq U$, thus $B^D(f,\varepsilon) \subseteq F(K,U)$. This proves $F(K,U) \in \tau_D$ and therefore $\tau_{\mathrm{co}} \subseteq \tau_D$.

In order to prove $\tau_D \subseteq \tau_{\mathrm{co}}$, it suffices to show $B^D(f,\varepsilon) \in \tau_{\mathrm{co}}$ for all f, ε. Since X is compact, so is $f(X) \subseteq Y$. Thus there are $x_1, \ldots, x_n$ such that $f(X) \subseteq \bigcup_i B(f(x_i), \varepsilon/3)$. The sets $K_i = f^{-1}(B(f(x_i),\varepsilon/4)) \subseteq X$ are closed, thus compact. Putting $U_i = B(f(x_i),\varepsilon/2) \subseteq Y$, we have

[13] The compact-open topology was introduced in 1945 by Ralph Fox (1913-1973), American topologist.

$f(K_i) = f(\overline{f^{-1}(B(f(x_i),\varepsilon/3))}) \subseteq \overline{f(f^{-1}(B(f(x_i),\varepsilon/3)))} \subseteq \overline{B(f(x_i),\varepsilon/3)} \subseteq U_i \ \forall i$. Thus clearly $f \in \bigcap_i F(K_i,U_i)$. If now $g \in \bigcap_i F(K_i,U_i)$ then $g(K_i) \subseteq U_i \ \forall i$. For every $x \in X$ we have $x \in K_i$ for some i, thus $\{f(x), g(x)\} \subseteq U_i$, implying $d(f(x),g(x)) < 2\varepsilon/2 = \varepsilon$. Thus $D(f,g) < \varepsilon$, so that $\bigcap_i F(K_i,U_i) \subseteq B^D(f,\varepsilon)$. This proves $B^D(f,\varepsilon) \in \tau_{\rm co}$, and thus $\tau_D \subseteq \tau_{\rm co}$. ■

Corollary 7.9.3 *If X is a compact space, Y a set and d_1, d_2 equivalent metrics on Y then the metrics $D_i(f,g) = \sup_x d_i(f(x),g(x))$, $i = 1,2$ on $C(X,Y)$ are equivalent.*

Proof. By definition, d_1 and d_2 induce the same topology on Y and therefore the same compact-open topology on $C(X,Y)$. According to the proposition, both metrics D_1, D_2 induce the compact open topology on $C(X,Y)$, thus $D_1 \simeq D_2$. ■

Remark 7.9.4 1. The proposition can easily be generalized: Whenever Y is metric, the compact-open topology on $C(X,Y)$ is the topology of uniform convergence on all compact subsets $K \subseteq X$. Thus a net $\{f_\iota\}$ in $C(X,Y)$ converges to f w.r.t. $\tau_{\rm co}$ if and only if it satisfies $\sup_{x\in K} d(f_\iota(x), f(x)) \to 0$ for every compact $K \subseteq X$.

2. It is natural to ask when the compact-open topology on $C(X,Y)$ is metrizable. Since the map $Y \to C(X,Y)$ that sends y to the constant function y is an embedding, metrizability of $(C(X,Y), \tau_{\rm co})$ implies that of Y. Thus metrizability of Y is necessary. Assume that X is hemicompact, and let $\{K_n\}_{n\in K}$ be a sequence of compact sets as required. Then

$$D(f,g) = \sum_{k=1}^{\infty} 2^{-k} \min(1, \sup_{x\in K_k} d(f(x),g(x)))$$

is a metric on $C(X,Y)$. A sequence $\{f_n \in C(X,Y)\}$ of functions converges w.r.t. D if and only if it converges uniformly on all the compact sets K_k. Since every compact K is contained in some K_n, this is equivalent to convergence on all compact subsets. By the first half of this remark, this is equivalent to convergence w.r.t. $\tau_{\rm co}$. Thus D is a metric for $\tau_{\rm co}$.

On the other hand, on its own metrizability of $C(X,Y)$ does not imply much: If X is such that every $f \in C(X,\mathbb{R})$ is constant, then every $f \in C(X,Y)$ with Y metric is constant, thus $C(X,Y) \cong Y$. But if $C(X,\mathbb{R})$ separates the points of X (X is 'completely Hausdorff') and the compact-open topology on it is first countable then X is hemicompact. Cf. [7, Theorem 8]. □

7.9.2 The exponential law

In Exercises 6.5.40 and 7.7.45, we considered the map $\Lambda : \mathrm{Fun}(X \times Y, Z) \to \mathrm{Fun}(X, \mathrm{Fun}(Y,Z))$ defined for $f : X \times Y \to Z$ by $\Lambda(f)(x) = f(x,\cdot)$, thus $\Lambda(f)(x)(y) = f(x,y)$. Writing Y^X instead of $\mathrm{Fun}(X,Y)$, we have $\Lambda : Z^{X\times Y} \to (Z^Y)^X$. For this reason, Λ is often called the exponential map. We now turn to the study of the topological behavior of this map, which is perhaps the most satisfactory aspect of the compact-open topology. Rather than imposing restrictive assumptions from the outset, we require stronger assumptions only when we need them, in order to make clear how they enter the proofs. In what follows all function spaces have the compact-open topologies.

Lemma 7.9.5 *Let X, Y, Z be arbitrary topological spaces and $f \in C(X \times Y, Z)$. Then $\Lambda(f) \in C(X, C(Y,Z))$.*

Proof. Since joint continuity implies separate continuity, we have $\Lambda(f)(x) = f(x,\cdot) \in C(Y,Z)$ for each $x \in X$. It remains to prove that $x \mapsto \Lambda(f)(x) = f(x,\cdot)$ is continuous. It is enough to show that $\Lambda^{-1}(F(K,U)) \subseteq X$ is open for $K \subseteq Y$ compact and $U \subseteq Z$ open. But

$$\Lambda^{-1}(F(K,U)) = \{x \in X \mid f(\{x\} \times K) \subseteq U\} = \{x \in X \mid \{x\} \times K \subseteq f^{-1}(U)\}.$$

Since $f : X \times Y \to Z$ is continuous, $f^{-1}(U)$ is open. Thus if $\{x\} \times K \subseteq f^{-1}(U)$ then Lemma 7.5.1 gives an open $W \subseteq X$ such that $x \in W$ and $W \times K \subseteq f^{-1}(U)$. But this means $x \in W \subseteq \Lambda^{-1}(F(K,U))$. Thus $\Lambda^{-1}(F(K,U))$ is open, and we are done. ■

Proposition 7.9.6 *If Y is locally compact Hausdorff then for all spaces X, Z we have:*

(i) The composition map $\Sigma : C(X,Y) \times C(Y,Z) \to C(X,Z),\ (f,g) \mapsto g \circ f$ is continuous.

(ii) The evaluation map $e : C(Y,Z) \times Y \to Z,\ (f,y) \mapsto f(y)$ is continuous.

(iii) For every $g \in C(X, C(Y,Z))$ we have $\Lambda^{-1}(g) \in C(X \times Y, Z)$. (Thus Λ is a bijection $C(X \times Y, Z) \to C(X, C(Y,Z))$.)

Proof. (i) It suffices to show that $\Sigma^{-1}(F(K,U))$ is open for each element $F(K,U)$ of the subbase for the compact-open topology on $C(X,Z)$. Thus let $K \subseteq X$ be compact and $U \subseteq Z$ open. If $(f,g) \in \Sigma^{-1}(F(K,U))$ then $g \circ f(K) \subseteq U$. This is equivalent to $f(K) \subseteq g^{-1}(U)$. Now $f(K) \subseteq Y$ is compact and $g^{-1}(U) \subseteq Y$ is open, thus since Y is locally compact Hausdorff, Proposition 7.8.25, provides an open $W \subseteq Y$ with compact closure $\overline{W}$ such that $f(K) \subseteq W \subseteq \overline{W} \subseteq g^{-1}(U)$. Thus $f \in F(K,W)$, $g \in F(\overline{W},U)$. Furthermore, $f' \in F(K,W)$, $g' \in F(\overline{W},U)$ implies $g' \circ f' = \Sigma(f',g') \in F(K,U)$. Thus $F(K,W) \times F(\overline{W},U) \subseteq C(X,Y) \times C(Y,Z)$ is an open product neighborhood of (f,g), proving that Σ is continuous.

(ii) This can be proven directly with a slightly simpler proof than the one for (i). Instead, we deduce the result from (i). For $X = \{x_0\}$ and any space T, the obvious bijection $T \to C(X,T)$ is a homeomorphism w.r.t. the compact-open on $C(X,T)$. By (i), $\Sigma : C(X,Y) \times C(Y,Z) \to C(X,Z)$ is continuous, and in view of the homeomorphisms $C(X,Y) \cong Y,\ C(X,Z) \cong Z$ this means the map $Y \times C(Y,Z) \to Z,\ (y,f) \mapsto f(y)$ is continuous. This is just the evaluation map.

(iii) Let $g \in C(X, C(Y,Z))$. Then $g \times \mathrm{id}_Y \in C(X \times Y, Y \times C(Y,Z))$. By assumption, the evaluation map $Y \times C(Y,Z) \to Z$ is continuous for every space Z. Thus $e \circ (g \times \mathrm{id}_Y) : X \times Y \to Z$ is continuous. Now the computation

$$[e \circ (g \times \mathrm{id}_Y)](x,y) = e(g(x),y) = g(x)(y) = [\Lambda^{-1}(g)](x,y)$$

shows that $\Lambda^{-1}(g) = e \circ (g \times \mathrm{id}_Y)$, which is continuous, as claimed. ■

Remark 7.9.7 1. The proof uses that Y is strongly locally compact. Thus if one assumes this, one does not need the Hausdorff property.

2. The local compactness is almost necessary: One can show that if X is completely regular and $C(X,\mathbb{R})$ admits a topology making the evaluation map $e : C(X,\mathbb{R}) \times X \to \mathbb{R}$ continuous then X is locally compact. Cf. [75, Exercise 3.4.A]. □

Lemma 7.9.8 *Let X be Hausdorff, Y a topological space and $\mathcal{S}$ a subbase for the topology of Y. Then the sets $F(K,V)$, where $K \subseteq X$ is compact and $V \in \mathcal{S}$, form a subbase for the compact-open topology on $C(X,Y)$.*

Proof. Let $K \subseteq X$ be compact, $U \subseteq Y$ open and $f \in F(K,U)$. Then $U = \bigcup_{i \in I} U_i$, where the U_i are in $\mathcal{B}$, the base consisting of finite intersections of elements of $\mathcal{S}$. By definition we have $f(K) \subseteq \bigcup_i U_i$, thus $\{U_i\}_{i \in I}$ is an open cover of the compact set $f(K)$. Thus there is subcover by finitely many U_i's, and we call these $U_1, \ldots, U_n$. Now $\{K \cap f^{-1}(U_i)\}_{i=1,\ldots,n}$ is a finite open cover of the compact Hausdorff, thus normal, space K. Thus by Exercise 8.1.48 or Lemma 8.1.51 there are closed (thus compact) subsets $K_i \subseteq K$ such that $K = \bigcup_{i=1}^n K_i$ and $K_i \subseteq f^{-1}(U_i)\ \forall i$. Thus $f(K_i) \subseteq U_i\ \forall i$, to wit $f \in F(K_i, U_i)\ \forall i$. Since $U_i \subseteq U\ \forall i$, we have $f \in \bigcap_{i=1}^n F(K_i, U_i) \subseteq F(K,U)$. Picking $V_{ij} \in \mathcal{S}$ such that $U_i = \bigcap_{j=1}^{n_i} V_{ij}$, we have

$$f \in \bigcap_{i=1}^n F(K_i, U_i) = \bigcap_{i=1}^n F(K_i, \bigcap_{j=1}^{n_i} V_{ij}) = \bigcap_{i=1}^n \bigcap_{j=1}^{n_i} F(K_i, V_{ij}) \subseteq F(K,U).$$

This proves that the $F(K,V)$ with $K \subseteq X$ compact and $V \in \mathcal{S}$ form a subbase for $C(X,Y)$. ■

Remark 7.9.9 In the proof of Lemma 7.9.8 we used the normality of compact Hausdorff spaces, alluded to in Remark 7.4.6, and a result proven only in Section 8.1.5. But the proof of the latter uses nothing but the equivalent characterizations of normality given in Lemma 8.1.4. □

Proposition 7.9.10 *Let X, Y be Hausdorff spaces and Z arbitrary. Then $\Lambda : C(X \times Y, Z) \to C(X, C(Y,Z))$ is an embedding.*

Proof. We already know that $\Lambda : \mathrm{Fun}(X \times Y, Z) \to \mathrm{Fun}(X, \mathrm{Fun}(Y,Z))$ is a bijection, and in view of Lemma 7.9.5, Λ restricts to an injective map $C(X \times Y, Z) \to C(X, C(Y,Z))$ that we again denote Λ. We prove continuity of Λ. If $K \subseteq X$, $K' \subseteq Y$ are compact, Proposition 7.9.6(iii) gives that $\Lambda' : C(K \times K', Z) \to C(K, C(K', Z))$ is a bijection. (Here we need that Y is T_2.) In particular

$$F(K \times K', U) = \Lambda^{-1}(F(K, F(K', U))) \tag{7.10}$$

for every open $U \subseteq Z$. Since $K \times K'$ is compact, we see that $\Lambda^{-1}(F(K, F(K', U))) = F(K \times K', U) \subseteq C(X \times Y, Z)$ is open for each $F(K, F(K', U))$. By Lemma 7.9.8 (for which we need X to be T_2) the latter form a subbase for $C(X, C(Y,Z))$ since the $F(K', U)$ are a subbase for $C(Y,Z)$. Thus Λ is continuous.

It remains to show that Λ is open as a function $C(X \times Y, Z) \to \Lambda(C(X \times Y, Z)) \subseteq C(X, C(Y,Z))$. From (7.10) we obtain

$$\Lambda(F(K \times K', U)) = \Lambda(C(X \times Y, Z)) \cap F(K, F(K', U)).$$

Thus openness of $\Lambda : C(X \times Y, Z) \to \Lambda(C(X \times Y, Z))$ follows as soon as we show that the sets $F(K \times K', U)$, where $K \subseteq X, K' \subseteq Y$ are compact and $U \subseteq Z$ is open, form a subbase for the compact-open topology on $C(X \times Y, Z)$.

Let $K \subseteq X \times Y$ be compact, $W \subseteq Z$ open and $f \in F(K, W)$. Now $K' = p_1(K) \subseteq X$, $K'' = p_2(K) \subseteq Y$ are compact Hausdorff. We have $K \subseteq f^{-1}(W)$, thus by definition of the product topology for every $(x,y) \in K$ there are open $U_{xy} \subseteq X, V_{xy} \subseteq Y$ such that $(x,y) \in U_{xy} \times V_{xy} \subseteq f^{-1}(W)$. Now $(x,y) \in K$ implies $x \in K'$, and $U_{xy} \cap K'$ is an open neighborhood in K' of x. By Lemma 7.4.19(i) there is an open $U'_{xy} \subseteq U_{xy}$ such that $x \in U'_{xy} \cap K' \subseteq \mathrm{cl}_{K'}(U'_{xy} \cap K') \subseteq U_{xy} \cap K'$. Similarly there is an open $V'_{xy} \subseteq V_{xy}$ such that $y \in V'_{xy} \cap K'' \subseteq \mathrm{cl}_{K''}(V'_{xy} \cap K'') \subseteq V_{xy} \cap K''$. Now $\{U'_{xy} \times V'_{xy}\}_{(x,y) \in K}$ is an open cover of K, so by compactness there is a finite subcover $\{U'_{x_i y_i} \times V'_{x_i y_i}\}_{i=1}^n$, which we abbreviate $U'_1 \times V'_1, \ldots, U'_n \times V'_n$. For $i = 1, \ldots, n$ write $K'_i = K' \cap \overline{U'_i}$, $K'_i = K'' \cap \overline{V'_i}$, which are compact subsets of X and Y, respectively. In view of $K \subseteq K' \times K''$ we have $K \subseteq \bigcup_i K'_i \times K''_i$, thus

$\bigcap_i F(K_i' \times K_i'', W) \subseteq F(K, W)$. On the other hand $K_i' = K' \cap \overline{U_i'} = \mathrm{cl}_{K'}(K' \cap U_i') \subseteq U_{x_i y_i}$ and similarly $K_i'' \subseteq V_{x_i y_i}$. Thus $K_i' \times K_i'' \subseteq U_{x_i y_i} \times V_{x_i y_i} \subseteq f^{-1}(W)\ \forall i$, which restates as $f \in \bigcap_{i=1}^n F(K_i' \times K_i'', W)$. This proves that the sets $F(K' \times K'', W)$ with $K' \subseteq X, K'' \subseteq Y$ compact form a subbase for $C(X \times Y, Z)$. ■

Theorem 7.9.11 *Let X be Hausdorff, Y locally compact Hausdorff and Z arbitrary. Then $\Lambda : C(X \times Y, Z) \to C(X, C(Y, Z))$ is a homeomorphism.*

Proof. This is immediate from Proposition 7.9.6(iii) and Proposition 7.9.10. ■

7.9.3 ⋆⋆ Compactly generated spaces = k-spaces

We briefly look at yet another generalization of compactness, with the aim of extending the range of the exponential law:

Definition 7.9.12 *A space X is called compactly generated or k-space if closedness of $A \cap K \subseteq X$ for every compact $K \subseteq X$ implies closedness of $A \subseteq X$.*

(The converse is automatic if X is Hausdorff: If $A \subseteq X$ is closed then so is $A \cap K$ for every compact K.) We will almost exclusively restrict to Hausdorff k-spaces.

Lemma 7.9.13 *Every first countable space is a k-space.*

Proof. Assume $A \subseteq X$ is not closed, pick $y \in \overline{A} \backslash A$ and a sequence $\{x_n\} \subseteq A$ such that $x_n \to y$. Now $K = \{y, x_1, x_2, \ldots\}$ is compact by Exercise 7.3.8, but $A \cap K = \{x_1, x_2, \ldots\} = A$ is non-closed. ■

(Actually every sequential space is a k-space since all one needs is the property in Corollary 5.1.8.)

Proposition 7.9.14 *A Hausdorff space X is a k-space if and only if it is (homeomorphic to) a quotient space of a locally compact Hausdorff space.*

Proof. ⇒ Let $\mathcal{K}$ be the family of all compact subsets $K \subseteq X$. Then the direct sum $Z = \bigoplus_{K \in \mathcal{K}} K$ is locally compact Hausdorff. For each $K \in \mathcal{K}$ we have the inclusion map $\iota_K : K \hookrightarrow X$. Putting them together using Proposition 6.3.5, we have a continuous map $p : Z \to X$. The latter clearly is surjective since $\{x\}$ is compact for every $x \in X$, thus contained in $\mathcal{K}$. In order to prove that p is a quotient map we must show that $p^{-1}(C) \subseteq Z$ is closed only if $C \subseteq X$ is closed. Since p is the aggregate of the inclusions ι_K, closedness of $p^{-1}(C)$ is equivalent to closedness of $K \cap C$ for each $K \in \mathcal{K}$. But by the k-space property, this is equivalent to closedness of C.

⇐ Let $p : Z \to X$ be a quotient map where Z is locally compact. Let $A \subseteq X$ be such that $A \cap K \subseteq K$ is closed for each compact $K \subseteq X$. If $z \in p^{-1}(A)$, local compactness of Z provides an open $U \ni z$ with compact $\overline{U}$. Then $p(\overline{U})$ is compact, thus $A \cap p(\overline{U})$ is closed, and so is $p^{-1}(A \cap f(\overline{U}))$ by continuity. In view of $p^{-1}(A) \cap \overline{U} \subseteq p^{-1}(A \cap p(\overline{U})) \subseteq p^{-1}(A)$ we have $z \in p^{-1}(A)$. Thus $p^{-1}(A) \subseteq Z$ is closed. Since p is a quotient map, this implies that $A \subseteq X$ is closed. ■

Corollary 7.9.15 *(i) Every locally compact Hausdorff space is a k-space.*

(ii) Every Hausdorff quotient of a k-space is a k-space.

The next two results are entirely analogous to corresponding facts for locally compact spaces:

Exercise 7.9.16 Prove:

(i) Every closed subspace of a k-space is a k-space.

(ii) Every open subspace of a Hausdorff k-space is a k-space.

Exercise 7.9.17 Prove that $X = \bigoplus_{i \in I} X_i$ is a k-space if and only if each X_i is a k-space.

The product of two k-spaces can fail to be a k-space! But:

Proposition 7.9.18 *The product of a k-space with a locally compact space is a k-space.*

Proof. Cf. e.g. [75, Theorem 3.3.27]. ■

Lemma 7.9.19 *Let X be a k-space and Y a topological space. Then $f : X \to Y$ is continuous if and only if $f \restriction K \to Y$ is continuous for every compact $K \subseteq X$.*

Proof. If f is continuous then also all its restrictions are continuous. If all $f \restriction K$ are continuous and $C \subseteq Y$ is closed then $(f \restriction K)^{-1}(C) = f^{-1}(C) \cap K \subseteq K$ is closed for each compact $K \subseteq X$. But then the k-space property implies that $f^{-1}(C)$ is closed, thus f is continuous. ■

Exercise 7.9.20 Let X be a topological space, Y a k-space and $f : X \to Y$ a function. Prove that f is open / closed / quotient if and only if the restriction $f^{-1}(Z) \to Z$ is open / closed / quotient for every compact $Z \subseteq Y$.

Theorem 7.9.21 *If $X \times Y$ is a Hausdorff k-space then for every topological space Z the exponential map $\Lambda : C(X \times Y, Z) \to C(X, C(Y, Z))$ is a homeomorphism.*

Proof. In view of Proposition 7.9.10 all we need to prove is $\Lambda^{-1}(g) \in C(X \times Y, Z)$ for every $g \in C(X, C(Y, Z))$. Since $X \times Y$ is a k-space, by Lemma 7.9.19 it suffices to show that $\Lambda^{-1}(g) \restriction K$ is continuous for every compact $K \subseteq X \times Y$. If such a K is given, $Y' = p_2(K) \subseteq Y$ is compact. Now Theorem 7.9.11 gives that $\Lambda' : C(X \times Y', Z) \to C(X, C(Y', Z))$ is a homeomorphism, thus in particular surjective. Defining $g' \in C(X, C(Y', Z))$ by $g'(x)(y') = g(x)(y')$ for $x \in X, y' \in Y'$, we have $\Lambda'^{-1}(g') = \Lambda^{-1}(g) \restriction X \times Y'$, thus $\Lambda^{-1}(g) \restriction X \times Y'$ – and a fortiori $\Lambda^{-1}(g) \restriction K$ – is continuous. ■

Remark 7.9.22 1. Theorem 7.9.21 clearly applies to many cases not covered by Theorem 7.9.11, e.g. when X, Y both are first countable Hausdorff spaces. But it does not contain Theorem 7.9.11 since the assumptions of the latter do not imply that $X \times Y$ is a k-space.

2. Because of their nice properties, as in the preceding theorem, k-spaces have been called a 'convenient category' for the purposes of algebraic topology. For more information cf. e.g. [96].

3. Given a Hausdorff space (X, τ), let $\mathcal{C}$ be the family of those $C \subseteq X$ for which $C \cap K$ is closed for every compact $K \subseteq X$. One easily checks that $\mathcal{C}$ satisfies the statements in Lemma 2.4.2 and thus is the family of closed sets for some topology τ' on X. The topology τ' is finer than τ, but has the same compact subspaces. Using Lemma 7.9.19 one proves that every continuous $f : X \to Y$ is also continuous w.r.t. the k-spaces topologies on X, Y. Thus we have a functor from Hausdorff spaces to k-spaces. This functor can be used to modify certain construction so that one stays in the category of k-spaces, for example via $X \times_k Y = k(X \times Y)$. □

We close by stating without proof a nice version of the Ascoli-Arzelà theorem for $(C(X, Y), \tau_{\text{co}})$. We need to modify the notion of equicontinuity:

Definition 7.9.23 *Let X, Y be topological spaces. A family $\mathcal{F} \subseteq \mathrm{Fun}(X, Y)$ is evenly continuous if for every $x \in X$, every $y \in Y$ and every open neighborhood V of y there are open neighborhoods $U \ni x$ and $W \ni y$ such that $f(U) \subseteq V$ for every $f \in \mathcal{F}$ that satisfies $f(x) \in W$.*

It is clear that every element of an evenly continuous family is continuous.

Theorem 7.9.24 *Let X be a k-space and Y a T_3-space (Remark 7.4.6). Let $\mathcal{F} \subseteq C(X, Y)$ be closed w.r.t. τ_{co}. Then $\mathcal{F}$ is compact if and only if $\mathcal{F}$ is evenly continuous and $\overline{\{f(x) \mid f \in \mathcal{F}\}}$ is compact for every $x \in X$.*

For the proof cf. [75, Theorem 3.4.20]. For much more on k-spaces see [96].

Chapter 8

Stronger separation axioms and their uses

8.1 T_3- and T_4-spaces

8.1.1 Basics

In a Hausdorff space, we can separate any two points by disjoint open neighborhoods. For many purposes, this is not sufficient. We actually have already met stronger separation properties in Lemma 7.4.1, Exercise 7.4.5 and Proposition 7.8.25(ii), but they involved compactness requirements. Now we will look at such stronger separation properties in a more systematic way. We begin with two definitions:

Definition 8.1.1 *A T_1-space (X, τ) is called regular or T_3-space if given any closed $C \subseteq X$ and $x \in X \backslash C$ there are $U, V \in \tau$ such that $x \in U$, $C \subseteq V$ and $U \cap V = \emptyset$.*

Definition 8.1.2 *A T_1-space (X, τ) is called normal or T_4-space if given disjoint closed sets $C, D \subseteq X$ there are $U, V \in \tau$ such that $C \subseteq U$, $D \subseteq V$ and $U \cap V = \emptyset$.*

Remark 8.1.3 1. One says: The 'open sets separate points from closed sets' and 'the open sets separate the closed sets'.

2. Many authors distinguish between T_3- and regular spaces (and similarly between T_4- and normal spaces) by defining one of the two terms as including T_1 but not the other one. Some, e.g. [237], define T_3=regular+T_1. Others write regular=$T_3 + T_1$, but this seems rather unreasonable since then $T_3 \not\Rightarrow T_2$. (What is the point of numbering axioms if the natural ordering has no consequences?) Since we will have little occasion to meet non-T_1-spaces in which closed sets can be separated from points or closed sets, we follow those (like [75, 77, 119, 28]) who use 'regular' and T_3 synonymously (including T_1!) and similarly for T_4.

3. Since we assume all T_i-axioms to include T_1, singletons are closed and therefore

$$T_4 \Rightarrow T_3 \Rightarrow T_2 \Rightarrow T_1 \Rightarrow T_0.$$

None of the converse implications is true. We already know that $T_0 \not\Rightarrow T_1 \not\Rightarrow T_2$. Exercise 2.7.7 gives a space that is T_2, but not T_3. T_3-spaces that are not T_4 will be encountered in Corollaries 8.1.29 and 8.1.39. □

Lemma 8.1.4 *For a T_1 space X, the following are equivalent:*

(i) X is T_4.

(ii) For every pair $U, V \subseteq X$ of open sets satisfying $U \cup V = X$ there are closed sets $C, D \subseteq X$ such that $C \subseteq U, D \subseteq V$ and $C \cup D = X$.

(iii) Whenever $C \subseteq U$ with C closed and U open, there is $V \in \tau$ such that $C \subseteq V \subseteq \overline{V} \subseteq U$.

Proof. The equivalence (i)⇔(ii) is proven by taking complements.

Assume (i) holds, and let $C \subseteq U$ with C closed and U open. Then $D = X\backslash U$ is closed, and $C \cap D = \emptyset$. By (i), there are open V, W such that $C \subseteq V$, $D \subseteq W$ and $V \cap W = \emptyset$. But this is equivalent to $C \subseteq V \subseteq X\backslash W \subseteq U$, and since $X\backslash W$ is closed, we have $\overline{V} \subseteq X\backslash W$, thus (iii) holds. Conversely, assume (iii) holds and $C, D \subseteq X$ are closed and disjoint. With $U = X\backslash D$, we have $C \subseteq U$, thus by (i) there is an open V such that $C \subseteq V \subseteq \overline{V} \subseteq U$. But then $W = X\backslash\overline{V}$ is open, disjoint from V and contains D. Thus (i) holds. ■

Lemma 8.1.5 *For a T_1 space, the following are equivalent:*

(i) X is T_3.

(ii) Whenever $x \in U \in \tau$, there is $V \in \tau$ such that $x \in V \subseteq \overline{V} \subseteq U$.
(Equivalently, each point has a neighborhood base consisting of closed neighborhoods.)

Proof. Proven as (i)⇔(iii) in the preceding proof, taking $C = \{x\}$. ■

The result of Lemma 7.4.19 actually holds for all T_3-spaces:

Corollary 8.1.6 *If X is T_3 and $x, y \in X$ with $x \neq y$ then there are open $U, V \subseteq X$ such that $x \in U$, $y \in V$ and $\overline{U} \cap \overline{V} = \emptyset$.*

Proof. Since X is T_2, there are disjoint open U', V' such that $x \in U'$, $y \in V'$. By (ii) in Lemma 8.1.5, we can find open U, V with $x \in U \subseteq \overline{U} \subseteq U'$, $y \in V \subseteq \overline{V} \subseteq V'$. Clearly $\overline{U} \cap \overline{V} = \emptyset$. ■

Exercise 8.1.7 Let X be a topological space, Y a T_3-space, $A \subseteq X$ dense and $f : A \to Y$ continuous. Prove that f has a (unique) continuous extension $\widehat{f} : X \to Y$ if and only if f extends continuously to $A \cup \{x\}$ for every $x \in X\backslash A$.

Lemma 7.4.1 immediately gives that a compact Hausdorff space is T_3. In fact:

Proposition 8.1.8 *Every compact Hausdorff space is normal.*

Proof. Let C, D be disjoint closed subsets. By Lemma 7.3.4, they are compact. Now apply Exercise 7.4.5. ■

Locally compact Hausdorff spaces need not be normal, cf. Corollary 8.1.29. But:

Corollary 8.1.9 *Locally compact Hausdorff spaces are T_3.*

Proof. If $x \in U \in \tau$ then Lemma 7.8.24 gives an open V and a compact K such that $x \in V \subseteq K \subseteq U$. Since X is Hausdorff, K is closed, thus $\overline{V} \subseteq \overline{K} = K \subseteq U$, implying $x \in V \subseteq \overline{V} \subseteq U$. By Lemma 8.1.5, this is equivalent to regularity. ■

The following generalizes Proposition 7.8.25(ii) to all T_3-spaces. We omit the proof, which is standard by now.

Lemma 8.1.10 *If X is a T_3-space, $K \subseteq X$ is compact, $C \subseteq X$ is closed and $K \cap C = \emptyset$, then there are two disjoint open sets containing K and C, respectively.*

Lemma 8.1.11 *Every metric space is normal.*

Proof. Let $C, D \subseteq (X, d)$ be disjoint closed subsets. Since C, D are closed and disjoint, we have $x \in C \Rightarrow \mathrm{dist}(x, D) > 0$ and $y \in D \Rightarrow \mathrm{dist}(y, C) > 0$. (See Exercise 2.1.20.) Thus

$$U = \bigcup_{x \in C} B\left(x, \frac{\mathrm{dist}(x, D)}{2}\right), \qquad V = \bigcup_{y \in D} B\left(y, \frac{\mathrm{dist}(y, C)}{2}\right)$$

are both open and $C \subseteq U$ and $D \subseteq V$. Assume $z \in U \cap V$. This means that there are $x \in C,\ y \in D$ such that

$$z \in B\left(x, \frac{\mathrm{dist}(x, D)}{2}\right) \cap B\left(y, \frac{\mathrm{dist}(y, C)}{2}\right).$$

But now the triangle inequality implies $d(x, y) \leq d(x, z) + d(y, z) < (\mathrm{dist}(x, D) + \mathrm{dist}(y, C))/2$, and using $\mathrm{dist}(x, D) \leq d(x, y) \geq \mathrm{dist}(y, C)$ we have $2d(x, y) < \mathrm{dist}(x, D) + \mathrm{dist}(y, C) \leq 2d(x, y)$, which is absurd. Thus $U \cap V = \emptyset$. ■

Remark 8.1.12 1. If d is a pseudometric such that τ_d is T_0 then d is a metric. Thus for the topology τ_d, either all separation axioms from T_0 to T_4 (actually up to T_6) hold or none.

2. Metric spaces and compact Hausdorff spaces are the most important classes of normal spaces. (In Section 8.5, we will prove that every product of a metric space with a compact Hausdorff space is normal.) □

Lemma 8.1.13 *Let X be a T_1-space such that given $C \subseteq U \subseteq X$, where C is closed and U open, there is a countable family $\{W_i\}_{i \in \mathbb{N}}$ of open sets such that $C \subseteq \bigcup_i W_i$ and $\overline{W_i} \subseteq U\ \forall i$. Then X is normal (T_4).*

Exercise 8.1.14 (Proof of Lemma 8.1.13) Let C, D be disjoint closed subsets of X.

(i) Use the hypothesis of the lemma to obtain two families $\{W_i\}_{i \in \mathbb{N}}, \{V_i\}_{i \in \mathbb{N}}$ of open sets such that

$$C \subseteq \bigcup_i W_i, \qquad \overline{W_i} \cap D = \emptyset\ \forall i, \qquad D \subseteq \bigcup_i V_i, \qquad \overline{V_i} \cap C = \emptyset\ \forall i.$$

(ii) Defining

$$U = \bigcup_{i=1}^{\infty} \left(W_i \backslash \bigcup_{j \leq i} \overline{V_j} \right), \qquad V = \bigcup_{i=1}^{\infty} \left(V_i \backslash \bigcup_{j \leq i} \overline{W_j} \right),$$

prove that U and V are open.

(iii) Prove $C \subseteq U,\ D \subseteq V$.

(iv) Prove $U \cap V = \emptyset$.

The following shows that the hypothesis of compactness in Proposition 8.1.8 can be replaced by the weaker Lindelöf property, at the expense of strengthening T_2 to T_3:

Proposition 8.1.15 *(i) Every countable T_3-space is T_4.*

(ii) Every second countable T_3-space is T_4.

(iii) Every Lindelöf T_3-space is T_4.

Proof. Let $C \subseteq U \subseteq X$, where C is closed and U open.

(i) For $x \in C$, use the T_3-property to find an open U_x such that $x \in U_x \subseteq \overline{U_x} \subseteq U$ and apply Lemma 8.1.13 to $\{U_x\}_{x \in C}$.

(ii) Let $\mathcal{B}$ be a countable base. As in (i), for every $x \in C$ choose an open V_x such that $x \in V_x \subseteq \overline{V_x} \subseteq U$. Since $\mathcal{B}$ is a base, there is a $U_x \in \mathcal{B}$ such that $x \in U_x \subseteq V_x$, which in turn implies $x \in U_x \subseteq \overline{U_x} \subseteq B$. Now the family $\{U_x \mid x \in C\} \subseteq \mathcal{B}$ is countable and satisfies the assumptions in Lemma 8.1.13.

(iii) By the T_3-property and Lemma 8.1.5, we can choose, for every $x \in C$ an open U_x such that $x \in U_x \subseteq \overline{U_x} \subseteq U$. Now $\{U_x\}_{x \in C} \cup \{X \backslash C\}$ is an open cover of X, which by the Lindelöf property has a countable subcover. Thus there is a countable subset $Y \subseteq X$ such that $\bigcup_{x \in Y} U_x \supseteq C$. Now the family $\{U_x\}_{x \in Y}$ satisfies the hypothesis of Lemma 8.1.13. ■

Remark 8.1.16 1. Claims (i) and (ii) of course follow from (iii) together with Exercise 7.1.5, but the proofs of (i) and (ii) arguments given above are more direct.

2. The proofs of (i) and (ii) actually made no use at all of the closed set C! This suggests that we actually have proven a stronger result. This is true, cf. Corollary 8.2.11.

3. Later we will use (ii) to prove that second countable T_3-spaces are even metrizable, cf. Theorem 8.2.33. For the spaces in (i) and (iii) this need not be true since they can fail to be first countable, as witnessed by the Arens-Fort space from Exercise 4.3.14 which is T_3 (thus T_6 by Corollary 8.2.11) and cubes of uncountable dimension, respectively. □

As usual, we need to study how the T_3- and T_4- properties behave w.r.t. the constructions from Section 6. Skipping over the trivial case of direct sums, we begin with continuous surjections. In Example 6.4.19 we have seen that neither the T_2- nor the T_3-property are preserved by continuous maps, not even closed ones. Luckily, the T_4-property is better behaved, allowing us to tie up one loose end from Section 6.4:

Proposition 8.1.17 *If (X, τ) is T_4 and $f : (X, \tau) \to (Y, \sigma)$ is surjective, continuous and closed then Y is T_4.*

Proof. X is T_1, thus singletons are closed. By closedness of f, also their images are closed and thus $f(X) = Y$ is T_1. Now let $U, V \in \sigma$ such that $U \cup V = Y$. Then $U' = f^{-1}(U), V' = f^{-1}(V)$ are open subsets of X and $U' \cup V' = X$. By Lemma 8.1.4, there are closed $C' \subseteq U', D' \subseteq V'$ such that $C' \cup D' = X$. Since f is closed, $C = f(C'), D = f(D')$ are closed subsets of Y. Now we have $C \cup D = f(C') \cup f(D') = f(C' \cup D') = f(X) = Y$, as well as $C = f(C') \subseteq f(U') = f(f^{-1}(U)) = U$ and similarly $D \subseteq V$. Invoking Lemma 8.1.4 again, we have that Y is T_4. ■

Corollary 8.1.18 *Let (X, τ) be compact Hausdorff and $\sim$ an equivalence relation on X. Then the (compact) quotient space $X/\sim$ is Hausdorff if and only if $\sim$ is closed.*

Proof. The $\Rightarrow$ direction was Corollary 7.4.12. The $\Leftarrow$ direction follows from Proposition 8.1.8, by which X is T_4 and Proposition 8.1.17. ■

The above corollary gives a simpler solution of Exercise 7.8.21(i): It suffices to note that closedness of $\sim$ follows from closedness of $\partial I^n \subseteq I^n$.

Exercise 8.1.19 Let (X, τ) be a topological space and $Y \subseteq X$. Define X/Y as in Exercise 6.4.16. Prove:

(i) If X is T_3 and Y is closed then X/Y is Hausdorff.

(ii) If X is T_4 and Y is closed then X/Y is T_4.

We already know that the T_1- and T_2-properties are hereditary and well behaved with respect to products. The same is true for the T_3-property:

Exercise 8.1.20 (i) Prove that the T_3-property is hereditary.

(ii) Let $X_i \neq \emptyset \ \forall i \in I$. Prove that $\prod_i X_i$ is a T_3-space if and only if X_i is a T_3-space $\forall i$. Hint: Remember Corollary 6.5.14.

Unfortunately, the T_4-property is neither hereditary nor is it true the a product of T_4-spaces must be T_4! Examples for both phenomena will be encountered soon. But at least, we have:

Lemma 8.1.21 *Every subspace Y of a T_4-space (thus in particular of a compact Hausdorff space) is T_3, and if Y is second countable then it is T_4.*

Proof. Every T_4-space is T_3, and by Exercise 8.1.20 T_3 is hereditary. By Proposition 8.1.8, compact Hausdorff spaces are T_4. The rest follows from Corollary 8.1.15. ■

Remark 8.1.22 1. For X locally compact Hausdorff, X_∞ is compact Hausdorff. Thus the preceding lemma gives another (more complicated) proof of Corollary 8.1.9.

2. In Section 8.3.1 we will encounter the $T_{3.5}$-axiom for which $T_4 \Rightarrow T_{3.5} \Rightarrow T_3$ and which behaves as nicely as T_3 w.r.t. products and subspaces. Thus subspaces and products of T_4-spaces are $T_{3.5}$. In particular locally compact Hausdorff spaces are $T_{3.5}$. □

Exercise 8.1.23 Prove that every regular (T_3) space is semiregular (cf. Example 4.1.23).

Exercise 8.1.24 Let $X = \{(x,y) \mid y \geq 0,\ (x,y) \neq (0,0)\} \subseteq \mathbb{R}^2$ with the topology induced from $\mathbb{R}^2$.

(i) Prove that $E = \{(x,0) \mid x < 0\}$ and $F = \{(x,0) \mid x > 0\}$ are closed subsets of X.

(ii) Construct explicitly a continuous function $f : X \to [0,1]$ such that $f \restriction E = 1$ and $f \restriction F = 0$.

(iii) Is X normal (T_4)? (With proof!)

8.1.2 Normality of subspaces. Hereditary normality (T_5)

We begin with the following easy observation:

Exercise 8.1.25 Prove: If X is normal and $Y \subseteq X$ is a closed subspace then Y is normal.

Exercise 8.1.26 Prove that every F_σ-set in a normal space is normal. Hint: Lemma 8.1.13.

We will see shortly that normality is not hereditary. This makes it meaningful to apply our meta-Definition 2.3.7 and speak of 'hereditarily normal' space. But also other names for this property are in widespread use:

Definition 8.1.27 *Hereditarily normal spaces, i.e. spaces all subspaces of which are normal (including of course the space itself), are also called completely normal or T_5-spaces.*

Trivially, the T_5-property is hereditary. Since metrizability implies normality and is hereditary, every metric space is T_5.

Exercise 8.1.28 For a set S, let $A(S) = S \cup \{\infty_S\}$ be the one-point compactification of the discrete space (S, τ_{disc}). Prove:

(i) For any S, the space $A(S)$ is compact Hausdorff (thus normal).

(ii) For any S, the space $A(S)$ is hereditarily normal (T_5).

(iii) For any S, T, the space $X = A(S) \times A(T)$ is normal.

(iv) For any S, T, the subspace $Y = X \backslash \{(\infty_S, \infty_T)\} \subseteq X$ is locally compact.

(v) For any S, T, the subsets $C = S \times \{\infty_T\}$ and $D = \{\infty_S\} \times T$ of Y are closed and disjoint.

(vi) If S is countably infinite and T is uncountable, then the above C and D cannot be separated by open sets. (Thus Y is not normal, whence $A(S) \times A(T)$ is not hereditarily normal.)

Corollary 8.1.29 *$T_3 \not\Rightarrow T_4 \not\Rightarrow T_5$. Compact Hausdorff $\not\Rightarrow T_5$. Lindelöf $T_3 \not\Rightarrow T_5$. Locally compact Hausdorff $\not\Rightarrow T_4$.*

Proof. The space $A(S) \times A(T)$ in Exercise 8.1.28 is compact Hausdorff, thus Lindelöf and T_4. Since the subspace $Y \subseteq X$ is non-normal, the T_4-property is not hereditary, thus $T_4 \not\Rightarrow T_5$. This also shows that neither compact Hausdorff spaces nor Lindelöf T_3-spaces are automatically T_5. By Exercise 8.1.20(i), Y is T_3, thus $T_3 \not\Rightarrow T_4$. And Y is open, thus locally compact Hausdorff, so that locally compact Hausdorff $\not\Rightarrow T_4$. ■

The definition of hereditarily normal spaces is unwieldy, but fortunately they have a nice characterization similar to the definitions of T_2, T_3, T_4-spaces:

Definition 8.1.30 *Two subsets $A, B \subseteq (X, \tau)$ are called separated if $\overline{A} \cap B = \emptyset = A \cap \overline{B}$.*

Notice that disjoint closed sets are separated, as are disjoint open sets.

Exercise 8.1.31 Let X be a T_1-space. Prove that the following are equivalent:

(i) X is hereditarily normal (T_5).

(ii) Every open subspace of X is normal.

(iii) Whenever $A, B \subseteq X$ are separated, there are disjoint open U, V with $A \subseteq U, B \subseteq V$.

Hint: For the implication (ii)⇒(iii), consider the subspace $Y = X \backslash (\overline{A} \cap \overline{B})$.

Exercise 8.1.32 (Hereditary normality (T_5) of the Sorgenfrey line $(\mathbb{R}, \tau_S)$)

(i) Let $A, B \subseteq \mathbb{R}$ be separated w.r.t. τ_S. Show that for every $a \in X \backslash \overline{B}$ there is an x_a such that $[a, x_a) \subseteq X \backslash \overline{B}$.

(ii) Define $U = \bigcup_{a \in A} [a, x_a) \supseteq A$ and define $V \supseteq B$ similarly. Show that $U, V \in \tau_S$.

(iii) Prove $U \cap V = \emptyset$.

Theorem 8.1.33 *For every totally ordered set $(X, \leq)$ the order topology $\tau_{\leq}$, cf. Definition 4.2.5, is hereditarily normal.*

Proof. See e.g. [46]. (Without the axiom of choice the theorem is false, cf. [146].) ■

Remark 8.1.34 More generally, one can prove hereditary normality for all generalized ordered spaces, unifying the above results (since the Sorgenfrey line is a generalized ordered space). □

8.1.3 Normality of finite products

We have seen that normality is not hereditary. This means that we cannot use Corollary 6.5.14 to conclude normality of the X_i from normality of $\prod_i X_i$. However, the conclusion still holds:

Exercise 8.1.35 Let $X_i \neq \emptyset$ for every $i \in I$, and assume that $\prod_i X_i$ is normal.

(i) Prove that X_i is T_1 for every i.

(ii) Prove that X_i is T_4 for every i.

Our next goal is to prove that the Sorgenfrey plane $(\mathbb{R}, \tau_S) \times (\mathbb{R}, \tau_S)$ is not normal despite the normality of $(\mathbb{R}, \tau_S)$ (Exercise 8.1.32). We need a technical tool:

Lemma 8.1.36 *If (X, τ) is separable and normal then the cardinality of a closed discrete subspace $D \subseteq X$ must be strictly smaller than $\mathfrak{c} = \#\mathbb{R}$.*

Proof. Let $D \subseteq X$ be discrete. Then any $S \subseteq D$ as well as $D \backslash S$ are closed in D and therefore closed in X since D is closed. Thus by normality of X, there are disjoint open sets U_S, V_S such that $S \subseteq U_S,\ D \backslash S \subseteq V_S$. By separability, we can choose a countable dense set $C \subseteq X$. We now define a map

$$f : P(D) \to P(C), \quad S \mapsto C \cap U_S.$$

Let $S, T \in D$ such that $S \neq T$. Thus either $S \not\subseteq T$ or $T \not\subseteq S$. We consider the first case, where we have $S \cap (D \backslash T) \neq \emptyset$ and therefore $U_S \cap V_T \neq \emptyset$. Since C is dense, we have $C \cap U_S \cap V_T \neq \emptyset$. On the other hand, we have $U_T \cap V_T = \emptyset$ by our choice of these sets, and thus $C \cap U_T \cap V_T = \emptyset$. Now it is clear that $C \cap U_S \neq C \cap U_T$, and similar argument works in the case $T \not\subseteq S$. We conclude that the map f is injective. But this implies $\#P(D) \leq \#P(C)$. Combining this with $\#D < \#P(D)$ and $\#P(C) = \#P(\mathbb{N}) = \#\mathbb{R}$, we obtain $\#D < \#\mathbb{R}$, as claimed. ■

Proposition 8.1.37 *The Sorgenfrey plane $(\mathbb{R}, \tau_S) \times (\mathbb{R}, \tau_S)$ is not normal.*

Proof. We know from Lemma 6.5.16 that the Sorgenfrey plane $(\mathbb{R}, \tau_S)^2$ is separable, but has a discrete subspace with the cardinality $\mathfrak{c}$ of the continuum. Now its non-normality follows from Lemma 8.1.36. ■

Remark 8.1.38 Since metrizable spaces are normal, we conclude that the Sorgenfrey plane is not metrizable. Since finite products of metrizable spaces are metrizable, this gives a new proof for the non-metrizability of $(\mathbb{R}, \tau_S)$. (The one obtained in Exercise 4.3.12 was more elementary.) □

Corollary 8.1.39 *(i) $T_3 \not\Rightarrow T_4$.*

(ii) Products of normal spaces need not be normal.

(iii) Products of Lindelöf spaces need not be Lindelöf.

Proof. (i) The Sorgenfrey plane is not T_4, but as a product of T_3-spaces it is T_3 by Exercise 8.1.20(ii).

(ii) By Exercise 8.1.32, the Sorgenfrey line $(\mathbb{R}, \tau_S)$ is T_5, but by Proposition 8.1.37 the Sorgenfrey plane $(\mathbb{R}, \tau_S)^2$ is not T_4.

(iii) We know that the Sorgenfrey line is Lindelöf and T_3, and also that products of T_3-spaces are T_3. If products of Lindelöf spaces were Lindelöf, it would follow that the Sorgenfrey plane is Lindelöf T_3, and thus normal by Proposition 8.1.15(iii), contrary to Proposition 8.1.37. ■

If X is discrete (thus normal) and Y is normal then $X \times Y$ is a direct sum of $\#X$ copies of Y and as such it is normal. The following result shows that the discrete spaces are *the only* normal spaces that are well-behaved under products with normal spaces:

Theorem 8.1.40 *If $X \times Y$ is normal for every normal space Y then X is discrete.*

The proof [200] is quite difficult and was given only in 1976 by M. E. Rudin[1]. (For an accessible review and further references see [9].) On the positive side, there is the following result:

Theorem 8.1.41 *The product of a compact Hausdorff space with a metrizable space is normal.*

The proof will be given in Section 8.5, using paracompactness, yet another – rather important – generalization of compactness. In Section 8.5.6, Theorem 8.1.41 is used to define a class of 'generalized metric spaces' that behaves well w.r.t. closed subspaces and countable products.

8.1.4 ⋆ Normality of infinite products

We have seen that a product of two normal spaces can fail to be normal. Yet, there are cases where even infinite products are normal:

Lemma 8.1.42 *A countable product of metrizable spaces is normal. Any product of compact Hausdorff spaces is normal.*

Proof. A countable product of metrizable spaces is metrizable by Corollary 6.5.36, thus normal by Lemma 8.1.11. Any product of compact Hausdorff spaces is Hausdorff and compact (by Tychonov), thus normal by Proposition 8.1.8. ■

On the other hand:

Proposition 8.1.43 *Any product of uncountably many infinite discrete spaces is non-normal.*

Proof. It suffices to prove this for an uncountable product of countably infinite discrete spaces. (If X_i is infinite, it has a countably infinite subset Y_i, which of course is closed. Then $\prod_i Y_i$ is closed in $\prod_i X_i$, so that normality of $\prod_i X_i$ would imply normality of $\prod_i Y_i$ by Exercise 8.1.25.) Let thus I be uncountable and $X = (\mathbb{N}, \tau_{\text{disc}})^I$.

For $n \in \mathbb{N}$, define

$$A_n = \{x \in X \mid \#\{i \in I \mid p_i(x) = k\} \leq 1 \ \forall k \neq n\}.$$

Thus x is in A_n if no integer other than n appears more than once among the coordinates $p_i(x)$. Each A_n is closed since its complement is given by

$$X \backslash A_n = \bigcup_{k \neq n} \bigcup_{\substack{i,j \in I \\ i \neq j}} p_i^{-1}(k) \cap p_j^{-1}(k),$$

which is open. If $n \neq m$ then $A_n \cap A_m$ consists of the points $x \in X$ for which no integer appears more than once as a coordinate $p_i(x)$, thus $I \to \mathbb{N}, i \mapsto p_i(x)$ is injective. But this is impossible since I is uncountable. Thus the A_n are mutually disjoint. Let $U \supseteq A_1$ be open.

[1] Mary Ellen Rudin (1924-2013). American mathematician, working mostly on point-set topology. Married to the analyst Walter Rudin.

Given $x \in X$ and a finite subset $\{i_1, \ldots, i_n\} \subseteq I$, we define

$$U(x; i_1, \ldots, i_n) = \bigcap_{k=1}^{n} p_{i_k}^{-1}(p_{i_k}(x)).$$

This is a neighborhood of x contained in the canonical base $\mathcal{B}_\Pi$ for the product topology. Let $x_1 \in X$ be the point with all coordinates equal to 1. Then $x_1 \in A_1 \subseteq U$, thus there exist $i_1, \ldots, i_{n_1} \in I$ such that $U(x_1; i_1, \ldots, i_{n_1}\} \subseteq U$. Now let $x_2 \in X$ be the point all of whose coordinates are 1 except that $p_{i_k}(x_2) = k$ for all $k = 1, \ldots, n_1$. Since no integer other than 1 appears more than once as a coordinate of x_2, we have $x_2 \in A_1$. Now find $i_{n_1+1}, \ldots, i_{n_2} \in I$ such that $U(x_2; i_1, \ldots, i_{n_1}, i_{n_1+1}, \ldots, i_{n_2}) \subseteq U$. (It is understood that the i_k are all different.) Iterating this, we obtain sequences $\{x_k\}$ in X, $\{n_k\}$ in $\mathbb{N}$ and $\{i_k\}$ in I such that all coordinates of x_k are 1 except that $p_{i_l}(x_k) = l$ for $l = 1, \ldots, n_k$ and such that $(x_k; i_1, \ldots, i_{n_k}) \subseteq U$.

Now let $y \in X$ be the point all of whose coordinates are 2 except that $p_{i_l}(y) = l$ for all $l \in \mathbb{N}$. Since no coordinate other than 2 appears repeatedly, $y \in A_2$. Let now $V \supseteq A_2$ be open. Then there exists $J = \{j_1, \ldots, j_m\} \subseteq I$ such that $U(y; j_1, \ldots, j_m) \subseteq V$. Since J is finite, there exists $s \in \mathbb{N}$ such that

$$J \cap \{i_1, i_2, \ldots\} = J \cap \{i_1, \ldots, i_{n_s}\}.$$

Now define $z \in X$ by

$$p_i(z) = \begin{cases} k & \text{if } \ i = i_k \in \{i_1, \ldots, i_{n_s}\} \\ 1 & \text{if } \ i \in \{i_{n_s+1}, \ldots, i_{n_{s+1}}\} \\ 2 & \text{otherwise} \end{cases}$$

Now, if $i_k \in J \cap \{i_1, \ldots, i_{n_s}\}$ then $p_{i_k}(z) = k = p_{i_k}(y)$, and for $i \in J\backslash\{i_1, \ldots, i_{n_s}\}$ we have $p_i(z) = 2$ since $J \cap \{i_{n_s+1}, \ldots, i_{n_{s+1}}\} = \emptyset$. Thus $p_i(z) = p_i(y)\ \forall i \in J$, so that $z \in U(y; j_1, \ldots j_m) \subseteq V$. Furthermore, we have $p_{i_k}(z) = k = p_{i_k}(x_{s+1})\ \forall i_k \in \{i_1, \ldots, i_{n_s}\}$ and $p_i(z) = 1 = p_i(x_{s+1})\ \forall i \in \{i_{n_s+1}, \ldots, i_{n_{s+1}}\}$, proving $z \in U(x_{s+1}; i_1, \ldots, i_{n_{s+1}}) \subseteq U$. We therefore have $z \in U \cap V$, thus X is not normal. ■

Corollary 8.1.44 *If an uncountable product $\prod_{i\in I} X_i$ of non-empty T_1-spaces is normal then at most countably many spaces X_i are not countably compact.*

Proof. Define $J = \{i \in I \mid X_i$ is not countably compact$\}$. By Exercise 7.7.8, every space X_j with $j \in J$ has an infinite closed discrete subspace Y_i. If J is uncountable then $\prod_{j\in J} Y_j$ is non-normal by Proposition 8.1.43. As before, this implies that $\prod_{j\in J} X_j$ is non-normal. Now Exercise 8.1.35 applied to $X \cong \prod_{i\in J} X_i \times \prod_{i\in I\backslash J} X_i$ shows that $\prod_{i\in I} X_i$ is non-normal. ■

Theorem 8.1.45 *A product $\prod_{i\in I} X_i$ of non-empty metrizable spaces is normal if and only if at most countably many X_i are non-compact.*

Proof. If $J = \{i \in I \mid X_i$ is non-compact$\}$ is countable then $\prod_{i\in J} X_i$ is metrizable, whereas $\prod_{i\in I\backslash J} X_i$ is compact Hausdorff. Now $X \cong \prod_{i\in J} X_i \times \prod_{i\in I\backslash J} X_i$ is normal by Theorem 8.1.41.

The converse follows from Corollary 8.1.44 since compactness and countable compactness are equivalent for metric spaces by Theorem 7.7.23. ■

Corollary 8.1.46 *A product of discrete spaces is normal if and only if at most countably many of them are infinite.*

Using Proposition 8.1.43, we can give another example of a normal space that is not hereditarily normal:

Corollary 8.1.47 *The space I^I, where $I = [0, 1]$, (cf. Example 7.7.13) is T_4 but not T_5.*

Proof. I^I is compact Hausdorff, thus normal. Let $A = \{1/n \mid n \in \mathbb{N}\} \subseteq I$. Equipped with the subspace topology, A is countable discrete and therefore homeomorphic to $(\mathbb{N}, \tau_{\text{disc}})$. Thus the subspace $A^I \subseteq I^I$ is homeomorphic to $(\mathbb{N}, \tau_{\text{disc}})^I$, which is non-normal by Proposition 8.1.43. ■

8.1.5 Normality and shrinkings of covers

The results of this section improve on the statement (ii) Lemma 8.1.4, which is equivalent to normality. The main application will be to partitions of unity, considered in Section 8.2.5.

Exercise 8.1.48 Prove that a T_1-space X is normal if and only if for every finite open cover $\{U_1, \ldots, U_n\}$ there is a closed cover $\{C_1, \ldots, C_n\}$ with $C_i \subseteq U_i$ $\forall i$.

Hint: Use Lemma 8.1.4(ii) and induction.

Definition 8.1.49 *Given a cover $\mathcal{U}$ of X, a shrinking of $\mathcal{U}$ is a cover $\mathcal{V} = \{V_U \subseteq X\}_{U \in \mathcal{U}}$ such that $V_U \subseteq U$ $\forall U \in \mathcal{U}$. If all V_U are open (closed) then $\mathcal{V}$ is called an open (closed) shrinking.*

Remark 8.1.50 1. If $\mathcal{U} \subseteq P(X)$ is a cover of X, a subcover $\mathcal{V} \subseteq \mathcal{U}$ is a special case of a shrinking, where for each $U \in \mathcal{U}$ we have either $V_U = U$ or $V_U = \emptyset$.

2. Exercise 8.1.48 just says that every finite open cover has a closed shrinking. □

Actually, one can do somewhat better than Exercise 8.1.48:

Lemma 8.1.51 *Let X be normal and $\{U_1, \ldots, U_n\}$ a finite open cover of X. Then there is an open cover $\{V_1, \ldots, V_n\}$ such that $\overline{V_k} \subseteq U_k$ $\forall k$.*

Proof. Let $W_1 = U_2 \cup \cdots \cup U_n$. Since $\{U_1, \ldots, U_n\}$ is a cover of X, we have $U_1 \cup W_1 = X$. This is equivalent to $F_1 \subseteq U_1$ where $F_1 = X \backslash W_1$ is closed. By Lemma 8.1.4(iii), there is an open V_1 such that $F_1 \subseteq V_1 \subseteq \overline{V_1} \subseteq U_1$. Now $F_1 \subseteq V_1$ is equivalent to $V_1 \cup W_1 = X$, thus $\{V_1, U_2, U_3, \ldots, U_n\}$ is an open cover of X. Now let $W_2 = V_1 \cup U_3 \cup \cdots \cup U_n$ and $F_2 = X \backslash W_2$. Now $F_2 \subseteq U_2$ and as above we can find an open V_2 such that $F_2 \subseteq V_2 \subseteq \overline{V_2} \subseteq U_2$. This is equivalent to $\{V_1, V_2, U_3, \ldots, U_n\}$ being an open cover. Going on like this, we consecutively construct open $V_k \subseteq U_k$ satisfying $\overline{V_k} \subseteq U_k$ and such that $\{V_1, \ldots, V_k, U_{k+1}, \ldots, U_n\}$ is an open cover. Once we have arrived at $k = n$, we are done. ■

It is now natural to ask whether Lemma 8.1.51 remains true for an infinite open cover $\mathcal{U}$. If X is compact, this is easy: Just choose a finite subcover $\mathcal{U}' \subseteq \mathcal{U}$, apply Lemma 8.1.51 to the latter to obtain a shrinking $\mathcal{V}' = \{V_U\}_{U \in \mathcal{U}'}$ of $\mathcal{U}'$ such that $\overline{V_U} \subseteq U$ $\forall U \in \mathcal{U}'$. Defining $V_U = \emptyset$ if $U \in \mathcal{U} \backslash \mathcal{U}'$, $\{V_U\}_{U \in \mathcal{U}}$ is a shrinking of $\mathcal{U}$. In the absence of compactness, one needs a certain weak finiteness condition on $\mathcal{U}$:

Definition 8.1.52 *An open cover $\mathcal{U}$ of a topological space (X, τ) is called point-finite if each $x \in X$ is contained in finitely many U's, i.e. $\#\{U \in \mathcal{U} \mid x \in U\} < \infty$ $\forall x \in X$.*

Proposition 8.1.53 *If X is normal and $\mathcal{U}$ is a point-finite open cover then there is an open cover $\mathcal{V} = \{V_U\}_{U \in \mathcal{U}}$ such that $\overline{V_U} \subseteq U$ $\forall U \in \mathcal{U}$.*

Proof. The generalization of Lemma 8.1.51 to countable covers could be proven using induction, but uncountable covers would require transfinite induction. We prefer to use Zorn's lemma instead. The idea is to consider open shrinkings $\mathcal{V} = \{V_i\}_{i\in I}$ of $\mathcal{U}$ such that (*) for each $i \in I$ either $V_i = U_i$ or $\overline{V_i} \subseteq U_i$ holds. Define a partial ordering on the set $\mathcal{F}$ of these shrinkings by declaring that $\mathcal{V} \le \mathcal{W}$ if $V_i \ne U_i \;\Rightarrow\; W_i = V_i$. Now let $\mathcal{C} \subseteq \mathcal{F}$ be a chain, i.e. a totally ordered subset. Define a family $\mathcal{W} = \{W_i\}_{i\in I}$ by $W_k = \bigcap_{\mathcal{V}\in\mathcal{C}} V_k$ for all $k \in I$. It is clear that (*) holds, but one must check that $\mathcal{W}$ is a cover, i.e. $\bigcup_i W_i = X$. Let $x \in X$. Since $\mathcal{U}$ is point-finite, the set $J_x = \{i \in I \mid x \in U_i\} \subseteq I$ is finite. If $W_j = U_j$ for some $j \in J_x$ then $x \in U_j = W_j \subseteq \bigcup \mathcal{W}$. It remains to consider the case where $\overline{W_j} \subseteq U_j$ for all $j \in J_x$. By definition of $\le$ and since $\mathcal{C}$ is a chain, for every $j \in J_x$ there is an element $\mathcal{V}^j \in \mathcal{C}$ such that $V_j = W_j$ for every $\mathcal{V} \ge \mathcal{V}^j$. Now let $\mathcal{Z} = \max_{j\in J_x} \mathcal{V}^j$. Since $\mathcal{Z} \in \mathcal{C}$ is a cover, we have $x \in \bigcup_i Z_i = \bigcup_i \bigcap_{j\in J_x} V_i^j = \bigcup_i W_i = \bigcup \mathcal{W}$. Thus $\mathcal{W}$ is cover, and it clearly is an upper bound for the chain $\mathcal{C}$.

Now Zorn's lemma applies and gives the existence of a maximal element $\mathcal{V} = \{V_i\}$ of $(\mathcal{F}, \le)$. If we can prove that $\overline{V_i} \subseteq U_i$ for all $i \in I$, we are done (since $\mathcal{V} \in \mathcal{F}$ by definition is a cover). So assume that instead there is an $i_0 \in I$ such that $V_{i_0} = U_{i_0}$. But then we can argue as in the proof of Lemma 8.1.51 to find an open V'_{i_0} such that $\overline{V'_{i_0}} \subseteq U_{i_0}$ and such that $\mathcal{V}$ with V_{i_0} replaced by V'_{i_0} still covers X. But this would be a cover $\mathcal{V}'$ strictly larger than $\mathcal{V}$ (w.r.t. $\le$), contradicting the maximality of the latter. ■

Exercise 8.1.54 Prove that a point-finite cover $\mathcal{U}$ has a subcover $\mathcal{V}$ from which no element can be omitted without losing the covering property. (One says $\mathcal{V}$ is irreducible.) Hint: Zorn's lemma.

8.2 Urysohn's "lemma" and its applications

The usefulness of the notion of normality was already seen in Proposition 8.1.17 and its Corollary 8.1.18. Now we will study some deeper ones, all of which rely on Urysohn's "lemma" proven next.

8.2.1 Urysohn's Lemma

Given a general topological space X, it is by no means obvious how to produce a non-constant continuous function $f : X \to \mathbb{R}$. The following result therefore is one of the most important (and beautiful) results in general topology:

Theorem 8.2.1 (Urysohn's Lemma) [2] *If (X, τ) is normal and $C, D \subseteq X$ are closed and disjoint, there exists a function $f \in C(X, [0, 1])$ such that $f\restriction C = 0,\ f\restriction D = 1$.*

Before beginning with the proof, we note that given $f \in C(X, [0, 1])$ and defining $U_r = f^{-1}([0, r))$ for $r \in (0, 1)$, the sets U_r are open and satisfy $r < s \Rightarrow \overline{U_r} \subseteq U_s$. Conversely, one can construct a $f \in C(X, [0, 1])$ from such a family of open sets, where it is enough to have U_r for r in any dense subset $S \subseteq (0, 1)$:

Lemma 8.2.2 *Let $S \subseteq (0, 1)$ be dense and let $\{V_r \subseteq X\}_{r\in S}$ be open sets satisfying*

$$r < s \Rightarrow \overline{V_r} \subseteq V_s. \tag{8.1}$$

[2] Pavel S. Urysohn (1898-1924), Russian mathematician. Drowned in the Atlantic off the French coast at age 26. The Jahrbuch website (`http://www.emis.de/MATH/JFM/JFM.html`) lists 38 publications, appearing until 1929! (Many were finished by U.'s collaborators and colleagues.)

Then

$$f(x) = \begin{cases} \inf\{s \in S \mid x \in V_s\} & \text{if } \ x \in \bigcup_{s\in S} V_s \\ 1 & \text{if } \ x \notin \bigcup_{s\in S} V_s \end{cases}$$

defines a continuous function $f : X \to [0,1]$ *such that*

$$f(x) = 0 \Leftrightarrow x \in \bigcap_{s\in S} V_s \quad \text{and} \quad f(x) = 1 \Leftrightarrow x \notin \bigcup_{s\in S} V_s.$$

Proof. If $x \in \bigcap_{s\in S} V_s$ then $f(x) = \inf(S) = 0$ since $S \subseteq (0,1)$ is dense. If $f(x) = 0$ then there are $s \in S$ arbitrarily close to zero such that $x \in V_s$. In view of (8.1) this implies $x \in \bigcap_{s\in S} V_s$. If $x \notin \bigcup_{s\in S} V_s$ then $f(x) = 1$ by definition. Since $x \in V_s$ implies $f(x) \le s$, we see that $f(x) = 1$ implies $x \notin \bigcup_{s\in S} V_s$. It remains to prove that f is continuous. Since $\{[0,a) \mid 0 < a \le 1\} \cup \{(b,1] \mid 0 \le b < 1\}$ is a subbase for the topology of $[0,1]$, it suffices to show that $f^{-1}([0,a))$ and $f^{-1}((b,1])$ are open. (I.e. f is upper and lower semicontinuous.)

If $T \subseteq \mathbb{R}$ and $a \in \mathbb{R}$ then $\inf(T) < a$ is equivalent to the existence of $t \in T$ with $t < a$. Thus

$$x \in f^{-1}([0,a)) \Leftrightarrow \inf\{s \in S \mid x \in V_s\} < a \Leftrightarrow \exists s \in S \cap (0,a) : x \in V_s \Leftrightarrow x \in \bigcup_{s\in S\cap(0,a)} V_s.$$

Thus $f^{-1}([0,a)) = \bigcup_{s\in S\cap(0,a)} V_s$, which is open. To consider $f^{-1}((b,1])$, assume $f(x) = \inf\{s \in S \mid x \in V_s\} > b$. Since $S \subseteq (0,1)$ is dense, we can choose $u, v \in S$ such that $b < u < v < f(x)$. Then we have $\overline{V_u} \subseteq V_v$ and $x \notin V_v$ (since $f(x) > v$), thus $x \in X\backslash V_v \subseteq X\backslash\overline{V_u} \subseteq \bigcup_{s\in S\cap(b,1)} X\backslash\overline{V_s}$. Conversely, if $x \in \bigcup_{s\in S\cap(b,1)} X\backslash\overline{V_s}$ then there is an $s \in S \cap (b,1)$ such that $x \notin \overline{V_s}$. Then clearly $x \notin V_s$ and thus $f(x) \ge s > b$. Thus $f^{-1}((b,1]) = \bigcup_{s\in S\cap(b,1)} X\backslash\overline{V_s}$, which clearly is open. ■

Proof of Theorem 8.2.1. If C, D are disjoint closed subsets, we have $C \subseteq U$ with $U = X\backslash D$ open. By normality and Lemma 8.1.4, there is an open set $V_{\frac{1}{2}}$ such that $C \subseteq V_{\frac{1}{2}} \subseteq \overline{V_{\frac{1}{2}}} \subseteq U$. Applying Lemma 8.1.4 to $C \subseteq V_{\frac{1}{2}}$ and to $\overline{V_{\frac{1}{2}}} \subseteq U$, we find open sets $V_{\frac{1}{4}}$ and $V_{\frac{3}{4}}$ such that

$$C \subseteq V_{\frac{1}{4}} \subseteq \overline{V_{\frac{1}{4}}} \subseteq V_{\frac{1}{2}} \subseteq \overline{V_{\frac{1}{2}}} \subseteq V_{\frac{3}{4}} \subseteq \overline{V_{\frac{3}{4}}} \subseteq U.$$

Going on like this, we find $V_{\frac{1}{8}}, V_{\frac{3}{8}}, V_{\frac{5}{8}}, V_{\frac{7}{8}}$ such that

$$C \subseteq V_{\frac{1}{8}} \subseteq \overline{V_{\frac{1}{8}}} \subseteq V_{\frac{2}{8}} \subseteq \cdots \subseteq \overline{V_{\frac{6}{8}}} \subseteq V_{\frac{7}{8}} \subseteq \overline{V_{\frac{7}{8}}} \subseteq U.$$

Iterating this, we obtain an open set V_s for every $s \in (0,1) \cap \mathbb{D}$, where $\mathbb{D}$ is the set of dyadic rationals $\mathbb{D} = \{\frac{n}{2^m} \mid n \in \mathbb{Z},\ m \in \mathbb{N}_0\}$. (Note that we have used ACDC.) By construction, we have

$$C \subseteq V_s \subseteq \overline{V_s} \subseteq U \qquad \text{and} \qquad s < t \ \Rightarrow\ \overline{V_s} \subseteq V_t. \tag{8.2}$$

Since the dyadic rationals are dense in $\mathbb{R}$, Lemma 8.2.2 applies and provides $f \in C(X,[0,1])$. Since $C \subseteq \bigcap_{s\in(0,1)\cap\mathbb{D}} U_s$, we have $f \restriction C = 0$, and $\bigcup_{s\in(0,1)\cap\mathbb{D}} U_s \subseteq X\backslash D$ implies $f \restriction D = 1$. ■

It is convenient to reformulate Urysohn's result in terms of the following:

Definition 8.2.3 *Let X be a topological space. Two subsets $A, B \subseteq X$ are completely separated if there is $f \in C(X,[0,1])$ such that $f \restriction A = 0$ and $f \restriction B = 1$.*

For $A, B \subseteq X$ we clearly have: A, B completely separated $\Rightarrow$ $\overline{A} \cap \overline{B} = \emptyset$ $\Rightarrow$ A, B separated (Definition 8.1.30) $\Rightarrow$ $A \cap B = \emptyset$. Urysohn's lemma gives:

Corollary 8.2.4 *In a normal space, we have* $\overline{A}\cap\overline{B}=\emptyset$ *if and only if* A, B *are completely separated.*

Remark 8.2.5 1. Obviously, the interval $[0,1]$ can be replaced by any bounded interval $[a,b]$.

2. All that was used about the field $\mathbb{R}$ is this proof were the facts that (i) it contains $\mathbb{Q}$ (i.e. characteristic zero), (ii) is ordered (this actually implies (i)) and that (iii) bounded sets have an infimum, i.e. the order completeness. But these properties already characterize $\mathbb{R}$ uniquely as a field.

3. Urysohn's Lemma has many applications, of which we will consider the following: Tietze's extension theorem (Section 8.2.3), the construction of 'partitions of unity' (Section 8.2.5), and the metrization results of Sections 8.2.4 and 8.5.4.

4. We have isolated a part of the proof as Lemma 8.2.2 since it has several other applications, e.g. in the proof of Theorem C.2.3.

5. Urysohn's lemma remains true as stated for spaces satisfying "$T_4 - T_1$". But then of course we cannot conclude that continuous functions separate points from each other or from closed sets. □

8.2.2 Perfect normality (T_6)

Urysohn's Lemma only claims that for disjoint closed C, D there is an $f \in C(X,[0,1])$ such that $f^{-1}(0) \supseteq C, f^{-1}(1) \supseteq D$. It is natural to ask whether one can choose f such that $f^{-1}(0) = C,\ f^{-1}(1) = D$. (Lemma 8.2.2 would give this if we knew that $\bigcap_r U_r = C$, $\bigcup_r = X\backslash D$, but that does not follow from the above construction of $\{U_r\}$.) We will see that there actually are normal spaces where this is false! As usual, this motivates another definition:

Definition 8.2.6 *A space X is called* <u>*perfectly normal*</u> *or T_6 if it is T_1 and for any two disjoint closed subsets C, D there exists an $f \in C(X,[0,1])$ such that $f^{-1}(0) = C$ and $f^{-1}(1) = D$.*

Note that (as in the definition of normality) we allow that C or D is empty. It is clear that $T_6 \Rightarrow T_4$. We will see that also $T_6 \Rightarrow T_5$. The following saves many words:

Definition 8.2.7 *A set $A \subseteq X$ is called a* <u>*zero-set*</u> *or* <u>*functionally closed*</u> *if there is a $f \in C(X,[0,1])$ such that $A = f^{-1}(0)$.*

Exercise 8.2.8 Let X be a topological space. Prove:

(i) Every zero-set is closed and G_δ.

(ii) If X is normal then every closed G_δ-set is a zero-set.

(iii) For a T_1-space X the following are equivalent:

- (α) X is perfectly normal (T_6).
- (β) Every closed set is a zero-set.
- (γ) X is normal and every closed set is G_δ.

Exercise 8.2.9 Prove:

(i) Every subspace of a T_6-space is T_6, thus the T_6-property is hereditary.

(ii) $T_6 \Rightarrow T_5$.

(iii) Metric spaces are T_6.

Some authors call spaces, in which every closed sets are G_δ, perfect. In view of (γ) above, this fits well with 'perfect normality', but it clashes with other uses of 'perfect'.

The following should be compared with Lemma 8.1.13:

Lemma 8.2.10 *A T_1-space X is perfectly normal (T_6) if and only if for every open U there is a countable family $\{W_n\}_{n\in\mathbb{N}}$ of open sets such that $\overline{W_n} \subseteq U$ $\forall n$ and $U = \bigcup_n W_n$.*

Proof. $\Rightarrow$ Since X is T_6, then by (β) of Exercise 8.2.8 there is an $f \in C(X, [0,1])$ such that $U = f^{-1}((0,1])$. Then $W_n = f^{-1}((1/n, 1])$ is open for each $n \in \mathbb{N}$, satisfies $\overline{W_n} \subseteq f^{-1}([1/n, 1]) \subseteq U$ and clearly $\bigcup_n W_n = U$.

$\Leftarrow$ Given $\{W_n\}_{n\in\mathbb{N}}$ as stated, we have $U = \bigcup_n W_n \subseteq \bigcup_n \overline{W_n} \subseteq U$, thus $U = \bigcup_n \overline{W_n}$. Thus every open set is F_σ, which is equivalent to every closed set being G_δ. If $C \subseteq U$ is closed then we trivially have $C \subseteq U = \bigcup_n W_n$. Now X is normal by Lemma 8.1.13, and therefore T_6 by Exercise 8.2.8. ■

Corollary 8.2.11 *A countable or second countable T_3-space is perfectly normal (T_6).*

Proof. Countable families $\{W_n\}_{n\in\mathbb{N}}$ satisfying the hypothesis of Lemma 8.2.10 were already obtained in the proofs of (i) and (ii) of Proposition 8.1.15. ■

Remark 8.2.12 1. Recall from Corollary 8.1.29 that there are Lindelöf T_3-spaces that are not T_5. Combining this with Exercise 8.2.9(ii) it is clear that Lindelöf $T_3 \not\Rightarrow T_6$.

2. The Sorgenfrey line is Lindelöf T_3, but not second countable. Nevertheless, one has: □

Exercise 8.2.13 Prove that the Sorgenfrey line is perfectly normal. Hint: Look at Exercise 7.1.8 for ideas.

The following provides simple examples of T_5-spaces that are not T_6:

Exercise 8.2.14 Let S be a set and let $A(S)$ be the one-point compactification of (S, τ_{disc}) (which is T_5 by Exercise 8.1.28). Prove that $A(S)$ is perfectly normal (T_6) if and only if S is countable.

Remark 8.2.15 For countable S, Corollary 8.2.40 will give the better result that $A(S)$ is metrizable. (By the above, $A(S)$ is not T_6 for uncountable S, thus not metrizable.) □

Exercise 8.2.16 Let X be a topological space. Prove:

(i) X is T_1 and first countable $\Rightarrow$ all singletons $\{x\}$ are G_δ-sets $\Rightarrow$ X is T_1.

(ii) If X is compact Hausdorff (or just countably compact and T_3) and all singletons $\{x\}$ are G_δ-sets then X is first countable.

(iii) Countably compact T_6 $\Rightarrow$ first countable.

Remark 8.2.17 1. Compare (iii) with compact metrizable $\Rightarrow$ second countable.

2. Spaces in which every singleton is G_δ are said to have 'countable pseudocharacter'.

3. Since metric spaces are first countable and T_1 and their closed sets are G_δ, the fact that first countable spaces are quite close to metric spaces (cf. Remark 5.2.28.2) and Exercise 8.2.16(i) could lead one to expect that every closed set in a first countable T_1-space is G_δ. This is not true: The lexicographical order topology on $[0,1]^2$ is Hausdorff by Exercise 4.2.7 (actually even T_5 by Theorem 8.1.33) and first countable by Exercise 4.3.11. Thus singletons are G_δ by Exercise 8.2.16(i). But

there are closed sets that are not G_δ. (This also provides another example of a T_5-space that is not T_6.)

4. By Exercise 8.2.13, the Sorgenfrey line $(\mathbb{R}, \tau_S)$ is T_6, but by Proposition 8.1.37 its square $(\mathbb{R}, \tau_S)^2$ is not even T_4. Thus also the T_6-property is not preserved by products.

5. It is not true that compact T_6 implies metrizable. Cf. [216, Space 95]. □

The following easy exercise shows that all separation axioms behave well w.r.t. direct sums:

Exercise 8.2.18 Let $X_i, i \in I$ be topological spaces. Show for every $p \in \{0,1,2,3,4,5,6\}$ that the direct sum $X = \bigoplus_i X_i$ is T_p if and only if each X_i is T_p.

Remark 8.2.19 1. Since the notions of of zero-sets and of complete separation of a pair of sets will play a rôle later, we use Lemma 8.2.2 to give 'intrinsic' characterizations of them:

(i) Two sets $C, D \subseteq X$ are completely separated if and only if there is a family $\{U_r\}_{r\in(0,1)\cap\mathbb{D}}$ of open sets satisfying (α) : $r, s \in (0,1) \cap \mathbb{D},\ r < s \Rightarrow \overline{U_r} \subseteq U_s$ and (β) : $C \subseteq U_r \subseteq X \backslash D\ \forall r$.

(ii) Similarly, $A \subseteq X$ is a zero-set if and only if there exists a family $\{U_r\}_{r\in(0,1)\cap\mathbb{D}}$ of open sets satisfying (α) and (γ) : $\bigcap_r U_r = A$.

2. Instead of $(0,1)\cap\mathbb{D}$ one could use any countable dense subset of $(0,1)$, but the dyadic rationals are the simplest choice. Since they are still somewhat complicated, one may ask whether the criteria $(\alpha)+(\beta)$ and $(\alpha)+(\gamma)$ really are more 'intrinsic' (see Remark 8.3.7.2 for a discussion of this notion) than being a zero set or completely separated. Consider the set $\{0,1\}^* = \{a \in \{0,1\}^{\mathbb{N}} \mid 1 \leq \#\{n \mid a_n = 1\} < \infty\}$ of finite sequences ending with 1, ordered lexicographically. (I.e. if $a \neq b$ and $i = \min\{j \mid a_j \neq b_j\}$ then $a < b \Leftrightarrow a_i = 0, b_i = 1$.) Now $a \mapsto \sum_{n=1}^{\infty} 2^{-n} a_n$ is an order preserving bijection $\{0,1\}^* \to (0,1)\cap\mathbb{D}$. This description of $(0,1)\cap\mathbb{D}$ seems simple enough to be called intrinsic. □

8.2.3 The Tietze-Urysohn extension theorem

Let X be topological space, $Y \subseteq X$ a subspace. It is natural to ask whether a continuous real valued function $f \in C(Y, \mathbb{R})$ on Y can be extended to a continuous function on X. More precisely: Is there an $\widehat{f} \in C(X, \mathbb{R})$ whose restriction to Y coincides with f? This question arises in many contexts.

Theorem 8.2.20 (Tietze-Urysohn extension theorem) [3] *Let (X, τ) be normal, $Y \subseteq X$ closed and $f \in C_b(Y, \mathbb{R})$. Then there exists $\widehat{f} \in C_b(X, \mathbb{R})$ such that $\widehat{f} \restriction Y = f$.*

We follow the exposition in [95] which is somewhat more conceptual than the standard one. Recall that for any topological space X and metric space Y, (2.6) defines a complete metric on $C_b(X, Y)$. For $Y = \mathbb{R}$, the metric D comes from the norm $\|f\| = \sup_x |f(x)|$, thus $(C_b(X, \mathbb{R}), \|\cdot\|)$ is Banach space. Now we observe that the restriction map $T : C_b(X, \mathbb{R}) \to C_b(Y, \mathbb{R}),\ f \mapsto f \restriction Y$ is a norm-decreasing map, and proving the extension theorem is equivalent to proving that T is surjective. Be begin with a general approximation result (used also in the proof of the Open Mapping Theorem G.2.9) which deduces surjectivity of a bounded map $A : E \to F$ between normed spaces from some form of 'almost surjectivity':

[3] H. F. F. Tietze (1880-1964) proved this for metric spaces. The generalization to normal spaces is due to Urysohn.

Lemma 8.2.21 *Let E be a Banach space, F a normed space (real or complex) and $T : E \to F$ a linear map. Assume also that there are $m > 0$ and $r \in (0,1)$ such that for every $y \in F$ there is an $x_0 \in E$ with $\|x_0\|_E \le m\|y\|_F$ and $\|y - Tx_0\|_F \le r\|y\|_F$. Then for every $y \in F$ there is an $x \in E$ such that $\|x\|_E \le \frac{m}{1-r}\|y\|_F$ and $Tx = y$. In particular, T is surjective.*

Proof. It suffices to consider the case $\|y\| = 1$. By assumption, there is $x_0 \in E$ such that $\|x_0\| \le m$ and $\|y - Tx_0\| \le r$. Applying the hypothesis to $y - Tx_0$ instead of y, we find an $x_1 \in E$ with $\|x_1\| \le rm$ and $\|y - T(x_0 + x_1)\| \le r^2$. Continuing this inductively, we obtain a sequence $\{x_n\}$ such that

$$\|x_n\| \le r^n m, \tag{8.3}$$

and

$$\|y - T(x_0 + x_1 + \cdots + x_n)\| \le r^{n+1}. \tag{8.4}$$

Now, (8.3) together with completeness of E implies[4] that $\sum_{n=0}^{\infty} x_n$ converges to an $x \in E$ with

$$\|x\| \le \|x_0\| + \|x_1\| + \cdots \le m + rm + r^2 m + \cdots = \frac{m}{1-r},$$

and taking $n \to \infty$ in (8.4) gives $Y = Tx$. ■

Proof of Theorem 8.2.20. As noted above, the restriction map $T : C_b(X, \mathbb{R}) \to C_b(Y, \mathbb{R}),\ f \mapsto f{\restriction}Y$ satisfies $\|Tf\| \le \|f\|$ w.r.t. the supremum norms on X and Y. Let $f \in C_b(Y, \mathbb{R})$, where we may assume $\|f\| = 1$, i.e. $f(Y) \subseteq [-1, 1]$. Let $A = f^{-1}([-1, -1/3])$ and $B = f^{-1}([1/3, 1])$. Then A, B are disjoint closed subsets of Y, which are also closed in X since Y is closed. Thus by Urysohn's Lemma, there is a $g \in C(X, [-1/3, 1/3])$ such that $g{\restriction}A = -1/3$ and $g{\restriction}B = 1/3$. Thus $\|g\|_X = 1/3$ and $\|Tg - f\|_Y \le 2/3$. (You should check this!) Now the above lemma is applicable with $m = 1/3$ and $r = 2/3$ and gives the existence of $\widehat{f} \in C(X, \mathbb{R})$ with $T\widehat{f} = f$ and $\|\widehat{f}\| = \|f\|$ (since $m/(1-r) = 1$). ■

Remark 8.2.22 Adding a constant to f we can arrange that $\inf f = -\sup f$. Together with $\|\widehat{f}\| = \|f\|$, this shows that if $f \in C(Y, [a, b])$ then $\widehat{f} \in C(X, [a, b])$. □

Exercise 8.2.23 Prove that for a T_1-space X, the following are equivalent: (i) X is T_4, (ii) the conclusion of Urysohn's Lemma holds, and (iii) the conclusion of Tietze's theorem holds.

Exercise 8.2.24 (i) Prove the following modification of Tietze's theorem: If (X, τ) is normal, $Y \subseteq X$ closed and $f \in C(Y, (0, 1))$, then there exists $\widehat{f} \in C(X, (0, 1))$ such that $\widehat{f}{\restriction}Y = f$.

(ii) Use (i) to prove that, for normal X and closed $Y \subseteq X$, every $f \in C(Y, \mathbb{R})$ has an extension $\widehat{f} \in C(X, \mathbb{R})$.

Remark 8.2.25 1. By the same argument, we can extend functions $f : Y \to [0, 1)$. (With Proposition 9.2.1, it will follow that there are extensions for every $f \in C(Y, Z)$ with $Z \subseteq \mathbb{R}$ connected.)

2. We now know the following: If X is normal, $Y \subseteq X$ is closed, and Z is $[0, 1]$, $[0, 1)$, $(0, 1)$ or $\mathbb{R}$ then every $f \in C(Y, Z)$ extends (non-uniquely) to $\widehat{f} \in C(X, Z)$. Since a continuous function $f \to \prod_i Z_i$ into a product space just is an aggregate of continuous functions $\{X \to Z_i\}_{i \in I}$ and the extension theorem can be applied to each Z_i individually, we have a similar extension theorem for

[4] Let X be a Banach space and $x_0, x_1, \ldots \in X$ such that $\sum_i \|x_i\| < \infty$. Thus $n \mapsto s_n = \sum_{i=0}^{n} \|x_i\|$ is convergent, thus a Cauchy sequence. Defining $y_n = \sum_{i=0}^{n} x_i$, we have $\|y_n - y_m\| = \|x_{m+1} + \cdots + x_n\| \le \|x_{m+1}\| + \cdots + \|x_n\| = s_n - s_m$. Thus $n \mapsto y_n$ is a Cauchy sequence in X, and therefore convergent by completeness. The limit is called $\sum_i x_i$.

every space Z that is a (possibly infinite) product of (open or closed) intervals and copies of $\mathbb{R}$. (For example, using $\mathbb{C} \cong \mathbb{R}^2$, Tietze's theorem also holds for complex-valued functions.) Here is a slight generalization: □

Exercise 8.2.26 Let X be normal, $Y \subseteq X$ closed, $Z \subseteq \mathbb{R}^n$ compact convex and $f \in C(Y, Z)$. Prove that there is a continuous extension $\widehat{f} : X \to Z$.

Remark 8.2.27 1. A more serious extension result than that of the exercise was given by Dugundji. In this result $\mathbb{R}^n$ is replaced by any locally convex vector space and f is not required to be bounded, at the expense of requiring X to be metrizable, cf. Theorem 8.5.36.

2. However, for spaces Z that are not products of intervals or copies of $\mathbb{R}$ or at least convex, the extension statement often fails! For example, while $X = \mathbb{R}$ is normal and $Y = (-\infty, -1] \cap [1, \infty) \subseteq \mathbb{R}$ is closed, the map $Y \to \{\pm 1\}$, $x \mapsto x/|x|$ cannot be continuously extended to $\mathbb{R}$. (We will see this as a consequence of the connectedness of $\mathbb{R}$.) Another example: The identity map $\mathrm{id} : S^1 \to S^1$ cannot be extended to a continuous map $f : D^2 \to S^1$, where $S^1 = \partial D^2 \subseteq D^2$. □

The following application of Tietze extension was already mentioned:

Lemma 8.2.28 *A pseudocompact normal space is countably compact.*

Proof. Assume that X is normal and pseudocompact but not countably compact. Then by Exercise 7.7.8, there is a countably infinite closed discrete subspace Y. Choosing a bijection $f : Y \to \mathbb{N}$ and considering f as a function $Y \to \mathbb{R}$, f is continuous. Thus by Tietze's extension theorem (for unbounded functions, cf. Exercise 8.2.24) there is a continuous extension $\widehat{f} : X \to \mathbb{R}$. Since f is unbounded, so is $\widehat{f}$, contradicting pseudocompactness. ■

8.2.4 Urysohn's metrization theorem

Given a topological space (X, τ), the metrization problem consists in finding a metric d inducing the topology τ. The following criterion for $\tau = \tau_d$ is quite trivial, but exactly what we need:

Lemma 8.2.29 *Let (X, τ) be a topological space and d a metric on X. Then the following are equivalent:*

(i) $\tau = \tau_d$.

(ii) Given $x \in U \in \tau$, there exists $r > 0$ such that $B^d(x, r) \subseteq U$, and $B^d(x, r) \in \tau$ $\forall x \in X, r > 0$.

(iii) Given $x \in U \in \tau$, there exists $r > 0$ such that $B^d(x, r) \subseteq U$, and $d : X \times X \to [0, \infty)$ is continuous w.r.t. the topology τ on X.[5]

Proof. That (i) implies (ii) and (iii) is immediate by the definition of τ_d. The two assumptions in (ii) imply $\tau \subseteq \tau_d$ and $\tau_d \subseteq \tau$, respectively, thus (i). And in view of $B^d(x, r) = f_x^{-1}((-\infty, r))$, (iii) implies (ii). ■

Here is a first application:

Exercise 8.2.30 Prove that for a compact Hausdorff space X, the following are equivalent:

[5] By Exercise 2.1.4, separate continuity already implies joint continuity.

(i) X is metrizable.

(ii) The diagonal $\Delta \subseteq X \times X$ is G_δ.

(iii) There exists $f \in C(X \times X, \mathbb{R})$ such that $f^{-1}(0) = \Delta$.

Hint: For (iii)⇒(i), prove that $d(x,y) = \sup_{z\in X} |f(x,z) - f(y,z)|$ is a metric and use Lemma 8.2.29.

Definition 8.2.31 *A family $\mathcal{D}$ of pseudometrics on a topological space (X,τ) separates points from closed sets if for every closed C and $x \in X\backslash C$ there is a $d \in \mathcal{D}$ such that $\inf_{y\in C} d(x,y) > 0$.*

Proposition 8.2.32 *If (X,τ) is a T_1-space and $\mathcal{D}$ is a countable family of continuous pseudometrics on X separating points from closed sets, then X is metrizable.*

Proof. If d is a (continuous) pseudometric then also $d'(x,y) = \min(1, d(x,y))$ is a (continuous) pseudometric. In view of $\inf_{y\in C} d'(x,y) = \min(1, \inf_{y\in C} d(x,y))$, also $\mathcal{D}' = \{d' \mid d \in \mathcal{D}\}$ separates points from closed sets. Thus we may assume that all $d \in \mathcal{D}$ are bounded by one. Choosing a bijection $\mathbb{N} \to \mathcal{D},\ n \mapsto d_n$ and defining

$$d(x,y) = \sum_{n=1}^{\infty} 2^{-n} d_n(x,y),$$

it is clear that d is a continuous pseudometric. If C is closed and $x \in X\backslash C$, by assumption there is an $n \in \mathbb{N}$ such that $\alpha = \inf_{y\in C} d_n(x,y) > 0$. Thus $\inf_{y\in C} d(x,y) \geq 2^{-n}\alpha > 0$. Applying this to the closed (by T_1) set $C = \{y\}$, we see that $x \neq y$ implies $d(x,y) > 0$, so that d is a metric. If now $x \in U \in \tau$, then $C = X\backslash U$ is open and $x \notin C$, thus by the above we have $\beta = \inf_{y\in C} d(x,y) > 0$. This implies $B(x,\beta) \cap C = \emptyset$, and therefore $B(x,\beta) \subseteq U$. Now the implication (iii)⇒(i) of Lemma 8.2.29 gives $\tau = \tau_d$. Thus X is metrizable. ■

Theorem 8.2.33 (Urysohn's Metrization Theorem) *A second countable T_3-space is metrizable.*

Proof. By Proposition 8.1.15, the space X is T_4. Let $\mathcal{B} = \{U_1, U_2, \ldots\}$ be a countable base for the topology, and let

$$S = \{(n,m) \in \mathbb{N}^2 \mid \overline{U_n} \subseteq U_m\}.$$

For every $(n,m) \in S$ use Urysohn's Lemma to find a function $f_{(n,m)} \in C(X,[0,1])$ such that $f_{(n,m)} \restriction \overline{U_n} = 0$, $f_{(n,m)} \restriction X\backslash U_m = 1$. Let $C \subseteq X$ be closed and $x \in X\backslash C$. With $U = X\backslash C$, we have $x \in U$, and since $\mathcal{B}$ is a base, there is $m \in \mathbb{N}$ such that $x \in U_m \subseteq U$. By normality and Lemma 8.1.4(iii), there is an open V such that $x \in V \subseteq \overline{V} \subseteq U_m \subseteq U$. Using again that $\mathcal{B}$ is a base, there is $n \in \mathbb{N}$ such that $x \in U_n \subseteq V \subseteq \overline{V} \subseteq U_m \subseteq U$. Thus $(n,m) \in S$, and we have $f_{(n,m)}(x) = 0$ (since $x \in \overline{U_n}$) and $f_{(n,m)} \restriction X\backslash U_m = 1$. In view of $U_m \subseteq U$, we have $C = X\backslash U \subseteq X\backslash U_m$, thus $f_{(n,m)} \restriction C = 1$. It is clear that $d_{n,m}(x,y) = |f_{n,m}(x) - f_{n,m}(y)|$ is a continuous pseudometric (bounded by 1). If $C \subseteq X$ is closed, $x \in X\backslash C$ and $(n,m) \in S$ is chosen as above, we have $\inf_{y\in C} d_{n,m}(x,y) = 1 > 0$. Thus $\mathcal{D} = \{d_{n,m} \mid (n,m) \in S\}$ is a countable family of pseudometrics separating points from closed sets, so that X is metrizable by Proposition 8.2.32. ■

Remark 8.2.34 An alternative way of deducing metrizability of X goes as follows: One first notes that the family $\mathcal{F} = \{f_{n,m} \mid (n,m) \in S\} \subseteq C(X,[0,1])$ 'separates points from closed sets' in the sense of Definition 8.3.14. Then by Proposition 8.3.15 the map

$$\iota_{\mathcal{F}} : X \to [0,1]^{\mathcal{F}} = \prod_{f\in\mathcal{F}} [0,1], \quad x \mapsto \prod_{f\in\mathcal{F}} f(x)$$

is an embedding. Since $\mathcal{F}$ is countable, $[0,1]^{\mathcal{F}}$ is metrizable, cf. Corollary 6.5.36. Thus the same holds for $X \cong \iota(X) \subseteq [0,1]^{\mathcal{F}}$.

In view of the way Corollary 6.5.36 is proven, this reasoning is not fundamentally different from the one given above. Both approaches have their virtues: The above $\iota_{\mathcal{F}}$ is an embedding irrespective of countability of $\mathcal{F}$, and this will be useful in Section 8.3.3. On the other hand, Proposition 8.2.32 will be reused in the proof of the Nagata-Smirnov metrization theorem, which also applies to spaces that are not second countable and thus do not embed into $[0,1]^{\mathbb{N}}$. □

Urysohn's theorem does not apply to non-second countable spaces, some of which are metrizable. Nevertheless, it gives rise to some if-and-only-if statements.

Corollary 8.2.35 *For a topological space X, the following are equivalent:*

(i) X is T_3 and second countable.

(ii) X is metrizable and separable ($\Leftrightarrow$ second countable $\Leftrightarrow$ Lindelöf).

Proof. In view of Urysohn's metrization theorem, we only need to observe that metrizable spaces are T_3 and that separability and second countability are equivalent for metric spaces. ■

From this one can draw the somewhat silly conclusion is that a second countable space is metrizable if and only if it is T_3. But for compact spaces one has a very satisfactory statement:

Corollary 8.2.36 *A compact Hausdorff space is metrizable if and only if it is second countable.*

Proof. By Proposition 8.1.8, a compact Hausdorff space is T_4, and the 'if' statement follows from Urysohn's metrization theorem. On the other hand, every compact metrizable space is second countable by Corollary 7.7.27. ■

Remark 8.2.37 We cannot replace 'second countable' by 'Lindelöf' in Theorem 8.2.33: This follows from the implications metrizable$\Rightarrow T_6 \Rightarrow T_5$ and the existence of Lindelöf T_3-spaces that are not T_5, like the Sorgenfrey line or a cube of uncountable dimension. (The non-metrizability of the latter follows already from its failure to be first countable, cf. Exercise 6.5.15(iii).) □

Now we can prove metrizability of certain quotient spaces:

Corollary 8.2.38 *Let X be compact metrizable.*

(i) If Y is Hausdorff and $f : X \to Y$ is a continuous surjection then Y is compact metrizable.

(ii) If $\sim$ is a closed equivalence relation on X then $X/\!\sim$ is compact metrizable. (If $\sim$ is not closed, $X/\!\sim$ is not even Hausdorff.)

Proof. (i) By Proposition 7.4.11, f is closed. By Corollary 7.7.27, X is second countable. Thus by Proposition 7.4.17, Y is second countable. Since Y is compact by Lemma 7.3.4 and Hausdorff by assumption, it is T_3. Now Y is metrizable by Urysohn's metrization theorem.

(ii) By closedness of $\sim$, the quotient map is closed. Thus by Proposition 8.1.17, $X/\!\sim$ is T_4. Now apply (i). (The last sentence follows from Proposition 7.4.11.) ■

This can be generalized a bit:

Corollary 8.2.39 *If X is metrizable and second countable (=separable) and $\sim$ is a closed equivalence relation such that each equivalence class $[x] \subseteq X$ is compact, then $X/\!\sim$ is metrizable.*

Proof. A metrizable space is normal, thus the quotient by a closed equivalence relation is normal. By Remark 7.4.18.1 and the compactness of $[x]$, $X/\sim$ is second countable. Now apply Theorem 8.2.33. ■

For a locally compact T_2-space X, the following provides necessary and sufficient conditions for the metrizability of X and X_∞:

Corollary 8.2.40 *Let X be locally compact Hausdorff space, and consider the following statements:*

(i) X_∞ is metrizable.

(ii) X_∞ is second countable.

(iii) X is second countable.

(iv) $X \cong \bigoplus_{i\in I} X_i$, where all X_i are second countable.

(v) X is metrizable.

Then (i) ⇔(ii) ⇔(iii)⇒(iv)⇔(v). Furthermore, (iv)⇏(iii).

Proof. The implications (ii)⇒(iii)⇒(iv) and (i)⇒(v) are obvious.

(i)⇔(ii) follows from Corollary 8.2.36 since X_∞ is compact Hausdorff.

(iii)⇒(ii) was Exercise 7.8.44(ii).

(iv)⇏(iii) Example: Any uncountable discrete space.

(iv)⇒(v) A locally compact Hausdorff space is T_3 by Corollary 8.1.9. Together with second countability this gives metrizability by Urysohn's metrization theorem. Thus each X_i is metrizable, and a direct sum of metrizable spaces is metrizable, cf. Exercise 6.3.9(i).

(v)⇒(iv) The proof involves the concept of paracompactness and several non-trivial results involving the latter. Cf. Corollary 8.5.21. ■

In Section 8.5 we will also prove a metrization theorem that applies to all spaces. (But there are many others.)

Exercise 8.2.41 Prove in a purely set-theoretic way (no continuous functions) that every compact Hausdorff space with G_δ diagonal $\Delta \subseteq X \times X$ is second countable. (Combined with Urysohn's theorem this gives a new proof for the non-trivial implication in Exercise 8.2.30.)

8.2.5 Partitions of unity. Locally finite families

Definition 8.2.42 *If X is a topological space and f is a real (or complex) valued function on X then the support of f is defined as* $\mathrm{supp}(f) = \overline{\{x \in X \mid f(x) \neq 0\}}$.

Definition 8.2.43 *Let $\mathcal{U} = \{U_i\}_{i\in I}$ be an open cover of a space X. Then a partition of unity subordinate to $\mathcal{U}$ is a family $\mathcal{F} = \{f_i\}_{i\in I} \subseteq C(X, [0, 1])$ satisfying*

(i) $\mathrm{supp}(f_i) \subseteq U_i \quad \forall i \in I$,

(ii) $\sum_{i\in I} f_i(x) = 1 \quad \forall x \in X$.

For the sake of transparency and in order to motivate later definitions, we first prove the following under restrictive assumptions:

Proposition 8.2.44 *If X is normal then given a finite open cover $\mathcal{U}$ of X there is a partition of unity subordinate to $\mathcal{U}$.*

Proof. Let $\mathcal{U} = \{U_i\}_{i\in I}$. By Lemma 8.1.51, we can find an open shrinking $\{V_i\}_{i\in I}$ such that $\overline{V_i} \subseteq U_i\ \forall i$. We can repeat this and find an open shrinking $\{W_i\}$ such that $\overline{W_i} \subseteq V_i\ \forall i$. Now, $\overline{W_i}$ and $X\backslash V_i$ are disjoint closed subsets. Thus by Urysohn's lemma, there are functions $g_i \in C(X, [0,1])$ such that $g_i \upharpoonright \overline{W_i} = 1$ and $g_i \upharpoonright X\backslash V_i = 0$. The latter implies $\mathrm{supp}(g_i) = \overline{\{x \in X \mid g_i(x) \neq 0\}} \subseteq \overline{V_i} \subseteq U_i$. Now, define $h(x) = \sum_{i\in I} g_i(x)$. As a finite sum of continuous functions, h is continuous. Since the g_i are non-negative, we have $h \geq g_i\ \forall i$. Together with $g_i \upharpoonright W_i = 1$ and the fact that $\{W_i\}$ is a cover of X we have $h(x) \geq 1\ \forall x$. Thus the functions $f_i = g_i/h$ are continuous, satisfy $\sum_i f_i = \sum_i g_i/h = h/h = 1$ and the support condition $\mathrm{supp}(f_i) \subseteq U_i$. ■

Remark 8.2.45 1. Partitions of unity are an extremely useful tool for many purposes, mostly in differential topology. The point is that they can be used to reduce a global construction to local ones which are easier, in particular when the elements $U \in \mathcal{U}$ of the covering have a simple structure, like $U \cong \mathbb{R}^n$.

2. Some authors define the support of a function $f : X \to \mathbb{R}$ without the closure, i.e. as $\mathrm{supp}(f) = X\backslash f^{-1}(0)$. If one does this, the above proof simplifies since one shrinking suffices instead of two. □

What about generalizing the proposition to infinite covers? Lemma 8.1.51 applies only to finite covers, but Proposition 8.1.53 gives the same conclusion provided we assume $\mathcal{U}$ to be point-finite. Now the above proof goes through up to the definition $h(x) = \sum_i g_i(x)$. Since $\mathcal{U}$ is point-finite, only finitely many $g_i(x)$ are non-zero for any $x \in X$. Thus h exists and satisfies $h \geq 1$. But trying to prove continuity of h we run into the problem that we have no control over the dependence of the sets $\{i \in I \mid g_i(x) \neq 0\}$ on x, thus cannot relate $h(x)$ to $h(y)$ for $x \neq y$. This motivates defining the following strengthening of the notion of point-finite family of sets:

Definition 8.2.46 *A family $\mathcal{U} \subseteq P(X)$ of subsets of a topological space X is called locally finite if every $x \in X$ has an open neighborhood V such that $\#\{U \in \mathcal{U} \mid U \cap V \neq \emptyset\} < \infty$. (One then says "$V$ meets only finitely many $U \in \mathcal{U}$".)*

Definition 8.2.47 *A family $\mathcal{F} \subseteq C(X, [0,1])$ of functions is called locally finite if the associated family $\{\mathrm{supp}(f) \mid f \in \mathcal{F}\}$ of subsets is locally finite.*

Lemma 8.2.48 *If X is a topological space and $\mathcal{F} \subseteq C(X, \mathbb{R})$ is a locally finite family of functions then $g(x) = \sum_{f\in\mathcal{F}} f(x)$ defines a continuous function.*

Proof. By local finiteness of $\mathcal{F}$, every $x \in X$ has a neighborhood U_x such that $\mathcal{F}_x = \{f \in \mathcal{F} \mid U_x \cap \mathrm{supp}(f) \neq \emptyset\}$ is finite. Thus for $y \in U_x$ we have $g(y) = \sum_{f\in\mathcal{F}_x} f(y)$, which is a finite sum of continuous functions, thus continuous. Since the U_x cover X and the functions $g \upharpoonright U_x$ coincide on intersections $U_x \cap U_y$, we conclude that g is continuous on X, cf. Corollary 6.2.6. ■

Theorem 8.2.49 *If X is normal then given a locally finite open cover $\mathcal{U}$ of X there is a partition of unity subordinate to $\mathcal{U}$.*

Proof. In view of the remarks preceding Definition 8.2.46, we only need to prove continuity of $h = \sum_i g_i$. But a partition of unity subordinate to a locally finite cover is locally finite, so that Lemma 8.2.48 applies. ■

Remark 8.2.50 It is natural to ask whether Theorem 8.2.49 can be generalized further. One can of course drop the local finiteness from Definition 8.2.43, interpreting $\sum_{f\in\mathcal{F}} f(x)$ in the sense of unordered summation, cf. Example 5.1.22. (It then becomes a non-trivial fact that the partial sums $x \mapsto \sum_{i\in J} f_i(x)$ are continuous for every $J \subseteq I$, cf. Exercise 5.2.32(viii).) But since local finiteness played a crucial rôle in the proof of Theorem 8.2.49, it is not clear how to construct a partition of unity without this assumption. If the open cover $\mathcal{U}$ admits a locally finite open shrinking $\mathcal{V}$ then there is a partition of unity subordinate to $\mathcal{V}$, which obviously also is subordinate to $\mathcal{U}$. Spaces where this is always possible are called paracompact. Cf. Section 8.5 for the basics of that subject. □

The following application of local finiteness should be viewed in the light of Remark 2.6.7.2:

Exercise 8.2.51 Let X be a topological space and $Y_i \subseteq X\ \forall i \in I$. Prove:

(i) If $\{Y_i\}_{i\in I}$ is locally finite then $\{\overline{Y_i}\}_{i\in I}$ is locally finite.

(ii) If $\{Y_i\}_{i\in I}$ is locally finite then $\overline{\bigcup_i Y_i} = \bigcup_i \overline{Y_i}$.

(iii) If $\{Y_i\}_{i\in I}$ is locally finite with all Y_i closed then $\bigcup_i Y_i$ is closed.

Local finiteness also gives new intrinsic characterizations of pseudocompactness and the missing implication in the proof of Proposition 7.7.33:

Proposition 8.2.52 *Let X be a topological space. Consider the following statements:*

(i) Every locally finite family of non-empty open sets in X is finite.

(ii) Every locally finite open cover of X by non-empty sets is finite.

(iii) Every locally finite open cover of X has a finite subcover.

(iv) X is pseudocompact.

Then (i)⇒(ii)⇒(iii)⇒(iv), and (i) implies statement (ii) in Proposition 7.7.33. If X is completely regular (Definition 8.3.1) then (iv)⇒(i).

Proof. (i)⇒(ii)⇒(iii) is obvious. (iii)⇒(iv) Let $f \in C(X,\mathbb{R})$ and for $i \in \mathbb{Z}$ define $U_i = f^{-1}((i-1,i+1))$. Then $\{U_i\}_{i\in\mathbb{Z}}$ is an open cover of X that is locally finite since $(i-1,i+1)\cap(j-1,j+1)$ is empty unless $|i-j| \le 1$. Now the existence of a finite subcover given by (iv) means that f is bounded.

(i)⇒ Proposition 7.7.33(ii): Let $\mathcal{F}$ be a family $\{U_i\}_{i\in\mathbb{N}}$ of non-empty open sets such that $U_{i+1} \subseteq U_i\ \forall i$. Then (ii) implies that this family is not locally finite. Thus there exists an $x \in X$ such that that every neighborhood of X meets infinitely many U_i. Since the U_i are decreasing, we conclude that $x \in \overline{U_i}\ \forall i$, thus $\bigcap_i \overline{U_i} \neq \emptyset$.

(iv)+complete regularity⇒(i) If $\mathcal{U}$ is an infinite locally finite family of non-empty open sets, we can pick an injective map $\mathbb{N} \to \mathcal{U}$, $n \mapsto U_n$. For each $n \in \mathbb{N}$, choose an $x_n \in U_n$. By regularity, we can find open V_n such that $x \in V_n \subseteq \overline{V_n} \subseteq U_n\ \forall n$. By complete regularity there are functions $f_n \in C(X,\mathbb{R})$ such that $f_n(x_n) = n$ and $f_n \upharpoonright X\backslash V_n = 0$. Thus $\mathrm{supp}(f_n) \subseteq U_n$. By local finiteness of $\mathcal{U}$ and Lemma 8.2.48, the function $f(x) = \sum_n f_n(x)$ is well-defined and continuous. In view of $f_n(x_n) = n$, f is unbounded, contradicting the pseudocompactness. ■

The natural continuation of the above discussion of locally finite families is Section 8.5 on paracompactness. But since a (minor) rôle in that discussion will be played by the notion complete regularity, we treat the latter first.

8.3 Completely regular spaces. Stone-Čech compactification

8.3.1 $T_{3.5}$: Completely regular spaces

One may wonder whether, analogously to Urysohn's Lemma for normal spaces, given a closed subset $C \subseteq X$ of a T_3-space and $x \in X \backslash C$, there is an $f \in C(X, [0,1])$ with $f(x) = 0$ and $f \upharpoonright C = 1$. Putting $U = X \backslash C$ and appealing to Lemma 8.1.5, one can find an open $V_{\frac{1}{2}}$ such that $x \in V_{\frac{1}{2}} \subseteq \overline{V_{\frac{1}{2}}} \subseteq U$. Applying Lemma 8.1.5 again, one can also find an open $V_{\frac{1}{4}}$ such that $x \in V_{\frac{1}{4}} \subseteq \overline{V_{\frac{1}{4}}} \subseteq V_{\frac{1}{2}}$. But in order to squeeze an open $V_{\frac{3}{4}}$ between $\overline{V_{\frac{1}{2}}}$ and U, one would need Lemma 8.1.4, which is not available unless X is T_4.

In fact, it turns out that the conclusion we were trying to prove is not true in all T_3-spaces! This motivates the following definition:

Definition 8.3.1 *A T_1-space X is called completely regular or $T_{3.5}$ if for every closed $C \subseteq X$ and $x \in X \backslash C$ there is $f \in C(X, [0,1])$ such that $f(x) = 0$ and $f \upharpoonright C = 1$.*

Remark 8.3.2 In analogy to Remark 8.1.3.2, many authors also consider spaces without the T_1-property. They would call our completely regular spaces 'Tychonov spaces' and use complete regularity for the analogous definition without the T_1-axiom. □

Lemma 8.3.3 $T_4 \;\Rightarrow\; T_{3.5} \;\Rightarrow\; T_3$.

Proof. The first implication follows from Urysohn's Lemma together with closedness of points (T_1), and the second is proven as (ii)⇒(i) in Exercise 8.2.23. ■

Exercise 8.3.4 Let X be a $T_{3.5}$-space, $C \subseteq X$ closed and $K \subseteq X$ compact with $C \cap K = \emptyset$. Construct a function $f \in C(X, [0,1])$ such that $f \upharpoonright C = 0$ and $f \upharpoonright K = 1$.

Hint: Do not attempt to use Urysohn's Lemma. It won't work.

Like the T_p-properties with $p \leq 3$, the $T_{3.5}$-property is well-behaved w.r.t. subspaces and products:

Exercise 8.3.5 Prove:

(i) The $T_{3.5}$-property is hereditary.

(ii) Let $X_i \neq \emptyset \; \forall i \in I$. Then $X = \prod_i X_i$ is a $T_{3.5}$-space if and only if each X_i is a $T_{3.5}$-space.

Corollary 8.3.6 *(i) Every product of T_4-spaces is completely regular.*

(ii) Every subspace of a T_4-space is completely regular.

(iii) Every subspace of a compact Hausdorff space is completely regular.

(iv) Every locally compact Hausdorff space is completely regular.

(v) $T_{3.5} \not\Rightarrow T_4$.

Proof. (i) All normal spaces are $T_{3.5}$, and products of $T_{3.5}$-spaces are $T_{3.5}$ by Exercise 8.3.5(ii).

(ii) A T_4-space is $T_{3.5}$, and the $T_{3.5}$-property is hereditary by Exercise 8.3.5(i).

(iii) Compact Hausdorff spaces are normal. Now apply (ii).

(iv) A locally compact Hausdorff space has a Hausdorff compactification. Now use (iii).

(v) In Exercise 8.1.28 we constructed a non-normal subspace Y of a compact Hausdorff (thus normal) space. By (ii) Y is $T_{3.5}$. ■

Thus in particular the Sorgenfrey plane $(\mathbb{R}, \tau_S) \times (\mathbb{R}, \tau_S)$ is $T_{3.5}$.

Remark 8.3.7 1. Also $T_3 \not\Rightarrow T_{3.5}$, but the usual counterexamples are a bit involved. Cf. e.g. [75, Example 1.5.9]. For a simpler one, cf. [179].

2. As evidenced by Section 8.2, the T_4-property has very useful and desirable properties but as we have seen, it is badly behaved w.r.t. subspaces and products. The $T_{3.5}$-property is somewhat weaker, but much better behaved. For this reason both axioms play central rôles in topology. In the remainder of Section 8.3, we will consider the most important applications and different characterization of complete regularity. (Cf. Theorem 8.3.20.)

3. Whereas the definitions of the other separation axioms are <u>intrinsic</u>, in that they refer only to the space and its topology, the $T_{3.5}$-axiom is <u>extrinsic</u>, since it involves also another space, here $[0, 1]$ with its standard topology. (Countability axioms are considered intrinsic even though they involve the smallest infinite cardinal number $\aleph_0 = \#\mathbb{N}$.) One usually prefers intrinsic definitions. There does in fact exist an equivalent intrinsic definition of complete regularity, cf. Exercise 8.3.11, but it is not very convenient. For this reason, in this case the extrinsic definition is more useful.

4. We quickly return to Remark 5.2.21.3 about the uselessness of the notion of a 'topological' property: While the fact of being 'topological' is utterly evident for all intrinsic properties, a very small amount of work may actually be needed to prove that an extrinsic property is topological, cf. the exercise below, but again this becomes a triviality if one has done it once. □

Exercise 8.3.8 Prove that complete regularity and metrizability are topological properties.

Definition 8.3.9 *A set $U \subseteq X$ is called a <u>cozero-set</u> or <u>functionally open</u> if there is a $f \in C(X, [0, 1])$ such that $U = f^{-1}((0, 1])$. (Equivalently, $X \backslash U$ is a zero-set.)*

Exercise 8.3.10 Prove that a T_1-space is completely regular if and only if it has a base consisting of cozero-sets.

The characterization of complete regularity in the preceding exercise is not entirely intrinsic, since the notion of (co)zero-set isn't. (But see Remark 8.2.19.) The following was discovered quite late (Frink 1964, Zaitsev 1967):

Exercise 8.3.11 Let (X, τ) be a T_1-space.

(i) Prove that X is completely regular ($T_{3.5}$) if and only if it admits a base $\mathcal{B}$ satisfying

- (α) If $x \in U \in \mathcal{B}$ then there is $V \in \mathcal{B}$ such that $x \notin V$ and $U \cup V = X$.
- (β) If $U, V \in \mathcal{B}$ such that $U \cup V = X$ then there are $U', V' \in \mathcal{B}$ such that $U' \cap V' = \emptyset$ and $X \backslash U \subseteq U', X \backslash V \subseteq V'$.

Hint: For $\Leftarrow$ use Lemma 8.2.2.

(ii) Show directly, i.e. not using (i) and Urysohn's lemma, that (α) and (β) are satisfied if X is normal and we take $\mathcal{B} = \tau$.

We briefly discuss the implications among the separation axioms between complete regularity and the Hausdorff property:

Definition 8.3.12 *A T_1-space X is called*

(i) strongly Hausdorff if given $x \neq y$ there are open $U \ni x, V \ni y$ such that $\overline{U} \cap \overline{V} = \emptyset$.

(ii) completely Hausdorff if given $x \neq y$ there is an $f \in C(X, \mathbb{R})$ such that $f(x) \neq f(y)$. We also say: The continuous functions separate the points of X.

(Some authors write $T_{2.5}$ for what we call completely Hausdorff, others take it to mean strongly Hausdorff. We prefer to avoid the notation.)

Lemma 8.3.13 *For a T_1-space, the following implications hold:*

$$\begin{tikzcd}
T_4 \arrow[r, Rightarrow] & T_{3.5} \arrow[r, Rightarrow] \arrow[d, Rightarrow] & T_3 \arrow[r, Rightarrow] \arrow[d, Rightarrow] & \text{semiregular} \arrow[d, Rightarrow] \\
& \overset{\text{completely}}{T_2} \arrow[r, Rightarrow] & \overset{\text{strongly}}{T_2} \arrow[r, Rightarrow] & T_2
\end{tikzcd}$$

Proof. The implication $T_3 \Rightarrow$ strongly T_2 was Corollary 8.1.6. If X is completely Hausdorff and $x \neq y$, pick $f \in C(X, \mathbb{R})$ such that $f(x) \neq f(y)$. If $\varepsilon = |f(x) - f(y)|/3$ and

$$U = f^{-1}((f(x) - \varepsilon, f(x) + \varepsilon)), \quad V = f^{-1}((f(y) - \varepsilon, f(y) + \varepsilon))$$

then $x \in U, y \in V$ and $\overline{U} \subseteq f^{-1}([f(x) - \varepsilon, f(x) + \varepsilon]), V = f^{-1}([f(y) - \varepsilon, f(y) + \varepsilon])$, implying $\overline{U} \cap \overline{V} = \emptyset$. That regularity ($T_3$) implies semiregularity was Exercise 8.1.23. Let X be semiregular and $x, y \in X,\ x \neq y$. By the T_1-property there is an open U such that $x \in U \not\ni y$. By semiregularity there is an open V such that $x \in V \subseteq \overline{V} \subseteq U \not\ni y$. Defining $W = X \backslash \overline{V}$, we have $y \in W$ and $V \cap W = \emptyset$, thus X is Hausdorff. ■

8.3.2 Embeddings into products

Definition 8.3.14 *Let X and Y_i, $i \in I$, be topological spaces. A family $\{f_i : X \to Y_i\}_{i \in I}$ of continuous functions separates points from closed sets if for every closed $C \subseteq X$ and $x \in X \backslash C$ there is an $i \in I$ such that $f_i(x) \notin \overline{f_i(C)}$.*

Proposition 8.3.15 *If the family $\{f_i : X \to Y_i\}_{i \in I}$ of continuous functions separates points from closed sets and all spaces are T_1 then the map $F : X \to \prod_i Y_i,\ x \mapsto (f_i(x))_{i \in I}$ is an embedding.*

Proof. If $x, y \in X$ with $x \neq y$ then $\{y\}$ is closed since X is T_1. Since the family $\{f_i\}$ separates points from closed sets, there is $i \in I$ such that $f_i(x) \notin \overline{f_i(\{y\})} = \overline{\{f_i(y)\}} = \{f_i(y)\}$, where we used that Y_i is T_1. This proves that F is injective, so that $F : X \to F(X)$ is a bijection. Continuity of F is clear by Proposition 6.5.2. Now let $C \subseteq X$ be closed, $x \in X$ and assume $F(x) \in \overline{F(C)}$. Then for each $i \in I$ we have

$$f_i(x) = p_i(F(x)) \in p_i(\overline{F(C)}) \subseteq \overline{p_i(F(C))} = \overline{f_i(C)},$$

where the inclusion $\subseteq$ is due to the continuity of p_i, cf. Exercise 5.2.8(v). Since this holds for all $i \in I$ and the family $\{f_i\}$ separates points from closed sets, we have $x \in C$. Thus $F(x) \in \overline{F(C)}$ implies $x \in C$, which means that F separates points from closed sets. Now F is an embedding by Lemma 6.2.9(ii). ■

Remark 8.3.16 Here is an alternative proof for the second half of the proposition, using nets instead of Lemma 6.2.9: The map $F : X \to F(X)$ being a bijection, by Proposition 5.2.5 it is a homeomorphism if and only if every net $\{x_\iota\}$ in X converges if and only if $\{F(x_\iota)\}$ converges. By Lemma 6.5.3, the second condition holds if and only if $\{f_i(x_\iota)\}$ converges in Y_i for every $i \in I$. It is clear that convergence of $\{x_\iota\}$ implies convergence of $\{f_i(x_\iota)\}$ for every i. So assume that the net $\{x_\iota\}$ in X does not converge. Then for every $x \in X$ there is an open neighborhood $U \ni x$ such that $x_\iota \not\in U$ frequently, thus $x_\iota \in C = X \backslash U$ frequently. Since the family $\{f_i\}$ separates points from closed sets, there is an f_i such that $f_i(x) \not\in \overline{f_i(C)}$. Thus $Y_i \backslash \overline{f_i(C)}$ is an open neighborhood of $f_i(x)$. Since $x_\iota \in C$ holds frequently, we frequently have $f_i(x_\iota) \in f_i(C) \subseteq \overline{f_i(C)}$, thus $p_i(F(x_\iota)) = f_i(x_\iota)$ frequently is not in $Y_i \backslash \overline{f_i(C)}$. This means that $F(x_\iota)$ frequently is not in the open neighborhood $p_i^{-1}(Y_i \backslash \overline{f_i(C)} = X \backslash p_i^{-1}(\overline{f_i(C)})$ of $F(x)$. Thus for every $x \in X$ there is an open neighborhood of $F(x)$ in which $F(x_\iota)$ frequently is not. This means that $F(x_\iota)$ does not converge to any point of $\overline{F(X)}$ (in which a limit of $F(x_\iota)$ would have to lie). □

8.3.3 The Stone-Čech compactification

Completely regular spaces by definition have enough continuous $[0,1]$-valued functions to separate points from closed sets. Taking into account Proposition 8.3.15, we are thus led to the following:

Definition 8.3.17 *If X is any topological space, we define a map*

$$\iota_X : X \to [0,1]^{C(X,[0,1])} = \prod_{f \in C(X,[0,1])} [0,1], \qquad x \mapsto \prod_{f \in C(X,[0,1])} f(x).$$

(Thus $\iota_X(x)$ is the map $C(X,[0,1]) \to [0,1]$ given by $f \mapsto f(x)$.)

Now we define $\beta X = \overline{\iota_X(X)} \subseteq [0,1]^{C(X,[0,1])}$.

It is clear that ι_X is continuous and βX is compact Hausdorff, thus normal, for any X.

Theorem 8.3.18 *The map $\iota_X : X \to [0,1]^{C(X,[0,1])}$ is an embedding if and only if X is completely regular. In this case, $(\beta X, \iota_X)$ is a Hausdorff compactification of X. If X is compact Hausdorff (thus completely regular) then $\beta X \cong X$.*

Proof. ⇒ By definition, $\iota_X(X)$ is a subspace of the compact Hausdorff space $[0,1]^{C(X,[0,1])}$, thus completely regular by Corollary 8.3.6(iii). Thus if $\iota_X : X \to \iota_X(X)$ is a homeomorphism then X is completely regular.

⇐ If X is completely regular, then given a closed set $C \subseteq X$ and $x \in X \backslash C$, complete regularity provides an $f \in C(X,[0,1])$ such that $f(x) = 0$ and $f \restriction C \equiv 1$. Thus $f(x) = 0 \not\in \{1\} = \overline{f(C)}$, so that the family $C(X,[0,1])$ separates points from closed sets in the sense of Definition 8.3.14. Now ι_X is an embedding by Proposition 8.3.15.

Since $\iota_X : X \to [0,1]^{C(X,[0,1])}$ is an embedding and $\beta X = \overline{\iota_X(X)}$ is compact Hausdorff with dense subspace $\iota_X(X) \cong X$, the pair $(\beta X, \iota_X)$ is a Hausdorff compactification of X.

If X is compact then $\iota_X(X) \subseteq [0,1]^{C(X,[0,1])}$ is compact, thus closed, implying that $\iota_X : X \to \iota_X(X) = \overline{\iota_X(X)} = \beta X$ is a homeomorphism. ■

Definition 8.3.19 *If X is completely regular, $(\beta X, \iota_X)$ is the Stone-Čech compactification*[6] *of X.*

Now we have the characterization of spaces which admit a Hausdorff compactification:

[6] Defined in 1937 by Marshall Harvey Stone (1903-1989) and Eduard Čech (1893-1960).

Theorem 8.3.20 *For a space (X, τ), the following are equivalent:*

(i) X is completely regular.

(ii) The topology τ coincides with the initial topology (also called weak topology) induced by $C(X, [0,1])$.

(iii) X is homeomorphic to a subspace of some cube I^N, where N is some cardinal number.

(iv) X is homeomorphic to a subspace of a compact Hausdorff space.

(v) X admits a Hausdorff compactification.

Proof. (i)⇔(ii) The product topology on $[0,1]^{C(X,[0,1])}$ by definition is the initial topology induced by the 'projections', which in this case are the evaluations in the points of X. Thus also the subspace topology on $\iota_X(X) \subseteq [0,1]^{C(X,[0,1])}$ is an initial topology.

(i)⇒(iii) This is the Stone-Čech compactification, cf. Theorem 8.3.18.

(iii)⇒(iv) Trivial.

(iv)⇒(v) If $\iota : X \hookrightarrow Y$ is an embedding with Y compact Hausdorff, then $\iota : X \to \overline{\iota(X)}$ is a Hausdorff compactification.

(v)⇒(i) This is Corollary 8.3.6(iii). ■

Remark 8.3.21 1. The equivalence (i)⇔(iv) should be compared with Corollary 7.8.30, according to which X is locally compact Hausdorff if and only if it is homeomorphic to an *open* subspace of a compact Hausdorff space.

2. For us, the main application of the Stone-Čech compactification is the proof of the implication (i)⇒(v) in the theorem above, which was already used in the proof of Theorem 8.5.25. But there are many others. Some of them are related to the fact that βX is the largest compactification of X (Corollary 8.3.32), others to the bijection in Corollary 8.3.38, yet others rely on deep connections to logic and model theory.

3. Using ideas from the proof of Urysohn's metrization theorem, one can show that every $T_{3.5}$ space (X, τ) embeds into $[0,1]^\chi$ whenever τ admits a base of cardinality $\leq \chi$. (I.e. if $\chi \geq w(X)$, where $w(X)$ is the weight of X.) □

Now we state an (imperfect) analogue for compact spaces of Proposition 3.2.7:

Corollary 8.3.22 *(i) If X is compact then $f(X) \subseteq Y$ is closed for every continuous injective $f : X \to Y$ with Y Hausdorff. (By Proposition 7.4.11, f automatically is an embedding.)*

(ii) If X is completely regular and $f(X) \subseteq Y$ is closed for every embedding $f : X \to Y$ into a Hausdorff space then X is compact.

Proof. (i) By Lemma 7.3.4, $f(X)$ is compact, thus by Lemma 7.4.2, $f(X) \subseteq Y$ is closed.

(ii) If X is completely regular but not compact, it embeds into the compact Hausdorff space βX as a non-closed subspace. Thus X is not universally closed. ■

Since $[0,1]^\chi$ is compact, the image of the embedding ι_X is closed if and only if X is compact. One may wonder which spaces admit a closed embedding into some $\mathbb{R}^\chi$.

Definition 8.3.23 *A topological space is called realcompact (or Hewitt complete) if it is homeomorphic to a closed subspace of $\mathbb{R}^\chi$, for some cardinal number χ.*

Lemma 8.3.24 *(i) Every realcompact space is completely regular ($T_{3.5}$).*

(ii) Compact Hausdorff $\Leftrightarrow$ realcompact + pseudocompact.

Proof. (i) Complete regularity is hereditary and preserved by products. Since $\mathbb{R}$ is completely regular, the same holds for $\mathbb{R}^\chi$ and any subspace of it.

(ii) $\Rightarrow$ A compact Hausdorff space X is pseudocompact (cf. Section 7.7.4) and completely regular. Thus by the above there is an embedding $\iota_X : X \to [0,1]^{C(X,[0,1])} \subseteq \mathbb{R}^{C(X,[0,1])}$. Since X is compact, so is $\iota_X(X)$, which thus is closed.

$\Leftarrow$ By realcompactness, there is an embedding $\iota : X \to \mathbb{R}^\chi$ with closed image. By pseudocompactness the coordinate functions $p_i \circ \iota : X \to \mathbb{R}$ of ι are all bounded, to wit $-\infty < \inf(p_i \circ \iota) < \sup(p_i \circ \iota) < \infty$. Thus $\prod_\iota [\inf(p_i \circ \iota), \sup(p_i \circ \iota)]$ is compact. Since $\iota(X)$ is closed, it is compact, thus X is compact. ■

Here some more facts:

- Obviously closed subspaces and arbitrary products of realcompact spaces are realcompact.
- The example $\mathbb{R}$ shows that realcompact $\not\Rightarrow$ compact.
- The subspace $\mathbb{N}^\mathbb{R} \subseteq \mathbb{R}^\mathbb{R}$ is closed, thus realcompact, but not normal by Proposition 8.1.43. Thus realcompact $\not\Rightarrow$ normal.
- Not every completely regular space is realcompact. For criteria equivalent to realcompactness of a completely regular space, cf. Theorems 3.11.10 and 3.11.11 of [75].
- Every Lindelöf T_3-space is realcompact. Thus the Sorgenfrey line and all separable metric spaces are realcompact.
- It is natural to ask whether every discrete space is realcompact. It turns out that this is true if and only if every cardinal number is 'non-measurable' (which we do not define). Whether this is the case depends on set-theoretic axioms beyond ZFC!

8.3.4 Topologies vs. families of pseudometrics

Pseudometrics provide an interesting perspective on complete regularity:

Lemma 8.3.25 *A topological space (X, τ) is completely regular if and only there is a family $\mathcal{D}$ of continuous pseudometrics on X that separates points from closed sets (in the sense of Definition 8.2.31). (We may assume all $d \in \mathcal{D}$ to be bounded by 1.)*

Proof. $\Rightarrow$ If X is completely regular then $\mathcal{F} = C(X,[0,1])$ separates points from closed sets. Now $\mathcal{D} = \{d_f(x,y) = |f(x) - f(y)|\}_{f \in \mathcal{F}}$ is a family of continuous pseudometrics on X that separates points from closed sets. Note that $d_f \in C(X \times X, [0,1])$.

$\Leftarrow$ If $\mathcal{D}$ is a family of continuous pseudometrics on X that separates points from closed sets then $\mathcal{F} = \{x \mapsto d(y,x)\}_{y \in X, d \in \mathcal{D}}$ is a family of continuous functions separating points from closed sets. Then $\iota_\mathcal{F} : X \to \mathbb{R}^\mathcal{F},\ x \mapsto \prod_{f \in \mathcal{F}} f(x)$ is an embedding. Since $\mathbb{R}^\mathcal{F}$ is completely regular, so is $X \cong \iota_\mathcal{F}(X) \subseteq \mathbb{R}^\mathcal{F}$. ■

As we saw in Section 8.3.3, if $\mathcal{F}$ is a family of continuous functions on (X, τ) that separates points from closed sets, the topology τ coincides with the initial topology induced by $\mathcal{F}$. This leads to the question whether we can define a topology in terms of a family of pseudometrics, as in the case of a single metric.

Lemma 8.3.26 *Let X be a set and $\mathcal{D}$ a family of pseudometrics on X. Let $\tau_{\mathcal{D}}$ be the initial topology on X induced by the family $\{x \mapsto d(x,y)\}_{y\in X, d\in\mathcal{D}} \subseteq \mathrm{Fun}(X,\mathbb{R})$. Then*

(i) An open neighborhood base for $x \in X$ is given by $\mathcal{U}_x = \{\bigcap_{i=1}^n B^{d_i}(x,r_i)\}_{d_1,\ldots,d_n\in\mathcal{D}, r_i>0}$.

(ii) The following are equivalent:

(α) $\mathcal{D}$ separates the points of X. (I.e., if $x \neq y$ then $d(x,y) > 0$ for some $d \in \mathcal{D}$.)

(β) $\tau_{\mathcal{D}}$ is T_0.

(γ) $\tau_{\mathcal{D}}$ is completely regular.

(iii) If $\mathcal{D}$ separates the points of X and is countable then $\tau_{\mathcal{D}}$ is metrizable, thus T_6.

Proof. (i) By construction, each $d \in \mathcal{D}$ is $\tau_{\mathcal{D}}$-continuous in one variable, thus in both by Exercise 2.1.4. Since $\mathbb{R}$ has the subbase $\{(-\infty,a)\} \cup \{(a,\infty)\}$, a subbase for $\tau_{\mathcal{D}}$ is given by

$$\mathcal{S} = \{\{x \in X \mid d(x,y) < r\}_{d\in\mathcal{D},y\in X,r>0} \cup \{\{x \in X \mid d(x,y) > r\}_{d\in\mathcal{D},y\in X,r>0}.$$

Now, $\{x \in X \mid d(x,y) > r\}$ is open w.r.t. the topology τ_d and thus a unit of balls $B^d(z,\varepsilon)$. Thus we can replace $\mathcal{S}$ by

$$\mathcal{S} = \{\{x \in X \mid d(x,y) < r\}_{d\in\mathcal{D},y\in X,r>0} = \{B^d(y,r)\}_{d\in\mathcal{D},y\in X,r>0}.$$

And if $x \in B^d(y,r)$, then $x \in B(x,s)$ for some $s > 0$, thus $\mathcal{U}_x$ is a neighborhood base for x.

(ii) $(\gamma) \Rightarrow (\beta)$ is trivial. $(\beta) \Rightarrow (\alpha)$. Assuming that (i) is false, there are $x, y \in X$ with $x \neq y$ such that $d(x,y) = 0$ for all $d \in \mathcal{D}$. This implies $y \in B^d(x,r)$ for all $d \in \mathcal{D}, r > 0$ and similarly for $x \leftrightarrow y$. Thus every open set containing one of the points contains both, so that $\tau_{\mathcal{D}}$ is not T_0.

$(\alpha) \Rightarrow (\gamma)$ Let C be $\tau_{\mathcal{D}}$-closed and $x \in X\backslash C =: U$. Then $x \in U \in \tau_D$, so that there are $d_1,\ldots,d_n \in \mathcal{D}$ and $r_1,\ldots,r_n > 0$ such that $\bigcap_{i=1}^n B^{d_i}(x,r_i) \subseteq U$. But this means that the continuous function $f : y \mapsto \max_{i=1}^n d_i(x,y)/r_i$ satisfies $\inf_{y\in C} f(y) \geq 1$ and $f(x) = 0$ trivially. Thus $C(X,\mathbb{R})$ separates points from closed sets, so that X is completely regular by Section 8.3.3.

(iii) By definition of $\tau_{\mathcal{D}}$, the $d \in \mathcal{D}$ are continuous and $\mathcal{D}$ separates points from closed sets. Now apply Proposition 8.2.32. ■

Remark 8.3.27 If (X,τ) is any space that is completely regular, but not normal, cf. Corollary 8.3.6, then the topology τ is the initial topology induced by $\mathcal{F} = C(X,[0,1])$, thus also $\tau = \tau_{\mathcal{D}}$, where $\mathcal{D} = \{(x,y) \mapsto |f(x) - f(y)|\}_{f\in\mathcal{F}}$. This shows that $\tau_{\mathcal{D}}$ need not be T_4. Why does the proof of Lemma 8.1.11 not extend to families of pseudometrics separating points? □

8.3.5 Functoriality and universal property of βX

In Section 7.8.5 we have seen that one-point compactification $X \mapsto X_\infty$ is functorial (only for proper maps!). We will now show also the Stone-Čech compactification $X \mapsto \beta X$ is a functor.

Theorem 8.3.28 *If X, Y are completely regular and $g \in C(X,Y)$ then there is a unique $\widehat{g} \in C(\beta X, \beta Y)$ such that $\widehat{g} \circ \iota_X = \iota_Y \circ g$.*

Proof. Uniqueness of $\widehat{g}$ follows from the fact that $\iota_X(X) \subseteq \beta X$ is dense and the prescription that $\widehat{g}$ be continuous and equal to $\iota_Y \circ g \circ \iota_X^{-1}$ on $\iota_X(X)$, cf. Exercise 6.5.18(iii). It remains to prove existence. If $f \in C(Y,[0,1])$ then $f \circ g \in C(X,[0,1])$. Recall that an element of $[0,1]^{C(X,[0,1])} = \prod_{f\in C(X,[0,1])}[0,1]$ is a map $e : C(X,[0,1]) \to [0,1],\ f \mapsto e(f)$ and similarly for $[0,1]^{C(Y,[0,1])}$. This allows to define a map

$$\gamma_g : [0,1]^{C(X,[0,1])} \to [0,1]^{C(Y,[0,1])}$$

by $\gamma_g(e)(f) = e(f \circ g)$ for every $e \in [0,1]^{C(X,[0,1])}$ and $f \in C(Y,[0,1])$. By construction, the diagram

$$\begin{tikzcd} X \arrow[r, "\iota_X"] \arrow[d, "g"'] & {[0,1]^{C(X,[0,1])}} \arrow[d, "\gamma_g"] \\ Y \arrow[r, "\iota_Y"'] & {[0,1]^{C(Y,[0,1])}} \end{tikzcd} \quad \text{commutes, since} \quad \begin{tikzcd} x \arrow[r, mapsto, "\iota_X"] \arrow[d, mapsto, "g"'] & (f \mapsto f(x)) \arrow[d, mapsto, "\gamma_g"] \\ g(x) \arrow[r, mapsto, "\iota_Y"'] & (f \mapsto f(g(x))) \end{tikzcd}$$

Now, by Lemma 6.5.3, a net e_i in $[0,1]^{C(X,[0,1])}$ converges to e if and only if $e_i(f) \to e(f)$ for all $f \in C(X,[0,1])$. But this implies that $e_i(f \circ g) \to e(f \circ g)$ for all $f \in C(Y,[0,1])$. Thus γ_g is continuous. Now,

$$\gamma_g(\beta X) = \gamma_g(\overline{\iota_X(X)}) \subseteq \overline{\gamma_g(\iota_X(X))} = \overline{\iota_Y(g(X))} \subseteq \overline{\iota_Y(Y)} = \beta Y,$$

where the first inclusion is due to continuity of γ_g, cf. Exercise 5.2.8(v). Thus if we define $\widehat{g} = \gamma_g \upharpoonright \beta X$, we have $\widehat{g}(\beta X) \subseteq \beta Y$. It is clear by construction that $\widehat{g} : \beta X \to \beta Y$ is the unique continuous function satisfying $\widehat{g} \circ \iota_X = \iota_Y \circ g$. ■

In analogy to Corollary 7.8.57 we now have:

Corollary 8.3.29 *Stone-Čech compactification $X \mapsto \beta X,\ f \mapsto \widehat{f}$ is a functor from the category $\mathcal{TOP}_{T_{3.5}}$ of completely regular spaces to the category $\mathcal{TOP}_{\rm cH}$ of compact Hausdorff spaces.*

(Both categories contain all continuous maps as morphisms and thus are full subcategories of the category $\mathcal{TOP}$ of all topological spaces and continuous maps.)

Proof. It only remains to show that $\widehat{\mathrm{id}_X} = \mathrm{id}_{\beta X}$ and $\widehat{h \circ g} = \widehat{h} \circ \widehat{g}$ for continuous $X \xrightarrow{g} Y \xrightarrow{h} Z$. The first statement is immediate since $g = \mathrm{id}_X$ implies $\gamma_g = \mathrm{id}_{[0,1]^{C(X,[0,1])}}$, and the second follows from $\gamma_{h\circ g} = \gamma_h \circ \gamma_g$ which is (almost) equally obvious. ■

Remark 8.3.30 1. The Stone-Čech compactification functor is 'better' than the one-point compactification in two respects: It is defined on a larger class of objects (completely regular spaces as opposed to locally compact Hausdorff spaces), and it is defined for all morphisms (continuous maps), not only the proper ones. On the other hand, whenever X is non-compact, βX is a *huge* space that is rather difficult to analyze. This is true even in the simplest case $\beta\mathbb{N}$, where $\mathbb{N}$ has the discrete topology. (There are entire books about the Stone-Čech compactification, e.g. [230]!) However, for discrete X there is a simpler interpretation of βX, cf. Theorem 11.1.82(iv).

2. The Stone-Čech compactification functor $\beta : \mathcal{TOP}_{T_{3.5}} \to \mathcal{TOP}_{\rm cH}$ is a left adjoint of the inclusion functor $\iota : \mathcal{TOP}_{T_{\rm cH}} \to \mathcal{TOP}_{3.5}$, thus the full subcategory $\mathcal{TOP}_{T_{\rm cH}} \subseteq \mathcal{TOP}_{3.5}$ is reflective. Cf. Appendix A.5 for definitions of these notions. □

The reason for studying βX despite its unwieldiness is provided by the following two results, which express the universal property of βX:

Corollary 8.3.31 *If X is completely regular, Y compact Hausdorff and $f \in C(X,Y)$ then there is a unique $\widehat{f} : \beta X \to Y$ such that $\widehat{f} \circ \iota_X = f$.*

Proof. Y is completely regular, and as already noted in Theorem 8.3.18, the map $\iota_Y : Y \to \beta Y$ is a homeomorphism. Identifying Y with βY, the claim follows from Theorem 8.3.28. ■

Corollary 8.3.32 *If (X, τ) is completely regular, the Stone-Čech compactification $(\beta X, \iota_X)$ is an initial object in the category $\mathcal{C}(X)$ of Hausdorff compactifications of X.*

Proof. If $(\widehat{X}, \iota)$ is some Hausdorff compactification of X, just apply Corollary 8.3.31 to $Y = \widehat{X}$ and $f = \iota$ to obtain a unique $\widehat{f} : \beta X \to \widehat{X}$ such that $\widehat{f} \circ \iota_X = \iota$. Thus $\widehat{f}$ is a morphism $(\beta X, \iota_X) \to (\widehat{X}, \iota)$ in $\mathcal{C}(X)$, and since it is unique, $(\beta X, \iota_X)$ is an initial object. ■

Remark 8.3.33 1. Corollary 8.3.32 should be compared to Corollary 7.8.59, where we showed that a terminal object in $\mathcal{C}(X)$ exists if and only if X is locally compact Hausdorff.

2. Since all initial objects in a category are isomorphic, cf. Exercise A.5.9, the Stone-Čech-compactification $(\beta X, \iota_X)$ is characterized (up to isomorphism) by its being an initial object in the category $\mathcal{C}(X)$ of Hausdorff compactifications of X. This characterization of the Stone-Čech compactification requires comparison with all other compactifications and therefore is neither intrinsic nor practical. Luckily, there is a much simpler and more intrinsic characterization of βX, which we now prove. □

Proposition 8.3.34 *Let X be completely regular and $(\widehat{X}, \iota)$ a Hausdorff compactification of X. Then:*

(i) If $C, D \subseteq X$ are such that $\overline{\iota(C)} \cap \overline{\iota(D)} = \emptyset$ then C, D are completely separated.

(ii) $(\widehat{X}, \iota)$ is isomorphic to $(\beta X, \iota_X)$ in the category $\mathcal{C}(X)$ if and only if complete separation of $C, D \subseteq X$ implies $\overline{\iota(C)} \cap \overline{\iota(D)} = \emptyset$.

Proof. (i) Since $\widehat{X}$ is compact Hausdorff, thus normal, Urysohn gives an $f \in C(\widehat{X}, [0,1])$ such that $f \upharpoonright \overline{\iota(A)} = 0$, $f \upharpoonright \overline{\iota(B)} = 1$. Restricting f to X shows that A, B are completely separated.

(ii) $\Rightarrow$ Let $C, D \subseteq X$ be separated by $f \in C(X, [0,1])$. Let $\widehat{f} \in C(\beta X, [0,1])$ be the unique function satisfying $\widehat{f} \circ \iota_X = f$, cf. Corollary 8.3.31. Assuming $\overline{\iota_X(C)} \cap \overline{\iota_X(D)} \neq \emptyset$, we have

$$\emptyset \neq \widehat{f}\left(\overline{\iota_X(C)} \cap \overline{\iota_X(D)}\right) \subseteq \widehat{f}\left(\overline{\iota_X(C)}\right) \cap \widehat{f}\left(\overline{\iota_X(D)}\right) \subseteq \overline{\widehat{f}\,(\iota_X(C))} \cap \overline{\widehat{f}\,(\iota_X(D))} = \overline{\{0\}} \cap \overline{\{1\}} = \emptyset,$$

which is absurd. (The first inclusion is just set theory, and the second is continuity of $\widehat{f}$.)

$\Leftarrow$ By Corollary 8.3.31, the continuous map $\iota : X \to \widehat{X}$ has a unique extension $\widehat{\iota} : \beta X \to \widehat{X}$. Now $\widehat{\iota}(\beta X) \subseteq \widehat{X}$ is closed since βX is compact and $\widehat{X}$ Hausdorff. On the other hand $\iota(X) \subseteq \widehat{X}$ is dense, thus $\widehat{\iota} : \beta X \to \widehat{X}$ is surjective. Thus $\widehat{\iota}$ is a homeomorphism, once we show that it is injective. If $x, y \in \beta X$, $x \neq y$, Urysohn's lemma gives $f \in C(\beta X, [0,1])$ with $f(x) = 0, f(y) = 1$. Let $g = f \upharpoonright X$ and $C = g^{-1}([0, 1/3])$, $D = g^{-1}([2/3, 1])$. Then $C, D \subseteq X$ are completely separated by the function $g'(x) = 3 \max(1/3, \min(2/3, g(x))) - 1$, so that $\overline{\iota(C)} \cap \overline{\iota(D)} = \emptyset$ holds by our assumption. By construction, $x \in \mathrm{Cl}_{\beta X}(C)$, thus $\widehat{\iota}(x) \in \widehat{\iota}(\mathrm{Cl}_{\beta X}(C)) \subseteq \overline{\widehat{\iota}(C)} = \overline{\iota(C)}$, and similarly $\widehat{\iota}(y) \in \overline{\iota(D)}$. This implies $\widehat{\iota}(x) \neq \widehat{\iota}(y)$, thus $\widehat{\iota}$ is injective. ■

Here is an immediate application (though not very useful) application:

Corollary 8.3.35 *Let X be completely regular and non-compact. Then the following are equivalent:*

(i) X has a unique (up to isomorphism) Hausdorff compactification.

(ii) X is locally compact and $X_\infty \cong \beta X$ in $\mathcal{C}(X)$.

(iii) X is locally compact and disjoint non-compact sets $C, D \subseteq X$ are cannot be separated by an $f \in C(X, [0,1])$ (in the sense of $f \restriction C = 0,\ f \restriction D = 1$).

Proof. In Remark 7.8.60.3 we have seen that existence of a unique Hausdorff compactification $\widehat{X}$ implies $\widehat{X} \cong X_\infty$ and therefore local compactness of X. Since βX is initial and X_∞ is terminal in $\mathcal{C}(X)$, the equivalence (i)⇔(ii) is now clear. It remains to show that, for locally compact X, the isomorphism $\beta X \cong X_\infty$ is equivalent to the condition under (iii). Applying the criterion in Proposition 8.3.34(ii) to X_∞ we see that $X_\infty \cong \beta X$ is equivalent to: Whenever $C, D \subseteq X$ are separated by an $f \in C(X, [0,1])$ we have $\overline{\iota_\infty(C)} \cap \overline{\iota_\infty(D)} = \emptyset$. It is clear that we can assume C, D to be closed (in X). By Exercise 7.8.18, the closure in X_∞ of a closed $C \subseteq X$ is either C or $C \cup \{\infty\}$ for compact and non-compact C, respectively. Thus for disjoint closed $C, D \subseteq X$, the intersection $\overline{\iota_\infty(C)} \cap \overline{\iota_\infty(D)}$ is either empty or $\{\infty\}$, the second alternative holding if and only if C and D are both non-compact. In view of this the condition in Proposition 8.3.34(ii) amounts to: Whenever closed $C, D \subseteq X$ are separated by a continuous function (in particular they are disjoint) then at least one of them is compact. This is the contraposition of (iii). ■

It is clear that a space satisfying (iii) cannot be normal.

Recall from Remark 8.2.19 that the notion of complete separation of sets involved in the above characterization can (in principle) be stated without invoking functions to $[0, 1]$. For normal spaces, we have a further simplification:

Corollary 8.3.36 *Let X be normal and $(\widehat{X}, \iota)$ a Hausdorff compactification of X. Then $(\widehat{X}, \iota) \cong (\beta X, \iota_X)$ in $\mathcal{C}(X)$ if and only if $\overline{\iota(C)} \cap \overline{\iota(D)} = \emptyset$ whenever $C, D \subseteq X$ are closed and disjoint.*

Proof. Immediate consequence of Corollary 8.2.4 and Proposition 8.3.34. ■

Remark 8.3.37 In Section 11.1.6, this result will be used to show that the Stone-Čech compactification βX is isomorphic in $\mathcal{C}(X)$ to a much simpler compactification γX, provided X is sufficiently disconnected. This is interesting since γX is closely related to Stone duality, which we will discuss in Section 11.1.11. □

Corollary 8.3.38 *For every completely regular space X, the map $C(\beta X, \mathbb{R}) \to C_b(X, \mathbb{R})$ given by $f \mapsto f \restriction X$ is a bijection and an isomorphism of commutative unital $\mathbb{R}$-algebras.*

Proof. Since βX is compact, every $f \in C(\beta X, \mathbb{R})$ is bounded, thus $f \restriction X \in C_b(X, \mathbb{R})$. Conversely, if $f \in C_b(X, \mathbb{R})$ then $f \in C(X, [a, b])$ for some $a < b$. Since $[a, b]$ is compact, Corollary 8.3.31 gives a unique extension $\widehat{f} \in C(\beta X, [a, b]) \subseteq C(X, \mathbb{R})$ such that $\widehat{f} \restriction X = f$. The restriction and extension maps are inverses of each other: For $f \mapsto \widehat{f} \mapsto f$ this is true by definition. Conversely, if $g \in C(\beta X, \mathbb{R})$ and $f = g \restriction X$ then g is one extension of f, and by uniqueness we have $\widehat{g \restriction X} = g$. The restriction map $f \mapsto f \restriction X$ clearly is a unit-preserving homomorphism of $\mathbb{R}$-algebras, thus by the above it is a unit-preserving algebra-isomorphism. ■

Remark 8.3.39 When X is locally compact but non-compact, the algebra $C_0(X)$ is not unital. In Exercise 7.8.83, we have obtained an isomorphism $C(X_\infty) \cong C_0(X) \oplus \mathbb{R}$ of $\mathbb{R}$-algebras, where the right hand side has a natural interpretation as the minimal unitization $(C_0(X))_\mathbf{1}$ of the algebra $C_0(X)$. But there are other unitizations: Every (decent, e.g. C^*-) algebra A (not necessarily commutative) has a multiplier algebra $M(A)$, which is a unital algebra containing A as an (essential) ideal and is commutative when A is. (Cf. e.g. [181, Sect. 2.1, 3.1].) If X is $T_{3.5}$ but not locally compact, $C_0(X)$ can be very small, but when X is locally compact Hausdorff, one can prove the algebra isomorphism $C_b(X) \cong M(C_0(X))$, cf. [181, Example 3.1.3]. Thus $C(\beta X) \cong M(C_0(X))$ in nice analogy to $C(X_\infty) \cong (C_0(X))_\mathbf{1}$. □

Exercise 8.3.40 Let X be completely regular.

(i) Use idempotents in $C_b(X, \mathbb{R})$ (i.e. functions satisfying $f^2 = f$) to prove that every clopen $C \subseteq X$ is of the form $D \cap X$ for a unique clopen $D \subseteq \beta X$.

(ii) Conclude that βX is connected if and only if X is connected.

Remark 8.3.41 The construction of βX given in Section 8.3.3 may be critizised for not being intrinsic, in that it uses the continuous functions $X \to [0, 1]$. We know from Exercise 8.3.11 that the $T_{3.5}$ property, which is necessary and sufficient for the existence of Hausdorff compactifications, has an intrinsic characterization. And Proposition 8.3.34 gives a characterization of βX that is as intrinsic as one can hope. (Corollary 8.3.36 for normal spaces is better in that the completely separated sets are replaced by disjoint closed sets.) It therefore is natural to ask whether βX can be defined without appeal to $[0, 1]$ or other spaces. This can indeed be done in terms of ultrafilters on X whose elements are zero-sets, cf. [90, 230], and the characterization of zero-sets given in Remark 8.2.19 seems intrinsic enough. Again, matters simplify for normal spaces since then zero-sets are precisely the closed G_δ-sets, cf. Exercise 8.2.8. In fact, for normal spaces one can define βX in terms of filters on X consisting of all closed subsets. (This construction is known as the Wallmann compactification ωX, cf. [75].) If X is strongly zero-dimensional, βX has a more economical description in terms of ultrafilters of clopen sets, cf. Section 11.1.11. For discrete X, this reduces to ultrafilters in $P(X)$ as in Section 7.5.4. □

8.3.6 Stone-Čech compactification via ultrafilters

To be written. Cf. [90] or [192].

8.4 ⋆ Applications of the Stone-Čech compactification

8.4.1 Čech-completeness

We already know that complete metric spaces are Baire spaces, i.e. countable intersections of dense open sets are dense. There is another large class of Baire spaces:

Theorem 8.4.1 *(i) Every compact Hausdorff space is a Baire space.*

(ii) Every G_δ-set in a compact Hausdorff space is a Baire space.

Proof. (i) Let $U_i \subseteq X$ be dense and open for every $i \in \mathbb{N}$. Let $\emptyset \neq W \in \tau$. By Lemma 2.7.9, $W \cap U_1 \neq \emptyset$, so that we can choose $x_1 \in W \cap U_1$. By regularity, we can find an open V_1 such that $x_1 \in V_1 \subseteq \overline{V_1} \subseteq W \cap U_1$. (The point x_1 only served as a stepping stone for getting a non-empty V_1.) Since U_2 is dense, $V_1 \cap U_2 \neq \emptyset$. As before, we find an open V_2 such that $\emptyset \neq V_2 \subseteq \overline{V_2} \subseteq V_1 \cap U_2$. Iterating this, we find non-empty open sets V_i such that $V_n \subseteq \overline{V_n} \subseteq V_{n-1} \cap U_n$ for all n. Now all $\overline{V_n}$ are non-empty and $\overline{V_1} \supseteq \overline{V_2} \supseteq \cdots$. The family $\{\overline{V_i}\}$ clearly has the finite intersection property, so that (countable) compactness gives $\bigcap_i \overline{V_i} \neq \emptyset$. By construction, $\bigcap_i \overline{V_i} \subseteq W \cap \bigcap_i U_i$, thus $W \cap \bigcap_i U_i \neq \emptyset$, proving density of $\bigcap_i U_i$.

(ii) Let X be compact Hausdorff and $Y \subseteq X$ G_δ, thus $Y = \bigcap_i W_i$ with $W_i \subseteq X$ open for every $i \in \mathbb{N}$, and let $U_i \subseteq Y$ be dense open for every $i \in \mathbb{N}$. Clearly $X' := \overline{Y}$ is compact Hausdorff and $Y \subseteq X'$ dense. The $W_i' = W_i \cap X' \subseteq X'$ are open and dense in X' since each contains Y. Furthermore, there are open $V_i \subseteq X'$ such that $U_i = Y \cap V_i$. Since Y is dense in X' and U_i is dense in Y, each U_i is dense in X', thus the opens $V_i \supseteq U_i$ are dense in X'. Now $Z = \bigcap_i U_i = \bigcap_i (Y \cap V_i) = \bigcap_i W_i \cap \bigcap_i V_i$ is a countable intersection of dense open sets in X', thus dense in X' by (i). Therefore $\mathrm{Cl}_Y(Z) = \overline{Z} \cap Y = X' \cap Y = Y$, so that Z is dense in Y. This proves that Y is Baire. ■

Corollary 8.4.2 *Locally compact Hausdorff spaces are Baire spaces.*

Proof. A locally compact Hausdorff space X is (by definition of τ_∞) an open subset of its one-point compactification X_∞, which is compact Hausdorff. Now apply Theorem 8.4.1(ii). ■

Remark 8.4.3 We have just proven that every countable intersection of dense open sets in a compact Hausdorff space is dense. The simplest form of Martin's axiom, mentioned in connection with the Souslin property (Section 6.5.2), is equivalent to the statement that every intersection of at most $\mathfrak{c} = \#\mathbb{R}$ dense open sets in a compact Hausdorff space with the Souslin property is dense. (Cf. [151, Theorem II.3.4].) □

The fact that complete metric spaces and G_δ-subsets of compact Hausdorff spaces are Baire can be traced back to a common property:

Definition 8.4.4 *A topological space (X, τ) is called Čech-complete if it is completely regular and $X \equiv \iota_X(X)$ is a G_δ-set in the Stone-Čech compactification βX.*

Proposition 8.4.5 *Locally compact T_2 $\Rightarrow$ Čech-complete $\Rightarrow$ Baire space.*

Proof. (i) A locally compact Hausdorff space X is completely regular. Since βX is Hausdorff, Proposition 7.8.31 implies that $X \subseteq \beta X$ is locally closed. Thus X is an open subset of $\overline{X} = \beta X$. Thus X is open in βX, thus a fortiori Čech-complete.

(ii) If X is Čech-complete then X is G_δ in βX. Since βX is compact Hausdorff, the result follows from Theorem 8.4.1(ii). ■

Proposition 8.4.6 *For a completely regular space X, the following are equivalent:*

(i) X is Čech-complete, i.e. G_δ in βX.

(ii) $f(X) \subseteq Z$ is G_δ whenever $f : X \to Z$ is a dense embedding with Z completely regular.

Proof. For (ii)⇒(i) it suffices to notice that $X \hookrightarrow \beta X$ is a dense embedding into a compact Hausdorff, thus completely regular, space.

(i)⇒(ii) Let $U_i \subseteq \beta X$ open for all $i \in \mathbb{N}$ such that $X = \bigcap_i U_i$, let $\widehat{X}$ be $T_{3.5}$ and $f : X \to Z$ a dense embedding. By functoriality there is an extension $\widehat{f} : \beta X \to \beta Z$. Since $f(X) \subseteq Z$ is dense, $\widehat{f}(\beta X)$ is dense in βZ, but since βX is compact, $\widehat{f}(\beta X)$ is closed. Thus $\widehat{f} : \beta X \to \beta Z$ is surjective. Now, $V_i = \beta Z \backslash \widehat{f}(\beta X \backslash U_i) \subseteq \beta Z$ is open since $\beta X \backslash U_i$ is closed, thus compact, so that $\widehat{f}(\beta X \backslash U_i)$ is compact, thus closed. Now for each i we have (using the first half of Lemma A.1.6)

$$\widehat{f}^{-1}(V_i) = \widehat{f}^{-1}(\beta Z \backslash \widehat{f}(\beta X \backslash U_i)) = \beta X \backslash \widehat{f}^{-1}(\widehat{f}(\beta X \backslash U_i)) \subseteq \beta X \backslash (\beta X \backslash U_i) = U_i. \tag{8.5}$$

Since $f : X \to Z$ is an embedding, the extension $\widehat{f} : \beta X \to \beta Z$ restricts to a homeomorphism $X \to f(X)$. Since also $X \subseteq \beta X$ is dense, Lemma 7.8.8 applies to the situation $X \subseteq \beta X \stackrel{\widehat{f}}{\to} \beta Y$ and gives $f(X) \cap \widehat{f}(\beta X \backslash X) = \emptyset$. In view of $X \subseteq U_i$ we conclude $f(X) \cap \widehat{f}(\beta X \backslash U_i) = \emptyset$. This is equivalent to $f(X) = \widehat{f}(X) \subseteq \beta Z \backslash \widehat{f}(\beta X \backslash U_i) = V_i$ and in turn to $X \subseteq \widehat{f}^{-1}(V_i)$. Combining this with (8.5), we have $X \subseteq \widehat{f}^{-1}(V_i) \subseteq U_i$ for all i. This implies $X \subseteq \bigcap_i \widehat{f}^{-1}(V_i) \subseteq \bigcap_i U_i = X$, thus $X = \bigcap_i \widehat{f}^{-1}(V_i) = \widehat{f}^{-1}\left(\bigcap_i V_i\right)$. Thus $\widehat{f}(X) = f\left(\widehat{f}^{-1}\left(\bigcap_i V_i\right)\right)$. This equals $\bigcap_i V_i$ by the surjectivity of $\widehat{f}$ shown above (and the second half of Lemma A.1.6). Thus $f(X) = \widehat{f}(X) = \bigcap_i V_i = \bigcap_i (V_i \cap Z)$. Since each $V_i \cap Z$ is open in Z, we conclude that $f(X) \subseteq Z$ is G_δ. ■

There also exists a characterization of Čech-complete spaces that does not involve the embedding into βX, but involving countable families of open covers it is not exactly simple:

Theorem 8.4.7 *A completely regular space X is Čech-complete if and only if there exists a countable family $\{\mathcal{U}_i\}_{i \in \mathbb{N}}$ of open covers of X such that $\bigcap \mathcal{F} \neq \emptyset$ holds for every family $\mathcal{F}$ of closed subsets of X having the finite intersection property and such that for every $i \in \mathbb{N}$ there is an $F \in \mathcal{F}$ and a $U \in \mathcal{U}_i$ such that $F \subseteq U$.*

Proof. See [75, Theorem 3.9.2]. ■

8.4.2 Characterization of completely metrizable spaces

Theorem 8.4.8 *For a metric space (X, d) the following are equivalent:*

(i) (X, τ_d) is Čech-complete, i.e. $X \subseteq \beta X$ is G_δ.

(ii) $\iota(X) \subseteq Z$ is G_δ whenever $\iota : X \to Z$ is a dense embedding with Z completely regular.

(iii) $\iota(X) \subseteq X'$ is G_δ for every isometry $\iota : (X, d) \to (X', d')$.

(iv) (X, d) is completely metrizable.

Proof. We recall first that metric spaces are normal, thus completely regular.

(i)⇒(ii) This was the non-trivial implication of Proposition 8.4.6.

(ii)⇒(iii) If $\iota : (X, d) \to (X', d')$ is an isometry then $\iota(X) \subseteq \overline{\iota(X)}$ is a dense embedding into a completely regular space, thus G_δ by (ii). As a closed set in a metric space, $\overline{\iota(X)} \subseteq X'$ is G_δ (Exercise 3.3.12). Thus $\iota(X) \subseteq X'$ is G_δ (Exercise 3.3.11(i)).

(iii)⇔(iv) is part of Theorem 3.4.20.

(iii)⇒(i) We want to prove that τ_d is Čech-complete. Since τ_d is not affected if we replace d by an equivalent metric, by Exercise 2.2.14 we may assume d to be bounded by some constant

C. Now Corollary 8.3.31 with $Y = [0, C]$ shows that for each $t \in X$ the continuous function $f_t : X \to [0, C]$, $x \mapsto d(t, x)$ extends to a continuous function $\widehat{f}_t : \beta X \to [0, C]$. Thus for all $x, y \in \beta X$ we can define[7]

$$\widetilde{d}(x, y) = \sup_{t \in X} |\widehat{f}_t(x) - \widehat{f}_t(y)| \leq 2C < \infty.$$

It is obvious that $\widetilde{d}(x, y) = \widetilde{d}(y, x)$ and $\widetilde{d}(x, x) = 0$. Furthermore,

$$\begin{aligned} \widetilde{d}(x, z) &= \sup_{t \in X} |\widehat{f}_t(x) - \widehat{f}_t(z)| \leq \sup_{t \in X} \Big(|\widehat{f}_t(x) - \widehat{f}_t(y)| + |\widehat{f}_t(y) - \widehat{f}_t(z)| \Big) \\ &\leq \sup_{t \in X} |\widehat{f}_t(x) - \widehat{f}_t(y)| + \sup_{t \in X} |\widehat{f}_t(y) - \widehat{f}_t(z)| = \widetilde{d}(x, y) + \widetilde{d}(y, z). \end{aligned}$$

Thus $\widetilde{d}$ is a pseudometric on βX, and (2.3) gives $\widetilde{d} \restriction X = d$.

Let $a \in X$, $b \in \beta X$ be such that $\widetilde{d}(a, b) = 0$. By definition of $\widetilde{d}$, this implies $\widehat{f}_t(b) = \widehat{f}_t(a) = d(t, a)$ for all $t \in X$, in particular for $t = a$. Since $X \subseteq \beta X$ is dense, we can choose a net $\{b'_\iota\}$ in X converging to b. By continuity of $b \mapsto \widehat{f}_a(b)$, we have $\lim d(b_\iota, a) = \lim \widehat{f}_a(b_\iota) = \widehat{f}_a(b) = d(a, a) = 0$. Thus $\lim d(b_\iota, a) = 0$, so that $b_\iota \to a$. Since βX is Hausdorff, we have $b = a$.

Let $\sim$ be the equivalence relation on βX defined by $a \sim b \Leftrightarrow \widetilde{d}(a, b) = 0$. If $p : \beta X \to \beta X/\!\sim =: K$ is the quotient map, we obtain a (true) metric $\widetilde{d}'$ on K. (Cf. Exercise 2.1.7.)

By what we have shown above, if $a \sim b$ for $a \in X$, $b \in \beta X$ then $a = b$. In particular, $\sim$ is trivial on X (as also follows directly from $\widetilde{d} \restriction X = d$). Thus p is injective on $X \subseteq \beta X$, and we have an isometric embedding $(X, d) \hookrightarrow (K, \widetilde{d}')$. And since $\sim$ identifies no point of X with an infinite point (i.e. of $\beta X \backslash X$), we have $p^{-1}(X) = X$.

Since we assume (iii), $X \subseteq K$ is G_δ. Now Exercise 3.3.11(iv) gives that $p^{-1}(X) \subseteq \beta X$ is G_δ. Combining this with $p^{-1}(X) = X$, we see that $X \subseteq \beta X$ is G_δ, proving (i). ■

Corollary 8.4.9 *A topological space (X, τ) is completely metrizable if and only if it is metrizable and Čech-complete.*

In order to have a really satisfactory characterization of completely metrizable spaces, we still need to characterize the metrizable spaces. (So far we only have the sufficient condition in Theorem 8.2.33.) One of many such characterizations will be provided in Theorem 8.5.31. But we already have:

Corollary 8.4.10 *For a topological space X, the following are equivalent:*

(i) X is completely metrizable and separable.

(ii) X is second countable, T_3 and Čech-complete.

Such spaces are called Polish spaces.

Proof. Follows from the above and Corollary 8.2.35. ■

[7] In [237] this is done with inf instead of sup. This is a mistake since then $\widetilde{d} \restriction X = d$ does not hold. A related problem is that Willard claims $\widetilde{d}$ to be continuous, which is not obvious.

8.4.3 From Hausdorff compactifications to Proximities

The criterion for the equivalence of Hausdorff compactifications provided by Proposition 7.8.9 is best understood in terms of the notion of proximity:

Definition 8.4.11 *A proximity on a set X is a binary relation δ on the powerset $P(X)$ satisfying*

(i) $A\delta B \ \Leftrightarrow\ B\delta A$.

(ii) $A\delta B \ \Rightarrow\ A \neq \emptyset \neq B$.

(iii) $A \cap B \neq \emptyset \ \Rightarrow\ A\delta B$.

(iv) $(A \cup B)\delta C \ \Leftrightarrow\ A\delta C \vee B\delta C$.

(v) $A \not\delta B \ \Rightarrow\ \exists E \subseteq X : A \not\delta E \wedge (X \backslash E) \not\delta B$.

A proximity δ is called separated if also

(vi) $x\delta y \ \Rightarrow\ x = y$.

(Here $x\delta y$ is short notation for $\{x\}\delta\{y\}$.)

Example 8.4.12 Let (X, d) be a metric space. For $A, B \subseteq$ define

$$A\delta B \ \Leftrightarrow\ A \neq \emptyset \neq B \ \wedge\ \mathrm{dist}(A, B) = 0.$$

One easily checks that δ is a separated proximity, called the metric proximity. □

Definition 8.4.13 *A pair (X, δ), where X is a set and δ a proximity on it, is called a proximity space.*

Exercise 8.4.14 Let $X, \delta)$ be a proximity space. Prove that the family $\{C \subseteq X \mid x\delta C \Rightarrow x \in C\}$ satisfies the axioms for closed sets and therefore defines a topology τ_δ on X.

A topology τ and a proximity δ on a set X are called compatible if $\tau = \tau_\delta$. (And every uniform structure $\mathcal{U}$ gives rise to a proximity $\delta_\mathcal{U}$ such that $\tau_\mathcal{U} = \tau_{\delta_\mathcal{U}}$. Thus proximity spaces are intermediate between uniform spaces and topological spaces.) There is an extensive theory of proximity spaces, cf. [183], but here we are only interested in the application to compactifications.

Lemma 8.4.15 *Let $\widehat{X}$ be a normal space and $X \subseteq \widehat{X}$. For $A, B \subseteq X$ define $A\delta B :\Leftrightarrow \overline{A} \cap \overline{B} \neq \emptyset$, where the closures are in $\widehat{X}$. Then δ is a separated proximity on X.*

Proof. (i) is obvious. (ii) follows from $\overline{\emptyset} = \emptyset$, (iii) from $\overline{A} \supseteq A$, and (iv) from $\overline{A \cup B} = \overline{A} \cup \overline{B}$. If $A \not\delta B$ then $\overline{A} \cap \overline{B} = \emptyset$. Then by normality of $\widehat{X}$ there are disjoint open sets $U, V \subseteq \widehat{X}$ such that $\overline{A} \subseteq U$, $\overline{B} \subseteq V$. By Lemma 2.6.3, we have $U \cap \overline{V} = \emptyset$ and thus $\overline{A} \cap \overline{V} = \emptyset$. Openness of V gives $\overline{\widehat{X} \backslash V} \cap \overline{B} = (\widehat{X} \backslash V) \cap \overline{B} = \emptyset$. Putting $E = V \cap X$, we have $A \not\delta E$ and $(X \backslash E) \not\delta B$, thus (v) holds. Separatedness is obvious since our definition of normality includes T_1, i.e. closedness of singletons. ■

In particular, taking $Y = X$ in Lemma 8.4.15 gives us a separated proximity δ_{n} on every normal space X. (This may fail for non-normal spaces since normality was used in the proof of (v).)

Corollary 8.4.16 *Let X be completely regular. Then:*

(i) *If $\widehat{X}$ is a Hausdorff compactification of X then $A\delta_{\widehat{X}}B \Leftrightarrow \overline{A}\cap\overline{B}\neq\emptyset$ defines a separated proximity $\delta_{\widehat{X}}$ on X.*

(ii) *Two Hausdorff compactifications $\widehat{X}_1, \widehat{X}_2$ of X are isomorphic in $\mathcal{C}(X)$ if and only if $\delta_{\widehat{X}_1} = \delta_{\widehat{X}_2}$.*

Proof. (i) $\widehat{X}$ is normal. Now apply Lemma 8.4.15 to $X \subseteq \widehat{X}$.

(ii) follows from Proposition 7.8.9. (In the latter, only disjoint closed sets A, B appear. But for non-disjoint A, B we always have $A\delta B$. And the closedness may be dropped since $\overline{\iota(A)} = \overline{\iota(\overline{A})}$ by continuity of ι.) ■

We now exhibit the proximities corresponding to the Stone-Čech and Alexandrov compactifications.

Lemma 8.4.17 *Let (X,τ) be completely regular (but not necessarily normal). Then*

(i) *Putting $A\delta B$ if and only if A and B are not completely separated defines a proximity δ_{cr} on X. If X is normal, then $\delta_{\mathrm{cr}} = \delta_{\mathrm{n}}$.*

(ii) *δ_{cr} coincides with the proximity $\delta_{\beta X}$ on X arising from Stone-Čech compactification βX.*

Proof. (i) That δ_{cr} satisfies axioms (i) and (iii) in Definition 8.4.11 is obvious. If A or B is empty, a complete separation exists, thus $A \not\delta_{\mathrm{cr}} B$. This gives axiom (ii). $(A\cup B)\not\delta C$ means that C is completely separated from $A\cup B$, thus also from A and B separately. Thus $(A\cup B)\not\delta C$ implies $A\not\delta C$ and $B\not\delta C$, which is the contraposition of implication $\Leftarrow$ in axiom (iv). Assume C is completely separated from A and from B. Then there are $f, g \in C(X,[0,1])$ such that $f\restriction C = g\restriction C = 1$ and $f\restriction A = g\restriction B = 0$. Then $f\cdot g\restriction(A\cup B) = 0$ and $f\cdot g\restriction C = 1$. Thus C is completely separated from $A\cup B$, proving axiom (iv). If $A\not\delta B$ then there is a function $f\in C(X,[0,1])$ with $f\restriction A = 0, f\restriction B = 1$. Define $E = f^{-1}([1/2,1])$. It is then clear that A, E are completely separated, thus $A\not\delta E$, and that B is completely separated from $X\backslash E = f^{-1}([0,1/2))$, thus $B\not\delta(X\backslash E)$. This proves axiom (v). Separatedness (vi) follows from the fact that in a completely regular space, continuous functions separate the points.

If X is normal then $\overline{A}\cap\overline{B}=\emptyset$ and complete separation of A and B are equivalent (Corollary 8.2.4), thus $\delta_{\mathrm{cr}} = \delta_{\mathrm{n}}$.

(ii) By Proposition 8.3.34(i)-(ii), for $\iota : X \to \beta X$ we have that complete separation of A and B is equivalent to $\overline{\iota(A)}\cap\overline{\iota(B)}=\emptyset$. Now for $A, B\subseteq X$ we have

$$A\delta_{\mathrm{cr}}B \overset{\text{def.}}{\Leftrightarrow} A, B \text{ not completely separated } \Leftrightarrow \overline{\iota(A)}\cap\overline{\iota(B)}\neq\emptyset \overset{\text{def.}}{\Leftrightarrow} A\delta_{\beta X}B,$$

proving $\delta_{\beta X} = \delta_{\mathrm{cr}}$. ■

Remark 8.4.18 1. If X is locally compact Hausdorff, the one-point compactification X_∞ is a Hausdorff compactification of X. Now Exercise 7.8.18(iii) immediately shows that the corresponding separated proximity δ_{X_∞} on X is given by:

$$A\delta_\infty B \quad\Leftrightarrow\quad \overline{A}\cap\overline{B}\neq\emptyset \ \vee\ \overline{A},\overline{B} \text{ are both non}-\text{compact}.$$

(If X is not locally compact Hausdorff, δ_∞ will still satisfy axioms (i)-(iv) (whose proof in Lemma 8.4.15 required no assumption on $\widehat{X}$), but not necessarily (v). It will be separated if and only if $x\neq y \Rightarrow \overline{\{x\}}\cap\overline{\{y\}}=\emptyset$, which is intermediate between T_0 and T_1.)

2. Comparing the proximity δ_∞ to δ_{cr} from Lemma 8.4.17, we obtain another proof of Corollary 8.3.35. □

8.4.4 From proximities to Hausdorff compactifications

Remarkably, one has the following converse of Lemma 8.4.15:

Theorem 8.4.19 (Smirnov) [8] *Let (X,τ) be completely regular. Then for every separated proximity δ on X there exists a Hausdorff compactification $((\widehat{X},\widehat{\tau}),\iota)$ of X such that the corresponding proximity on X equals δ.*

Corollary 8.4.20 *For every completely regular space X there is a bijection between separated proximities on X and isomorphism classes of Hausdorff compactifications of X.*

Remark 8.4.21 The theorem can be proven from scratch, cf. Smirnov's papers or the book [183]. Alternatively, one can use compactification w.r.t. a uniformity (due to P. Samuel), [126] or [75, Section 8.4]. We will follow the more economical (but apparently little known) route [167] of constructing the compactification $\widehat{X}$ corresponding to δ as a quotient of βX. □

Lemma 8.4.22 *Let (X,τ) be completely regular and δ a separated proximity on X. Consider X as a subspace of βX. Define a binary relation $\sim$ on βX by*

$$p \sim q \quad \Leftrightarrow \quad \text{there do not exist } A,B \subseteq X \text{ such that } p \in \overline{A},\ q \in \overline{B} \text{ and } A \not\delta B.$$

(The closures are in βX.) Then $\sim$ is a closed equivalence relation.

Proof.

■

Let $(X,\tau),\delta$ and the equivalence relation as in the lemma. Define $\widehat{X} = \beta X/\sim$, and let $p : \beta X \to \widehat{X}$ be the quotient map. βX is compact Hausdorff, thus $\widehat{X}$ is compact, and since $\sim$ is closed, $\widehat{X}$ is Hausdorff by Corollary 8.1.18. We have an obvious map $\iota : X \to \widehat{X}$ given as composition $X \hookrightarrow \beta X \xrightarrow{p} \widehat{X}$. Since $X \subseteq \beta X$ is dense, $\iota(X) \subseteq \widehat{X}$ is dense. Now the proof of Theorem 8.4.19 is completed by the following:

Proposition 8.4.23 *Let $(X,\tau),\delta$ be given as in the lemma and $\sim,\widehat{X},p$ as defined above. Then*

(i) The restriction of $p : \beta X \to \widehat{X}$ to $X \subseteq \beta X$ is injective.

(ii) For $A,B \subseteq X$, we have $\mathrm{cl}(\iota(A)) \cap \mathrm{cl}(\iota(B)) \neq \emptyset \ \Leftrightarrow\ A\delta B$, where cl denotes closure in $\widehat{X}$.

Thus $(\widehat{X},\iota)$ is a Hausdorff compactification of X, and the corresponding proximity equals δ.

Proof.

■

[8] Yuri Mikhailovich Smirnov (1921-2007). Russian topologist. Also known as one of the discoverers of the Bing-Nagata-Smirnov metrization theorem.

8.4.5 Freudenthal compactification of rimcompact spaces

As an easy application of the formalism of the preceding section, we discuss the 'end compactification' due to Freudenthal[9]. There are very few expositions of the latter in the textbook literature (the only ones known to the author are [16, p. 57-70] and [126, p. 109-116], which however makes heavy use of uniform spaces), possibly because of the considerable effort necessary for a direct approach.

Definition 8.4.24 *A space (X, τ) is rim-compact if τ admits a base consisting of open sets with compact boundaries.*

Lemma 8.4.25 *(i) Every locally compact Hausdorff space is rim-compact.*

(ii) Every zero-dimensional space is rim-compact. (Cf. Definition 11.1.15: The topology has a base consisting of clopen sets.)

Proof. (i) By Lemma 7.8.24, τ has a base of open sets with compact closures. Now the claim follows from, since compactness of $\overline{U}$ implies compactness of ∂U.

(ii) Follows immediately from the fact that a set is clopen if and only if its boundary is empty. ■

8.5 Paracompactness and some of its uses

The notion of paracompactness has many applications, some of which do not even involve paracompactness in their statements, like Theorem 8.1.41 or Corollary 8.5.21. It bears on the problem of normality of product spaces, allowing to prove Theorem 8.1.41, and thus is responsible for one direction of Theorem 8.1.45. Metric spaces are paracompact, and notions related to paracompactness appear in many characterizations of metrizable spaces. (This subsection was not part of Section 7 on compactness and its generalizations because it strongly relies on Sections 8.1-8.2 on separation axioms, both technically and for motivation.)

8.5.1 The basic facts

In Section 8.2.5 we have encountered the notion of local finiteness of a family of sets, in particular of a cover, in connection with the existence of partitions of unity. This raises the question which spaces allow to find locally finite subcovers for all open covers. The following shows that this condition is too strong, since it is equivalent to compactness:

Lemma 8.5.1 *Let X be a topological space.*

(i) If X is non-compact then for every $x \in X$ there is an open cover $\mathcal{U}$ of X such that no subcover $\mathcal{V} \subseteq \mathcal{U}$ is point-finite at x.

(ii) If every open cover of X has a point-finite subcover then X is compact.

(iii) If every open cover of X has a locally finite subcover then X is compact.

Proof. (i)⇒(ii)⇒(iii) are obvious.

(i) [Due to Mike F on stackexchange] Assume X is non-compact and $x \in X$. Assume that $X \backslash U$ is compact for every open neighborhood U of x. Let $\mathcal{U}$ be an open cover. Then there is a $U \in \mathcal{U}$ with $x \in U$. By assumption $X \backslash U$ is compact, thus it is covered by a finite subfamily $\mathcal{V} \subseteq \mathcal{U}$. Now

[9] Hans Freudenthal (1905-1990). German-Dutch mathematician.

$\mathcal{V} \cup \{U\} \subseteq \mathcal{U}$ is a finite subcover of X. Since $\mathcal{U}$ was arbitrary, we have a contradiction with the assumption that X is non-compact. We conclude that there is an open neighborhood of x such that $X \backslash U$ is non-compact. The latter fact means that there is an open cover $\mathcal{V} \subseteq \tau_{X \backslash U}$ of $X \backslash U$ that admits no finite subcover. Choosing $\mathcal{W}_0 \subseteq \tau$ such that $\mathcal{V} = \{W \cap (X \backslash U) \mid W \in \mathcal{W}_0\}$, $\mathcal{W}$ has no finite subfamily whose union contains $X \backslash U$. Now $\mathcal{W} = \{W \cup U \mid W \in \mathcal{W}_0\}$ is an open cover of X. Since x is contained in every element of $\mathcal{W}$, every subcover that is point-finite at x is finite. But by construction, no finite subfamily of $\mathcal{W}$ can cover X. This proves the claim. ■

We therefore need to work with a generalization of subcovers, like the notion of shrinkings that we have already met. It is advantageous to generalize that even further:

Definition 8.5.2 *A refinement of a cover $\mathcal{U}$ of X is a cover $\mathcal{V}$ such that for every $V \in \mathcal{V}$ there is a $U \in \mathcal{U}$ such that $V \subseteq U$. We call $\mathcal{V}$ open (closed) if every $V \in \mathcal{V}$ is open (closed).*

Remark 8.5.3 The notion of refinement of a cover generalizes that of a shrinking (which in turn generalizes subcovers): Instead of having one subset $V_U \subseteq U$ of each $U \in \mathcal{U}$, we only require that every $V \in \mathcal{V}$ is contained in some $U \in \mathcal{U}$. (In the notation $\mathcal{U} = \{U_i\}_{i \in I}, \mathcal{V} = \{V_j\}_{j \in J}$ this means that instead of $V_i \subseteq U_i \ \forall i$ we have a map $\alpha : J \to I$ such that $V_j \subseteq U_{\alpha(j)} \ \forall j \in J$.) □

Lemma 8.5.4 *If an open cover admits a finite refinement then it also has a finite subcover.*

Proof. If $\mathcal{V}$ is a finite refinement of $\mathcal{U}$ (open or not), for each $V \in \mathcal{V}$ choose a $U_V \in \mathcal{U}$ such that $V \subseteq U_V$. Now $\{U_V\}_{V \in \mathcal{V}} \subseteq \mathcal{U}$ is a finite subcover since $\bigcup_V U_V \supseteq \mathcal{V} = X$. ■

Thus the existence of finite shrinkings or refinements (open or not) of all open covers is again equivalent to compactness. But the following statements will turn out to be weaker:

Lemma 8.5.5 *For a topological space X, the following are equivalent:*

(i) Every open cover of X has a locally finite open shrinking.

(ii) Every open cover of X has a locally finite open refinement.

Proof. (i)⇒(ii) Again this is obvious since a shrinking in particular is a refinement.

(ii)⇒(i) Let $\mathcal{U} = \{U_i\}_{i \in I}$ be an open cover of X. Choose a locally finite open refinement $\mathcal{W} = \{W_j\}_{j \in J}$ of $\mathcal{U}$ and a map $f : J \to I$ such that $W_j \subseteq U_{f(j)} \ \forall j \in J$. Now for $i \in I$ define $V_i = \bigcup_{j \in f^{-1}(i)} W_j$. This is open, and in view of $\bigcup_{i \in I} V_i = \bigcup_{i \in I} \bigcup_{j \in f^{-1}(i)} W_j = \bigcup_{j \in J} W_j = X$, $\mathcal{V}$ covers X. We have $V_i \subseteq U_i$ by the very definition of f, thus $\mathcal{V}$ is an open shrinking of $\mathcal{U}$. By local finiteness of $\mathcal{W}$, every $x \in X$ has an open neighborhood N such that $K = \{j \in J \mid N \cap W_j \neq \emptyset\}$ is finite. Now $N \cap V_i \neq \emptyset$ holds if and only if $N \cap W_j \neq \emptyset$ for some $j \in f^{-1}(i)$, which is equivalent to $i \in f(K)$. Since $f(K)$ is finite, $\mathcal{V}$ is locally finite. ■

Definition 8.5.6 *A topological space X is called paracompact[10] if every open cover of X has a locally finite open refinement.*

[10]Paracompactness was invented/discovered in 1944 by Jean Dieudonné (1906-1992). He was incredibly productive: Besides writing > 15 books on his own and Elements de Géometrie Algébrique I (joint with Grothendieck), he was responsible for the final version of almost everything published by the Bourbaki group.

Remark 8.5.7 1. Lemma 8.5.5 shows that paracompactness could be defined in terms of shrinkings, dispensing with refinements. But the added generality of refinements is often useful.

2. It is clear that every compact space is paracompact. The converse is not true. For example, we will see that every metrizable space is paracompact.

3. As in the case of compactness, some authors include the Hausdorff property in the definition of paracompactness, but we do not do this. □

Since paracompactness is weaker than compactness, the following improves Proposition 8.1.8:

Proposition 8.5.8 *Every paracompact Hausdorff space is normal.*

Proof. Let $C \subseteq X$ be closed and $x \in X \backslash C$. Since X is Hausdorff, for every $y \in C$ there are disjoint open $U_y \ni x, V_y \ni y$. This implies $x \notin \overline{V_y}$. Now $\mathcal{V} = \{V_y \mid y \in C\} \cup \{X \backslash C\}$ is an open cover of X. Pick a locally finite refinement $\mathcal{W}$ of $\mathcal{V}$ and define $V = \bigcup\{W \in \mathcal{W} \mid W \cap C \neq \emptyset\}$. Then V is open and contains C. Since $\{W \in \mathcal{W} \mid W \cap C \neq \emptyset\}$ is locally finite, Exercise 8.2.51(ii) gives $\overline{V} = \bigcup\{\overline{W} \mid W \in \mathcal{W},\ W \cap C \neq \emptyset\}$. $\mathcal{W}$ being a refinement of $\mathcal{V}$, every W with $W \cap C \neq \emptyset$ is contained in some V_y (since it clearly is not contained in $X \backslash C$), thus $\overline{W} \subseteq \overline{V_y}$. Thus $\overline{V} = \bigcup\{\overline{W} \mid W \in \mathcal{W},\ W \cap C \neq \emptyset\} \subseteq \bigcup_{y \in C} \overline{V_y}$, which implies $x \notin \overline{V}$. Thus $U = X \backslash \overline{V}$ is an open neighborhood of x disjoint from V, so that X is T_3.

In order to prove that X is T_4, let C, D be disjoint closed subsets. By the T_3-property just proven, every for every $y \in C$ we can find an open $V_y \ni y$ such that $\overline{V_y} \cap D = \emptyset$. Repeating the above argument, we can construct an open $V \supseteq C$ such that $\overline{V} \cap D = \emptyset$, proving normality. ■

The following should be compared with Lemma 8.2.28 to the effect that pseudocompact normal spaces are countably compact:

Corollary 8.5.9 *A pseudocompact paracompact Hausdorff space is compact.*

Proof. By Proposition 8.5.8, a paracompact Hausdorff space is normal, thus completely regular. If $\mathcal{U}$ is an open cover, paracompactness provides a locally finite open refinement $\mathcal{V}$. By pseudocompactness, complete regularity and Proposition 8.2.52, the locally finite cover $\mathcal{V}$ is finite. Now Lemma 8.5.4 shows that $\mathcal{U}$ has a finite subcover. ■

Remark 8.5.10 1. A topological space X is called metacompact[11] if every open cover of X has a point-finite open refinement. Obviously paracompact ⇒ metacompact. Since point-finiteness of a cover is not sufficient for the existence of a subordinated partition of unity, metacompactness is much less important than paracompactness. Nevertheless, it has some interesting applications.

2. Example: A (weakly) countably compact and metacompact space is compact.
Proof. Let $\mathcal{U}$ be an open cover. By metacompactness, it has a point-finite open refinement $\mathcal{V}$. By Exercise 8.1.54, there is an irreducible subcover $\mathcal{W} \subseteq \mathcal{V}$. For every $W \in \mathcal{W}$, we have $\bigcup(\mathcal{W} \backslash \{W\}) \subsetneq X$, thus we can choose a point $x_W \in X \backslash (\bigcup(\mathcal{W} \backslash \{W\}))$. Then $\{x_W \mid W \in \mathcal{W}\}$ is a closed discrete subset of X, which must be finite by definition of weak countable compactness (implied by countable compactness, cf. Exercise 7.7.8). Thus $\mathcal{W}$ is a finite refinement of $\mathcal{U}$, and by Lemma 8.5.4 $\mathcal{U}$ has a finite subcover.

3. To compare the above with Corollary 8.5.9, recall that paracompact ⇒ metacompact and countably compact ⇒ pseudocompact. One can actually show that already complete regularity, metacompactness and pseudocompactness imply compactness, cf. [233]. □

Now we can answer the question posed in Remark 8.2.50:

[11] Metacompactness was introduced in 1950 by Richard F. Arens (1919-2000) and James Dugundji (1919-1985).

Theorem 8.5.11 *For a Hausdorff space X, the following are equivalent:*

(i) X is paracompact.

(ii) For every open cover $\mathcal{U}$ of X there is a locally finite partition of unity subordinate to $\mathcal{U}$.

(iii) For every open cover $\mathcal{U}$ of X there is a family $\mathcal{F} \subseteq C(X,[0,1])$ such that $\{\mathrm{supp}(f) \mid f \in \mathcal{F}\}$ is a locally finite (closed) refinement of $\mathcal{U}$ and $\sum_{f\in\mathcal{F}} f(x) = 1\ \forall x \in X$.

Proof. (i)⇒(ii) Assume X is paracompact, and let $\mathcal{U}$ be an open cover. By Lemma 8.5.5 we can find a locally finite open shrinking $\mathcal{V}$ of $\mathcal{U}$. Since X is normal by Proposition 8.5.8, Theorem 8.2.49 provides a partition of unity $\mathcal{F}$ subordinate to $\mathcal{V}$. It is clear that $\mathcal{F}$ is locally finite and subordinate to $\mathcal{U}$.

(ii)⇒(iii): This is trivial.

(iii)⇒(i): Let $\mathcal{U}$ be an open cover of X. Pick a family $\mathcal{F}$ with the properties under (iii). Since the $f \in \mathcal{F}$ are continuous and satisfy $\sum_{f\in\mathcal{F}} f(x) = 1\ \forall x$, we see that $\mathcal{W} = \{X\backslash f^{-1}(0) \mid f \in \mathcal{F}\}$ is an open cover of X. It obviously is a shrinking of $\mathcal{V} = \{\mathrm{supp}(f) \mid f \in \mathcal{F}\}$, which by assumption is a locally finite (closed) refinement of $\mathcal{U}$. Thus $\mathcal{W}$ is a locally finite open refinement of $\mathcal{U}$, proving paracompactness. ■

We have met three classes of spaces that are automatically normal: (i) Compact T_2-spaces, (ii) Lindelöf T_3-spaces and (iii) metrizable spaces. For the first class (which is contained in the second) paracompactness is trivial. For the other two classes, we have the following strengthenings of Proposition 8.1.15(iii) and Lemma 8.1.11:

Proposition 8.5.12 *Lindelöf T_3-spaces are paracompact.*

Proof. Let $\mathcal{U} = \{U_i\}_{i\in I}$ be an open cover. For each $x \in X$ choose an $i_x \in I$ such that $x \in U_{i_x}$. Since X is T_3, there are open V_x, W_x such that $x \in V_x \subseteq \overline{V_x} \subseteq W_x \subseteq \overline{W_x} \subseteq U_x$. (Begin with W_x or use that X is T_4 by Proposition 8.1.15.) Now $\mathcal{V} = \{V_x\}_{x\in X}$ is an open cover of X, which by the Lindelöf property has a countable subcover $\{V_{x_1}, V_{x_2}, \ldots\}$. Take $T_1 = W_{x_1}$ and, for $n \geq 2$ define inductively $T_n = W_{x_n} \cap X\backslash(\overline{V}_{x_1} \cup \cdots \cup \overline{V}_{x_{n-1}})$, which is open. Paracompactness will follow once we prove that $\mathcal{T} = \{T_n\}_{n\in\mathbb{N}}$ is a locally finite refinement of $\mathcal{U}$. By definition, $T_n \subseteq W_{x_n} \subseteq U_{x_n} \in \mathcal{U}$, thus $\mathcal{T}$ refines $\mathcal{U}$. By construction, $T_n \supseteq W_{x_n}\backslash(W_{x_1} \cup \cdots \cup W_{x_{n-1}})$, thus $\bigcup_{k=1}^n T_k \supseteq \bigcup_{k=1}^n W_{x_k}$. Since $\{V_{x_i}\}$ covers X, so does $\{W_{x_i}\}$ and therefore $\{T_n\}$. Every x is contained in some V_{x_k}, and by construction $V_{x_k} \cap T_l = \emptyset$ whenever $l \geq k$. Thus V_k is an open neighborhood of x meeting only finitely many T_k, and $\{T_k\}$ is locally finite. ■

Theorem 8.5.13 *Metrizable spaces are paracompact.*

Proof. Let $\mathcal{U} = \{U_i\}_{i\in I}$ be an open cover of a metric space (X,d). By the Well-Ordering Principle (which is equivalent to the axiom of choice by Theorem A.3.26) there is a well-ordering $\leq$ on I. Recall that this means that every $J \subseteq I$ has a $\leq$-smallest element $\min(J)$. In particular, for every x there is a unique $i \in I$, namely $i = \min(\{j \in I \mid x \in U_j\})$, such that $x \in U_i \backslash \bigcup_{j<i} U_j$.

For $i \in I, n \in \mathbb{N}$ define, inductively over n,

$$V_{i,n} = \bigcup_{x\in X_{i,n}} B(x, 2^{-n}), \quad \text{where} \quad X_{i,n} = \{x \in X \mid B(x, 3\cdot 2^{-n}) \subseteq U_i, \tag{8.6}$$

$$x \notin \bigcup_{j<i} U_j \ \cup \bigcup_{j\in I, k<n} V_{j,k}\}. \tag{8.7}$$

We claim that $\mathcal{V} = \{V_{i,n}\}_{i\in I, n\in\mathbb{N}}$ is a locally finite open refinement of $\mathcal{U}$, proving paracompactness. Openness of the $V_{i,n}$ is obvious.

By (8.6), $B(x, 2^{-n}) \subseteq B(x, 3\cdot 2^{-n}) \subseteq U_i$ for every ball contributing to $V_{i,n}$, thus $V_{i,n} \subseteq U_i$, and $\mathcal{V}$ refines $\mathcal{U}$.

For $x \in X$, put $i = \min\{j \in I \mid x \in U_j\}$ and choose $n \in \mathbb{N}$ such that (8.6) holds. Then by (8.7), either $x \in V_{j,k}$ for some $j \in I$ and $k < n$ or we have $x \in X_{i,n} \subseteq V_{i,n}$. Thus $\mathcal{V}$ covers X.

For $x \in X$, put $i = \min\{j \in I \mid x \in \bigcup_{n\in\mathbb{N}} V_{j,n}\}$. Then we can choose $n, k \in \mathbb{N}$ such that $B(x, 2^{-k}) \subseteq V_{i,n}$. We will prove:

(i) If $l \geq n+k$ then $B(x, 2^{-n-k})$ intersects no $V_{j,l}$.

(ii) If $l < n+k$ then $B(x, 2^{-n-k})$ intersects $V_{j,l}$ for at most one $j \in I$.

These claims imply that the open neighborhood $B(x, 2^{-n-k})$ of x can meet at most $n+k-1$ elements of $\mathcal{V}$, thus $\mathcal{V}$ is locally finite.

Proof of (i): Let $y \in X_{j,l}$. In view of $l > n$, (8.7) implies that $y \notin V_{i,n}$. Together with $B(x, 2^{-k}) \subseteq V_{i,n}$, this implies $d(x,y) \geq 2^{-k}$. We have $l \geq k+1$ (by assumption) and $n+k \geq k+1$ (trivially). Now $z \in B(x, 2^{-n-k}) \cap B(y, 2^{-l})$ would imply

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

which is a contradiction. Thus $B(x, 2^{-n-k}) \cap B(y, 2^{-l}) = \emptyset$, i.e. $B(x, 2^{-n-k})$ is disjoint from the balls $\{B(y, 2^{-l}), y \in X_{j,l}\}$ whose union is $V_{j,l}$.

Proof of (ii): Let $x \in V_{i,l}, y \in V_{j,l}$ where $i < j$. By definition of the V's, there are x', y' such that $x \in B(x', 2^{-l}) \subseteq V_{i,l}$, $y \in B(y', 2^{-l}) \subseteq V_{j,l}$. By (8.6), $B(x', 3\cdot 2^{-l}) \subseteq U_i$, but by (8.7) $y' \notin U_i$. This implies $d(x', y') \geq 3 \cdot 2^{-l}$, and with the triangle inequality we have

$$3\cdot 2^{-l} \leq d(x', y') \leq d(x', x) + d(x,y) + d(y, y') < d(x,y) + 2^{-l} + 2^{-l},$$

thus $d(x,y) > 2^{-l}$. This obviously implies $V_{i,l} \cap V_{j,l} = \emptyset$ whenever $i \neq j$. It also implies that every ball of radius 2^{-l-1} intersects $V_{i,l}$ for at most one i, and in view of $l < n+k$, i.e. $n+k \geq l+1$, this conclusion a fortiori holds for every $B(x, 2^{-n-k})$. ■

Remark 8.5.14 1. The original proof of this result by A. H. Stone (1948) was quite complicated. The above proof was given in 1969 by M. E. Rudin [199].

2. The above proof still applies to (X, τ_d) when d is only a pseudometric, since axiom (iii) of Definition 2.1.1 (equivalent to τ_d being T_2) was used nowhere.

3. The cover $\mathcal{V} = \{V_{i,n}\}_{i\in I, n\in\mathbb{N}}$ can be written as $\mathcal{V} = \bigcup_{n\in\mathbb{N}} \mathcal{V}_n$, where $\mathcal{V}_n = \{V_{i,n}\}_{i\in I}$. As noted in the proof of Theorem 8.5.13, the elements of $\mathcal{V}_n$ are mutually disjoint. Thus the family $\mathcal{V}_n$ is discrete for each n, and as a union of countably many discrete families, $\mathcal{V}$ is σ-discrete.

Thus we have: If the topology of a space arises from a pseudometric then every open cover has a refinement that is both locally finite and σ-discrete. □

Remark 8.5.15 1. We now know that (i) Lindelöf T_3-spaces and (ii) metric spaces are paracompact. Since there are metric spaces that are not Lindelöf and Lindelöf T_3-spaces that are not metrizable, paracompactness (even together with T_2) does not imply metrizability or Lindelöf.

2. The preceding results might lead one to think that every normal space is paracompact, but this is false! The counterexamples are complicated. Cf. e.g. Example 20.11 and Problem 20H in [237].

3. Just as normality, paracompactness is neither hereditary nor preserved by products: □

Example 8.5.16 1. By Corollary 8.1.29, a subspace Y of a compact Hausdorff (thus paracompact) space X can fail to be normal. Since Y in any case is Hausdorff, such a Y is not paracompact by Proposition 8.5.8. Thus paracompactness is not hereditary (so that one studies hereditarily paracompact spaces).

2. As already noted in Remark 8.2.37, the Sorgenfrey line $(\mathbb{R}, \tau_S)$ is Lindelöf and T_3, thus paracompact by Proposition 8.5.12. By Proposition 8.1.37, the Sorgenfrey plane $(\mathbb{R}, \tau_S)^2$ is not normal (but Hausdorff), thus not paracompact. This shows that paracompactness is not preserved by products. □

At least, in analogy to Exercise 8.1.25, we have:

Lemma 8.5.17 *Closed subspaces of paracompact spaces are paracompact.*

Proof. Let (X, τ) be paracompact and $Y \subseteq X$ closed. Let $\mathcal{U} \subseteq \tau_Y$ be an open cover of Y. For every $U \in \mathcal{U}$, choose an open $V_U \in \tau$ such that $U = Y \cap V_U$. Then $\mathcal{U}' = \{V_U \mid U \in \mathcal{U}\} \cup \{X \backslash Y\}$ is an open cover of X. By paracompactness of X, there is a locally finite refinement $\mathcal{V}' \subseteq \tau$ of $\mathcal{V}$. Now $\mathcal{V} = \{V \cap Y \mid V \in \mathcal{V}'\}$ is an open refinement of $\mathcal{U}$. If $x \in Y$ then by local finiteness of $\mathcal{V}'$ there is a τ-open neighborhood W of x meeting only finitely many $V \in \mathcal{V}'$. Then $W \cap Y$ is a τ_Y-open neighborhood of x meeting only finitely many $V \in \mathcal{V}$, so that $\mathcal{V}$ is locally finite. ■

(In analogy to Exercise 8.1.26, every F_σ-subset of a paracompact space is paracompact, cf. e.g. [237].)

Like normality, paracompactness behaves well under closed maps, at least when combined with the Hausdorff property. For the proof (which simplifies if f is also proper) see [237, 75].

Proposition 8.5.18 *Let X be paracompact Hausdorff and $f : X \to Y$ continuous, closed and surjective. Then Y is paracompact Hausdorff.*

Exercise 8.5.19 Prove that $\bigoplus_{i \in I} X_i$ is paracompact if and only if each X_i is paracompact.

8.5.2 Paracompactness and local compactness

Proposition 8.5.20 *A locally compact Hausdorff space X is paracompact if and only if $X \cong \bigoplus_{i \in I} X_i$, where all X_i are Lindelöf (equivalently: σ-compact).*

Proof. $\Leftarrow$ Since $X \cong \bigoplus_{i \in I} X_i$ is locally compact Hausdorff, each X_i is locally compact Hausdorff (Exercise 7.8.35). Now, locally compact Hausdorff spaces are T_3 (Corollary 8.1.9), so that the claim follows from the paracompactness of Lindelöf T_3-spaces (Proposition 8.5.12) and of arbitrary direct sums of paracompact spaces (Exercise 8.5.19). (The equivalence of σ-compactness and the Lindelöf property for locally compact spaces was shown in Exercise 7.8.43(i-ii).)

$\Rightarrow$ Using local compactness pick, for every $x \in X$, an open U_x and a compact K_x such that $x \in U_x \subseteq K_x$. Since X is Hausdorff, K_x is closed. Thus $\overline{U_x} \subseteq K_x$ is compact. Now by paracompactness, the cover $\mathcal{U} = \{U_x\}_{x \in X}$ has a locally finite refinement $\mathcal{V}$, and it is automatic that $\overline{V}$ is compact for every $V \in \mathcal{V}$. For every $x \in X$, let $W_x \ni x$ be an open set meeting only finitely many V's. If $V \in \mathcal{V}$ and $x \in \overline{V}$ then $V \subseteq \overline{V} \subseteq \bigcup_{x \in \overline{V}} W_x$. Since $\overline{V}$ is compact, we conclude that V is contained in the union of finitely many W_x's. Thus each $V \in \mathcal{V}$ meets only finitely many other elements of $\mathcal{V}$. (This is stronger than local finiteness.)

For $V_0 \in \mathcal{V}$ and $k \in \mathbb{N}$, consider chains $V_1, V_2, \ldots, V_k \in \mathcal{V}$ satisfying $V_i \cap V_{i+1} \neq \emptyset$ for $i = 0, \ldots, k-1$. The above finiteness property of $\mathcal{V}$ implies that only finitely many $V_k \in \mathcal{V}$ are connected to V_0 by a chain of length k. Thus the family $F(V_0)$ of $V's$ connected to V_0 by a chain of finite length is at most

countable. The union $S(V_0) = \bigcup F(V_0)$ of these $V's$ is open, and for any $V, V' \in \mathcal{V}$ it is clear from this construction that either $S(V) = S(V')$ or $S(V) \cap S(V') = \emptyset$. Since $\mathcal{V}$ is a cover of X, each $S(V)$ is the complement of the union of the $S(V') \neq S(V)$, and therefore is closed. Thus X is a disjoint union of clopen subsets, thus a direct sum.

Now, $S(V) = \overline{S(V)} = \overline{\bigcup\{V \mid V \in F(V_0)\}} = \bigcup\{\overline{V} \mid V \in F(V_0)\}$, where the last identity is due to the fact that $F(V_0) \subseteq \mathcal{V}$ is locally finite and Exercise 8.2.51(ii). Thus each $S(V)$ is a countable union of compact subsets of X, i.e. σ-compact. As such it also is Lindelöf, cf. Exercise 7.8.43(i). ■

Now we can prove the missing implication (v)⇒(iv) in Corollary 8.2.40.

Corollary 8.5.21 *A locally compact Hausdorff space X is metrizable if and only if $X \cong \bigoplus_{i \in I} X_i$, where all X_i are second countable.*

Proof. ⇐ This was implication (iv)⇒(v) in Corollary 8.2.40. ⇒ Since X is metrizable, it is paracompact by Theorem 8.5.13. Thus Proposition 8.5.20 gives $X \cong \bigoplus_{i \in I} X_i$, where all X_i are Lindelöf. Each X_i is a subspace of X, thus metrizable and therefore second countable by Exercise 7.1.9. ■

That direct sums appear in Proposition 8.5.20 and Corollary 8.5.21 simply is a consequence of the fact that local compactness, paracompactness and metrizability are preserved under arbitrary direct sums, but not compactness, second countability and the Lindelöf property.

Corollary 8.5.22 *A connected locally compact Hausdorff space is ...*

(i) paracompact if and only if it is Lindelöf (equivalently, σ-compact),

(ii) metrizable if and only if it is second countable.

Proof. This follows directly from Proposition 8.5.20 and Corollary 8.5.21 since connectedness implies that only one summand appears in the direct sum. ■

8.5.3 Paracompactness and normality of product spaces

Despite Example 8.5.16.2, paracompactness behaves somewhat better with respect to products than normality does:

Proposition 8.5.23 *If X is paracompact and Y is compact then $X \times Y$ is paracompact. If, in addition, X and Y are Hausdorff then $X \times Y$ is normal.*

Proof. Let $\mathcal{U}$ be an open cover of $X \times Y$ and let $x \in X$. Since $Y \cong \{x\} \times Y$ is compact, there are finitely many $U_{x,1}, \ldots, U_{x,n_x} \in \mathcal{U}$ such that $\{x\} \times Y \subseteq \bigcup_{i=1}^{n_x} U_{x,i}$. By Lemma 7.5.1 and compactness of Y, there is an open $V_x \ni x$ such that $V_x \times Y \subseteq \bigcup_{i=1}^{n_x} U_{x,i}$. Clearly $\mathcal{V} = \{V_x\}_{x \in X}$ is an open cover of X, which by paracompactness has a locally finite refinement $\mathcal{W}$. For every $W \in \mathcal{W}$ choose an $x_W \in X$ such that $W \subseteq V_{x_W}$. Now define

$$\mathcal{S} = \{(W \times Y) \cap U_{x_W,i} \mid W \in \mathcal{W},\ i = 1, \ldots, n_{x_W}\}.$$

Every $S \in \mathcal{S}$ is open and contained in some $U \in \mathcal{U}$. Since $\mathcal{W}$ is a cover of X, every $x \in X$ is contained in some $W \in \mathcal{W}$. By construction, this W is contained in V_{x_W}, and $V_{x_W} \times Y \subseteq \bigcup_{i=1}^{n_{x_W}} U_{x_W,i}$. Thus $\bigcup \mathcal{S} = X \times Y$, and $\mathcal{S}$ is an open refinement of $\mathcal{U}$. Finally, if $(x, y) \in X \times Y$, there is an open neighborhood $T \subseteq X$ of x meeting only finitely many $W \in \mathcal{W}$ (by local finiteness of $\mathcal{W}$). Then $T \times Y$ is an open neighborhood of (x, y) that can meet only those elements $(W \times Y) \cap U_{x_W,i}$ of $\mathcal{S}$

for which $T \cap W \neq \emptyset$. Since these are finitely many, $\mathcal{S}$ is locally finite, proving paracompactness of $X \times Y$.

If X and Y are Hausdorff, then so is $X \times Y$ and combining this with the paracompactness proven above, Proposition 8.5.8 gives normality of $X \times Y$. ■

Combining this with Theorem 8.5.13 gives the following, which was used in Section 8.5.6:

Corollary 8.5.24 *If X is metrizable and Y is compact Hausdorff then $X \times Y$ is paracompact Hausdorff, thus normal.*

The preservation of normality under products with compact Hausdorff spaces actually characterizes paracompact Hausdorff spaces:

Theorem 8.5.25 *A space X is paracompact Hausdorff if and only if $X \times Y$ is normal for every compact Hausdorff space Y.*

Proof. $\Rightarrow$ Cf. Proposition 8.5.23. $\Leftarrow$ Taking Y to be a one-point space, which is compact Hausdorff, we have $X \cong X \times Y$, thus the assumption implies that X is normal. Thus X is completely regular and therefore admits a Hausdorff compactification $\widehat{X}$ by Theorem 8.3.20. By our assumption, $X \times \widehat{X}$ is normal, so that the following Proposition 8.5.26 gives the result. ■

Proposition 8.5.26 (H. Tamano (1960)) *If X admits a Hausdorff compactification $\widehat{X}$ such that $X \times \widehat{X}$ is normal then X is paracompact.*

Proof. Since $\widehat{X}$ is T_2, the diagonal $\Delta = \{(x,x) \mid x \in X\} \subseteq X \times \widehat{X}$ is closed, cf. Exercise 8.2.30(i). Let $\mathcal{U} = \{U_i\}_{i \in I}$ be an open cover of X. Considering X as a subspace of $\widehat{X}$, for each i we can pick an open $V_i \subseteq \widehat{X}$ such that $U_i = X \cap V_i$. Now $Z = \widehat{X} \backslash \bigcup_{i \in I} V_i \subseteq \widehat{X} \backslash X$ is closed. Thus $X \times Z$ and Δ are disjoint closed subsets of the space $X \times \widehat{X}$. The latter being normal by assumption, Urysohn's lemma gives an $f \in C(X \times \widehat{X}, [0,1])$ such that $f \restriction \Delta = 0$ and $f \restriction X \times Z = 1$. Then

$$d(x,y) = \sup_{z \in \widehat{X}} |f(x,z) - f(y,z)|,$$

is finite by compactness of $\widehat{X}$ and continuity of f. It is easy to check (as in Exercise 8.2.30) that d is a pseudometric on X. Exercise 7.7.45(i) implies $d(x_\iota, x_0) \to 0$ as $x_\iota \to x_0$. Thus every net that converges w.r.t. the original topology τ is τ_d-convergent, which means that τ_d is coarser than τ. By Remark 8.5.14, the cover $\{B^d(x,1/2)\}_{x \in X}$ of X has a locally finite open refinement refinement $\{W_t\}_{t \in T}$, where both 'locally finite' and 'open' refer to τ_d. Since τ_d is coarser than τ, $\{W_t\}_{t \in T}$ is also open and locally finite w.r.t. τ. If $x \in X$ and $y \in B^d(x,1/2)$, we have $f(x,y) = |f(x,y) - f(y,y)| \leq d(x,y) < 1/2$ (we used $f \restriction \Delta = 0$). Thus $f(x,y) \leq 1/2$ for all $y \in \mathrm{Cl}_{\widehat{X}}(B^d(x,1/2))$. For $t \in T$ there is an $x \in X$ such that $W_t \subseteq B^d(x,1/2)$. Then $\mathrm{Cl}_{\widehat{X}}(W_t) \subseteq \mathrm{Cl}_{\widehat{X}}(B^d(x,1/2))$, so that $y \in \mathrm{Cl}_{\widehat{X}}(W_t)$ implies $f(x,y) \leq 1/2$. On the other hand, by construction we have $f(x,y) = 1$ whenever $y \in Z$. This implies $\mathrm{Cl}_{\widehat{X}}(W_t) \cap Z = \emptyset$, so that we have $\mathrm{Cl}_{\widehat{X}}(W_t) \subseteq \bigcup_{i \in I} V_i$ $\forall t \in T$. Since $\mathrm{Cl}_{\widehat{X}}(W_t)$ is compact, there is a finite set $J_t \subseteq I$ such that $\mathrm{Cl}_{\widehat{X}}(W_t) \subseteq \bigcup_{i \in J_t} V_i$, and since $W_t \subseteq X \cap \mathrm{Cl}_{\widehat{X}}(W_t)$, we have $W_t \subseteq \bigcup_{i \in J_t} U_i$. This implies that $\{W_t \cap U_i\}_{t \in T, i \in J_t}$ is an open cover of X refining $\mathcal{U}$. This cover is locally finite, as follows from the local finiteness of $\{W_t\}$ and the finiteness of the J_t. ■

Remark 8.5.27 1. Theorem 8.5.25 together with the existence of normal spaces that are not paracompact shows that there are normal spaces X such that $X \times Y$ is non-normal for some compact Hausdorff Y.

2. The last few results show that the notion of paracompactness is relevant not only for the existence of partitions of unity subordinate to any open cover, but also for intrinsic questions of general topology, like the difficult problem of normality of product spaces. Cf. Subsection 8.5.6 for an application of this circle of ideas.

3. A T_1-space is called collectionwise normal if given a family $\{C_i \subseteq X\}$ of mutually disjoint closed sets, there are mutually disjoint open sets $\{U_i\}$ such that $U_i \supseteq C_i\ \forall i$. One can prove that every paracompact Hausdorff space is collectionwise normal.

4. One can prove that the following are equivalent for a normal space X, cf. [237]:

(i) Every countable open cover of X has a locally finite refinement.

(ii) For every countable open cover $\{U_n\}_{n\in\mathbb{N}}$ of X there is a locally finite open shrinking $\{V_n \subseteq U_n\}$ such that $\overline{V_n} \subseteq U_n\ \forall n$.

(iii) If $\{C_n\}_{n\in\mathbb{N}}$ are closed sets in X such that $\bigcap_n C_n = \emptyset$ then there are open sets (or G_δ-sets) $U_n \supseteq C_n\ \forall n$ such that $\bigcap_n U_n = \emptyset$.

(iv) $X \times [0,1]$ is normal.

(v) $X \times Y$ is normal for every Y that is second countable compact T_2 (=compact metrizable).

(vi) The conclusion of Theorem 8.5.33 holds. (Cf. [75, Problem 5.5.20].)

Such spaces are called countably paracompact. Clearly every countably compact space is countably paracompact.

5. There exist spaces that are countably compact, thus countably paracompact, but not paracompact, cf. e.g. [75, Example 5.1.21].

6. Countably paracompact normal spaces are also called binormal. A normal space that is not binormal is called a Dowker space. That such spaces exist was first shown by M. E. Rudin by a complicated construction. For a simpler (and smaller) example cf. [14].

7. Every perfectly normal (T_6) space is countably paracompact, thus binormal. This follows from (iii) since closed sets are G_δ. □

We summarize some of the implications involving compactness, metrizability, (countable) paracompactness and the separation axioms $T_i,\ i \geq 4$:

$$\begin{tikzcd}
\begin{smallmatrix}\text{compact}\\ \text{metrizable}\end{smallmatrix} \arrow[r, Rightarrow] \arrow[d, Rightarrow] & \overset{\text{separable metr.=}}{\text{2nd cnt. } T_3} \arrow[r, Rightarrow] \arrow[d, Rightarrow] & \text{metrizable} \arrow[r, Rightarrow] \arrow[d, Rightarrow] & T_6 \arrow[r, Rightarrow] \arrow[d, Rightarrow] & T_5 \arrow[d, Rightarrow] \\
\text{compact } T_2 \arrow[r, Rightarrow] & \text{Lindelöf } T_3 \arrow[r, Rightarrow] & \text{paracompact } T_2 \arrow[r, Rightarrow] & \begin{smallmatrix}\text{countably}\\ \text{paracompact}\end{smallmatrix} T_4 \arrow[r, Rightarrow] & T_4
\end{tikzcd}$$

Remark 8.5.28 1. None of the implications is reversible. The space I^I is compact T_2, but not T_5. This already provides a counterexample for all vertical implications. $\mathbb{R}$ is in the second column, but not in the first. Every uncountable discrete space is in the third column, but not in the second. The Sorgenfrey line is T_6, but not metrizable. Spaces that are T_5, but not T_6 were encountered in Exercise 8.2.14.

2. For the first two squares we actually have that a space with the properties in the upper right and lower left corner also has the property in the upper left corner. But in the right half of the diagram, this is not the case. E.g., the Sorgenfrey line is Lindelöf and T_6 (thus also paracompact) but not metrizable. □

8.5.4 The Nagata-Smirnov metrization theorem

Definition 8.5.29 *If X is a topological space, a family $\mathcal{U} \subseteq P(X)$ is called* <u>σ-locally finite</u> *if $\mathcal{U} = \bigcup_{i\in\mathbb{N}} \mathcal{U}_i$ where the $\mathcal{U}_i$ are locally finite families.*

Lemma 8.5.30 *A T_3-space admitting a σ-locally finite base is T_6.*

Proof. We will prove that for every open $U \subseteq X$ there is a countable family $\{W_n\}_{n\in\mathbb{N}}$ of open sets such that $\overline{W_n} \subseteq U\ \forall n$ and $U = \bigcup_n W_n$. Given that, we have $U = \bigcup_n W_n \subseteq \bigcup_n \overline{W_n} \subseteq U$, thus $U = \bigcup_n \overline{W_n}$. Thus every open set is F_σ, which is equivalent to every closed set being G_δ. Furthermore, the condition in Lemma 8.1.13 is clearly satisfied (since it would even hold for $C = W$). Thus X is normal and therefore T_6 by Exercise 8.2.8.

Let $\mathcal{V}$ be a σ-locally finite base and $\mathcal{V} = \bigcup_{n\in\mathbb{N}} \mathcal{V}_n$, where each $\mathcal{V}_n = \{V_{n,\imath}\}_{i\in I_n}$ is locally finite. Let $U \subseteq X$ be open. Since X is regular, for every $x \in U$ there is an open V_x such that $x \in V_x \subseteq \overline{V_x} \subseteq U$. Since $\mathcal{V}$ is a base, there are $n_x \in \mathbb{N}, i_x \in I_{n_x}$ such that $x \in V_{n,i} \subseteq V_x$. Thus also $\overline{V_{n_x,i_x}} \subseteq U$. Defining $W_n = \bigcup\{V_{n_x,i_x} \mid x \in X, n_x = n\}$, we have a family $\{W_n\}_{n\in\mathbb{N}}$ of open sets such that $U = \bigcup_{n\in\mathbb{N}} W_n$. Since $\{V_{n_x,i_x} \mid n_x = n\} \subseteq \mathcal{V}_n$ is locally finite, Exercise 8.2.51(ii) gives $\overline{W_n} = \overline{\bigcup\{V_{n_x,i_x} \mid n_x = n\}} = \bigcup\{\overline{V_{n_x,i_x}} \mid n_x = n\} \subseteq U\ \forall n$, and we are done. ■

Theorem 8.5.31 (Nagata-Smirnov metrization theorem (1950/1)) *A topological space X is metrizable if and only if it is T_3 and its topology has a σ-locally finite base.*

Proof. $\Rightarrow$ By Lemma 8.1.11, a metric space is T_3. By Remark 8.5.14, the open cover $\mathcal{U}_n = \{B(x, 1/n)\}_{x\in X}$ has a σ-discrete open refinement $\mathcal{V}_n$. Then $\mathcal{V} = \bigcup_{n\in\mathbb{N}} \mathcal{V}_n$ is σ-discrete. Now, if $x \in U \in \tau$ then there is an $n \in \mathbb{N}$ such that $B(x, 1/n) \subseteq U$. Since $\mathcal{V}_{2n}$ is a cover of X, there is a $V \in \mathcal{V}_{2n}$ such that $x \in V$. Since $\mathcal{V}_{2n}$ is subordinate to $\mathcal{U}_{2n}$, there is $y \in X$ such that $V \subseteq B(y, 1/2n)$. Thus $d(z, x) < 2/2n = 1/n\ \forall z \in V$, to wit $V \subseteq B(x, 1/n) \subseteq U$, so that $\mathcal{V}$ is a σ-discrete base, thus also σ-locally finite.

$\Leftarrow$: Let $\mathcal{V}$ be a σ-locally finite base and $\mathcal{V} = \bigcup_{n\in\mathbb{N}} \mathcal{V}_n$, where each $\mathcal{V}_n = \{V_{n,i}\}_{i\in I_n}$ is locally finite. Since X is T_6 by Lemma 8.5.30, Exercise 8.2.8 gives functions $f_{n,i} \in C(X, [0, 1])$ such that $X\backslash V_{n,i} = f_{n,i}^{-1}(0)$. Since $\{V_{n,i}\}_{i\in I_n}$ is a locally finite family in X, the family $\{(V_{n,i} \times X) \cup (X \times V_{n,i})\}_{i\in I_n}$ in $X \times X$ is locally finite. Thus

$$d_n(x, y) = \sum_{i\in I_n} |f_{n,i}(x) - f_{n,i}(y)|$$

is a continuous function function $d_n : X \times X \to [0, \infty)$ by Lemma 8.2.48, and it clearly is a pseudometric on X. Let $C \subseteq X$ be closed and $x \in X\backslash C =: U$. Then $x \in U \in \tau$, and since $\mathcal{V}$ is a base, there exist $n \in \mathbb{N}$ and $i \in I_n$ such that $x \in V_{n,i} \subseteq U$. By construction, $f_{n,i}(x) > 0$ and $f_{n,i} = 0$ on $X\backslash V_{n,i} \supseteq C$. Thus $\inf_{y\in C} d_n(x, y) = f_{n,i}(x) > 0$, so that $\mathcal{D} = \{d_n\}_{n\in\mathbb{N}}$ is a countable family of continuous seminorms that separates points from closed sets. Thus X is metrizable by Proposition 8.2.32. ■

Remark 8.5.32 1. Urysohn's metrization theorem (Theorem 8.2.33) follows immediately from Theorem 8.5.31: By second countability there is a countable base $\mathcal{B} = \{U_i\}_{i\in\mathbb{N}}$. Now $\mathcal{B} = \bigcup_{i\in\mathbb{N}} \mathcal{B}_i$, where $\mathcal{B}_i = \{U_i\}$ trivially is a locally finite family, thus $\mathcal{B}$ is σ-locally finite.

2. The family $\mathcal{F} = \{f_{n,i}\} \subseteq C(X, [0, 1])$ constructed in the proof separates points from closed sets, but if X is not second countable, $\mathcal{F}$ must be uncountable, as noted in Remark 8.2.34. The essential idea of the above proof is to use the σ-local finiteness of $\mathcal{V}$ to produce a countable family

$\{d_n\}$ of continuous functions separating points from closed sets, but they live on $X \times X$ rather than X ! (This is related to the notion of 'uniform structures'.)

3. Many other statements equivalent to metrizability have been found, but they involve concepts beyond the scope of these notes. (E.g. 'developments', which are certain countable families of covers.) Cf. e.g. [237, 75] and [106, Sections e02, e03]. □

8.5.5 Two applications of paracompactness

Here is an application to semicontinuous functions:

Theorem 8.5.33 *Let X be paracompact. Then given functions $f, g : X \to \mathbb{R}$ that are upper and lower semicontinuous, respectively, with $f(x) < g(x)\ \forall x \in X$, there exists a continuous $h : X \to \mathbb{R}$ such that $f(x) < h(x) < g(x)\ \forall x \in X$.*

Proof. For each $r \in \mathbb{Q}$, define $U_r = f^{-1}((-\infty, r)) \cap g^{-1}((r, \infty))$, which is open by the semicontinuity properties. We have $U_r = \{x \in X \mid f(x) < r < g(x)\}$, implying that $\{U_r\}_{r \in \mathbb{Q}}$ is an open cover of X. By Theorem 8.5.11 there is a locally finite partition of unity $\{f_r\}_{r \in \mathbb{Q}}$ subordinate to $\{U_r\}$. The local finiteness implies that $h(x) = \sum_{r \in \mathbb{Q}} r f_r(x)$ is continuous. Every $x \in X$ is contained in finitely many U_r. If $a_x = \min\{s \in \mathbb{Q} \mid x \in U_s\}$, $b_x = \max\{s \in \mathbb{Q} \mid x \in U_s\}$ then $f(x) < a_x \leq b_x < g(x)$. Since $h(x)$ is a convex combination of $\{s \in \mathbb{Q} \mid x \in U_s\}$ we have $a_x \leq h(x) \leq b_x$. Combining these inequalities, we obtain $f(x) < h(x) < g(x)\ \forall x$. ■

Our second application of paracompactness is more substantial in that it uses not only the definition, but also the non-trivial fact that metric spaces are paracompact.

Definition 8.5.34 *A topological vector space V (over $\mathbb{R}$ or $\mathbb{C}$) is called <u>locally convex</u> if it is T_0 and $0 \in V$ has a neighborhood base consisting of convex open sets.*

Locally convex topological vector spaces will be studied in a bit more depth in Section G.5. For now, we just observe the following:

Lemma 8.5.35 *If V is a vector space over $\mathbb{R}$ or $\mathbb{C}$ and $\|\cdot\|$ is a norm on it, the norm topology (coming from the metric $d(x, y) = \|x - y\|$) is locally convex.*

Proof. The balls $B(0, r)$ form a neighborhood base at 0 and are convex by Lemma 7.7.59. ■

Theorem 8.5.36 (Dugundji's Extension Theorem (1951)) *Let X be a metrizable space, $A \subseteq X$ closed and L a locally convex topological vector space. Then*

(i) There is a linear map $C(A, L) \to C(X, L), f \mapsto \widehat{f}$ satisfying $\widehat{f} \restriction A = f$ and $\widehat{f}(X) \subseteq \mathrm{conv}(f(A))$.

(ii) If $K \subseteq L$ is convex then every $f \in C(A, K)$ has an extension $\widehat{f} \in C(X, K)$.

Proof. (i) Let d be a metric on X compatible with the topology. Since A is closed, for each $x \in X \backslash A$ we have $\varepsilon_x := \mathrm{dist}(x, A)/4 > 0$. Then $B(x, \varepsilon_x) \subseteq X \backslash A$. Since $X \backslash A$ is paracompact, the open cover $\{B(x, \varepsilon_x)\}_{x \in X \backslash A}$ has a locally finite refinement $\mathcal{U}$. Recall the notation A_ε from (3.1). Now, if $x \notin A_{2\eta}$ (i.e. $\mathrm{dist}(x, A) \geq 2\eta$) then $B(x, \varepsilon_x) \cap A_\eta = \emptyset$ by our choice of ε_x. Thus (*) if $U \in \mathcal{U}$ intersects A_η and $U \subseteq B(x, \varepsilon_x)$ (such an x exists since $\mathcal{U}$ refines $\{B(x, \varepsilon_x)\}$) then $x \in A_{2\eta}$, implying $\mathrm{diam}(U) \leq 2\varepsilon_x = \mathrm{dist}(x, A)/2 \leq \eta$.

For each non-empty $U \in \mathcal{U}$ pick any $x_U \in U$ and then choose $a_U \in A$ such that $d(x_U, a_U) < 2\mathrm{dist}(x_U, A)$. Now the cover $\mathcal{U}$ and the points $\{a_U\}_{U \in \mathcal{U}}$ satisfy the following claim (**):

If $a \in A$ and $W \subseteq X$ is an open neighborhood of a then there is an open $V \subseteq W$ containing a such that $U \in \mathcal{U},\ U \cap V \neq \emptyset$ implies $a_U \in A \cap W$.

It suffices to prove this for $W = B(a, \varepsilon)$. Put $V = B(a, \varepsilon/6)$. If now $U \in \mathcal{U}$ intersects V then it intersects $A_{\varepsilon/6}$, thus by (*) we have $\mathrm{diam}(U) \leq \varepsilon/6$. This implies $U \subseteq B(a, \varepsilon/4)$, which gives $d(a, x_U) < \varepsilon/4$ and therefore $\mathrm{dist}(x_U, A) < \varepsilon/4$. Now,

$$d(a_U, a) \leq d(a_U, x_U) + d(x_U, a) \leq 2\mathrm{dist}(x_U, A) + \mathrm{dist}(x_U, A) \leq 3\mathrm{dist}(x_U, A) \leq \frac{3}{4}\varepsilon < \varepsilon.$$

This proves $a_U \in W$, thus the claim.

Let now $\{\phi_U\}_{U \in \mathcal{U}}$ be a partition of unity subordinate to the cover $\mathcal{U}$. For $f \in C(A, L)$ define $\widehat{f} : X \to L$ by

$$\widehat{f}(x) = \begin{cases} f(x) & x \in A \\ \sum_{U \in \mathcal{U}} \phi_U(x) f(a_U) & x \in X \backslash A \end{cases}$$

From this definition it is clear that $\widehat{f} \upharpoonright A = f$, that the assignment $f \mapsto \widehat{f}$ is linear and that $\widehat{f}$ takes values in the convex hull of $f(A)$. (Note that no closure is needed since the family $\{\phi_U\}$ is locally finite!) It remains to prove continuity of $\widehat{f}$. Each point of the open set $X \backslash A$ has a neighborhood V on which only finitely many ϕ_U are non-zero. Thus $\widehat{f} \upharpoonright V$ is a finite sum of products of continuous functions, thus continuous on V and therefore on $X \backslash A$.

Let now $a \in A$ and $S \subseteq L$ an open neighborhood of $f(a)$. Since L is locally convex and f continuous, there is a convex open $C \subseteq S$ and an open $W \subseteq X$ containing a such that $f(W \cap A) \subseteq C \subseteq S$. Let now $V \subseteq W$ be as provided by (**). It is clear that $\widehat{f}(V \cap A) = f(V \cap A) \subseteq C \subseteq S$. And if $x \in V \backslash A$ then x belongs to at most finitely many $U_1, \ldots, U_n \in \mathcal{U}$. Each of these U_i intersects V, so that (**) gives $a_{U_i} \in A \cap W\ \forall i = 1, \ldots, n$. Thus $f(a_{U_i}) \in C\ \forall i$, and since $\widehat{f}(x)$ is a convex combination of the $f(a_{U_i})$, we conclude $\widehat{f}(x) \in C \subseteq S$. Thus $\widehat{f}(V) \subseteq S$, proving continuity of $\widehat{f}$ at each point of A, and therefore everywhere on X.

(ii) Obvious consequence of (i). ∎

8.5.6 ⋆⋆ A glimpse at generalized metric spaces

There have been many attempts at defining a notion of 'generalized metric spaces' by enlarging the class of metric spaces while maintining as many of its desirable properties as possible. Here we briefly look at the most accessible of these attempts, based on Corollary 8.5.24:

Definition 8.5.37 *A paracompact M-space is a topological space that is homeomorphic to a closed subspace of a space $M \times K$ with K compact Hausdorff and M metrizable. The class of paracompact M-spaces is denoted by $\mathcal{PM}$.*

Corollary 8.5.38 *(i) All spaces in $\mathcal{PM}$ are paracompact Hausdorff, thus normal.*

(ii) The class $\mathcal{PM}$ is closed w.r.t. passage to closed subspaces and countable products.

Proof. (i) Follows from Corollary 8.5.24 and Lemma 8.5.17.

(ii) Follows from the following facts: Closed subspaces and arbitrary products of compact spaces are compact, countable products of metrizable spaces are metrizable, and a product $\prod_i Y_i$ of closed subspaces $Y_i \subseteq X_i$ is closed in $\prod_i X_i$. ∎

It clearly follows that $\mathcal{P}$ is the smallest class of spaces containing the metrizable and the compact Hausdorff spaces and being closed w.r.t. closed subspaces and countable products. There is an alternative characterization:

Proposition 8.5.39 *A topological space X is a paracompact M-space if and only if it is completely regular and admits a continuous closed proper map $f : X \to M$ with M metrizable.*

Proof. $\Rightarrow$ Let $X \subseteq M \times K$ be closed with K compact Hausdorff and M metrizable. Clearly X is $T_{3.5}$. Now $p_1 : M \times K \to M$ is closed by Exercise 7.5.5. Since X is closed, every closed $C \subseteq X$ is closed in $M \times K$, thus $f = p_1 \upharpoonright X$ is closed. If $C \subseteq M$ is compact then $f^{-1}(C) = X \cap (C \times K)$ is a closed subset of the compact space $C \times K$ and therefore compact. Thus $f : X \to M$ is continuous, closed and proper with M is metrizable.

$\Leftarrow$ Let X be completely regular and $f : X \to M$ continuous, closed and proper with M metric. Then the map $g : X \to M \times \beta X,\ x \mapsto (f(x), \iota_X(x))$ (where βX denotes the Stone-Čech compactification) is continuous and injective. Since $\iota_X : X \to \beta X$ is an embedding, the same is true for $g = (f, \iota_X)$ (since $p_2 \circ g = \mathrm{id}_X$). Since M is metric and βX compact Hausdorff, $X \cong g(X)$ is a paracompact M-space provided we show that $g(X) \subseteq M \times \beta X$ is closed.

Thus assume $g(X)$ is non-closed and define $\widehat{X} := \overline{g(X)} \subseteq M \times \beta X$. Then $g = (f, \iota)$, considered as a map $X \to \widehat{X}$ is an embedding. Clearly, $p_1 \circ g = f$. Identifying X with $g(X) \subseteq \widehat{X}$, this means that $p_1 : \widehat{X} \to M$ is a continuous extension of the closed proper map $f : X \to M$ to a strictly larger Hausdorff space $\widehat{X} \supsetneq X$. This contradicts Lemma 7.8.68, thus $g(X) \subseteq M \times \beta X$ must be closed. ■

Remark 8.5.40 The argument in the last paragraph of Proposition 8.5.39 works for any Hausdorff compactification, not just βX. The statement also has a converse (whose proof is similar to the proof of $\Rightarrow$ in the proposition). This leads to the following characterization of closed proper maps: If X is completely regular, then a continuous $f : X \to Y$ is closed and proper if and only if the embedding $g = (f, \iota)$ has closed image for some (equivalently, every) Hausdorff compactification $X \overset{\iota}{\hookrightarrow} \widehat{X}$. □

8.6 Summary of the generalizations of compactness

We briefly collect the generalizations of the notion of compactness that we have encountered. Most of them are implied by compactness. (For sequential compactness we also need first countability, for rimcompactness and realcompactness we need the Hausdorff property.)

0. Lindelöf property: Every open cover has a countable subcover.
1. compactness: Every open cover has a finite subcover.
2. sequential compactness: Every sequence has a convergent subsequence.
3. countable compactness: Every countable open cover has a finite subcover.
4. weak countable compactness: Every closed discrete subspace is finite.
5. pseudocompactness: Every continuous $\mathbb{R}$-valued function is bounded. Equivalently (for $T_{3.5}$-spaces), every locally finite open cover is finite.
6. realcompactness: There is an embedding into $\mathbb{R}^\chi$ for some cardinal number χ.
7. σ-compactness: There is a countable compact cover.
8. hemicompactness: There is a countable compact cover such that every compact set is contained in an element of the cover.

9. local compactness: Every point has a compact neighborhood.
10. strong local compactness: There is a base whose elements have compact closures.
11. compact generation (=k-spaces): A subset is closed if and only if its intersection with every compact set is closed.
12. rimcompactness: There is a base whose elements have compact boundaries.
13. metacompactness: Every open cover has a point-finite refinement.
14. paracompactness: Every open cover has a locally finite refinement.
15. countable paracompactness: Every countable open cover has a locally finite refinement.

Part III:

Connectedness. Steps towards algebraic topology

Chapter 9

Connectedness: Fundamentals

So far, connectedness has been encountered only passingly. We have postponed the full treatment of the concept up to this point since we consider it as a part of algebraic topology, or at least located on the boundary between general and algebraic topology. This is not often stated explicitly, but it is quite obvious: Connectedness has a functorial nature (Theorem 11.1.4) and admits a 'higher-dimensional' generalization that is close to (co)homology, cf. Section 10. Furthermore, connectedness is related to (but different from) path-connectedness, whose higher-dimensional extensions (of which we only consider the first, i.e. fundamental group and groupoid) belong to homotopy theory.

9.1 Connected spaces and components

9.1.1 Basic results

Recall that a space (X, τ) is called connected if it has no clopen subsets other than $\emptyset$ and X. Now we will go deeper into the subject, and we begin by studying the behavior of connectedness under the various constructions. Connectedness is preserved by continuous functions:

Lemma 9.1.1 *If $f : (X, \tau) \to (Y, \sigma)$ is continuous and (X, τ) is connected then the subspace $f(X) \subseteq Y$ is connected. In particular, quotients of connected spaces are connected.*

Proof. If the subspace $f(X) \subseteq Y$ is not connected, it contains a non-trivial clopen subset C. But then $f^{-1}(C) \subseteq X$ is non-trivial and clopen, contradicting the connectedness of X. ■

We summarize what we know so far about connectedness:

(i) Lemma 9.1.1 in particular implies that quotient spaces of connected spaces are connected.

(ii) Connectedness of products will be the subject of Exercise 9.1.5.

(iii) By Exercise 2.5.9, connectedness is not hereditary.

(iv) In Section 6.3 we saw that a non-trivial direct sum $X \cong X_1 \oplus X_2$ never is connected, and that the converse is also true, i.e. every non-connected space is a direct sum.

(v) If X_1, X_2 are nontrivial and connected then $X_1 \oplus X_2 = \iota_1(X_1) \cup \iota_2(X_2)$ is non-connected. Thus connectedness of $Y_1, Y_2 \subseteq X$ does not imply connectedness of $Y_1 \cup Y_2$. But:

Lemma 9.1.2 *If $Y_i \subseteq (X, \tau)$ is connected for each $i \in I$ and there exists $i_0 \in I$ such that $Y_i \cap Y_{i_0} \neq \emptyset$ $\forall i$ then $Y = \bigcup_i Y_i$ is connected. (A simpler sufficient condition is $\bigcap_i Y_i \neq \emptyset$.)*

Proof. Assume that we have $Y = U_1 \cup U_2$ with $U_1, U_2 \in \tau_Y$ and $U_1 \cap U_2 = \emptyset$. For each $i \in I$ it follows that $V_i = U_1 \cap Y_i$ and $W_i = U_2 \cap Y_i$ are disjoint open sets such that $V_i \cup W_i = Y_i$. Since each Y_i is connected, we must have $V_i = Y_i, W_i = \emptyset$ or $\leftrightarrow$. In the first case we have $U_1 \cap Y_i = Y_i$, thus $Y_i \subseteq U_1$. In the second case, $Y_i \subseteq U_2$. Since $U_1 \cap U_2 = \emptyset$ and each Y_i has non-trivial intersection with Y_{i_0}, all Y_i must be contained in the same U_k ($k \in \{1, 2\}$) as Y_{i_0}. But this means that either $U_2 = \emptyset$ or $U_1 = \emptyset$, thus $Y = \bigcup_i Y_i$ is connected. If $\bigcap_i Y_i \neq \emptyset$, we clearly have $Y_i \cap Y_j \neq \emptyset \ \forall i, j$. Thus the above holds for any choice of $i_0 \in I$. ■

Exercise 9.1.3 Let $\emptyset \neq Y_i \subseteq X$ be connected for each $i \in I$. Define an undirected graph G by $V(G) = I$ and $E(G) = \{(i, j) \mid Y_i \cap Y_j \neq \emptyset\}$. Let $Y = \bigcup_{i \in I} Y_i$. Prove

(i) If G is connected then Y is connected.

(ii) If all Y_i are open and Y is connected then G is connected.

Lemma 9.1.4 *If $Y \subseteq (X, \tau)$ is connected then $\overline{Y}$ is connected.*

Proof. It is sufficient to prove this in the case that Y is dense in X, i.e. $\overline{Y} = X$. Assume that we have $\overline{Y} = U_1 \cup U_2$ with $U_1, U_2 \in \tau$ and $U_1 \cap U_2 = \emptyset$. Let $V_i = U_i \cap Y$. Then V_1, V_2 are disjoint open subsets in (Y, τ_Y) whose union is Y. Since Y is connected, we must have $V_1 = Y, V_2 = \emptyset$ or $\leftrightarrow$. We may assume that the first alternative holds, which means $U_1 \cap Y = Y$ and $U_2 \cap Y = \emptyset$. This implies $Y \subseteq U_1$, and since U_1 is clopen, we have $\overline{Y} \subseteq U_1$. Therefore $U_2 = \emptyset$, thus $\overline{Y}$ is connected. ■

Exercise 9.1.5 (Connectivity of products) (i) Assume that $X_i \neq \emptyset \ \forall i \in I$. Prove that if $X = \prod_i X_i$ is connected then each X_i is connected.

(ii) Let X, Y be non-empty and connected. Prove that $X \times Y$ is connected. Hint: Apply Lemma 9.1.2 to $(X \times \{y_0\}) \cup \bigcup_{x \in X} (\{x\} \times Y)$.

(iii) Let $X_i \neq \emptyset$ be connected $\forall i$. Choose $z = (z_i) \in \prod_i X_i$. For $J \subseteq I$ define $X_J = \{x \in X \mid x_i = z_i \ \forall i \in I \backslash J\}$. Prove that X_J is connected for every finite $J \subseteq I$.

(iv) Continuing (iii), prove that $Y = \displaystyle\bigcup_{\substack{J \subseteq I \\ \#J < \infty}} X_J \subseteq X$ is dense and connected, so that Lemma 9.1.4 implies connectedness of X.

Exercise 9.1.6 Let X be a topological space and $A, B \subseteq X$ connected subspaces. Prove or disprove the following claims:

(a) If $\overline{A} \cap \overline{B} \neq \emptyset$ then $A \cup B$ is connected.

(b) If $\overline{A} \cap B \neq \emptyset$ then $\overline{A} \cup B$ is connected.

(c) If $\overline{A} \cap B \neq \emptyset$ then $A \cup B$ is connected.

Exercise 9.1.7 Prove that a topological space X is connected if and only if it does not admit an open cover $\mathcal{U}$ such that $\#\mathcal{U} \geq 2$ and all elements of $\mathcal{U}$ are non-empty and mutually disjoint.

Remark 9.1.8 The preceding exercise is just the starting point of the beautiful (more so than the singular theory, in this author's view) Čech homology and cohomology theories, cf. e.g. [231, Chapters 5-7]. □

9.1.2 Connected components and local connectedness

Lemma 9.1.9 *Let X be a topological space. For $x, y \in X$ say $x \sim y$ if and only if there exists a connected $Y \subseteq X$ such that $\{x, y\} \subseteq Y$. Then*

(i) $\sim$ is an equivalence relation.

(ii) The $\sim$ equivalence class $[x]$ of x is connected, closed, and equals the union of all connected subsets containing x.

Instead of $[x]$ we write $C(x)$, which we call the connected component of x. A subset $Y \subseteq X$ is called a connected component of X if $Y = C(x)$ for some $x \in X$.

Proof. (i) Reflexivity and symmetry of $\sim$ are obvious. If $x \sim y \sim z$, there are connected sets $Y, Z \subseteq X$ such that $\{x, y\} \subseteq Y$ and $\{y, z\} \subseteq Z$. Then $Y \cap Z$ contains y, thus $Y \cup Z$ is connected by Lemma 9.1.2. Thus $x \sim z$.

(ii) By definition, $y \in [x]$ if and only if there is a connected $Y \subseteq X$ containing x and y. Thus $[x]$ is the union of the connected Y that contain x. In particular $[x]$ is itself connected, again by Lemma 9.1.2. By Lemma 9.1.4, $\overline{[x]}$ is connected and thus contained in $[x]$ by definition. Thus $[x]$ is closed. ■

Since $\sim$ is an equivalence relation, any two equivalence classes $C(x), C(y)$ are either disjoint or equal. Obviously, X is connected if and only if $C(x) = X \ \forall x \in X$.

As we have seen, connected components are always closed. Are they always open?

Example 9.1.10 Consider $X = \{1/n \mid n \in \mathbb{N}\} \cup \{0\} \subseteq [0, 1]$, which is a closed, thus compact subspace of $[0, 1]$. (We know this space from Exercises 5.2.18, 5.2.22 and Remark 7.8.16.3.) All the singletons $\{1/n\}$ are clopen, whereas $\{0\}$ is closed but not open since every neighborhood of 0 contains infinitely many $1/n$. Thus X is not discrete, yet we have $C(x) = \{x\} \ \forall x \in X$. The component $C(0) = \{0\}$ is not open, thus connected components need not be open! As the same time, this example shows that a compact space can have infinitely many connected components! (The two phenomena are related, cf. Proposition 9.1.13. See also Exercise 11.1.23.) □

Definition 9.1.11 *A topological space (X, τ) is (weakly) locally connected if every $x \in X$ has a connected open neighborhood.*

Proposition 9.1.12 *For a space (X, τ), the following are equivalent:*

(i) All connected components $C(x)$ are open.

(ii) X is locally connected.

(iii) There is a homeomorphism $(X, \tau) \cong \bigoplus_i (X_i, \tau_i)$, where all (X_i, τ_i) are connected subspaces of X.

Proof. (i)⇒(ii) For each x, the component $C(x)$ is a connected open neighborhood.

(ii)⇒(i) Fix $x \in X$. For each $y \in C(x)$ there is a connected open neighborhood U_y. Lemma 9.1.2 implies that $C(x) \cup U_y$ is connected and therefore contained in $C(x)$. Thus $U_y \subseteq C(x)$. This implies $C(x) = \bigcup_{y \in C(x)} U_y$, which is open.

(i)⇔(iii) This is clear in view of the discussion in Section 6.3. ■

Proposition 9.1.13 *Let X be a topological space and consider the following statements:*

(i) X is compact and locally connected.

(ii) X has finitely many connected components.

(iii) X is locally connected.

Then (i)⇒(ii)⇒(iii). For compact spaces, (ii)⇔(iii).

Proof. (i)⇒(ii): If X is locally connected, then by Proposition 9.1.12, we have $X \cong \bigoplus_{i\in I} X_i$ where the X_i are mutually disjoint connected subspaces. Since $\{X_i\}_{i\in I}$ is an open cover of X, compactness implies $\#I < \infty$.

(ii)⇒(iii): Since the components $C(x_i)$ form a partitioning of X, for each $i \in \{1,\ldots,n\}$ we have $X\backslash C(x_i) = \bigcup_{j\neq i} C(x_j)$. This is a finite union of closed sets, thus closed. Therefore $C(x_i)$ is open. ■

Remark 9.1.14 Finiteness of the number of connected components follows neither from local connectedness (consider an infinite discrete space) nor from compactness, as we saw in Example 9.1.10. (And (ii) does not imply compactness, since we will soon see that $\mathbb{R}$ is connected.) □

Exercise 9.1.15 Let X be compact and locally connected. Prove that there is a finite subset $A \subseteq X$ such that the quotient space X/A obtained by identifying the points of A is connected.

Exercise 9.1.16 Let $(X_i, \tau_i) \neq \emptyset \; \forall i \in I$. Give necessary and sufficient conditions in order for $\prod_k X_k$ to be (a) locally connected, (b) strongly locally connected.

In Example 9.1.10, we have seen a compact space with infinitely many components. In Section 11.1, we will consider badly disconnected spaces. In particular, we will see that a compact space can be second countable and yet have uncountably many connected components! (Such a space clearly cannot be discrete.) Before we turn to such (apparent) oddities, we will consider Euclidean spaces.

9.1.3 ⋆ Quasi-components

The following relative of the notion of connected component has a number of applications:

Definition 9.1.17 *The quasi-component $Q(x)$ of a point x in a topological space X is defined as the intersection of all clopen sets $C \subseteq X$ that contain x. (Thus each $Q(x)$ is closed.)*

Exercise 9.1.18 Prove that $C(x) \subseteq Q(x) \; \forall x \in X$.

Proposition 9.1.19 *If X is compact Hausdorff then $Q(x) = C(x)$ for all $x \in X$.*

Proof. We will show that $Q(x)$ is connected. This implies $Q(x) \subseteq C(x)$, and the converse being known from Exercise 9.1.18, we have $Q(x) = C(x)$. So assume $Q(x) = A \cup B$ where A, B are clopen (in $Q(x)$) and $A \cap B = \emptyset$. A is closed in $Q(x)$ and $Q(x)$ is closed in X, thus A is closed in X, and the same is true for B. X is compact T_2, thus normal. Therefore, there are open $U, V \subseteq X$ such that $A \subseteq U,\; B \subseteq V,\; U \cap V = \emptyset$. Thus

$$Q(x) = \bigcap \{C \subseteq X \mid C \text{ clopen}, \; x \in C\} \subseteq U \cup V.$$

Since $X\backslash(U \cup V)$ is closed, thus compact, by Remark 7.3.3 we conclude that there are finitely many clopens $C_1, \ldots, C_n$ such that $x \in C_i$ and $C := \bigcap_{i=1}^n C_i \subseteq U \cup V$. As the intersection of finitely many clopen sets, C is clopen, and by construction we have $Q(x) \subseteq C$. Now,

$$\overline{U \cap C} \subseteq \overline{U} \cap \overline{C} = \overline{U} \cap C = \overline{U} \cap C \cap (U \cup V) = U \cap C,$$

since $\overline{U} \cap V = \emptyset$. Thus $U \cap C$ is clopen, and the same holds for $V \cap C$. Since $x \in Q(x)$, we either have $x \in A \subseteq U$ or $x \in B \subseteq V$. In the first case, we have $Q(x) \subseteq U$ (since $U \cap C$ is clopen and contains x) thus $B = \emptyset$. Similarly, in the second case we have $A = \emptyset$. Thus $Q(x)$ is connected. ■

Corollary 9.1.20 *Let X be compact Hausdorff. Then whenever $x \not\sim y$ there is a clopen $C \subseteq X$ with $x \in C \not\ni y$.*

Proof. $y \not\sim x$ is equivalent to $y \notin C(x)$, and Proposition 9.1.19 gives $y \notin Q(x)$. But then the definition of $Q(x)$ means that there is a clopen C with $x \in C \not\ni y$. ■

Proposition 9.1.21 *For a compact Hausdorff space X the equivalence relation $\sim$ from Lemma 9.1.9 is closed.*

Proof. We need to show that for every closed $A \subseteq X$ the $\sim$-saturation $A^\sim$ is closed. To this purpose, let $z \in X \backslash A^\sim$. This means that $z \not\sim x$ for every $x \in A$. Then Corollary 9.1.20 gives, for every $x \in A$, a clopen $C_x \subseteq X$ such that $x \in C_x \not\ni z$. Now $A \subseteq X$ is closed, thus compact. Thus the open cover $\{C_x\}_{x \in A}$ of A has a finite subcover. To wit, there are $x_1, \ldots, x_n \in A$ such that $A \subseteq \bigcup_{i=1}^n C_{x_i} =: D$. As a finite union of clopens, D is closed. Thus $X \backslash D$ is open, and in view of $z \in X \backslash D \subseteq X \backslash A^\sim$, we have shown that $A^\sim$ is closed. ■

9.2 Connectedness of Euclidean spaces

9.2.1 Basics

By Euclidean space we mean the spaces $\mathbb{R}^n$ and their subspaces, equipped with the standard metric topologies. Applied to a subset $X \subseteq \mathbb{R}$, convexity simply means that $a, b \in X, a < b$ implies $[a, b] \subseteq X$, thus 'X has no holes'.

Proposition 9.2.1 *A subspace $X \subseteq \mathbb{R}$ is connected if and only if X is convex.*

Proof. If X is non-convex then there are $a < b < c$ such that $a, c \in X$, but $b \notin X$. But then $X = (X \cap (-\infty, b)) \cup (X \cap (b, \infty))$. The sets $X \cap (-\infty, b)$ and $X \cap (b, \infty)$ are both non-empty and are both open (in X) since $(-\infty, b)$ and (b, ∞) are open subsets of $\mathbb{R}$. Thus we have a decomposition of X into non-empty disjoint open subsets. Thus X is not connected.

We now prove that $[a, b]$, where $a < b$, is connected. Assume that $C \subseteq [a, b]$ is clopen and $\emptyset \neq C \neq [a, b]$. Let $D = [a, b] \backslash C$. We either have $a \in C$ or $a \in D$. We may assume that $a \in C$. With $s = \inf(D)$, we have $s \in \overline{D} = D$ by closedness of D. Since $a \notin D$, we have $s > a$. Thus $[a, s) \subseteq C$. Now closedness of C implies $s \in C$, thus $s \in C \cap D$, contradicting $C \cap D = \emptyset$.

Now let $X \subseteq \mathbb{R}$ be an arbitrary convex subset. Let $x, y \in X$. If $x < y$ put $Y = [x, y]$ and $Y = [y, x]$ otherwise. By convexity of X we have $Y \subseteq X$, and by the above Y is connected. This shows $y \in C(x)$, and since y was arbitrary, we have $C(x) = X$. Thus X is connected. ■

Remark 9.2.2 1. It should be clear that a convex subset of $\mathbb{R}$ is of one of the following types: $\emptyset, \mathbb{R}, [a, \infty), (a, \infty), (-\infty, a], (-\infty, a), (a, b), (a, b], [a, b), [a, b]$.

2. Proposition 9.2.1 again shows that subspaces of connected spaces and unions of connected subspaces need not be connected.

3. We will later see that convex subsets of $\mathbb{R}^n$ are connected. But the converse is false for $n \geq 2$, as follows already from the connectedness of $S^n \subseteq \mathbb{R}^{n+1}$ for $n \geq 1$: □

Proposition 9.2.3 *$I^n, \mathbb{R}^n, S^n$ are connected for all $n \in \mathbb{N}$.*

Proof. Connectedness of I^n and $\mathbb{R}^n$ is immediate by Proposition 9.2.1 and Exercise 9.1.5. Define $S^n_\pm = \{x \in S^n \subseteq \mathbb{R}^{n+1} \mid \pm x_{n+1} \geq 0\}$. For $x \in S^n_\pm$, we have $x_{n+1} = \pm\sqrt{1 - \sum_{i=1}^n x_i^2}$. Thus the continuous maps $p_\pm : S^n_\pm \to D^n$, $(x_1, \ldots, x_{n+1}) \mapsto (x_1, \ldots, x_n)$ are bijective, and therefore homeomorphisms by Proposition 7.4.11. Thus $S^n_\pm \cong D^n \cong I^n$ is connected. Now, $S^n_+ \cap S^n_- = \{x \in S^n \mid x_{n+1} = 0\} \cong S^{n-1} \neq \emptyset$, so that $S^n = S^n_+ \cup S^n_-$ is connected by Lemma 9.1.2. (Connectedness of S^n also follows from the fact that S^n can be considered as a quotient space of I^n, cf. Exercise 7.8.21(iii).) ■

Connectedness can be used to distinguish spaces, i.e. prove them to be non-homeomorphic.

Definition 9.2.4 *If X is connected and $x \in X$ then x is called a <u>cut-point</u> if $X\backslash\{x\}$ is not connected.*

Exercise 9.2.5 (i) If $f : X \to Y$ is a homeomorphism, prove that $x \in X$ is cut-point $\Leftrightarrow$ $f(x) \in Y$ is cut-point.

(ii) Use (i) to prove that the spaces $[0,1], [0,1), (0,1)$ are pairwise non-homeomorphic.

Exercise 9.2.6 Prove the following claims:

(i) For all $n \geq 1$ we have $\mathbb{R}^n\backslash\{0\} \cong S^{n-1} \times (0, \infty)$. ($S^0 = \{\pm 1\}$.)

(ii) $\mathbb{R}^n\backslash\{x\}$ is connected if and only if $n \geq 2$.

(iii) $\mathbb{R}^n \not\cong \mathbb{R}$ when $n \geq 2$.

(iv) For $n \geq 2$ we have $S^1 \not\cong S^n$.

(v) For $n \geq 2$ we have $I \not\cong I^n$.

The following exercise provides a generalization of Proposition 9.2.1 to ordered spaces:

Exercise 9.2.7 Let $(X, \leq)$ be a totally ordered set. Prove that the order topology, cf. Definition 4.2.5, is connected if and only if both of the following hold:

(i) whenever $a < b$, we have $(a,b) \neq \emptyset$. (I.e. $a < c < b$ for some $c \in X$.)

(ii) $(X, \leq)$ is complete, i.e. every subset that is bounded above has a supremum.

Hint: For the 'if' part, adapt the proof of Proposition 9.2.1.

Exercise 9.2.8 Prove that for every continuous function $f : S^2 \to S^1$ <u>exactly one</u> of the following statements is true:

(i) f is surjective.

(ii) f is constant.

(iii) $f(S^2) \cong [0,1]$.

9.2.2 Intermediate value theorem and applications

The following shows that the intermediate value theorem from classical analysis is best understood as a consequence of the connectedness of intervals:

Corollary 9.2.9 (Intermediate Value Theorem) *Let $f \in C([a,b],\mathbb{R})$ and let $y \in [f(a), f(b)]$ or $y \in [f(b), f(a)]$ (depending on whether $f(a)$ or $f(b)$ is bigger). Then there is $x \in [a,b]$ such that $f(x) = y$.*

Proof. By Proposition 9.2.1, the space $[a,b]$ is connected. Thus by Lemma 9.1.1, $f([a,b])$ is connected and therefore, using Proposition 9.2.1 again, convex. Thus every y that lies between $f(a)$ and $f(b)$ is contained in $f([a,b])$, meaning that there is $x \in [a,b]$ such that $f(x) = y$. ■

Remark 9.2.10 1. The intermediate value theorem has the following immediate consequences, where $I = [0,1]$:

(i) If $f \in C(I,\mathbb{R})$ satisfies $f(0) \le 0,\ f(1) \ge 0$ then there is $x \in [0,1]$ with $f(x) = 0$.

(ii) If $f \in C(I,I)$ satisfies $f(0) = 0,\ f(1) = 1$ then $f(I) = I$.

It is easy to see that each of these statements implies Corollary 9.2.9.

2. Both (i) and (ii) have natural higher-dimensional versions, cf. Corollaries 10.3.1 and 10.3.8. □

Definition 9.2.11 *A point x such that $f(x) = x$ is a fixed point of f. A topological space (X,τ) has the fixed point property if every $f \in C(X,X)$ has a fixed point.*

Lemma 9.2.12 $[0,1]$ *has the fixed point property.*

Proof. Given $f : I \to I$, define $g : I \to \mathbb{R},\ x \mapsto x - f(x)$. Clearly, $g(0) \le 0$ and $g(1) \ge 0$. Thus there exists $x \in I$ such that $g(x) = 0$. This is equivalent to $f(x) = x$. ■

The following notion has many applications, in particular in algebraic topology:

Definition 9.2.13 *Let (X,τ) be a topological space and $Y \subseteq X$. A retraction of X to Y is a continuous map $r : X \to Y$ such that $r{\restriction}Y = \mathrm{id}_Y$.*

Remark 9.2.14 1. If $y_0 \in Y$ then the map $X \times Y \to X \times \{y_0\} \subseteq X \times Y,\ (x,y) \mapsto (x, y_0)$ is a retraction.

2. The map $r : x \mapsto \frac{x}{\|x\|}$ is a retraction $\mathbb{R}^n \backslash \{0\} \to S^{n-1}$, and by restriction we have a retraction $r : D^n \backslash \{0\} \to S^{n-1} = \partial D^n$. (In view of Exercise 9.2.6(i), both of these examples can be considered as applications of 1.)

3. Later we will prove that there is **no** retraction $r : D^n \to \partial D^n$ for any $n \ge 1$. For $n = 1$, this is easy: A retraction $r : [0,1] \to \{0,1\}$ is a map $r : I \to I$ fixing the endpoints, thus the intermediate value theorem, cf. the version given in Remark 9.2.10(ii), implies $r(I) = I$, contradicting $r(I) \subseteq \{0,1\}$. □

Lemma 9.2.15 *If X is Hausdorff and $r : X \to Y$ is a retraction then Y is closed.*

Proof. If $x \in Y$ then $r(x) = x$. If $x \in X \backslash Y$ then $r(x) \in Y$ implies $r(x) \ne x$. Thus $Y = \{x \in X \mid x = r(x)\}$, and the claim follows from Exercise 6.5.18(ii). ■

Lemma 9.2.16 *If $r : X \to Y \subseteq X$ is a retraction and X has the fixed point property then Y has the fixed point property.*

Proof. Given $f \in C(Y, Y)$, define $g = f \circ r \in C(X, X)$. Since X has the fixed point property, there is $x \in X$ such that $x = g(x) = f(r(x))$. Since f takes values in Y, we have $x \in Y$. Since r is a retraction, this implies $x = r(x)$. Thus $x = g(x) = f(r(x)) = f(x)$, so that x is a fixed point of f. Since $f \in C(Y, Y)$ was arbitrary, Y has the fixed point property. ■

The above results can be used to obtain a simple inverse function theorem:

Exercise 9.2.17 Let $f : [a, b] \to \mathbb{R}$ be continuous and strictly increasing, i.e. $x < x' \Rightarrow f(x) < f(x')$. Prove:

(i) There are real numbers $c \leq d$ such that $f([a, b]) = [c, d]$.

(ii) $f : [a, b] \to [c, d]$ is a bijection.

(iii) The inverse function f^{-1} is continuous, i.e. f is a homeomorphism.

Remark 9.2.18 1. If f is also differentiable with $f'(x) > 0$ everywhere, one proves that also $g = f^{-1}$ is differentiable with $g'(y) = 1/f'(f^{-1}(y))$. (This identity follows from $g \circ f(x) = x$ and the chain rule.)

2. Generalizing the proposition to higher dimensions is not trivial since there is no obvious analogue of monotonicity. But the following provides a hint: If $f : [a, b] \to \mathbb{R}$ is continuously differentiable (C^1) and there is an $x_0 \in (a, b)$ with $f'(x_0) \neq 0$ then by continuity $f' \neq 0$ on some open neighborhood (a', b') of x_0. Then $f \upharpoonright (a', b')$ is strictly increasing, thus $f \upharpoonright [a'', b'']$, where $a' < a'' < x_0 < b'' < b$, satisfies the assumption of Exercise 9.2.17. The condition $f'(x_0) \neq 0$ does generalize to higher dimensions. Cf. Appendix **??** for several inverse function theorems involving such conditions. □

9.2.3 n-th roots in $\mathbb{R}$ and $\mathbb{C}$

If F is a field, $x \in F$ and $n \in \mathbb{N}$ then by an n-th root of x we mean a $w \in F$ such that $w^n = x$.

Exercise 9.2.19 Use the Intermediate Value Theorem to prove:

(i) For $x > 0$ and $n \in \mathbb{N}$ there is a unique positive n-th root $x^{1/n}$.

(ii) Every polynomial $P \in \mathbb{R}[x]$ with real coefficients and odd degree has a zero.

(iii) Every $z \in \mathbb{C}$ has a square root $z^{1/2}$. Do not use the polar representation $z = re^{i\phi}$!

Corollary 9.2.20 *(i) If $n \in \mathbb{N}$ is odd then every $z \in \mathbb{C}$ has an n-th root $z^{1/n}$.*

(ii) For $z \in \mathbb{C}$, $n \in \mathbb{N}$ there is an n-th root $z^{1/n}$.

Proof. (i) In view of Exercise 9.2.19(ii), the claim is true for $z \in \mathbb{R}$ so that we may assume $z \notin \mathbb{R}$. Writing $z = |z|z'$, we have $|z'| = 1$, and now (i) or (ii) provides $|z|^{1/n}$. We thus may and will assume $z \notin \mathbb{R}$ and $|z| = 1$. By (iii) we can find $c \in \mathbb{C}$ with $c^2 = z$. Clearly $|c| = 1$ and $c \notin \mathbb{R}$. Define

$$P(x) = i(c(x + i)^n - \overline{c}(x - i)^n).$$

This is a polynomial in x, and for $x \in \mathbb{R}$ we have $\overline{P(x)} = P(x)$. Thus $P \in \mathbb{R}[x]$. The highest order term $ix^n(c - \bar{c})$ is non-zero since $c \notin \mathbb{R}$. Thus P is of odd order, and by (ii) there is $x \in \mathbb{R}$ with $P(x) = 0$. Thus $c(x+i)^n = \bar{c}(x-i)^n$. Since $|c| = 1$ implies $\bar{c} = c^{-1}$, we have

$$\left(\frac{x-i}{x+i}\right)^n = c^2 = z,$$

so that $(x-i)/(x+i)$ is an n-th root of z.

(ii) Write $n = 2^k m$ with m odd. Now use (i) and Exercise 9.2.19(iii) to define $w = (((z^{1/m})^{1/2})^{1/2})^{\cdots}$ (k square roots), which clearly satisfies $w^n = z$. ■

Remark 9.2.21 1. The above proof is from [64, Chapter 3, §5].

2. The standard proof of (ii) uses the fact that for every $z \in \mathbb{C}$ with $|z| = 1$ there is a $t \in \mathbb{R}$ with $z = \exp(it)$. Thus writing $z = r\exp(it)$ with $r = |z| > 0$, an n-th root of z is given by $r^{1/n}\exp(it/n)$. But the above argument is more elementary since it does not require the somewhat involved proof of surjectivity of the map $\mathbb{R} \to S^1, t \mapsto \exp(it)$.

3. The result of Exercise 9.2.19 was already used to prove the algebraic closedness of $\mathbb{C}$, cf. Theorem 7.7.57. In Exercise 13.7.19, another topological proof of that result will be given, using some more information about the map $t \mapsto \exp(it), \mathbb{R} \to S^1 \subseteq \mathbb{C}$, namely that the latter is a covering map. □

Chapter 10

Higher-dimensional connectedness

10.1 Introduction

We have just seen that one can get considerable milage out of the simple fact that the interval $I = [0,1]$ is connected, namely the intermediate value theorem, the fixed point property of I, and non-homeomorphisms $\mathbb{R}^n \not\cong \mathbb{R}$, $I^n \not\cong I$, $S^n \not\cong S^1$ for $n \neq 1$. The aim of this section is to show that all these results generalize to arbitrary (finite) dimension. (Versions of Brouwer's fixed point theorem even hold in infinite dimension!)

These generalizations are usually proven using (co)homology theory, which belongs to algebraic topology. As nicely stated by Spanier [213, p. 155], "*[Homology] measures higher dimensional connectedness, and some of the applications of homology are to prove higher dimensional analogues of results obtainable in low dimensions by using connectedness considerations.*"

However, (co)homology theory is a sophisticated machinery, whose proper definition and study takes weeks of lectures. It therefore is very satisfactory that one can actually base fairly elementary proofs of the desired higher-dimensional results on the essential idea contained in the cited phrase, which is using a notion of higher-dimensional connectedness:

Definition 10.1.1 *Let $I = [0,1]$ and $n \in \mathbb{N}$. The faces of the n-cube I^n are given by*

$$I_i^- = \{x \in I^n \mid x_i = 0\}, \qquad I_i^+ = \{x \in I^n \mid x_i = 1\} \qquad (i = 1, \ldots, n).$$

Theorem 10.1.2 (Higher connectedness of I^n) *Let $H_i^+, H_i^- \subseteq I^n$, $i = 1, \ldots, n$, be closed sets such that $I_i^\pm \subseteq H_i^\pm$ and $H_i^- \cup H_i^+ = I^n \ \ \forall i$. Then $\bigcap_i (H_i^- \cap H_i^+) \neq \emptyset$.*

For $n = 1$, this is just the statement that there is no clopen $C \subseteq [0,1]$ such that $0 \in C \not\ni 1$. This immediately follows from the connectedness of I proven earlier (and is equivalent to it). But for $n \geq 2$, the theorem is new, since it gives more than just the connectedness of I^n.

For $n = 2$, there is a fairly easy proof based on Theorem 13.2.4 and Exercise 13.1.15. But for arbitrary n, the proof requires a certain amount of combinatorics (2.5 pages). It should not be conceiled that the mentioned combinatorics are closely related to those of cohomology theory, of which they in a sense represent the essential core, cf. [127]. The same is true, cf. [128, 129], for proofs of Brouwer's fixed point theorem using calculus methods like [63, Section V.12] or [155].

10.2 The cubical Sperner lemma. Proof of Theorem 10.1.2

We begin with some more notations:

- For $k \in \mathbb{N}$, we put $\mathbb{Z}_k = k^{-1}\mathbb{Z} = \{n/k \mid n \in \mathbb{Z}\}$. Clearly $\mathbb{Z}_k^n \subseteq \mathbb{R}^n$.
- $e_i \in \mathbb{Z}_k^n$ is the vector whose coordinates are all zero, except the i-th, which is $1/k$.
- $C(k) = I^n \cap \mathbb{Z}_k^n = \left\{0, \frac{1}{k}, \ldots, \frac{k-1}{k}, 1\right\}^n$. (The combinatorial n-cube.)
- $C_i^\pm(k) = I_i^\pm \cap \mathbb{Z}_k^n$. (The faces of the combinatorial n-cube $C(k)$.)
- $\partial C(k) = \bigcup_i (C_i^+(k) \cup C_i^-(k))$. (The boundary of the combinatorial n-cube $C(k)$.)
- A subcube of $C(k)$ is a set $C = \{z_0 + \sum_{i=1}^n a_i e_i \mid a \in \{0,1\}^n\} \subseteq C(k)$, where $z_0 \in C(k)$.

Proposition 10.2.1 (Cubical version of Sperner's Lemma) [1] *Let $\varphi : C(k) \to \{0, \ldots, n\}$ be a map such that*

(i) $x \in C_i^-(k) \Rightarrow \varphi(x) < i$.

(ii) $x \in C_i^+(k) \Rightarrow \varphi(x) \neq i - 1$.

Then there is a subcube $C \subseteq C(k)$ such that $\varphi(C) = \{0, \ldots, n\}$.

Proof of Theorem 10.1.2 assuming Proposition 10.2.1. We define $F_0 = I^n$ and $F_i = H_i^+ \backslash I_i^-$ for all $i \in \{1, \ldots, n\}$. Now define a map $\varphi : I^n \to \{0, \ldots, n\}$ by

$$\varphi(x) = \max\left\{ j : \ x \in \bigcap_{k=0}^{j} F_k \right\}.$$

Since $I_i^- \cap F_i = \emptyset$, we have $x \in I_i^- \ \Rightarrow\ \varphi(x) < i$. On the other hand, if $x \in I_i^+$ then $\varphi(x) \neq i - 1$. Namely, $\varphi(x) = i - 1$ would mean that $x \in \bigcap_{k=0}^{i-1} F_k$ and $x \notin F_i$. But this is impossible since $x \in I_i^+ \subseteq H_i^+$ and $I_i^+ \cap I_i^- = \emptyset$, thus $x \in F_i$.

By the above, the restriction of φ to $C(k) \subseteq I^n$ satisfies the assumptions of Proposition 10.2.1. Thus for every $k \in \mathbb{N}$ there is a subcube $C_k \subseteq C(k)$ such that $\varphi(C_k) = \{0, \ldots, n\}$. Now, if $\varphi(y) = i \in \{1, \ldots, n\}$ then $y \in F_i = H_i^+ \backslash I_i^- \subseteq H_i^+$. On the other hand, if $\varphi(x) = i-1 \in \{0, \ldots, n-1\}$ then $x \notin F_i = H_i^+ \backslash I_i^-$, which is equivalent to $x \notin H_i^+ \vee x \in I_i^-$. The first alternative implies $x \in H_i^-$ (since $H_i^+ \cup H_i^- = I^n$), as does the second (since $I_i^- \subseteq H_i^-$). In either case, $x \in H_i^-$. Combining these facts, we find that $\varphi(C_k) = \{0, \ldots, n\}$ implies $C_k \cap H_i^+ \neq \emptyset \neq C_k \cap H_i^- \ \forall i$, thus the subcube C_k meets all the $H_i^\pm$.

We clearly have $\mathrm{diam}(C_k) = \sqrt{n}/k$, and since k was arbitrary, Lemma 7.7.46 applied to $X = I^n,\ S_k = C_k$ and $\{K_1, \ldots, K_{2n}\} = \{H_i^\pm\}$ completes the proof. ■

Definition 10.2.2 *An n-simplex in $\mathbb{Z}_k^n$ is an ordered set $S = [z_0, \ldots, z_n] \subseteq \mathbb{Z}_k^n$ such that*

$$z_1 = z_0 + e_{\alpha(1)}, \quad z_2 = z_1 + e_{\alpha(2)}, \quad \ldots, \quad z_n = z_{n-1} + e_{\alpha(n)},$$

where α is a permutation of $\{1, \ldots, n\}$. The subset $F_i(S) = [z_0, \ldots, z_{i-1}, z_{i+1}, \ldots, z_n] \subseteq S$, where $i \in \{0, \ldots, n\}$, is called the i-th face of the n-simplex S.

A finite ordered set $F \subseteq \mathbb{Z}_k^n$ is called a face if it is a face of some simplex.

Note that the faces $F_i(S)$ are $(n-1)$-simplices in the above sense only if $i = 0$ or $i = n$.

[1] Emanuel Sperner (1905-1980). German mathematician.

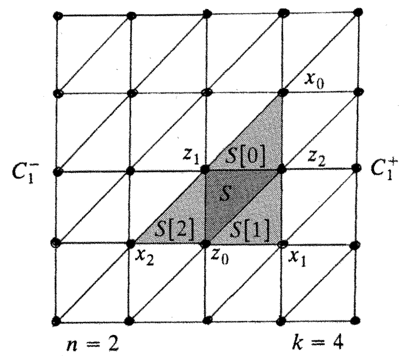


Figure 10.1: The neighbors of a simplex in $\mathbb{Z}_4^2$. From [148] in Amer. Math. Monthly.

Lemma 10.2.3 *(i) Let $S = [z_0, \ldots, z_n] \subseteq \mathbb{Z}_k^n$ be an n-simplex. Then for every $i \in \{0, \ldots, n\}$ there is a unique n-simplex $S[i]$, the <u>i-th neighbor</u> of S, such that $S \cap S[i] = F_i(S)$.*

(ii) If $S \subseteq C(k)$ and $i \in \{0, \ldots, n\}$ then $S[i] \subseteq C(k)$ holds if and only if $F_i(S) \not\subseteq \partial C(k)$.

Proof. (i) Existence: We define the i-th neighbor $S[i]$ as follows:

(a) $S[0] = [z_1, \ldots, z_n, x_0]$, where $x_0 = z_n + (z_1 - z_0)$.

(b) $0 < i < n$: Take $S[i] = [z_0, \ldots, z_{i-1}, x_i, z_{i+1}, \ldots, z_n]$, where $x_i = z_{i-1} + (z_{i+1} - z_i)$.

(c) $S[n] = [x_n, z_0, \ldots, z_{n-1}]$, where $x_n = z_0 - (z_n - z_{n-1})$.

It is obvious that $\#(S \cap S[i]) = n$ in all three cases. In the three cases, the distances between consecutive points of $S[i]$ are given by (a) $e_{\alpha(2)}, \ldots, e_{\alpha(n)}, e_{\alpha(1)}$, (c) $e_{\alpha(n)}, e_{\alpha(1)}, \ldots, e_{\alpha(n-1)}$, and (b) $e_{\alpha(1)}, \ldots, e_{\alpha(i-1)}, e_{\alpha(i+1)}, e_{\alpha(i)}, e_{\alpha(i+2)}, \ldots, e_{\alpha(n)}$. (Figure 10.1 should make this quite clear.) Thus $S[i]$ is a legal n-simplex for each $i \in \{0, \ldots, n\}$. Uniqueness: It remains to show that these are the only ways of defining $S[i]$ consistently with $S \cap S[i] = \{z_0, \ldots, z_{i-1}, z_{i+1}, \ldots, z_n\}$. In the cases $i = 0$ or $i = n$, the latter condition implies that $S[i]$ has a string of n consecutive z_i's in common with S, and therefore also their differences given by $n-1$ mutually different vectors e_j. This means that only one such vector is left, and the only way to use it so that $S[i]$ is an n-simplex different from S is to use it at the other end of the string of z's. This shows the uniqueness of the above definitions in cases (a) and (c). In the case $0 < i < n$, S and $S[i]$ have two corresponding substrings of z's. A little thought shows that the order of these two substrings must be the same in $S[i]$ as in S, so that all we can do is exchange two adjacent difference vectors $e_{\alpha(i)}, e_{\alpha(i+1)}$, as done in the definition of $S[i]$ in case (b).

(ii) We must check whether $S[i] \subseteq C(k)$, which amounts to checking whether the new point x_i is in I^n. In case (a), we have $S = [z_1 - e_{\alpha(1)}, z_1, z_2, \ldots, z_n]$ and $S(0) = [z_1, \ldots, z_n, z_n + e_{\alpha(1)}]$. If $F_0(s) = [z_1, \ldots, z_n] \subseteq C_j^\varepsilon(k)$ then $z_1, \ldots, z_n$ all have the same j-coordinate c, thus we must have $\alpha(1) = j$ and $c = 1$ (since $S \subseteq I^n$). But then $S[0] \not\subseteq I^n$. Conversely, if both S and $S[0]$ are in I^n, then z_1 must have $\alpha(1)$-th coordinate > 0 and z_n must have $\alpha(1)$-th coordinate < 1. All other coordinates of $z_1, \ldots, z_n$ are non-constant since the vectors $e_{\alpha(2)}, \ldots, e_{\alpha(n)}$ appear as differences. Thus $F_0(S)$ is not contained in any face I_i^ε. The cases (b) and (c) are checked similarly. ■

Proof of Proposition 10.2.1. For later use, we note the following fact (*): If $S \subseteq I^n$ satisfies $\varphi(S \cap I_i^\varepsilon) = \{0, \ldots, n-1\}$ then $i = n$ and $\varepsilon = -$. [The statement $\varphi(S \cap I_i^\varepsilon) = \{0, \ldots, n-1\}$ is contradicted by assumption (ii) if $\varepsilon = +$ and by (i) if $\varepsilon = -$ and $i < n$.]

We call a subset $S \subseteq C(k)$ with $l+1$ elements <u>full</u> if $\varphi(S) = \{0, \ldots, l\}$. By (vi), a full n-simplex S meets all $H_i^\pm$. We will prove that the number N_k of full n-simplices in $C(k)$ is odd, thus non-zero,

for all k. The proof of $N_k \equiv 1 \pmod 2$ proceeds by induction over the dimension n of $C(k)$ (for fixed k). For $n = 0$ we have $C(k) = \{0\}$, and there is exactly one full n-simplex, namely $S = [z_0 = 0]$. Thus $N_0 = 1$.

For an n-simplex $S \subseteq C(k)$, let $N(S)$ denote the number of full $(n-1)$-faces of S. If S is full then $N(S) = 1$. [Since $\varphi(S) = \{0, \dots, n\}$ and the only full $(n-1)$-face is obtained by omitting the unique z_i for which $\varphi(z_i) = n$.] If S is not full then $N(S) = 0$ in the case $\{0, \dots, n-1\} \not\subseteq \varphi(S)$ [since omitting a z_i cannot give a full (n-1)-face] or $N(S) = 2$ in the case $\varphi(S) = \{0, \dots, n-1\}$ [since there are $i \neq i'$ such that $z_i = z_{i'}$, so that S becomes full upon omission of either z_i or $z_{i'}$]. Thus

$$N_k \equiv \sum_S N(S) \pmod 2, \tag{10.1}$$

where the summation extends over all n-simplices in $C(k)$.

Now by the Lemma, an $(n-1)$-face $F \subseteq C(k)$ belongs to one or two n-simplices in $C(k)$, depending on whether $F \subseteq \partial C(k)$ or not. Thus only the full faces $F \subseteq \partial C(k)$ contribute to (10.1):

$$N_k \equiv \#\{F \subseteq \partial C(k) \text{ full } (n-1)\text{−face}\} \pmod 2.$$

If $F \subseteq \partial C(k)$ is a full $(n-1)$-face then (*) implies $F \subseteq C_n^-(k)$. We can identify $C_n^-(k) = C(k) \cap I_n^-$ with $C_{n-1}(k)$, and under this identification F is a full $(n-1)$-simplex in $\mathbb{Z}_k^{n-1}$. Thus $N_k \equiv N_{k-1} \pmod 2$. By the induction hypothesis, N_{k-1} is odd, thus N_k is odd.

Thus there is a full n-simplex $S = [z_0, \dots, z_n]$, and if $C = \{z_0 + \sum_{i=1}^n a_i e_i \mid a \in \{0,1\}^n\}$ we have $S \subseteq C$ so that $\varphi(C) = \{0, \dots, n\}$. ■

Remark 10.2.4 Sperner's lemma in its original form [214, 142] works with standard simplices, i.e. $\Delta_n = \{(x_0, \dots, x_n) \in \mathbb{R}_{\geq 0}^{n+1} \mid x_0 + \dots + x_n = 1\}$. Cubical versions of Sperner's lemma were first considered in [147] and differently in [239, Lemma 1]. The proof given above is due to Kulpa [148]. □

10.3 The theorems of Poincaré-Miranda and Brouwer

Theorem 10.1.2 has many important classical results about continuous functions on I^n as corollaries, for all $n \in \mathbb{N}$. Since it expresses a higher-dimensional form of connectedness, it implies a higher-dimensional generalization of the version of the **intermediate value theorem** given in Remark 9.2.10(i):

Corollary 10.3.1 (Poincaré-Miranda theorem) [2] *Let $f = (f_1, \dots, f_n) \in C(I^n, \mathbb{R}^n)$ satisfy $f_i(I_i^-) \subseteq (-\infty, 0]$, $f_i(I_i^+) \subseteq [0, \infty)$ for all i. Then there is $x \in I^n$ such that $f(x) = 0$.*

Proof. Put $H_i^- = f_i^{-1}((-\infty, 0]), H_i^+ = f_i^{-1}([0, \infty))$. Then clearly $I_i^\pm \subseteq H_i^\pm$ and $H_i^- \cup H_i^+ = I^n$, for all i. By Theorem 10.1.2, there exists $x \in \bigcap_i (H_i^- \cap H_i^+)$. This means $f_i(x) \in (-\infty, 0] \cap [0, +\infty)$ for all i, thus $f(x) = 0$. ■

Remark 10.3.2 The one-dimensional intermediate value theorem provides points $x^1, \dots, x^n \in I^n$ such that $f_i(x^i) = 0$ for each i. But the existence of a point x where all $f_i(x)$ vanish simultaneously is much deeper! □

[2] Henri Poincaré (1854-1912) was the greatest French mathematician of around 1900 (often compared to Hilbert). He was one of the fathers of algebraic topology, but he worked in almost all fields of mathematics and mathematical physics.

Exercise 10.3.3 Deduce Theorem 10.1.2 from Corollary 10.3.1.

Just as for $n = 1$, the intermediate value theorem implies the fixed point property of I^n:

Corollary 10.3.4 (Brouwer's fixed point theorem) [3] *Let $g \in C(I^n, I^n)$. Then there exists $x \in I^n$ such that $g(x) = x$. (I.e., I^n has the fixed point property.)*

Proof. Put $f(x) = x - g(x)$. Now $x_i = 0 \Rightarrow f_i(x) = 0 - g_i(x) \le 0$ and $x_i = 1 \Rightarrow f_i(x) = 1 - g_i(x) \ge 0$. Thus the assumptions of Corollary 10.3.1 are satisfied, so that there is an $x \in I^n$ for which $f(x) = 0$. Thus $g(x) = x$. ■

Remark 10.3.5 1. The history of these two results is quite convoluted and interesting. See the introduction of [148] for a glimpse. Brouwer's proof was given in 1912, cf. [33].

2. Brouwer's fixed point theorem is often deduced from the non-existence of retractions $D^n \to S^{n-1}$, see below, but the deduction from the Poincaré-Miranda theorem is simpler, and it is just the higher-dimensional version of the proof for $n = 1$.

3. The fixed point theorem is the best known of a whole cluster of results that are mutually 'equivalent' in the sense that each is easily deducible from the others, while it is much harder to prove that any one of these statements actually is true.

These results have countlessly many applications in pure and applied mathematics. E.g., they can be used to prove the Perron-Frobenius theorem of linear algebra, cf. e.g. [24], and the existence of Nash equilibria in game theory, cf. [184], as well as many results in the 'general equilibrium theory' of Walras. We will encounter several applications within topology, e.g. Exercise 10.3.11 and the 'invariance of dimension' (Corollary 10.4.7). □

Corollary 10.3.6 *Let $X \subseteq \mathbb{R}^n$ be compact and convex. Then X has the fixed point property.*

Proof. By Proposition 7.7.62, there is a homeomorphism $X \xrightarrow{\cong} D^m \cong I^m$ for some $m \le n$. Now use Corollary 10.3.4. ■

Remark 10.3.7 1. The convexity assumption cannot be omitted. E.g., the reflection map $S^n \to S^n$, $x \mapsto -x$ has no fixed point.

2. Corollary 10.3.6 has extensions to infinite dimensional spaces of increasing degree of generality, cf. Section G.4 for the simplest. Such results are important in non-linear functional analysis and have applications to the theory of partial differential equations. □

Corollary 10.3.8 *Let $g \in C(I^n, I^n)$ satisfy $g(I_i^\pm) \subseteq I_i^\pm$ $\forall i$. Then $g(I^n) = I^n$.*

Proof. Let $p \in I^n$, and put $f(x) = g(x) - p$. Now $x_i = 0 \Rightarrow f_i(x) = 0 - p \le 0$ and $x_i = 1 \Rightarrow f_i(x) = 1 - p \ge 0$. Thus f satisfies the assumptions of Corollary 10.3.1, so that there is $x \in I^n$ with $f(x) = 0$. This means $g(x) = p$, so that g is surjective. ■

Also the version of the **intermediate value theorem** given in Remark 9.2.10(ii) and the **no-retraction** result (Remark 9.2.14.3) generalize to higher dimensions:

Corollary 10.3.9 *Let $X \subseteq \mathbb{R}^n$ be compact and convex with $X^0 \neq \emptyset$. Then*

(i) If $f \in C(X, X)$ satisfies $f \upharpoonright \partial X = \mathrm{id}$ then $f(X) = X$.

[3] Luitzen Egbertus Jan Brouwer (1881-1966). Dutch mathematician. Many important contributions to topology. Founder of a controversial philosophy of mathematics (intuitionism).

(ii) There are no retractions $X \to \partial X$.

Proof. We prove the claims for $X = I^n$. The general result follows from the homeomorphism $(X \supseteq \partial X) \cong (D^n \supseteq \partial D^n = S^{n-1})$ proven in Section 7.7.6.

(i) If $g \in C(I^n, I^n)$ satisfies $g \restriction \partial I^n = \mathrm{id}$ then $g(I_i^{\pm}) \subseteq I_i^{\pm}$ $\forall i$, thus Corollary 10.3.8 applies.

(ii) If $r : X \to \partial X$ is a retraction, it satisfies the hypothesis of (i), thus satisfies $r(X) = X$, which is absurd since $X^0 \neq \emptyset$ implies $X \neq \partial X$. ■

The preceding results generalize to compact subsets of $\mathbb{R}^n$ that are not necessarily convex:

Corollary 10.3.10 *Let $X \subseteq \mathbb{R}^n$ be compact with $X^0 \neq \emptyset$.*

(i) If $f \in C(X, X)$ satisfies $f \restriction \partial X = \mathrm{id}_{\partial X}$ then $f(X) = X$.

(ii) There exists no retraction $r : X \to \partial X$.

(If $X^0 = \emptyset$ then $X = \partial X$, in which case (i) is trivially true and (ii) is false.)

Proof. (i) Choose $a > 0$ such that $X \cup f(X) \subseteq (-a, a)^n$. Let $C = [-a, a]^n$ and define $\widehat{f} : C \to C$ by $\widehat{f}(x) \restriction X = f$ and $\widehat{f}(x) = x$ for $x \in C \backslash X$. Since X and $\overline{C \backslash X}$ are closed, $\widehat{f}$ is continuous on both subsets of C and both prescriptions coincide on the intersection $X \cap \overline{C \backslash X} = \partial X$, $\widehat{f}$ is continuous, cf. Exercise 6.2.5. Furthermore, $\widehat{f} \restriction \partial C = \mathrm{id}$ since $\partial C \cap X = \emptyset$. Now Corollary 10.3.9 implies $\widehat{f}(C) = C$, and since $\widehat{f}(C \backslash X) \cap X = \emptyset$, we have $f(X) = X$.

(ii) This follows from (i) in the same way as in the preceding corollary. ■

Brouwer's theorem has a nice application to <u>vector fields</u>, which for our purposes just are continuous maps $\mathbb{R}^n \supseteq X \to \mathbb{R}^n$:

Exercise 10.3.11 Let $f : D^n \to \mathbb{R}^n$ be a continuous function satisfying $f(x) \cdot x > 0$ for all $x \in S^{n-1} = \partial D^n$ (or $f(x) \cdot x < 0$ for all $x \in S^{n-1}$). Prove:

(i) There is a constant $0 \neq \lambda \in \mathbb{R}$ such that $g_\lambda(x) = x - \lambda f(x)$ maps D^n into itself.

(ii) There is a $x \in (D^n)^0$ such that $f(x) = 0$.

(iii) If $n \geq 2$, the hypothesis can be weakened to $f(x) \cdot x \neq 0$ for all $x \in S^{n-1}$. (For $n = 1$ this fails, as $f \equiv 1 : D^1 = [0, 1] \to \mathbb{R}$ shows.)

For a different approach (but also ultimately relying on Brouwer's theorem) to proving the existence of zeros of vector fields, cf. Theorem 13.1.16.

10.4 The dimensions of I^n and $\mathbb{R}^n$

Definition 10.4.1 *If $A, B, C \subseteq X$ are closed sets such that $X \backslash C = U \cup V$, where U, V are disjoint open sets such that $A \subseteq U$ and $B \subseteq V$, we say that C <u>separates</u> A and B.*

The following result plays an essential rôle in virtually all accounts of dimension theory. While it is usually derived from Brouwer's fixed point theorem, we obtain it more directly as an obvious corollary of Theorem 10.1.2.

Corollary 10.4.2 *Whenever $C_1, \ldots, C_n \subseteq I^n$ are closed sets such that C_i separates I_i^- and I_i^+ for each i, then $\bigcap_i C_i \neq \emptyset$.*

Proof. In view of Definition 10.4.1, we have open sets $U_i^\pm$ such that $I_i^\pm \subseteq U_i^\pm$, $U_i^+ \cap U_i^- = \emptyset$ and $U_i^+ \cup U_i^- = X \backslash C_i$ for all i. Define $H_i^\pm = U_i^\pm \cup C_i$. Then $X \backslash H_i^\pm = U_i^\mp$, thus $H_i^\pm$ is closed. By construction, $I_i^\pm \subseteq H_i^\pm$ and $H_i^+ \cup H_i^- = I^n$, $H_i^+ \cap H_i^- = C_i$, for all i. Now Theorem 10.1.2 gives $\bigcap_i C_i = \bigcap_i (H_i^- \cap H_i^+) \neq \emptyset$. ■

The preceding result will provide a lower bound on the dimension of I^n. The next result, taken from [169], will provide the upper bound:

Proposition 10.4.3 *Let $A_1, B_1, \ldots, A_{n+1}, B_{n+1} \subseteq I^n$ be closed sets such that $A_i \cap B_i = \emptyset\ \forall i$. Then there exist closed sets C_i separating A_i and B_i for all i and satisfying $\bigcap_i C_i = \emptyset$.*

Proof. Pick real numbers $r_1, r_2, \ldots$ such that $r_i - r_j \notin \mathbb{Q}$ for $i \neq j$. (It suffices to take $r_k = k\sqrt{2}$.) Then the sets $E_i = r_i + \mathbb{Q}$ are mutually disjoint dense subsets of $\mathbb{R}$.

Let $A, B \subseteq I^n$ be disjoint closed sets and $E \subseteq \mathbb{R}$ dense. Then for every $x \in A$ we can find an open neighborhood $U_x = I^n \cap \prod_{i=1}^n (a_i, b_i)$ with $a_i, b_i \in E$ such that $\overline{U_x}$ is disjoint from B. Since $A \subseteq I^n$ is closed, thus compact, there are $x_1, \ldots, x_k \in A$ such that $U = U_{x_1} \cup \cdots \cup U_{x_k} \supseteq A$. Now $C = \partial U \subseteq I^n$ is closed and $X \backslash C = U \cup V$, where $V = I^n \backslash \overline{U}$. U, V are open and disjoint such that $A \subseteq U, B \subseteq V$, thus C separates A and B. If $x \in \partial(I^n \cap \prod_i (a_i, b_i))$ then at least one of the coordinates x_i of x equals a_i or b_i, and thus is in E. Now, $C = \partial U \subseteq \partial U_{x_1} \cup \cdots \cup \partial U_{x_k} \subseteq \{x \in I^n \mid \exists j \in \{1, \ldots, n\} : x_j \in E\}$.

We can thus find, for each pair (A_i, B_i) a closed set $C_i \subseteq \{x \in I^n \mid \exists j : x_j \in E_i\}$ that separates A_i and B_i. Let now $x \in \bigcap_i C_i$. Then for every $i \in \{1, \ldots, n+1\}$ there is a $j_i \in \{1, \ldots, n\}$ such that $x_{j_i} \in E_i$. By the pigeonhole principle there are $i, i' \in \{1, \ldots, n+1\}$ such that $i \neq i'$ and $j_i = j_{i'} = j$. But this means that $x_j \in E_i \cap E_{i'} \in \emptyset$, which is absurd. Thus $\bigcap_i C_i = \emptyset$. ■

Proposition 10.4.3 should be compared with Corollary 10.4.2. In order to this systematically, the following is convenient:

Definition 10.4.4 *Let X be a topological space. We define the* <u>separation-dimension</u> $s\text{-}\dim(X) \in \{-1, 0, 1, \ldots, \infty\}$ *as follows:*

- *We put $s\text{-}\dim(X) = -1$ if and only if $X = \emptyset$.*
- *If $X \neq \emptyset$ and $n \in \mathbb{N}_0$, we say that $s\text{-}\dim(X) \leq n$ if, given closed sets $A_1, B_1, \ldots, A_{n+1}, B_{n+1}$ such that $A_i \cap B_i = \emptyset\ \forall i$, there exist closed C_i separating A_i and B_i and satisfying $\bigcap_i C_i = \emptyset$. (This is consistent: If $s\text{-}\dim(X) \leq n$ and $n < m$ then $s\text{-}\dim(X) \leq m$.)*
- *If $s\text{-}\dim(X) \leq n$ holds, but $s\text{-}\dim(X) \leq n-1$ does not, we say $s\text{-}\dim(X) = n$.*
- *If there is no $n \in \mathbb{N}$ such that $s\text{-}\dim(X) \leq n$ then $s\text{-}\dim(X) = \infty$.*

Lemma 10.4.5 *If X and Y are homeomorphic then $s\text{-}\dim(X) = s\text{-}\dim(Y)$.*

Proof. Obvious since $s\text{-}\dim(X)$ is defined in terms of the topology of X. ■

Theorem 10.4.6 $s\text{-}\dim(I^n) = n$.

Proof. In view of Definition 10.4.4, Proposition 10.4.3 amounts to the statement $s\text{-}\dim(I^n) \leq n$. On the other hand, it is clear that $s\text{-}\dim(X) \geq n$ holds if and only if there are closed sets $A_1, B_1, \ldots, A_n, B_n \subseteq X$ satisfying $A_i \cap B_i = \emptyset\ \forall i$ such that any closed sets C_i separating A_i and B_i satisfy $\bigcap_i C_i \neq \emptyset$. This is exactly what is asserted for $X = I^n$ by Corollary 10.4.2. ■

The preceding result shows that the intuitive dimension of the cube I^n, namely n, can indeed be recovered from the topology of I^n via the notion of separation-dimension. Since the latter is a topological invariant, we have proven of the 'invariance of dimension':

Corollary 10.4.7 *We have $I^n \cong I^m$ if and only if $n = m$.*

Remark 10.4.8 The first correct proof of this result was given by Brouwer [32] in 1911. His proof did not (explicitly) involve associating a topologically defined dimension to I^n. This was for the first time done in [34]. With some delay, this led to the rapid development of 'dimension theory'. This is a branch of point-set topology (as opposed to algebraic topology), but in order to compute the dimension of Euclidean spaces, one cannot avoid invoking some result equivalent to Brouwer's fixed point theorem. Cf. [76] for a comprehensive contemporary account of dimension theory. □

In order to compute the dimension of S^n and $\mathbb{R}^n$, some more work is needed.

Proposition 10.4.9 *If $Y \subseteq X$ is closed, we have $s\text{-dim}(Y) \leq s\text{-dim}(X)$.*

Proof. Assume $s\text{-dim}(X) \leq n$, and let $A_i, B_i \subseteq Y$, $i = 1, \ldots, n+1$, be closed sets satisfying $A_i \cap B_i = \emptyset$. Since Y is closed, A_i, B_i are closed in X. Thus there exist closed sets $C_i \subseteq X$ separating A_i and B_i and $\bigcap_i C_i = \emptyset$. Now the sets $D_i = C_i \cap Y \subseteq Y$ are closed and separate A_i and B_i in Y. Thus $s\text{-dim}(Y) \leq n$. ■

The above already implies $s\text{-dim}(\mathbb{R}^n) \geq n$ and $n \leq s\text{-dim}(S^n) \leq n+1$. (Since S^n has the closed subspaces $S^n_\varepsilon = \{x \in S^n \subseteq \mathbb{R}^{n+1} \mid \varepsilon x_{n+1} \geq 0\} \cong D^n \cong I^n$ and, in turn, is a closed subspace of $[-1,1]^{n+1} \cong I^{n+1}$.) In order to go further, we need the following:

Proposition 10.4.10 (Sum Theorem) *Let X be normal. If $X = \bigcup_{k \in \mathbb{N}} Y_k$, where each $Y_k \subseteq X$ is closed with $s\text{-dim}(Y_k) \leq n$ then $s\text{-dim}(X) \leq n$.*

Proof. Omitted. The only proof known to the author proceeds by proving the 'sum-theorem' for the covering dimension $\text{cov}(X)$ defined in Section 10.5, and the identity $s\text{-dim}(X) = \text{cov}(X)$. These results are Theorems 3.1.8 and 3.2.6 in [76], respectively. It would be desirable to work out a direct proof of the sum-theorem for the separation dimension s-dim. ■

Corollary 10.4.11 *$s\text{-dim}(\mathbb{R}^n) = s\text{-dim}(S^n) = n$. Thus $\mathbb{R}^n \cong \mathbb{R}^m \;\Leftrightarrow\; n = m \;\Leftrightarrow\; S^n \cong S^m$.*

Proof. By Proposition 10.4.9, $s\text{-dim}(\mathbb{R}^n) \geq s\text{-dim}(I^n) = n$. On the other hand, $\mathbb{R}^n$ is the union of countably many closed cubes $\prod_i [a_i, a_i+1]$ with $a = (a_i) \in \mathbb{Z}^n$. Using Theorem 10.4.6 and Proposition 10.4.10, we have $s\text{-dim}(\mathbb{R}^n) \leq n$. Similarly, $S^n = S^n_+ \cup S^n_-$. Since $S^n_\varepsilon \cong D^n$, we have $s\text{-dim}(S^n_\varepsilon) = n$, and Proposition 10.4.10 gives $s\text{-dim}(S^n) \leq n$. ■

Corollary 10.4.12 *If X is T_6 and $U \subseteq X$ is open then $s\text{-dim}(U) \leq s\text{-dim}(X))$.*

Proof. A T_6 space is normal and every closed set is G_δ. Thus every open set is F_σ, so that there are closed sets $\{C_i\}_{i \in \mathbb{N}}$ such that $U = \bigcup_i C_i$. By Proposition 10.4.9, $s\text{-dim}(C_i) \leq s\text{-dim}(X)$ $\forall i$. Now Proposition 10.4.10 implies $s\text{-dim}(U) = s\text{-dim}(\bigcup_i C_i) \leq n$. ■

Since $\mathbb{R}^n \cong (0,1)^n$, which is an open subspace of $[0,1]^n$, this again gives $s\text{-dim}(\mathbb{R}^n) \leq n$.

Exercise 10.4.13 Prove:

(i) $s\text{-dim}(X) = 0$ holds if and only if $X \neq \emptyset$ and for any pair of disjoint closed sets $A, B \subseteq X$ there exists a clopen set C such that $A \subseteq C$ and $B \subseteq X \backslash C$. ($T_1$-spaces with this property will be called strongly zero-dimensional, cf. Definition 11.1.38.)

(ii) If X is T_1 and $s\text{-dim}(X) = 0$ then every point has a neighborhood base of clopen sets. (Thus X is zero-dimensional in the sense of Definition 11.1.15.)

Remark 10.4.14 1. When it comes to distinguishing topological spaces, the power of dimension theory is quite limited. After all, $\dim(I^n) = \dim(S^n) = \dim(\mathbb{R}^n) = n$, while $I^n, S^n, \mathbb{R}^n$ are mutually non-homeomorphic. (E.g., since I^n, S^n are compact while $\mathbb{R}^n$ is not, and $I^n, \mathbb{R}^n$ are contractible, but S^n is not, cf. Theorem 13.1.11.) Another huge disadvantage of topological dimensions is that they are not functorial. The more useful invariants of topological spaces are functorial. Cf. the functors $\pi_c, \pi_0, \pi_1, \Pi_1$ defined in these notes.

2. The methods used to prove Theorems 10.1.2 and its applications like Corollary 10.3.1 and Theorem 10.4.6 can be pushed further to prove other results that are usually obtained using algebraic topology, in particular 'domain invariance'. Cf. e.g. [149, 125]. □

10.5 Other notions of topological dimension

Our definition of the 'separation-dimension' of a space X is rather non-standard (but motivated by the treatment in [169]). In dimension theory, a branch of general topology, various other notions of dimension have been defined and studied intensely. We briefly consider those definitions and summarize some known comparison results. Cf. e.g. [123, 76] for the full story.

Definition 10.5.1 *Let (X, τ) be a topological space. We define the* <u>small inductive dimension</u> $\mathrm{ind}(X) \in \{-1, 0, 1, \ldots, \infty\}$ *as follows:*

- *We put $\mathrm{ind}(X) = -1$ if and only if $X = \emptyset$.*
- *If $X \neq \emptyset$ and $n \in \mathbb{N}_0$, we say that $\mathrm{ind}(X) \leq n$ if, whenever $x \in U \in \tau$, there exists $V \in \tau$ such that $x \in V \subseteq \overline{V} \subseteq U$ and $\mathrm{ind}(\partial V) \leq n - 1$.*
- *If $\mathrm{ind}(X) \leq n$ holds, but $\mathrm{ind}(X) \leq n - 1$ does not, we say $\mathrm{ind}(X) = n$.*
- *If there is no $n \in \mathbb{N}$ such that $\mathrm{ind}(X) \leq n$ then $\mathrm{ind}(X) = \infty$.*

Remark 10.5.2 1. Since $\partial V = \emptyset$ if and only if V is clopen, it is immediate that a T_1-space X satisfies $\mathrm{ind}(X) = 0$ if and only if it is zero-dimensional in the sense of Definition 11.1.15.

2. Since every neighborhood of $x \in \mathbb{R}^n$ contains a closed ball $\overline{B}(x, \varepsilon)$, whose boundary is an $(n-1)$-sphere, it is very easy to see that $\mathrm{ind}(\mathbb{R}^n) \leq n$.

3. The small inductive dimension is the most intuitive of the notions of topological dimension. But it has some technical disadvantages, which motivate the next definition. □

Definition 10.5.3 *The* <u>large inductive dimension</u> $\mathrm{Ind}(X) \in \{-1, 0, 1, \ldots, \infty\}$ *is defined like ind(X) except that in the second clause, $x \in U \in \tau$ is replaced by $C \subseteq U \in \tau$, where C is closed.*

Remark 10.5.4 1. It is obvious that $\mathrm{ind}(X) \leq \mathrm{Ind}(X)$ for every T_1-space X.

2. We have $\mathrm{Ind}(X) = 0$ if and only given $C \subseteq U \subseteq X$ with C closed and U open, there is a clopen V such that $C \subseteq V \subseteq U$. This is equivalent to the statement that for any two disjoint closed sets C, D there is a clopen that contains C and is disjoint to D. Combined with the T_1-axiom, this again is the property of strong zero-dimensionality. □

Definition 10.5.5 *An open cover $\mathcal{U}$ of a topological space X has* <u>order $\leq n$</u> *if every $x \in X$ is contained in at most $n+1$ elements of $\mathcal{U}$.*

Definition 10.5.6 *Let (X,τ) be a topological space. We define the* covering dimension $\mathrm{cov}(X) \in \{-1,0,1,\dots,\infty\}$ *as follows:*

- *We put $\mathrm{cov}(X)=-1$ if and only if $X=\emptyset$.*
- *If $X \neq \emptyset$ and $n \in \mathbb{N}_0$, we say that $\mathrm{cov}(X) \le n$ if, whenever $\mathcal{U}$ is an open cover of X, there is an open refinement $\mathcal{V}$ of $\mathcal{U}$ of order $\le n$.*
- *If $\mathrm{cov}(X) \le n$ holds, but $\mathrm{cov}(X) \le n-1$ does not, we say $\mathrm{cov}(X)=n$.*
- *If there is no $n \in \mathbb{N}$ such that $\mathrm{cov}(X) \le n$ then $\mathrm{cov}(X)=\infty$.*

Remark 10.5.7 1. It is clear that $\mathrm{cov}(X)=0$ if and only if every open cover $\mathcal{U}$ has a refinement $\mathcal{V}$ consisting of disjoint open sets. Such a space obviously has no connected subspaces with more than one point and thus is 'totally disconnected'.

2. One can show that $\mathrm{cov}(\mathbb{R}^n) \le n$ with just a little more effort than for the small inductive dimension. Cf. e.g. [180]. □

Naturally one needs to understand the relationships between the four notions of dimension that we have defined. They are summarized in the following

Theorem 10.5.8 *Let X be a topological space.*

(i) If X is normal then $s\text{-}\mathrm{dim}(X)=\mathrm{cov}(X)$.

(ii) If X is metrizable then $s\text{-}\mathrm{dim}(X)=\mathrm{cov}(X)=\mathrm{Ind}(X)$.

(iii) If X is separable metrizable (=second countable T_3) then $s\text{-}\mathrm{dim}(X)=\mathrm{cov}(X)=\mathrm{Ind}(X)=\mathrm{ind}(X)$. In this case we write dim *instead.*

Proof. (i) See [76, Theorem 3.2.6].

(ii) The fact that $\mathrm{cov}(X)=\mathrm{Ind}(X)$ for every metrizable space X was proven independently by Katetov and Morita in the first half of the 1950s. This can be found in most expositions of dimension theory, including the concise ones in [75] and [50].

(iii) That $\mathrm{cov}(X)=\mathrm{Ind}(X)=\mathrm{ind}(X)$ for separable metric spaces is even more classical. Cf. again the short treatments in [75, 50] or the full-blown books [76]. ■

We will see in Corollary 11.1.44 that a second countable zero-dimensional space embeds into $\mathbb{R}$. This has a generalization to higher dimensions:

Theorem 10.5.9 *If X is a second countable T_3-space with $\mathrm{dim}(X) \le n$ then there is an embedding $X \hookrightarrow \mathbb{R}^{2n+1}$.*

(Compare with Whitney's embedding theorem, according to which a smooth manifold of dimension n embeds smoothly into $\mathbb{R}^{2n+1}$.)

Chapter 11

Highly disconnected spaces. Peano curves

11.1 Highly disconnected spaces

11.1.1 Totally disconnected spaces. The connected component functor π_c

Definition 11.1.1 *A topological space is called totally disconnected if $C(x) = \{x\}$ for each x. Equivalently, X has no connected subspaces other than singletons and $\emptyset$.*

Remark 11.1.2 1. Clearly, every discrete space is totally disconnected, but not conversely as Example 9.1.10 shows. Another example is provided by the subspace $\mathbb{Q} \subseteq \mathbb{R}$. Since $\mathbb{Q}$ contains no intervals, it is totally disconnected by Proposition 9.2.1. But $\mathbb{Q}$ is dense-in-itself and thus is far from being discrete.

2. Some authors write 'hereditarily disconnected' instead of totally disconnected, since every subspace (with more than one point) is disconnected.

3. If X is totally disconnected then every $Y \subseteq X$ with $\#Y \geq 2$ is totally disconnected.

4. Every totally disconnected space is T_1 (since $\{x\} = C(x)$ $\forall x$ and $C(x)$ is closed). But a totally disconnected space need not be T_2, cf. Exercise 11.1.10.

5. Total disconnectedness is not a very convenient condition to check since it involves all subspaces of the space. (It is clear that X is totally disconnected if and only if for every $S \subseteq X$ with $\#S \geq 2$ there are open $U, V \subseteq X$ such that $U \cup V \supseteq S$, $U \cap S \neq \emptyset \neq V \cap S$ and $U \cap V \cap S = \emptyset$.) Therefore also other notions of disconnectedness have been defined, some of which we will encounter below. □

Totally disconnected spaces arise in a very natural way:

Proposition 11.1.3 *Let (X, τ) be a topological space and $\sim$ the equivalence relation of being in the same connected component.*

(i) The quotient space $X/\!\sim$ is T_1 and totally disconnected.

(ii) If X is compact Hausdorff then $X/\!\sim$ is Hausdorff (and compact).

(iii) The quotient space $X/\!\sim$ is discrete if and only if X is locally connected.

(iv) The map $p : X \to X/\!\sim$ is a homeomorphism if and only if X is totally disconnected.

Proof. By Lemma 6.4.12, $X/\sim$ is T_1 (respectively discrete) if and only if the $\sim$-equivalence classes, in this case the connected components, are closed (respectively open). Connected components are always closed, and by Proposition 9.1.12 openness of the connected components is equivalent to local connectedness. This proves (iii) and the first half of (i).

We must prove that $X/\sim$ does not have connected subspaces with more than one point. Let $p : X \to X/\sim$ be the quotient map and let $S \subseteq X/\sim$ with $\#S \geq 2$. Then $p^{-1}(S) \subseteq X$ is the union of at least two connected components of X and therefore not connected. There thus is a clopen $T \subseteq p^{-1}(S)$ with $\emptyset \neq T \neq p^{-1}(S)$. This T is itself a union of connected components of X. (Assume $Y \subseteq X$ is connected with $Y \cap T$ and $Y \backslash T$ both non-empty. Then Y is the union of two non-empty clopen subsets, contradicting connectedness.) Thus $T = p^{-1}(p(T))$. In view of the definition of the quotient topology it is clear that a subset Z of $X/\sim$ is clopen if and only if $p^{-1}(Z) \subseteq X$ is clopen. In view of $T = p^{-1}(p(T))$ and the fact that T is clopen, we thus have that $p(T)$ is clopen. It is clear that $p(T) \subseteq S$ with $\emptyset \neq p(T) \neq S$. Therefore S is not connected.

(ii) For compact Hausdorff X, it is clear that $X/\sim$ is compact. Furthermore, the equivalence relation $\sim$ is closed by Proposition 9.1.21. Now Corollary 8.1.18 implies that $X/\sim$ is Hausdorff (in fact normal).

(iv) If X is totally disconnected then the equivalence relation $\sim$ is trivial, thus the quotient map p is a homeomorphism. By (i), $X/\sim$ is totally disconnected. Thus if p is a homeomorphism then X is totally disconnected. ■

In fact, the assignment $X \rightsquigarrow X/\sim$ is only the object-part of a functor from the category $\mathcal{TOP}$ to the category $\mathcal{TOP}_{\text{totdisc}}$ of totally disconnected spaces:

Theorem 11.1.4 *Write $\pi_c(X) = X/\sim$. Every $f \in C(X,Y)$ defines a map $\pi_c(f) : \pi_c(X) \to \pi_c(Y)$ that is continuous w.r.t. the quotient topologies on both spaces and such that the diagram*

$$\begin{tikzcd} X \arrow[r, "f"] \arrow[d, "p_X"'] & Y \arrow[d, "p_Y"] \\ \pi_c(X) \arrow[r, "\pi_c(f)"'] & \pi_c(Y) \end{tikzcd} \tag{11.1}$$

commutes. The assignments $X \mapsto \pi_c(X)$, $f \mapsto \pi_c(f)$ constitute a functor $\pi_c : \mathcal{TOP} \to \mathcal{TOP}_{\text{totdisc}}$.

Proof. Let $f \in C(X,Y)$. If $x, x' \in E \subseteq X$, where E is connected, then $f(E) \subseteq Y$ is connected by Lemma 9.1.1. Thus $f(x), f(x')$ are in the same connected component, so that $p_Y(f(x)) = p_Y(f(x'))$. Thus the diagonal map $g = p_Y \circ f : X \to \pi_c(Y)$, which clearly is continuous, is constant on the connected components of X. Since the connected components are precisely the $\sim$-equivalence classes, Proposition 6.4.8 provides a unique continuous map $\widetilde{g} : \pi_c(X) \to \pi_c(Y)$ such that $\widetilde{g} \circ p_X = g = p_Y \circ f$. Calling this map $\pi_c(f)$, we have the commutativity of (11.1).

It is entirely obvious that $\pi_c(\mathrm{id}_X) = \mathrm{id}_{\pi_c(X)}$. If $g \in C(Y,Z)$, by the first part of the proof there is a unique continuous $\pi_c(g) : \pi_c(Y) \to \pi_c(Z)$ such that $p_Z \circ g = \pi_c(g) \circ p_Y$. Then

$$p_Z \circ g \circ f = \pi_c(g) \circ p_Y \circ f = \pi_c(g) \circ \pi_c(f) \circ p_X,$$

thus $p_Z \circ (g \circ f) = (\pi_c(g) \circ \pi_c(f)) \circ p_X$. But we also know that there is a unique $\pi_c(g \circ f)$ such that $p_Z \circ (g \circ f) = \pi_c(g \circ f) \circ p_X$. This proves $\pi_c(g \circ f) = \pi_c(g) \circ \pi_c(f)$. Thus π_c is a functor from topological spaces to topological spaces. ■

Exercise 11.1.5 Let $f : X \to Y$ be continuous, where Y is totally disconnected. Prove:

(i) f is constant on the connected components of X.

(ii) There is a unique continuous $\widetilde{f} : X/\sim \to Y$ such that $f = \widetilde{f} \circ p_X$.

(iii) Use this to give another proof of Theorem 11.1.4.

(Notice that the first proof of the theorem is simpler in that it does not make explicit use of the total disconnectedness of $\pi_c(Y)$,which was the hardest part of Proposition 11.1.3.)

The functor π_c is a variation of the path-component functor π_0, which we will study later, but since it is topologically better behaved it also has applications unrelated to algebraic topology:

Exercise 11.1.6 Let $(G, \cdot, e)$ be a topological group, cf. Definition 7.8.23.

(i) Prove that $G_e := C(e)$, the connected component of the unit element, is a closed normal subgroup.

(ii) Show that the equivalence relations $\sim$ given by connectedness and by group theory, cf. $x \sim_G y \Leftrightarrow gh^{-1} \in G_e$, coincide.

(iii) Conclude that the group theoretical quotient map $G \to G/G_e$ can be identified with the topological quotient $G \to \pi_c(G)$.

(iv) Conclude that every topological group G has a connected closed normal subgroup G_e such that the quotient group G/G_e is totally disconnected, the quotient being discrete if and only if $G_e \subseteq G$ is open.

11.1.2 Totally separated spaces

As noted before, the notion of total separation is not very convenient. The following related concept is better behaved:

Definition 11.1.7 *A space X is called* <u>*totally separated*</u> *if for any $x, y \in X$, $x \neq y$ there is a clopen set C such that $x \in C \not\ni y$.*

Exercise 11.1.8 (i) Prove that every $X \subseteq \mathbb{R}$ with $X^0 = \emptyset$ is totally separated.

(ii) Use (i) to prove that $\mathbb{Q}$ and $\mathbb{R}\backslash\mathbb{Q}$ are totally separated.

Lemma 11.1.9 *A totally separated space is Hausdorff and totally disconnected.*

Proof. If $x \neq y$ and $x \in C \not\ni y$ with C clopen, then C and $X\backslash C$ are disjoint open neighborhoods of x, y, respectively. If $\{x, y\} \subseteq S$ then $S \cap C$ is a clopen subset of S different from $\emptyset, S$, thus S is not connected. Thus $x \not\sim y$. ■

The converse is not true:

Exercise 11.1.10 Consider $X = \mathbb{Q} \times \{0, 1\}$, where $\mathbb{Q}$ has the (totally separated) topology as above. Let $Y = X/\sim$ where $\sim$ identifies $(x, 0)$ and $(x, 1)$ provided $x \neq 0$. Then Y is totally disconnected, but not Hausdorff, thus not totally separated. (Compare Example 6.4.17.)

However, for compact Hausdorff spaces, the two notions are equivalent, the key being:

Exercise 11.1.11 Prove that X is totally separated if and only if $Q(x) = \{x\}$ for every $x \in X$, where $Q(x)$ is the quasi-component.

Corollary 11.1.12 *A totally disconnected compact Hausdorff space is totally separated.*

Proof. If X is compact Hausdorff and totally disconnected, then $x \neq y \Rightarrow x \not\sim y$, so that the identity $Q(x) = C(x)$ from Proposition 9.1.19 implies total separation. ■

Exercise 11.1.13 Let X_i be totally separated for each i. Prove that $X = \prod_i X_i$ is totally separated.

Exercise 11.1.14 For a topological space X, let $x \sim_q y \Leftrightarrow Q(x) = Q(y)$, thus $\pi_q(X) = X/{\sim_q}$ is the set of quasi-components of X.

(i) Show that the quotient topology on $\pi_q(X)$ is totally separated.

(ii) If $f : X \to Y$ is continuous, where Y is totally separated, show that f is constant on the quasi-components of X.

(iii) Given a continuous $f : X \to Y$, use (i)-(ii) to canonically define $\pi_q(f) : \pi_q(X) \to \pi_q(Y)$, giving rise to a functor from topological spaces to totally separated spaces.

11.1.3 Zero-dimensional spaces. Stone spaces

Definition 11.1.15 *A space (X, τ) is called zero-dimensional if it is T_0 and τ has a base consisting of clopen sets.*

Remark 11.1.16 The above definition can be restated as follows: If $x \in U \in \tau$ then there is a clopen set C such that $x \in C \subseteq U$. Then, of course, $D = X \backslash C$ is clopen, disjoint from C and it contains the closed set $X \backslash U$. Thus a T_0-space is zero-dimensional if and only if "points are separated from closed sets by clopen sets". □

Exercise 11.1.17 Prove that every element $[a, b)$ of the canonical base of the Sorgenfrey topology τ_S on $\mathbb{R}$ is clopen. Deduce that $(\mathbb{R}, \tau_S)$ is zero-dimensional.

Lemma 11.1.18 *A zero-dimensional space is totally separated and completely regular ($T_{3.5}$).*

Proof. If $x \neq y$, we can find an open U containing precisely one of the two points. If that point is x, zero-dimensionality gives us a clopen C such that $x \in C \subseteq U$. Clearly $y \notin C$, and similarly if $y \in U$. Thus the space is totally separated and thereby T_2. If now $x \in U \in \tau$, zero dimensionality gives a clopen V such that $x \in V \subseteq U$. Since V is closed, we have $x \in V \subseteq \overline{V} \subseteq U$. In view of Lemma 8.1.5, X is T_3. The above also shows that, given a closed $D \subseteq X$ and $x \in X \backslash D$, there is a clopen C such that $x \in C$, $D \subseteq X \backslash C$. Since C is clopen, the characteristic function χ_C is continuous and $\chi_C(x) = 1$, $\chi_C \restriction D = 0$. Thus X is completely regular. ■

Thus we have zero-dimensional $\Rightarrow$ totally separated $\Rightarrow$ totally disconnected.

Exercise 11.1.19 Prove that the properties of being totally separated and zero-dimensional are both hereditary.

Exercise 11.1.20 Prove: If (X, τ) is a compact Hausdorff space and totally separated then X is zero-dimensional. Hint: Show that the clopen sets in X form the base for a Hausdorff topology τ' weaker than τ. Then prove $\tau' = \tau$.

Combining Lemma 11.1.9, Corollary 11.1.12, Lemma 11.1.18 and Exercise 11.1.20, we find:

Corollary 11.1.21 *For compact Hausdorff spaces, the properties of total disconnectedness, total separation and zero-dimensionality are equivalent.*

Definition 11.1.22 *A Stone space is a totally disconnected compact Hausdorff space.*

The space from Example 9.1.10 is a Stone space. The following exercise provides many more:

Exercise 11.1.23 Prove that the one-point compactification X_∞ of every discrete space X is a Stone space.

Exercise 11.1.24 Let X be a discrete space.

(i) Prove that the linear span of the idempotents (cf. Exercise 8.3.40) in $C_b(X,\mathbb{R})$ is dense in $C_b(X,\mathbb{R})$ (with the supremum metric).

(ii) Prove that the Stone-Čech compactification βX is a Stone space.

Remark 11.1.25 Stone spaces are important for many reasons: In the next section we will see that infinite products of finite discrete spaces are Stone spaces, cf. Theorem 11.1.33. More generally, Stone spaces are precisely the 'inverse limits' (a generalization of direct products) of finite discrete spaces. Furthermore, Stone spaces are closely related to Boolean algebras via 'Stone duality'[1], discussed in the Section 11.1.11. Stone duality provides a nice interpretation of the Stone-Čech compactification βX of a discrete space X as the Stone dual of the Boolean algebra $P(X)$. □

11.1.4 Extremely disconnected spaces. Stonean spaces

Definition 11.1.26 *A topological space X is called extremely disconnected if the closure of every open set is open (and, of course, closed, thus clopen).*

An extremely disconnected compact Hausdorff space is called Stonean space.

- Some authors write 'extremally' disconnected, even though this word does not exist.
- Obviousy, every discrete and every indiscrete space is extremely disconnected.
- Indiscrete spaces show that extreme disconnectedness alone does not imply any separation properties and can even coexist with connectedness (since indiscrete spaces are connected). This can also happen for T_1-spaces: If X is infinite, the cofinite topology on X is irreducible, and therefore connected (by Exercise 2.8.3 (iii) and (i)). On the other hand:

Exercise 11.1.27 Prove that the cofinite topology on any set is extremely disconnected.

But combined with a separation propery $\geq T_2$, extreme disconnectedness really is a strong form of disconnectedness:

Exercise 11.1.28 Let X be extremely disconnected. Prove:

(i) If X is T_2 then it is totally separated.

(ii) If X is T_3 then it is zero-dimensional.

[1] The same M. H. Stone as in Stone-Čech compactification.

Whether a space is extremely disconnected can be characterized in terms of the Boolean algebra $\mathrm{Clop}(X)$ of clopen sets in X:

Definition 11.1.29 *A Boolean algebra A is complete if every subset $B \subseteq A$ has a least upper bound (for the partial order defined by $a \leq b \Leftrightarrow a \vee b = b$), denoted $\bigvee B$ or $\sup(B)$.*

Proposition 11.1.30 *Let X be a topological space.*

(i) If X is extremely disconnected then $\mathrm{Clop}(X)$ is complete.

(ii) If X is zero-dimensional and $\mathrm{Clop}(X)$ is complete then X is extremely disconnected.

Proof. (i) Let $\mathcal{F} \subseteq \mathrm{Clop}(X)$. Then $U = \bigcup \mathcal{F} = \bigcup_{V \in \mathcal{F}} V$ is open. Since X is extremely disconnected, $C = \overline{U}$ is clopen. Since C contains every $V \in \mathcal{F}$ and clearly is the smallest clopen with that property, C is the supremum of $\mathcal{F}$, and $\mathrm{Clop}(X)$ is complete.

(ii) By assumption, X has a base consisting of clopens. Thus for every open $U \subseteq X$ there is a family $\mathcal{F} \subseteq \mathrm{Clop}(X)$ such that $U = \bigcup \mathcal{F}$. By completeness of $\mathrm{Clop}(X)$, the family $\mathcal{F}$ has a least upper bound $C = \sup(\mathcal{F})$. Thus there is a smallest clopen C such that $C \supseteq \bigcup \mathcal{F} = U$. Thus extreme disconnectedness of X follows, if we show that $C = \overline{U}$. Since C is closed, we have $\overline{U} \subseteq \overline{C} = C$. On the other hand, if $x \notin \overline{U}$ then $x \in V = X \backslash \overline{U}$. Now by zero-dimensionality, there is a clopen D with $x \in D \subseteq V = X \backslash \overline{U}$. Thus $\overline{U} \subseteq X \backslash D$. Since C is the smallest clopen containing $\overline{U}$, we have $C \subseteq X \backslash D$, and therefore $x \notin C$. This completes the proof of $\overline{U} = C$. ■

In Section 11.1.11 we will prove that every Boolean algebra is isomorphic to $\mathrm{Clop}(X)$ for some Stone space X. This, together with Proposition 11.1.30 and the fact that there are non-complete Boolean algebras, proves that extreme disconnectedness is stronger than the three disconnectedness properties considered earlier, even in the compact Hausdorff case where the latter are equivalent. I.e., there are Stone spaces that are not Stonean.

11.1.5 Infinite products of discrete spaces

We have already extensively studied infinite products in connection with compactness (Section 7.5) and normality (Section 8.1.4). Now we will study the connectedness properties of infinite products.

Exercise 11.1.31 Prove that a finite direct product of discrete spaces is discrete.

Assume I is infinite and X_i is discrete and finite with $\#X_i \geq 2$ $\forall i \in I$. By Tychonov's theorem, $X = \prod_i X_i$ is compact, but $\#X \geq 2^{\mathbb{N}}$ is infinite. Since compact discrete spaces are finite, X cannot be discrete. In fact, the following is a stronger result under weaker assumptions:

Lemma 11.1.32 *Let I be infinite and X_i discrete with $\#X_i \geq 2$ for all $i \in I$. Then every neighborhood of every point of $X = \prod_i X_i$ has cardinality at least $2^{\mathbb{N}} = \#\mathbb{R}$. In particular, X is dense-in-itself.*

Proof. Let $x = (x_i) \in X$. If U is an open neighborhood of x, there is a basic open set $V \subseteq U$ with

$$V = p_{i_1}^{-1}(U_{i_1}) \cap \cdots \cap p_{i_n}^{-1}(U_{i_n}) = \prod_{i \in \{i_1,\ldots,i_n\}} U_i \times \prod_{i \notin \{i_1,\ldots,i_n\}} X_i.$$

Since I is infinite and $\#X_i \geq 2$ for all $i \in I$, the factor $\prod_{i \notin \{i_1,\ldots,i_n\}} X_i$ is infinite. Thus every neighborhood of $x \in X$ contains at least $2^{\mathbb{N}} = \#\mathbb{R}$ points, thus uncountably many. Thus obviously, no point is isolated. ■

We collect (most of) what we know about infinite products of discrete spaces:

Theorem 11.1.33 *Let X_i be discrete with $\#X_i \geq 2\ \forall i \in I$. Then $X = \prod_i X_i$ is totally separated. Furthermore:*

(i) I is finite $\Leftrightarrow$ X is discrete.

(ii) I is infinite $\Leftrightarrow$ X is perfect.

(iii) All X_i are finite $\Leftrightarrow$ X is compact, thus a Stone space.

(iv) I is countable $\Leftrightarrow$ X is metrizable.

(v) At most countably many X_i are infinite $\Leftrightarrow$ X is paracompact $\Leftrightarrow$ X is normal.

(vi) I is countable and all X_i are countable $\Leftrightarrow$ X is second countable.

Remark 11.1.34 Statement (iii) of course does not mean that every Stone space is (homeomorphic to) an infinite direct product of finite discrete spaces. But: Every zero-dimensional space embeds into a product $\{0,1\}^X$, cf. Lemma 11.1.36. And every Stone space is an inverse (=projective) limit of finite discrete spaces. Cf. Section 11.1.10. □

The simplest infinite product of discrete spaces clearly is

$$K = \{0,1\}^{\mathbb{N}} := \prod_{n=1}^{\infty} \{0,1\},$$

thus the set of $\{0,1\}$-valued sequences. By Theorem 11.1.33, K is a second countable perfect Stone space. Remarkably, K is characterized by these properties:

Theorem 11.1.35 *If X is a second countable perfect Stone space then $X \cong K$.*

This is somewhat surprising since, e.g., the space $L = \{1,2,3\}^{\mathbb{N}}$ satisfies the assumptions while it is not obvious how a homeomorphism $K \stackrel{\cong}{\to} L$ could look like. We do not give the proof (cf. e.g. [237, 50]), but Proposition 11.1.41, while simpler, goes in the same direction. Cf. also Theorem 11.1.55 for another universal property of K.

11.1.6 Zero-dimensional spaces: βX and embeddings $X \hookrightarrow K$

In Section 8.3.3, we defined a map $\iota_X : X \to [0,1]^{C(X,[0,1])}$ and showed that it is an embedding if and only if X is completely regular. In this section, we study the analogous construction, where $C(X,[0,1])$ is replaced by its subset $C(X,\{0,1\})$. In view of the obvious bijection between $C(X,\{0,1\})$ and $\mathrm{Clop}(X)$ given by $f \mapsto f^{-1}(1)$ and $C \mapsto \chi_C$, we parametrize the functions by the associated clopens.

Lemma 11.1.36 *For a space X, define*

$$\widetilde{\iota}_X : X \to \{0,1\}^{\mathrm{Clop}(X)}, \qquad x \mapsto \prod_{C \in \mathrm{Clop}(X)} \chi_C(x).$$

Then

(i) $\widetilde{\iota}_X$ is injective if and only if X is totally separated.

(ii) $\widetilde{\iota}_X$ is an embedding if and only if X is a zero-dimensional space.

Proof. (i) Just as for ι_X, it is clear that $\widetilde{\iota}_X$ is injective if and only if for every $x \neq y$ there is a $C \in \mathrm{Clop}(X)$ such that $\chi_C(x) \neq \chi_C(y)$. Clearly this is the case if and only if for any $x \neq y$ there is a clopen C such that $x \in C \not\ni y$, which is the definition of total separatedness.

(ii) The space $\{0,1\}^{\mathrm{Clop}(X)}$ is compact Hausdorff and, by Exercise 11.1.13, totally separated. Thus by Exercise 11.1.20, it is zero-dimensional. (One can also show directly that products of zero-dimensional spaces are zero-dimensional). Thus if $\widetilde{\iota}_X$ is an embedding, X is zero-dimensional by Exercise 11.1.19.

Conversely, let X be zero-dimensional. By Lemma 11.1.18, X is totally separated, so that $\widetilde{\iota}_X$ is injective by (i), and $T_{3.5}$. But the proof of Lemma 11.1.18 also showed that given a closed $D \subseteq X$ and $x \in X \backslash D$ there is an $f \in C(X, \{0,1\})$ such that $f(x) = 1,\ f \restriction D = 0$. This shows that the $\{0,1\}$-valued continuous functions separate points from closed sets. Thus by Proposition 8.3.15, $\widetilde{\iota}_X$ is an embedding. ■

Corollary 11.1.37 *If X is zero-dimensional and $\gamma X = \overline{\widetilde{\iota}_X(X)}$ then $(\gamma X, \widetilde{\iota}_X)$ is a Hausdorff compactification of X.*

Proof. By Tychonov's theorem, $\{0,1\}^{\mathrm{Clop}(X)}$ is compact Hausdorff, thus γX is compact Hausdorff, and by Lemma 11.1.36(ii), $\widetilde{\iota}$ is an embedding. ■

Since a zero-dimensional space X is $T_{3.5}$ by Lemma 11.1.18, it also has a Stone-Čech compactification βX, and it is natural to ask how γX is related to βX.

Definition 11.1.38 *A space X is called <u>strongly zero-dimensional</u> if it is T_1 and for any two disjoint closed sets $A, B \subseteq X$ there is a clopen C such that $A \subseteq C \subseteq X \backslash B$.*

It is clear that a strongly zero-dimensional space is zero-dimensional and normal. Every discrete space being strongly zero-dimensional, the following generalizes Exercise 11.1.24(ii):

Proposition 11.1.39 *If X is strongly zero-dimensional then $\beta X \cong \gamma X$ in $\mathcal{C}(X)$, and βX is a Stone space.*

Proof. As noted, X is zero-dimensional, thus by the above, $\widetilde{\iota}_X : X \to \gamma X$ is an embedding. If $A, B \subseteq X$ are disjoint closed sets and C is a clopen such that $A \subseteq C \subseteq X \backslash B$, then the characteristic function χ_C separates A and B, and as in the proof of Proposition 8.3.34, one concludes $\overline{\widetilde{\iota}_X(A)} \cap \overline{\widetilde{\iota}_X(B)} = \emptyset$. Since X also is normal, Corollary 8.3.36 implies $\gamma X \cong \beta X$. Now, γX by construction is a closed subspace of $\{0,1\}^{\mathrm{Clop}(X)}$, thus totally separated and compact Hausdorff, thus Stone. ■

Remark 11.1.40 1. Leaving aside total disconnectedness, the other disconnectedness assumptions above form a hierarchy similar to that of T_2, T_3, T_4: The weakest is the statement that points can be separated by clopens (total separatedness). The next is that X be T_0 and clopens separate points from closed sets, i.e. zero-dimensionality. Finally T_0 combined with separation of disjoint closed sets by clopens, i.e. strong zero-dimensionality. It is tempting to call these properties D_2, D_3, D_4 since $D_k \Rightarrow T_k$. Notice that $D_3 + T_4$ does not imply D_4! (Cf. [75, Example 6.2.20].) But in analogy to Proposition 8.1.15, one can show that D_3+Lindelöf $\Rightarrow D_4$, cf. [75, Theorem 6.2.7]. Thus in particular Stone spaces are strongly zero-dimensional.

2. γX is closely related to Stone duality and ultrafilters, cf. Theorem 11.1.82. □

The compactification γX was defined by embedding X into $\{0,1\}^{\mathrm{Clop}(X)}$, similar to the Stone-Čech compactification. If X is second countable, one may ask whether X embeds into $K = \{0,1\}^{\mathbb{N}}$, similar to what was done in proving Urysohn's metrization theorem.

Proposition 11.1.41 *A zero-dimensional space embeds into K if and only if it is second countable.*

Proof. By Lemma 11.1.18, X is $T_{3.5}$. Zero dimensionality means that every $x \in X$ has a neighborhood base $\mathcal{N}_x$ consisting of clopens. Now $\mathcal{V} = \bigcup_x \mathcal{N}_x$ is a base for τ that consists of clopens. Thus Proposition 7.1.10 provides a countable base $\{V_1, V_2, \ldots\} = \mathcal{V}_0 \subseteq \mathcal{V}$, so that X has a countable base consisting of clopens.

Now define a map $\iota : X \to K = \{0,1\}^{\mathbb{N}}$ by $\iota(x) = (\iota(x)_1, \iota(x)_2, \ldots)$, where $\iota(x)_n = \chi_{V_n}(x)$. Again, this map is continuous since the V_n are clopen. If $C \subseteq X$ is closed and $x \in X \backslash C =: U$ then $x \in U \subseteq \tau$. Since $\mathcal{V}_0$ is a base, there is an i such that $x \in V_i \subseteq U$. Thus $\mathcal{V}_0$ separates points from closed sets, so that ι is an embedding by Proposition 8.3.15. The converse follows from the fact that K is second countable, together with heriditarity of this property. ■

Remark 11.1.42 We will soon see that K embeds into $\mathbb{R}$. By $X \hookrightarrow K \hookrightarrow \mathbb{R}$, X embeds into $\mathbb{R}$. This is the simplest case of the general result, proven in topological dimension theory, that a space of dimension N embeds into $\mathbb{R}^{2N+1}$. □

11.1.7 Embeddings $K \hookrightarrow \mathbb{R}$. The Cantor set

Proposition 11.1.43 *Let $c > 1$ and define*

$$F_c : K \to \mathbb{R}, \quad (a_1, a_2, \ldots) \mapsto (c-1) \sum_{n=1}^{\infty} c^{-n} a_n.$$

(For $c = 2$ this is the well-known binary expansion $0.x_1x_2x_3\ldots$ of $x \in [0,1]$.) Then

(i) $F_c(K) \subseteq [0,1]$.

(ii) $F_c : K \to \mathbb{R}$ is continuous, thus $F_c(K) \subseteq [0,1]$ is compact and closed.

(iii) $F_c : K \to [0,1]$ is not a bijection.

(iv) If $c > 2$ then F_c is injective, thus an embedding.

(v) If $c \in (1,2]$ then F_c is surjective, thus a quotient map.

(vi) F_c is not injective if $c \in (1,2]$ and not surjective if $c > 2$.

Proof. (i) This follows from

$$\sum_{n=1}^{\infty} c^{-n} = \frac{1}{1-\frac{1}{c}} - 1 = \frac{1}{c-1}. \tag{11.2}$$

(ii) For $a, b \in K$, we have $|F_c(a) - F_c(b)| = (c-1)\sum_{n=1}^{\infty} c^{-n}|a_n - b_n|$. If $a_n = b_n$ for $n \le N$, this implies $|F_c(a) - F_c(b)| \le 2(c-1)\sum_{n=N+1}^{\infty} c^{-n} = 2c^{-N}$. Since $c > 1$, for every $\varepsilon > 0$ we can find $N \in \mathbb{N}$ such that $2c^{-N} < \varepsilon$. Now $U_N(a) = \{x \in K \mid x_i = a_i \ \forall i \le N\}$ is an open neighborhood of a, and we have $x \in U_N(a) \ \Rightarrow \ |F_c(a) - F_c(x)| < \varepsilon$. Thus F_c is continuous. As continuous image of the compact space K, $F_c(K)$ is compact. Since $\mathbb{R}$ is Hausdorff, $F_c(K)$ is closed.

(iii) If F_c was a bijection, it would be a homeomorphism by Proposition 7.4.11. But this is not possible since $[0,1]$ is connected, while K is totally disconnected.

(iv) Let $a, b \in K$ with $a \neq b$. Define $N = \min\{n \in \mathbb{N} \mid a_n \neq b_n\}$. Then $a_N \neq b_N$, and we may assume $a_N = 0, b_N = 1$. Now

$$F_c(a) \leq (c-1)\left(\sum_{n=1}^{N-1} c^{-n} a_n + \sum_{n=N+1}^{\infty} c^{-n}\right), \qquad F_c(b) \geq (c-1)\left(\sum_{n=1}^{N-1} c^{-n} a_n + c^{-N}\right),$$

thus

$$F_c(b) - F_c(a) \geq (c-1)(c^{-N} - \sum_{n=N+1}^{\infty} c^{-n}) = (c-1)c^{-N} - c^{-N}(c-1)\sum_{n=1}^{\infty} c^{-n} = c^{-N}(c-2).$$

Thus for $c > 2$ and $a \neq b$ we have $F_c(a) \neq F_c(b)$, thus injectivity. Now Proposition 7.4.11(iii) implies that F_c is an embedding.

(v) Let $x \in [0,1]$, and put $y = x/(c-1)$. Thus $y \leq 1/(c-1)$. Put $a_n = 0$ for all $n \in \mathbb{N}$ and $N = 1$. Now consider the following algorithm[2]:

(*) If $x = 0$ then we stop. If $x \geq 1/c$, put $a_N = 1$ and replace x by $x - 1/c$, else do nothing. Now multiply x by c, increase N by 1 and return to (*).

If the algorithm stops at a certain value of N, it should be clear that $y = \sum_{n=1}^{N-1} c^{-n} a_n$, thus $x = F_c(a)$, where the sequence a is finite (in the sense of having finitely many non-zero a_n's).

Now consider the case where the algorithm does not stop. We first verify that the variable x of the algorithm always stays bounded by $1/(c-1)$: If $x < 1/c$ (thus $cx < 1$) at the beginning of one iteration then x gets replaced by $cx < 1 \leq 1/(c-1)$ (since $c \leq 2$). Whereas if $x \geq 1/c$ then the new value of x is $c(x - 1/c) = cx - 1$, which is bounded by $c/(c-1) - 1 = 1/(c-1)$. After N steps, the algorithm has computed $a_1, \ldots, a_N$, and we know that $0 \leq y - \sum_{n=1}^{N} c^{-n} a_n \leq c^{-N}/(c-1)$. Since this tends to zero as $N \to \infty$, we have $\lim \sum_{n=1}^{N} c^{-n} a_n = y$, thus $F_c(a) = x$.

(vi) Now we have proven that F_c is surjective for $c \in (1,2]$ and injective for $c > 2$. Combining this with (iii), we know that F_c is not injective for $c \in (1,2]$ and not surjective for $c > 2$. ∎

Corollary 11.1.44 *Every second countable zero-dimensional space embeds into* $\mathbb{R}$.

Proof. Combine the embeddings $X \hookrightarrow K$ of Proposition 11.1.41 and $F_c : K \hookrightarrow \mathbb{R}$, where $c > 2$, of Proposition 11.1.43(iii). ∎

Theorem 11.1.45 *Let $c > 2$. Then*

(i) $F_c(K)$ *is a perfect Stone space.*

(ii) $F_c(K) \subseteq \left[0, \frac{1}{c}\right] \cup \left[1 - \frac{1}{c}, 1\right]$.

(iii) $F_c(K)$ *is* *self-similar* *in the sense of*

$$c\left(F_c(K) \cap \left[0, \frac{1}{c}\right]\right) = F_c(K) = c\left(F_c(K) \cap \left[1 - \frac{1}{c}, 1\right] - \left(1 - \frac{1}{c}\right)\right). \tag{11.3}$$

(iv) $F_c(K)$ *is the maximal set satisfying (ii) and (iii).*

[2] This is an example of a 'greedy' algorithm.

Proof. (i) Obvious consequence of $F_c(K) \cong K$.

(ii) Let $a \in K$ with $a_1 = 0$. Then $F_c(A) \leq (c-1)\sum_{n=2}^{\infty} c^{-N} = c^{-1}(c-1)\sum_{n=1}^{\infty} c^{-N} = c^{-1}$, using (11.2). If $a_1 = 1$ then clearly $F_c(a) \geq (c-1)c^{-1} = 1 - 1/c$.

(iii) As seen in (ii), if $x = F_c(K) \cap [0, 1/c]$ then $x = F_c(a)$, where $a_1 = 0$. Now

$$cF_c(a) = c(c-1)\sum_{n=2}^{\infty} c^{-n} a_n = (c-1)\sum_{n=2}^{\infty} c^{-n+1} a_n = F_c(\sigma(a)),$$

where $\sigma(a) = (a_2, a_3, \ldots)$. If $x = F_c(K) \cap [1 - 1/c, 1]$ then $x = F_c(a)$ where $a_1 = 1$. Then

$$c\left(F_c(a) - \left(1 - \frac{1}{c}\right)\right) = c(c-1)\sum_{n=2}^{\infty} c^{-n} a_n = F_c(\sigma(a)).$$

Now (11.3) follows from the obvious fact $K = \sigma(K)$.

(iv) Assuming that $x \in [0,1]$ satisfies (ii) and (iii), we must show $x \in F_c(K)$. Similar to the proof of Proposition 11.1.43(ii), we do this by giving an algorithm that computes $a \in K$ such that $x = F_c(a)$. Put $N = 1$. (*) By assumption (ii), we have $x \in [0, 1/c]$ or $x \in [1 - 1/c, 1]$. According to which is the case, define $a_N = 0$ or $a_N = 1$. Now assumption (iii) implies that $x' = c(x - a_N/c) \in F_c(K)$. Replace x by x', increase N by one, and return to (*).

The above algorithm defines a map $\beta_c : F_c(a) \to K$. It is obvious that $\beta_c \circ F_c = \mathrm{id}_K$. That $F_c \circ \beta_c = \mathrm{id}_{F_c(K)}$ is seen by the same convergence argument as in Proposition 11.1.43(ii). ■

Remark 11.1.46 We have proven that $F_c(K)$ is the largest subset of $\mathbb{R}$ satisfying (ii) and (iii) of Theorem 11.1.45. One can actually prove the stronger result that already (iii) alone has only three solutions: $\emptyset, F_c(K), \mathbb{R}$. This can be generalized considerably and is proven using a contraction principle for compact subsets rather than points. Cf. Section D.3. □

The result of (iv) can be reformulated equivalently as follows: Let $C_0 = [0,1]$ and define the open interval $U_0 = (1/c, 1 - 1/c)$. Define $C_1 = C_0 \backslash U_0$. Clearly, $C_1 = [0, 1/c] \cup [1 - 1/c, 1]$, which consists of two closed intervals of length $1/c$. Now

$$U_1 = c^{-1}(1/c, 1 - 1/c) \cup \left(c^{-1}(1/c, 1 - 1/c) + (1 - 1/c)\right)$$

is the union of two open intervals, which upon rescaling $[0, 1/c]$ and $[1 - 1/c, 1]$ to $[0,1]$ become $(1/c, 1 - 1/c)$. Then $C_2 = C_1 \backslash U_1$ a union of four closed intervals of length c^{-2}. By now it should be clear that this construction can be iterated. We obtain open sets $U_n \subseteq [0,1]$ such that each U_n is the union of 2^n open intervals of length $c^{-(n+1)}$. Furthermore, $C_n = [0,1] \backslash (\bigcup_{k=0}^{n-1} U_k)$ is the union of 2^n closed intervals of length c^{-n}. Now

$$F_c(K) = \bigcap_{n=0}^{\infty} C_n = [0,1] \backslash \bigcup_{k=0}^{\infty} U_k.$$

The above construction simplifies slightly when $c = 3$, since then $(1/c, 1 - 1/c) = (1/3, 2/3)$ is the middle open third of $[0,1]$. For this (and only this) reason, $F_c(K)$ plays a distinguished rôle and has a name:

Definition 11.1.47 *The subset $\Gamma = F_3(K) \subseteq [0,1]$ is called Cantor's middle third set or just the Cantor set.*

Remark 11.1.48 In the proof of Theorem 11.1.45(iii), we constructed an inverse $F_c(K) \to K$ of F_c. Since $F_3 : K \to \Gamma$ is a homeomorphism, $\beta_c = F_c^{-1}(K) : F_c(K) \to K$ is continuous. This can be shown directly: If a sequence $x_i \in F_c(K)$ converges to $x \in F_c(K)$, then x_i is eventually either in $[0, 1/c]$ or in $[1 - 1/c, 1]$. Since $1/c < 1 - 1/c$, $\beta_c(x_i)_1$ is eventually constant. Since $\{0, 1\}$ is discrete, this is equivalent to $\beta_c(x_i)_1 \to \beta_c(x)_1$.) By the same inductive reasoning as above, we see that $\lim_i \beta_c(x_i)_j = \beta_c(x)_j$ for each $j \in \mathbb{N}$. By the definition of the product topology on $K = \{0, 1\}^{\mathbb{N}}$, this implies that $\beta_c(x_i) \to \beta_c(x)$ in K. Thus $\beta_c : \Gamma \to K$ is continuous.

By the alternative construction $F_c(K) = \bigcap_i C_i$ is a closed subset of $[0, 1]$, it is compact. Since $\beta : F_c(K) \to K$ is a continuous bijection (and K is Hausdorff), β is a homeomorphism. This gives a proof of the compactness of $K = \{0, 1\}^{\mathbb{N}}$ that is independent of Tychonov's theorem! □

Exercise 11.1.49 Use the alternative construction $F_c(K) = \bigcap_i C_i$ (thus not the homeomorphism $F_c(K) \cong K$) to prove that $F_c(K)$ is

(i) totally disconnected,

(ii) perfect.

We now use the homeomorphism $K \cong F_c(K)$ for $c > 2$ to prove a result about K.

Lemma 11.1.50 *Consider $\Gamma = F_3(K)$.*

(i) Let (l_i, r_i) be one of the open intervals removed from I in the construction of the Cantor set Γ. Let $a = \beta(l_i)$, $b = \beta(r_i)$, where $\beta = F_3^{-1} : \Gamma \to K$ is as above. Then there is $N \in \mathbb{N}_0$ such that

$$a = (a_1, a_2, \ldots, a_N, 0, 1, 1, 1, \ldots), \qquad b = (a_1, a_2, \ldots, a_N, 1, 0, 0, 0, \ldots).$$

(Here $N = 0$ is interpreted as absence of $a_1, \ldots, a_N$.)

(ii) Let $x, y, z \in \Gamma$ satisfy $x < y < z$ and $y = (x + z)/2$. (Equivalently, $y - x = z - y$, i.e. x, y, z is an arithmetic progression.) Let $a = \beta(x)$, $b = \beta(y)$, $c = \beta(z)$. Then there is $N \in \mathbb{N}_0$ such that either

$$a = (a_1, \ldots, a_N, 0, 0, 0, 0, \ldots),\ b = (a_1, \ldots, a_N, 0, 1, 1, 1, \ldots),\ c = (a_1, \ldots, a_N, 1, 0, 0, 0, \ldots)$$

or

$$a = (a_1, \ldots, a_N, 0, 1, 1, 1, \ldots),\ b = (a_1, \ldots, a_N, 1, 0, 0, 0, \ldots),\ c = (a_1, \ldots, a_N, 1, 1, 1, 1, \ldots)$$

In either case, y is a boundary point of one of the intervals (l_i, r_i).

Remark 11.1.51 For $N = 0$, we have $(l, r) = (1/3, 2/3)$ in (1), and in (2) we find that (x, y, z) equals either $(0, 1/3, 2/3)$ or $(1/3, 2/3, 1)$. □

Exercise 11.1.52 Prove Lemma 11.1.50. (The preceding remark should help.)

Part (ii) of the lemma has the following remarkable consequence:

Proposition 11.1.53 *Every closed subset $F \subseteq \Gamma$ of the Cantor set admits a retraction $r : \Gamma \to F$. The same holds for K.*

Proof. For $y \in F$, we must define $r(y) = y$. If $y \in \Gamma \backslash F$, let $d = \inf\{|x - y| \mid x \in F\} > 0$. Clearly, this infimum is assumed either by one point $x \in F$ or by two points x, z, in which case we may assume $x < y < z$. In the first case, we define $r(y) = x$. In the second case, we apply part (ii) of Lemma 11.1.50 and see that either $y = l_i$ or $y = r_i$ for one of the open intervals $(l_i, r_i) \subseteq I \backslash \Gamma$. ($y$ cannot have removed intervals to its left and right, since that would mean that it is an isolated point, contradicting the perfectness of Γ.) If $y = l_i$, we define $r(y) = x$ and if $y = r_i$ we define $r(y) = z$. This uniquely defines a map $r : \Gamma \to F$ satisfying $r \restriction F = \mathrm{id}_F$.

It remains to show continuity of r. Consider a sequence $\{x_i\} \in \Gamma$ converging to $x_0 \in \Gamma$. This means that for every $m \in \mathbb{N}$ there is an $N \in \mathbb{N}$ such that for $n \geq N$ the first m components of $\beta(x_n) \in K$ do not change anymore. I.e., x_n remains in one of the 2^m closed intervals of width 3^{-m} that are left of the interval I (and whose union is called C_m) after the first m steps of the construction of Γ. As long as these successively shorter closed intervals in C_m contain points of F, $r(x_n)$ stays close to x_n (concretely, $|x_n - r(x_n)| \leq 3^{-m}$). If x_n eventually lands in a closed interval containing no point of F, $r(x_n)$ stops changing. In any case, $r(x_n)$ is a Cauchy series, and thus convergent (since F is closed in Γ and thus in I and therefore complete).

That the same conclusion holds for K is obvious in view of $K \cong \Gamma$. ■

11.1.8 K maps onto every compact metrizable space

As we saw in Proposition 11.1.43, the map $f_1 \equiv F_2 : K \to [0,1]$ defined by

$$f_1(a) = \sum_{n=1}^{\infty} 2^{-n} a_n \tag{11.4}$$

is continuous and surjective, but not injective. (The non-injectivity can also be seen directly, since the points $a = (1, 0, 0, 0, \ldots)$ and $b = (0, 1, 1, 1, \ldots)$ of K have the same image 1/2.)

This construction can be generalized: For every $d \in \mathbb{N}$ we can define a continuous surjective map $f_d : K \to I^d$ by

$$f_d(a) = (f_{d,1}(a), \ldots, f_{d,d}(a)), \quad \text{where} \quad f_{d,i}(a) = \sum_{j=1}^{\infty} 2^{-j} a_{d(j-1)+i}.$$

Using a bijection $k : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ one can even define a continuous surjective map $f_{\aleph_0} : K \to I^{\mathbb{N}}$ by

$$f_{\aleph_0}(a) = (f_{\aleph_0,1}(a), f_{\aleph_0,2}(a), \ldots), \quad \text{where} \quad f_{\aleph_0,i}(a) = \sum_{j=1}^{\infty} 2^{-j} a_{k(i,j)}.$$

Thus:

Proposition 11.1.54 *For every $d \in \mathbb{N} \cup \{\aleph_0\}$, there is a continuous surjection $f_d : K \to [0,1]^d$.*

We now combine the above results to prove that K (or Γ) is a ‘universal space’ among the second countable compact Hausdorff spaces (= compact metrizable spaces):

Theorem 11.1.55 (Hausdorff) *For every second countable compact Hausdorff space X there exists a continuous surjection $f : K \to X$. Conversely, if X is Hausdorff and continuous image of K then X is compact and second countable.*

Proof. As in the proof of Urysohn's metrization theorem we have an embedding $\iota : X \hookrightarrow I^{\mathbb{N}}$. Since X is compact, $\iota(X) \subseteq I^{\mathbb{N}}$ is compact, thus closed. By the preceding discussion, we have a continuous surjective map $f_{\aleph_0} : K \to I^{\mathbb{N}}$. Now by continuity, $F = f_{\aleph_0}^{-1}(\iota(X)) \subseteq K$ is closed. By Proposition 11.1.53, there is a retraction $r : K \to F$. Now $f = \iota^{-1} \circ f_{\aleph_0} \circ r : K \to X$ is the desired surjection.

The converse follows from the fact that K is compact Hausdorff, the Hausdorff property of X and Corollary 8.2.38. ■

Remark 11.1.56 1. This theorem can also be proven without embedding X into a cube, cf. [237].

2. In a similar way one shows that every compact Hausdorff space is the continuous image of $\{0,1\}^\chi$ for χ a sufficiently large cardinal number. □

11.1.9 Projective limits of topological spaces

******************** define projective limits. projective limits of compact (totally disc) spaces are compact (totally disc)

Definition 11.1.57 *Let J be a small category and $F : J \to \mathrm{Top}$ be a functor. Let X be a space*

*** cofiltered ****

Proposition 11.1.58 *Let P be a property of topological spaces that is preserved under products and closed subspaces. Then any projective limits of spaces with property P has property P.*

Examples: $T_0, T_1, T_2, T_3, T_{3.5}$, compactness, total separation

Corollary 11.1.59 *A projective limit of Stone spaces is Stone. In particular, a projective limit of finite discrete spaces is Stone.*

Remark 11.1.60 Projective limit of finite discrete groups (=profinite group) is Stone space and topological group. Application to Galois theory of infinite algebraic field extensions. Cf. e.g. [174, Chapter 17]. □

11.1.10 Stone spaces = profinite spaces. Profinite groups

Definition 11.1.61 *A topological space is profinite if it is (homeomorphic to) a projective limit of finite discrete spaces.*

Theorem 11.1.62 *A topological space is a Stone space if and only if it is profinite.*

Proof. By Theorem 11.1.33, any product of finite discrete spaces is a Stone space. By construction, an inverse limit $\lim_{\leftarrow} X_i$ is a closed subspace of the product $\prod_i X_i$. Since total disconnectedness is hereditary, we find that every profinite space is compact Hausdorff and totally disconnected, thus Stone.

Now let X be a Stone space. Let E be the set of equivalence relations $\sim$ on X for which $X/\sim$ is a finite discrete space. (We have $\sim \subseteq X \times X$, thus every $\sim$ is a point in $P(X \times X)$ and $E \subseteq P(X \times X)$ clearly is a set.) If $\sim_1, \sim_2 \in E$ and $x \sim_1 y \Rightarrow x \sim_2 y$, we write $\sim_2 \geq \sim_1$. ($\sim_2$ is 'stronger' since it

identifies more.) It is easy to see that $\le$ is a partial order on E. For every $\sim\in E$, we have a canonical quotient map $p : X \mapsto X/\sim$. And if $\sim_1\geq\sim_2$, there is a canonical map $q_{12} : X/\sim_1\to X/\sim_2$ such that

$$\begin{tikzcd} X \arrow[r, "q_1"] \arrow[dr, "q_2"'] & X/\sim_1 \arrow[d, "q_{12}"] \\ & X/\sim_2 \end{tikzcd}$$

commutes. This shows that we have a family $\{X_i = X/\sim_i\}_{\sim\in E}$ of finite discrete spaces and connecting maps $q_{ij} : X_i \to X_j$ whenever $j \geq i$, thus an inverse system, and maps $p_i : X \to X_i$ compatible with the maps $q_{ij} : X_i \to X_j$. Let $Y = \lim_{\leftarrow} X_i$ be the projective limit of the system $\{X_i, q_{ij}\}$, and let $p_i : Y \to X_i$ be the canonical projection maps. Then we have a canonical continuous map $p : X \to Y$ such that

$$\begin{tikzcd} X \arrow[r, "q_i"] \arrow[d, "p"'] \arrow[dr, "q_j"'] & X_i \arrow[d, "q_{ij}"] \\ Y \arrow[r, "p_j"'] \arrow[ur, "p_i"] & X_j \end{tikzcd}$$

proficommutes whenever $i \le j$. By construction, Y is a profinite space. Thus we are done if we prove that $p : X \to Y$ is a homeomorphism. By construction p is continuous. Since X is compact and Y Hausdorff, p automatically is a homeomorphism, provided it is a bijection. (Cf. Proposition 7.4.11(ii).) By assumption X is Stone, thus totally separated. Thus if $x \neq y$, there is a clopen $C \subseteq X$ such that $x \in C \not\ni y$. Let $\sim$ be the equivalence relation on X having C and $X\backslash C$ as equivalence classes. Then clearly $X/\sim$ is a two-point space that is discrete (since the equivalence classes are closed), and the quotient map $q : X \to X/\sim$ satisfies $q(x) \neq q(y)$. Thus already the maps from X to discrete two-point spaces separate the points of X. Thus certainly the map $q = \prod_i q_i : X \to \prod_i X_i$ is injective. Since $Y = \lim_{\leftarrow} X_i \subseteq \prod_i X_i$ is contained in the image of p and the map $p : X \to Y$ is nothing but q considered as a map into $\lim_{\leftarrow} X_i$, also p is injective.

Since X is compact, $p(X) \subseteq Y$ is closed. In order to prove surjectivity of p it therefore suffices to show that $p(X) \subseteq Y$ is dense. A point of Y is a point in $x \in \prod_i X_i$ such that $q_{ij}(x_i) = x_j$ whenever $i \le j$.

****

■

If X is compact Hausdorff but not necessarily Stone, we have $\lim_{\leftarrow} X_i \cong X/\sim$, the Stone space $\pi_c(X)$ of connected components!! (In the next section, we will give yet another interpretation of this space, namely as the Stone dual of the Boolean algebra $\mathrm{Clop}(X)$.)

Definition 11.1.63 *A profinite group is an in an inverse limit of finite groups, where the connecting maps of the inverse system are group homomorphisms.*

It is almost immediately obvious that the inverse limit of an inverse system consisting of groups and group homomorphisms again has a group structure. Thus a profinite group actually is both a group and a Stone space. It is natural to ask for a characterization of the topological groups arising in this way.

Corollary 11.1.64 *Let G be a compact T_0-topological group whose topology such that for every $e \neq g \in G$ there is a closed normal subgroup N of finite index such that $e \notin N$. Then G is a profinite group, i.e. an inverse limit (as group and as topological space) of finite discrete groups.*

Proof. By Theorem C.2.3 (or simply by assumption), the topology τ is Hausdorff. By the last assumption, the family of quotient maps $G \to G/N$ with finite discrete quotients separates the points of G. Now we redo the proof of Theorem 11.1.62, replacing the set E of all equivalence relations on G with finite discrete $G/\sim$ by the subset $E_n \subseteq E$ of equivalence relations arising from closed normal subgroups N of finite index. Thus for every $\sim \in E_n$, the quotient map $G \to G/\sim$ is a surjective homomorphism to a finite discrete group. As before, one shows that the canonical map $G \to \lim_{\leftarrow} G_i$, which now is a group homomorphism, is a homeomorphism. (And thus an isomorphism of topological groups.) ■

************** this is not the best result: it suffices to assume that τ is Stone!!

Remark 11.1.65 1. There are situation, e.g. in number theory, where one needs a generalization of profiniteness. A topological space is called locally profinite if it is locally compact Hausdorff and totally disconnected. In such a space every point has an open neighborhood base whose elements have compact closures and thus are Stone and therefore profinite. Examples are provided by the p-adic numbers $\mathbb{Q}_p$, the completions of $\mathbb{Q}$ w.r.t. the norms $\|\cdot\|_p$ from Example 2.1.9, and finite dimensional vector spaces over them. (Thus also finite algebraic extensions $k \supseteq \mathbb{Q}_p$ and finite dimensional vector spaces over them.)

2. The most immediate examples of profinite spaces actually are profinite groups. But in the next section we will see that every Boolean algebra gives rise to a Stone space, and thus by Theorem 11.1.62 to a profinite space. In these cases, there is no natural group structure. □

11.1.11 Stone duality. Connections to βX

If C, D are clopen subsets of a space X then $C \cap D$, $C \cup D$ and $X \backslash C$ are clopen. Since $\emptyset, X$ are clopen, it turns out that the set of clopen $C \subseteq X$ forms a Boolean algebra.

Definition 11.1.66 *A Boolean algebra is a sextuple $(A, \wedge, \vee, \neg, \mathbf{0}, \mathbf{1})$, where A is a set, $\wedge$ and $\vee$ are binary operations on A, $\neg$ is a unary operation and $\mathbf{0} \neq \mathbf{1} \in A$ are constants. These data must satisfy the following axioms (not all independent):*

1. *$\vee$ and $\wedge$ are associative and commutative.*
2. *$x \vee (y \wedge z) = (x \vee y) \wedge (x \vee z)$ and $x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z)$ (Distributivity).*
3. *$x \vee x = x = x \wedge x$ (Idempotency).*
4. *$x \vee (x \wedge y) = x = x \wedge (x \vee y)$ (Absorption).*
5. *$\neg\neg x = x$*
6. *$\neg(x \vee w) = \neg x \wedge \neg w, \quad \neg(x \wedge w) = \neg x \vee \neg w$ (de Morgan laws).*
7. *$x \vee \neg x = \mathbf{1},\ x \wedge \neg x = \mathbf{0}$.*
8. *$x \vee \mathbf{0} = x,\ x \wedge \mathbf{1} = x,\ x \vee \mathbf{1} = \mathbf{1},\ x \wedge \mathbf{0} = \mathbf{0}$.*

For $a, b \in A$, we write $a \leq b$ if $a \wedge b = a$. (Thus $a \leq a \vee b$ and $a \wedge b \leq a$.)

Example 11.1.67 For every set X, the power set $P(X)$ is a Boolean algebra (with $\vee = \cup, \wedge = \cap, \mathbf{0} = \emptyset, \mathbf{1} = X$ and $\neg Y = X \backslash Y$). □

Example 11.1.68 Let (X, τ) be a topological space. Then $A = \mathrm{Clop}(X)$, equipped with the same operations as in the preceding example, is a Boolean algebra. □

It is clear that in general the assignment $X \mapsto \mathrm{Clop}(X)$ is not injective. E.g., whenever X is connected, $\mathrm{Clop}(X)$ is the minimal Boolean algebra $\{\mathbf{0}, \mathbf{1}\}$. However, we will see that if X is a Stone space, it be recovered from the Boolean algebra $\mathrm{Clop}(X)$.

Exercise 11.1.69 Let R be a commutative ring with unit 1. Let $A = \{p \in R \mid p^2 = p\}$ be the set of idempotents in R. For $p, q \in A$, define $p \wedge q = pq,\ p \vee q = p + q - pq,\ \neg p = 1 - p$. Prove that $(A, \vee, \wedge, \neg, 0, 1)$ is a Boolean algebra.

Remark 11.1.70 If R is a unital ring (non-necessarily commutative) and p a non-trivial ($p \notin \{0, 1\}$) idempotent in the center of R then $x \mapsto (px, (1-p)x),\ R \mapsto pR \oplus (1-p)R$ is a ring-isomorphism. Conversely, if R, S are unital rings then $(1, 0), (0, 1)$ are complementary central idempotents in $R \oplus S$. Thus a ring has a non-trivial direct sum decomposition if and only if it has non-trivial central idempotents. (For this reason, rings without non-trivial central idempotents are called connected.) There are many analogies with the topological situation, beginning with Proposition 6.3.7. Just as a topological space need not be a direct sum of connected subspaces, not even in the compact case, a ring need not be a direct sum of connected rings. If X is compact Hausdorff, then the space $\pi_c(X)$ of connected components is a Stone space by Proposition 11.1.3(i)-(ii). On the other hand, Theorem 11.1.74 below will associate a Stone space to the Boolean algebra of idempotents of a commutative ring R. This space is called the Pierce spectrum of R. It plays an important rôle in the Galois theory of commutative rings, cf. [162, 23]. □

Definition 11.1.71 *Let $(A, \wedge, \vee, \neg, \mathbf{0}, \mathbf{1})$ be a Boolean algebra. A subset $I \subseteq A$ is an ideal if*

(i) $\mathbf{0} \in I$.

(ii) If $a, b \in I$ then $a \vee b \in I$.

(iii) If $b \in I$ and $a \leq b$ then $a \in I$.

An ideal I is called proper if $\mathbf{1} \notin I$. A proper ideal is called maximal if it is not properly contained in another proper ideal.

If $I \subseteq A$ is a proper ideal we cannot have $a \in I$ and $\neg a \in I$, since axiom (ii) would imply $\mathbf{1} = a \vee \neg a \in I$, contradicting properness. If $a \in I$ then $a \wedge b \leq a$, thus $a \wedge b \in I$, and similarly if $b \in I$. The next lemma shows when the converse of these statements holds:

Lemma 11.1.72 *If A is a Boolean algebra and $I \subseteq A$ is an ideal then the following are equivalent:*

(i) I is maximal.

(ii) For every $a \in A$ exactly one of the statements $a \in I$ and $\neg a \in I$ holds.

(iii) $a \wedge b \in I$ holds if and only if $a \in I$ or $b \in I$.

Proof. (ii)⇒(i) Let J be a proper ideal properly containing I. If $a \in J \backslash I$, then $\neg a \in I \subseteq J$. Thus $\mathbf{1} = a \vee \neg a \in J$, which contradicts properness of J.

(i)⇒(ii) If $a \notin I$ and $\neg a \notin I$, define $J = \{x \vee z \mid z \leq a \vee y,\ x, y \in I\}$. Then J is a proper ideal containing J and a.

(ii)⇒(iii) Assume $a \wedge b \in I$. If $a \in I$ we are done, so assume $a \notin I$. By (ii), $\neg a \in I$. By axiom (ii), $I \ni (a \wedge b) \vee \neg a = (a \vee \neg a) \wedge (b \vee \neg a) = b \vee \neg a \geq b$. Thus $b \in I$ by axiom (iii).

(iii)⇒(ii) For every a, we have $a \wedge \neg a = \mathbf{0} \in I$. Now by (iii), we have $a \in I$ or $\neg a \in I$. ■

Lemma 11.1.73 *Every proper ideal is contained in a maximal ideal.*

Proof. The set of ideals is partially ordered under inclusion $\subseteq$, and it is easy to see that if $\mathcal{C}$ is a chain (totally ordered family) of proper ideals then $\bigcup \mathcal{C}$ is a proper ideal. Now the claim follows from Zorn's lemma. ■

Theorem 11.1.74 *For every Boolean algebra* $(A, \wedge, \vee, \neg, \mathbf{0}, \mathbf{1})$ *there is a Stone space* $\mathrm{Spec}(A)$*, the spectrum of* A*, such that* $A \cong \mathrm{Clop}(\mathrm{Spec}(A))$ *as Boolean algebras. (*$\mathrm{Spec}(A)$ *is not to be confused with the spectrum* $\mathrm{Spec}(R)$ *of a commutative ring.)*

Proof. Let $\mathrm{Spec}(A) \subseteq P(A)$ be the set of all maximal ideals in A. For each $a \in A$ we define

$$V(a) = \{M \in \mathrm{Spec}(A) \mid a \notin M\}.$$

We have $V(\mathbf{0}) = \emptyset$ and $V(\mathbf{1}) = \mathrm{Spec}(A)$ (since every ideal contains $\mathbf{0}$ and maximal ideals are proper). If I is any ideal then axioms (ii) and (iii) for ideals imply that $a \vee b \in I \;\Leftrightarrow\; a \in I$ and $b \in I$. And if M is a maximal ideal then by Lemma 11.1.72(i), $a \wedge b \in M \;\Leftrightarrow\; a \in M$ or $b \in M$. Thus:

$$\begin{aligned} V(a) \cup V(b) &= \{M \in \mathrm{Spec}(A) \mid a \notin M \text{ or } b \notin M\} = \{M \in \mathrm{Spec}(A) \mid \neg(a \in M \text{ and } b \in M)\} \\ &= \{M \in \mathrm{Spec}(A) \mid \neg(a \vee b \in M)\} = V(a \vee b). \qquad (11.5) \\ V(a) \cap V(b) &= \{M \in \mathrm{Spec}(A) \mid a \notin M \text{ and } b \notin M\} = \{M \in \mathrm{Spec}(A) \mid \neg(a \in M \text{ or } b \in M)\} \\ &= \{M \in \mathrm{Spec}(A) \mid \neg(a \wedge b \in M)\} = V(a \wedge b). \qquad (11.6) \end{aligned}$$

Thus the family $\mathcal{B} = \{V(a)\}_{a \in A}$ is closed w.r.t. intersection and contains $\mathrm{Spec}(A) = V(\mathbf{1})$. By Proposition 4.1.20, the family of all unions of elements of $\mathcal{B}$ is a topology τ on $\mathrm{Spec}(A)$ (the Stone topology) having $\mathcal{B}$ as a base. By Lemma 11.1.72(ii), for $a \in A$ and $M \in \mathrm{Spec}(A)$, we either have $a \in M$ or $\neg a \in M$. Thus

$$V(\neg a) = \{M \in \mathrm{Spec}(A) \mid \neg a \notin M\} = \{M \in \mathrm{Spec}(A) \mid a \in M\} = \mathrm{Spec}(A) \backslash V(a), \qquad (11.7)$$

which implies that the elements of the base $\mathcal{B}$ are clopen, thus $\mathcal{B} \subseteq \mathrm{Clop}(\mathrm{Spec}(A))$. This means that $\mathrm{Spec}(A)$ has a base $\mathcal{B}$ consisting of clopens, thus $\mathrm{Spec}(A)$ is zero-dimensional.

If $M_1, M_2 \in \mathrm{Spec}(A)$ with $M_1 \neq M_2$ we cannot have $M_1 \subseteq M_2$ since M_1 is maximal, thus there is $a \in M_1 \backslash M_2$. By the above considerations, we have $\neg a \notin M_1, \neg a \in M_2$, thus $\neg a \in M_2 \backslash M_1$. Now $M_1 \in V(\neg a)$ and $M_2 \in V(a) = \mathrm{Spec}(A) \backslash V(\neg a)$, and since the V's are open, $(\mathrm{Spec}(A), \tau)$ is Hausdorff.

Consider now a cover of $\mathrm{Spec}(A)$ by elements of $\mathcal{B}$, i.e. we have $\{b_t \in A\}_{t \in T}$ be such that $\bigcup_{t \in T} V(b_t) = \mathrm{Spec}(A)$. Let I be the set of $a \in A$ for which there are finitely many $t_1, \ldots, t_n$ such that $a \leq b_{t_1} \vee \cdots \vee b_{t_n}$. It is easy to check that I is an ideal. If I is proper, we can find a maximal ideal M containing I, as shown above. By definition, we have $b_t \in I \subseteq M$ for each $t \in T$. Since $\{V(b_t)\}_{t \in T}$ covers $\mathrm{Spec}(A)$, we have $M \in V(b_t)$ for some $t \in T$. But this means $b_t \notin M$, and we

have a contradiction. Thus $I = A$. In particular $\mathbf{1} \in I$, so that there are $t_1, \ldots, t_n \in T$ such that $1 = b_{t_1} \vee \cdots \vee b_{t_n}$. But this implies

$$V(b_{t_1}) \cup \cdots \cup V(b_{t_n}) = V(b_{t_1} \vee \cdots \vee b_{t_n}) = V(\mathbf{1}) = \mathrm{Spec}(A).$$

Thus every cover of $\mathrm{Spec}(A)$ by elements of the base $\mathcal{B}$ has a finite subcover, and by Exercise 7.2.4, $\mathrm{Spec}(A)$ is compact. Therefore $\mathrm{Spec}(A)$ is a Stone space.

Let $U \subseteq \mathrm{Spec}(A)$ be clopen. Since U is open, it is a union of elements of the base $\mathcal{B}$. Since U is closed, it is compact, thus a finite number of elements of $\mathcal{B}$ suffices. Thus

$$U = V(a_1) \cup \cdots \cup V(a_n) = V(a_1 \vee \cdots \vee a_n) \in \mathcal{B},$$

so that $\mathcal{B} \subseteq \mathrm{Clop}(\mathrm{Spec}(A))$ actually exhausts $\mathrm{Clop}(\mathrm{Spec}(A))$.

Now in view of (11.5-11.6) and $V(\mathbf{1}) = \mathrm{Spec}(A)$, $V(\mathbf{0}) = \emptyset$, the map $f : A \to \mathrm{Clop}(\mathrm{Spec}(A))$, $a \mapsto V(a)$ is a homomorphism of Boolean algebras, whose surjectivity is obvious since $\mathcal{B} := \{V(a) \mid a \in A\} = \mathrm{Clop}(\mathrm{Spec}(A))$. If $a \neq \mathbf{0}$ then $I = \{b \leq \neg a\}$ is a proper ideal. If M is a maximal ideal containing I, we have $\neg a \in I \subseteq M$, thus $a \notin M$ and therefore $M \in V(a)$. Thus $a \neq \mathbf{0} \Rightarrow V(a) \neq \emptyset$. Assume now $a \neq b$. Then $a \wedge \neg b \neq \mathbf{0}$ or $b \wedge \neg a \neq 0$. Assume the former. Then $V(a) \backslash V(b) = V(a) \cap V(\neg b) = V(a \wedge \neg b) \neq \emptyset$, thus $V(a) \neq V(b)$. The same conclusion holds if $b \wedge \neg a \neq 0$. Thus $f : A \to \mathrm{Clop}(\mathrm{Spec}(A))$ is injective, and therefore an isomorphism of Boolean algebras. ■

Corollary 11.1.75 *If A_1, A_2 are Boolean algebras then $A_1 \cong A_2 \Leftrightarrow \mathrm{Spec}(A_1) \cong \mathrm{Spec}(A_2)$.*

Proof. In view of the definition of $\mathrm{Spec}(A)$ it is obvious that $A_1 \cong A_2 \;\Rightarrow\; \mathrm{Spec}(A_1) \cong \mathrm{Spec}(A_2)$. As to the converse, assume $\mathrm{Spec}(A_1) \cong \mathrm{Spec}(A_2)$. Using the trivial implication $X_1 \cong X_2 \;\Rightarrow\; \mathrm{Clop}(X_1) \cong \mathrm{Clop}(X_2)$ and $\mathrm{Clop}(\mathrm{Spec}(A)) \cong A$, we have $A_1 \cong \mathrm{Clop}(\mathrm{Spec}(A_1)) \cong \mathrm{Clop}(\mathrm{Spec}(A_2)) \cong A_2$. ■

Corollary 11.1.76 *Every finite Boolean algebra is isomorphic to $P(X) = \{0,1\}^{\#X}$ for some finite X and thus has 2^n elements for some $n \in \mathbb{N}$.*

Proof. Let A be a finite Boolean algebra and $X = \mathrm{Spec}(A)$. Finiteness of A and Stone duality imply that the Stone space X has finitely many clopens. But that implies that X is a direct sum of connected subspaces. Since X is totally separated, the latter all are singletons, thus X is discrete, and $\mathrm{Clop}(X) = P(X)$. Now compactness implies that X is finite. The Boolean algebra of clopens of a finite set X is isomorphic to $0, 1^{\#X}$. ■

If X is any topological space, then $\mathrm{Clop}(X)$ is a Boolean algebra (whose elements we will denote by C instead a), so that $\mathrm{Spec}(\mathrm{Clop}(X))$ is a Stone space. By construction, $\mathrm{Spec}(\mathrm{Clop}(X)) \subseteq P(\mathrm{Clop}(X))$.

Proposition 11.1.77 *Let X be a topological space. Define a map $X \to P(\mathrm{Clop}(X))$ by*

$$F : X \to P(\mathrm{Clop}(X)), \quad x \mapsto \{C \in \mathrm{Clop}(X) \mid x \notin C\}.$$

(i) F is injective if and only if X is totally separated.

(ii) $F(x)$ is a maximal ideal in $\mathrm{Clop}(X)$ for every $x \in X$, thus $F : X \to \mathrm{Spec}(\mathrm{Clop}(X))$.

(iii) $F : X \to \mathrm{Spec}(\mathrm{Clop}(X))$ is continuous.

(iv) $F(X)$ is dense in $\mathrm{Spec}(\mathrm{Clop}(X))$.

(v) If X is compact then $F : X \to \mathrm{Spec}(\mathrm{Clop}(X))$ is surjective.

(vi) $F : X \to \mathrm{Spec}(\mathrm{Clop}(X))$ is a homeomorphism if and only if X is Stone.

Proof. (i) From the definition of $F : X \to P(\mathrm{Clop}(X))$ it is clear that F is injective if and only if $x \neq y$ implies the existence of a clopen C such that $x \in C \not\ni y$. But this is precisely the definition of total separatedness.

(ii) We verify the axioms in Definition 11.1.71: (i) Since $x \notin \emptyset$, we have $\mathbf{0} = \emptyset \in F(x)$. (ii) If $C, D \in F(x)$ then $x \notin C, X \notin D$, thus $x \notin C \cup D$, so that $C \vee D \in F(x)$. (iii) If $C \subseteq D \in F(x)$ then $x \notin D$, thus $x \notin C$, and therefore $C \in F(x)$. Thus $F(x) \subseteq \mathrm{Clop}(X)$ is an ideal. Since $x \notin C \cap D$ holds if and only if $x \notin C$ or $x \notin D$, thus if and only if $C \in F(x)$ or $D \in F(x)$, Lemma 11.1.72 implies that $F(x)$ is a maximal ideal.

(iii) Since the topology of $\mathrm{Spec}(\mathrm{Clop}(X))$ has $\mathcal{B} = \{V(C) \mid C \in \mathrm{Clop}(X)\}$ as base, it is sufficient to show that $F^{-1}(V(C)) \subseteq X$ is open for each $C \in \mathrm{Clop}(X)$. Let thus $C \in \mathrm{Clop}(X)$. Now

$$\begin{aligned} F^{-1}(V(C)) &= \{x \in X \mid F(x) \in V(C)\} = \{x \in X \mid C \notin F(x)\} \\ &= \{x \in X \mid C \notin \{D \in \mathrm{Clop}(X) \mid x \notin D\}\} = C. \end{aligned} \tag{11.8}$$

Since $C \subseteq X$ is clopen, thus open, F is continuous.

(iv) In view of Lemma 2.7.9, it is enough to show that every element $V(C) \neq \emptyset$ of the base of the Stone topology contains $F(x)$ for some $x \in X$. But in view of (11.8), we have $F(x) \in V(C)$ whenever $x \in C$. In view of $C \neq \emptyset \Leftrightarrow V(C) \neq \emptyset$, we are done.

(v) If X is compact then $F(X) \subseteq \mathrm{Spec}(\mathrm{Clop}(X))$ is compact by (iii), thus closed since $\mathrm{Spec}(A)$ is Hausdorff for every A. With (iv), $F(X) = \overline{F(X)} = \mathrm{Spec}(\mathrm{Clop}(X))$, i.e. surjectivity.

(vi) $\Leftarrow$ By (i) and (v), F is a bijection. As a continuous bijection between compact Hausdorff spaces, F is a homeomorphism. The converse is obvious. ■

Corollary 11.1.78 *If X_1, X_2 are Stone spaces then $X_1 \cong X_2 \Leftrightarrow \mathrm{Clop}(X_1) \cong \mathrm{Clop}(X_2)$.*

Proof. $\Rightarrow$ is trivial. And $\mathrm{Clop}(X_1) \cong \mathrm{Clop}(X_2)$ implies $\mathrm{Spec}(\mathrm{Clop}(X_1)) \cong \mathrm{Spec}(\mathrm{Clop}(X_2))$. Proposition 11.1.77(vi) now gives $X_1 \cong \mathrm{Spec}(\mathrm{Clop}(X_1)) \cong \mathrm{Spec}(\mathrm{Clop}(X_2)) \cong X_2$. ■

Remark 11.1.79 The maps $A \mapsto \mathrm{Spec}(A)$, $X \mapsto \mathrm{Clop}(X)$ can be extended to full and faithful contraviant functors between the categories of Boolean algebras and Stone spaces. Corollaries 11.1.75 and 11.1.78 then amount to essential surjectivity of the functors, which thus establish a contravariant equivalence between the two categories. □

Lemma 11.1.80 *For every topological space X, define $\alpha_X : P(\mathrm{Clop}(X)) \to \{0,1\}^{\mathrm{Clop}(X)}$ by*

$$\alpha_X(\mathcal{F})(C) = 1 - \chi_{\mathcal{F}}(C).$$

Then α_X is a bijection, and the diagram (recall that $F(X) \subseteq \mathrm{Spec}(\mathrm{Clop}(X))$)

$$\begin{tikzcd} X \arrow[r, "F"] \arrow[dr, "\widetilde{\iota}_X"'] & \mathrm{Spec}(\mathrm{Clop}(X)) \arrow[d, "\alpha_X"] \\ & \{0,1\}^{\mathrm{Clop}(X)} \end{tikzcd} \tag{11.9}$$

commutes, where F is as in Lemma 11.1.77 and $\widetilde{\iota}_X$ as in Lemma 11.1.36. The restriction of α_X to $\mathrm{Spec}(\mathrm{Clop}(X)) \subseteq P(X)$ is continuous (w.r.t. the Stone topology on $\mathrm{Spec}(\mathrm{Clop}(X))$ and the product topology on $\{0,1\}^{\mathrm{Clop}(X)}$).

Proof. We identify $P(\mathrm{Clop}(X))$ with $\{0,1\}^{\mathrm{Clop}(X)}$ by the usual identification of functions $\{0,1\}^X$ with subsets $Y \subseteq X$ via $Y \rightsquigarrow \chi_Y$ and $f \rightsquigarrow f^{-1}(1)$. In the definition $F : x \mapsto \{C \in \mathrm{Clop}(X) \mid x \notin C\} = \{C \in \mathrm{Clop}(X) \mid \chi_C(x) = 0\}$ a negation appears, whereas there is none in $\widetilde{\iota}_X(x) = \{C \mapsto \chi_C(x)\}$. Correcting this by the exchange $0 \leftrightarrow 1$, it is clear that α_X is a bijection and that (11.9) commutes.

In view of Exercise 5.2.8, in order to prove continuity of α_X it is sufficient to show that $\alpha_X^{-1}(S)$ is open for the elements of the canonical subbase $\mathcal{S}_\Pi$ for the product topology on $\{0,1\}^{\mathrm{Clop}(X)}$, i.e. for $S = p_C^{-1}(v)$, where $C \in \mathrm{Clop}(X)$ and $v \in \{0,1\}$. Now, $p_C^{-1}(1) = \{\ h : \mathrm{Clop}(X) \to \{0,1\} \mid h(C) = 1\}$. Thus $\alpha_X^{-1}(p_C^{-1}(1)) = \{\mathcal{F} \subseteq \mathrm{Clop}(X) \mid C \notin \mathcal{F}\} = V(C)$, which is open by definition of the Stone topology. Similarly, $\alpha_X^{-1}(p_C^{-1}(0)) = \{\mathcal{F} \subseteq \mathrm{Clop}(X) \mid C \in \mathcal{F}\} = \mathrm{Spec}(\mathrm{Clop}(X))\backslash V(C) = V(X\backslash C)$ by (11.7). Thus $\alpha_X \restriction \mathrm{Spec}(\mathrm{Clop}(X))$ is continuous. ■

The following result relates $F : X \to \mathrm{Spec}(\mathrm{Clop}(X))$ to the compactifications γX and βX and provides an interpretation of βX for discrete X in terms of ultrafilters on X.

Definition/Proposition 11.1.81 *For a set X, let $\mathcal{UF}(X)$ denote the set of ultrafilters on X. Then there is a topology τ' on $\mathcal{UF}(X)$ that has $\{V(C) = \{D \subseteq X \mid D \supseteq C\} \mid C \subseteq X\}$ as base. $G(x) = \{C \subseteq X \mid x \in C\}$ is an ultrafilter, called principal. $G : X \to \mathcal{UF}(X)$ is injective.*

Theorem 11.1.82 *Let X be any topological space. Then*

(i) $\alpha_X \restriction \mathrm{Spec}(\mathrm{Clop}(X))$ is a homeomorphism $\mathrm{Spec}(\mathrm{Clop}(X)) \to \gamma X \subseteq \{0,1\}^{\mathrm{Clop}(X)}$.

(ii) If X is zero-dimensional then $(\mathrm{Spec}(\mathrm{Clop}(X)), F)$ is a Hausdorff compactification of X, and $(\mathrm{Spec}(\mathrm{Clop}(X)), F) \cong (\gamma X, \widetilde{\iota}_X)$ in $\mathcal{C}(X)$.

(iii) If X is strongly zero-dimensional then $(\mathrm{Spec}(\mathrm{Clop}(X)), F) \cong (\gamma X, \widetilde{\iota}_X) \cong (\beta X, \iota_X)$.

(iv) If X is discrete then $(\mathcal{UF}(X), G) \cong (\mathrm{Spec}(P(X)), F) \cong (\gamma X, \widetilde{\iota}_X) \cong (\beta X, \iota_X)$.

Proof. (i) Consider the diagram

$$\begin{tikzcd}
X \arrow[r, "F"] \arrow[dr, "\widetilde{\iota}_X"'] & F(X) \arrow[r, hook] \arrow[d, "\alpha_X"] & \mathrm{Spec}(\mathrm{Clop}(X)) \arrow[r, hook] \arrow[d, dashed, "\alpha_X"] & P(\mathrm{Clop}(X)) \arrow[d, "\alpha_X"] \\
& \widetilde{\iota}_X(X) \arrow[r, hook] & \gamma X \arrow[r, hook] & \{0,1\}^{\mathrm{Clop}(X)}
\end{tikzcd}$$

where all horizontal arrows except the one labelled F are inclusions. (Ignore the middle vertical arrow for a minute.) By the Lemma, the left triangle commutes, as well as the boundary of the diagram. The map $\alpha_X \restriction \mathrm{Spec}(\mathrm{Clop}(X))$ is a homeomorphism of $\mathrm{Spec}(\mathrm{Clop}(X))$ onto its image, which is closed in $\{0,1\}^{\mathrm{Clop}(X)}$. Thus $\gamma X = \overline{\widetilde{\iota}_X(X)} \subseteq \alpha_X(\mathrm{Spec}(\mathrm{Clop}(X))$. And since $F(X)$ is dense in $\mathrm{Spec}(\mathrm{Clop}(X))$ and $\alpha_X(F(X)) = \widetilde{\iota}_X(X)$, we have $\overline{\widetilde{\iota}_X(X)} \supseteq \alpha_X(\mathrm{Spec}(\mathrm{Clop}(X))$, implying $\alpha_X(\mathrm{Spec}(\mathrm{Clop}(X)) = \gamma X$. This justifies the middle vertical arrow.

(ii) If X is zero-dimensional, $\widetilde{\iota}_X : X \to \gamma X$ is an embedding, thus $(\gamma X, \widetilde{\iota}_X)$ is a Hausdorff compactification. Now the claim follows from commutativity of the above diagram and the homeomorphism $\alpha_X : \mathrm{Spec}(\mathrm{Clop}(X)) \to \gamma X$. (That $F : X \to \mathrm{Spec}(\mathrm{Clop}(X))$ is an embedding for zero-dimensional X can be shown directly, using (11.8).)

(iii) If X is strongly zero-dimensional then $(\beta X, \iota_X) \cong (\gamma X, \widetilde{\iota}_X)$ by Proposition 11.1.39. Combine this with (ii).

(iv) Since a discrete space is strongly zero-dimensional and $\mathrm{Clop}(X) = P(X)$, in view of (iii) it suffices to establish a bijection between $\mathrm{Spec}(P(X))$ and the set of ultrafilters on X. Recall that

a maximal ideal $M \in \mathrm{Spec}(P(X))$ is a family $M \subseteq P(X)$ that contains the empty set, is closed under union and subsets, does not contain X, and is maximal w.r.t. these properties. Then the family $\mathcal{F} = \{X \backslash C \mid C \in M\}$ is an ultrafilter, cf. Definitions 5.1.34 and 7.5.14. It is clear that this establishes a bijection between maximal ideals in $P(X)$ and ultrafilters on X. ■

Every topological space X gives rise to a Boolean algebra $\mathrm{Clop}(X)$ of clopens and thus to a Stone space $\Omega_X := \Omega(\mathrm{Clop}(X))$. It is more than natural to ask whether there is a more direct relation between X and Ω_X. If X is compact Hausdorff, this is easy:

Proposition 11.1.83 *If X is compact Hausdorff then the Stone space $\Omega(\mathrm{Clop}(X))$ is homeomorphic to $X/\sim$, where $\sim$ is the connectedness equivalence relation. (That $X/\sim$ is Stone was proven directly in Proposition 9.1.21.)*

Proof. (Sketch) For the compact Hausdorff space X we have $C(x) = Q(x)$, i.e. connected components and quasi-components coincide. Thus $x \sim y$ if and only if x, y are contained in the same clopens. Thus the set of clopens is in canonical bijection to $X/\sim$. ■

Proposition 11.1.84 *If X is completely regular, we have a homeomorphism $\Omega(\mathrm{Clop}(X)) \cong \beta X/\sim$, where $\sim$ is the connectedness relation on (the compact Hausdorff space) βX.*

Proof. From Exercise 8.3.40 we have a bijection between clopens $C \subseteq X$, idempotents $p^2 = p \in C_b(X, \mathbb{R}) \cong C(\beta X, \mathbb{R})$ and clopens in βX. Since every isomorphism between commutative algebras gives rise to an isomorphism of the respective Boolean algebras of idempotents, $\mathrm{Clop}(\beta X) \to \mathrm{Clop}(X)$, $D \mapsto D \cap X$ actually is an isomorphism of Boolean algebras. By the preceding proposition, we have $\Omega(\mathrm{Clop}(\beta X)) \cong \beta X/\sim$. Thus $\Omega(\mathrm{Clop}(X)) \cong \Omega(\mathrm{Clop}(\beta X)) \cong \beta X/\sim$. ■

We summarize:

- If X is Stone then $\Omega(\mathrm{Clop}(X)) \cong X$.
- If X is compact T_2 then $\Omega(\mathrm{Clop}(X)) \cong \pi_c(X) = X/\sim$.
- If X is completely regular then $\Omega(\mathrm{Clop}(X)) \cong \pi_c(\beta X) = \beta X/\sim$.
- If X is strongly zero dimensional then $\Omega(\mathrm{Clop}(X)) \cong \beta X$.

Warning: if X is completely regular and zero-dimensional it does NOT follow that βX is zero-dimensional!

Thus βX is given by ultrafilters of z-sets if X is $T_{3.5}$, by ultrafilters of closed sets if X is T_4, by ultrafilters of clopens if X is strongly zero-dimensional (and thus by ultrafilters of all subsets if X is discrete).

11.2 Peano curves and the problem of dimension

It was known since Cantor's work on set theory in the 1870's that the sets $\mathbb{R}^n$ have the same cardinality as $\mathbb{R}$ for all finite n, thus there are bijections between them. From the point of view of topology, this is not particularly disturbing, since topology is (almost) only interested in continuous maps. It was much more disturbing and surprising when in 1890 Peano[3] constructed a continuous surjection $f : I \to I^2$. This motivates the following

[3]Giuseppe Peano (1858-1932). Italian mathematician, also known for his axiomatization of $\mathbb{N}$ and many other things. In 1888 he gave the modern definition of a vector space which unfortunately was ignored for 30 years.

Definition 11.2.1 *A Peano curve in a space X is a continuous surjective map $f : I \to X$.*

(Note that a curve is the same as a path.) Since then, many other constructions have been given. The beautiful overview [205] is highly recommended. Constructions of Peano curves tend to be either quite arithmetic, based e.g. on the binary expansion of real numbers, or rather geometric, the more satisfactory ones being those that can be interpreted either way.

Exercise 11.2.2 Assume *some* continuous surjection $f_2 : I \to I^2$ is given.

(i) Use this to define continuous surjections $f_d : I \to I^d$ for all $d \in \mathbb{N}$.

(ii) Prove the existence of continuous surjections $S^1 \to S^d$ and $\mathbb{R} \to \mathbb{R}^d$ for all $d \in \mathbb{N}$.

At first encounter, the existence of Peano maps, i.e. continuous surjections $I \to I^d, S^1 \to S^d, \mathbb{R} \to \mathbb{R}^d$, is quite troublesome since it throws doubt on the whole concept of dimension. It is true that the spaces $\{\mathbb{R}^n\}_{n\in\mathbb{N}}$ are mutually non-isomorphic as vector spaces, since $\dim_{\mathbb{R}}(\mathbb{R}^n) = n$. But topology is not restricted to linear maps, so that there might well be non-linear homeomorphisms $\mathbb{R}^n \xrightarrow{\cong} \mathbb{R}^m$ for $n \neq m$. As we saw in Section 10.4, that this is not the case was only proven by Brouwer in 1911, cf. [32].

In the next three subsections, we explain three different constructions of Peano curves, all of which take the continuous surjections $f_d : K \to I^d$ of the Section 11.1.8 as starting point.

11.2.1 Peano curves using the Cantor set and Tietze extension

Theorem 11.2.3 (Peano) *There are continuous surjective maps $\widehat{f_d} : I \to I^d$ for every $d \in \mathbb{N} \cup \{\aleph_0\}$.*

Proof. In Section 11.1.8, we have constructed continuous surjective maps $f_d : K \to I^d$, where $K = \{0,1\}^{\mathbb{N}}$ and $d \in \mathbb{N} \cup \{\aleph_0\}$. Since K is homeomorphic to the Cantor set $\Gamma \subseteq [0,1]$, we can interpret f_d as a function $\Gamma \to I^d$. Since $[0,1]$ is metric, thus normal, and $\Gamma \subseteq [0,1]$ is closed, we can apply Tietze's extension theorem and obtain continuous functions $\widehat{f_d} : [0,1] \to I^d$ such that $\widehat{f_d}\restriction\Gamma = f_d$. Since f_d is already surjective, so is $\widehat{f_d} : I \to I^d$. ■

Remark 11.2.4 1. Tietze's extension theorem is quite non-constructive, since the construction of the extension $\widehat{f}$ involves many choices. (The same is true for Urysohn's Lemma, which is used in the proof.) In particular, apart from continuity, we know nothing about the behaviors of $\widehat{f_d}$ on $I\backslash K$. In the next two subsections we will consider two much more concrete ways of extending $f_d : \Gamma \to I^d$ to $I \supseteq \Gamma$. These two methods are complementary: The first is such that $\widehat{f_d}$ is smooth (infinitely differentiable) on the open set $I\backslash\Gamma$, whereas the second can be proven to give functions $\widehat{f_d}$ that are nowhere differentiable!

2. In view of $K \cong \Gamma$, we have a continuous surjection $g : \Gamma \to X$. Since Γ is a closed subset of $[0,1]$, one would like to extend g to a continuous map $\widehat{g} : [0,1] \to X$. However, Tietze's extension theorem applies only to maps $X \supseteq A \to I^d$, where d is some cardinal number. Applying it to a map $[0,1] \supseteq \Gamma \to \iota(X) \subseteq I^d$, there is no reason why the image $\widehat{g}([0,1]) \subseteq I^d$ of the extension should still be contained in $\iota(X)$. Yet, with the additional assumptions that X is connected and locally connected, and using the metrizability of X one can show that every continuous surjection $g : \Gamma \to X$ has an extension $\widehat{g} : [0,1] \to X$, giving the Hahn-Mazurkiewicz theorem, mentioned in Section 12. □

11.2.2 Lebesgue's construction and the Devil's staircase

Most space-filling curves are nowhere differentiable, but not all! To see this, we now describe a different way, due to Lebesgue, of producing continuous extensions $\widehat{f_d} : I \to I^d$ of the continuous surjection $f_d : \Gamma \to I^d$. (The construction works for all $d \in \mathbb{N}$, and we drop the subscript d.) We recall that Γ is obtained by removing countably many open intervals from I:

$$I\backslash\Gamma = (1/3, 2/3) \cup (1/9, 2/9) \cup (7/8, 8/9) \cup \cdots = \bigcup_{i=1}^{\infty} U_i, \quad U_i = (l_i, r_i).$$

Now we define

$$\widehat{f} : I \to I^d, \qquad t \ \mapsto \left\{ \begin{array}{ll} f(t) & \text{for} \quad t \in \Gamma \\ \frac{f(l_i)(r_i - t) + f(r_i)(t - l_i)}{r_i - l_i} & \text{for} \quad t \in (l_i, r_i) \end{array} \right. \tag{11.10}$$

i.e. by linear interpolation between $f(l_i)$ and $f(r_i)$ on (l_i, r_i).

Theorem 11.2.5 $\widehat{f} \in C(I, I^d)$.

Proof. It is clear that $\widehat{f}$ is continuous – in fact C^∞ – on each of the open intervals (l_i, r_i) and thus on $\bigcup_i (l_i, r_i) = I\backslash\Gamma$. On the other hand, despite the fact that $f : \Gamma \to I^d$ is continuous and $\widehat{f} \restriction \Gamma = f$, it is by no means obvious that $\widehat{f}$ is continuous at every $t \in \Gamma$. The point is that Γ has empty interior and therefore every neighborhood of $t_0 \in \Gamma$ contains points from Γ and from $I\backslash\Gamma$. If t_0 is the left boundary l_i of one of the intervals (l_i, r_i), continuity of $\widehat{f}$ at t_0 from the right is clear by construction. Similarly for continuity on the left if $t_0 = r_i$. The other cases require proof. We only prove continuity from the right.

Let $t_0 \in \Gamma$ and $\varepsilon > 0$. Since $f \in C(\Gamma, I^d)$ and Γ is compact, f is uniformly continuous, cf. Proposition 7.7.38. Thus there is $\delta > 0$ such that $\|f(t_1) - f(t_2)\| < \varepsilon$ for all $t_1, t_2 \in \Gamma$ with $|t_1 - t_2| < \delta$. If now $t \in I\backslash\Gamma$ then $t \in (l_i, r_i)$ for some i, thus

$$\begin{aligned} \widehat{f}(t) - \widehat{f}(t_0) &= \frac{f(l_i)(r_i - t) + f(r_i)(t - l_i)}{r_i - l_i} - f(t_0) \\ &= \frac{(f(l_i) - f(t_0))(r_i - t) + (f(r_i) - f(t_0))(t - l_i)}{r_i - l_i} \end{aligned}$$

By Lemma 11.1.50 we may assume $t_0 \neq l_i$ for any i, considering $t \in (t_0, t_0 + \delta)$ sufficiently close to t_0, we will have $t \in (l_i, r_i)$, where $t_0 < l_i < r_i < t_0 + \delta$, and therefore

$$\|\widehat{f}(t) - \widehat{f}(t_0)\| \leq \frac{\|f(l_i) - f(t_0)\|(r_i - t) + \|f(r_i) - f(t_0)\|(t - l_i)}{r_i - l_i} \leq \frac{\varepsilon(r_i - t + t - l_i)}{r_i - l_i} = \varepsilon.$$

Continuity from the left of $t_0 \in \Gamma$ is shown in the same way. This proves that $\widehat{f}$ is continuous on Γ, and thus on I. ■

Remark 11.2.6 Notice that Lebesgue's $\widehat{f}$ is differentiable (even smooth) on $I\backslash\Gamma$, i.e. almost everywhere! In Section 11.2.4 we will show that a space-filling curve cannot be differentiable everywhere. □

Our aim in this section was to produce continuous surjections $\widehat{f_d} : I \to I^d, d \geq 2$. But Lebesgue's extension $\widehat{f_d} \in C(I, I^d)$ of $f_d \in C(\Gamma, I^d)$ is interesting even when $d = 1$. The resulting function $\widehat{f_1} : I \to I$ is called 'the devil's staircase' (or more prosaically the Cantor-Lebesgue function) in

view of its strange properties, which we briefly study. (As before, we drop the superscript 1.) By construction, $\widehat{f}$ is a continuous surjection $I \to I$, which it is smooth on $I\backslash\Gamma$, but we can say more. To begin, it satisfies $\widehat{f}(0) = 0$ and $\widehat{f}(1) = 1$.

Corollary 11.2.7 *If $f = f_1 : I \to I$ and (l_i, r_i) is one of the open intervals constituting $I\backslash\Gamma$ then $f(l_i) = f(r_i)$. Therefore $\widehat{f}$ it is constant on each of the intervals (l_i, r_i). Furthermore, $\widehat{f}$ is monotonously increasing.*

Proof. In view of (i) of Lemma 11.1.50 and the definition of $f : \Gamma \to I$, we have

$$f(l_i) = \sum_{n=1}^{N} a_n 2^{-n} + \sum_{n=N+2}^{\infty} 2^{-n} = \sum_{n=1}^{N} a_n 2^{-n} + 2^{-(N+1)} = f(r_i).$$

Since the extension $\widehat{f}$ of f is defined by linear interpolation on (l_i, r_i), the first claim follows.

In view of the definition of Γ and of the bijection $\beta : K \to \Gamma$, it is clear that $\beta(a) > \beta(b)$ for $a, b \in K$ if and only if $a > b$ in the lexicographic ordering, i.e. there exists $N \in \mathbb{N}_0$ such that $a_i = b_i \ \forall i \leq N$ and $a_{N+1} > b_{N+1}$ (which amounts to $a_{N+1} = 1, b_{N+1} = 0$). Now, the definition of f implies that the map $K \to I$ is weakly order preserving in the sense that $a > b \ \Rightarrow f(a) \geq f(b)$. Since $\widehat{f}$ is constant on $I\backslash\Gamma$, also $\widehat{f}$ is weakly monotonous. ■

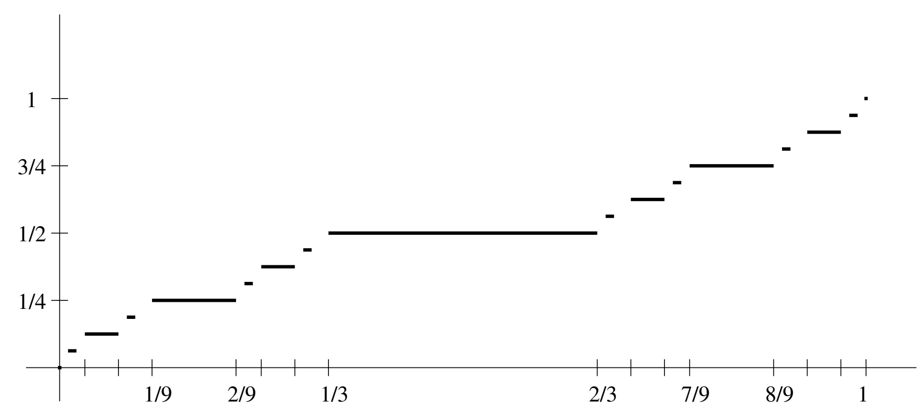


Figure 11.1: The Cantor-Lebesgue function $\widehat{f}^{(1)}$

Remark 11.2.8 1. For the graph of $\widehat{f}$ see Figure 11.1.

2. Since $\widehat{f}$ is constant on $I\backslash\Gamma$, it is differentiable almost everywhere, with $f'(x) = 0$. Obviously the Lebesgue integral $\int_0^1 f'(x)dx$ is zero and fails to coincide with $f(1) - f(0) = 1$. Thus, in order for the 'fundamental theorem of calculus' $\int_a^b f'(x)dx = f(b) - f(a)$ to hold it is not sufficient that f is continuous and almost everywhere differentiable, even if the almost-everywhere-defined function f' is Lebesgue integrable. $\int_a^b f'(x)dx = f(b) - f(a)$ does hold if f is 'absolutely continuous'. One can show that this is equivalent to f being (i) continuous, (ii) of bounded variation and (iii) 'negligent'. Our $\widehat{f}$ is continuous, and being monotonous, it has bounded variation. But 'negligent' means that $f(S)$ has measure zero whenever S has measure zero, a property that out $\widehat{f}$ manifestly doesn't have since it maps the Cantor set Γ to $[0, 1]$. For the missing definitions and proofs, cf. e.g. [218]. □

11.2.3 Schöneberg's construction

Theorem 11.2.9 *There exists a continuous surjective map $I \to I^2$.*

Proof. We use the identification $K \cong \Gamma$ to interpret $f_2 : K \to I^2$ as a map $\Gamma \to I^2$. Since $\Gamma \subseteq [0,1]$, the only problem is to extend f_2 to I in a continuous way. We define a function $p : \mathbb{R} \to \mathbb{R}$ with period 2 by

$$p(t) = \left\{ \begin{array}{ccc} 0 & \text{for} & 0 \leq t \leq \frac{1}{3} \\ 3t-1 & \text{for} & \frac{1}{3} \leq t \leq \frac{2}{3} \\ 1 & \text{for} & \frac{2}{3} \leq t \leq 1 \end{array} \right\}, \quad p(-t) = p(t), \quad p(t+2) = p(t). \tag{11.11}$$

See also Figure 11.2.

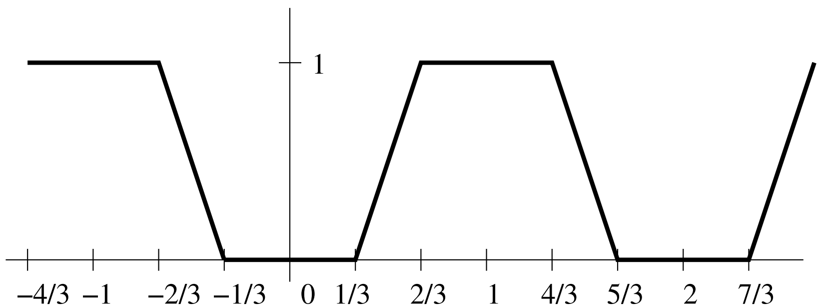


Figure 11.2: Schöneberg's function p

Now we define

$$f_1(t) = \sum_{k=1}^{\infty} 2^{-k} p(3^{2k-2} t), \qquad f_2(t) = \sum_{k=1}^{\infty} 2^{-k} p(3^{2k-1} t).$$

Since $p \in C(\mathbb{R}, [0,1])$, the sums converge uniformly and we have $f_1, f_2 \in C(\mathbb{R}, [0,1])$. Defining $\widehat{f}(t) = (f_1(t), f_2(t)) \in I^2$, we claim that $\widehat{f} \restriction \Gamma = f^{(2)}$. By definition,

$$f^{(2)} : \Gamma \to I^2, \quad 2\sum_{k=1}^{\infty} a_n 3^{-n} \ \mapsto \ \left(\sum_{k=1}^{\infty} 2^{-k} a_{2k-1}, \sum_{k=1}^{\infty} 2^{-k} a_{2k} \right).$$

Thus $\widehat{f} \restriction \Gamma = f^{(2)}$ follows if

$$p\left(3^{2k-2} (2 \sum_{n=1}^{\infty} a_n 3^{-n}) \right) = a_{2k-1}, \qquad p\left(3^{2k-1} (2 \sum_{n=1}^{\infty} a_n 3^{-n}) \right) = a_{2k} \tag{11.12}$$

for all $k \in \mathbb{N}$. We have

$$p\left(3^{2k-2} (2 \sum_{k=1}^{\infty} a_n 3^{-n}) \right) = p\left(2 \sum_{n=1}^{\infty} a_n 3^{2k-n-2} \right)$$

The terms in the sum over $n \in \mathbb{N}$ for which $2k - n - 2 \geq 0$ can be omitted since they contribute an integer multiple of 2 to the sum and p has period 2. Thus

$$= p\left(2 \sum_{n=2k-1}^{\infty} a_n 3^{2k-n-2} \right) = p\left(\frac{2}{3} a_{2k-1} + \frac{2}{9} a_{2k+1} + \frac{2}{27} a_{2k+3} + \cdots \right).$$

Now,
$$\frac{2}{9}a_{2k+1} + \frac{2}{27}a_{2k+3} + \cdots \le \frac{2}{9}\sum_{k=0}^{\infty} 3^{-k} = \frac{2}{9}\frac{1}{1-1/3} = \frac{1}{3},$$
and therefore $2\sum_{n=2k-1}^{\infty} a_n 3^{2k-n-2}$ lies in $[0,1/3]$ if $a_{2k-1}=0$ and in $[2/3,1]$ if $a_{2k-1}=1$. With the definition (11.11) of p, this proves the first identity in (11.12). The proof of the second is entirely similar.

We thus have proven $\widehat{f} \in C(\mathbb{R}, I^2)$ and $\widehat{f}\restriction\Gamma = f^{(2)}$. In view of $f^{(2)}(\Gamma) = I^2$, this implies $\widehat{f}(I) = I^2$, thus $\widehat{f} : I \to I^2$ is a Peano curve. ■

Remark 11.2.10 1. One can prove by completely elementary means that the two functions $f_1, f_2 : I \to I$ that make up Schöneberg's $\widehat{f} = (f_1, f_2) : I \to I^2$ are nowhere differentiable, cf. [205, Theorem 7.2]. (We don't do this here since we already know, albeit non-constructively, that nowhere differentiable functions are even dense in the continuous functions. The proofs of nowhere-differentiability of specific functions are usually not very illuminating – unless they use important general techniques like 'lacunarity'.)

2. It is not hard to generalize Schöneberg's construction such as to obtain continuous surjections $I \to I^d$ and $I \to I^{\mathbb{N}}$, cf. [205], but we leave the matter here. □

11.2.4 There are no differentiable Peano maps

All known Peano curves $f : I \to I^d$ are manifestly non-differentiable, cf. [205]. In this section we will see that this is not due our lack of ingenuity. We need the following notion:

Definition 11.2.11 *By a cube of edge λ in $\mathbb{R}^n$ we mean a product $D = \prod_{i=1}^n [a_i, b_i]$ of n intervals in $\mathbb{R}$ with $|a_i - b_i| = \lambda$ for all i. We write $|D| = \lambda^n$. Now, a set $C \subseteq \mathbb{R}^n$ has measure zero if for every $\varepsilon > 0$ there exist countably many cubes $\{D_i \subseteq \mathbb{R}^n\}_{i\in\mathbb{N}}$ such that*
$$C \subseteq \bigcup_{i=1}^{\infty} D_i \qquad \text{and} \qquad \sum_{i=1}^{\infty} |D_i| < \varepsilon.$$

Remark 11.2.12 1. It is important to understand that measure zero is a relative notion. The interval $I = [0,1] \subseteq \mathbb{R}$ has non-zero measure, but $I \times 0 \subseteq \mathbb{R}^2$ has measure zero!

2. We would arrive at the same notion of measure zero if we replace closed by open cubes. Since the ratio of the volumes of a cube and the circumscribed ball depends only on n, $U \subseteq \mathbb{R}^n$ has measure zero if and only if it can be covered by countably many balls of arbitrarily small total volume. Similarly, one could use rectangles, balls, etc. □

Lemma 11.2.13 *The Cantor set $\Gamma \subseteq \mathbb{R}$ has measure zero.*

Proof. We can cover Γ by one interval of length 1 or by two intervals of length 1/3 or, more generally, by 2^n intervals of length 3^{-n}, thus total length $(2/3)^n$. Since this is true for any $n \in \mathbb{N}$, Γ has measure zero. ■

Lemma 11.2.14 *Let $\{C_i \subseteq \mathbb{R}^n\}_{i\in\mathbb{N}}$ be sets of measure zero. Then $\bigcup_i C_i$ has measure zero.*

Proof. Let $\varepsilon > 0$. Since C_i has measure zero we can pick a sequence $\{D_i^j, j \in \mathbb{N}\}$ of cubes such that $C_i \subseteq \bigcup_j D_i^j$ and $\sum_j |D_i^j| < 2^{-i}\varepsilon$. Then $\{D_i^j, i,j \in \mathbb{N}\}$ is a countable cover of $\bigcup_i C_i$ and we have $\sum_{i,j} |D_i^j| < \varepsilon \sum_i 2^{-i} = \varepsilon$. ■

Exercise 11.2.15 Prove: If $m < n$ then $\mathbb{R}^m \cong \mathbb{R}^m \times 0 \subseteq \mathbb{R}^n$ has measure zero.

Proposition 11.2.16 *Let $U \subseteq \mathbb{R}^m$ be open and $f : U \to \mathbb{R}^m$ differentiable (C^1). If $C \subseteq U$ has measure zero then $f(C) \subseteq \mathbb{R}^m$ has measure zero.*

Proof. Let $\|\cdot\|$ be the Euclidean norm on $\mathbb{R}^m$. Every $p \in U$ belongs to an open ball $B \subseteq U$ such that $\|f'(q)\|$ (where $f'(q) \in \operatorname{End}\mathbb{R}^m$) is uniformly bounded on B, say by $\kappa > 0$. Then

$$\|f(x) - f(y)\| \leq \kappa\|x - y\|$$

for all $x, y \in B$. Thus, if $C \subseteq B$ is an m-cube of edge λ then $f(C)$ is contained in an m-cube of edge less than $\sqrt{m}\kappa\lambda$. It follows that $f(C)$ has measure zero if C has measure zero. Writing U as a countable union of such C, the claim follows by Lemma 11.2.14. ■

Remark 11.2.17 A function $\mathbb{R}^n \supseteq U \to \mathbb{R}^m$ is called Lipschitz-continuous if there is $C > 0$ such that $\|f(x) - f(y)\| \leq \|x - y\|\ \forall x, y \in U$. A function is called locally Lipschitz if every point $x \in U$ has a neighborhood restricted to which f is Lipschitz. Ex: $f : \mathbb{R} \to \mathbb{R},\ x \mapsto x^2$ is locally Lipschitz (like every C^1-function), but not (globally) Lipschitz. The proof of Proposition 11.2.16 clearly works for all locally Lipschitz functions. □

Proposition 11.2.18 *Let $U \subseteq \mathbb{R}^m$ be open and $f : U \to \mathbb{R}^n$ differentiable, where $n > m$. Then $f(U) \subseteq \mathbb{R}^n$ has measure zero. (Thus in particular, empty interior.)*

Proof. Define $\widehat{f} : U \times \mathbb{R}^{n-m} \to \mathbb{R}^n$ by $\widehat{f}(x, y) = f(x)$. Since $U \times \{0\} \subseteq \mathbb{R}^n$ is open and has measure zero, Lemma 11.2.16 implies that $f(U) = \widehat{f}(U \times \{0\}) \subseteq \mathbb{R}^n$ has measure zero. ■

Corollary 11.2.19 *If $n < m$, there is no differentiable surjective map $f : \mathbb{R}^n \to \mathbb{R}^m$.*

Remark 11.2.20 1. The conclusion of Proposition 11.2.18 also holds if f is continuous and $\overline{S}$ is countable, where $S = \{x \in U \mid f \text{ is not differentiable at } x\}$. However, requiring S to have measure zero does not suffice, as shown by Lebesgue's $\widehat{f}^{(n)} : I \to I^n$ almost everywhere differentiable surjections.

2. There are much stronger results: If $f = (f_x, f_y) : I \to I^2$ is such that at each $t \in I$ at least one of $f'_x(t), f'_y(t)$ exists, then $f(I) \subseteq I^2$ has measure zero, cf. [175]. □

11.2.5 Digression: Sard's theorem and other uses of measure zero

Proposition 11.2.18 (and its easy generalization to differentiable manifolds) is called the 'trivial case' of Sard's theorem. Since this theorem is one of the cornerstones of 'differential topology', we allow ourselves the digression of discussing it briefly. In order to state it, we recall that if $f : \mathbb{R}^m \supseteq U \to \mathbb{R}^n$ is differentiable at $x \in U$ then $f'(x)$ is a linear map $\mathbb{R}^m \to \mathbb{R}^n$. If $f'(x)$ is not surjective, one calls $x \in U$ a critical point and $f(x)$ a critical value. Now one has:

Theorem 11.2.21 (Morse-Sard) [4] *If $n, m \in \mathbb{N},\ r \geq \max(1, m - n + 1),\ U \subseteq \mathbb{R}^m$ is open and $f \in C^r(U, \mathbb{R}^n)$ then the set*

$$C = \{f(x) \mid x \in U,\ f'(x) : \mathbb{R}^m \to \mathbb{R}^n \text{ is not surjective}\} \subseteq \mathbb{R}^n$$

of critical values has measure zero.

Warning: The set of critical points need not have measure zero. E.g. if f is constant then every $x \in U$ is a critical point.

Theorem 11.2.21 clearly contains Proposition 11.2.18, since no linear map $\mathbb{R}^m \to \mathbb{R}^n$ with $m < n$ is surjective. Thus $C = \{f(x) \mid x \in U\} = f(U)$ has measure zero.

There seems to be essentially only one proof of Sard's theorem. Involving a double induction, it is not pretty, e.g. [117]. But the theorem is non-trivial already for $n = m = 1$, in which case there is a nice and instructive proof. We cannot resist the urge to state it.

Theorem 11.2.22 (Sard's Theorem for $n = m = 1,\ r = 2$) *If $f \in C^2(U, \mathbb{R})$, where $U \subseteq \mathbb{R}$ is open, then the set*

$$f\left(f'^{-1}(0)\right) = \{f(x) \mid x \in U,\ f'(x) = 0\} \subseteq \mathbb{R}$$

of critical values of f has measure zero.

Proof. Let $C \subseteq U$ be the set of critical points. We first show that $f(C \cap [a, b])$ has measure zero for each closed interval $[a, b]$ with $a < b$. Since f'' is continuous, it is bounded by some constant M on the compact set $[a, b]$. Let $N \in \mathbb{N}$ and put $h = (b - a)/N$. Define $I_n = [a + nh, a + (n + 1)h],\ n = 0, \dots, N - 1$, and let $A = \{n \in \{0, \dots, N - 1\} \mid I_n \cap C \neq \emptyset\}$. Then $f(C \cap [a, b]) = \bigcup_{n\in A} f(C \cap I_n)$. By construction, each I_n contains an x with $f'(x) = 0$. Then the Mean Value Theorem, applied to f', gives that $|f'(x)| \leq Mh$ for all $x \in I_n$. Applying the Mean Value Theorem to f, this implies that $f(I_n)$ is contained in an interval of length at most Mh^2. Thus $f(C \cap [a, b])$ is contained in a union of at most N intervals of length $Mh^2 = M\left(\frac{b-a}{N}\right)^2$. The sum of these lengths is bounded by

$$NM\left(\frac{b-a}{N}\right)^2 = \frac{M(b-a)^2}{N},$$

which can be made arbitrarily small by making N large. This proves that $f(C \cap [a, b])$ has measure zero. If now $U = \bigcup_{i\in I}[a_i, b_i]$, where different intervals $[a_i, b_i]$ overlap at most in one (boundary) point, then I can be at most countable. (As is seen, e.g., by choosing a rational number $q_i \in (a_i, b_i)$ for each $i \in I$. These q_i are all different since the intervals have disjoint interiors.) Thus Lemma 11.2.14 implies that $f(C)$ has measure zero. ■

Remark 11.2.23 1. Theorem 11.2.22 is intuitively quite natural if one looks at the extreme cases: If f is constant then the set $f'^{-1}(0) = U$ is big, but f assumes only one value. If we try to make f assume more values, the set $f'^{-1}(0)$ – and therefore also $f(f'^{-1}(0))$ – gets smaller. However, giving a rigorous general proof is a different matter.

2. Sets of measure zero arise at many places in analysis, not only in Lebesgue integration theory. They appear in the criterion for Riemann integrability, cf. Remark 5.1.23. If $f : [a, b] \to \mathbb{R}$ is monotonous, it is easy to show that f is continuous on $[a, b]\backslash S$, where S is at most countable. In fact, under the same assumptions one can prove that f is differentiable on $[a, b]\backslash T$, where $T \supseteq S$ has measure zero (but may be uncountable). This is 'Lebesgue's differentiation theorem' and is a good deal harder to prove than Theorem 11.2.22.

3. Simplifying a bit, one can say that continuity considerations tend to lead to G_δ-sets, at least when metric spaces are involved. On the other hand, when differentiation and/or integration are involved, sets of measure zero are more relevant. Cf. also [187]. □

[4] Anthony Perry Morse (1911-1984), Arthur Sard (1909-1980). American mathematicians. (A.P. Morse should not be confused with Marston Morse (1892-1977), the founder of 'Morse theory'.)

Chapter 12

Paths in topological and metric spaces

12.1 Paths. Path components. The π_0 functor

As noted in Remark 10.4.14, the methods of Section 10 are of limited scope and should be replaced by functorial ones. In Section 11.1.1 we already encountered the connected component functor π_c. We now turn to the closely related notion of path-connectedness, which is more intuitive than connectedness (at the expense of being less intrinsic and technically less well-behaved). Like the 'higher dimensional' notion of connectedness considered in Section 10, there are higher versions of the path-component functor π_0. In Section 13 we will extensively study π_1 (and encounter π_n in an exercise).

Definition 12.1.1 *Let (X,τ) be a topological space and $x,y \in X$. A path from x to y is a continuous function $p : [0,1] \to X$ such that $p(0) = x,\ p(1) = y$.*

Definition 12.1.2 *(i) If p is a path from x to y, then the reversed path p^{-1} is given by $t \mapsto p(1-t)$. It is a path from y to x.*

(ii) If p is a path from x to y and q is a path from y to z, the composite path $q \bullet p$ is the path from x to z given by

$$(q \bullet p)(t) = \begin{cases} p(2t) & \text{for } t \in [0,1/2] \\ q(2t-1) & \text{for } t \in [1/2,1] \end{cases}$$

Remark 12.1.3 In the topological literature, there is no agreement about the way composite paths are denoted. Some authors would write $p \bullet q$ instead of $q \bullet p$. In Sections 13.4 we will interpret composition of paths in categorical terms, which is why already here we write compositions of paths from right to left, as is customary for composition of morphisms in a category.[1] □

Now the following is obvious:

Lemma 12.1.4 *The relation $\sim_p$ on X defined by $x \sim_p y \Leftrightarrow$ 'there is a path from x to y' is an equivalence relation.*

As always, the equivalence relation R gives rise to a decomposition of X.

Definition 12.1.5 *The $\sim_p$-equivalence classes in X are the path-components of X. For $x \in X$, we write $P(x) = [x]_{\sim_p} = \{y \in X \mid \exists p \in C([0,1],X),\ p(0) = x, p(1) = y\}$.*

The set $X/\!\sim_p$ of path-components is denoted $\pi_0(X)$.

---

[1] There are actually (some) authors who denote the composite morphism $X \xrightarrow{f} Y \xrightarrow{g} Z$ by $f \circ g$ instead of $g \circ f$. One may find this convention more readable, but it is awkward for composite functions: $g(f(x)) = (fg)(x)$!

The following is obvious:

Lemma 12.1.6 *For a topological space X, the following are equivalent:*

(i) For any $x, y \in X$ there is a path from x to y.

(ii) $P(x) = X \quad \forall x \in X$.

(iii) X has only one path-component, i.e. $\#\pi_0(X) = 1$.

Definition 12.1.7 *A topological space satisfying the equivalent statements in Lemma 12.1.6 is called path-connected.*

Definition 12.1.8 *A set $X \subseteq \mathbb{R}^n$ is called star-shaped if there is an $x_0 \in X$ such that for every $x \in X$ and $t \in [0,1]$ one has $tx + (1-t)x_0 \in X$.*

Exercise 12.1.9 For a subset $X \subseteq \mathbb{R}^n$, prove: convex $\Rightarrow$ star-shaped $\Rightarrow$ path-connected.

Lemma 12.1.10 *Let $f : X \to Y$ be continuous. Then*

(i) If $x, y \in X$ are in the same path-component of X then $f(x), f(y)$ are in the same path-component of Y.

(ii) If X is path-connected then $f(X) \subseteq Y$ is path-connected.

Proof. For (i), it suffices to observe that if $p : I \to X$ is a path from x to y then $f \circ p : I \to Y$ is a path from $f(x)$ to $f(y)$. (ii) is an obvious consequence. ■

Proposition 12.1.11 *A continuous function $f : X \to Y$ defines a map $\pi_0(f) : \pi_0(X) \to \pi_0(Y)$ given by $[x] \mapsto [f(x)]$. Now the assignments $X \mapsto \pi_0(X)$, $f \mapsto \pi_0(f)$ constitute a functor $\pi_0 : \mathcal{TOP} \to \mathcal{SET}$.*

Proof. If $C \in \pi_0(X)$, pick any $x \in C$ and define $\pi_0(f)(C) = [f(x)] \in \pi_0(Y)$. By Lemma 12.1.10(i), this is independent of the choice of $x \in C$, thus well-defined. It is clear that $\pi_0(\mathrm{id}_X) = \mathrm{id}_{\pi_0(X)}$, and $\pi_0(g \circ f) = \pi_0(g) \circ \pi_0(f)$ follows from the observation $(g \circ f) \circ p = g \circ (f \circ p)$. We omit the easy details. ■

Corollary 12.1.12 *If $f : X \to Y$ is a homeomorphism, $\pi_0(f) : \pi_0(X) \to \pi_0(Y)$ is a bijection.*

Proof. Since f is a homeomorphism, there exists a continuous $g : Y \to X$ such that $g \circ f = \mathrm{id}_X$ and $f \circ g = \mathrm{id}_Y$. Functoriality of π_0 implies $\pi_0(g) \circ \pi_0(f) = \pi_0(g \circ f) = \pi_0(\mathrm{id}_X) = \mathrm{id}_{\pi_0(X)}$ and similarly $\pi_0(f) \circ \pi_0(g) = \mathrm{id}_{\pi_0(Y)}$. Thus $\pi_0(f)$ is a bijection. ■

Remark 12.1.13 1. Applying the functor π_0 to a topological space clearly entails a huge loss of information. E.g. $\pi_0(X)$ is a one-point space for every path-connected X. But this is exactly the philosophy of algebraic topology where one studies functors from some category of topological spaces to a simpler one. An immediate consequence of the corollary is that two spaces whose π_0's are different (i.e. have different cardinality) cannot be homeomorphic. In the case of π_0 this is rather trivial, but the functors considered in algebraic topology are much more sophisticated. (We will meet the connected component functor π_c and the fundamental group and groupoid functors π_1, Π_1.)

2. If one gives $\pi_0(X), \pi_0(Y)$ the quotient topologies, Proposition 6.4.8 shows that $\pi_0(f) : \pi_0(X) \to \pi_0(Y)$ is continuous. But there is not much point in doing so since there are no good general results

about the quotient-topology on $\pi_0(X) = X/\sim_p$. We will see that path-components need not be open or closed, so that $\pi_0(X)$ need not even be T_1. (For the space Z of Example 12.2.3, $\pi_0(Z)$ is our Example 2.8.5 for a T_0-space that is not T_1.) □

Exercise 12.1.14 Let $\{X_i\}_{i\in I}$ be topological spaces and $X = \prod_{i\in I} X_i$. Prove:

(i) There is a bijection $\pi_0(X) \to \prod_{i\in I} \pi_0(X_i)$.

(ii) If $X_i \neq \emptyset\ \forall i \in I$ then X is path-connected if and only if each X_i is path-connected.

Exercise 12.1.15 Prove that S^n with $n \geq 1$ and $\mathbb{R}^n\backslash\{x\}$ with $n \geq 2$ are path-connected.

12.2 Path-connectedness vs. connectedness

Paths are defined in terms of $I = [0, 1]$. The fact that I is connected provides an implication between path-connectedness and connectedness:

Lemma 12.2.1 *If X is path-connected, it is connected.*

Proof. Assume $C \subseteq X$ is clopen and non-trivial, i.e. $\emptyset \neq C \neq X$. Choose points $x \in C, y \in X\backslash C$. By path-connectedness, there is a path $f : [0, 1] \to X$ from x to y. Now $A = f^{-1}(C)$ is a clopen subset of the interval $[0, 1]$ that is non-trivial since $0 \in A \not\ni 1$. But this contradicts the connectedness of $[0, 1]$ proven in Proposition 9.2.1. ■

Corollary 12.2.2 *For $X \subseteq \mathbb{R}$, connectedness and path-connectedness are equivalent.*

Proof. In view of Lemma 12.2.1, it is enough to show that connectedness implies path-connectedness. This follows from Proposition 9.2.1 and Exercise 12.1.9. ■

This result does not generalize to subsets of $\mathbb{R}^n$, as the next example shows:

Example 12.2.3 (The topologist's sine curve) Define $X, Z \subseteq \mathbb{R}^2$ by

$$X = \left\{ \left(x, \sin\frac{1}{x} \right) \mid 0 < x \leq 1 \right\}, \qquad Z = \overline{X}.$$

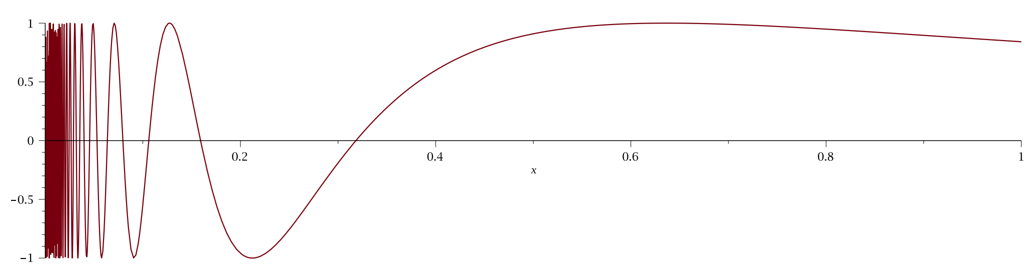


Figure 12.1: Topologist's sine curve (drawn by Maple©)

As the image of the path-connected space $(0, 1]$ under the continuous map $x \mapsto (x, \sin(1/x))$, X is path-connected, thus connected. Therefore, by Lemma 9.1.4, $Z = \overline{X}$ is connected. Now, using Lemma 2.7.3, one easily shows that $Z = X \cup Y$, where $Y = \{0\} \times [-1, 1]$ is the straight line

segment from $(0, -1)$ to $(0, +1)$. It is clear that Y is path-connected, so that Z has at most two path-components. We now claim that there is no path in Z connecting a point in X to a point in Y. Thus Z is not path-connected and has precisely two path-components, X and Y.

To prove this, assume that $f : [0, 1] \to Z$ is a path that begins in X and ends in Y. Writing $f(t) = (f_1(t), f_2(t))$, this means $f_1(0) > 0$ and $f_1(1) = 0$. With $s = \inf\{t \in [0, 1] \mid f_1(s) = 0\}$, continuity of f implies $s > 0$ and $f_1(s) = 0$. For $t \in [0, s)$ we have $f_1(t) > 0$ and thus $f_2(t) = \sin \frac{1}{f_1(t)}$ (since $f(t) \in X$). Now, as $t \nearrow s$, we have $f_1(t) \to 0$, but $f_2(t) = \sin \frac{1}{f_1(t)}$ clearly has no limit as $t \nearrow s$ since for every $\varepsilon > 0$ we have $f_2((s - \varepsilon, s)) = [-1, 1]$. This contradicts the continuity of f, and thus no such path exists. □

Thus connectedness does not imply path-connectedness! In view of Lemma 12.2.1, we have $P(x) \subseteq C(x)$ for every x, but a connected component $C(x)$ can consist of several path components $P(y)$. The example Z also shows that path-components need be neither open nor closed: $X \subseteq Z$ is open but not closed, whereas $Y \subseteq Z$ is closed but not open!

By our standard terminological conventions, a space is locally path-connected if every point has a path-connected neighborhood and strongly locally path-connected if every point has an neighborhood base consisting of path-connected sets.

Lemma 12.2.4 *If X is (weakly) locally path-connected then the path-components are clopen and coincide with the connected components. Thus X is a direct sum of path-connected spaces.*

Proof. Weak local path-connectedness means that for every $y \in P(x)$ there is a path-connected neighborhood N_y. Clearly $N_y \subseteq P(x)$, and since N_y contains an open neighborhood of y, it follows that $P(x)$ is open. Since $P(x)$ is the complement of the union of all other path-components, which is open, $P(x)$ is also closed. Now this implies that $C(x)$, which a priori could be larger then $P(x)$, must equal $P(x)$ since otherwise the latter would be a non-trivial clopen subspace of $C(x)$, contradicting connectedness of the latter. The last statement now is a consequence of the discussion in Section 6.3. ■

Corollary 12.2.5 *If (X, τ) is connected and (weakly) locally path-connected, then it is path-connected.*

The following class of spaces was already briefly encountered in Section 7.8.4:

Definition 12.2.6 *A topological space is locally Euclidean if every point has an open neighborhood that is homeomorphic to $\mathbb{R}^n$ for some $n \in \mathbb{N}$. (n may depend on x.)*

Lemma 12.2.7 *Locally Euclidean spaces are strongly locally path-connected.*

Proof. Follows from the fact that the open balls $B(x, r)$ are convex, thus path-connected. ■

Remark 12.2.8 If $X \subseteq \mathbb{R}^n$ is open then it is locally Euclidean, thus strongly locally path-connected. Thus connected components and path-components coincide. In particular, X is path-connected if and only if it is connected. □

Exercise 12.2.9 Let (X, τ) be a smallest neighborhood space, cf. Section 2.8.3 (for example a finite topological space). Let U_x be the smallest neighborhood U_x of each $x \in X$.

(i) For every $y \in U_x$, construct a path from x to y (continuous of course).

(ii) Conclude that U_x is path-connected.

(iii) Deduce that X strongly locally path-connected.

Thus connected components and path components coincide for smallest neighborhood spaces.

It is trivial that a path-connected space is weakly locally path-connected. But there are spaces that are path-connected, but not strongly locally path-connected!

Exercise 12.2.10 (The topologist's comb) Consider the following subspace of the first quadrant in the plane:

$$X = \{(x, y) \in [0, 1]^2 \mid y = 0 \ \vee \ x = 0 \ \vee \ x = 1/n \text{ with } n \in \mathbb{N}\},$$

cf. Figure 12.2. Prove:

(i) X, with the topology induced from $\mathbb{R}^2$, is path-connected, thus connected.

(ii) X is neither strongly locally path-connected nor strongly locally connected.

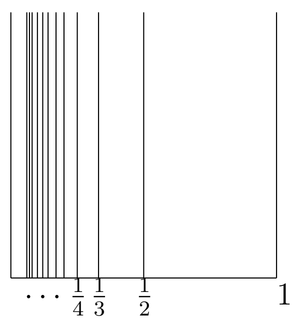


Figure 12.2: The topologist's comb

Remark 12.2.11 Recall that comparing the separation axiom $T_{3.5}$ with the other ones, we made the distinction between intrinsic and extrinsic properties. Clearly, connectedness is intrinsic, but path-connectedness is extrinsic, involving the space $[0, 1]$. Corollary 12.2.5 allows us to replace a global extrinsic condition by a global intrinsic and a local extrinsic condition. This is still not very satisfactory. One would like to have a purely intrinsic criterion equivalent to path-connectedness, or at least implying it. We cite one such result without proofs. (Two others, of a more geometric flavor, will be discussed in Section 12.4.) □

Theorem 12.2.12 (Hahn-Mazurkiewicz ∼1915) [2] *Let (X, τ) be compact, Hausdorff, second countable (=compact metrizable), and strongly locally connected. Then*

(i) X is strongly locally path-connected. (Thus connected components and path-components coincide and are clopen, thus direct summands.)

(ii) If X is connected then (it is path connected and) there exists a continuous surjective map $[0, 1] \to X$.

Conversely, a Hausdorff space that is a continuous image of $[0, 1]$ has the above properties.

[2] Hans Hahn (1879-1934, German), Stefan Mazurkiewicz (1888-1945, Polish)

Remark 12.2.13 1. Spaces X satisfying the conditions in (ii) of Theorem 12.2.12 are called Peano-spaces since they admit a Peano map, i.e. a continuous surjective map $[0, 1] \to X$. For proofs see any of the books [237, 50, 77]. There are generalizations to locally compact spaces, but then $[0, 1]$ must must be replaced by a tree that depends on X. Cf. e.g. [12].

2. The first step in the proof of Theorem 12.2.12 is noting that X is metrizable. The rest of the proof strongly relies on metric space methods. □

Using similar methods, one can give intrinsic characterizations of the spaces $[0, 1]$ and S^1:

Theorem 12.2.14 *Let X be second countable compact Hausdorff (=compact metrizable) and connected.*

(i) If X has precisely two non-cut-points then $X \cong [0, 1]$.

(ii) If X has no cut-points but becomes disconnected upon removal of two points then $X \cong S^1$.

12.3 The Jordan curve theorem

Having met the notion of paths, we briefly consider a closely related notion:

Definition 12.3.1 *A Jordan curve in a topological space X is a continuous injective map $f : S^1 \to X$.*

Remark 12.3.2 1. When X is Hausdorff, which is the only case we are interested in, compactness of S^1 implies that $f : S^1 \to X$ is an embedding, cf. Proposition 7.4.11(iii). Thus $C = f(S^1) \subseteq X$ is a closed subspace homeomorphic to S^1.

2. Actually, some authors mean by 'Jordan curve' not a map $f : S^1 \to X$, but a subspace $Y \subseteq X$ homeomorphic to S^1. They might call our notion of Jordan curve 'parametrized'. □

If f is a Jordan curve in $\mathbb{R}^n$, then $C = f(S^1)$ is closed. Thus $\mathbb{R}^n \backslash C$ is open and therefore locally path connected, thus its connected components and path-components coincide. But how many are there? For $n \geq 3$, the answer is 'one', whereas:

Theorem 12.3.3 (Jordan Curve Theorem) *Let $f : S^1 \to \mathbb{R}^2$ be a Jordan curve. Then $\mathbb{R}^2 \backslash C$, where $C = f(S^1)$, has exactly two connected components, both having C as boundary. One of these is bounded and one unbounded.*

This theorem was first stated by Jordan[3] in the 1890s. While it may appear obvious, the first correct proof was only given by Veblen[4] in 1905. (Also proving connectedness of $\mathbb{R}^n \backslash f(S^1)$ for $n \geq 3$ is non-trivial!)

Exercise 12.3.4 (i) Prove the Jordan curve theorem for $C = S^1$ (the unit circle).

(ii) Use (i) to prove the JCT for $C = \partial X$, where $X \subseteq \mathbb{R}^2$ is compact convex with $X^0 \neq \emptyset$.

In this section we explain the very short and elegant proof given by Maehara [161]. The only modification is that instead of Brouwer's fixed point theorem we use two other results from Section 10.3, namely Corollaries 10.3.1 and 10.3.8(i), which are more convenient. The proof uses two lemmas. The first of these is of a very general nature:

[3] Camille Jordan (1838-1922). Also responsible for the Jordan normal form and much more.

[4] Oswald Veblen (1880-1960). American mathematician.

Lemma 12.3.5 *Let $f, g : I \to I^2$ be paths such that $f(0) \in I_1^-, f(1) \in I_1^+, g(0) \in I_2^-, g(1) \in I_2^+$. Then there are $s, t \in I$ such that $f(s) = g(t)$.*

Proof. Define $h : I^2 \to I^2$ by $h_1(s,t) = f_1(s) - g_1(t),\ h_2(s,t) = g_2(s) - f_2(t)$. Then

$$h_1(I_1^-) \subseteq [-1,0], \quad h_1(I_1^+) \subseteq [0,1], \quad h_1(I_2^-) \subseteq [-1,0], \quad h_2(I_2^+) \subseteq [0,1],$$

and by Poincaré-Miranda (Corollary 10.3.1) there is $(s,t) \in I^2$ such that $h(s,t) = 0$, thus $f(s) = g(t)$. ■

Remark 12.3.6 1. The statement of the lemma may seem even more obvious than the Jordan curve theorem, but this impression is false. The connectedness of intervals is not sufficient to prove it. We obtained it as a corollary of Theorem 10.1.2 in the case $n = 2$, but it can also be deduced from the much simpler Theorem 13.2.4.

2. The Jordan curve theorem, or at least Lemma 12.3.5, plays a crucial rôle in the discussion of planarity of graphs, in particular in proving that the complete graph K_5 and the complete bipartite graph $K_{3,3}$ cannot be embedded into the plane. □

Lemma 12.3.7 *If $\mathbb{R}^2 \backslash C$ is not connected, then $\partial U = C$ for each component U of $\mathbb{R}^2 \backslash C$.*

Proof. Let V be a connected component of $\mathbb{R}^2 \backslash C$. Since $\mathbb{R}^2 \backslash C$ is assumed disconnected, it has another component $W \neq V$. Since V, W are open and disjoint, $W \cap \overline{V} = \emptyset$, and thus $W \cap \partial V = \emptyset$. Since this holds for every component W of $\mathbb{R}^2 \backslash C$ and since $\partial V \cap V = \emptyset$, we must have $\partial V \subseteq C$. Assume $\partial V \subsetneq C$. Then there exists an arc $A \subseteq C$ such that $\partial V \subseteq A$. If V is bounded, choose a point $p \in V$. Otherwise choose $p \in W$, where W is a bounded component. Let D be a disk with center p that is large enough to contain C. Then ∂D is contained in the unbounded component U of $\mathbb{R}^2 \backslash C$. Since the arc A is homeomorphic to $[0,1]$, the identity map $A \to A$ has a continuous extension to $r : D \to A$ by Tietze's Theorem 8.2.20. Now define $q : D \to D$ by

$$q(z) = \begin{cases} r(z) & \text{if} \quad z \in \overline{V} \\ z & \text{if} \quad z \in D \backslash V \end{cases} \qquad \text{if } V \text{ is bounded, thus } V \subseteq D$$

$$q(z) = \begin{cases} z & \text{if} \quad z \in D \cap \overline{V} \\ r(z) & \text{if} \quad z \in D \backslash V \end{cases} \qquad \text{if } V \text{ is unbounded, i.e. } V = U.$$

On $\partial V = \overline{V} \cap (X \backslash V)$ this is well-defined, since r is the identity on $A \supseteq \partial V$. Now q is continuous by Exercise 6.2.5. For bounded V, we have $q(D) = A \cup D \backslash V$, and for unbounded V we have $q(D) = A \cup (D \cap \overline{V})$. In either case, we by construction have $q \in D \backslash q(D)$. Thus q is not surjective, but since it is the identity on ∂D, this contradicts Corollary 10.3.8(i). This proves $\partial V = C$. ■

Proof of Theorem 12.3.3. (1) The compact space $C = p(S^1) \subseteq \mathbb{R}^2$ is closed and bounded. Thus $\mathbb{R}^2 \backslash C$ is open, and locally path-connected. Thus $\mathbb{R}^2 \backslash C$ is a union of path-connected open subsets of $\mathbb{R}^2$. Since C is bounded, precisely one component U is unbounded.

(2) Preliminaries: Since C is compact, there are points $a, b \in C$ such that $d(a,b) = \sup_{x,y \in C} d(x,y)$, where d is the Euclidean distance. Rotating and rescaling, we may assume that $a = (-1,0), b = (1,0)$. Now the curve C is contained in $\overline{B}((-1,0),2) \cap \overline{B}((1,0),2) \subseteq [-1,1] \times [-2,2] =: R$, and a, b are the only points of C on the boundary ∂R. The curve $C \cong S^1$ is separated by a, b into two arcs, each of which connects a to b. Let $n = (0,2),\ s = (0,-2)$ be the midpoints of the 'northern' and 'southern' boundaries of R, cf. Figure 12.3. Lemma 12.3.5 implies that the line segment $\overline{ns}$ intersects both arcs. Denote J_n the arc of C whose intersection with $\overline{ns}$ has the largest y-coordinate, and let J_s be the

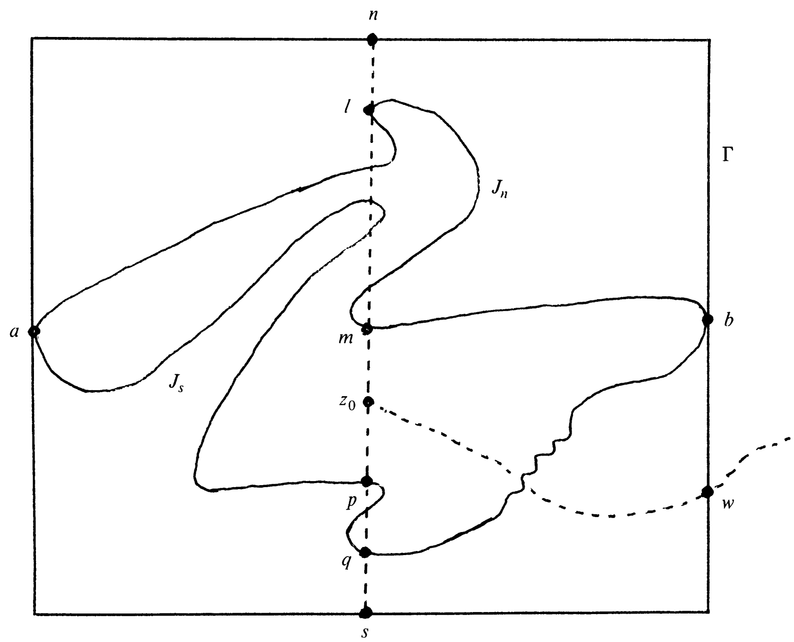


Figure 12.3: From [161] in Amer. Math. Monthly.

other arc. Let l and m be the points in $J_n \cap \overline{ns}$ with maximal, respectively minimal, y-coordinate. (We have $l = m$ if and only if $\#(J_n \cap \overline{ns}) = 1$.) Now the line segment $\overline{ms}$ must meet J_s, since otherwise $\overline{nl} + \widehat{lm} + \overline{ms}$ would be a path from n to s that does not intersect J_s, contradicting Lemma 12.3.5. (Here $\widehat{lm}$ denotes the subarc of J_n between l and m.) Let p and q denote the points in $J_s \cap \overline{ms}$ with maximal, respectively minimal, y-coordinate. (Notice that J_s may intersect $\overline{ns}$ above m, cf. the figure.) Let z be the midpoint of the segment $\overline{mp}$, and call V be the component of $\mathbb{R}^2 \backslash C$ that contains z.

(3) We claim that V is bounded: If this is not the case, and since V is path-connected, there must be a path α from z that leaves the rectangle R. Let w be the first point where α meets the boundary ∂R, and let α_w be the subpath of α from z to w. If w is in the lower half of ∂R (i.e. has y-coordinate < 0) then there is a path $\widehat{ws} \subseteq \partial R$ from w to s containing neither a nor b. Now $\overline{nl} + \widehat{lm} + \overline{mz} + \alpha_w + \widehat{ws}$ is a path from n to s avoiding J_s, which is impossible by Lemma 12.3.5. Similarly, if w is in the upper half of ∂R then we can find $\widehat{wn} \subseteq \partial R$ such that $\overline{sz} + \alpha_w + \widehat{wn}$ avoiding J_n, which again is impossible by Lemma 12.3.5. Thus V must be bounded and in particular $V \neq U$. Therefore $\mathbb{R}^2 \backslash C$ has at least the two components U and V.

(4) We claim that V is the only bounded component: assume there was another bounded component $W \neq U$ of $\mathbb{R}^2 \backslash C$. We would have $W \subseteq R$. Let $\beta = \overline{nl} + \widehat{lm} + \overline{mp} + \widehat{pq} + \overline{qs}$. Since $\overline{nl}$ and $\overline{qs}$ lie in the unbounded component, $\widehat{lm}$ and $\widehat{pq}$ on C and $\overline{mp}$ in U, β is disjoint from W. Since β is closed and does not contain a, b, there are open neighborhoods $N_a \ni a, N_b \ni b$ disjoint from β. But Lemma 12.3.7 implies $\{a, b\} \subseteq C \subseteq \overline{W}$, thus $N_a \cap W \neq \emptyset \neq N_b \cap W$ by Lemma 2.7.3. Choose $a' \in W \cap N_a$, $b' \in W \cap N_b$ and a path $\widehat{a'b'}$ in W. Then the path $\overline{aa'} + \widehat{a'b'} + \overline{b'a}$ is a path from a to b that does not meet the path β from n to s. Since this contradicts Lemma 12.3.5, there is no third component W.

(5) Now that we know that $\mathbb{R}^2 \backslash C$ has precisely two components, Lemma 12.3.7 implies that each of them has C as boundary, and we are done. ∎

Remark 12.3.8 1. According to a theorem of Schoenflies, the Jordan curve theorem can be im-

proved considerably: Whenever $S^1 \cong C \subseteq \mathbb{R}^2$, there is a homeomorphism $\alpha : \mathbb{R}^2 \to \mathbb{R}^2$ that maps the bounded component V of $\mathbb{R}^2 \backslash C$ to the open unit disc $B(0,1)$ and $C = \partial V$ to $S^1 = \partial B$. Fairly few books contain the proof, cf. e.g. [50].

2. For $d > 2$, there is the following analogue of the Jordan curve theorem: Whenever $S^{n-1} \cong C \subseteq \mathbb{R}^n$, the complement $\mathbb{R}^n \backslash C$ has two connected components U, V such that $\partial U = \partial V = C$, with V bounded and U unbounded. Usually this is proven using homology theory, cf. e.g. [28], but there are also more elementary proofs, cf. e.g. [77].

3. However, the analogue of the Schoenflies theorem is false for $d > 2$: In that case, if V is bounded and $\partial V \cong S^{n-1}$, it is not necessarily true that $\overline{V} \cong D^n$, since there are 'wild embeddings' $\iota : S^{n-1} \hookrightarrow \mathbb{R}^n$. (For examples cf. [77, Section 4.3].) If one wants to conclude that the bounded component of $\mathbb{R}^n \backslash \iota(S^{n-1})$ is homeomorphic to a ball, one needs to make stronger assumptions on ι. Cf. e.g. [28, Theorem IV.19.11]. □

12.4 Paths in metric spaces. Geodesic spaces. Length spaces

12.4.1 Geodesic metric spaces. Menger's theorem

In this subsection and the next, we trivially generalize our definition of paths as continuous maps to X defined on $[0, M]$ instead of $[0, 1]$. The following property clearly is a strong form of path-connectedness:

Definition 12.4.1 *Let (X, d) be a metric space. If $x, y \in X$, a map $f : [0, d(x, y)] \to X$ that satisfies $f(0) = x,\ f(d(x, y)) = y$ and is isometric ($d(f(s), f(t)) = |s - t|$) is called a geodesic (or geodesic path) from x to y.*

The space (X, d) is called geodesic if there exists a geodesic from x to y for any $x, y \in X$.

Exercise 12.4.2 If $(V, \|\cdot\|)$ is a normed space and $d(x, y) = \|x - y\|$, prove that (V, d) is geodesic.

We now discuss two simpler conditions:

Definition 12.4.3 *A metric space (X, d) has strict midpoints if for every $x, y \in X$ there exists $z \in X$ with $d(x, z) = d(z, y) = d(x, y)/2$.*

Definition 12.4.4 *Let (X, d) be a metric space.*

- *If $x, y, z \in X$ are all different and $d(x, z) + d(z, y) = d(x, y)$ then 'z lies between x and y'.*
- *If for every $x, y \in X,\ x \neq y$ there is a $z \in X,\ x \neq z \neq y$ between x and y then (X, d) is called metrically convex.*

Exercise 12.4.5 Let $X \subseteq \mathbb{R}^n$ and d the Euclidean metric. Prove that (X, d) is metrically convex if and only if X is convex.

It is obvious that (X, d) is geodesic $\Rightarrow$ (X, d) has strict midpoints $\Rightarrow$ (X, d) is metrically convex. For complete metric spaces the converse implications are true. The first is easy:

Lemma 12.4.6 *Every complete metric space with strict midpoints is geodesic.*

Proof. Let $x \neq y$ and write $T = d(x,y)$. Since (X,d) has strong midpoints, we can find a point $x_{T/2} \in X$ such that $d(x,x_{T/2}) = d(x_{T/2},y) = T/2$, thus $x_{T/2}$ lies between x and y. In the same way we find points $x_{T/4}, x_{3T/4}$ such that $d(x,x_{T/4}) = d(x_{T/4},d_{T/2}) = d(x_{T/2},x_{3T/4}) = d(x_{3T/4},y) = T/4$. Iterating this in the obvious manner we find a point x_r for each $r \in (\mathbb{D} \cap [0,1])T$, where $\mathbb{D}$ are the dyadic rationals, such that $d(r,s) = |r-s|$ for all $r,s \in (\mathbb{D}\cap[0,1])T$. Thus the map $f : (\mathbb{D}\cap[0,1])T \to X,\ t \mapsto x_t$ is an isometry. Since an isometry is uniformly continuous and $(\mathbb{D} \cap [0,1])T$ is dense in $[0,T]$, Corollary 3.4.13 gives a unique $\widehat{f} \in C([0,T],X)$ extending f. By continuity, also $\widehat{f}$ is an isometry. Clearly $\widehat{f}(0) = x, \widehat{f}(T) = y$, thus $\widehat{f}$ is a geodesic from x to y. ■

Notice the close formal similarity between this proof and that of Urysohn's lemma (Theorem 8.2.1)! Since the condition of metric convexity tells us nothing a priori about the position of the points whose existence it asserts, it looks much weaker than that of having strict midpoints. But in combination with completeness we actually have:

Lemma 12.4.7 *Let (X,d) be a metrically convex complete metric space. Then for any $x,y \in X$ with $x \neq y$ and $0 < \lambda < d(x,y)$ there is a $z \in X$ between x and y such that $d(x,z) = \lambda$.*

Combining Lemmas 12.4.6 and 12.4.7 we immediately obtain:

Theorem 12.4.8 (Menger 1928) [5] *Every metrically convex complete metric space is geodesic. (Thus a complete metric space is geodesic if and only if it is metrically convex.)*

It remains to prove Lemma 12.4.7. We write (xzy) when z is between x and y. We will need the following elementary and classical lemma:

Lemma 12.4.9 (Transitivity of betweenness) *In every metric space we have*

$$(pqr) \ \wedge \ (prs) \quad \Leftrightarrow \quad (pqs) \ \wedge \ (qrs).$$

Proof. Assume (pqr) and (prs). Then the points p,q,r,s are all different except perhaps $q = s$. But that would imply $(pqr) \wedge (prq)$. Thus $d(p,q) + d(q,r) = d(p,r), d(p,r) + d(r,q) = d(p,q)$, which in turn imply $d(p,q) < d(p,r), d(p,r) < d(p,q)$, thus a contradiction. Thus $q \neq s$.

Combining $d(p,q) + d(q,r) = d(p,r)$ and $d(p,r) + d(r,s) = d(p,s)$, we obtain

$$d(p,q) + d(q,r) + d(r,s) = d(p,s) \leq d(p,q) + d(q,s).$$

Adding $d(p,q)$ to the triangle inequality $d(q,s) \leq d(q,r) + d(r,s)$, we have

$$d(p,q) + d(q,s) \leq d(p,q) + d(q,r) + d(r,s).$$

Combining the two inequalities, we have

$$d(p,q) + d(q,r) + d(r,s) = d(p,s) \leq d(p,q) + d(q,s) \leq d(p,q) + d(q,r) + d(r,s).$$

Since the left and right most terms agree, the inequalities in fact are equalities, and this gives (pqs) and (qrs), as claimed. The converse implication is proven similarly. ■

If $x \neq y$ and $0 < \lambda < d(x,y)$ we put

$$\begin{aligned} B(x,y) &= \{z \in X \mid (xzy)\}, \\ S(x,y,\lambda) &= \{z \in B(x,y) \mid d(x,z) \leq \lambda\} \cup \{x\}. \end{aligned}$$

[5] Karl Menger (1902-1985), Austrian-American mathematician. Many contributions to dimension theory, metric geometry, mathematical economics, in particular game theory.

Lemma 12.4.10 *Let (X, d) be a complete metric space, $x, y \in X$ with $x \neq y$ and $0 < \lambda < d(x, y)$. Then there is a $z_\lambda \in S(x, y, \lambda)$ such that $u \in B(x, y) \ \wedge \ (xz_\lambda u) \ \Rightarrow \ d(x, u) > \lambda$.*

Proof. The set $\{z \in X \mid d(x, z) + d(z, y) = d(x, y)\}$ is closed (by continuity of d), thus the same is true for its intersection with $\overline{B}(x, \lambda)$, which equals $S(x, y, \lambda) =: S$. Thus (S, d) is complete.

The function $\phi : S \to \mathbb{R}, \ z \mapsto -d(z, x)$ is continuous and bounded below by $-\lambda$. Thus Ekeland's Variational 'Principle' D.2.6(i) implies the existence of a $z \in S$ such that

$$S \ni u \neq z \quad \Rightarrow \quad \phi(u) - \phi(z) > -d(z, u).$$

With our definition of ϕ this means that $S \ni u \neq z$ implies $d(x, z) - d(x, u) > -d(z, u)$, which is the same as $d(x, z) + d(z, u) > d(x, u)$. Thus $u \in S(x, y, \lambda)\backslash\{z\} \Rightarrow \neg(xzu)$. This is equivalent to impossibility of the combination $u \in B(x, y), d(x, u) \leq \lambda, (xzu)$. This in turn is equivalent to $u \in B(x, y) \wedge (xzu) \Rightarrow d(x, u) > \lambda$. Thus $z_\lambda = z$ does the job. ■

Proof of Lemma 12.4.7. Lemma 12.4.10 gives a point $z_\lambda \in S(x, y, \lambda)$ such that:

$$u \in B(x, y) \ \wedge \ (xz_\lambda u) \quad \Rightarrow \quad d(x, u) > \lambda. \tag{12.1}$$

Another invocation of Lemma 12.4.10 with $\lambda' = d(x, y) - \lambda$ gives a $y_{\lambda'} \in S(y, z_\lambda, \lambda')$ such that

$$u \in B(y, z_\lambda) \ \wedge \ (yy_{\lambda'}u) \quad \Rightarrow \quad d(y, u) > \lambda'. \tag{12.2}$$

We claim that $z_\lambda = y_{\lambda'}$. Assume $z_\lambda \neq y_{\lambda'}$. Then by metric convexity of (X, d) there is $w \in X$ such that $(z_\lambda w y_{\lambda'})$. Thus we have the betweenness statements $(xz_\lambda y), (z_\lambda y_{\lambda'} y), (z_\lambda w y_{\lambda'})$. Now the Transitivity Lemma 12.4.9 gives that also $(xwy), xz_\lambda w), (z_\lambda wy), (wy_{\lambda'}y)$ hold. Now, $(xwy) + (xz_\lambda w)+$(12.1) implies $d(x, w) > \lambda$, and $(ywz_\lambda)+(yy_{\lambda'}w)+$(12.2) implies $d(w, y) > \lambda'$. Thus $d(x, y) = d(x, w) + d(w, y) > \lambda + \lambda' = d(x, y)$, which is a contradiction.

With $z_\lambda = y_{\lambda'}$ we have $d(x, z_\lambda) \leq \lambda$ and $d(z_\lambda, y) \leq \lambda' = d(x, y) - \lambda$. Now $d(x, z_\lambda) < \lambda$ would imply the contradiction $d(x, y) \leq d(x, z_\lambda) + d(z_\lambda, y) < \lambda + \lambda' = d(x, y)$. Thus $d(x, z_\lambda) = \lambda$, and we are done. ■

Remark 12.4.11 The above proof of Lemma 12.4.10 is inspired by the one in [93], which however uses Caristi's fixed point theorem instead of Ekeland's variational principle, making it less transparent and less constructive since it then needs the full axiom of choice. □

12.4.2 Path lengths

Definition 12.4.12 *Let (X, d) be a metric space and $f : [a, b] \to X$ a path in X. The length of the path f is defined as*

$$L(f) = \sup \left\{ \sum_{i=0}^{N-1} d(f(t_i), f(t_{i+1})) \mid a = t_0 < t_1 < \cdots < t_N = b \right\} \in [0, \infty]. \tag{12.3}$$

If $L(f) < \infty$, the path is called rectifiable.

Remark 12.4.13 The set $\mathcal{T}$ of finite sequences $\{t_i\}$ containing the endpoints $\{0, M\}$ is partially ordered by inclusion $\{t_0, t_1, \ldots, t_N\} \subseteq \{s_0, s_1, \ldots, s_{N'}\}$ and directed (exactly as in the discussion of

Example 5.1.23). Thus the map $\{t_0, t_1, \ldots, t_N\} \mapsto \sum_{i=0}^{N-1} d(f(t_i), f(t_{i+1}))$ is a net $\mathcal{T} \to [0,\infty)$. If $T \le S$, the triangle inequality gives (via an easy induction)

$$\sum_{i=0}^{N-1} d(f(t_i), f(t_{i+1})) \le \sum_{i=0}^{N'-1} d(f(s_i), f(s_{i+1})).$$

Thus the net is increasing, which implies that its limit equals its supremum. □

Exercise 12.4.14 Let $f : [0, M] \to \mathbb{R}^n$ be a path such that the coordinate functions $f_i = p_i \circ f : [0, M] \to \mathbb{R}$ are continuously differentiable. Prove that

$$L(f) = \int_0^M \sqrt{\sum_{i=1}^n f_i'(t)^2 dt}, \tag{12.4}$$

thus f is rectifiable. (The integral is the Riemann integral.)

Remark 12.4.15 Differentiability of the f_i is a rather strong condition. One can show that a path f in $\mathbb{R}^n$ is rectifiable if and only if all the f_i have bounded variation (in addition go being continuous, of course), cf. e.g. [218, Theorem III.3.1]. Under this condition, all f_i are differentiable almost everywhere, but (12.4) need not hold, even if understood as Lebesgue integral. (This is already clear for $n = 1$ if f_1 is the Cantor-Lebesgue function, cf. Section 11.2.2. The latter has derivative zero almost everywhere, so that (12.4) vanishes, whereas $L(f) = d(f_1(0), f_1(1)) = |1 - 0| = 1$.) However, (12.4) does hold under the stronger assumption that each f_i is absolutely continuous, cf. e.g. [218, Theorem III.4.1]. (Absolute continuity of a function f is equivalent to f being continuous, having bounded variation and mapping sets of measure zero to sets of measure zero. The latter condition is clearly violated by the Cantor-Lebesgue function since it maps to Cantor set Γ to $[0,1]$.) □

Lemma 12.4.16 *Let (X, d) be a metric space and $f : [0, M] \to X$ a path from x to y.*

(i) We have $L(f) \ge d(x, y)$.

(ii) If $d(f(t), f(t')) \le C|t - t'|\ \forall t, t'$ then f is rectifiable and $L(f) \le CM$.

(iii) If $0 < N < M$ then $L(f) = L(f \restriction [0, N]) + L(f \restriction [N, M])$. In particular f is rectifiable if and only if $f \restriction [0, N]$ and $f \restriction [N, M]$ are rectifiable.

(iv) If $f : [a', b'] \to X$ is a path and $h : [a, b] \to [a', b']$ is continuous, non-decreasing and surjective then $L(f \circ h) = L(f)$.

(v) If f is rectifiable then the map $\ell : [0, M] \to [0, L(f)],\ t \mapsto L(f \restriction [0, t])$ is non-decreasing, continuous and surjective.

(vi) The map ℓ from (v) is constant on $[a, b] \subseteq [0, M]$ if and only if f is constant on $[a, b]$.

(vii) The map $C([0, M], X) \to [0, \infty],\ f \mapsto L(f)$ is lower semicontinuous (w.r.t. either pointwise or uniform convergence of paths).

Proof. (i) $L(f)$ is defined as a supremum over all partitions of $[0,M]$. But taking the minimal partition $\{0,M\}$, the right hand side of (12.3) is just $d(x,y)$.

(ii) The assumption implies, for every partition $\{t_0,\dots,t_N\}\in\mathcal{T}$, that

$$\sum_{i=0}^{N-1} d(f(t_i),f(t_{i+1}))\le C\sum_{i=0}^{N-1}|t_i-t_{i+1}|=CM.$$

Thus also the supremum $L(f)$ of this over $\mathcal{T}$ is bounded by CM.

(iii) Since $L(f)$ is defined as the supremum over the set of all partitions $\mathcal{T}$ and the net $\mathcal{T}\to[0,\infty]$ is non-decreasing, we may as well restrict ourselves to partitions containing the point N. But then the additivity is obvious.

(iv) $f'=f\circ h$ is continuous, thus a path. If $\{a=t_0<t_1<\dots<t_N=b\}$ is a partition of $[a,b]$, in computing $L(f\circ h)$ we encounter the sum

$$\sum_{i=0}^{N-1} d(f'(t_i),f'(t_{i+1}))=\sum_{i=0}^{N-1} d(f(h(t_i)),f(h(t_{i+1}))). \tag{12.5}$$

Since h is non-decreasing, $\{h(t_i),i=0,\dots,N\}$ is a partition of $[a',b']$, and since h is surjective, the supremum of (12.5) over the partitions of $[a,b]$ coincides with the supremum over the partitions of $[a',b']$. Thus $L(f\circ h)=L(f)$.

(v) Let $0\le t\le t'$. Then by (iii), $L(f\restriction[0,t'])=L(f\restriction[0,t])+L(f\restriction[t,t'])$. Since lengths are non-negative, this implies $L(f\restriction[0,t'])\ge L(f\restriction[0,t])$.

Let $s\in(0,M]$ and $\varepsilon>0$. Choose a partition $0=t_0,\dots,t_N=s$ such that $d(f(t_{N-1},f(s))<\varepsilon/2$ and $\sum_{i=0}^{N-1}d(f(t_i,t_{t+1})>L(f\restriction[0,s])-\varepsilon/2=\ell(s)-\varepsilon/2$. Then

$$\begin{aligned}\ell(s) &< \varepsilon/2+\sum_{i=0}^{N-2}d(f(t_i),f(t_{i+1}))+d(f(t_{N-1}),f(t_N))\\ &\le \varepsilon/2+L(f\restriction[0,t_{N-1}])+\varepsilon/2=\ell(t_{N-1})+\varepsilon.\end{aligned}$$

Thus with $s'=t_{N-1}<s$ we have $\ell(s')>\ell(s)-\varepsilon$, proving continuity of ℓ from the left. Now the additivity $\ell(t)=\ell(s)+L(f\restriction[s,t])$ proves that $s\mapsto\ell(s)$ is also continuous from the right.

It is clear that $\ell(0)=0$ and $\ell(M)=L(f)$. Since ℓ is continuous, the intermediate value theorem implies $\ell([0,M])=[0,L(f)]$, i.e. surjectivity.

(vi) In view of additivity (iii), it is enough to show that $L(f)=0\Leftrightarrow f$ is constant. The implication $\Leftarrow$ is obvious. If f is non-constant, e.g. $f(t)\ne f(t)$ for some $t<t'$, then $L(f)\ge L(f\restriction[t,t'])\ge d(f(t),f(t'))>0$.

(vii) We will prove $\liminf L(f_i)\ge L(f)$ whenever $\{f_\iota\}$ is a net of rectifiable paths converging pointwise to f. This implies the claim for both topologies on the $C([0,M],X)$. Let $\varepsilon>0$. Pick a partition $0=t_0,\dots,t_N=M$ such that $\sum_{k=0}^{N-1}d(f(t_k),f(t_{k+1}))>L(f)-\varepsilon$. Since $f_\iota\to f$ pointwise (and the partition has finitely many points), there is a ι such that $d(f_\iota(t_k),f(t_k))<\varepsilon/2N$ for all $k=0,\dots,N$. Then $d(f(t_k),f(t_{k+1}))\le d(f_\iota(t_k),f_\iota(t_{k+1}))+\varepsilon/N$, thus

$$L(f)\le\varepsilon+\sum_{k=0}^{N-1}d(f(t_k),f(t_{k+1}))\le\varepsilon+\varepsilon+\sum_{k=0}^{N-1}d(f_\iota(t_k),f_\iota(t_{k+1}))\le 2\varepsilon+L(f_\iota).$$

Since ε was arbitrary, the claim follows. (For the uniform topology, which is metric and thus first countable, one may replace 'net' by 'sequence'.) ■

By Lemma 12.4.16, reparametrizations $f\rightsquigarrow f\circ h$, where h is continuous non-decreasing, do not change the length of a curve. It is therefore natural to ask whether there is a best way to parametrize a curve.

Definition 12.4.17 *If a path $f : [0, M] \to (X, d)$ satisfies $L(f \restriction [0, t]) = t$ for all $t \in [0, M]$, it is* <u>*parametrized by arc-length*</u>*. (Obviously, then $L(f) = M < \infty$, thus f is rectifiable.)*

Lemma 12.4.18 *A path f from x to y is geodesic if and only if it is arc-length parametrized and $L(f) = d(x, y)$.*

Proof. $\Rightarrow$ A geodesic f is isometric, thus satisfies the assumption of Lemma 12.4.16 (ii) with $C = 1$. This gives $L(f) \leq M$. Combining this with Lemma 12.4.16(i), we have $d(x, y) \leq L(f) \leq M$. Since f is geodesic, we have $d(x, y) = d(f(0), f(M)) = M$, and thus $L(f) = M = d(x, y)$. Applying the same consideration to $f \restriction [0, t]$ gives $L(f \restriction [0, t]) = d(f(0), f(t)) = t$, thus f is parametrized by arc-length.

$\Leftarrow$ If $f : [0, M] \to X$ is parametrized by arc-length then clearly $M = L(f)$. By assumption, $L(f) = d(x, y)$, thus $M = d(x, y)$. Again by arc-length parametrization, we have

$$t = L(f \restriction [0, t]) \geq d(x, f(t)) \quad \text{and} \quad d(x, y) - t = L(f \restriction [t, d(x, y)]) \geq d(f(t), y)$$

for all $t \in [0, d(x, y)]$. Adding these two inequalities and using the triangle inequality we have

$$d(x, y) \geq d(x, f(t)) + d(f(t), y) \geq d(x, y),$$

thus $d(x, f(t)) + d(f(t), y) = d(x, y)$ for every t. (In the terminology of the preceding section: $f(t)$ is between x and y.) Combining this with $d(x, f(t)) \leq t$, $d(f(t), y)) \leq d(x, y) - t$ we have $d(x, f(t)) = t$ and $d(f(t), y) = d(x, y) - t$. If now $t \leq t'$ then

$$d(f(t), f(t')) \geq d(x, y) - d(x, f(t)) - d(f(t'), y) = d(x, y) - t - (d(x, y) - t') = t' - t$$

by the triangle inequality. On the other hand, $d(f(t), f(t')) \leq L(f \restriction [t, t']) = t' - t$ by Lemma 12.4.16(i) and arc-length parametrization. Thus $d(f(t), f(t')) = |t' - t|$, so that f is geodesic. ∎

Example 12.4.19 Consider the path given by $f : [0, 2\pi] \to \mathbb{R}^2$, $t \mapsto (\cos t, \sin t)$. This is well-known to be parametrized by arc-length with $L(f) = 2\pi$, but f is not geodesic since $d(f(0), f(2\pi)) = 0 \neq L(f)$. □

Proposition 12.4.20 *Let $f : [0, M] \to X$ be rectifiable. Define $\ell(t) = L(f \restriction [0, t])$. Then there is a rectifiable path $g : [0, L(f)] \to X$ that is parametrized by arc-length and satisfies $f = g \circ \ell$.*

Proof. Consider the diagram

$$\begin{tikzcd}
[0, M] \arrow[d, "\ell"'] \arrow[dr, "f"] & \\
[0, L(f)] \arrow[r, "g"'] & X
\end{tikzcd}$$

where f is given and ℓ is as in Lemma 12.4.16(vi), and thus continuous and surjective. If $t < t'$ and $\ell(t) = \ell(t')$ then the monotonicity of ℓ implies that ℓ is constant on $[t, t']$, and Lemma 12.4.16(vii) implies that f is constant on $[t, t']$. Now Proposition 6.4.8 implies that there is unique a continuous function $g : [0, L(f)] \to X$ such that $f = g \circ \ell$. (The topology on $[0, L(f)]$ is the quotient topology defined by ℓ since $[0, M]$ is compact.)

Lemma 12.4.16(v-vi) imply $\ell(t) = L(f \restriction [0, t]) = L(g \circ \ell \restriction [0, t]) = L(g \restriction [0, \ell(t)])$ for all $t \in [0, M]$. Since ℓ is surjective, this implies $L(g \restriction [0, t]) = t$, thus g is parametrized by arc-length. ∎

Remark 12.4.21 1. By Lemma 12.4.16(vii), $\ell(t) = L(f \restriction [0,t])$ is strictly increasing if and only if f is not constant on any interval. In this case ℓ has a continuous inverse function ℓ^{-1}, and we have $g = f \circ \ell^{-1}$.

2. If f is a path from x to y such that $L(f) = d(x,y)$ then its arc-length reparametrization g is geodesic. (This follows from Lemma 12.4.18 since $L(g) = L(f)$.)

3. Some authors, e.g. of [13], call our geodesic paths 'minimal geodesic' and use 'geodesic' for paths parametrized by arc-length (possibly rescaled). □

12.4.3 Length spaces. The Hopf-Rinow theorem

Definition 12.4.22 *A metric space (X,d) is a length space if for any two $x, y \in X$ and $\varepsilon > 0$ there is a path f from x to y such that $L(f) < d(x,y) + \varepsilon$.*

Remark 12.4.23 1. By the very definition, every length space is path connected. And in view of Lemma 12.4.18 every geodesic space is a length space.

2. $\mathbb{R}^n$ with one point removed is not geodesic, but a length space. □

Exercise 12.4.24 Recall from Exercise 2.2.11 that

$$\overline{B(x,r)} = \overline{B}(x,r) \quad \forall x \in X, \ r > 0 \tag{12.6}$$

does not hold in every metric space.

(i) Prove that (12.6) is true in all length spaces.

(ii) Give an example showing that (12.6) does not follow from path-connectedness.

Definition 12.4.25 *A metric space (X,d) has approximate midpoints if for every $x, y \in X$ and $\varepsilon > 0$ there is $z \in X$ such that $\max(d(x,z), d(z,y)) \le \frac{d(x,y)}{2} + \varepsilon$.*

Proposition 12.4.26 *(i) Every length space has approximate midpoints.*

(ii) Every complete metric space having approximate midpoints is a length space.

Proof. (i) Let $x, y \in X$ and $\varepsilon > 0$. By assumption there is a path $f : [0,M] \to X$ from x to y with $L(f) < d(x,y) + 2\varepsilon$. By Lemma 12.4.16(vi), the map $\ell : t \mapsto L(f \restriction [0,t])$ is surjective onto $[0, L(f)]$. Thus there is a $t \in [0,M]$ such that $\ell(t) = L(f \restriction [0,t]) = L(f)/2$. By additivity of lengths, also $L(f \restriction [t,M]) = L(f)/2$. Defining $z = f(t)$ we have $d(x,z) \le L(f \restriction [0,t]) = L(f)/2 < d(x,y)/2 + \varepsilon$, and similarly $d(z,y) < d(x,y)/2 + \varepsilon$.

(ii) The proof is quite similar to that of Lemma 12.4.6, but a bit more involved since we have to take some ε's into account. Let $x, y \in X$ and $\varepsilon_k > 0$ for all $k \in \mathbb{N}$ such that $C = \prod_{k=1}^{\infty}(1+\varepsilon_k) < \infty$. (We know from analysis that this is equivalent to $\sum_k \varepsilon_k < \infty$.) By the existence of approximate midpoints, we can find $z_{1/2} \in X$ such that $\max(d(x,z_{1/2}), d(z_{1/2},y)) \le (1+\varepsilon_1)\frac{d(x,y)}{2}$. Then we choose points $z_{1/2}, z_{3/4}$ such that

$$\max(d(x,z_{1/4}), d(z_{1/4},z_{1/2}), d(z_{1/2},z_{3/4}), d(z_{3/4},y)) \le (1+\varepsilon_2)\frac{(1+\varepsilon_1)\frac{d(x,y)}{2}}{2}.$$

Iterating this construction we find points $z_t \in X$ for all $t \in \mathbb{D} \cap (0,1)$ such that $d(z_t, z_{t'}) \le C d(x,y)|t - t'| \ \forall t, t' \in \mathbb{D} \cap [0,1]$ (where $z_0 = x, z_1 = y$). Then the map $f : \mathbb{D} \cap [0,1] \to X, \ t \mapsto z_t$ is uniformly

continuous, and since $\mathbb{D} \cap [0,1]$ is dense in $[0,1]$, Corollary 3.4.13 gives us a (unique) extension $\widehat{f} : [0,1] \to X$ satisfying $d(\widehat{f}(t), \widehat{f}(t')) \le Cd(x,y)|t-t'|\ \forall t,t' \in [0,1]$. By Lemma 12.4.16(ii), f is a rectifiable path of length $L(f) \le Cd(x,y)$. By appropriate choice of the ε_k we can bring $C > 1$ arbitrarily close to 1. Thus (X,d) is a length space. ■

Theorem 12.4.27 (Hopf-Rinow) [6] *Let (X,d) be a locally compact length space. Then the following are equivalent:*

(i) (X,d) is proper. (I.e. closed bounded subsets are compact.)

(ii) (X,d) is complete.

(iii) (X,d) is geodesically complete: Every isometry $f : [0,M) \to X$ can be extended to $[0,M]$.

(iv) There is a point $x_0 \in X$ such that every isometry $f : [0,M) \to X$ with $f(0) = x_0$ can be extended to $[0,M]$.

These (equivalent) conditions imply that (X,d) is geodesic.

Proof. (i)⇒(ii) By Lemma 7.8.78, every proper metric space is complete (and locally compact).

(ii)⇒(iii) Since isometries are uniformly continuous, this is an immediate consequence of Corollary 3.4.13.

(iii)⇒(iv) Obvious.

(iv)⇒(i) Defining $R = \sup\{r \mid \overline{B}(x_0,r) \text{ is compact}\}$, local compactness of X implies $R > 0$. Properness of X, i.e. compactness of all closed balls, is equivalent to $R = \infty$. Assuming $R < \infty$, we will use (iii) to prove that $\overline{B(x_0,R)}$ is compact.

****

In view of Exercise 12.4.24 this implies that $\overline{B}(x_0,R)$ is compact. Since X is locally compact, for every $x \in \overline{B}(x_0,R)$ there is $\varepsilon_x > 0$ such that $\overline{B}(x,\varepsilon_x) = \overline{B(x,\varepsilon_x)}$ is compact. Since $\overline{B}(x_0,R)$ is compact, there are $x_1,\ldots,x_n$ such that $\overline{B}(x_0,R) \subseteq \bigcup_{i=1}^n B(x_i,\varepsilon_{x_i})$. Now, $\bigcup_{i=1}^n \overline{B}(x_i,\varepsilon_{x_i})$ is compact, and this contains $\overline{B}(x_0,R+\varepsilon)$ where $\varepsilon = \min(\varepsilon_{x_1},\ldots,\varepsilon_{x_n}) > 0$, which therefore is compact. This contradicts the definition of R. Thus $R = \infty$, proving properness

Now we prove that every proper length space is geodesic. Let $x \ne y$. By the length space property we can choose, for every $n \in \mathbb{N}$ a path f_n from x to y such that $L(f_n) \le d(x,y) + 1/n$. By Proposition 12.4.20 we can first parametrize the f_n by arc-length, and then rescale linearly so that they are defined on $[0,1]$. Then

$$|t-t'| = \frac{L(f_n \restriction [t,t'])}{L(f_n)} \ge \frac{d(f_n(t), f_n(t'))}{d(x,y)+1},$$

implying that the family $\{f_n\}$ is equicontinuous (even uniformly). All paths f_n begin at x and have length $\le d(x,y)+1$, thus they live in $\overline{B}(x,d(x,y)+1) \subseteq Y$, which is compact by properness. Thus the assumptions of the Ascoli-Arzelà Theorem 7.7.67 are satisfied, and we obtain a subsequence $\{f_{n_m}\}$ of $\{f_n\}$ that converges uniformly to some $g \in C([0,1],X)$. By Lemma 12.4.16(viii) the map $f \mapsto L(f)$ is lower semicontinuous, thus in view of Exercise 5.2.33 we have $L(g) \le \liminf L(f_{n_m}) = d(x,y)$. Since the converse $L(g) \ge d(x,y)$ holds for every path from x to y, we have $L(g) = d(x,y)$. It is clear from the construction that g is parameterized by arc-length, rescaled to $[0,1]$. Defining $g'(t) = g\left(\frac{t}{d(x,y)}\right)$, it follows that g' is parametrized by arc-length. This does not affect the length, thus $L(g') = L(g) = d(x,y)$, so that Lemma 12.4.18 implies that g' is a geodesic. ■

[6] Heinz Hopf (1894-1971), Willi Rinow (1907-1979). German mathematicians.

Corollary 12.4.28 *A length space is proper if and only if it is locally compact and complete.*

Remark 12.4.29 1. The Hopf-Rinow theorem was proved in 1931 (by Hopf and Rinow) for Riemannian manifolds and generalized to metric spaces by S. Cohn-Vossen.

2. The construction of geodesics in the second half of the proof can be considered a part of variational calculus, where one looks for the function f minimizing a certain functional F, here the length $L(f)$. This is a very classical subject going back to the 18th century (the Bernoullis, Euler, Lagrange) and was originally approached by transforming the variational problem into a (roughly) equivalent differential equation. This works for sufficiently smooth problems like those considered originally (the isoperimetric problem and the 'brachistochrone problem'). However, in our metric setting one needs to use the 'direct method' of variational calculus, which was invented much later: One picks a sequence $\{f_i\}$ of functions such that $F(f_i)$ converges to the infimum of F, uses a compactness argument to find a subsequence of $\{f_i\}$ that converges uniformly to some function f and finally invokes lower semicontinuity of F to conclude $F(f) = \inf F$. (The first to propose this was Riemann with his 'Dirichlet principle' (1851, 1857), whose rigorous proof was only given in 1900 by Hilbert.) For more on direct methods in variational calculus, cf. e.g. [157] or the book-length treatment [91].

3. The subject of length and geodesic spaces discussed above belongs to 'metric geometry' or the 'geometry of metric spaces'. People started being interested in metric spaces for the purposes of *geometry* at the same time when metric spaces were displaced by topological spaces within *topology*. One motivation was to generalize results from differential geometry (Riemannian manifolds) to the simpler and more general setting of metric spaces. For the classical results see books like [22, 197] from the 1950s. More recently, metric geometry became quite popular again, in particular the study of metric spaces with 'curvature' bounded above or below, cf. e.g. [41, 39, 31]. This subject has many ramifications towards combinatorial/geometric group theory (hyperbolic groups), to analysis on groups and 'non-commutative geometry' [186, 198] and analysis on metric spaces [4, 110]. We (reluctantly) leave this beautiful subject here. □

Chapter 13

Homotopy. The Fundamental Group(oid). Coverings

13.1 Homotopy of maps and spaces. Contractibility

In this section, $I = [0,1]$.

Definition 13.1.1 *Let X, Y be topological spaces and $f, g \in C(X,Y)$. A homotopy from f to g, occasionally denoted $h : f \to g$, is a continuous function $h : X \times I \to Y$, $(x,t) \mapsto h(x,t)$ such that $h(x,0) = f(x)$ and $h(x,1) = g(x)$ $\forall x \in X$. We usually write $h_t(x)$ instead of $h(x,t)$.*

Two functions $f, g \in C(X,Y)$ are called homotopic, $f \sim g$, if there exists a homotopy from f to g.

Homotopies are, in a sense, arrows between continuous maps that have the same source and target. A good way to visualize the situation is this:

$$\begin{tikzcd} X \arrow[r, bend left=50, "f"] \arrow[r, bend right=50, "g"'] \arrow[r, phantom, "\Downarrow h"] & Y \end{tikzcd}$$

Lemma 13.1.2 *(i) If $f, g, k \in C(X,Y)$ and $h : f \to g$ and $h' : g \to k$ are homotopies then*

$$h''_t(x) = \begin{cases} h_{2t}(x) & \text{if } t \in [0,1/2], \\ h'_{2t-1}(x) & \text{if } t \in [1/2,1] \end{cases}$$

defines a homotopy $h'' = h' \circ h : f \to k$. Symbolically:

$$\begin{tikzcd} X \arrow[r, bend left=70, "f"] \arrow[r, "g"] \arrow[r, bend right=70, "k"'] \arrow[r, phantom, shift left=3ex, "\Downarrow h"] \arrow[r, phantom, shift right=3ex, "\Downarrow h'"] & Y \end{tikzcd} \quad \rightsquigarrow \quad \begin{tikzcd} X \arrow[r, bend left=50, "f"] \arrow[r, bend right=50, "k"'] \arrow[r, phantom, "\Downarrow h''"] & Y \end{tikzcd}$$

(ii) For all X, Y, homotopy is an equivalence relation on $C(X,Y)$.

Proof. (i) This is quite obvious: We have $h''_0 = h_0 = f$ and $h''_1 = h'_1 = k$. Continuity at $t = 1/2$ follows from $h_1(x) = h'_0(x) = g(x)$.

(ii) If $f \in C(X,Y)$ then defining $h_t(x) = f(x)$, it is clear that h is a homotopy from f to f, thus $f \sim f$. If h is a homotopy from f to g then h_{1-t} is a homotopy from g to f, thus $f \sim g$ implies $g \sim f$. That $f \sim g \sim k$ implies $f \sim k$ follows from (i). ■

Remark 13.1.3 If one topologizes $C(X,Y)$ suitably it turns out that a homotopy between $f, g \in C(X,Y)$ is nothing but a continuous path in $C(X,Y)$ from f to g. See Section 7.9, in particular Theorem 7.9.11. (In the rather special case where X is compact and Y metrizable, this follows from the bijection $C(Z \times X, Y) \leftrightarrow C(Z, C(X,Y))$ established in Exercise 7.7.45 by taking $Z = [0,1]$.) Then the composition of homotopies reduces to the composition of paths in $C(X,Y)$ and the lemma reduces to the fact that the path relation $\sim_p$ from Section 12.1 is an equivalence relation. □

Lemma 13.1.4 *If $f, g : X \to Y$, $f', g' : Y \to Z$ and $h : f \to g$ and $h' : f' \to g'$ are homotopies then $h''_t = h'_t \circ h_t$ is a homotopy $f' \circ f \to g' \circ g$.*

$$\begin{tikzcd} X \arrow[r, bend left=50, "f"] \arrow[r, bend right=50, "g"'] \arrow[r, phantom, "\Downarrow h"] & Y \arrow[r, bend left=50, "f'"] \arrow[r, bend right=50, "g'"'] \arrow[r, phantom, "\Downarrow h'"] & Z \end{tikzcd} \quad \rightsquigarrow \quad \begin{tikzcd} X \arrow[r, bend left=50, "f'\circ f"] \arrow[r, bend right=50, "g'\circ g"'] \arrow[r, phantom, "\Downarrow h'\circ h"] & Z \end{tikzcd}$$

Proof. Let $h : X \times I \to Y$ be a homotopy from f to g and $h' : Y \times I \to Z$ a homotopy from f' to g'. Define $h''_t(x) = h'_t \circ h_t(x)$. Now $(x,t) \mapsto h_t(x)$ is a continuous function $X \times I \to Z$ that clearly satisfies $h''_0(x) = f' \circ f(x)$ and $h''_1(x) = g' \circ g(x)$. Thus $g' \circ g \sim f' \circ f$. ■

Remark 13.1.5 We thus have defined two different ways of composing homotopies: In Lemma 13.1.2 we have the <u>vertical</u> composition of homotopies $h : f \to g$ and $h' : g \to k$, where $f, g, k \in C(X,Y)$, giving rise to a homotopy $h'' : f \to k$. While in Lemma 13.1.4 we consider the <u>horizontal</u> composition of homotopies $h : f \to g$ and $h' : f' \to h'$, where $f, g \in C(X,Y)$ and $f', g' \in C(Y,Z)$, giving a homotopy $h'' : f' \circ f \to g' \circ g$. These compositions should not be confused! (Yet we denote both by $\circ$.) □

Definition 13.1.6 *A map $f \in C(X,Y)$ is called a <u>homotopy equivalence</u> if there is a map $g \in C(Y,X)$ such that $g \circ f \in C(X,X)$ is homotopic to id_X and $f \circ g \in C(Y,Y)$ is homotopic to id_Y. Two spaces X, Y are called <u>homotopy equivalent</u>, denoted $X \sim Y$, if there is a homotopy equivalence $f : X \to Y$.*

Exercise 13.1.7 Show that homotopy equivalence of spaces is an equivalence relation.

Remark 13.1.8 There is a third way of composing homotopies! If $h_1 : f_1 \to g_1$ is a homotopy between $f_1, g_1 \in C(X_1, Y_1)$ and $h_2 : f_2 \to g_2$ is a homotopy between $f_2, g_2 \in C(X_2, Y_2)$ then $h_{3,t} = h_{1,t} \times h_{2,t}$ is a homotopy between $f_1 \times f_2$ and $g_1 \times g_2$, both of which are maps $X_1 \times X_2 \to Y_1 \times Y_2$.

Using this one easily shows that the existence of homotopy equivalences $X \sim X'$, $Y \sim Y'$ implies a homotopy equivalence $X \times Y \sim X' \times Y'$.

(In modern language this means that topological spaces, continuous maps and homotopies form a 'monoidal 2-category'...) □

Since classification of topological spaces up to homeomorphism is hopeless, one may hope to classify them up to homotopy equivalence. This is still too hard, but there is some hope. But we will see that homotopy ideas can also be used to distinguish certain spaces that are homotopy equivalent (but not homeomorphic), like the spaces $\mathbb{R}^n$ for different $n \in \mathbb{N}$.

Definition 13.1.9 *A space (X, τ) is* <u>*contractible*</u> *if it is homotopy equivalent to a one-point space.*

Exercise 13.1.10 (i) Show that a space (X, τ) is contractible if and only if id_X is homotopic to a constant map.

(ii) Deduce that a contractible space is path-connected.

(iii) Prove that every star-shaped subspace $X \subseteq \mathbb{R}^n$ is contractible.

Combining with Exercise 12.1.9, we have the implications

$$\text{convex} \Rightarrow \text{star-shaped} \;\Rightarrow\; \text{contractible} \;\Rightarrow\; \text{path-connected}.$$

The converses of the above implications are far from true. This is easy to see for the first two implications, but proving non-contractibility for a given path-connected space can be quite hard – even when it is intuitively obvious as in the following case.

Theorem 13.1.11 *S^n is not contractible for all $n \geq 0$. ($S^0 = \{\pm 1\}$.)*

Proof. By Exercise 13.1.10, contractibility of S^{n-1} is equivalent to existence of a homotopy h_t : $S^{n-1} \to S^{n-1}$ satisfying $h_0(\cdot) = x_0$ and $h_1 = \mathrm{id}_{S^{n-1}}$. But then the map $r : D^n \to \partial D^n$ defined by

$$r(x) = \begin{cases} x_0 & \|x\| \leq \frac{1}{2} \\ h_{2\|x\|-1}(\frac{x}{\|x\|}) & \|x\| \geq \frac{1}{2} \end{cases}$$

is a retraction, contradicting Corollary 10.3.8(ii). ■

Remark 13.1.12 For $n = 0$, this is just connectedness of $[0, 1]$. For $n = 1$ a short direct proof is given below, and yet another one follows from the computation of the fundamental group $\pi_1(S^1)$. □

Corollary 13.1.13 *If $X \subseteq \mathbb{R}^n$ is compact convex then ∂X is not contractible (unless $\#X = 1$).*

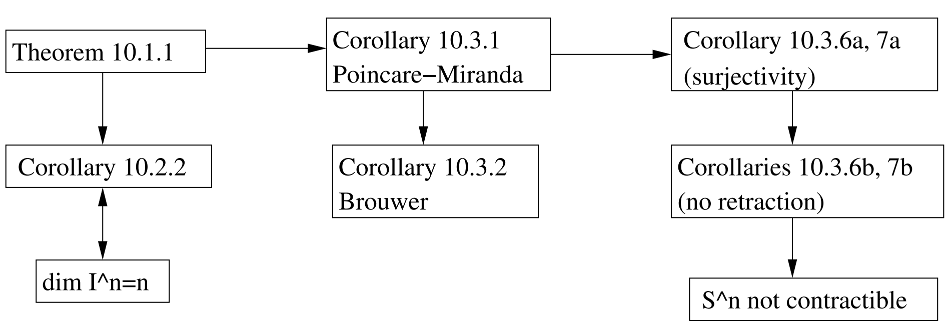


Figure 13.1: Implications involving higher connectedness

Remark 13.1.14 1. The preceding results, as well as those of Sections 10.3 and 10.4 were all deduced from Theorem 10.1.2, cf. Figure 13.1. Usually these results are proven using homology theory, which most naturally yields the non-existence of a retraction $D^n \to S^{n-1}$. The following exercise shows that the implications in Figure 13.1 can be reversed. □

Exercise 13.1.15 Prove the following implications:

(i) $\mathrm{id}_{S^{n-1}}$ not homotopic to constant map $\Rightarrow$ non-existence of retractions $D^n \to S^{n-1}$. (Hint: $h_t(x) = r(tx)$.)

(ii) Non-existence of retractions $D^n \to S^{n-1}$ $\Rightarrow$ if $f : D^n \to D^n$ satisfies $f \upharpoonright \partial D^n = \mathrm{id}$ then $f(D^n) = D^n$. (If $f \upharpoonright \partial D^n = \mathrm{id}$, but $f(D^n) \neq D^n$, use f to produce a retraction.)

(iii) Non-existence of retractions $D^n \to S^{n-1}$ $\Rightarrow$ Brouwer's fixed point theorem. (Assume $f : D^n \to D^n$ has no fixed point, and define $r(x)$ as the point where the ray from $f(x)$ to x meets S^{n-1}.)

(iv) Brouwer's fixed point theorem $\Rightarrow$ Poincaré-Miranda theorem. (Define $g_i(x) = x_i + \varepsilon_i f_i(x)$ and prove that g has a fixed point for suitable choice of the ε_i.)

(v) Poincaré-Miranda theorem $\Rightarrow$ Theorem 10.1.2. (Hint: $f_i(x) = d(x, H_i^-) - d(x, H_i^+)$.)

Theorem 13.1.11 can be used to give a new proof of the statement in Exercise 10.3.11:

Theorem 13.1.16 *Let $n \in \mathbb{N}$ and $f \in C(D^n, \mathbb{R}^n)$ such that $f(x) \cdot x > 0$ for all $x \in S^{n-1}$ or $f(x) \cdot x < 0$ for all $x \in S^{n-1}$. Then there is $x \in (D^n)^0$ such that $f(x) = 0$.*

Proof. We consider the first case. We have $f(x) \neq 0$ for all $x \in S^{n-1}$, so that $g : x \mapsto \frac{f(x)}{\|f(x)\|}$ defines a continuous map $S^{n-1} \to S^{n-1}$. We claim that g is homotopic to $\mathrm{id}_{S^{n-1}}$. To see this, observe that for all $x \in S^{n-1}$, $t \in [0,1]$ we have

$$[(1-t)f(x) + tx] \cdot x = (1-t)f(x) \cdot x + tx \cdot x = (1-t)f(x) \cdot x + t > 0,$$

thus $(1-t)f(x) + tx \neq 0$. Thus $h(x,t) = \frac{(1-t)f(x)+tx}{\|(1-t)f(x)+tx\|}$ defines a continuous map $S^{n-1} \times [0,1] \to S^{n-1}$. Clearly $h(\cdot, 0) = g$ and $h(\cdot, 1) = \mathrm{id}_{S^{n-1}}$, thus h is a homotopy from g to $\mathrm{id}_{S^{n-1}}$.

Assuming $f(x) \neq 0$ for all $x \in D^n$, we can define $k : S^{n-1} \times [0,1] \to S^{n-1}$, $(x,t) \mapsto \frac{f(tx)}{\|f(tx)\|}$. Then $k(\cdot, 0) = \frac{f(0)}{\|f(0)\|} =: c$, and $k(x, 1) = \frac{f(x)}{\|f(x)\|} = g(x)$. Thus k is a homotopy from the constant map $x \mapsto c$ to g.

Thus there is a composite homotopy $h \circ k$ from the constant map $x \mapsto c$ to $\mathrm{id}_{S^{n-1}}$, contradicting Theorem 13.1.11. Thus f must have a zero, which then must lie in $\mathrm{Int}(D^n)$. ■

Exercise 13.1.17 Let X, Y be topological spaces. Prove:

(i) If $f, g : X \to Y$ are continuous functions that are homotopic then $\pi_0(f) = \pi_0(g)$.

(ii) If $X \sim Y$ (homotopy equivalence) then $\pi_0(X) \cong \pi_0(Y)$ (bijection).

13.2 Alternative proof of non-contractibility of S^1. Borsuk-Ulam for S^1, S^2

We begin with an easy observation:

Lemma 13.2.1 *There is no embedding $S^1 \hookrightarrow \mathbb{R}$.*

Proof. Let $f : S^1 \to \mathbb{R}$. Let $W = (-1,0), O = (1,0)$ and $a = f(W), b = f(O)$. Recall that $S^1_\pm = \{x \in S^1 \mid \pm x_2 \geq 0\} \cong I$. Therefore the intermediate value theorem (Corollary 9.2.9) implies that $f(S^1_+)$ and $f(S^1_-)$ both contain the entire interval between a and b. Thus f clearly is not injective. ■

Intuitively, this is not surprising, but the next result, which is slightly stronger, perhaps is:

Lemma 13.2.2 (Borsuk-Ulam Theorem for S^1) [1] *If $f \in C(S^1, \mathbb{R})$ then there is $x \in S^1$ such that $f(x) = f(-x)$.*

Proof. Assume $f(x) \neq f(-x)\ \forall x \in S^1$. Then $k : S^1 \to S^0 = \{\pm 1\},\ x \mapsto \frac{h(x)-h(-x)}{|h(x)-h(-x)|}$ is continuous. Since it satisfies $k(-x) = -k(x)$ it assumes both values ± 1, contradicting the connectedness of S^1. ■

Proposition 13.2.3 *Let X be compact metrizable.*

(i) $f \in C(X, S^1)$ is homotopic to the constant function 1 if and only if there exists $h \in C(X, \mathbb{R})$ such that $f(x) = e^{ih(x)}\ \forall x \in X$.

(ii) If $f, g \in C(X, S^1)$ then $f \sim g$ if and only if $f/g \sim 1$, where $(f/g)(x) = f(x)/g(x)$.

Proof. (i) $\Leftarrow$ If $f(x) = e^{ih(x)}$ then $h_t(x) = e^{ith(x)}$ is a homotopy from 1 to f.

$\Rightarrow$ If $f, g \in C(X, S^1)$, where $f(x) = e^{ih(x)}$ (we say f is of exponential form) and $|f(x) - g(x)| < 2$ for all $x \in S^1$ then $f(x) \neq -g(x)$ for all x, thus $\frac{g(x)}{f(x)} \neq -1$. Thus there is $k \in C(X, (-\pi, \pi))$ such that $g(x)/f(x) = e^{ik(x)}$, and therefore $g(x) = e^{i(h(x)+k(x))}$. Thus g is of exponential form.

Let now $h : I \times S^1 \to S^1$ be a homotopy from 1 to id_{S^1}. Since I and S^1 are compact, Exercise 7.7.45(iii) gives that $h_t(x)$ is continuous in t uniformly in x. Thus there is an $\varepsilon > 0$ such that $|t - t'| < \varepsilon$ implies $|h_t(x) - h_{t'}(x)| < 2$ for all $x \in S^1$. Let $0 = t_0 < t_1 < t_2 < \cdots < t_N = 1$ be a subdivision such that $t_{i+i} - t_i < \delta\ \forall i$. Since $h_0 = 1 = e^{it0}$ is of exponential form, successive application of the above observation implies that each $x \mapsto h_{t_i}(x)$ is of exponential form, thus h_1 is.

(ii) h_t is a homotopy from f to g if and only if h_t/g is a homotopy from f/g to 1. ■

Theorem 13.2.4 *S^1 is not contractible.*

Proof. If S^1 were contractible, by Exercise 13.1.10 there would be a homotopy from id_{S^1} to a constant function, which we clearly can take to be 1. Then Proposition 13.2.3 implies $x = e^{ih(x)}\ \forall x \in S^1$ for some $h \in C(S^1, \mathbb{R})$. Since id_{S^1} is injective, h must be injective. But this is impossible by Lemma 13.2.1 or Lemma 13.2.2. ■

Theorem 13.2.5 (Borsuk-Ulam Theorem for S^2) *If $f \in C(S^2, \mathbb{R}^2)$ then there is $x \in S^2$ such that $f(x) = f(-x)$.*

[1] Karol Borsuk (1905-1982), Stanislaw Ulam (1909-1984). Polish mathematicians. (Ulam later became US citizen and invented the hydrogen bomb.)

Proof. Assume $f(x) \neq f(-x)\ \forall x \in S^2$. Then $k : S^2 \to S^1,\ x \mapsto \frac{h(x)-h(-x)}{\|h(x)-h(-x)\|_2}$ is continuous and satisfies $k(-x) = -k(x)$. Since $S^2_+ \cong D^2$, the restriction $k \upharpoonright S^2_+$ is a homotopy between $k \upharpoonright \partial S^2_+$, where $\partial S^2_+ = \{x \in S^2 \subseteq \mathbb{R}^3 \mid x_3 = 0\} \cong S^1$, and the constant map of value $k(N), N = (0,0,1)$. With the inclusion map $\iota : S^1 \hookrightarrow S^2,\ (x_1, x_2) \to (x_1, x_2, 0)$, Proposition 13.2.3 gives that the map $l : S^1 \to S^1,\ x \mapsto k(\iota(x))$ is exponential: $l(x) = e^{ih(x)}$ for some $h \in C(S^1, \mathbb{R})$. The property $k(-x) = -k(x)$ translates to $e^{ih(-x)} = -e^{ih(x)}$, which is equivalent to $e^{i(h(x)-h(-x))} = -1\ \forall x$. Since $\{t \in \mathbb{R} \mid e^{it} = -1\} = i\pi(1 + 2\mathbb{Z})$ is discrete, this implies that $x \mapsto h(x) - h(-x)$ is constant: $h(x) - h(-x) = c\ \forall x$. But $h(x) - h(-x)$ is odd, thus $c = 0$. This gives $h(x) - h(-x) = 0\ \forall x$, but this contradicts $e^{i(h(x)-h(-x))} = -1\ \forall x$. ■

Remark 13.2.6 1. The Borsuk-Ulam theorem holds for all $n \in \mathbb{N}$. The above proof for S^2 suggests that the result for S^n can be proven by induction over n. This is indeed the case, cf. [42], but the proof uses analytic methods.

2. There is a nice proof of Borsuk-Ulam by combinatorial methods similar to those of Section 10.2, but more involved, cf. [61, Section II.5] and [165], which is a book entirely dedicated to the Borsuk-Ulam theorem and its many applications.

3. If $n > m$ then every open $U \subseteq \mathbb{R}^n$ has subspaces homeomorphic to S^m. Now the Borsuk-Ulam theorem for $C(S^m, \mathbb{R}^m)$ implies that no $f \in C(U, \mathbb{R}^m)$ is injective, providing a nice proof of the invariance of dimension for $\mathbb{R}^n$. Lemma 13.2.2 and Theorem 13.2.5 thus prove $\mathbb{R}^m \not\cong \mathbb{R}^n$ for $n > m \leq 2$. □

13.3 Path homotopy. Algebra of paths

In Section 12, we discussed paths as a more intuitive and geometric approach to studying connectivity of spaces. We defined composition and reversal of paths in order to show that the existence of a path connecting two points defines an equivalence relation. Our aim now is to use paths in a more substantial manner, mainly in order to distinguish spaces. This makes it desirable to study composition of paths algebraically. If p is a path from x to y and q a path from y to z then $q \circ p$ is a path from x to z. This clearly is reminiscent of composition of morphisms in a category, but not entirely: Composition of morphisms in a category is required to be associative, but composition of paths is not! Consider paths $x \stackrel{p_1}{\to} y \stackrel{p_2}{\to} z \stackrel{p_3}{\to} w$. (Meaning that $p_i \in C(I, X), p_1(0) = x, p_1(1) = y$, etc.) Recalling that we write composition of paths from right to left, $p = (p_3 \bullet p_2) \bullet p_1$ and $q = p_3 \bullet (p_2 \bullet p_1)$ are paths from x to w, and $p(I) = q(I)$ as subsets of X. But as maps $I \to X$, usually p and q are different: By definition of $\bullet$, we have:

t	$p_3 \bullet (p_2 \bullet p_1)$	$(p_3 \bullet p_2) \bullet p_1$
0	x	x
1/4	y	$p_1(1/2)$
1/2	z	y
3/4	$p_3(1/2)$	z
1	w	w

Recall that c_x denotes the function $I \to X$ of constant value $x \in X$. If now p is a path from x to y then we have $p \bullet c_x \neq p \neq c_y \bullet p$ and $p^{-1} \bullet p \neq c_x,\ p \bullet p^{-1} \neq c_y$ unless $x = y$ and $p = c_x$. It turns out that composition of paths has much better properties if considered up to path homotopy.

Definition 13.3.1 *Let $x, y \in X$ and $p, q : I \to X$ be paths $x \to y$. Let $h \in C(I \times I, X)$, $(s,t) \mapsto h_t(s)$ a homotopy from p to q. (I.e., $h_0 = p, h_1 = q$.) h is called a path homotopy if $h_t(0) = x$ and $h_t(1) = y$ for all $t \in I$. (I.e., each h_t is a path $x \to y$.)*

Being a sharpening of homotopy, also path homotopy is an equivalence relation.

Example 13.3.2 If $X \subseteq \mathbb{R}^n$ is convex, $x, y \in X$ and p, q are paths $x \to y$ then $h_t(s) = tq(s) + (1-t)p(s)$ is a path homotopy from p to q. □

Definition 13.3.3 *A reparametrization of a path $p : I \to X$ is a composite map $I \xrightarrow{f} I \xrightarrow{p} X$, where $f : I \to I$ is continuous and satisfies $f(0) = 0,\ f(1) = 1$. (This condition implies that the path $p \circ f$ has the same start and end points as p.)*

Lemma 13.3.4 *If $p : I \to X$ is a path, then every reparametrization $p \circ f$ is path-homotopic to p.*

Proof. Since I is convex, we have a path homotopy $h_t(s) = (1-t)s + tf(s)$ from the identity map id_I to f. Now $k_t(s) = p(h_t(s))$ clearly is a path homotopy from $p = p \circ \mathrm{id}_I$ to $p \circ f$. ■

Exercise 13.3.5 Let paths $x \xrightarrow{p_1} y \xrightarrow{p_2} z \xrightarrow{p_3} w$ in X be given. Define a map $f : I \to I$ by

$$f(s) = \begin{cases} s/2 & \text{if} \quad s \in [0, 1/2] \\ s - 1/4 & \text{if} \quad s \in [1/2, 3/4] \\ 2s - 1 & \text{if} \quad s \in [3/4, 1] \end{cases}$$

Prove that $((p_3 \bullet p_2) \bullet p_1)(s) = (p_3 \bullet (p_2 \bullet p_1))(f(s)) \ \ \forall s \in [0,1]$.

Corollary 13.3.6 *Given paths $x \xrightarrow{p_1} y \xrightarrow{p_2} z \xrightarrow{p_3} w$, there is a path homotopy $(p_3 \bullet p_2) \bullet p_1 \to p_3 \bullet (p_2 \bullet p_1)$.*

Proof. By Exercise 13.3.5, the paths $(p_3 \bullet p_2) \bullet p_1$ and $p_3 \bullet (p_2 \bullet p_1)$ are reparametrizations of each other and therefore path-homotopic by Lemma 13.3.4. ■

Lemma 13.3.7 *If p is a path $x \to y$ then p is path-homotopic to $c_y \bullet p$ and to $p \bullet c_x$.*

Proof. Define $f_1, f_2 : I \to I$ by $f_1(s) = \min(1, 2s)$ and $f_2(s) = \max(0, 2s-1)$. Now $p \circ f_1 = c_y \bullet p$ and $p \circ f_2 = p \bullet c_x$, thus $c_y \bullet p$ and $p \bullet c_x$ are path-homotopic to p by Lemma 13.3.4. ■

Lemma 13.3.8 *Let $p : x \to y$ be a path. Then there is a path homotopy $c_x \to p^{-1} \bullet p$.*

Proof. Define $f : I \to I$ by $f(s) = 2s$ for $s \in [0, 1/2]$ and $f(s) = 2 - 2s$ for $s \in [1/2, 1]$. Clearly, $p \circ f = p^{-1} \bullet p$. By Example 13.3.2, f is path-homotopic to the map $c_0 : I \to I, s \mapsto 0$. Thus $p^{-1} \bullet p = p \circ f$ is path-homotopic to $p \circ c_0 = c_x$. ■

Lemma 13.3.9 *(i) If p, p' are paths from x to y and $h_t(s)$ is a path homotopy from p to p' then $h_t(1-s)$ is a path homotopy from p^{-1} go p'^{-1}.*

(ii) If $p, p' : x \to y$ and $q, q' : y \to z$ are paths and $u : p \to p'$ and $v : q \to q'$ are path homotopies, then $q \bullet p$ and $q' \bullet p'$ are path-homotopic.

Proof. (i) Obvious. (ii) Define

$$w_t(s) = \begin{cases} u_t(2s) & \text{for} \quad s \in [0, 1/2] \\ v_t(2s-1) & \text{for} \quad s \in [1/2, 1] \end{cases}$$

Then w is a path homotopy from $q \bullet p$ to $q' \bullet p'$. ■

For a path $p : x \to y$, we write $[p]$ for the set of paths $x \to y$ that are path-homotopic to p. Then Lemma 13.3.9 amounts to $[p] = [p'] \Rightarrow [p^{-1}] = [p'^{-1}]$ and $[p] = [p'], [q] = [q'] \Rightarrow [q \bullet p] = [q' \bullet p']$. This allows us to define reversal and composition of path homotopy classes by $[p]^{-1} = [p^{-1}]$ and $[q] \bullet [p] := [q \bullet p]$. I.e. the reversal or composition of path homotopy classes is defined by choosing representer(s), reversing or composing them and passing to the path homotopy class(es).

Corollary 13.3.10 *Composition of path-homotopy classes of paths has the following properties:*

(i) Strict Inverses: Let $[p]$ be a path homotopy class of paths $x \to y$. Then $[p]^{-1} \bullet [p] = [c_x]$ and $[p] \bullet [p]^{-1} = [c_y]$.

(ii) Associativity: If $[p], [q], [r]$ are path homotopy classes of paths $x \to y \to z \to t$, then $([r] \bullet [q]) \bullet [p] = [r] \bullet ([q] \bullet [p])$.

Proof. Immediate consequence of Lemma 13.3.9. ■

13.4 The fundamental groupoid functor Π_1 : Top $\to$ Grpd

Since passing to path homotopy-classes fixes the problem with associativity of composition of paths, we have:

Proposition 13.4.1 *Let X be a topological space. Then there is a category $\Pi_1(X)$, called the fundamental groupoid of X, defined by*

- $\mathrm{Obj}(\Pi_1(X)) = X$. *(I.e., the points of X).*
- *For $x, y \in X = \mathrm{Obj}(\Pi_1(X))$, we define $\mathrm{Hom}_{\Pi_1(X)}(x, y)$ as the set of path homotopy classes of paths $x \to y$.*
- *The identity morphism of x is the (path homotopy class of the) constant path $[c_x]$.*
- *Composition of morphisms is given by composition of path homotopy classes as above. Thus if $[p] \in \mathrm{Hom}_{\Pi_1(X)}(x, y), [q] \in \mathrm{Hom}_{\Pi_1(X)}(y, z)$ then $[q] \bullet [p] \in \mathrm{Hom}_{\Pi_1(X)}(x, z)$ is defined by $[q] \bullet [p] = [q \bullet p]$.*

Every morphism $[p] \in \mathrm{Hom}_{\Pi_1(X)}(x, y)$ has an inverse $[p]^{-1} \in \mathrm{Hom}_{\Pi_1(X)}(y, x)$.

Proof. Corollary 13.3.10 gives the associativity of composition of morphisms and the fact that $[c_x]$ behaves under composition as a unit should. Given $[f] \in \mathrm{Hom}_{\Pi_1(X)}(x, y)$, $[f]^{-1}$ is an inverse morphism of $[f]$. ■

The fact that every morphism in $\Pi_1(X)$ has an inverse means that $\Pi_1(X)$ actually is a groupoid:

Definition 13.4.2 *A groupoid is a category that is small, in the sense that the objects form a set (and not just a class), and in which all morphisms have inverses. The category of groupoids, which is a full subcategory of* Cat, *is denoted* Grpd.

Remark 13.4.3 1. Let $\mathcal{C}$ be a category and $X \in \mathrm{Obj}\,\mathcal{C}$. If $s, t \in \mathrm{End}_{\mathcal{C}}(X, X) := \mathrm{Hom}_{\mathcal{C}}(X, X)$ (the set of endomorphisms of X) then s and t can be composed, the composition is associative and has id_X as a unit. Thus $\mathrm{End}_{\mathcal{C}}(X)$ is a monoid for each $X \in \mathrm{Obj}\,\mathcal{C}$. If $\mathcal{C}$ is a groupoid, then in addition every $s \in \mathrm{End}_{\mathcal{C}}(X)$ has an inverse, thus $\mathrm{End}_{\mathcal{C}}(X)$ is a group for each $X \in \mathrm{Obj}\,\mathcal{C}$.

2. If a category $\mathcal{C}$ has only one object, say $\mathrm{Obj}\,\mathcal{C} = \{X\}$, then all its information is contained in the monoid $\mathrm{End}_{\mathcal{C}}(X)$. In this sense, the notion of 'category' is a generalization of the notion of 'monoid', and similarly groupoids are a generalization of groups.

3. For someone reasonably familiar with categories, this really is the most transparent definition of groupoids. However, we briefly describe two alternative definitions, leaving it to the reader to convince him/herself that the three definitions are equivalent.

Definition 2: A groupoid consists of two sets G, B, where B is called the base space, two maps $s, t : G \to B$, the source and target maps, a unit map from $u : B \to G$, an inverse map $G \to G, g \mapsto g^{-1}$ and a product map

$$m : G \times_B G := \{(g, h) \in G \times G \mid s(g) = t(h)\} \to G.$$

These data have to satisfy some axioms that are easy to figure out via the correspondence to the categorical Definition 13.4.2. If $\mathcal{G}$ is a groupoid in the sense of the latter, we obtain the data of Definition 2 by putting $B = \mathrm{Obj}(\mathcal{G})$ and $G = \bigoplus_{X,Y \in \mathrm{Obj}(\mathcal{G})} \mathrm{Hom}_{\mathcal{G}}(X, Y)$. In view of the definition of the disjoint union $\oplus$, cf. Section A.2, we have a map $G \to B \times B$, which associates to every $g \in G$ its source and target in B. The unit map is given by $X \mapsto \mathrm{id}_X \in G$, etc.

Definition 3: A groupoid is a set G, together with an inverse map $G \to G, g \mapsto g^{-1}$ and a partially defined multiplication map $m : G_2 \to G$, where G_2 is a subset of $G \times G$. Among the axioms, one has: $(g, h) \in G_2$ and $(m(g, h), k) \in G_2$ hold if and only if $(h, k) \in G_2$ and $(g, m(h, k)) \in G_2$. In this case, $m(m(g, h), k) = m(g, m(h, k))$. It is probably clear that G_2 corresponds to the $G \times_B G$ appearing in Definition 2. □

Applying Remark 13.4.3.1 to $\mathcal{C} = \Pi_1(X)$, the fundamental groupoid of a space X, leads to:

Definition 13.4.4 *Let X be a topological space and $x_0 \in X$. Then the fundamental group*[2] $\pi_1(X, x_0)$ *for basepoint x_0 is the group* $\mathrm{Hom}_{\Pi_1(X)}(x_0, x_0)$.

Lemma 13.4.5 *Every path $p : x_0 \to x_1$ induces an isomorphism $\pi_1(X, x_0) \to \pi_1(X, x_1)$ via $\alpha_{[p]} : [q] \mapsto [p] \bullet [q] \bullet [p]^{-1}$. This map depends only on the path-homotopy class $[p]$.*

Proof. This map is a homomorphism:

$$\alpha_{[p]}([q]) \bullet \alpha_{[p]}([r]) = ([p] \bullet [q] \bullet [p]^{-1}) \bullet ([p] \bullet [r] \bullet [p]^{-1}) = [p] \bullet [q] \bullet [r] \bullet [p]^{-1} = \alpha_{[p]}([q] \bullet [r]).$$

That $\alpha_{[p]}$ is a bijection follows by consideration of the map $\pi_1(X, x_1) \to \pi_1(X, x_0)$, $[q] \mapsto [p]^{-1} \bullet [q] \bullet [p]$, which clearly is an inverse of $\alpha_{[p]}$. ■

Remark 13.4.6 1. Lemma 13.4.5 just is a special case of the general fact that in a groupoid every $s \in \mathrm{Hom}(X, Y)$ gives rise to a group isomorphism $\mathrm{End}(X) \to \mathrm{End}(Y)$ via $t \mapsto s \circ t \circ s^{-1}$.

2. For path-connected X, Lemma 13.4.5 gives us isomorphisms $\pi_1(X, x_0) \stackrel{\cong}{\to} \pi_1(X, x_1)$ for all x_0, x_1. This allows to omit the base point x_0 from the notation and to simply write $\pi_1(X)$. But strictly speaking, this is not a concrete group, but only an isomorphism class of groups!

[2] The fundamental group was defined by Poincaré in 1895. Groupoids were introduced in 1927 by Heinrich Brandt (1886-1954), and the fundamental groupoid was introduced by Ronald Brown (1935-) in 1967.

3. A path f in X such that $f(0) = f(1) = x_0$ is called a loop based at x_0. It is clear that a loop in X based at x_0 is the same as a map $S^1 \to X$ mapping $1 \in S^1$ to x_0. Now it is clear that $\pi_1(X, x_0)$ can be defined directly as the set of path-homotopy classes of loops based at x_0, the group structure given by composition and reversion of homotopy classes.

But other than to avoid groupoids, there is no good reason to do so. On the one hand, $\pi_1(X, x_0)$ depends on x_0, and if x_0, x_1 are contained in different path-components of X, the fundamental groups $\pi_1(X, x_0), \pi_1(X, x_1)$ can be totally unrelated. Thus in order not to waste information provided by π_1, one needs to consider the whole family $\{\pi_1(X, x_0) \mid x_0 \in X\}$ of groups. The latter information is, of course, contained in the fundamental groupoid $\Pi_1(X)$. But the latter also contains the information about the path-components of X. Namely, for $x, y \in X$ we have $x \sim_p y$ if and only if $\mathrm{Hom}_{\Pi_1(X)}(x, y) \neq \emptyset$. Since all morphisms in $\Pi_1(X)$ are isomorphisms, there is an obvious (and 'natural' in the technical categorical sense) bijection between the set $\pi_0(X) = X/\sim_p$ of path-components of X (Definition 12.1.5) and the set $\mathrm{Obj}(\Pi_1(X))/\cong$ of isomorphism classes of objects in the groupoid $\Pi_1(X)$. □

Definition 13.4.7 *A topological space is simply connected if it is path-connected and $\pi_1(X, x_0) = 0$ for some (thus every) $x_0 \in X$.*

Proposition 13.4.8 *S^1 is path-connected, but not simply connected.*

Proof. We already know the path-connectedness. Thus the fundamental groups $\pi(S^1, x_0)$ are all isomorphic and we may consider $x_0 = 1$. Since the identity map id_{S^1} defines a loop in S^1 based at 1, simple connectedness would mean that id_{S^1} is path homotopic, thus homotopic, to the constant loop c_1. Exercise 13.1.10(i) then would imply that S^1 is contractible, contradicting Theorem 13.1.11 (or Theorem 13.2.4). ■

Remark 13.4.9 For many applications, like the proof of the Jordan curve theorem, it is sufficient to know that $\pi_1(S^1)$ is nontrivial, which we denote $\pi_1(S^1) \neq 0$. For others, however one needs to know it explicitly, and we will compute $\pi_1(S^1)$ in Section 13.7. □

Let $f : X \to Y$ be a continuous map. By definition, it associates a point $f(x) \in Y$ to every $x \in X$. But as noted in Section 12, we have more: If p is a path in X from x to x' then $f \circ p$ is a path in Y from $f(x)$ to $f(x')$. If h_t is a path-homotopy from p to a second path p' from x to y, then $f \circ h_t$ is a path-homotopy between the paths $f \circ p$ and $f \circ p'$ in Y. Thus $[p] = [p'] \;\Rightarrow\; [f \circ p] = [f \circ p']$. This allows us to define a map

$$\Pi_1(f) : \mathrm{Hom}_{\Pi_1(X)}(x, y) \to \mathrm{Hom}_{\Pi_1(Y)}(f(x), f(y)), \quad [p] \mapsto [f \circ p]. \tag{13.1}$$

Proposition 13.4.10 *The maps $\mathrm{Obj}(\Pi_1(X)) = X \xrightarrow{f} Y = \mathrm{Obj}(\Pi_1(Y))$ and (13.1) define a functor $\Pi_1(f) : \Pi_1(X) \to \Pi_1(Y)$ between the fundamental groupoids $\Pi_1(X)$ and $\Pi_1(Y)$.*

Proof. It only remains to show that $\Pi_1(f)$ maps identity morphisms to identity morphisms and commutes with composition of morphisms. (Cf. (i) and (ii) in Definition A.5.12.)

If c_x is the constant path at $x \in X$ then $f \circ c_x$ is the constant path at $f(x)$. In view of $\mathrm{id}_x = [c_x]$, we have $\Pi_1(f)(\mathrm{id}_x) = \Pi_1(f)([c_x]) = [c_{f(x)}] = \mathrm{id}_{f(x)}$. Furthermore,

$$\Pi_1(f)([p \bullet p']) = [f \circ (p \bullet p')] = [(f \circ p) \bullet (f \circ p')] = [(f \circ p)] \bullet [(f \circ p')] = \Pi_1([p]) \bullet \Pi_1([p']),$$

thus $\Pi(f)$ commutes with composition of morphisms. ■

Theorem 13.4.11 Π_1 *defines a functor* $\mathcal{TOP} \to \mathcal{GRPD}$ *from topological spaces to groupoids.*

Proof. Let $f : X \to Y,\ g : Y \to Z$ be continuous maps. By Proposition 13.4.10, we have a functor $\Pi_1(f) : \Pi_1(X) \to \Pi_1(Y)$. It remains to show that the map $C(X,Y) \to \mathrm{Fun}(\Pi_1(X), \Pi_1(Y)),\ f \mapsto \Pi_1(f)$ is functorial, cf. Definition A.5.12. If $X = Y$ and $f = \mathrm{id}_X$ it is clear that $\Pi_1(f)$ acts identically on the objects and morphisms of $\Pi_1(X)$, thus $\Pi_1(\mathrm{id}_X) = \mathrm{id}_{\Pi_1(X)}$, the identity functor of $\Pi_1(X)$.

Now consider continuous maps $X \stackrel{f}{\to} Y \stackrel{g}{\to} Z$. If $x \in X = \mathrm{Obj}(\Pi_1(X))$ then $\Pi_1(g)(\Pi_1(f)(x)) = g(f(x)) = \Pi_1(g \circ f)(x)$, thus $\Pi_1(g \circ f) = \Pi_1(g) \circ \Pi_1(f)$ holds on objects. On the other hand, let $[p] \in \mathrm{Hom}_{\Pi_1(X)}(x, y)$ be a homotopy class of paths in X. By definition of Π_1, we have $\Pi_1(f)([p]) = [f \circ p]$ and $\Pi_1(g)(\Pi_1(f)([p])) = [g \circ f \circ p] = \Pi_1(g \circ f)([p])$. Thus $\Pi_1(g \circ f) = \Pi_1(g) \circ \Pi_1(f)$ also holds on morphisms, and Π_1 indeed is a functor from $\mathcal{TOP} \to \mathcal{GRPD}$. ■

Corollary 13.4.12 *If $f : X \to Y$ is a homeomorphism then $\Pi_1(f) : \Pi_1(X) \to \Pi_1(Y)$ is an isomorphism of groupoids. (Cf. Definition A.5.18.)*

Proof. Since f is a homeomorphism, there is a continuous map $g : Y \to X$ such that $g \circ f = \mathrm{id}_X$ and $f \circ g = \mathrm{id}_Y$. Now functoriality of Π_1 implies $\Pi_1(g) \circ \Pi_1(f) = \mathrm{id}_{\Pi_1(X)}$ and $\Pi_1(f) \circ \Pi_1(g) = \mathrm{id}_{\Pi_1(Y)}$. Thus $\Pi_1(f)$ is an isomorphism of categories (and thus of groupoids). ■

It is clear that this result can be used to prove that certain spaces are not homeomorphic to each other. Examples will be given as soon as we have computed some fundamental group(oid)s.

Focusing our attention on the fundamental groups, we have the following consequences:

Corollary 13.4.13 *A continuous map $f : X \to Y$ gives rise to a homomorphism $\pi_1(f) : \pi_1(X, x_0) \to \pi_1(Y, f(x_0))$ of groups for every $x_0 \in X$.*

If f is a homeomorphism then the above homomorphisms $\pi_1(f)$ are isomorphisms.

Corollary 13.4.13 looks somewhat like the statement that π_1 is a functor from topological spaces to groups, but there are two (related) problems: (i) The natural domain of a functor π_1 is not given by topological spaces, but by spaces together with a chosen point. (ii) The fundamental group in which $\pi_1(f)([p])$, where $[p] \in \pi_1(X, x_0)$, lives depends on $f(x_0)$. This motivates:

Definition 13.4.14 *The category $\mathcal{TOP}_*$ of <u>pointed topological spaces</u> is the category whose objects are pairs (X, x_0), where X is a topological space and $x_0 \in X$. The morphisms are given by*

$$\mathrm{Hom}_{\mathcal{TOP}_*}((X, x_0), (Y, y_0)) = \{f \in \mathrm{Hom}_{\mathcal{TOP}}(X, Y) = C(X, Y) \mid f(x_0) = y_0\}.$$

Corollary 13.4.15 *We have a functor $\pi_1 : \mathcal{TOP}_* \to \mathrm{Grp}$ given by $(X, x_0) \mapsto \pi_1(X, x_0)$ and* $\mathrm{Hom}_{\mathcal{TOP}_*}((X, x_0), (Y, y_0)) \to \mathrm{Hom}_{\mathrm{Grp}}(\pi_1(X, x_0), \pi_1(Y, y_0)),\ [p] \mapsto [f \circ p]$.

Exercise 13.4.16 Let $y_0 \in Y \subseteq X$ and let $\iota : Y \to X$ be the inclusion map.

(i) Give an example of this situation such that $\pi_1(\iota) : \pi_1(Y, y_0) \to \pi_1(X, y_0)$ is not injective.

(ii) Prove that $\pi_1(\iota) : \pi_1(Y, y_0) \to \pi_1(X, y_0)$ is injective if there exists a retraction $X \to Y$.

Exercise 13.4.17 (Products) (i) For $(X, x_0), (Y, y_0) \in \mathrm{Top}_*$, prove the isomorphism $\pi_1(X \times Y, (x_0, y_0)) \cong \pi_1(X, x_0) \times \pi_1(Y, y_0)$.

(ii) For $X, Y \in \mathcal{T\emptyset P}$, prove $\Pi_1(X \times Y) \cong \Pi_1(X) \times \Pi_1(Y)$. (This includes defining the product $\mathcal{C} \times \mathcal{D}$ of two categories $\mathcal{C}, \mathcal{D}$, but it should be clear how to do this.)

Exercise 13.4.18 $\pi_1(S^n) = 0$ for $n \geq 2$.

(i) Consider $S^n \subseteq \mathbb{R}^{n+1}$ as a metric space via the Euclidean metric of $\mathbb{R}^{n+1}$ and cover it by open balls of diameter 1. If $p : [0,1] \to S^n$ is a path (with $p(0) = p(1)$), show that there is a partition $0 = t_0 < t_1 < \cdots < t_n = 1$ such that $p([t_i, t_{i+1}])$ is contained in one of the balls, for each i.

(ii) Show that $p \upharpoonright [t_i, t_{i+1}]$ is homotopic to a path $p(t_i) \to p(t_{i+1})$ along a great circle.

(iii) Conclude that p is homotopic to a path p' consisting of finitely many arcs on great circles and conclude that $p'([0,1]) \neq S^n$.

(iv) Show that p', thus p is homotopic to a constant loop.

The spheres S^2 and S^3 are actually characterized by their simple connectedness:

Theorem 13.4.19 *Let $d \in \{2, 3\}$ and let X be compact, Hausdorff, second countable, simply connected and such that every $x \in X$ has an open neighborhood homeomorphic to $\mathbb{R}^d$. Then $X \cong S^d$.*

Remark 13.4.20 For $d = 2$, this is part of the classification of surfaces, which has been known since the 1860s, even though rigorous proofs appeared only in the 20th century. Cf. e.g. [50, 77, 180]. For $d = 3$, this was conjectured by Poincaré in 1904 (correcting a false conjecture from 1900) and proven by Perelman[3] around 2002-3. The proof uses Riemannian geometry (which belongs to differential geometry, thus differential topology) and nonlinear partial differential equations! A reasonably self-contained exposition of the proof occupies an entire book like [176] (plus the fact that every topological 3-manifold admits a smooth structure, proven in the 1950s by Moise and Bing). □

Exercise 13.4.21 ("Eckmann-Hilton argument") Consider $(S, e, \bullet, \times)$, where S is a set, $e \in S$ and $\bullet, \times$ are binary operations $S \times S \to S$ that satisfy the distributivity law

$$(g \bullet h) \times (g' \bullet h') = (g \times g') \bullet (h \times h') \qquad \forall g, g', h, h' \tag{13.2}$$

and have e as two-sided unit:

$$g \times e = e \times g = g \bullet e = e \bullet g = g \qquad g. \tag{13.3}$$

Prove that $g \times h = h \times g = g \bullet h = h \bullet g$ for all $g, h \in S$. (I.e. the two operations coincide and are abelian.)

Exercise 13.4.22 ($\pi_1(X)$ is abelian for topological monoids) (i) Let $(M, \cdot, \mathbf{1})$ be a topological monoid. If g, h are loops in M based at $\mathbf{1}$, define $g \times h$ by $(g \times h)(t) = g(t) \cdot h(t)$. Show that $g \times h$ is a loop based at $\mathbf{1}$. Show that if there are path homotopies $g \cong g', h \cong h'$ then there is a path homotopy $g \times h \cong g' \times h'$. Conclude that $\times$ descends to an operation on $\pi_1(M, \mathbf{1})$ having $[c_{\mathbf{1}}]$ as unit.

(ii) Use (i) and Exercise 13.4.21 to prove that $\pi_1(M, \mathbf{1})$ is abelian.

(iii) If G is a topological group, prove that $\pi_1(G, g)$ is abelian for every $g \in G$.

[3]Grigori Perelman (1966-) Russian mathematician, working mainly in Riemannian geometry. Fields medal (declined).

Exercise 13.4.23 (Higher homotopy groups) Let X be a topological space and $x_0 \in X$. Let $I = [0,1]$. For $n \geq 2$, let

$$P_n = \{f \in C(I^n, X) \mid f \restriction \partial I^n = x_0\}.$$

(i) Let $f, g \in P_n$ and $k \in \{1,,\dots,n\}$. Define $f \circ_k g : I^n \to X$ by

$$f \circ_k g(x) = \left\{ \begin{array}{cll} f(x_1,\dots,2x_k,\dots,x_n) & \text{if} & x_k \in [0,1/2] \\ g(x_1,\dots,2x_k-1,\dots,x_n) & \text{if} & x_k \in [1/2,1] \end{array} \right.$$

Prove that $f \circ_k g \in P_n$. (A picture helps!)

(ii) Define an equivalence relation $\sim$ on P_n by $f \sim g \Leftrightarrow$ there is a homotopy h_t from f to g such that $h_t \restriction \partial I^n = x_0$ $\forall t \in [0,1]$.

For $i \in \{1,\dots,n\}$ and $f, f', g, g' \in P_n$ with $f \sim f'$ and $g \sim g'$, prove that $f \circ_i g \sim f' \circ_i g'$.

Use this to define, for each $i \in \{1,\dots,n\}$, a unique binary operation $\bullet_i$ on $\pi_n = P_n/\sim$ such that $[f \circ_i g] = [f] \bullet_i [g]$.

(iii) Show that the class $[x_0]$ of the constant function is a unit for all operations $\bullet_i$.

(iv) Use Exercise 13.4.21 to prove that $f \bullet_i g = f \bullet_j g$ $\forall i, j \in \{1,\dots,n\}$, so that we can define $f \bullet g := f \bullet_i g$ for i arbitrary, and that $f \bullet g = g \bullet f$.

(v) Show that $(\pi_n, \bullet, [x_0])$ is an abelian group. We denote it $\pi_n(X, x_0)$.

Exercise 13.4.24 (π_n as functors) Let $n \geq 2$.

(i) Extend the definition of $\pi_n(X, x_0)$ to functors $\pi_n : \mathrm{Top}_* \to \mathrm{Ab}$. (Recall that Ab is the category of abelian groups.)

(ii) Prove that if $f : X \to Y$ is a homotopy equivalence then $(\pi_n)_*(f) : \pi_n(X, x_0) \to \pi_n(Y, f(x_0))$ is an isomorphism of abelian groups.

(iii) Prove that $\pi_n(S^n) \neq 0$ $\forall n \geq 2$. Hint: Use Theorem 13.1.11.

Remark 13.4.25 Since S^0 is the discrete space $\{\pm 1\}$, it is clear that $\pi_0(S^0)$ is a two-element set. We already know that $\pi_1(S^1) \neq 0$, and in the next subsection we will prove that $\pi_1(S^1) \cong \mathbb{Z}$. Similarly, one can prove $\pi_n(S^n) \cong \mathbb{Z}$ for all $n \geq 2$, but this is beyond the scope of this course. □

13.5 Homotopy invariance of π_1 and Π_1

In Exercise 13.1.17 we have seen that the functor π_0 is homotopy invariant, in the sense that it sends homotopic maps $f, g : X \to Y$ to identical maps $\pi_0(f) = \pi_0(g) : \pi_0(X) \to \pi_0(Y)$. In this section we will prove analogous (but more complicated) results for the functors π_1 and Π_1.

13.5.1 Homotopy invariance of π_1

According to Corollary 13.4.12 the functor $\Pi_1(f) : \Pi_1(X) \to \Pi_1(Y)$ is an isomorphism of categories when $f : X \to Y$ is a homeomorphism. As we have seen, the notion of homeomorphism is vastly generalized by that of homotopy equivalence. This raises the question what can be said about $\Pi_1(f)$ if $f : X \to Y$ is a homotopy equivalence and, more generally, how $\Pi_1(f)$ and $\Pi_1(g)$ are related if $f, g : X \to Y$ are homotopic. This question has a very simple and satisfactory answer, cf. Proposition 13.5.5 below.

Since the latter is somewhat abstract, we first consider the question at the level of fundamental groups. Let $x_0 \in X$. The existence of a homotopy h between $f, g : X \to Y$ does not imply $f(x_0) = g(x_0)$. Thus the homomorphisms $\pi_1(f) : \pi_1(X, x_0) \to \pi_1(Y, f(x_0))$ and $\pi_1(g) : \pi_1(X, x_0) \to \pi_1(Y, g(x_0))$ typically have different groups as images, so that we cannot hope to prove homotopy invariance in the form $\pi_1(f) = \pi_1(g)$. We begin with the following

Lemma 13.5.1 *Let h_t be a homotopy between the maps $f, g : X \to Y$ and let p be a path in X from x_0 to x_1. Let $u(t) = h_t(x_0)$, $v(t) = h_t(x_1)$. Then the composite paths*

$$f(x_0) \xrightarrow{u} g(x_0) \xrightarrow{g\circ p} g(x_1) \quad \text{and} \quad f(x_0) \xrightarrow{f\circ p} f(x_1) \xrightarrow{v} g(x_1)$$

are path-homotopic.

Proof. Consider the diagram

$$\begin{tikzcd} f(x_0) \arrow[r, phantom, "="] & h_0(p(0)) \arrow[r, "{[t\mapsto h_t(p(0))]}"] \arrow[d, "{[s\mapsto h_0(p(s))]}"'] & h_1(p(0)) \arrow[r, phantom, "="] \arrow[d, "{[s\mapsto h_1(p(s))]}"] & g(x_0) \\ f(x_1) \arrow[r, phantom, "="] & h_0(p(1)) \arrow[r, "{[t\mapsto h_t(p(1))]}"'] & h_1(p(1)) \arrow[r, phantom, "="] & g(x_1) \end{tikzcd} \qquad (13.4)$$

where the arrows represent paths in Y. Let q_1, q_2 be the paths in $I \times I$ that go from $(0, 0)$ to $(1, 1)$ at constant speed, with q_1 passing through $(1, 0)$ and q_2 through $(0, 1)$. Since $I \times I$ is convex, Example 13.3.2 implies the path-homotopy $q_1 \sim q_2$. Defining $k : I \times I \to Y, (s, t) \mapsto h_t(p(s))$ we thus have $k \circ q_1 \sim k \circ q_2$. Now, the path $k \circ q_1$ in Y coincides with the composite path in (13.4) along the left and bottom edges, thus $k \circ q_1 = v \bullet (f \circ p)$, and similarly $k \circ q_2 = (g \circ p) \bullet u$. We thus have $[v] \bullet [f \circ p] = [g \circ p] \bullet [u]$. ■

Proposition 13.5.2 *Let h be a homotopy between the maps $f, g : X \to Y$ and $x_0 \in X$. Then the diagram*

$$\begin{tikzcd} \pi_1(X, x_0) \arrow[r, "\pi_1(f)"] \arrow[dr, "\pi_1(g)"'] & \pi_1(Y, f(x_0)) \arrow[d, "\alpha_{[u]}"] \\ & \pi_1(Y, g(x_0)) \end{tikzcd} \qquad (13.5)$$

commutes, where $\alpha_{[u]}$ is the change-of-basepoint isomorphism, cf. Lemma 13.4.5, arising from the path $u : t \mapsto h_t(x_0)$ from $f(x_0)$ to $g(x_0)$.

Proof. When dealing with fundamental groups, we consider loops, i.e. $x_0 = x_1$. Thus the functions u, v in the Lemma coincide, and we have $[u] \bullet [f \circ p] = [g \circ p] \bullet [u]$. This is equivalent to $\pi_1(g)([p]) = [u] \bullet \pi_1(f)([p]) \bullet [u]^{-1}$. In view of Lemma 13.4.5, the right hand side coincides with $\alpha_{[u]}(\pi_1(f)([p]))$. ■

Corollary 13.5.3 *(i) If $f : X \to Y$ is a homotopy equivalence and $x_0 \in X$ then $\pi_1(f) : \pi_1(X, x_0) \to \pi_1(Y, f(x_0))$ is an isomorphism.*

(ii) If X is contractible then it is simply connected.

Proof. (i) Pick a $g : Y \to X$ such that $g \circ f \sim \mathrm{id}_X$ and $f \circ g \sim \mathrm{id}_Y$. Then $\pi_1(g) \circ \pi_1(f) = \pi_1(g \circ f) = \alpha \circ \pi_1(\mathrm{id}_X) = \alpha$, where α is the isomorphism from Lemma 13.5.1. Thus $\pi_1(f) : \pi_1(X, x_0) \to \pi_1(Y, f(x_0))$ is injective (and $\pi_1(g)$ is surjective). Similarly, we have

$$\pi_1(f) \circ \pi_1(g) = \pi_1(f \circ g) = \beta \circ \pi_1(\mathrm{id}_Y) = \beta \quad \text{as maps} \quad \pi_1(Y, f(x_0)) \to \pi_1(X, fgf(x_0)),$$

where β is an isomorphism. This implies that $\pi_1(f) : \pi_1(X, gf(x_0)) \to \pi_1(Y, fgf(x_0))$ is surjective. Now, the homotopy $gf \sim \mathrm{id}_X$ gives rise to a path $p : gf(x_0) \to x_0$. Since the diagram

$$\begin{tikzcd}
\pi_1(X, gf(x_0)) \arrow[r, "\pi_1(f)"] \arrow[d, "\alpha_{[p]}"'] & \pi_1(Y, fgf(x_0)) \arrow[d, "\alpha_{[f \circ p]}"] \\
\pi_1(X, x_0) \arrow[r, "\pi_1(f)"'] & \pi_1(Y, f(x_0))
\end{tikzcd}$$

commutes, the vertical arrows are isomorphisms and the top morphism is a surjection, also the bottom morphism $\pi_1(f) : \pi_1(X, x_0) \to \pi_1(Y, f(x_0))$ is a surjection, thus an isomorphism.

(ii) As a contractible space, X is path-connected, cf. Exercise 13.1.10(ii). Now (i) implies $\pi_1(X, x_0) = 0$ for any $x_0 \in X$. ■

Remark 13.5.4 The converse is in general not true. In Exercise 13.4.18 it is shown that $\pi_1(S^n) = 0$ for all $n \geq 2$, thus the spheres $S^2, S^3, \ldots$ are simply connected. But by Theorem 13.1.11 the spheres are not contractible. (S^0 is not even path-connected, and S^1 is path-connected but not simply connected since $\pi_1(S^1) \neq 0$ as we have seen in Section 13.2. □

It remains to explicitly compute $\pi_1(S^1)$. This can be done by elementary but somewhat ad-hoc methods, cf. e.g. [237, Section 34]. We will instead give two conceptual ones which proceed via methods that have a much wider applicability, namely via a Seifert-van Kampen theorem for the fundamental groupoid in Section 13.6 and the more conventional one via covering spaces in Section [?].

13.5.2 Homotopy invariance of Π_1

Now we turn to the – conceptually more satisfactory – Π_1-version of the above results:

Proposition 13.5.5 *Every homotopy h between maps $f, g : X \to Y$ gives rise to a natural isomorphism $\alpha : \Pi_1(f) \to \Pi_1(g)$ between the functors $\Pi_1(f), \Pi_1(g) : \Pi_1(X) \to \Pi_1(Y)$.*

Proof. In order to improve readability of the proof and to make contact with the definitions in Section A.5, we write $\mathcal{C} = \Pi_1(X), \mathcal{D} = \Pi_1(Y), F = \Pi_1(f), G = \Pi_1(g)$. Now F and G are functors $\mathcal{C} \to \mathcal{D}$, and we want a natural isomorphism $\alpha : F \to G$, i.e. a family of isomorphisms $\{\alpha_X : F(X) \to G(X)\}_{X \in \mathcal{C}}$ such that (A.3) commutes. The objects of $\Pi_1(X)$ and of $\Pi_1(Y)$ are the points of X, Y, respectively, and a morphism $s \in \mathrm{Hom}_{\mathcal{C}}(x, x')$ is a path-homotopy class $s = [p]$ of a path in X from x to x'. By definition, $F(s) = [f \circ p]$, $G(s) = [g \circ f]$ are homotopy classes of paths $f(x) \to f(x')$ and $g(x) \to g(x')$ in Y. The diagram (A.3) thus specializes to:

$$\begin{tikzcd} f(x) \arrow[r, "\alpha_x"] \arrow[d, "{[f\circ p]}"'] & g(x) \arrow[d, "{[g\circ p]}"] \\ f(x') \arrow[r, "\alpha_{x'}"'] & g(x') \end{tikzcd} \tag{13.6}$$

Thus in order to define a natural isomorphism, we must choose for each $x \in X$ a (homotopy class of a) path α_x from $f(x)$ to $g(x)$ such that the diagram commutes. An obvious choice is to define $\alpha_x = [t \mapsto h_t(x)]$. In view of $h_0 = f, h_1 = g$, α_x is a path from $f(x)$ to $g(x)$, for every $x \in X$. It remains to show that this choice of α makes (A.3) commute for every (homotopy class of a) path $[p]$ in X from x to x'. Inserting our choice of α_x in (13.6), replacing f and g by h_0 and h_1, as well as x and x' by $p(0)$ and $p(1)$, respectively, the diagram becomes (13.4). By Lemma 13.5.1, the diagram commutes up to homotopy, which gives the commutativity of (13.6). ■

Now we can prove the following important generalization of Corollary 13.4.12:

Corollary 13.5.6 *If $f : X \to Y$ is a homotopy equivalence then the functor $\Pi_1(f) : \Pi_1(X) \to \Pi_1(Y)$ is an equivalence of groupoids. (Cf. Definition A.5.21.)*

Proof. By definition, there are a map $g : Y \to X$ and homotopies $h : g \circ f \to \mathrm{id}_X,\ h' : f \circ g \to \mathrm{id}_Y$. By Proposition 13.5.5, there are natural isomorphisms $\alpha_h : \Pi_1(g) \circ \Pi_1(f) = \Pi_1(g \circ f) \to \Pi_1(\mathrm{id}_X) = \mathrm{id}_{\Pi_1(X)}$ and $\alpha_{h'} : \Pi_1(f) \circ \Pi_1(g) = \Pi_1(g \circ g) \to \Pi_1(\mathrm{id}_Y) = \mathrm{id}_{\Pi_1(Y)}$. Thus $\Pi_1(f)$ is an equivalence of categories. ■

The above can be used to give an alternative proof of Corollary 13.5.3:

Proof. We already know that X is path-connected. By Corollary 13.5.6, $\Pi_1(X)$ is equivalent to $\Pi_1(\{z\})$, which consists only of the object z and its identity morphism. Now Proposition A.5.22 implies $\pi_1(X, x_0) = \mathrm{Hom}_{\Pi_1(X)}(x_0, x_0) \cong \mathrm{Hom}_{\Pi_1(\{z\})}(z, z) = 0$. ■

Remark 13.5.7 1. Provided one knows the categorical prerequisites, the above proofs are simpler and more natural than those for π_1. In particular, the proof of Corollary 13.5.3(i) is just an ad-hoc proof, in this special situation, of Proposition A.5.22.

2. In terms of diagrams, the situation in Proposition 13.5.5 can be represented as follows:

$$\begin{tikzcd} X \arrow[r, bend left=50, "f"{name=U}] \arrow[r, bend right=50, "g"'{name=D}] \arrow[from=U, to=D, Rightarrow, "h"] & Y \end{tikzcd} \quad \rightsquigarrow \quad \begin{tikzcd} \Pi_1(X) \arrow[r, bend left=50, "\Pi_1(f)"{name=U}] \arrow[r, bend right=50, "\Pi_1(g)"'{name=D}] \arrow[from=U, to=D, Rightarrow, "\alpha"] & \Pi_1(Y) \end{tikzcd}$$

What lies behind these diagrams is the following: We know that topological spaces and continuous maps form the category Top and that (small) categories together with functors form the category Cat. But taking also homotopies into account as '2-morphisms', we obtain a '2-category' **Top**. Similarly, the category Cat of categories and functors becomes a 2-category **Cat** with natural transformations as 2-morphisms. Now the proposition amounts to the statement that the functor $\Pi_1 : \mathrm{Top} \to \mathrm{Grpd}$ extends to a '2-functor' **Top**→**Cat** between 2-categories. □

13.6 The Seifert-van Kampen theorem

Here a traditional proof (via manipulation of paths) of Seifert-van Kampen for fundamental groupoids. Cf. [37, Section 6.7] or [226, Sections 2.6-2.7]. Application: Computation of $\pi_1(S^1)$ without use of coverings.

13.7 Covering spaces and applications

13.7.1 Covering spaces. Lifting of paths and homotopies

Definition 13.7.1 *Let $p : \widehat{X} \to X$ be continuous. An open subset $U \subseteq X$ is evenly covered by p if $p^{-1}(U) \subseteq \widehat{X}$ is a union of disjoint open sets, called sheets, each of which if mapped homeomorphically to U by p. Equivalently, there is a homeomorphism $\alpha : p^{-1}(U) \to U \times J$, where J is a discrete space, such that the diagram*

$$\begin{tikzcd} p^{-1}(U) \arrow[r, "\alpha_U"] \arrow[d, "p"'] & U \times J \arrow[d, "p_1"] \\ U \arrow[r, equal] & U \end{tikzcd}$$

commutes. Such an α_U is called a local trivialization of p.

The map p is called a covering map if it is surjective and every $x \in X$ has an open neighborhood that is evenly covered by p. The discrete subspace $p^{-1}(x) \subseteq \widehat{X}$ is the fiber of x.

Example 13.7.2 The quotient map $p : S^n \to \mathbb{RP}^n = S^n/\sim$, where $x \sim y \Leftrightarrow x = \pm y$, is a covering map. As a quotient map, it is continuous and surjective by construction. If $y \in \mathbb{RP}^n$ and $x \in S^n$ such that $p(x) = y$, we can find an open $V \subseteq S^n$ containing x and such that $V \cap (-V) = \emptyset$. Now $U = p(V)$ is an open neighborhood of y that is covered by $p^{-1}(U) = V \cup -V$. Restricted to V or $-V$, p is is injective and thus a homeomorphism. The fiber of each $y \in \mathbb{RP}^n$ has two elements. □

Example 13.7.3 The map $p : \mathbb{R} \to S^1,\ x \mapsto e^{2\pi i x}$ is a covering map. Continuity is obvious, and surjectivity will be assumed known. Given $y \in S^1$, pick $x \in \mathbb{R}$ such that $y = p(x)$ and $\varepsilon \in (0, 1/2)$. Then $U = p((x-\varepsilon, x+\varepsilon))$ is open and $p^{-1}(U) = \bigcup_{n\in\mathbb{Z}}(x-\varepsilon+n, x+\varepsilon+n) \cong U \times \mathbb{Z}$. Since $p : (x-\varepsilon+n, x+\varepsilon+n) \to U$ is a homeomorphism for each $n \in \mathbb{Z}$, p is a covering map. Thus the fiber $p^{-1}(y) = x + \mathbb{Z}$ of each $y \in S^1$ has countably many elements. □

These two examples can be generalized, and we will do so later.

Definition 13.7.4 *If $p : \widehat{Y} \to Y$ is a covering map and $f : X \to Y$ is continuous, a* lift *of f is a continuous map $\widehat{f} : X \to \widehat{Y}$ such that the following commutes:*

$$\begin{tikzcd} & \widehat{Y} \arrow[d, "p"] \\ X \arrow[ur, "\widehat{f}"] \arrow[r, "f"'] & Y \end{tikzcd}$$

Lemma 13.7.5 *Let $p : \widehat{Y} \to Y$ be a covering map, X connected and $f : X \to Y$ continuous. If $\widehat{f}_1, \widehat{f}_2 : X \to \widehat{Y}$ are lifts of f such that $\widehat{f}_1(x) = \widehat{f}_2(x)$ for some $x \in X$ then $\widehat{f}_1 = \widehat{f}_2$.*

Proof. We define $C = \{x \in X \mid \widehat{f}_1(x) = \widehat{f}_2(x)\}$. We will show that C is clopen. Since X is connected and $C \neq \emptyset$ by assumption, we have $C = X$, implying $\widehat{f}_1 = \widehat{f}_2$.

For every $x \in X$, we can find an open neighborhood U of $f(x) \in Y$ that is evenly covered by p. Let $\alpha_U : p^{-1}(U) \to U \times J$ be a local trivialization of p, cf. Definition 13.7.1. Since f is continuous, $V = f^{-1}(U) \subseteq X$ is an open neighborhood of x. In view of $p \circ \widehat{f}_i = f$, we have $\widehat{f}_i(V) \subseteq p^{-1}(U)$. Thus we can consider the composites $\alpha_U \circ \widehat{f}_i : V \to U \times J$ for $i = 1, 2$. Now define

$$C_V = \{x \in V \mid p_2 \circ \alpha_U \circ \widehat{f}_1(x) = p_2 \circ \alpha_U \circ \widehat{f}_2(x)\}.$$

This is the subset of V on which the two lifts $\widehat{f}_1, \widehat{f}_2$ take values in the same sheet above U. In view of the definitions of C and C_V, we have $C_V = C \cap V$. Since the space J labeling the sheets is discrete and the maps under consideration are continuous, C_V is a clopen subset of V. Now the following lemma implies that C is clopen, and we are done. ■

Lemma 13.7.6 *If X is a topological space, $C \subseteq X$ and every $x \in X$ has an open neighborhood U_x such that $U_x \cap C$ is clopen (in U_x), then C is clopen.*

Proof. We have $C = \bigcup_x (U_x \cap C)$, which is open. On the other hand, $X \backslash C = \bigcup_x (U_x \backslash C)$ is open since $U_x \backslash C \subseteq U_x$ is open. Thus C is closed. ■

Proposition 13.7.7 (Path lifting) *Let $p : \widehat{X} \to X$ be a covering map, $f : I \to X$ a path, and $\widehat{x_0} \in p^{-1}(f(0))$. Then there is a unique continuous lift $\widehat{f} : I \to \widehat{X}$ of f such that $\widehat{f}(0) = \widehat{x_0}$.*

Proof. For each $y \in X$, choose an open neighborhood U_y that is evenly covered by p. The open sets $f^{-1}(U_y)$ form an open cover of I. Since I is compact metric, we can find a Lebesgue number $\lambda > 0$. Choose $1/\lambda < M \in \mathbb{N}$. Then the partitioning $0 = s_0 < s_1 < \cdots < s_M = 1$ of $I = [0, 1]$ into M intervals of length $1/M$ has the property that $f([s_i, s_{i+i}])$ is contained in some U_{y_i}. Since $p(\widehat{x_0}) = f(0) \in U_{y_0}$, there is an open neighborhood V_0 of $\widehat{x_0}$ that is mapped homeomorphically to U_{x_0} by p. Now define $\widehat{f}$ on the interval $[0, s_1]$ as $(p \restriction V_0)^{-1} \circ f$. Then $\widehat{f}(0) = \widehat{x_0}$ and $p \circ \widehat{f} = f$ on $[0, s_1]$.

Now we have $p \circ \widehat{f}(s_1) = f(s_1) \in U_{y_1}$, and since $f([s_1, s_2]) \subseteq U_{y_1}$, there is an open neighborhood V_1 of $\widehat{f}(s_1)$ that is mapped homeomorphically to U_{y_1} by p. Now define $\widehat{f}$ on $[s_1, s_2]$ as $(p \restriction V_1)^{-1} \circ f$. Continuing in this way, we obtain the lift $\widehat{f}$ in M steps.

Uniqueness of $\widehat{f}$ follows from Lemma 13.7.5. ■

We will need a generalization of path-lifting to families of paths continuously parametrized by a space Z, i.e. continuous maps $h : Z \times I \to X$.

Theorem 13.7.8 *Let $p : \widehat{X} \to X$ be a covering map, $h \in C(Z \times I, X)$ and $k \in C(Z, \widehat{X})$ such that $p(k(x)) = h(x,0)$ $\forall x \in Z$. Then there is a unique continuous lift $\widehat{h} : Z \times I \to \widehat{X}$ of h such that*

$$\begin{tikzcd} Z \arrow[r, "k"] \arrow[d, "\iota_0"'] & \widehat{X} \arrow[d, "p"] \\ Z \times I \arrow[r, "h"'] \arrow[ur, "\widehat{h}"] & X \end{tikzcd} \tag{13.7}$$

commutes, where $\iota_0(x) = (x,0)$. (Thus $p \circ \widehat{h} = h$ and $\widehat{h}(z,0) = k(z)$.)

Proof. For every $z \in Z$, we have a path $h_z : I \to X$ defined by $h_z(s) = h(z,s)$. By Proposition 13.7.7, there is a unique lift $\widehat{h_z} : I \to \widehat{X}$ such that $p(\widehat{h_z}(s)) = h_z(s) = h(z,s)$ $\forall s \in I$ and $\widehat{h_z}(0) = k(z)$. Defining $\widehat{h} : Z \times I \to \widehat{X}$ by $\widehat{h}(z,s) = \widehat{h_z}(s)$, it is clear that (13.7) commutes. It remains to show that $\widehat{h}$ is continuous. (So far, we only have continuity w.r.t. $s \in I$ for every fixed $z \in Z$.)

Let $\widehat{x} \in \widehat{X}$. Since p is a covering map, there is an open neighborhood $W \ni \widehat{x}$ such that $p(W) \subseteq X$ is open and $p : W \to p(W)$ is a homeomorphism. In view of $p \circ \widehat{h} = h$, we have $\widehat{h}^{-1}(W) = h^{-1}(p(W))$, which is open by continuity of h. Since every open $V \subseteq \widehat{X}$ can be represented as a union of W's as above, $\widehat{h}^{-1}(V)$ is open for every open $V \subseteq \widehat{X}$. ■

Remark 13.7.9 A continuous map $p : \widehat{X} \to X$ is called <u>fibration</u> if given X, h, k as in Theorem 13.7.8 there is a (not necessarily unique) $\widehat{h}$ making (13.7) commute. Thus Proposition 13.7.8 says that every covering map is a fibration with unique diagonals. But not every fibration is a covering map, and the added generality of fibrations is important in higher homotopy theory. □

Corollary 13.7.10 (Homotopy lifting) *Let $p : \widehat{X} \to X$ be a covering map. Assume f, f' are paths in X from x_0 to x_1, and h is a path-homotopy from f to f'. If $\widehat{x_0} \in p^{-1}(x_0)$ and $\widehat{f}, \widehat{f}'$ are the unique lifts of f, f' beginning at $\widehat{x_0}$ then $\widehat{f}(1) = \widehat{f}'(1)$, and h lifts to a path-homotopy $\widehat{h}$ from $\widehat{f}$ to $\widehat{f}'$. (Thus if f, f' are path-homotopic paths and $\widehat{f}, \widehat{f}'$ are lifts beginning in the same fiber over x_0 then they end in the same fiber over x_1, and they are path-homotopic.)*

Proof. Let h_t be a path-homotopy $h : f \to f'$, thus $h_0 = f, h_1 = f'$. Now take $Z = I$, $k = c_{\widehat{x_0}}$ and $h(t,s) = h_t(s)$. Then the hypothesis in Theorem 13.7.8 is satisfied since $p(k(t)) = p(\widehat{x_0}) = x_0 = h(t,0)$ $\forall t \in I = Z$. (We used that h_t is a path-homotopy, thus $h(t,0) = x_0$ for all t.) Thus the theorem gives us a lift $\widehat{h} : I \times I \to \widehat{X}$ of h satisfying $\widehat{h}(t,0) = k(t) = \widehat{x_0}$ $\forall t$ and $p \circ \widehat{h} = h$. In particular, $\widehat{h}(t,1) \in p^{-1}(h(t,1)) = p^{-1}(x_1)$ (again since h is a path-homotopy of paths from x_0 to x_1). Since $t \mapsto \widehat{h}(t,1)$ is continuous and the fiber $p^{-1}(x_1)$ is discrete, we find that $\widehat{h}(t,1)$ is constant, and we call this value $\widehat{x_1}$. Thus the paths $\widehat{h}_t = \widehat{h}(t,\cdot)$ are lifts of h_t for every t, all beginning in $\widehat{x_0}$ and ending in $\widehat{x_1}$. Since $\widehat{h}_0 = \widehat{f}$ and $\widehat{h}_1 = \widehat{f}'$ by construction, we find that $\widehat{h}$ is a path-homotopy from $\widehat{f}$ to $\widehat{f}'$. ■

Exercise 13.7.11 Let $p : \widehat{X} \to X$ be a covering map and $\widehat{x_0} \in \widehat{X}$. Prove that the homomorphism $\pi_1(p) : \pi_1(\widehat{X}, \widehat{x_0}) \to \pi_1(X, p(\widehat{x_0}))$ is injective.

13.7.2 Computation of some fundamental groups

Proposition 13.7.12 *Let $p : \widehat{X} \to X$ be a covering map and $\widehat{x_0} \in \widehat{X}$. Then with $x_0 = p(\widehat{x_0})$, there is a unique map $\alpha : \pi_1(X, x_0) \to p^{-1}(x_0)$ defined by $\alpha([f]) = \widehat{f}(1)$, where $\widehat{f}$ is the lift of f satisfying $\widehat{f}(0) = \widehat{x_0}$.*

If $\widehat{X}$ is path-connected then α is surjective. If $\widehat{X}$ is simply connected then α a bijection.

Proof. Let f be a loop in X based at x_0. By Proposition 13.7.7, there is a unique lift $\widehat{f} : I \to \widehat{X}$ of f such that $\widehat{f}(0) = \widehat{x_0}$. We have $p(\widehat{f}(1)) = f(1) = x_0$, thus $\widehat{f}(1) \in p^{-1}(x_0)$. By Corollary 13.7.10, if two paths f, f' are path-homotopic, their lifts satisfying $\widehat{f}(0) = \widehat{f}'(0) = \widehat{x_0}$ automatically satisfy $\widehat{f}(1) = \widehat{f}'(1)$. Thus $\widehat{f}(1)$ depends only on the homotopy class $[f] \in \pi_1(X, x_0)$, and $\alpha : [f] \mapsto \widehat{f}(1)$ is well-defined.

Now assume that $\widehat{X}$ is path-connected, and let $\widehat{x_1} \in p^{-1}(x_0)$. We can choose a path g from $\widehat{x_0}$ to $\widehat{x_1}$. Now $f = p \circ g : I \to X$ is a path in X from $p(\widehat{x_0}) = x_0$ to $p(\widehat{x_1}) = x_0$, thus a loop based at x_0, and $[f] \in \pi_1(X, x_0)$. The unique lift $\widehat{f}$ of f satisfying $\widehat{f}(0) = \widehat{x_0}$ equals g by uniqueness of path lifting. Thus $\alpha([f]) = \widehat{f}(1) = g(1) = \widehat{x_1}$, so that α is surjective.

Finally, let $\widehat{X}$ be simply connected, and assume that f, f' are loops in X based at x_0 such that $\alpha([f]) = \alpha([f'])$. I.e., the lifts $\widehat{f}, \widehat{f'}$ beginning at $\widehat{x_0}$ satisfy $\widehat{f}(1) = \widehat{f}'(1)$. Then $\widehat{f'}^{-1} \circ \widehat{f}$ is a loop in $\widehat{X}$ based at $\widehat{x_0}$. Since $\widehat{X}$ is simply connected, there is a path-homotopy $\widehat{h}$ from this loop to the constant loop at $\widehat{x_0}$. But then $h = p \circ \widehat{h}$ is a path-homotopy from $f'^{-1} \circ f$ to the constant loop at x_0. This means that $[f'^{-1}][f] = [f'^{-1} \circ f] = [c_{x_0}] = \mathbf{1}$, thus $[f'] = [f]$ in $\pi_1(X, x_0)$. This proves the injectivity of α. ■

We can give the first application:

Corollary 13.7.13 *For every $n \geq 2$, we have $\pi_1(\mathbb{R}\mathrm{P}^n, x_0) \cong \mathbb{Z}/2\mathbb{Z}$. Thus $\mathbb{R}\mathrm{P}^n \not\cong S^n$.*

Proof. For $n \geq 2$, S^n is simply connected by Exercise 13.4.18. Since p is a double covering, i.e. all fibers have two elements, Proposition 13.7.12 implies that $\pi_1(\mathbb{R}\mathrm{P}^n, x_0)$ has two elements. The claim follows from the fact that every two-element group is isomorphic to $\mathbb{Z}/2\mathbb{Z}$. The second claim follows from $\pi^1(S^n) = 0 \ \forall n \geq 2$, cf. Exercise 13.4.18. ■

Applying the same reasoning to the covering map $p : \mathbb{R} \to S^1,\ x \mapsto e^{2\pi i x}$ and noting that $\mathbb{R}$ is simply connected, we see that $\pi_1(S^1, x_0)$ has countably infinitely many elements, but this does not uniquely determine the group structure, and an additional argument is needed. Since S^1 is connected, the choice of the base-point does not matter, and for simplicity we take $x_0 = 1$. Then $p^{-1}(1) = \mathbb{Z}$, thus by Proposition 13.7.12 there is a bijection $\pi_1(S^1, 1) \to \mathbb{Z}$.

Proposition 13.7.14 *The bijection $\pi_1(S^1, 1) \to \mathbb{Z}$ is an isomorphism of groups.*

Proof. Let $f_1, f_2 : I \to S^1$ be loops based at 1. Let $h_1, h_2 : I \to \mathbb{R}$ be lifts to $\mathbb{R}$ beginning at 0, i.e. $f_i(t) = e^{2\pi i h_i(t)}$. Define

$$h_3(t) = \begin{cases} h_1(2t) & \text{if } \ 0 \leq t \leq 1/2 \\ h_1(1) + h_2(2t-1) & \text{if } \ 1/2 \leq t \leq 1 \end{cases}$$

It is easy to see that $f_3(t) = e^{2\pi i h_3(t)}$ is the composite path $f_1 \bullet f_2$, and h_3 is the lift of f_3. Now $h_3(1) = h_1(1) + h_2(2)$. Since $h_i(1)$ is the image of $[f_i]$ under the map $\pi_1(S^1, 1) \to \mathbb{Z}$, we have proven that the latter is a homomorphism. ■

Remark 13.7.15 The number $\alpha([f]) \in \mathbb{Z}$ associated to a loop f based at 1 is called the index or winding number. □

Definition 13.7.16 *The n-torus is defined by $T^n = (S^1)^{\times n}$, i.e. as the product of n circles S^1.*

Now we can state a first application of the fundamental group(oid) to the classification of topological spaces:

Corollary 13.7.17 *For $n, m \in \mathbb{N}_0$, we have $T^n \cong T^m$ if and only if $n = m$.*

Proof. It is is trivial that $n = m \Rightarrow T^n \cong T^m$. Since S^1 is path-connected, so is T^n for every n, so that it makes sense to write $\pi_1(T^n)$. Using $\pi_1(S^1) \cong \mathbb{Z}$ and Exercise 13.4.17(i), we have $\pi_1(T^n) \cong \mathbb{Z}^n$. For $n \neq m$ we know from group theory that $\mathbb{Z}^n \ncong \mathbb{Z}^m$, thus $\pi_1(T^n) \ncong \pi_1(T^m)$, implying $T^n \ncong T^m$. ■

Remark 13.7.18 1. Similar to the proof of s-dim$(S^n) = n$ one can show s-dim$(T^n) = n$. Thus the above corollary also follows from dimension theory. But the latter is useless for the classification of surfaces mentioned next, since they all have dimension 2.

2. Another class of spaces which can be distinguished by their fundamental groups are the compact orientable (2-dimensional) surfaces without boundary. The simplest examples are the 2-sphere S^2 and the 2-torus T^2. The latter can be considered as the 2-sphere with one 'handle' attached. Attaching n handles to S^2 gives the surface M_g with 'g holes'. One can show that $\pi_1(M_g)$ is isomorphic to the finitely presented group $G_g = \langle a_1, \ldots, b_g \mid [a_1, b_1] \cdots [a_g, b_g] = e \rangle$. (For $g = 0$ this is the trivial group and for $g = 1$ this is $\mathbb{Z}^2$, consistent with $M_0 = S^2, M_1 = T^2$.) For $g \geq 2$, G_g is non-abelian and somewhat complicated. (The Poincaré conjecture in three dimensions has a purely algebraic reformulation involving the groups G_g and the free groups F_n. The only known proof of this algebraic statement is via Perelman's proof of the Poincaré conjecture!) But it is easy to see that the abelianization of G_g is $\mathbb{Z}^{2g}$, so that the spaces M_g are mutually non-homeomorphic. □

The homotopy theory developed thus far can be used to give another proof of the algebraic closedness of $\mathbb{C}$, i.e. the fact that every non-constant polynomial $P \in \mathbb{C}[x]$ has a zero.

Exercise 13.7.19 (Fundamental theorem of algebra) As usual, write $S^1 = \{z \in \mathbb{C} \mid |z| = 1\}$. Let $P(z) = a_n z^n + \ldots + a_0$ be a polynomial of degree $n \geq 1$ without zeros. We may assume that $a_n = 1$.

(i) For $R \geq 0$ define
$$f_R : S^1 \to S^1, \quad z \mapsto \frac{P(Rz)}{|P(Rz)|}.$$
Use this to prove that $f_1 : S^1 \to S^1$ is homotopic with a constant map.

(ii) Define $g : S^1 \times [0, 1] \to \mathbb{C}$ by
$$g(z, t) = \begin{cases} f_{1/t}(z) & t \in (0, 1] \\ z^n & t = 0 \end{cases}$$
Prove that g is a homotopy between f_1 and the map $z \mapsto z^n$.

(iii) Show that the conclusion of (ii) contradicts results we proved above.

13.7.3 Properly discontinuous actions of discrete groups

The above computation of $\pi_1(S^1)$ can be generalized considerably:

Definition 13.7.20 *Let X be a topological space and G a (discrete) group. An action of G on X is a map $G \times X \to X, (g,x) \mapsto gx$ such that $x \mapsto gx$ is continuous for each $g \in G$, and $(gh)x = g(hx)$ and $ex = x$ for all $g, h \in G,\ x \in X$.*

An action of G on X is called properly discontinuous if every $x \in X$ has a neighborhood V such that $gV \cap V = \emptyset\ \forall g \neq e$.

Remark 13.7.21 In view of $g^{-1}(gx) = (g^{-1}g)x = x$, the map $x \mapsto gx$ is a homeomorphism for each g. Therefore, an action of a discrete group G on X is the same as a group homomorphism of G to the group $\mathrm{Aut}(X)$ of homeomorphisms of X (i.e. automorphisms of X in the category Top). □

Example 13.7.22 1. $G = \mathbb{Z}$ acts properly discontinuously on $X = \mathbb{R}$ by $(n,x) = n + x$.

2. $G = \{+,-\} \cong \mathbb{Z}/2\mathbb{Z}$ acts properly discontinuously on $X = S^n$ by $(\pm, x) = \pm x$. □

An action of G on X defines an equivalence relation by $x \sim_G y \Leftrightarrow x \in Gy$. We write X/G instead of $X/{\sim_G}$.

Exercise 13.7.23 Prove that $\sim_G$ is an equivalence relation.

Proposition 13.7.24 *(i) Let X be a space acted upon properly discontinuously by the group G. Then the quotient map $p : X \to X/G$ is a covering map.*

(ii) For every $x_0 \in X/G$ we have a homomorphism $\pi_1(X/G, x_0) \to G$, which is is surjective if X is path-connected and bijective if X is simply connected.

Proof. (i) The map $p : X \to X/G$ is continuous and surjective by construction. Since G acts properly discontinuously, for every $x \in X$ we can find an open neighborhood $V \subseteq X$ such that $gV \cap V = \emptyset\ \forall g \neq e$. This clearly implies $gV \cap hV = \emptyset$ whenever $g \neq h$. Then $U = p(V)$ is a neighborhood of $p(x)$ evenly covered by p, thus p is a covering map, the sheets over U being $\{gV\}_{g \in G}$.

(ii) Given $x_0 \in X/G$, pick $\widehat{x_0} \in p^{-1}(x_0)$. Now Proposition 13.7.12 provides a map $\alpha : \pi_1(X/G, x_0) \to p^{-1}(x_0)$ such that $\alpha([f]) = \widehat{f}(1)$ (terminology as in that proposition). Since G acts properly discontinuously and by definition of the equivalence relation $\sim$, for every $x \in p^{-1}(x_0)$ there is a unique $g \in G$ such that $x = g\widehat{x_0}$. This defines a map $\beta : p^{-1}(x_0) \to G$, and now the composite $\gamma = \beta \circ \alpha : \pi_1(X/G, x_0) \to G$ is the desired map. Since β is a bijection, it follows from Proposition 13.7.12 that γ is a surjection if X is path-connected and a bijection if X is simply connected. That γ is a group homomorphism is shown in essentially the same way as in the proof of Proposition 13.7.14. ■

Corollary 13.7.25 $\pi_1(S^1, x_0) \cong \mathbb{Z}$ *for every* $x_0 \in S^1$.

Proof. As noted above, $\mathbb{Z}$ acts properly discontinuously on $\mathbb{R}$ by $(n,x) \mapsto n + x$. The quotient space $\mathbb{R}/\mathbb{Z}$ is homeomorphic to S^1. Since $\mathbb{R}$ is simply connected, Proposition 13.7.24 implies $\pi_1(S^1) \cong \mathbb{Z}$. ■

In the computation of $\pi_1(S^1)$ we have more structure than in the computation of $\pi_1(\mathbb{R}\mathrm{P}^n)$: $\mathbb{R}$ is a topological group containing $\mathbb{Z}$ as a subgroup. This motivates the following generalization:

Corollary 13.7.26 *If G is a topological group and $N \subseteq G$ a discrete normal subgroup then the quotient map $p : G \to G/N$ is a covering map and there is a homomorphism $\pi_1(G/N, e) \to N$. The latter is surjective if G is path-connected and bijective if G is simply connected.*

Proof. We let N act on G by $(n, g) \mapsto ng$. The discreteness of N implies that the action is properly discontinuous, and the rest follows from Proposition 13.7.24. ■

We have seen in Exercise 13.4.21 that the fundamental group of a topological group is abelian. The preceding corollary thus implies that every discrete normal subgroup $N \subseteq G$ of a path-connected topological group is abelian. Actually, it is easy to give a direct proof of a stronger result:

Lemma 13.7.27 *If G is a connected topological group and $N \subseteq G$ is a discrete normal subgroup then $N \subseteq Z(G)$. ($Z(G) = \{g \in G \mid gh = hg \ \forall h \in G$ is the center of G.)*

Proof. Fix $n \in N$ and consider the map $G \to N, \ g \mapsto gng^{-1}$. Since N is discrete and G connected, this map must be constant, thus $gng^{-1} = n$ for all $g \in G$. But this is equivalent to $n \in Z(G)$. Since this holds for all $n \in N$, we have $N \subseteq Z(G)$. ■

Example 13.7.28 1. The group $SU(n)$ can be shown to be simply connected for every $n \geq 2$. Its center is given by $Z(SU(n)) = \{c\mathbf{1} \mid c^n = 1\} \cong \mathbb{Z}/n\mathbb{Z}$. Thus if $A \subseteq Z(n)$ is a subgroup, we have $\pi_1(SU(n)/A) \cong A$.

2. It is well known that there is a double covering of $SO(3)$ by $SU(2)$. Thus $SO(3) \cong SU(2)/Z(SU(2))$, "explaining" the fact $\pi_1(SO(3)) = \mathbb{Z}/2\mathbb{Z}$.

3. More generally, for every $n \geq 3$ there is a simply connected compact group $\mathrm{Spin}(n)$ whose center is $\mathbb{Z}/2\mathbb{Z}$ and such that $\mathrm{Spin}(n)/Z(\mathrm{Spin}(n))/ \cong SO(n)$. (In fact, $\mathrm{Spin}(3) \cong SU(2)$.) □

13.7.4 The universal covering space

Here the standard construction of the universal covering of a semilocally simply connected space. Perhaps some comments on non-connected spaces.

13.7.5 Classification of covering spaces

Here the classification of coverings of a semilocally simply connected space in terms of subgroups of π_1. (Merge this with Section 13.7.3.) Connections to Galois theory. ([23, 141, 209, 221, 228])

13.7.6 Seifert-van Kampen via coverings

Here a Grothendieck-style proof of Seifert-van Kampen (in the connected case) via covering space theory. References: [85, 26, 59, 143, 208]

Appendix A

Background on sets and categories

A.1 Reminder of basic material

The basic material from (naive) set theory recalled in this subsection is fundamental and is used all the time. One should not attempt to study topology before everything, including the proofs (most of which are not given here) appears utterly evident.

A.1.1 Notation. Sets. Cartesian products

The following is a rapid summary of the notions, notations and facts that are assumed known.

Definition/Proposition A.1.1 • $\vee$ *stands for 'or',* $\wedge$ *for 'and',* $\neg$ *for 'not'.*

- *The contraposition of* $A \Rightarrow B$ *is the equivalent statement* $\neg B \Rightarrow \neg A$.
- *We write* $\mathbb{N} = \{1, 2, \ldots\}$ *(thus 0 is NOT considered a natural number),* $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$ *and* $\mathbb{R}_{\geq 0} = [0, \infty)$.
- *If* A, B *are sets,* $A \subseteq B$ *means* $x \in A \Rightarrow x \in B$*, thus* $\subseteq$ *denotes not necessarily strict inclusion. Strict (or proper) inclusion (*$A \subseteq B$*, but* $A \neq B$*) is denoted by* $A \subsetneq B$.[1]
- *The set of all subsets of* X*, the powerset of* X*, is denoted by* $P(X)$.
- *If* $\mathcal{F} \subseteq P(X)$ *is a family of subsets of* X *then* $\bigcup \mathcal{F}$ *denotes the union of all* $F \in \mathcal{F}$*, i.e.*

$$\bigcup \mathcal{F} = \bigcup_{F \in \mathcal{F}} F = \{x \in X \mid \exists F \in \mathcal{F} : x \in F\}.$$

- *If* $Y_i \subseteq X \ \forall i$ *then De Morgan's formulae*[2] *hold:*

$$X \backslash \bigcup_i Y_i = \bigcap_i (X \backslash Y_i), \qquad X \backslash \bigcap_i Y_i = \bigcup_i (X \backslash Y_i).$$

[1] It might be more "logical" or more consistent with the use of $<$ vs. $\leq$ to simply write $\subset$ instead of $\subsetneq$ to denote strict inclusion. It is a fact, however, that a vast majority of mathematicians uses $\subset$ to denote *non-strict* inclusion, so that doing otherwise (as some authors insist on doing) can only create confusion. We therefore prefer to avoid $\subset$ and $\supset$ altogether, as done e.g. in [224].

[2] Augustus de Morgan (1806-1871). British mathematician and logician.

- *The cardinality of a set X is denoted by $\#X$. $\aleph_0 := \#\mathbb{N}$ and $\mathfrak{c} = \#\mathbb{R} = \#P(\mathbb{N}) = 2^{\aleph_0}$.*

- *"X is countable" means $\#X \le \aleph_0$. Occasionally we write 'at most countable', but this is only for emphasis. When $\#X = \aleph_0$ is intended, we always write "X is countably infinite".*

- *Given two sets X, Y, there is another set, the Cartesian (or direct) product $X \times Y$, having as elements the pairs (x, y), where $x \in X, y \in Y$. Finite direct products $X_1 \times \cdots \times X_n$ are defined by induction.*

- *A function (or map) $f : X \to Y$ is a subset $G(f) \subseteq X \times Y$ such that $\#\{y \in Y \mid (x, y) \in G(f)\} = 1$ for each $x \in X$. The unique $y \in Y$ such that $(x, y) \in G(f)$ is denoted $f(x)$. (We can consider f and $G(f)$ as the same thing, which is probably better than considering $G(f)$ as the graph of the function f, the latter being considered as primary in some unclear sense.)*

- *If $f : X \to Y$ is a function and $Z \subseteq X$ then $f \upharpoonright Z$ or $f_{|Z}$ denotes the restriction of f to Z, i.e. the function $Z \to Y$ defined by $z \mapsto f(z)$ for $z \in Z$. Clearly $G(f \upharpoonright Z) = G(f) \cap (Z \times Y)$.*

- *There are projection maps $p_1 : X \times Y \to X$, $p_2 : X \times Y \to Y$ such that $p_1((x, y)) = x, p_2((x, y)) = y$. (Alternatively, we may write p_X, p_Y.)*

- *The set of all functions from X to Y is denoted $\mathrm{Fun}(X, Y)$ or Y^X.*

- *If X is a set, the characteristic function of X is defined by $\chi_X(x) = 1$ if $x \in X$ and $\chi_X(x) = 0$ if $x \notin X$.*

- *If $Y \subseteq X$ then $\chi_Y \in \{0, 1\}^X$. (Strictly speaking, we should write $\chi_Y \upharpoonright X$, but we don't.) Conversely, if $\chi \in \{0, 1\}^X$ then $Y = \{x \in X \mid \chi(x) = 1\} \subseteq X$. One easily checks that this establishes a bijection $P(X) \leftrightarrow \{0, 1\}^X$.*

- *A union of countably many countable sets is countable.*

- *A product of finitely many countable sets is countable.*

- *Cantor's diagonal argument: $\#P(X) > \#X$ for every set X. In particular, $P(\mathbb{N})$ is uncountable.*

- *It is not true that countable products (defined later) of countable sets are countable! In fact, if I is infinite and $\#X \ge 2$ then $\#X^I \ge \#\{0, 1\}^I \ge \#\{0, 1\}^{\mathbb{N}} = \#P(\mathbb{N}) > \#\mathbb{N}$, thus X^I is uncountable. (This will be generalized in the next subsection.)*

- *If X is infinite and Y is finite then there is a bijection $X \times Y \cong X$. In terms of cardinal numbers: $n\chi = \chi$ whenever $n \in \mathbb{N}$ and $\chi \ge \#\mathbb{N}$.*

Remark A.1.2 Central as the notion of products is, experience tells that there are widespread misconceptions involving it and its difference to the (disjoint) union of sets. This leads to many mistakes. Here is a list of the most frequent ones:

- If $A = \emptyset$ or $B = \emptyset$ then $A \times B = \emptyset$. Everything else is nonsense. (Like: If $A = \emptyset$ then $A \times B$ 'does not exist' or '$A \times B = \{b \mid b \in B\}$'.)

- If $A \subseteq X$ and $B \subseteq Y$ then $A \times B \subseteq X \times Y$. But 'most' subsets of $X \times Y$ are not of this form, as the example of the plane shows (where $X = Y = \mathbb{R}$). If $C \subseteq X \times Y$ and one defines $A = p_1(C) \subseteq X, B = p_2(C) \subseteq Y$ then $C \subsetneq A \times B$, but equality only holds if C is of the form $A' \times B'$ in the first place.

- If $A, B \subseteq X$ and $C, D \subseteq Y$ then $(A \times C) \cap (B \times D) = (A \cap B) \times (C \cap D)$. But: $\boxed{(A \times C) \cup (B \times D) \subsetneq (A \cup B) \times (C \cup D)}$. Here $\subsetneq$ can be replaced by $=$ only in special cases, like $A = B$ or $C = D$ which lead to the distributivity laws $(A \times C) \cup (A \times D) = A \times (C \cup D)$ and $(A \times C) \cup (B \times C) = (A \cup B) \times C$, as well as some uninteresting degenerate cases like $(A = X$ and $C = Y)$ or $(A = \emptyset$ and $C \subseteq D)$ and some others.

- If $A \subseteq X, B \subseteq Y$ then

$$\{(x, y) \mid x \notin A \wedge y \notin B\} = \boxed{(X \backslash A) \times (Y \backslash B) \subsetneq (X \times Y) \backslash (A \times B)} = \{(x, y) \mid x \notin A \vee y \notin B\}.$$

  Equality holds only in very degenerate cases, namely when at least one of the following conditions holds: (1) $X = \emptyset$, (2) $Y = \emptyset$, (3) $A = B = \emptyset$, (4) $A = X$, $B = Y$.

- Let $X_1 = X_2 = \{a, b\}$ and $X = X_1 \times X_2$. If $Y = \{(a, a), (a, b), (b, b)\} \subseteq X$ then $X \backslash Y = \{(b, a)\}$. Now $p_1(Y) = \{a, b\}$ and $p_1(X \backslash Y) = \{b\}$. Thus it is not true that $X_i \backslash p_i(Y) = p_i(X \backslash Y)$ for each $Y \subseteq X$.

- Two corollaries of the preceding point are: If $Y, Z \subseteq X = \prod_i X_i$ satisfy $Y \cap Z = \emptyset$ then it does not follow that $p_i(Y) \cap p_i(Z) = \emptyset$. (This is related to the fifth point of Lemma A.1.7.) In particular $x \notin Y$ does not imply $p_i(x) \notin p_i(Y)$.

Why these mistakes are made is mysterious. After all, everyone understands that the family of rectangles in the plane (with edges parallel to the axes) is closed under intersection, but not under union or complements. It is even more evident that not every subset of $\mathbb{R}^2$ is such a rectangle. $\square$

A.1.2 More on Functions

Definition A.1.3 *If $f : X \to Y$ is a function, $y \in Y$ and $A \subseteq Y$, we define*

$$f^{-1}(y) = \{x \in X \mid f(x) = y\}, \qquad f^{-1}(A) = \{x \in X \mid f(x) \in A\}.$$

Clearly, the two definitions are consistent in that $f^{-1}(\{x\}) = f^{-1}(x)$.

Definition A.1.4 *Let $f : X \to Y$ be a function.*

- *f is injective if $f(x) = f(y)$ implies $x = y$. (I.e., $\#f^{-1}(y) \leq 1 \ \forall y \in Y$.)*
- *f is surjective if for every $y \in Y$ we have $f^{-1}(y) \neq \emptyset$. (I.e., $\#f^{-1}(y) \geq 1 \ \forall y \in Y$.)*
- *f is bijective if it is injective and surjective.*

Lemma A.1.5 *Let $f : X \to Y$ and $g : Y \to Z$ be functions.*

- *If f and g are injective then $g \circ f$ is injective. If f and g are surjective then $g \circ f$ is surjective.*
- *If $g \circ f$ is injective then f and $g \restriction f(X)$ are injective, but g need not. If $g \circ f$ is surjective then g is surjective, but f need not.*
- *f is bijective (=injective and surjective) if and only if it has an inverse function, i.e. a function $h : Y \to X$ such that $h \circ f = \mathrm{id}_X$, $f \circ h = \mathrm{id}_Y$.*

Lemma A.1.6 *Let $f : X \to Y$ be a function. Then*

- $\boxed{S \subseteq X \Rightarrow S \subseteq f^{-1}(f(S)).}$ *Equality holds for all $S \subseteq X$ if and only if f is injective.*
- $\boxed{T \subseteq Y \Rightarrow f(f^{-1}(T)) \subseteq T.}$ *Equality holds for all $T \subseteq Y$ if and only if f is surjective.*

Lemma A.1.7 *Let $f : X \to Y$ be a function, $A, A_i, B \subseteq X$ and $C, C_i, D \subseteq Y$. Then*

- $f^{-1}\left(\bigcap_{i\in I} C_i\right) = \bigcap_{i\in I} f^{-1}(C_i)$.
- $f^{-1}\left(\bigcup_{i\in I} C_i\right) = \bigcup_{i\in I} f^{-1}(C_i)$.
- $f^{-1}(C)\backslash f^{-1}(D) = f^{-1}(C\backslash D)$.
- $f\left(\bigcup_{i\in I} A_i\right) = \bigcup_{i\in I} f(A_i)$.
- $\boxed{f(A\cap B) \subseteq f(A)\cap f(B).}$ *Equality holds for all A, B if and only if f is injective.*
- $\boxed{f(A)\backslash f(B) \subseteq f(A\backslash B).}$ *Equality holds for all A, B if and only if f is injective.*

A.1.3 Relations

Definition A.1.8 *If X is a set, a (binary) relation on X is a subset $R \subseteq X \times X$. Instead of $(x, y) \in R$ one often writes xRy. A relation R on X is called*

(i) reflexive if xRx for all $x \in X$,

(ii) transitive if the combination of xRy and yRz implies xRz,

(iii) symmetric if $xRy \Leftrightarrow yRx$,

(iv) antisymmetric if the combination of xRy and yRx implies $x = y$,

(v) preorder if R is reflexive and transitive,

(vi) partial order if R is reflexive, transitive, and antisymmetric.

(viii) equivalence relation if R is reflexive, transitive, and symmetric,

Equivalence relations are usually denoted by one of the symbols $\sim, \simeq, \cong$, depending on the context. Partial orders are usually denoted by $\leq$.

Remark A.1.9 More generally, a relation on sets $X_1, \dots, X_n$ is a subset R of $X_1 \times \cdots \times X_n$. In particular, functions are special cases of relations: A relation R on $X \times Y$ is a function if and only if it is left-total (for every $x \in X$ there is at least one $y \in Y$ such that $(x, y) \in R$) and right-unique (for every $x \in X$ there is at most one $y \in Y$ such that $(x, y) \in R$). □

For more on partial orders, cf. Section A.3.3. Here we focus on equivalence relations.

Exercise A.1.10 Let $\leq$ be a preorder on X. Prove that $x \sim y :\Leftrightarrow (x \leq y \wedge y \leq x)$ defines an equivalence relation.

For every subset $S \subseteq X \times X$, there is a smallest equivalence relation $\sim$ on X such that $(x,y) \in S \Rightarrow x \sim y$. Just define $\sim$ to be the intersection of all equivalence relations $R \subseteq X \times X$ such that $S \subseteq R$.

Let $\sim$ be an equivalence relation on a set X. For $x \in X$ define $[x] = \{y \in X \mid y \sim x\}$, the equivalence class of x. We write $X/\sim = \{[x] \mid x \in X\}$ for the set of $\sim$-equivalence classes. For $x, y \in X$ one either has $[x] = [y]$ or $[x] \cap [y] = \emptyset$. Thus the equivalence relation gives rise to a partitioning of X in disjoint subsets. Conversely, a partitioning $X = \bigcup_{i \in I} X_i$, where $i \neq j \;\Rightarrow\; X_i \cap X_j = \emptyset$, defines an equivalence relation $\sim$ via $x \sim y \;\Leftrightarrow\; \exists i : \{x, y\} \subseteq X_i$.

There is an obvious surjective map $p : X \to X/\sim, x \mapsto [x]$. On the other hand, every function $f : X \to Y$ defines an equivalence relation $\sim$ on X via $x \sim y \Leftrightarrow f(x) = f(y)$. The equivalence classes are the non-empty sets of the form $f^{-1}(y)$. If f is surjective then the 'non-empty' can be dropped and there is a unique map $g : X/\sim \to Y$ such that $g \circ p = f$. The above constructions essentially give a bijection between equivalence relations on X and surjective maps $f : X \to Y$. [In order to make the 'essentially' precise, we call surjective maps $f_1 : X \to Y_1$, $f_2 : X \to Y_2$ isomorphic if there is a bijection $h : Y_1 \to Y_2$ such that $h \circ f_1 = f_2$. Then there is a bijection between equivalence relations on X and isomorphism classes of surjections $f : X \to Y$.]

Given a function $f : X \to Y$ and an equivalence relation $\sim$ on X, it is important to know when f 'descends' to a function $X/\sim \to Y$:

Lemma A.1.11 *(i) If X, Y are sets and $\sim$ is an equivalence relation on X then there is a bijection between maps $g : X/\sim \to Y$ and maps $f : X \to Y$ that are constant on equivalence classes (i.e. $x \sim x'$ implies $f(x) = f(x')$) such that $f = g \circ p$, i.e. the diagram*

$$\begin{tikzcd} X \arrow[d, "p"'] \arrow[dr, "f"] & \\ X/\sim \arrow[r, "g"'] & Y \end{tikzcd} \tag{A.1}$$

commutes.

(ii) g is injective if and only if $f(x) = f(y) \;\Rightarrow\; x \sim y$.

(iii) g is surjective if and only if $f : X \to Y$ is surjective.

Proof. (i) Let $p : X \to X/\sim$ be the quotient map and $g : X/\sim \to Y$ a map. Clearly $f := g \circ p : X \to Y$ is constant on equivalence classes. Conversely, given $f : X \to Y$ constant on equivalence classes, define $g : X/\sim \to Y$ as follows: For $c \in X/\sim$ take a representer $x \in c$ and define $g(c) = f(x)$. It is clear that the choice of $x \in c$ does not matter. It is easy to see that these two constructions are inverses of each other.

(ii) By assumption on f, we have $x \sim y \;\Rightarrow\; f(x) = f(y)$. If $f(x) = f(y)$ then $g(p(x)) = g(p(y))$, thus if g is injective we have $p(x) = p(y)$, which is equivalent to $x \sim y$. Conversely, assume $f(x) = f(y) \;\Rightarrow\; x \sim y$, and let $c, d \in X/\sim$ satisfy $g(c) = g(d)$. Then $f(x) = f(y)$ whenever $x \in c, y \in d$. But then the assumption on f implies $x \sim y$ and thus $c = d$. Thus g is injective.

(iii) In view of $f = g \circ p$ and the surjectivity of p, this is immediate by Lemma A.1.5. ■

A.2 Disjoint unions and direct products

A.2.1 Disjoint unions

Occasionally, given a family $\{X_i\}_{i\in I}$ of sets, we need a set X "containing each X_i as a subset", or rather more precisely a set X together with injective maps $\iota_i : X_i \to X$. The obvious solution is $X = \bigcup_{i\in I} X_i$ with ι_i the inclusion maps. However, this leads to a 'loss of points' if the X_i are not all mutually disjoint: $\{1,2,3\}\cup\{3,4,5\} = \{1,2,3,4,5\}$ has 5 elements instead of $6 = 3+3$. What one really wants is a set X together with injective maps $\iota_i : X_i \to X$ such that $X = \bigcup_i \iota_i(X)$ and $\iota_i(X_i)\cap\iota_j(X_j) = \emptyset$ whenever $i \neq j$. The simplest solution is obtained by keeping track of from which X_i a point $x \in X$ originates:

Definition A.2.1 *Let X_i be a set for every $i \in I$. Then the <u>disjoint union</u> is defined by*

$$\bigoplus_{i\in I} X_i = \{(i,x) \in I \times \bigcup_{k\in I} X_k \mid i \in I,\ x \in X_i\}.$$

For every $k \in I$, we define the <u>inclusion map</u>

$$\iota_i : X_i \to \bigoplus_{k\in I} X_k, \quad x \mapsto (i,x).$$

We also define $b : \bigoplus_{k\in I} X_k \to I$ by $(i,x) \mapsto i$.

Remark A.2.2 1. The disjoint union of sets is also called the <u>direct sum</u> or <u>coproduct</u>, since its properties are dual to that of the product, cf. below. Frequently the symbols $\coprod$ or $\uplus$ are used instead of $\oplus$. We will consistently use $\oplus$ for sets and topological spaces.

2. The point of this construction is that $\bigoplus_{i\in I} X_i$ is a set that contains all the $X_i, i \in I$ as subsets, keeping track of 'from which X_i a point x comes', even if $X_i \cap X_j \neq \emptyset$. The following Lemma makes this precise. □

Lemma A.2.3 *Let X_i be a set for every $i \in I$. Then*

(i) The maps $\iota_i : X_i \to \bigoplus_{k\in I} X_k$ are injective.

(ii) If $i \neq j$ then $\iota_i(X_i) \cap \iota_j(X_j) = \emptyset$.

(iii) $\displaystyle\bigcup_{i\in I} \iota_i(X_i) = \bigoplus_{k\in I} X_k.$

(iv) $b(\iota_i(x)) = i\ \forall x \in X_i.$

(v) The map b is surjective if and only if $X_i \neq \emptyset$ for each $i \in I$.

Proof. Obvious. ■

Lemma A.2.4 *Let $A_i, B_i \subseteq X_i\ \forall i \in I$. Then*

$$(\bigoplus_i A_i) \cup (\bigoplus_i B_i) = \bigoplus_i (A_i \cup B_i), \qquad (\bigoplus_i A_i) \cap (\bigoplus_i B_i) = \bigoplus_i (A_i \cap B_i)$$

$$(\bigoplus_i X_i)\backslash(\bigoplus_i A_i) = \bigoplus_i (X_i \backslash A_i).$$

The following is the 'universal property' of the disjoint union of set and characterizes it (up to isomorphism):

Proposition A.2.5 *Let $\{X_i\}_{i\in I}$ be a family of sets and Y a set. Then there is a bijection between maps $f : \bigoplus_k X_k \to Y$ and families of maps $\{f_i : X_i \to Y\}_{i\in I}$.*

Proof. If $f : \bigoplus_k X_k \to Y$ is given, define $f_i = f \circ \iota_i$. If a family $\{f_i : X_i \to Y\}_{i\in I}$ is given, define $f : \bigoplus_k X_k \to Y$ by $f((i,x)) = f_i(x)$. It is clear that these two constructions are inverses of each other. ■

A.2.2 Arbitrary direct products

Here we generalize the notion of a direct product of any family $\{X_i\}_{i\in I}$ of sets. Since we write $X_1 \times X_2$ in infix notation, it may seem natural to write $\bigtimes_{i\in I} X_i$ for general products, but very few authors (like those of [77]) do this. We follow the standard practice of writing $\prod_{i\in I} X_i$. (Conversely, $X_1 \prod X_2$ is even less common.)

Definition A.2.6 *Let X_i be a set for every $i \in I$. Then the direct product is defined by*

$$\prod_{i\in I} X_i = \left\{ f : I \to \bigcup_{k\in I} X_k \mid f(i) \in X_i \ \forall i \in I \right\}.$$

The projection maps are defined by

$$p_i : \prod_{k\in I} X_k \to X_i, \quad f \mapsto f(i).$$

Again, there is a universal property:

Proposition A.2.7 *Given a set X and a family of sets $\{Y_i, i \in I\}$, there is a bijection between maps $f : X \to \prod_k Y_k$ and families of maps $\{f_i : X \to Y_i\}_{i\in I}$.*

Proof. Given $f : X \to \prod_k Y_k$, define $f_i = p_i \circ f$. If a family $\{f_i : X \to Y_i\}_{i\in I}$ is given, define $f : X \to \prod_k Y_k$ by $f(x) : I \to \bigcup_k Y_k$, $i \mapsto f_i(x)$. Again, it is easy to see that the constructions are inverses of each other. ■

Lemma A.2.8 *Let $A_i, B_i \subseteq X_i \ \forall i \in I$. Then*

$$(\prod_i A_i) \cap (\prod_i B_i) = \prod_i (A_i \cap B_i), \qquad (\prod_i A_i) \cup (\prod_i B_i) \subseteq \prod_i (A_i \cup B_i),$$

$$\prod_i (X_i \backslash A_i) \subseteq (\prod_i X_i) \backslash (\prod_i A_i).$$

As in the case of finite products, replacing the $\subseteq$ by $=$ gives formulas that are hardly ever true! (It is instructive to work out when this is the case.)

Remark A.2.9 IMPORTANT: Let I and $\{X_i\}_{i\in I}$ be finite sets. Then

$$\#\left(\bigoplus_{i\in I} X_i\right) = \sum_i \#X_i, \qquad \#\left(\prod_{i\in I} X_i\right) = \prod_i \#X_i,$$

cf. e.g. [171, Exercise 3 (!)]. It is therefore obvious that the two construction are not the same – just as addition and multiplication of natural numbers are not the same! As recalled in the preceding subsection, a countable union of countable sets is countable and a finite product of countable sets is countable. But if I is infinite and $\#X_i \geq 2\ \forall i \in I$ then

$$\#\prod_{i\in I} X_i \geq \#\{0,1\}^I \geq \#\{0,1\}^{\mathbb{N}} = \#P(\mathbb{N}) > \#\mathbb{N},$$

as everyone has learned in the basic courses on set theory, basic logic (cf. e.g. [171, Proposition 1.1.5]) and analysis, cf. e.g. [223, Section 8.3]. (In this topology course, it reappears e.g. in the surjection $\{0,1\}^{\mathbb{N}} \to [0,1]$ of (11.4) and in Lemma 11.1.32.) Thus: An infinite product (countable or not) of non-trivial (in the sense of $\#X_i \geq 2\ \forall i$) sets is uncountable! [3] □

It is clear that $\prod_{i\in I} X_i = \emptyset$ if $X_i = \emptyset$ for some $i \in I$ (since in that case we cannot satisfy $f(i) \in X_i$). At least when I is infinite, the usual axioms of set theory (whatever they are, like Zermelo-Frenkel) do not imply that $\prod_{i\in I} X_i \neq \emptyset$, even when $X_i \neq \emptyset\ \forall i \in I$. This leads to the Axiom of Choice, which will be discussed next.

A.3 Choice axioms and their equivalents

A.3.1 Three formulations of the Axiom of Choice

Definition A.3.1 *Axiom of Choice (AC): $\prod_{i\in I} X_i \neq \emptyset$ whenever $X_i \neq \emptyset\ \forall i \in I$.*

We will give two equivalent versions of the Axiom of Choice that are often useful. This requires some more definitions.

Definition A.3.2 *Let $f : X \to Y$ and $g : Y \to X$ be functions such that $g \circ f = \mathrm{id}_X$. In this situation, g is called a left-inverse of f, and f is a right-inverse or section of g.*

Lemma A.3.3 *(i) Let $f : X \to Y,\ g : Y \to X$ satisfy $g \circ f = \mathrm{id}_X$. Then f is injective and g is surjective.*

(ii) If $f : X \to Y$ is injective then it has a left-inverse $g : Y \to X$.

Proof. (a) Lemma A.1.5(ii). (b) Choose some $x_0 \in X$. Now define $g : Y \to X$ by

$$g(y) = \begin{cases} x & \text{if } f(x) = y \\ x_0 & \text{if } y \notin f(X) \end{cases}$$

The second case clearly poses no problem, and the first case is well defined since f is injective. It is evident that $g \circ f = \mathrm{id}_X$. ■

Notice that in the above proof, we had to make a single choice, that of x_0, for which no Axiom of Choice is needed!

Now the only remaining question is whether every surjective map $f : X \to Y$ has a right-inverse or section $g : Y \to X$. When Y is infinite, this involves infinitely many choices! In fact, we have the following

[3] If this remark seems superfluous, I know otherwise from experience with an exam where less than 10% of the students got this unambiguously right, after ≥ 2.5 years of study!

Theorem A.3.4 *The following statements are equivalent:*

(i) Axiom of Choice.

(ii) Whenever $X \neq \emptyset$, there is a map $h : P(X)\backslash\{\emptyset\} \to X$ such that $h(S) \in S$ for every $S \in P(X)\backslash\{\emptyset\} \to X$. (I.e., h assigns to every non-empty subset S of X a point in S.)

(iii) For every surjective $f : X \to Y$ there is a map $g : Y \to X$ such that $f \circ g = \mathrm{id}_Y$.

Proof. (i)⇒(ii) Take $I = P(X)\backslash\{\emptyset\}$ and $X_S = S$ for every $S \in I$. Now the Axiom of Choice gives a map that assigns to every $S \in I$ (thus $\emptyset \neq S \subseteq X$) an element of $X_S = S$.

(ii)⇒(iii) Since (ii) holds, we have a function $h : P(X)\backslash\{\emptyset\} \to X$ such that $h(Y) \in Y$ for $\emptyset \neq Y \subseteq X$. Now, surjectivity of f means that for every $y \in Y$ we have $K_y := f^{-1}(y) \neq \emptyset$. Thus each K_y is a non-empty subset of X. Now we define $g : Y \to X$ by $g(y) = h(K_y)$. By definition, $g(y)$ is in X and satisfies $f(g(y)) = y$, thus g is a right inverse for f.

(iii)⇒(i). Let $X_i \neq \emptyset$ $\forall i \in I$. By Lemma A.2.3.(v), the map $b : \bigoplus_i X_i \to I$ is surjective. Thus by (iii) it has a section, i.e. a function $s : I \to \bigoplus_i X_i$ such that $b(s(i)) = i$ $\forall i \in I$. But this means that $s(i) \in X_i$ $\forall i \in I$. Thus s is an element of $\prod_i X_i$, proving AC. ■

Remark A.3.5 In the literature, one can find all three statements above called the Axiom of Choice. In view of the theorem, this can cause no problem. □

A.3.2 Weak versions of the Axiom of Choice

For some purposes, the following weaker axiom is sufficient. Some authors find that acceptable, but not AC. (This is an attitude that the present author fails to understand.)

Definition A.3.6 *The Axiom of Countable Choice (ACC) coincides with AC, except that I must be countable.*

It is easy to prove that ACC is equivalent to the existence of sections for surjective maps $f : X \to Y$ with countable Y. There is no ACC-analogue for (ii) in Theorem A.3.4 since $P(X)$ is either finite or uncountable.

We will encounter another weakened version of AC:

Definition A.3.7 *The Axiom of Countable Dependent Choice (ACDC) is the following statement: Let X be a set and $R \subseteq X \times X$ a relation such that for every $x \in X$ there exists $y \in X$ with xRy. (I.e. R is left-total.) Then for every $x_1 \in X$ there exist $\{x_n\}_{n\geq 2}$ such that $x_n R x_{n+1}$ for all $n \in \mathbb{N}$.*

Exercise A.3.8 (i) Prove the implications AC⇒ACDC⇒ACC.

(ii) Can you prove ACC⇒ACDC?

Remark A.3.9 What we called the Axiom of Countable Dependent Choice (ACDC) actually differs from what most authors mean by this term. The standard definition, which we call ACDC', is the following statement: If X is a non-empty set and $R \subseteq X \times X$ is a relation such that for every $x \in X$ there exists $y \in X$ with xRy, then there exists a sequence $\{x_n\}_{n\in\mathbb{N}}$ such that $x_n R x_{n+1}$ for all $n \in \mathbb{N}$. Thus there is no restriction on the beginning x_1 of the sequence. It is trivial to see that ACDC⇒ACDC'. The converse requires some work: □

Exercise A.3.10 Prove that ACDC'⇒ACDC. Hint: The proof involves constructing new pair (X', R'), such that ACDC' for (X', R') implies ACDC for (X, R).

A.3.3 Zorn's Lemma

The material concerning Zorn's lemma in this subsection is needed for the proof of Tychonov's theorem in Section 7.5, but also in some other places. In fact, it plays an important rôle in many parts of mathematics. We begin with a heuristic discussion.

Often in mathematics, one wants to prove the existence of a set that is maximal w.r.t. a certain property. In favorable cases, this can be done by taking the union of all sets with that property. For example, the interior Y^0 of $Y \subseteq X$, to wit the maximal open set contained in Y, can be obtained as the union of all open U with $U \subseteq Y$. This works since the union of an arbitrary number of open sets is open. However, the property that the maximal object is supposed to have is not always preserved under arbitrary unions. Consider the problem of finding a basis B in a vector space. While a basis is the same as a maximal linearly independent subset, it is clear that B cannot be obtained as the union of all finite linearly independent subsets, since a union of linearly independent sets has no reason to be linearly independent. On the other hand, the union over a *totally ordered* family of linearly independent sets actually is linearly independent, and this is the exactly the situation for which Zorn's lemma was proved.

Recall the definition of partial orders given as part of Definition A.1.8. Here we write $\leq$ instead of R.

Definition A.3.11 *A (partial) order $\leq$ where for all $x, y \in X$ we have $x \leq y$ or $y \leq x$ is called a total order or linear order.*

A partially (totally) ordered set is a pair $(X, \leq)$, where X is a set and $\leq$ is a partial (total) ordering on X. A subset $S \subseteq X$ is understood to be ordered by the restriction of $\leq$. Partially (totally) ordered sets are often called posets (respectively, chains). In particular, a chain in a poset is a totally ordered subset.

Definition A.3.12 *If $(X, \leq)$ is a partially ordered set and $S \subseteq X$, then $y \in X$ is called an upper bound for S if $x \leq y$ for all $x \in S$. (An upper bound for S need not satisfy $y \in S$, but if it does it is called greatest element of S. One easily sees that greatest elements are unique.)*

If $(X, \leq)$ is a partially ordered set, $m \in X$ is called maximal if $m \leq x \in X$ implies $x = m$.

Lower bounds, smallest and minimal elements are defined analogously.

Remark A.3.13 A greatest element of a poset X also is a maximal element. If X is totally ordered, a maximal element also is (the unique) greatest element, but this is not true for general posets. Maximality then just means that there are no truly bigger elements. And of course, an upper bound for a subset $S \subseteq X$ need not be a maximal element of X. □

Lemma A.3.14 (Zorn (1935)) [4] *Let $(X, \leq)$ be a partially ordered set in which every chain has an upper bound. Then X has a maximal element.*

Assuming Zorn's Lemma, we can finish the discussion of bases in vector spaces.

Proposition A.3.15 *Every vector space has a basis.*

Proof. Let V be a vector space and define

$$B = \{S \subseteq V \mid \text{the elements of } S \text{ are linearly independent}\}.$$

[4] Max August Zorn (1906-1993), German mathematician

The set B is partially ordered by inclusion $\subseteq$. We claim that every chain in $(B, \subseteq)$ has a maximal element: Just take the union $\widehat{S}$ of all sets in the chain. Since any finite subset of the union over a chain of sets is contained in some element of the chain, every finite subset of $\widehat{S}$ is linearly independent. Thus $\widehat{S}$ is in B and clearly is an upper bound of the chain. Thus the assumption of Zorn's Lemma is satisfied and $(B, \subseteq)$ has a maximal element M. We claim that M is a basis for V: If this was false, we could find a $v \in V$ not contained in the span of M. But then $M \cup \{v\}$ would be a linearly independent set strictly larger than M, contradicting the maximality of M. ■

Using Zorn's Lemma in order to prove existence of certain maximal structures typically is quite easy: The choice of the partially ordered set $(X, \leq)$ usually is obvious, and proving that every chain has an upper bound easy. In some cases, the existence of a maximal element asserted by Zorn's lemma is exactly what one wants to show, as in the claim that every ideal in a ring is contained in a maximal ideal. In other cases another short argument is required, as in Proposition A.3.15 or in the following:

Proposition A.3.16 *Zorn's Lemma implies the Axiom of Choice.*

Proof. Let $X_i \neq \emptyset \ \forall i \in I$. A partial choice function is a pair (J, f), where $J \subseteq I$ and $f : J \to \bigcup_k X_k$ satisfies $f(j) \in X_j \ \forall j \in J$. The set of partial choice functions is partially ordered as follows:

$$(J, f) \leq (J', f') \ :\Leftrightarrow \ J \subseteq J' \ \text{ and } \ f' \restriction J = f.$$

To see that every chain of partial choice functions has an upper bound, let $\widehat{J}$ be the union of all the J's in the chain and define $\widehat{f} : \widehat{J} \to \bigcup_k X_k$ by saying $\widehat{f}(i) = f(i)$ for any member (J, f) of the chain such that $i \in J$. (This is well defined is clear in view or the definition of $\leq$.) Clearly $(\widehat{J}, \widehat{f})$ is an upper bound for the chain. Thus by Zorn's lemma, there is a partial choice function (M, f) that is maximal. We claim that $M = I$, such that f is a choice function as required by the Axiom of Choice. If we had $M \subsetneq I$, we could pick a $j' \in I \backslash M$ and an $x \in X_{j'}$ and extend $\widehat{f}$ to $M \cup \{j'\}$ by mapping j' to x. But this would contradict the fact that (M, f) is a maximal element in the set of partial choice functions. ■

The converse implication AC $\Rightarrow$ Zorn is also true, but the proof is somewhat trickier and requires the introduction of some further terminology. It is given in the following (optional) subsection.

Remark A.3.17 1. The second application of Zorn's lemma shows that the discussion at the beginning of the section was somewhat simplistic: The elements of the partially ordered set to which Zorn's lemma is applied need not be subsets of some big set X. Often they are families of subsets of X (thus again subsets of $P(X)$), or something more general.

2. Apart from Zorn's Lemma, the Axiom of Choice is equivalent to many other statements, e.g. the Maximal Chain Principle, the Well-Ordering Principle and the Teichmüller-Tukey Lemma. (For the first two, cf. Section A.3.5.) It is quite misleading that these equivalent statements are called Axiom, Principle, Lemma, respectively, since they are logically on exactly the same footing. However, it seems natural to consider the three elementary statements considered in Theorem A.3.4 as more 'fundamental', in particular the desirable and 'intuitively obvious' ones "Every product of non-empty sets is non-empty" and "Every surjection admits a section".

3. The equivalent purely set (and order) theoretic statements listed above have many applications in algebra, general topology and analysis. Some of the facts deduced from the axiom of choice actually also imply it, e.g. the fact that every vector space has a basis, cf. [21], Krull's "every commutative unital ring has a maximal ideal", cf. [120], and Tychonov's theorem ("every product of compact topological spaces is compact"), cf. Theorem 7.5.12. □

A.3.4 Proof of AC $\Rightarrow$ Zorn

We will follow [156]; for alternative approaches cf. e.g. [103, 189].

Definition A.3.18 *A poset $(X, \leq)$ is called* <u>*well-ordered*</u> *if every non-empty subset $S \subseteq X$ has a smallest element.*

Remark A.3.19 1. A well-ordered set is totally ordered: Consider its subsets of the form $\{x, y\}$.

2. $\mathbb{N}$ is well-ordered, but $\mathbb{Z}$ is not: $\{\dots, -2, -1, 0\}$ has no smallest element. Similarly, $\mathbb{Z}, \mathbb{Q}, \mathbb{R}$ are not well-ordered, but also $\mathbb{R}_{\geq 0}$ is not: An open subset (a, b) has a lower bound (namely a), but no smallest element.

3. For more on well-orders cf. the next section. □

If $(X, \leq)$ is a poset, $C \subseteq X$ a chain and $x \in C$ we define

$$P(C, x) = \{y \in C \mid y < x\}.$$

Subsets of this form are called initial segments in C.

Proof of AC $\Rightarrow$ Zorn. Let $(X, \leq)$ be a poset in which every chain has an upper bound. In order to arrive at a contradiction, we assume that X has no maximal element, i.e. for every $x \in X$ there is a $y \in X$ such that $y > x$ (i.e. $x \leq y$ and $x \neq y$). If C is a chain in X, the assumptions imply that the set $S(C) = \{y \in X \mid y > x \ \forall x \in C\}$ of *strict upper bounds* of C is non-empty. By AC, there is a function f that assigns to every chain $C \subseteq X$ a strict upper bound $f(C) \in S(C)$.

With these preparations in place, a subset A of X is called *conforming* if $(X, \leq)$ is well-ordered and $x = f(P(A, x))$ for every $x \in A$.

Lemma A.3.20 *Under the above assumptions, let $A \neq B$ be conforming subsets of X. Then one of these subsets is an initial segment of the other.*

Proof. We assume that $A \backslash B \neq \emptyset$. (Otherwise we have $A \subseteq B$ and are done.) Let x be the smallest element of $A \backslash B$, which exists by the well-orderedness. Then $P(A, x) \subseteq B$. We claim that $P(A, x) = B$. To obtain a contradiction, assume that $B \backslash P(A, x) \neq \emptyset$, and let y be the smallest element of $B \backslash P(A, x)$. Given any element $u \in P(B, y)$ and any element $v \in A$ such that $v < u$, it is clear that $v \in P(B, y)$. Therefore if z is the smallest member of $A \backslash P(B, y)$, we have $P(A, z) = P(B, y)$. Note that $z \leq x$. But since

$$z = f(P(A, z)) = f(P(B, y)) = y,$$

and since $y \in B$, we cannot have $z = x$. Therefore $z < x$, and we conclude that $y = z \in P(A, x)$, contradicting the choice of y. ■

Using the property of comparability of conforming subsets that we have just proved, we now observe that if A is a conforming subset of X and $x \in A$, then whenever $y < x$, either $y \in A$ or y does not belong to any conforming set. It now follows easily that the union U of all the conforming subsets of X is conforming, and we deduce from this fact that if $x = f(U)$, then the set $U \cup \{x\}$ is conforming. Therefore, $x \in U$, contradicting the fact that x is a strict upper bound of U. ■

A.3.5 Some other statements equivalent to AC and Zorn's Lemma

There are several other statements that are equivalent to the Axiom of Choice and Zorn's Lemma that can be stated without introducing further terminology.

Definition A.3.21 *Hausdorff's Maximal Chain Principle (MCP): Every poset $(X, \leq)$ contains a maximal chain, i.e. a chain that is not properly contained in another one.*

Proposition A.3.22 *MCP $\Leftrightarrow$ ZL.*

Proof. $\Rightarrow$: Let $(X, \leq)$ satisfy the hypothesis in ZL, i.e. every totally ordered subset $S \subseteq X$ has an upper bound. By MCP there is a maximal totally ordered subset M. By assumption, the latter has an upper bound $y \in X$. We claim that y is a maximal element of X: Let $z \in X$ such that $y \leq z$. If $z \notin M$ then $M \cup \{z\}$ is a totally ordered set properly containing M. This contradicts the maximality of M, and therefore we must have $z \in M$. But since y is an upper bound for M, we also have $z \leq y$ and therefore $z = y$. Thus y is maximal.

$\Leftarrow$ Given an ordered set $(X, \leq)$, let Y be the set of totally ordered subsets of X, ordered by inclusion. If $\mathcal{C}$ is a chain in the poset $(Y, \subseteq)$, $\bigcup \mathcal{C}$ is a subset of X. If $x, y \in \bigcup \mathcal{C}$, then by the total order of $\mathcal{C}$ there is a $Y \in \mathcal{C}$ such that $x, y \in Y$, implying that x and y are comparable. Thus $\bigcup \mathcal{C}$ is totally ordered, thus in Y and it clearly is an upper bound for $\mathcal{C}$. Now Zorn's lemma immediately gives a maximal element. ■

In view of the above, an alternative if less direct proof of AC $\Rightarrow$ Zorn can be obtained by proving AC $\Rightarrow$ MCP, as is done in [28, App. B] and [201, Appendix].

There is a closely related equivalent statement:

Definition A.3.23 *MCP': Every chain in a poset is contained in a maximal chain.*

Clearly MCP' implies MCP, and proving MCP' from ZL works just as the proof of MCP, except that one applies the ZL to the set of chains containing the given one.

The other statement we consider uses the notion of well (thus also totally) ordered sets:

Definition A.3.24 *Well-Ordering Principle (WOP): Every set X admits a partial order $\leq$ such that $(X, \leq)$ is well-ordered.*

Remark A.3.25 1. The order $\leq$ provided by the WOP typically has nothing to do with a previously given 'natural' order on X.

2. The WOP is useful since well-orderings are the basis for 'transfinite induction', which generalizes ordinary mathematical induction (over $\mathbb{N}$), cf. Section A.3.6. □

Theorem A.3.26 *WOP is equivalent to AC and ZL.*

Proof. It is trivial to prove WOP$\Rightarrow$AC, version (ii): Give X a well-ordering and define the choice function $f : P_0(X) \to X$ by mapping $S \in P_0(X)$ to its smallest element.

Also proving ZL$\Rightarrow$WOP is not hard: Just apply ZL to the set of well-ordered subsets of X, cf. [189, Section 1.1] or [28, App. B]. (Together with WOP$\Rightarrow$AC, this provides an alternative proof of ZL$\Rightarrow$AC.) ■

A.3.6 Transfinite induction. The Δ-system lemma

Recall that for an ordered set $(X, \leq)$ and $y \in X$ we define $L(y) \equiv \{z \in X \mid z < y\}$.

An important application of well-orderings is the following:

Lemma A.3.27 (Transfinite Induction) *Let $(X, \leq)$ we a well-ordered set. If $Y \subseteq X$ is such that $L(y) \subseteq Y$ implies $y \in Y$ then $Y = X$.*

Proof. Assume $Y \neq X$, thus $X \backslash Y \neq \emptyset$. Let y be the smallest element of $X \backslash Y$. By definition of y we have $z \in Y$ for all $z < y$, so that the assumption gives $y \in Y$, which is a contradiction. ■

Proposition A.3.28 (Transfinite recursion) *Let a well-ordered set $(X, \leq)$, a set Y and a function $f : \bigoplus_{x \in X} Y^{L(x)} \to Y$ [thus f assigns an element of Y to every Y-valued function defined on some $L(x)$, where $x \in X$] be given.*

Then there is a unique function $F : X \to Y$ such that $F(x) = f(F \restriction L(x))\ \forall x \in X$.

Proof. Uniqueness: Let $F, G : X \to Y$ be functions satisfying the assertion. Let $Z = \{x \in X \mid F(x) = G(x)\}$. Assume $L(x) \subseteq Z$, thus $F \restriction L(x) = G \restriction L(x)$. Then

$$F(x) = f(F \restriction L(x)) = f(G \restriction L(x)) = G(x),$$

thus $x \in Z$. Now transfinite induction (Lemma A.3.27) gives $Z = X$, thus $F = G$.

Also the proof of existence of F uses transfinite induction, but it is more work. An elementary proof (avoiding ordinal numbers) is given in [103, Sect. 18]. ■

Lemma A.3.29 (Δ-system lemma) *If $\mathcal{A}$ is an uncountable family of finite sets then there exists an uncountable subfamily $\mathcal{A}_0 \subseteq \mathcal{A}$ and a finite set A such that $X \cap Y = A$ for all $X, Y \in \mathcal{A}_0$ with $X \neq Y$.*

Proof. Let $\mathcal{A}_n = \{A \in \mathcal{A} \mid \#A = n\}$. Since $\mathcal{A} = \bigcup_n \mathcal{A}_n$ is uncountable, there must be an $n \in \mathbb{N}$ such that $\mathcal{A}_n$ is uncountable. Replacing $\mathcal{A}$ by such an $\mathcal{A}_n$, we may assume that $\#A = n$ for all $A \in \mathcal{A}$. We will prove the lemma by induction over n. Since $n = 0$ would mean that the elements of $\mathcal{A}$ are empty, thus $\#\mathcal{A} \leq 1$, we begin with $n = 1$. A family of singletons clearly satisfies $X \neq Y \Rightarrow X \cap Y = A$ with $A = \emptyset$. Assume that the lemma is true for families of sets with cardinality $< m$. Let $\mathcal{A}$ be an uncountable family of sets that have cardinality m. Now there are two cases:

Case (i): There is a $B \in \mathcal{A}$ that intersects uncountably many $A \in \mathcal{A}$. Since B is finite, there is an $x \in B$ that is contained in uncountably many $A \in \mathcal{A}$. Defining $\mathcal{A}' = \{A \backslash \{x\} \mid x \in A \in \mathcal{A}\}$, each element of $\mathcal{A}'$ has cardinality $m - 1$, thus by the induction hypothesis there is an uncountable subfamily $\mathcal{A}'' \subseteq \mathcal{A}'$ and finite set A' such that $X, Y \in \mathcal{A}''$, $X \neq Y \Rightarrow X \cap Y = A'$. Now $\mathcal{A}_0 = \{A \mid x \in A \in \mathcal{A}\}$ is an uncountable subfamily of $\mathcal{A}$, and $X, Y \in \mathcal{A}_0$, $X \neq Y \Rightarrow X \cap Y = A' \cup \{x\} =: A$, as desired.

Case (ii): For every $B \in \mathcal{A}$ we have $\#\{A \in \mathcal{A} \mid A \cap B \neq \emptyset\} \leq \aleph_0$. We use the fact from set theory that the first uncountable cardinal number $\aleph_1$ is a well ordered uncountable set $(I, \leq)$ such that $L(i)$ is countable for every $i \in I$. Our aim is to use transfinite recursion to define a map $F : I \to \mathcal{A}$, which we denote $i \mapsto A_i$ for readability. An element of $\mathcal{A}^{L(i)}$, where $i \in I$, is the same as a a family $\{A_j \in \mathcal{A} \mid j < i\}$. The hypothesis of Case (ii) together with the countability of $L(i)$ for each $i \in I$ implies that there are at most countably many $B \in \mathcal{A}$ that have non-trivial intersection with some A_j where $j < i$. Since $\mathcal{A}$ is uncountable, there exists a $B \in \mathcal{A}$ such that $B \cap \bigcup_{j<i} A_j = \emptyset$. Using the axiom of choice, we therefore can define a map $f : \bigoplus_{i \in I} \mathcal{A}^{L(i)} \to \mathcal{A}$ such that $f(\{A_j \in \mathcal{A} \mid j < i\}) \cap \bigcup_{j<i} A_j = \emptyset$. Now Proposition A.3.28 gives a map $F : I \to \mathcal{A}$ such

that $F(i) = f(F \restriction L(i))$. By construction, we have $A_i = F(i) = f(F \restriction L(i)) \subseteq \bigcup \mathcal{A} \setminus \bigcup_{j<i} A_j$, thus $A_i \cap A_j = \emptyset$ whenever $j < i$, thus whenever $i \neq j$. Therefore $\mathcal{A}_0 = \{A_i \mid i \in I\}$ is an uncountable family of mutually disjoint sets, so that the lemma holds (with $A = \emptyset$). ■

A.4 Lattices. Boolean algebras

Definition A.4.1 *A lattice is a poset (partially ordered set) $(L, \leq)$ in which each two-element subset has an infimum and a supremum. If $a, b \in L$, we write $a \vee b = \sup\{a, b\}$, $a \wedge b = \inf\{a, b\}$.*

Comments:

- Equivalently, we could require that every finite subset has an infimum and a supremum.
- Since the infimum and supremum (of any subset, provided they exist) are unique, we can consider $(a, b) \mapsto a \vee b$ and $(a, b) \mapsto a \wedge b$ as binary operations on L.
- It is obvious that the operations $\vee, \wedge$ are commutative and associative and satisfy idempotency: $a \vee a = a = a \wedge a$.
- We have $a \wedge b \leq a \leq a \vee b$ and therefore for all a, b the absorption identities

$$a \vee (a \wedge b) = a = a \wedge (a \vee b) \tag{A.2}$$

  hold.
- We have $a \leq b \;\Leftrightarrow\; a \wedge b = a \;\Leftrightarrow\; a \vee b = b$.

Definition A.4.2 *An algebraic lattice is a triple $(L, \vee, \wedge)$, where L is a set and $\vee, \wedge$ are binary operations $L \times L \to L$ satisfying associativity, commutativity and absorption (A.2).*

Lemma A.4.3 *Let $(L, \vee, \wedge)$ be an algebraic lattice. Then:*

(i) Idempotency holds: $a \vee a = a = a \wedge a$.

(ii) We have $a \wedge b = a \;\Leftrightarrow\; a \vee b = b$.

(iii) Defining $a \leq b :\Leftrightarrow a \wedge b = a$, the relation $\leq$ is a partial order on L.

(iv) The partially ordered set $(L, \leq)$ is a lattice, and $\sup_{\leq}\{a, b\} = a \vee b$, $\inf_{\leq}\{a, b\} = a \wedge b$.

Proof.

■

Thus for every set L, there is a bijection between lattice structures $\leq$ on L and algebraic structures $(L, \vee, \wedge)$ satisfying (a),(b).

Definition A.4.4 *A lattice $(L, \leq)$ is called bounded if $(L, \leq)$ has a smallest and a largest element, called 0 and 1 respectively.*

It is clear that $0, 1$ are unique, if they exist and that we have

$$a \vee 0 = a = a \wedge 1, \qquad a \wedge 0 = 0, \quad a \vee 1 = 1$$

for all $a \in L$.

Definition A.4.5 *A lattice $(L, \leq)$ is called distributive if for all $a, b, c \in L$ we have*

$$a \vee (b \wedge c) = (a \vee b) \wedge (a \vee c), \qquad a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c).$$

Definition A.4.6 *A Boolean algebra is a bounded distributive lattice $(L, \leq, 0, 1)$ coming with a complementation, i.e. a map $L \to L,\ a \mapsto \neg a$ such that*

****

A.5 Basic definitions on categories

Example A.5.1 Let $\mathcal{G}$ be the set of all groups. (There is a set-theoretic problem here, related to Russel's paradox. This problem can be avoided in several ways, e.g. by taking $\mathcal{G}$ to be the *class* of all groups. But this discussion is only marginally relevant here, and we ignore it.) If $G, H \in \mathcal{G}$, thus G, H are groups, we denote by $\mathrm{Hom}(G, H)$ the set of all group homomorphisms from G to H. (This really is a set, since $\mathrm{Hom}(G, H) \subseteq \mathrm{Fun}(G, H) \subseteq P(G \times H)$.) Instead of $\alpha \in \mathrm{Hom}(G, H)$ we also write $\alpha : G \to H$. For every $G \in \mathcal{G}$, we have a distinguished morphism $\mathrm{id}_G \in \mathrm{Hom}(G, G)$, namely the identity map. If $\alpha \in \mathrm{Hom}(G, H), \beta \in \mathrm{Hom}(H, K)$ then the composite map $\beta \circ \alpha$ is an element, clearly uniquely defined, in $\mathrm{Hom}(G, K)$. It is clear that for $\alpha : G \to H$ we have $\mathrm{id}_H \circ \alpha = \alpha = \alpha \circ \mathrm{id}_G$, thus the id_G are identities for the composition $\circ$. Finally, if $G \stackrel{\alpha}{\to} H \stackrel{\beta}{\to} K \stackrel{\gamma}{\to} L$ then we have the associativity $\gamma \circ (\beta \circ \alpha) = (\gamma \circ \beta) \circ \alpha$ which holds for any composition of functions. Thus the operation $\circ$ is associative.

Now observe that if we pick any field k and replace $\mathcal{G}$ by the 'set' $\mathcal{V}_k$ of k-vector spaces and by $\mathrm{Hom}(V, W)$ we mean k-linear maps from V to W, the rest of the discussion goes through unchanged. It should be clear by now that a similar situation arises if we talk about all sets with all functions as morphisms, or about topological spaces together with continuous maps. This situation arises virtually everywhere in mathematics, which leads to the following definition. □

Definition A.5.2 *A category $\mathcal{C}$ consists of a class $\mathrm{Obj}\,\mathcal{C}$ of objects and, for every pair $X, Y \in \mathrm{Obj}\,\mathcal{C}$ of objects, a set $\mathrm{Hom}_{\mathcal{C}}(X, Y)$ of morphisms from X to Y. If $s \in \mathrm{Hom}_{\mathcal{C}}(X, Y)$, we also write $s : X \to Y$ or $X \stackrel{s}{\to} Y$. For every $X \in \mathrm{Obj}\,\mathcal{C}$, there is a distinguished morphism $\mathrm{id}_X : X \to X$. Whenever $X \stackrel{f}{\to} Y \stackrel{g}{\to} Z$, there is a morphism $g \circ f : X \to Z$, the composite of f and g. These data satisfy:*

- *If $X \stackrel{f}{\to} Y \stackrel{g}{\to} Z \stackrel{h}{\to} W$, we have associativity: $h \circ (g \circ f) = (h \circ g) \circ f$.*
- *If $f : X \to Y$ then $f \circ \mathrm{id}_X = f = \mathrm{id}_Y \circ f$.*

We write $\mathrm{End}_{\mathcal{C}}(X) := \mathrm{Hom}_{\mathcal{C}}(X, X)$. When there is no risk of confusion, we write $\mathrm{Hom}(X, Y)$ and $\mathrm{End}(X)$ instead of $\mathrm{Hom}_{\mathcal{C}}(X, Y)$ and $\mathrm{End}_{\mathcal{C}}(X)$.

Remark A.5.3 1. Categories (and functors and natural transformations, cf. below) were invented (discovered?) in the years 1942-43 for the purposes of topology and homological algebra and formally defined in 1945, cf. [67]. They are now an indispensable part of the mathematical toolkit, without which a large part of modern (1945-) mathematics would be unthinkable. (Some people even think that categories should replace sets as foundation of mathematics, but I wouldn't go that far.)

2. Usually we will use upper-case letters for objects and lower-case letters for morphisms. We will mostly write $X \in \mathcal{C}$ instead of $X \in \mathrm{Obj}\,\mathcal{C}$. □

Definition A.5.4 *A morphism $f : X \to Y$ is called* isomorphism *if it admits an inverse, i.e. a morphism $g : Y \to X$ such that $g \circ f = \mathrm{id}_X$ and $f \circ g = \mathrm{id}_Y$. (Such a g is unique.)*

Two objects X, Y are called isomorphic*, $X \cong Y$, if an isomorphism between them exists.*

Definition A.5.5 *A morphism $f : X \to Y$ is called ...*

- monic *(or* monomorphism*) if $f \circ g = f \circ h$, where $g, h : Z \to X$, implies $g = h$.*
- epic *(or* epimorphism*) if $g \circ g = h \circ f$, where $g, h : Y \to Z$, implies $g = h$.*

Exercise A.5.6 (i) Prove that every isomorphism is both monic and epic.

(ii) For a morphism f in the category $\mathcal{SET}$ of sets, prove: f is monic $\Leftrightarrow$ f is injective, and f is epic $\Leftrightarrow$ f is surjective.

(iii) Conclude that in $\mathcal{SET}$ we have: f is an isomorphism $\Leftrightarrow$ f is a bijection $\Leftrightarrow$ f is monic and epic. (Warning: There are categories where f monic+epi does not imply that f is an isomorphism!)

Definition A.5.7 *If $\mathcal{C}$ is a category, an object $X \in \mathcal{C}$ is called* initial *(resp.* terminal*) if for every $Y \in \mathcal{C}$, there is exactly one morphism $f : X \to Y$ (resp. $f : Y \to X$).*

Example A.5.8 If $\mathcal{C} = \mathrm{Set}$ then the empty set $\emptyset$ is initial and every one-point set (=singleton) $\{x\}$ is terminal. In $\mathcal{C} = \mathrm{Grp}$, the trivial group $\{e\}$ is both initial and terminal. □

Exercise A.5.9 Prove that any two initial objects are isomorphic, and any two terminal objects are isomorphic.

Definition A.5.10 *A* subcategory *of a category $\mathcal{C}$ is a category $\mathcal{D}$ such that $\mathrm{Obj}\,\mathcal{D} \subseteq \mathrm{Obj}\,\mathcal{C}$ and, $\mathrm{Hom}_{\mathcal{D}}(X, Y) \subseteq \mathrm{Hom}_{\mathcal{C}}(X, Y)$ for all $X, Y \in \mathrm{Obj}\,\mathcal{D}$. The composition of morphisms in $\mathcal{D}$ is inherited from $\mathcal{C}$, and the unit morphisms of $\mathcal{D}$ are those of $\mathcal{C}$.*

A subcategory $\mathcal{D} \subseteq \mathcal{C}$ is called full *if $\mathrm{Hom}_{\mathcal{D}}(X, Y) = \mathrm{Hom}_{\mathcal{C}}(X, Y)$ for all $X, Y \in \mathrm{Obj}\,\mathcal{D}$.*

Definition A.5.11 *Let $\mathcal{C}$ be a category. The* opposite category *$\mathcal{C}^{\mathrm{op}}$ is defined by*

$$\mathrm{Obj}\,\mathcal{C}^{\mathrm{op}} = \mathrm{Obj}\,\mathcal{C}, \qquad \mathrm{Hom}_{\mathcal{C}^{\mathrm{op}}}(X^{\mathrm{op}}, Y^{\mathrm{op}}) = \mathrm{Hom}_{\mathcal{C}}(Y, X).$$

An object $X \in \mathrm{Obj}\,\mathcal{C}$ is written X^{op} when it is considered as an object of $\mathcal{C}^{\mathrm{op}}$. When $s \in \mathrm{Hom}_{\mathcal{C}}(X, Y)$ is interpreted as an element of $\mathrm{Hom}_{\mathcal{C}^{\mathrm{op}}}(Y^{\mathrm{op}}, X^{\mathrm{op}})$ it is written s^{op}. If $s \in \mathrm{Hom}_{\mathcal{C}}(X, Y) = \mathrm{Hom}_{\mathcal{C}^{\mathrm{op}}}(Y^{\mathrm{op}}, X^{\mathrm{op}})$ and $t \in \mathrm{Hom}_{\mathcal{C}}(Y, Z) = \mathrm{Hom}_{\mathcal{C}^{\mathrm{op}}}(Z^{\mathrm{op}}, Y^{\mathrm{op}})$ then the composite $s^{\mathrm{op}} \circ t^{\mathrm{op}}$ is given by $(t \circ s)^{\mathrm{op}}$. The identity morphisms are given by $\mathrm{id}_{X^{\mathrm{op}}} = (\mathrm{id}_X)^{\mathrm{op}}$.

Categories are just a type of algebraic structure like groups or vector spaces. Since there is a notion of homomorphism of groups and of vector spaces, it is only natural to ask for a notion of homomorphisms between categories. Since a homomorphism of algebraic objects should preserve all the structures on that object (composition, units, etc.), we are naturally led to the following (except for the name of the notion):

Definition A.5.12 *If $\mathcal{C}, \mathcal{D}$ are categories, a* functor *from $\mathcal{C}$ to $\mathcal{D}$, denoted $F : \mathcal{C} \to \mathcal{D}$, consists of a map $F : \mathrm{Obj}\,\mathcal{C} \to \mathrm{Obj}\,\mathcal{D}$ and, for every pair $X, Y \in \mathrm{Obj}\,\mathcal{C}$, of a map $F : \mathrm{Hom}_{\mathcal{C}}(X, Y) \to \mathrm{Hom}_{\mathcal{D}}(F(X), F(Y))$. These maps satisfy*

(i) For every $X \in \mathrm{Obj}\,\mathcal{C}$, we have $F(\mathrm{id}_X) = \mathrm{id}_{F(X)}$.

(ii) Given $X \xrightarrow{f} Y \xrightarrow{g} Z$*, we have* $F(g \circ f) = F(g) \circ F(f)$*.*

The class of all functors $\mathcal{C} \to \mathcal{D}$ *is denoted* $\mathrm{Fun}(\mathcal{C}, \mathcal{D})$*.*

Lemma A.5.13 *There is a category* Cat *whose objects are categories and such that* $\mathrm{Hom}_{\mathrm{Cat}}(\mathcal{C}, \mathcal{D})$ *is given by the functors* $F : \mathcal{C} \to \mathcal{D}$*.*

Proof. It is clear that functors $\mathcal{C} \xrightarrow{F} \mathcal{D} \xrightarrow{G} \mathcal{E}$ can be composed to give a functor $G \circ F : \mathcal{C} \to \mathcal{E}$, and that this, together with identity functors $\mathrm{id}_{\mathcal{C}}$, satisfies the axioms of a category. ■

Remark A.5.14 In order to avoid set theoretic problems, one should define Cat as the category of *small* categories, but we won't bother. □

Example A.5.15 1. There are many forgetful functors, like the functor F : Grp→Sets, which sends any group to its underlying set, forgetting the group structure, and every group homomorphism to the underlying set map.

2. There are the functors F : Set→Ab and G_k : Set→Vect$_k$ which associate to every set X the free abelian group which has the elements of X as generators, respectively the k-vector space that has X as basis (the "finite linear combinations of elements of X with coefficients in k"). Every set map $f : X \to Y$ induces a homomorphism $F(f) : F(X) \to F(Y)$ of abelian groups, respectively a k-linear map $G_k(f) : G_k(X) \to G_k(Y)$. Both are determined by specifying that $F(f)$ sends the generator (resp. basis element) $x \in F(X)$ to the generator (basis element) $f(x) \in F(Y)$.

3. An example from topology: There is a functor T from the category of metric spaces and continuous maps to the category $\mathcal{TOP}$ of topological spaces, defined by $T((X, d)) = (X, \tau_d)$ and as the identity map on morphisms. The functor T is injective on morphisms, but not on objects (since there are equivalent metrics).

4. More interesting functors will be encountered later, their existence being the main reason why we introduce the categorical language. □

Definition A.5.16 *Let* $F : \mathcal{C} \to \mathcal{D}$ *be a functor.* F *is called*

- *faithful if the maps* $\mathrm{Hom}_{\mathcal{C}}(X, Y) \to \mathrm{Hom}_{\mathcal{D}}(F(X), F(Y))$ *are injective for all* $X, Y \in \mathcal{C}$*.*
- *full if the maps* $\mathrm{Hom}_{\mathcal{C}}(X, Y) \to \mathrm{Hom}_{\mathcal{D}}(F(X), F(Y))$ *are surjective for all* $X, Y \in \mathcal{C}$*.*
- *essentially surjective if for every* $Y \in \mathcal{D}$ *there is an* $X \in \mathcal{C}$ *such that* $F(X) \cong Y$*.*

Remark A.5.17 If $\mathcal{D} \subseteq \mathcal{C}$ is a subcategory, then obvious inclusion functor $\iota : \mathcal{D} \hookrightarrow \mathcal{C}$ is always faithful, and it is full if and only if $\mathcal{D}$ is a full subcategory. □

In view of Lemma A.5.13 and Definition A.5.4 it is clear what an isomorphism of categories is, but we state it explicitly anyway:

Definition A.5.18 *A functor* $F : \mathcal{C} \to \mathcal{D}$ *is an isomorphism of categories if there exists a functor* $G : \mathcal{D} \to \mathcal{C}$ *such that* $G \circ F = \mathrm{id}_{\mathcal{C}}$ *and* $F \circ G = \mathrm{id}_{\mathcal{D}}$*. If an isomorphism* $\mathcal{C} \to \mathcal{D}$ *exists, we write* $\mathcal{C} \cong \mathcal{D}$*.*

Definition A.5.19 *If* $\mathcal{C}, \mathcal{D}$ *are categories, a contravariant functor* $F : \mathcal{C} \to \mathcal{D}$ *is a functor* $F : \mathcal{C} \to \mathcal{D}^{\mathrm{op}}$ *(or, equivalently,* $F : \mathcal{C}^{\mathrm{op}} \to \mathcal{D}$*).*

For many purposes the notion of isomorphism of categories is too restrictive and therefore of limited usefulness. (E.g., compare Corollaries 13.4.12 and 13.5.6.) In order to define a generalization (as well as for many other purposes) we need a new concept:

Definition A.5.20 *Let $\mathcal{C},\mathcal{D}$ be categories and $F,G:\mathcal{C}\to\mathcal{D}$ functors. A natural transformation from F to G is a family of morphisms $\{\alpha_X : F(X)\to G(X)\}_{X\in\mathcal{C}}$ such that the diagram*

$$\begin{tikzcd} F(X) \arrow[r, "\alpha_X"] \arrow[d, "F(s)"'] & G(X) \arrow[d, "G(s)"] \\ F(Y) \arrow[r, "\alpha_Y"'] & G(Y) \end{tikzcd} \tag{A.3}$$

commutes for every $s\in \mathrm{Hom}_{\mathcal{C}}(X,Y)$. A natural isomorphism is a natural transformation α where α_X is an isomorphism for every $X\in\mathcal{C}$.

Definition A.5.21 *A functor $F:\mathcal{C}\to\mathcal{D}$ is called an equivalence of categories if there exist a functor $G:\mathcal{D}\to\mathcal{C}$ and natural isomorphisms $G\circ F\to \mathrm{id}_{\mathcal{C}}$ and $F\circ G\to\mathrm{id}_{\mathcal{D}}$.*

Two categories $\mathcal{C},\mathcal{D}$ are called equivalent, $\mathcal{C}\simeq\mathcal{D}$ if an equivalence $F:\mathcal{C}\to\mathcal{D}$ exists.

There is a very useful criterion for a functor to be an equivalence, cf. [160]:

Proposition A.5.22 *A functor $F:\mathcal{C}\to\mathcal{D}$ is an equivalence if and only if it is full, faithful, and essentially surjective.*

Proof. Assume F to be an equivalence. Then we have a functor $G:Y\to X$ and natural isomorphisms $\alpha: G\circ F\to\mathrm{id}_{\mathcal{C}}$ and $\beta: F\circ G\to \mathrm{id}_{\mathcal{D}}$. The isomorphisms $GF(X)\cong X$ for $X\in\mathcal{C}$ imply essential surjectivity of G, and similarly for F. $G\circ F$ maps $\mathrm{Hom}_{\mathcal{C}}(X,Y)$ to $\mathrm{Hom}_{\mathcal{C}}(GF(X),GF(Y))$, and for $s\in\mathrm{Hom}_{\mathcal{C}}(X,Y)$ we have $s\circ\alpha_X=\alpha_Y\circ GF(s)$. Since α_X,α_Y are isomorphisms, the map $s\mapsto GF(s)$ is a bijection. This implies that $F:\mathrm{Hom}_{\mathcal{C}}(X,Y)\to\mathrm{Hom}_{\mathcal{D}}(F(X),F(Y))$ is injective and that $G:\mathrm{Hom}_{\mathcal{D}}(F(X),F(Y))\to\mathrm{Hom}_{\mathcal{C}}(GF(X),GF(Y))$ is surjective. A similar reasoning applied to $F\circ G$ shows that $G:\mathrm{Hom}_{\mathcal{D}}(X,Y)\to\mathrm{Hom}_{\mathcal{C}}(G(X),G(Y))$ is injective and $F:\mathrm{Hom}_{\mathcal{C}}(G(X),G(Y))\to\mathrm{Hom}_{\mathcal{D}}(FG(X),FG(Y))$ is surjective. Thus F and G are faithful. Fullness now follows using essential surjectivity and the bijections $\mathrm{Hom}(X,Y)\cong\mathrm{Hom}(X',Y')$ when $X\cong X',Y\cong Y'$.

We omit the proof of the converse implication, since we will not need it. ■

Definition A.5.23 *Let $\mathcal{C},\mathcal{D}$ be categories and $F:\mathcal{C}\to\mathcal{D}$, $G:\mathcal{D}\to\mathcal{C}$ functors. We say that F is a left adjoint of G and G a right adjoint of F if the following holds: For every $X\in\mathrm{Obj}\,\mathcal{C}$, $Y\in\mathrm{Obj}\,\mathcal{D}$ there is a bijection $\alpha_{X,Y}:\mathrm{Hom}_{\mathcal{D}}(F(X),Y)\to\mathrm{Hom}_{\mathcal{C}}(X,G(Y))$ that is natural w.r.t. X,Y. I.e. if $s\in\mathrm{Hom}_{\mathcal{C}}(X',X)$, $t\in\mathrm{Hom}_{\mathcal{D}}(Y,Y')$ then the diagrams*

$$\begin{tikzcd} \mathrm{Hom}_{\mathcal{C}}(F(X),Y) \arrow[r, "\alpha_{X,Y}"] \arrow[d, "t_*"'] & \mathrm{Hom}_{\mathcal{D}}(X,G(Y)) \arrow[d, "(Gt)_*"] \\ \mathrm{Hom}_{\mathcal{C}}(F(X),Y') \arrow[r, "\alpha_{X,Y'}"'] & \mathrm{Hom}_{\mathcal{D}}(X,G(Y')) \end{tikzcd} \qquad \begin{tikzcd} \mathrm{Hom}_{\mathcal{C}}(F(X),Y) \arrow[r, "\alpha_{X,Y}"] \arrow[d, "(Fs)^*"'] & \mathrm{Hom}_{\mathcal{D}}(X,G(Y)) \arrow[d, "s^*"] \\ \mathrm{Hom}_{\mathcal{C}}(F(X'),Y) \arrow[r, "\alpha_{X',Y}"'] & \mathrm{Hom}_{\mathcal{D}}(X',G(Y)) \end{tikzcd}$$

commute. Here the vertical maps are given by

$$t_*: u\mapsto t\circ u,\quad (Gt)_*: w\mapsto G(t)\circ w,\quad F(s)^*: v\mapsto v\circ F(s),\quad s^*: k\mapsto k\circ s.$$

It is not hard to show that if $F : \mathcal{C} \to \mathcal{D}$ is an equivalence then the functor $G : \mathcal{D} \to \mathcal{C}$ showing this (cf. Definition A.5.21) is both a left and a right adjoint of F. But there are many adjoint pairs of functors that are not equivalences.

Definition A.5.24 *A full subcategory $\mathcal{D} \subseteq \mathcal{C}$ of category $\mathcal{C}$ is called reflective if the inclusion functor $\iota : \mathcal{D} \hookrightarrow \mathcal{C}$ has a left adjoint.*

Appendix B

Spectra of commutative rings. Spectral spaces

A commutative ring is called unital if it has a unit $\mathbf{1} \in R$. A prime ideal $\mathfrak{p}$ is a proper ideal such that $ab \in \mathfrak{p} \;\Rightarrow\; a \in \mathfrak{p} \vee b \in \mathfrak{p}$.

Definition/Proposition B.1 *If R is a commutative unital ring, the* <u>*prime spectrum*</u> $\mathrm{Spec}(R)$ *is the set of prime ideals in R. If I is any ideal in R, we define*

$$V(I) = \{\mathfrak{p} \in \mathrm{Spec}(R) \mid \mathfrak{p} \supseteq I\} \;\subseteq\; \mathrm{Spec}(R).$$

Then the family $\{V(I) \mid I \subseteq R \text{ an ideal}\}$ *satisfies the axioms for the closed sets in a topological space, cf. Lemma 2.4.2, and thus defines a topology τ on* $\mathrm{Spec}(R)$*, the Zariski topology*[1].

Proof. We have $V(\{0\}) = \mathrm{Spec}(R)$ and $V(R) = \emptyset$ (since no *proper* ideal contains R). Furthermore, given any family I_i of ideals, we define

$$\sum_i I_i = \left\{ \sum_i a_i \mid a_i \in I_i \ \forall i,\ \#\{i \mid a_i \neq 0\} < \infty \right\},$$

which again is an ideal. (Obviously, $\sum_k I_k \supseteq I_i \ \forall i$, and $\sum_k I_k$ is the smallest ideal containing all I_i.) Now we have

$$\bigcap_i V(I_i) = \{\mathfrak{p} \in \mathrm{Spec}(R) \mid \mathfrak{p} \supseteq I_i \ \forall i\} = V(\sum_i I_i).$$

Given ideals I_1, I_2, we define

$$I_1 \cdot I_2 = \left\{ \sum_{k=1}^{n} x_k y_k \mid x_k \in I_1,\ y_k \in I_2 \right\},$$

the ideal generated by I_1 and I_2. Clearly $I_1 \cdot I_2 \subseteq I_1 \cap I_2$. (But the inclusion can be proper.) Thus if $\mathfrak{p} \supseteq I_1$ or $\mathfrak{p} \supseteq I_2$ then $\mathfrak{p} \supseteq I_1 \cap I_2 \supseteq I_1 \cdot I_2$, and we have $V(I_1) \cup V(I_2) \subseteq V(I_1 \cdot I_2)$. In order to prove the converse inclusion $V(I_1 \cdot I_2) \subseteq V(I_1) \cup V(I_2)$, let $\mathfrak{p} \in V(I_1 \cdot I_2)$. If $\mathfrak{p} \in V(I_1)$, we are done, so assume $\mathfrak{p} \notin V(I_1)$. This means $\mathfrak{p} \not\supseteq I_1$, so that we can find $a \in I_1$ with $a \notin \mathfrak{p}$. Now, if $b \in I_2$, we

[1] The topology on $\mathrm{Spec}(R)$ generalizes the Zariski topology on k^n that we have met in Section 2.4. In principle, it should be called after Alexander Grothendieck (1928-2014) who proposed to consider the prime spectrum as a topological space. But the term Grothendieck topology is in use for something else – outside topology as understood here!

have $ab \in I_1 \cdot I_2 \subseteq \mathfrak{p}$, since $\mathfrak{p} \in V(I_1 \cdot I_2)$. Since $a \notin \mathfrak{p}$ and $\mathfrak{p}$ is prime, we have $b \in \mathfrak{p}$. This is true for all $b \in I_2$, thus $I_2 \subseteq \mathfrak{p}$, which is equivalent to $\mathfrak{p} \in V(I_2)$. ■

In order to see whether Spec(R) is T_1, we compute the closure of a point (=one-point set):

$$\overline{\{\mathfrak{p}\}} = \bigcap\{V(I) \mid \{\mathfrak{p}\} \in V(I)\} = \bigcap\{V(I) \mid \mathfrak{p} \supseteq I\} = V(\mathfrak{p}) = \{\mathfrak{q} \in \mathrm{Spec}(R) \mid \mathfrak{q} \supseteq \mathfrak{p}\}. \tag{B.1}$$

Lemma B.2 Spec(R) *is* T_0. *It is* T_1 *if and only if every prime ideal in* R *is maximal.*

Proof. By (B.1), the singleton $\{\mathfrak{p}\}$ is closed if and only if the ideal $\mathfrak{p}$ is maximal. Also, $\mathfrak{q} \in \overline{\{\mathfrak{p}\}}$ holds if and only if $\mathfrak{q} \supseteq \mathfrak{p}$. Thus if $\mathfrak{p} \neq \mathfrak{q}$ we have $\mathfrak{q} \notin \overline{\{\mathfrak{p}\}}$ or $\mathfrak{p} \notin \overline{\{\mathfrak{q}\}}$, thus $\overline{\{\mathfrak{p}\}} \neq \overline{\{\mathfrak{q}\}}$. Thus Spec$(R)$ is T_0. ■

Very few rings satisfy the restrictive condition that prime ideals be maximal. E.g., in an integral domain the zero ideal $\mathbf{0}$ is prime, and it is clear that $\overline{\{\mathbf{0}\}} = \mathrm{Spec}(R)$. And indeed, $\mathbf{0}$ is maximal if and only if R is a field, in which case Spec(R) is a one-point space.

Proposition B.3 Spec(R) *is compact. If* R *is Noetherian then* Spec(R) *is hereditarily compact.*

Proof. We first note that $V(I) = \emptyset$ holds if and only if $I = R$, since every proper ideal is contained in a maximal ideal, thus in a prime ideal. Furthermore, an ideal is proper if and only if it does not contain the unit $\mathbf{1}$. Now let $\{I_i\}$ be a family of ideals such that the family $\{V(I_i)\}$ of closed sets has the finite intersection property. Since $\bigcap_i V(I_i) = V(\sum_i I_i)$ for any family of ideals, the finite intersection property means that sum of any finite number of the ideals I_i is a proper ideal, i.e. does not contain $\mathbf{1}$. But by definition, an element of $\sum_i I_i$ is finite sum of elements of the I_i and therefore already contained in a finite sum of I_i's, which is proper as we saw. Therefore no element of $\sum_i I_i$ equals $\mathbf{1}$, thus $\sum_i I_i$ is proper, and $\bigcap_i V(I_i) \neq \emptyset$. Thus Spec$(R)$ is compact.

By Exercise 7.4.9, Spec(R) is hereditarily compact if and only if every strictly decreasing chain of closed sets is finite. So let $V(I_1) \supsetneq V(I_2) \supsetneq \cdots$. In view of $V(I_1) \cap V(I_2) = V(I_1 + I_2)$, we have $V(I_k) = V(J_k)$, where the ideals $J_k = \sum_{i=1}^k I_i$ satisfy $J_1 \subsetneq J_2 \subsetneq \cdots$. If R is Noetherian, i.e. satisfies the ascending chain condition for ideals, this chain of ideals must terminate. Thus the decreasing chain of closed subsets is finite, and Spec(R) is hereditarily compact. ■

Exercise B.4 Is Noetherianness of R necessary in order for Spec(R) to be hereditarily compact?

Definition B.5 *Let* R *be a commutative unital ring. Then the* maximal spectrum Spec$_m(R)$ *of* R *is the subspace of* Spec(R) *consisting of the closed points, i.e. the maximal ideals.*

Obviously, Spec$_m(R)$ is a T_1-space for any R. In order to make the connection with the more classical Zariski topology on affine space, take $R = k[x_1, \ldots, x_n]$, which is Noetherian. Thus Spec(R) is hereditarily compact, and so is the subspace Spec$_m(R)$. Every point in $z \in \mathbb{A}^n$ gives rise to a maximal ideal M_z, generated by $\{x_1 - z_1, x_2 - z_2, \ldots, x_n - z_n\}$, i.e. consisting of the polynomials vanishing in z. The map $z \mapsto M_z$ is injective. If k is algebraically closed, Hilbert's Nullstellensatz implies that every maximal ideal in R is of this form. This gives a bijection between Spec$_m(R)$ and $\mathbb{A}^n$. Under this bijection, the closed set $V_m(I) := V(I) \cap \mathrm{Spec}_m(R)$ corresponds to $\{x \in \mathbb{A}^n \mid p(x) = 0 \ \forall p \in I\}$, which is the classical definition of the algebraic set defined by I. Since the Zariski-closed sets in $\mathbb{A}^n$ are precisely these algebraic sets, we obtain a homeomorphism

$$\mathrm{Spec}_m(k[x_1, \ldots, x_n]) \cong \mathbb{A}^n$$

between the subspace $\mathrm{Spec}_m(k[x_1, \ldots, x_n]) \subseteq \mathrm{Spec}(k[x_1, \ldots, x_n])$ and affine space $\mathbb{A}^n$ with the (classical) Zariski topology. The latter is well known to be T_1 and hereditarily compact, but this follows directly from our more general considerations on Spec(R).

Exercise B.6 Let R be a commutative unital ring. Show that $\operatorname{Spec}(R)$ is irreducible if and only if $I_1 \cdot I_2 \neq \{0\}$ whenever I_1, I_2 are non-zero ideals.

Appendix C

More on Topological Groups

C.1 Basics

Definition C.1.1 *A topological group is a group $(G, \cdot, e)$ equipped with a topology τ such that the group operations are τ-continuous. (It suffices to require that the map $(g, h) \mapsto gh^{-1}$ be jointly continuous.)*

It is important the the group operation be jointly continuous!

Exercise C.1.2 1. Let $G = (\mathbb{R}, +, 0)$ and τ the cocountable topology. Prove that $+$ is continuous w.r.t. both arguments, but not jointly. Thus $(\mathbb{R}, +, 0, \tau)$ is not a topological group!

2. Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and $n \in \mathbb{N}$. Let $GL(n, \mathbb{F})$ be the set of invertible $n \times n$-matrices with entries in $\mathbb{F}$, equipped with the topology induced from $\mathbb{F}^{n \times n}$. Then $(GL(n, \mathbb{F}), \cdot, e)$ is a topological group w.r.t. matrix multiplication and e the unit matrix. (The product of two matrices is polynomial in the entries of the matrices, thus jointly continuous. The fact that inversion of invertible matrices is continuous follows from Cramer's formula and continuity of $A \mapsto \det(A)$.) Now any subgroup $G \subseteq GL(n, \mathbb{F})$ also is a topological group.

If G is a topological group and $A, B \subseteq G$ then we write $AB := \{ab \mid a \in A, b \in B\}$. gH and Hg are definied analogously. If $h \in G$, the maps $i : g \mapsto g^{-1}$, $l_h : g \mapsto hg$ and $r_g : g \mapsto gh$ are continuous bijections with continuous inverses i, $l_{h^{-1}}$, $r_{h^{-1}}$, respectively, thus homeomorphisms. Thus if U is open then so are $U^{-1} \equiv \{g^{-1} \mid g \in U\}$ and hU, Uh.

Exercise C.1.3 If $\{G_i\}_{i \in I}$ are topological groups, prove that $\prod_i G_i$ equipped with the product group structure and the product topology is a topological group.

Lemma C.1.4 *Let G be a topological group.*

(i) Every open subgroup is closed, thus clopen.

(ii) If $H \subseteq G$ is a (normal) subgroup then $\overline{H}$ is a (normal) subgroup.

Proof. (i) If H is open, so are all cosets gH since l_g is a homeomorphism. Thus $U = \bigcup\{gH \mid gH \neq H\}$ is open, and $H = G - U$ is closed.

(ii) Since $i : G \to G$ is a homeomorphism, we have $\overline{H}^{-1} = i(\overline{H}) = \overline{i(H)} = \overline{H^{-1}} = \overline{H}$, thus $\overline{H}$ is closed under inverses. Since l_g is a homeomorphism, we have $g\overline{H} = l_g(\overline{H}) = \overline{l_g(H)} = \overline{gH}$. For $h \in H$ we have $hH = H$ so that $h\overline{H} = \overline{H}$. Thus $\overline{H}$ is closed under left multiplication with elements of H. Analogously one proves $\overline{H}h = \overline{H}$, thus $H\overline{H} = \overline{H} = \overline{H}H$. Let now $h, h' \in \overline{H}$.

Then $hh' \in h\overline{H} = l_h(\overline{H}) = \overline{l_h(H)} = \overline{hH} \subseteq \overline{\overline{H}H} = \overline{H}$, thus $\overline{H}\,\overline{H} = \overline{H}$, so that $\overline{H}$ is closed under multiplication. Thus $\overline{H}$ is a subgroup. ■

Lemma C.1.5 *Let G be a topological group.*

(i) If $U, V \subseteq G$ and U (or V) is open then $UV \subseteq G$ is open.

(ii) If $K, L \subseteq G$ are compact then KL is compact.

(iii) $\overline{A} = \bigcap AV = \bigcap VA$, where the intersection is over the open sets V containing e.

Proof. (i) If U is open then $UV = \bigcup_{v \in V} Uv$, which is a union of open sets, thus open. Similarly if V is open.

(ii) $K \times L$ is compact, thus $KL = m(K \times L)$ is compact since $m : G \times G \to G$ is continuous.

(iii) We have $x \in \overline{A}$ if and only if $A \cap U \neq \emptyset$ for every open neighborhood U of x. Every open $U \ni x$ is of the form Vx for some open $V \ni e$. Now, $A \cap Vx \neq \emptyset$ is equivalent to $V^{-1}A \cap \{x\} \neq \emptyset$, i.e. $x \in V^{-1}A$. Since $V \mapsto V^{-1}$ is an involution on the set of open neighborhoods of e, we have $\overline{A} = \bigcap_V VA$. Replacing Vx by xV, one obtains $\overline{A} = \bigcap_V AV$. ■

Definition C.1.6 *Let G be a topological group. A subset U called symmetric if $e \in U$ and $U = U^{-1}$.*

Lemma C.1.7 *Let G be a topological group. Then for every open neighborhood U of e there is a symmetric open neighborhood V of e such that $VV \subseteq U$. (I.e. $g, h \in V \Rightarrow gh \in U$.) Any such V satisfies $\overline{V} \subseteq U$.*

Proof. Write $m : G \times G \to G$ for the multiplication map. Then $m^{-1}(U) \subseteq G \times G$ is open and contains (e, e). By definition of the product topology, there open U_1, U_2 containing e such that $U_1 \times U_2 \subseteq m^{-1}(U)$. Then $V_0 = U_1 \cap U_2$ is open, contains e, and $V_0 \times V_0 \subseteq U_1 \times U_2 \subseteq m^{-1}(U)$ implies $V_0 V_0 \subseteq U$. Now take $V = V_0 \cap V_0^{-1}$, which still is an open neighborhood of e.

Lemma C.1.5(iii) in particular gives $\overline{A} \subseteq AV$ for any A and any open V, thus $\overline{U} \subseteq U^2$ for open U. ■

C.2 Separation axioms and metrizability for topological groups

Proposition C.2.1 *Every T_0-topological group is T_3.*

Proof. Let $e \neq g \in G$. We claim that there is an open U such that $e \in U \not\ni g$. By the T_0-axiom, there is an open U such that $\#(U \cap \{e, g\}) = 1$. In the case $e \in U \not\ni g$ we are done. In the case $g \in U \not\ni e$, note that $g^{-1}U$ contains e but not g^{-1}. Then $(g^{-1}U)^{-1}$ contains e, but not g, and again we are done. Let now $g \neq h$. By the above, there is an open neighborhood U of e not containing $g^{-1}h$. Then gU is an open neighborhood of g not containing h. Thus G is T_1.

Lemma C.1.7 tells us that given $e \in U \in \tau$, there is an open V such that $e \in V \subseteq \overline{V} \subseteq U$. Let now $g \in U \in \tau$ be given. Now given $g \in U \in \tau$, there is an open V such that $e \in V \subseteq \overline{V} \subseteq g^{-1}U$, and it follows that $g \in gV \subseteq g\overline{V} \subseteq U$. Thus G is T_3 by Lemma 8.1.5. ■

Exercise C.2.2 Let G be a T_0-topological group, $K \subseteq G$ compact and $L \subseteq G$ closed. Prove that KL and LK are closed. Hint: Use Exercise 7.5.5 or nets.

The combination of the axioms of group theory and of topology has some very strong consequences:

Theorem C.2.3 *Every T_0-topological group is completely regular ($T_{3.5}$).*

Proof. The T_1-property was proven in Proposition C.2.1. For complete regularity, it suffices to prove that given an open $U \ni e$ there is $f \in C(G,[0,1])$ such that $f(e)=0$ and $f \upharpoonright G\backslash U = 1$. The general situation $g \in U$ then follows by translation as in the proof of Proposition C.2.1. Define $U_0 = U$ and use Lemma C.1.7 to inductively pick symmetric open neighborhoods U_i of e satisfying $U_{i+1}^2 \subseteq U_i$. If $n < N$ then

$$\begin{aligned} U_{n+1}\cdots U_{N-1}U_N &\subseteq U_{n+1}\cdots U_{N-1}U_{N-1} = U_{n+1}\cdots U_{N-2}U_{N-1}^2 \subseteq U_{n+1}\cdots U_{N-2}U_{N-2} \\ &\subseteq U_{n+1}\cdots U_{N-3}^2 \subseteq \cdots \subseteq U_{n+1}^2 \subseteq U_n, \end{aligned}$$

thus

$$U_{n+1}U_{n+2}\cdots U_N \subseteq U_n. \tag{C.1}$$

We write

$$\{0,1\}^* = \{a \in \{0,1\}^{\mathbb{N}} \mid \ell(a) := \max\{i \mid a_i = 1\} < \infty\} \subseteq \{0,1\}^{\mathbb{N}}.$$

To every $a \in \{0,1\}^*$ we associate a $V_a \subseteq G$ by

$$V_a = U_1^{a_1}U_2^{a_2}\cdots U_\ell^{a_\ell}, \quad \ell = \ell(a),$$

where $U^1 = U$ and $U^0 = \{e\}$. Ordering the elements of $\{0,1\}^*$ lexicographically as in Remark 8.2.19, we claim that $a,b \in \{0,1\}^*,\ a < b \Rightarrow \overline{V_a} \subseteq V_b$.

To prove this assume that $a_i = b_i \ \forall i < n$ and $a_n = 0$, $b_n = 1$. Now,

$$\begin{aligned} V_a &= U_1^{a_1}U_2^{a_2}\cdots U_{n-1}^{a_{n-1}}\ \{e\}\ U_{n+1}^{a_{n+1}}\cdots U_{\ell(a)}^{a_{\ell(a)}} \\ &\subseteq U_1^{a_1}U_2^{a_2}\cdots U_{n-1}^{a_{n-1}}\ \{e\}\ U_{n+1}\cdots U_{\ell(a)} \\ &\subseteq U_1^{a_1}U_2^{a_2}\cdots U_{n-1}^{a_{n-1}}\ U_n \\ &\subseteq U_1^{a_1}U_2^{a_2}\cdots U_{n-1}^{a_{n-1}}\ U_n\ U_{n+1}^{b_{n+1}}\cdots U_{\ell(b)}^{b_{\ell(b)}} = V_b, \end{aligned}$$

where the second inclusion is due to (C.1). Thus $a < b \Rightarrow V_a \subseteq V_b$. Now define $a' \in \{0,1\}^*$ by $a' = (a_1,\ldots,a_{\ell(a)},1)$. We then have $a' < b$, thus $V_{a'} \subseteq V_b$. By definition of V_a, we have $V_{a'} = V_aU_{\ell(a)+1}$. Lemma C.1.5(iii) gives $\overline{A} \subseteq AV$ for every open $V \ni e$, thus

$$\overline{V_a} \subseteq V_aU_{\ell(a)+1} = V_{a'} \subseteq V_b,$$

completing the proof of $\overline{V_a} \subseteq V_b$.

Now we invoke the order-preserving bijection $\alpha : (0,1)\cap\mathbb{D} \to \{0,1\}^*$ to define $W_r = V_{\alpha(r)}$. We clearly have $e \in \bigcap_r W_r$ and $\bigcup_r W_r \subseteq U$. Thus Lemma 8.2.2 provides a function $f \in C(G,[0,1])$ such that $f(e)=0$ and $f\upharpoonright G\backslash U = 1$. ■

Example C.2.4 The space $G = \mathbb{R}^{\mathbb{R}} = \prod_{x\in\mathbb{R}}\mathbb{R}$ is a topological group and a topological vector space (w.r.t. coordinatewise operations). Now G is $T_{3.5}$ by Exercise 8.3.5, but non-normal by Corollary 8.1.44. □

Remark C.2.5 1. Theorem C.2.3 is hard to find in the literature (the most accessible account probably being the one in [116]), which is a pity in view of its generality. The main reason probably is that most texts on topological groups quickly specialize to locally compact groups, for which the complete regularity follows without further reference to the group structure from local compactness and the Hausdorff property (which is either assumed or deduced from T_0 or T_1 as we did).

2. Combining Theorem C.2.3 with the results of Section 8.3.4, we see that the topology of a T_0 topological group can be described in terms of a family of pseudometrics. In the special case of a topological vector space it therefore is natural to ask whether its topology arises from a family of seminorms. In Section G.5 we will see that this is the case precisely for the locally convex vector spaces, briefly encountered in Section 8.5.5. □

No natural necessary and sufficient condition for normality of a topological group seems to be known, but here is a sufficient one:

Proposition C.2.6 *A locally compact T_0-group is normal and paracompact.*

Proof. By local compactness, there are $e \in V \subseteq K$ with V open and K compact. Then $U = V \cap V^{-1}$ is symmetric and $\overline{U}$ is compact. By construction, $H = \bigcup_{n\in\mathbb{N}} U^n$ is closed under multiplication and inversion (since U is symmetric). Thus $H \subseteq G$ is an open subgroup, thus closed by Lemma C.1.4(i). By Lemma C.1.7, $\overline{U} \subseteq U^2$. Thus $H = \bigcup_n U^n = \bigcup_n \overline{U}^n$. By Lemma C.1.5(ii), $\overline{U}^n$ is compact, thus H is σ-compact, thus Lindelöf. Since G is T_3 by Proposition C.2.1, Propositions 8.1.15 and 8.5.12 give normality and paracompactness of H. The same hold for the cosets gH. Since G topologically is a direct sum of cosets gH, G is normal and paracompact. ■

For the stronger property of metrizability there is a very satisfactory criterion:

Theorem C.2.7 *(i) A topological group is metrizable if and only if it is T_0 and the unit e has a countable neighborhood base. In that case, the metric can be taken to be left-invariant, i.e. $d(kg, kh) = d(g, h)\ \forall g, h, k$.*

(ii) If G is compact metrizable, d can be chosen two-sided invariant, i.e. $d(kgl, khl) = d(g, h)$.

Appendix D

The fixed point theorems of Banach and Caristi

D.1 Banach's contraction principle and variations

Definition D.1.1 *If (X, d) is a metric space, a map $f : X \to X$ is a contraction if there is a constant K such that $0 \leq K < 1$ and $d(f(x_1), f(x_2)) \leq K d(x_1, x_2)$ for all $x_1, x_2 \in X$.*

Obviously, a contraction is continuous. We denote the m-fold composition $f \circ \cdots \circ f$ as f^m.

Theorem D.1.2 (Banach's fixed point theorem, contraction principle) *Every contraction f in a non-empty complete metric space (X, d) has a unique fixed point z.*

More precisely, given any $x \in X$ we have $f^n(x) \to z$ at the rate given by

$$d(f^n(x), z) \leq \frac{K^n}{1-K} d(x, f(x)) \quad \forall n. \tag{D.1}$$

Proof. Let $f : X \to X$ have contraction constant $K < 1$. If x_1, x_2 are fixed points then the assumption implies $d(x_1, x_2) = d(f(x_1), f(x_2)) \leq K d(x_1, x_2)$. Thus $d(x_1, x_2) = 0$ and therefore $x_1 = x_2$. This proves the uniqueness.

Induction gives

$$d(f^m(x_1), f^m(x_2)) \leq K^m d(x_1, x_2). \tag{D.2}$$

The triangle inequality and the fact that f is a contraction give

$$\begin{aligned} d(x_1, x_2) &\leq d(x_1, f(x_1)) + d(f(x_1), f(x_2)) + d(f(x_2), x_2) \\ &\leq d(x_1, f(x_1)) + K d(x_1, x_2) + d(f(x_2), x_2), \end{aligned}$$

thus

$$(1-K) d(x_1, x_2) \leq d(x_1, f(x_1)) + d(f(x_2), x_2).$$

In view of $K < 1$ we have $1 - K > 0$, so that dividing by $1 - K$ gives

$$d(x_1, x_2) \leq \frac{d(x_1, f(x_1)) + d(x_2, f(x_2))}{1-K} \quad \forall x_1, x_2. \tag{D.3}$$

Let $x \in X$ be arbitrary. Putting $x_1 = f^n(x),\ x_2 = f^m(x)$ in (D.3) and using (D.2), we have

$$\begin{aligned} d(f^n(x), f^m(x)) &\leq \frac{d(f^n(x), f^n(f(x))) + d(f^m(x), f^m(f(x)))}{1-K} \\ &\leq \frac{K^n + K^m}{1-K}\, d(x, f(x)) \end{aligned} \tag{D.4}$$

Since $K < 1$, it follows immediately that the sequence $\{f^n(x)\}$ is Cauchy. Since X is complete, this sequence converges to some $z \in X$. Using continuity of f, we have $f(z) = f(\lim_n f^n(x)) = \lim_n f^{n+1}(x) = z$, thus z is a fixed point. Taking the limit $m \to \infty$ in (D.4), we get (D.1). ■

Remark D.1.3 1. The above proof, due to Palais [188], avoids summing a geometric series, making it marginally simpler than the standard textbook proof (and perhaps also prettier).

2. The first statement of a general contraction principle (albeit not quite in the above form) appears in the 1922 Ph.D. thesis of Stefan Banach. But the idea already underlied the 19th century work of Picard, Lindelöf and Lipschitz on the existence and uniqueness of solutions for ordinary differential equations. The contraction principle still plays a central rôle in the theory of ordinary and partial differential equations. Cf. Section **??** for the simplest such result. □

There are many variations on the contraction principle, cf. e.g. [140, Section 3.2]. We consider two of them.

Corollary D.1.4 *Let (X, d) be a complete metric space and $f : X \to X$ such that f^n is a contraction for some $n \in \mathbb{N}$. Then f has a unique fixed point.*

Proof. By the contraction principle, there is a unique $x \in X$ such that $f^n(x) = x$. Now $f(x) = f(f^n(x)) = f^n(f(x))$, thus $f(x)$ is a fixed point of f^n. But then uniqueness of the latter implies $f(x) = x$. Thus x is a fixed point of f. If x, y are fixed points of f, we have $x = f^n(x)$ and $y = f^n(y)$, and uniqueness of the fixed point of f^n implies $x = y$. ■

Definition D.1.5 *Let (X, d) be a metric space. A map $f : X \to X$ is called a weak contraction if $d(f(x), f(y)) < d(x, y)$ whenever $x \neq y$.*

Obviously, every contraction is a weak contraction, but the converse is not true. In fact, a weak contraction of a complete metric space need not have a fixed point. But:

Theorem D.1.6 (Edelstein 1962 [65]) *Let (X, d) be a non-empty compact metric space and $f : X \to X$ a weak contraction. Then*

(i) f has a unique fixed point z.

(ii) For every $x \in X$ we have $f^n(x) \to z$.

Proof. (i) Let $x \neq y$ be fixed points. Then $d(x, y) = d(f(x), f(y)) < d(x, y)$, which is absurd. This proves uniqueness.

Every weak contraction is continuous. Thus $g : X \to \mathbb{R},\ x \mapsto d(x, f(x))$ is continuous. By Corollary 7.7.30, $C = \inf g(X)$ is finite and there is a $z \in X$ such that $g(z) = C$. Assuming $C > 0$, we have $C = g(z) = d(z, f(z)) > 0$, thus $z \neq f(z)$. Then by assumption $g(f(z)) = d(f(z), f(f(z))) < d(z, f(z)) = C$, contradicting the fact that C is the infimum of g. Thus $C = 0$. But now $0 = C = g(z) = d(z, f(z))$ implies that $f(z) = z$, i.e. z is a fixed point.

(ii) Let $x \in X$. If $f^n(x) = z$ for some n, the sequence is stationary, thus $f^n(x) \to z$. We may therefore assume that $f^n(x) \neq z$ for all n. Then

$$d(f^{n+1}(x), z) = d(f^{n+1}(x), f(z)) < d(f^n(x), z),$$

thus the non-negative sequence $\{d(f^n(x), z)\}$ is strictly decreasing and therefore converges to some $r \geq 0$. Since compact metric spaces are sequentially compact, the sequence $\{f^n(x)\}$ has a subsequence $\{f^{n_m}(x)\}$ that converges to some $z' \in X$. Since this implies

$$d(z', z) = \lim_m d(f^{n_m}(x), z) = r = \lim_m d(f^{n_m+1}(x), z) = d(f(z'), z),$$

we have $d(z', z) = d(f(z'), z) = d(f(z'), f(z))$. Since f is a weak contraction, we must have $z' = z$, so that $f^{n_m}(x) \to z$. Thus the non-increasing sequence $\{d(f^n(x), z)\}$ has a subsequence $\{d(f^{n_m}(x), z)\}$ converging to zero and thus converges to zero itself. This proves $f^n(x) \to z$. ■

For a nice application of Theorem D.1.6, cf. [206] where it is used to give a quick proof of the Perron-Frobenius theorem. This approach has the advantage of avoiding Brouwer's fixed point theorem, whose proof invariably involves a certain amount of combinatorics.

D.2 Caristi's fixed point theorem

The following easy result is proved mainly to motivate what follows and may be skipped.

Proposition D.2.1 *Let (X, d) be a non-empty complete metric space, $\phi : X \to \mathbb{R}$ bounded below and $f : X \to X$ a continuous function satisfying*

$$d(x, f(x)) \le \phi(x) - \phi(f(x)) \quad \forall x \in X. \tag{D.5}$$

Then $\{f^n(x)\}$ converges to a fixed point of f for every x in X.

Proof. The assumption (D.5) implies $\phi(f(x)) \le \phi(x)$. Thus the sequence $\{\phi(f^n(x))\}$ is weakly decreasing and bounded below, thus converges to some $r \in \mathbb{R}$. If $n < m$ we have

$$d(f^n(x), f^m(x)) \le \sum_{k=n}^{m-1} d(f^k(x), f^{k+1}(x)) \le \sum_{k=n}^{m-1} \phi(f^k(x)) - \phi(f^{k+1}(x)) = \phi(f^n(x)) - \phi(f^m(x))$$

by the triangle inequality, (D.5) and 'telescoping'. For $n, m \to \infty$ the right hand side tends to $r - r = 0$, implying that $\{f^n(x)\}$ is a Cauchy sequence, thus convergent to some $z \in X$ by completeness. Now $f(z) = z$ follows from continuity of f as at the end of the proof of Theorem D.1.2. ■

Note that uniqueness of the fixed point is not claimed and that continuity of f is used only in the last step where, however, it is indispensable. Surprisingly, replacing the continuity of f by lower semicontinuity of ϕ, one can still prove the following:

Theorem D.2.2 (Caristi 1976 [43]) *Let (X, d) be a non-empty complete metric space, $\phi : X \to \mathbb{R}$ bounded below and lower semicontinuous, and let $f : X \to X$ satisfy*

$$d(x, f(x)) \le \phi(x) - \phi(f(x)) \quad \forall x \in X. \tag{D.6}$$

Then f has a fixed point.

The brunt of proving this is borne by the following widely applicable result:

Proposition D.2.3 *Let (X, d) be a non-empty complete metric space and $\phi : X \to \mathbb{R}$ lower semicontinuous and bounded below. Define a binary relation $\le$ on X by*

$$x \le y \quad \Leftrightarrow \quad d(x, y) \le \phi(x) - \phi(y).$$

(i) Then $\le$ is a preorder (i.e. reflexive and transitive).

(ii) This preorder has a maximal element.

Proof. (i) If $x \le y \le z$ then $d(x,y) \le \phi(x) - \phi(y)$ and $d(y,z) \le \phi(y) - \phi(z)$. Adding these inequalities we have $d(x,z) \le d(x,y) + d(y,z) \le \phi(x) - \phi(z)$, thus $x \le z$ so that $\le$ is transitive. Reflexivity of $\le$ is obvious.

(ii) For every $x \in X$, define the set of (nonstrict) majorants by

$$\overline{M}(x) = \{y \in X \mid y \ge x\} = \{y \in X \mid \phi(y) + d(x,y) \le \phi(x)\}.$$

Since ϕ is lower semicontinuous and d is (jointly) continuous, the function $\phi_x : y \mapsto \phi(y) + d(x,y)$ is lower semicontinuous. Thus $\phi_x^{-1}((a,\infty))$ is open for any a, thus $\overline{M}(x) = \phi_x^{-1}((-\infty, \phi(x)])$ is closed. Choose $x_0 \in X$ arbitrarily and inductively define a sequence $\{x_n\}$ as follows: Given x_n, choose $x_{n+1} \in \overline{M}(x_n)$ such that $\phi(x_{n+1}) < \inf(\phi(\overline{M}(x_n))) + 1/n$. (This can be done since the infimum is finite, ϕ being bounded below.) By construction $x_{n+1} \ge x_n$, and for $x \in \overline{M}(x_{n+1}) \subseteq \overline{M}(x_n)$ (transitivity!) we have

$$\phi(x) \ge \inf(\phi(\overline{M}(x_n))) > \phi(x_{n+1}) - \frac{1}{n},$$

thus $\phi(x_{n+1}) - \phi(x) < 1/n$. Since $x \ge x_{n+1}$, we have $d(x, x_{n+1}) \le \phi(x_{n+1}) - \phi(x)$ and thus $d(x, x_{n+1}) < 1/n$. This implies $\mathrm{diam}(\overline{M}(x_{n+1})) \le 2/n$. Combining this with the closedness of the sets $X \supseteq \overline{M}(x_0) \supseteq \overline{M}(x_1) \supseteq \cdots$ and completeness of (X,d), Cantor's Intersection Theorem (Exercise 3.1.6) gives a $z \in X$ such that $\bigcap_n \overline{M}(x_n) = \{z\}$. This means $z \ge x_n \ \forall n$, so that $x \ge z$ implies $x \in \overline{M}(x_n)$ for all n and therefore $x = z$. This means that z is maximal w.r.t. $\le$. ∎

Proof of Theorem D.2.2. Defining $\le$ as in Proposition D.2.3, the latter gives a maximal element $z \in X$. Since (D.6) is equivalent to $f(x) \ge x \ \forall x$, we have $f(z) \ge z$, so that maximality of z gives $f(z) = z$. ∎

Remark D.2.4 1. Theorem D.2.2 asserts neither uniqueness of the fixed point nor convergence of $\{f^n(x)\}$ for every (or any) $x \in X$.

2. It is also true that Caristi's theorem implies Proposition D.2.3(ii): If $\le$ had no maximal element, we could pick an $f(x) > x$ for every $x \in X$. This implies $d(x, f(x)) \le \phi(x) - \phi(f(x)) \ \forall x$, so that Caristi's theorem provides a fixed point $z = f(z)$, contradicting the construction of f.

3. The above proof of Proposition D.2.3, which closely followed [190], clearly uses the axiom of countable dependent choice (ACDC). (Cantor's intersection theorem only uses countable choice.) It has been shown [38] that Proposition D.2.3 also implies the axiom of countable dependent choice and therefore is equivalent to it. (As noted earlier, the same holds for Baire's Theorem 3.3.1.)

4. Since Theorem D.2.2 follows from Proposition D.2.3 without invoking choice axioms, it follows from ACDC. Remarkably, Theorem D.2.2 can be proven without using only the axiom ACC of countable choice, cf. [35]. This is no contradiction to the preceding remark, since the deduction of the Proposition from the Theorem clearly used the full AC. (There are papers claiming to prove Caristi's theorem [210, 163] without any use of choice, but they are not convincing. Even [35] uses ACC without pointing it out!)

5. The existence part of Theorem D.1.2 (but not $f^n(x) \to z \ \forall x$) can be deduced from Caristi's Theorem D.2.2: If $f : X \to X$ is a contraction of a complete metric space (X,d) with contraction constant $0 \le K < 1$, it is obvious that $\phi : X \to \mathbb{R}$ by $\phi(x) = \frac{d(x,f(x))}{1-K}$ is continuous and bounded below. In view of

$$d(x, f(x)) = \frac{d(x,f(x)) - Kd(x,f(x))}{1-K} \le \frac{d(x,f(x)) - d(f(x), f^2(x))}{1-K} = \phi(x) - \phi(f(x)),$$

Caristi's theorem gives that f has a fixed point.

6. So far, Caristi's fixed point theorem seems to have made it only into specialized texts on metric fixed point theory [93, 140] and into advanced books on non-linear (functional) analysis like [10, 56]. Indeed, most of its applications seem to be in non-linear analysis and infinite dimensional geometry. (We already used Caristi's theorem for proving Theorem 12.4.8, following [93].) □

If $\phi : X \to \mathbb{R}$ is bounded below, $\inf \phi$ is finite, but it does not follow that infimum is assumed, i.e. that there is $z \in X$ with $\phi(z) = \inf \phi$. We know that compactness of X together with continuity of ϕ is sufficient, cf. Corollary 7.7.30. Using Proposition D.2.3 one has another set of sufficient conditions:

Corollary D.2.5 (Takahashi [222]) *Let (X, d) be a complete metric space and $\phi : X \to \mathbb{R}$ bounded below and lower semicontinuous. Suppose that*

$$x \in X,\ \phi(x) > \inf \phi \quad \Rightarrow \quad \exists y \neq x : \phi(y) < \phi(x) - d(x, y). \tag{D.7}$$

Then ϕ assumes its infimum.

Proof. Proposition D.2.3 provides a maximal element z for the preorder $\leq$ defined by ϕ. By assumption (D.7), $\phi(x) > \inf \phi$ implies the existence of $y \neq x$ such that $\phi(y) < \phi(x) - d(x, y)$. This implies $y > x$. For the maximal element z no $y > z$ can exist, thus $\phi(z) = \inf \phi$. ■

Imposing an additional condition on the function ϕ whose infimum one would like to be assumed is not an option in most cases. Part (ii) of the next result shows that even without a condition like (D.7) one can always find good approximate minima. The preliminary result (i) is nothing but a less snappy (but equivalent) restatement of Proposition D.2.3:

Theorem D.2.6 (Ekeland's Variational 'Principle' 1974 [70]) *Let (X, d) be a non-empty complete metric space and $\phi : X \to \mathbb{R}$ lower semicontinuous and bounded below. Then*

(i) There exists $z \in X$ such that

$$x \neq z \quad \Rightarrow \quad \phi(x) - \phi(z) > -d(z, x).$$

(ii) For every $\varepsilon, \delta > 0$ and $y \in X$ with $\phi(y) \leq \inf \phi + \varepsilon$ there exists $z \in X$ such that

$$d(z, y) \leq \delta, \qquad \phi(z) \leq \phi(y), \qquad x \neq z \ \Rightarrow\ \phi(x) > \phi(z) - \frac{\varepsilon}{\delta} d(x, z).$$

Proof. (i) Let $z \in X$ be a maximal element for the preorder $\leq$ as provided by Proposition D.2.3. Maximality of z means that $x > z$, to wit $d(x, z) \leq \phi(z) - \phi(x) \wedge x \neq z$, is impossible. Thus $x \neq z$ implies $d(x, z) > \phi(z) - \phi(x)$. This is precisely what we claimed.

(ii) Since ϕ is lower semicontinuous, $\phi^{-1}((-\infty, \phi(y)]) \subseteq X$ is closed. Also $\overline{B}(y, \delta)$ is closed, thus $X' = \phi^{-1}((-\infty, \phi(y)]) \cap \overline{B}(y, \delta) \subseteq X$ is closed (and non-empty since $y \in X'$), thus complete. In view of $\delta > 0$, $d'(x, y) = d(x, y)/\delta$ is a metric equivalent to d, thus (X', d') is complete. Since $\phi'(x) = \phi(x)/\varepsilon$ lower semicontinuous and bounded below, (X', d') and ϕ' satisfy the assumptions of (i), so that we obtain a point $z \in X' \subseteq X$. By construction, z satisfies $d(z, y) \leq \delta$ and $\phi(z) \leq \phi(y)$. And if $x \neq z$, by the conclusion of (i) we have

$$\frac{\phi(x) - \phi(z)}{\varepsilon} = \phi'(x) - \phi'(z) > -d'(z, x) = \frac{d(z, x)}{\delta},$$

which gives the last of the claims in (ii). ■

Remark D.2.7 1. It is obvious that the result generalizes to functions $\phi : X \to \mathbb{R} \cup \{+\infty\}$ that are not identically $+\infty$.

2. Ekeland's variational 'principle' actually predates Caristi's theorem. It is clear that Proposition D.2.3, Theorem D.2.2 and Theorem D.2.6 all imply each other.

3. Several other equivalent statements and many applications can be found in [30]. Unsurprisingly, Ekeland's variational principle has applications to the direct approach to variational problems, cf. [91]. □

D.3 Application: Iterated function systems

Definition D.3.1 *An* <u>*iterated function system*</u> *on a metric space* (X, d) *is a family* $S_1, \ldots, S_m$ *of contractions of* (X, d).

The aim of this section is to prove the following theorem:

Theorem D.3.2 *Let* (X, d) *be a complete metric space and* $S_1, \ldots, S_m$ *an iterated function system on* (X, d). *Then there is a unique non-empty compact set* $F \subseteq X$, *called the* <u>*attractor*</u> *of the iterated function system* $S_1, \ldots, S_m$, *such that*

$$F = \bigcup_{i=1}^{m} S_i(F).$$

It is easy to see that for $m = 1$ one gets $F = \{z\}$, where z is the unique fixed point of the contraction S_1. But for $n \geq 2$, one can obtain fairly interesting attractors. In order to prove the theorem we need some tools.

Definition D.3.3 *Let* (X, d) *be a metric space. The Pompeiu-Hausdorff distance of two non-empty subsets* $A, B \subseteq X$ *is defined as*

$$\mathrm{Dist}(A, B) = \max\left(\sup_{a \in A} \mathrm{dist}(a, B), \ \sup_{b \in B} \mathrm{dist}(b, A)\right).$$

Remark D.3.4 1. The Pompeiu-Hausdorff distance $\mathrm{Dist}(A, B)$ must not be confused with the ordinary distance $\mathrm{dist}(A, B) = \inf\{d(a, b) \mid a \in A, \ b \in B\}$. Note that we always have $\mathrm{dist}(A, B) \leq \mathrm{Dist}(A, B)$.

2. Equivalently one could define $\mathrm{Dist}(A, B) = \inf\{\varepsilon > 0 \mid A \subseteq B_\varepsilon \wedge B \subseteq A_\varepsilon\}$. □

Proposition D.3.5 *For every metric space* (X, d), *the Pompeiu-Hausdorff distance is a metric on the family* $\mathcal{B}$ *of bounded subsets of* X.

Proof. If A is bounded, the triangle inequality gives

$$\sup_{a \in A} \mathrm{dist}(a, B) \leq \mathrm{dist}(A, B) + \mathrm{diam}(A) < \infty.$$

Similarly boundedness of B gives $\sup_{b \in B} \mathrm{dist}(b, A) < \infty$, and thus $\mathrm{Dist}(A, B) < \infty$.

It is clear that $\mathrm{Dist}(A, B) = \mathrm{Dist}(B, A)$. By definition, we have $\mathrm{Dist}(A, B) = 0$ if and only if $\sup_{a \in A} \mathrm{dist}(a, B) = \sup_{b \in B} \mathrm{dist}(b, A) = 0$, which in turn is equivalent to $\mathrm{dist}(a, B) = \mathrm{dist}(b, A) = 0$ for all $a \in A, \ b \in B$. By closedness of A and B, this implies $A \subseteq B$, $B \subseteq A$, thus $A = B$. It is even clearer that $\mathrm{Dist}(A, A) = 0$.

***** triangle ineq.

■

Remark D.3.6 ************ check !!! usually this is stated for compact subsets ! *******

One can show, cf. [112], that if (X, d) is complete, resp. totally bounded, then $(\mathcal{B}, \mathrm{Dist})$ is complete, resp. totally bounded. Thus if (X, d) is compact then so is $(\mathcal{B}, \mathrm{Dist})$. We will not need this in the sequel. □

Remark D.3.7 1. In Theorem D.3.2 the metric space (X, d) was arbitrary. More can be said if $X = \mathbb{R}^n$ with the Euclidean metric. To every subset $F \subseteq \mathbb{R}^n$ one can associate its Hausdorff dimension $\dim_H(F)$ which can (a priori) assume every value in $[0, n]$. Under somewhat stronger assumptions on the contractions S_i, one can compute $\dim_H(F)$ in terms of the contraction constants $C_1, \ldots, C_n$. Namely $\dim_H(F)$ equals the unique real number $s \geq 0$ for which

$$\sum_{i=1}^{m} C_i^s = 1.$$

(For a proof cf. [78, Theorem 9.3].) If $C_i = C \;\forall i$, this gives $\dim_H(F) = \frac{\log m}{\log(1/c)}$. In particular we see that $m = 1$ implies $\dim_H(F) = 0$, consistent with F being a single point.

2. The standard Cantor $C \subseteq [0, 1]$ set is the attractor of the iterated function system

$$X = [0, 1], \quad m = 2, \quad S_1(x) = x/3, \quad S_2(x) = \frac{x+2}{3},$$

where $C_1 = C_2 = 1/3$. Then 1. gives $\dim_H(C) = \frac{\log 2}{\log 3} \approx 0.631$. □

Appendix E

The sequence spaces $\ell^p(S)$

In this section, we will prove a few instructive results for the infinite dimensional spaces equipped with norms $\|\cdot\|_p$.[1]

E.1 Definition and completeness

Definition E.1.1 (ℓ^p-Spaces) *If $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, S is a set and f is a function from S to $\mathbb{F}$, define $\|f\|_\infty = \sup_{s\in S} |f(s)|$, and for $1 \le p < \infty$ define*

$$\|f\|_p = \left(\sum_{s\in S} |f(s)|^p\right)^{1/p} \equiv \sup_{\substack{T\subseteq S\\ \text{finite}}} \left(\sum_{t\in T} |f(t)|^p\right)^{1/p} \in [0,\infty].$$

Now put

$$\ell^p(S,\mathbb{F}) := \{f : S \to \mathbb{F} \mid \|f\|_p < \infty\}.$$

Proposition E.1.2 *For every set S and $1 \le p < \infty$ we have that $\ell^p(S,\mathbb{F})$ is a vector space and $\|\cdot\|_p$ is a norm on it.*

Proof. If $f, g \in \ell^\infty(S)$ then obviously $f + g \in \ell^\infty(S)$. If $1 \le p < \infty$ and $f, g \in \ell^p(S,\mathbb{F})$ then

$$\begin{aligned}
\|f+g\|^p &= \sup_{\substack{T\subseteq S\\ \text{finite}}} \left(\sum_{t\in T} |f(t)+g(t)|^p\right)^{1/p} \le \sup_{\substack{T\subseteq S\\ \text{finite}}} \left[\left(\sum_{t\in T} |f(t)|^p\right)^{1/p} + \left(\sum_{t\in T} |f(g)|^p\right)^{1/p}\right] \\
&\le \sup_{\substack{T\subseteq S\\ \text{finite}}} \left(\sum_{t\in T} |f(t)|^p\right)^{1/p} + \sup_{\substack{T\subseteq S\\ \text{finite}}} \left(\sum_{t\in T} |g(t)|^p\right)^{1/p} = \|f\|_p + \|g\|_p < \infty,
\end{aligned}$$

where the first $\le$ is the Minkowski inequality for the finite dimensional space $\mathrm{Fun}(T,\mathbb{F}) \cong \mathbb{F}^{\#T}$. Thus $f + g \in \ell^p(S)$. It is obvious that $\|cf\|_p = |c|\|f\|_p$ for every $f \in \ell^p(S,\mathbb{F})$ and $c \in \mathbb{F}$. Thus $\ell^p(S)$ is a vector space and $\|\cdot\|_p$ is a norm on it. ■

Lemma E.1.3 *Let $p \in [1,\infty]$ and $d_p(x,y) = \|x-y\|_p$. Then $(\ell^p(S,\mathbb{F}), d_p)$ is complete for every $\mathbb{F} \in \{\mathbb{R},\mathbb{C}\}$, set S.*

[1] As exlained in the Preface, the author would disagree with the view that these spaces belong to functional analysis and therefore have no place in an introduction to topology. There is an entire branch of topology, infinite-dimensional topology, cf. e.g. [169], concerned with such spaces, in particular $\ell^2(\mathbb{N})$.

Proof. In this proof, we write $\|\cdot\|$ instead of $\|\cdot\|_p$. Let $\{f_n\}$ be a Cauchy sequence in $\ell^p(S,\mathbb{F})$. As in Lemma 3.1.5 one proves that $f_n(s) \stackrel{n\to\infty}{\longrightarrow} g(s)$ for every $s \in S$, but now the proof of $g \in \ell^p(S)$ and $\|f_n - g\| \to 0$ requires more work. Since $\{f_n\}$ is Cauchy, there is an $N \in \mathbb{N}$ such that $n, m \geq N$ implies $\|f_n - f_m\| \leq 1$. In particular, for all $n \geq N$ we have $\|f_n - f_N\| \leq 1$ and thus $\|f_n\| \leq \|f_N\| + 1 =: C$. Now

$$\|g\| = \sup_{\substack{T\subseteq S\\ \text{finite}}} \left(\sum_{t\in T} |g(t)|^p\right)^{1/p} = \sup_{\substack{T\subseteq S\\ \text{finite}}} \left(\sum_{t\in T} |\lim f_n(t)|^p\right)^{1/p} = \sup_{\substack{T\subseteq S\\ \text{finite}}} \lim_n \left(\sum_{t\in T} |f_n(t)|^p\right)^{1/p} \leq C$$

since $\sum_{t\in T} |f_n(t)|^p \leq C$ for all $n \in \mathbb{N}$ and finite $T \subseteq S$. Thus $g \in \ell^p(S)$. Since $\{f_n\}$ is Cauchy, for every $\varepsilon > 0$ there is N such that $n, m \geq N \Rightarrow \|f_n - f_m\| < \varepsilon$. Thus we have $\sum_{t\in T} |f_n(t) - f_m(t)|^p < \varepsilon^p$ for every finite $T \subseteq S$. Since the sum is finite and $f_n(t) \to g(t)$ for every t, we can take $m \to \infty$ and obtain $\sum_{t\in T} |f_n(t) - g(t)|^p \leq \varepsilon^p$. Taking the supremum over T, we have $\|f_n - g\| \leq \varepsilon$ and therefore $f_n \stackrel{\|\cdot\|_p}{\longrightarrow} g$. ■

Remark E.1.4 Another important class of function spaces equipped with a metric, besides the spaces $C(X,Y)$ and ℓ^p, is provided by the L^p-spaces. Since integration on measure spaces goes beyond the scope of this text, we only give the definition: If $(X, \mathfrak{A}, \mu)$ is a measure space and $f : X \to \mathbb{F}$ is a measurable function, then $\|f\|_p = (\int |f(x)|^p d\mu(x))^{1/p} \in [0,\infty]$. Then $\widetilde{L}^p(X,\mu) = \{f : X \to \mathbb{F} \text{ measurable} \mid \|f\|_p < \infty\}$ is a vector space and $\|\cdot\|_p$ is a seminorm on it. Defining $x \sim y \Leftrightarrow \|x - y\|_p = 0$, one proves that $\widetilde{L}^p(X,\mu)/\!\sim$ is a (complete) normed space, called $L^p(X,\mu)$. □

E.2 Separability

Proposition E.2.1 *Let $p \in [1,\infty)$. The metric space $(\ell^p(S,\mathbb{F}), d_p)$, where $d_p(f,g) = \|f - g\|_p$, is separable ($\Leftrightarrow$ second countable) if and only if the set S is countable.*

Proof. We prove this for $\mathbb{F} = \mathbb{R}$, from which the claim for $\mathbb{F} = \mathbb{C}$ is easily deduced. For $f : S \to \mathbb{R}$, let $\text{supp}(s) := \{s \in S \mid f(s) \neq 0\} \subseteq S$ be the support of f. Now, if S is countable, then $Y = \{g : S \to \mathbb{Q} \mid \#\text{supp}(g) < \infty\} \subseteq \ell^p(S)$ is countable, and we claim that $\overline{Y} = \ell^p(S)$. To prove this, let $f \in \ell^p(S)$ and $\varepsilon > 0$. Since $\|f\|^p = \sum_{s\in S} |f(s)|^p < \infty$, there is a finite subset $T \subseteq S$ such that $\sum_{s\in S\backslash T} |f(s)|^p < \varepsilon/2$. On the other hand, since $\mathbb{Q}^{\#T} \subseteq \mathbb{R}^{\#T}$ is dense, we can choose $g : T \to \mathbb{Q}$ such that $\sum_{t\in T} |f(t) - g(t)|^p < \varepsilon/2$. Defining g to be zero on $S\backslash T$, we have $g \in Y$ and

$$\|f - g\|^p = \sum_{t\in T} |f(t) - g(t)|^p + \sum_{s\in S\backslash T} |f(s)|^p < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,$$

and since $\varepsilon > 0$ was arbitrary, $Y \subseteq \ell^p(S)$ is dense.

For the converse, assume that S is uncountable. By Lemma E.2.2 below $\text{supp}(s)$ is countable for every $f \in \ell^p(S)$. Thus if $Y \subseteq \ell^p(S)$ is countable then $T = \bigcup_{f\in Y} \text{supp}(f) \subseteq S$ is a countable union of countable sets and therefore countable. Thus all functions $f \in Y$ vanish on $S\backslash T \neq \emptyset$, and the same holds for $f \in \overline{Y}$ since the coordinate maps $f \mapsto f(s)$ are continuous in view of $|f(s)| \leq \|f\|$. Thus Y cannot be dense. ■

Lemma E.2.2 *If S is a set and $f : S \to [0,\infty)$ satisfies $\sum_{s\in S} f(s) < \infty$ then $T = \{s \in S \mid f(s) \neq 0\}$ is countable.*

Proof. Defining $T_n = \{s \in S \mid f(s) > 1/n\}$, we clearly have $T = \bigcup_{n\in\mathbb{N}} T_n$. Now

$$\#T_n \frac{1}{n} \le \sum_{t\in T_n} f(t) \le \sum_{s\in S} f(s) < \infty$$

implies $\#T_n < \infty$ for each n. Thus T is a countable union of finite sets, thus countable. ■

Exercise E.2.3 Prove:

(i) The space $\ell^\infty(S,\mathbb{F})$ is separable if and only if S is finite.

(ii) The set of 'functions vanishing at infinity'

$$c_0(S,\mathbb{F}) = \{f : S \to \mathbb{F} \mid \forall\varepsilon > 0 \;\exists \text{ finite } F \subseteq S : x \in S\backslash F \Rightarrow |f(x)| < \varepsilon\}$$

is a closed subset (w.r.t. $\|\cdot\|_\infty$) of $\ell^\infty(S,\mathbb{F})$.

(iii) The space $(c_0(S,\mathbb{F}), d_\infty)$ is separable if and only if S is countable.

Hint: For (i), consider $\{0,1\}^S \subseteq \ell^\infty(S)$.

Remark E.2.4 1. The vector space $\ell^2(S)$ of Definition E.1.1 is a Hilbert[2] space with inner product $\langle f, g\rangle = \sum_{s\in S} f(s)\overline{g(s)}$. On the other hand, every orthonormal basis $B \subseteq \mathcal{H}$ of an (abstract) Hilbert space $\mathcal{H}$ gives rise to a unitary isomorphism $\mathcal{H} \cong \ell^2(B)$. In view of the above we thus see that a Hilbert space $\mathcal{H}$ is separable (in the sense of Definition 4.1.9) if and only if it admits a countable orthonormal basis. For this reason, the latter property is often taken as the definition of separability of a Hilbert space.

2. The question of separability of the metric space $(C_b(X,\mathbb{F}), D)$ will be addressed in Section F.1.3. □

E.3 Compactness in function spaces II: $\ell^p(S,\mathbb{F})$ and $L^p(\mathbb{R}^n)$

As done for the spaces $C(X,Y)$ in Section 7.7.7, we want to identify the compact subsets of the sequence spaces $\ell^p(S)$. Since the set S is discrete, there is no question of (equi)continuity of $\mathcal{F} \subseteq \ell^p(S)$, but the fact that S typically is infinite, thus non-compact leads to a new conditon on $\mathcal{F}$:

Definition E.3.1 *A family $\mathcal{F} \subseteq \ell^p(S)$ has equi-small tails if*

$$\lim_{\substack{T\nearrow S\\ T \text{ finite}}} \sup_{f\in\mathcal{F}} \|f \cdot \chi_{S\backslash T}\|_p = 0.$$

Equivalently, for every $\varepsilon > 0$ there is a finite subset $T \subseteq S$ such that $\sum_{s\in S\backslash T} |f(s)|^p < \varepsilon^p$ for every $f \in \mathcal{F}$.

Theorem E.3.2 *Consider $\ell^p(S)$ with the norm $\|f\|_p = (\sum_{s\in S} |f(s)|^p)^{1/p}$, where $1 \le p < \infty$, cf. Definition E.1.1. Then $\mathcal{F} \subseteq \ell^p(S)$ is relatively compact if and only if*

- *$\mathcal{F}$ is pointwise bounded, i.e. $\sup_{f\in\mathcal{F}} |f(s)| < \infty \;\forall s \in S$,*

[2] David Hilbert (1862-1943). German mathematician who worked on many different subjects (consecutively).

- *$\mathcal{F}$ has equi-small tails.*

Proof. $\Rightarrow$ If $\mathcal{F} \subseteq \ell^p(S)$ is totally bounded, then $\overline{\mathcal{F}}$ is compact, so that $\{f(s) \mid f \in \overline{\mathcal{F}}\}$ is bounded for every $s \in S$ by the same argument as in the proof of Theorem 7.7.67. Let $\varepsilon > 0$. Since $\mathcal{F}$ is totally bounded, there are $g_1, \ldots, g_n \in \ell^p(S)$ such that $\ell^p(S) = \bigcup_i B^D(g_i, \varepsilon/2)$. Since $g_i \in \ell^p(S)$, there is a finite $T_i \subseteq S$ such that $\|g_i \cdot \chi_{S\backslash T_i}\|_p < \varepsilon/2$. Take $T = \bigcup_{i=1}^n T_i$. If $f \in \ell^p(S)$ then $f \in B^D(g_i, \varepsilon/2)$ for some i. Clearly, $\|(f - g_i) \cdot \chi_{S\backslash T}\|_p \le \|f - g_i\|_p < \varepsilon/2$, which implies

$$\|f \cdot \chi_{S\backslash T}\|_p \le \|g_i \cdot \chi_{S\backslash T}\|_p + \frac{\varepsilon}{2} \le \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Since $f \in \ell^p(S)$ was arbitrary, $\mathcal{F}$ has equi-small tails.

$\Leftarrow$ Assume $\mathcal{F}$ is pointwise bounded and has equi-small tails. Let $\varepsilon > 0$. By the second assumption, there is a finite $T \subseteq S$ such that $\sup_{f\in\mathcal{F}} \|f \cdot \chi_{S\backslash T}\|_p < \varepsilon$. Write $T = \{t_1, \ldots, t_n\}$ and define $h : \mathcal{F} \to \mathbb{F}^n : f \mapsto (f(t_1), \ldots, f(t_n))$. Now by pointwise boundedness of $\mathcal{F}$, the image $h(\mathcal{F}) \subseteq \mathcal{F}^n$ is bounded, thus totally bounded by Lemma 7.7.47. If $f, g \in \mathcal{F}$ satisfy $\|h(f) - h(g)\|_p = \|(f - g) \cdot \chi_T\|_p < \varepsilon$ then

$$\begin{aligned} \|f - g\|_p &\le \|(f - g) \cdot \chi_T\|_p + \|(f - g) \cdot \chi_{S\backslash T}\| \\ &\le \|(f - g) \cdot \chi_T\|_p + \|f\chi_{S\backslash T}\|_p + \|g\chi_{S\backslash T}\|_p \le \varepsilon + 2\varepsilon = 3\varepsilon. \end{aligned}$$

Thus the assumptions of Lemma 7.7.66 are satisfied, and we obtain total boundedness of $\mathcal{F}$. ■

Remark E.3.3 If S is finite, $\mathcal{F}$ automatically has equi-small tails. Then $\mathcal{F}$ is relatively compact w.r.t. $\|\cdot\|_p$ if and only if it is bounded w.r.t. $\|\cdot\|_\infty$. We could of course have obtained this from the Heine-Borel theorem and the fact that all norms on a finite dimensional space are equivalent to the Euclidean norm. □

Exercise E.3.4 For $g \in \ell^p(S)$, define $\mathcal{F}_g = \{f \in \ell^p(S) \mid |f(s)| \le g(s)\ \forall s \in S\} \subseteq \ell^p(S)$. Prove that $\mathcal{F}_g$ is compact and homeomorphic to the product space $\prod_{s\in S} \overline{B}(0, g(s))$ (which is compact by Tychonov's theorem).

Remark E.3.5 1. In view of $\sum_{n\in\mathbb{N}} 1/n^2 < \infty$, we have as a special case the compactness of $\prod_{n\in\mathbb{N}}[-1/n, 1/n] \subseteq \ell^2(\mathbb{N}, \mathbb{R})$, which is known as the <u>Hilbert cube</u>. (The point of course is not that a countable product of closed intervals is compact, but that the ℓ^2-topology on the Hilbert cube coincides with the product topology.)

2. There actually also is a homeomorphism between $\ell^2(\mathbb{N})$ and the infinite product $\mathbb{R}^\mathbb{N} = \prod_{n\in\mathbb{N}} \mathbb{R}$. For the proof (due to R. D. Anderson (1966)), cf. [169, §6.6]. □

Finally we state a result on compact subsets of $L^p(\mathbb{R}^n)$, the Kolmogorov[3]-Riesz[4]-Fréchet theorem:

Theorem E.3.6 *Let $1 \le p < \infty$. If $\mathcal{F} \subseteq L^p(\mathbb{R}^n)$ then $\overline{\mathcal{F}}$ is compact if and only if*

(i) $\mathcal{F}$ is bounded w.r.t. $\|\cdot\|_p$,

(ii) $\lim_{z\to 0} \sup_{f\in\mathcal{F}} \|f_z - f\|_p = 0$, where $f_z(x) = f(x - z)$,

(iii) $\lim_{r\to\infty} \sup_{f\in\mathcal{F}} \|f \cdot \chi_{\mathbb{R}^n\backslash B(0,r)}\|_p = 0$.

[3] Andrey Nikolaevich Kolmogorov (1903-1987), Russian mathematician.
[4] Marcel Riesz (1886-1969), Hungarian mathematician. Brother of Frigyes Riesz.

Note that now we have two conditions besides pointwise boundedness: (ii) is an L^p-version of equicontinuity. (For $p = \infty$ it reduces to uniform equicontinuity.) Condition (iii) is analogous to the 'equi-small tail' condition above, and is trivially true if all $f \in \mathcal{F}$ are supported in some bounded $\Omega \subseteq \mathbb{R}^n$.

Theorem E.3.6 has been generalized in many ways, for example to an arbitrary locally compact group instead of $\mathbb{R}^n$, cf. [235, §12]. Since the theorem and its generalizations involve Lebesgue integration, discussing them any further would lead us too far afield, cf. [105], [29, Theorem 4.26].

Appendix F

Between topology and functional analysis: $C_0(X, \mathbb{F})$

At several instances, we have met functions from a (locally) compact Hausdorff (or just completely regular) space with values in $\mathbb{R}$ (or $\mathbb{C}$). (Recall the discussion of $C_0(X)$ in Section 7.8.6 and the rôle of $\mathbb{R}$-valued functions in the discussion of completely regular spaces and the Stone-Čech compactification.) In this section we discuss several major results involving $\mathbb{R}$-valued functions, all of which lie on the boundary of point set topology and functional analysis.

F.1 The Stone-Weierstrass theorem

F.1.1 The main result

The following corollary follows immediately from Weierstrass' classical Theorem F.1.10, but in order to be self-contained, we will give a direct proof, which may be skipped without harm.

Corollary F.1.1 *There exists a sequence $\{p_n\}_{n\in\mathbb{N}}$ of real polynomials that converges uniformly on $[0,1]$ to the function $x \mapsto \sqrt{x}$.*

Proof. Inductively define polynomials p_n by $p_0 = 0$ and

$$p_{n+1}(x) = p_n(x) + \frac{x - p_n(x)^2}{2}. \tag{F.1}$$

An easy induction proves that $p_n(x) \leq \sqrt{x}$ for all $n \in \mathbb{N}, x \in [0,1]$. With this and (F.1) we see that $p_{n+1}(x) \geq p_n(x) \geq 0$ for all n and $x \in [0,1]$. Thus $p_n(x)$ converges pointwise to a function $g(x) \leq \sqrt{x}$. Taking $n \to \infty$ in (F.1) gives $0 = x - g(x)^2$, and since $g \geq 0$ we have $g(x) = \sqrt{x}$. Since the convergence $p_n \to g$ is monotonous, it is uniform by Dini's Proposition 7.7.37. ■

Exercise F.1.2 Prove the claim $p_n(x) \leq \sqrt{x}$ for all $n \in \mathbb{N}, x \in [0,1]$.

Theorem F.1.10 says that the polynomials, restricted to $[0,1]$ are uniformly dense in $C([0,1])$. Our aim is to generalize this, replacing replacing $[0,1]$ by (locally) compact Hausdorff spaces. In order to see what should take the place of polynomials, notice that a polynomial on $\mathbb{R}$ is a linear combination of powers x^n, and the latter can be seen as powers f^n (under pointwise multiplication) of the identity function $f = \mathrm{id}_{\mathbb{R}}$. Thus the polynomials are the unital subalgebra $P \subseteq C(\mathbb{R}, \mathbb{R})$ generated by the single element $\mathrm{id}_{\mathbb{R}}$. Now, if X is a topological space and $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ then $C(X, \mathbb{F})$

is a unital algebra, and we will consider subalgebras (not necessarily singly generated) $A \subseteq C(X, \mathbb{F})$. Since the functions on a (locally) compact Hausdorff space separate points, we clearly need to impose the following if we want to prove $\overline{A} = C(X)$:

Definition F.1.3 *A subalgebra $A \subseteq C(X, \mathbb{F})$ separates points if for any $x, y \in X$, $x \neq y$ there is a $f \in A$ such that $f(x) \neq f(y)$.*

Theorem F.1.4 (M. H. Stone 1937) *If X is compact Hausdorff and $A \subseteq C(X, \mathbb{R})$ is a unital subalgebra separating points then $\overline{A} = C(X, \mathbb{R})$.*

Proof. Replacing A by $\overline{A}$, the claim is equivalent to showing that $A = C(X)$. We proceed in several steps. We claim that $f \in A$ implies $|f| \in A$. Since f is bounded by Corollary 7.7.30(i), it clearly is enough to prove this under the assumption $|f| \leq 1$. With the p_n of Corollary F.1.1, we have $(x \mapsto p_n(f^2(x))) \in A$ since A is a unital algebra. Since $p_n \circ f^2$ converges uniformly to $\sqrt{f^2} = |f|$, closedness of A implies $|f| \in A$. In view of

$$\max(f, g) = \frac{f + g + |f - g|}{2}, \qquad \min(f, g) = \frac{f + g - |f - g|}{2},$$

and the preceding result, we see that $f, g \in A$ implies $\min(f, g), \max(f, g) \in A$. By induction, this extends to pointwise minima/maxima of finite families of elements of A.

Now let $f \in C(X)$. Our goal is to find $f_\varepsilon \in A$ satisfying $\|f - f_\varepsilon\| < \varepsilon$ for each $\varepsilon > 0$. Since A is closed, this will give $A = C(X)$.

If $a \neq b$, the fact that A separates points gives us an $h \in A$ such that $h(a) \neq h(b)$. Thus the function $h_{a,b}(x) = \frac{h(x) - h(a)}{h(b) - h(a)}$ is in A, continuous and satisfies $h(a) = 0$, $h(b) = 1$. Thus also $f_{a,b}(x) = f(a) + (f(b) - f(a))h_{a,b}(x)$ is in A, and it satisfies $f_{a,b}(a) = f(a)$ and $f_{a,b}(b) = f(b)$. This implies that the sets

$$U_{a,b,\varepsilon} = \{x \in X \mid f_{a,b}(x) < f(x) + \varepsilon\}, \qquad V_{a,b,\varepsilon} = \{x \in X \mid f_{a,b}(x) > f(x) - \varepsilon\}$$

are open neighborhoods of a and b, respectively, for every $\varepsilon > 0$. Thus keeping b, ε fixed, $\{U_{a,b,\varepsilon}\}_{a \in X}$ is an open cover of X, and by compactness we find a finite subcover $\{U_{a_i,b,\varepsilon}\}_{i=1}^n$. By the above preparation, the function $f_{b,\varepsilon} = \min(f_{a_1,b,\varepsilon}, \ldots, f_{a_n,b,\varepsilon})$ is in A. If $x \in U_{a_i,b,\varepsilon}$ then $f_{b,\varepsilon}(x) \leq f_{a_i,b,\varepsilon}(x) < f(x) + \varepsilon$ for all $x \in X$, and since $\{U_{a_i,b,\varepsilon}\}_{i=1}^n$ covers X, we have $f_{b,\varepsilon}(x) < f(x) + \varepsilon \; \forall x$. For all $x \in V_{b,\varepsilon} = \bigcap_{i=1}^n V_{a_i,b,\varepsilon}$, we have $f_{a_i,b,\varepsilon}(x) > f(x) - \varepsilon$, and therefore $f_b(x) = \min_i(f_{a_i,b,\varepsilon}) > f(x) - \varepsilon$. Now $\{V_{b,\varepsilon}\}_{b \in X}$ is an open cover of X, and we find a finite subcover $\{V_{b_j,\varepsilon}\}_{j=1}^n$. Then $f_\varepsilon = \max(f_{b_1,\varepsilon}, \ldots, f_{b_n})$ is in A. Now $f_\varepsilon(x) = \max_j(f_{b_j,\varepsilon}) \leq f(x) + \varepsilon$ holds everywhere, and for $x \in V_{b_j,\varepsilon}$ we have $f_\varepsilon(x) \geq f_{b_j,\varepsilon} > f(x) - \varepsilon$. Since $\{V_{b_j,\varepsilon}\}_j$ covers X, we conclude that $f_\varepsilon(a) \in (f(x) - \varepsilon, f(x) + \varepsilon)$ for all x, to wit $\|f - f_\varepsilon\| < \varepsilon$. ■

F.1.2 Generalizations

Having proven Theorem F.1.4, it is easy to generalize it to locally compact spaces or/and subalgebras of $C_{(0)}(X, \mathbb{C})$.

Definition F.1.5 *A subalgebra $A \subseteq C(X, \mathbb{C})$ is selfadjoint if $f \in A$ implies $f^* \in A$, where $f^*(x) := \overline{f(x)}$.*

Corollary F.1.6 *If X is compact Hausdorff and $A \subseteq C(X, \mathbb{C})$ is a selfadjoint unital subalgebra separating points then $\overline{A} = C(X, \mathbb{C})$.*

Proof. Define $B = A \cap C(X, \mathbb{R})$. Let $f \in A$. Since $f^* \in A$, we also have $\mathrm{Re}(f) = \frac{f+f^*}{2} \in B$ and $\mathrm{Im}(f) = \frac{f-f^*}{2i} = -\mathrm{Re}(if) \in B$. Thus $A = B + iB$. It is obvious that $\mathrm{Re}(A) \subseteq C(X, \mathbb{R})$ is a unital subalgebra. If $x \neq y$ then there is $f \in C(X, \mathbb{C})$ such that $f(x) \neq f(y)$. Thus $\mathrm{Re}(f)(x) \neq \mathrm{Re}(f)(y)$ or $\mathrm{Re}(if)(x) \neq \mathrm{Re}(if)(y)$ (or both). Since $\mathrm{Re}(f), \mathrm{Re}(if) \in B$, we see that B separates points. Thus $\overline{B} = C(X, \mathbb{R})$ by Theorem F.1.4, implying $\overline{A} = \overline{B + iB} = \overline{B} + i\overline{B} = C(X, \mathbb{R}) + iC(X, \mathbb{R}) = C(X, \mathbb{C})$. ■

Definition F.1.7 *A subalgebra $A \subseteq C_0(X, \mathbb{F})$ vanishes at no point if for every $x \in X$ there is an $f \in A$ such that $f(x) \neq 0$.*

Corollary F.1.8 *If X is locally compact Hausdorff and $A \subseteq C_0(X)$ is a subalgebra separating points and vanishing at no point then $\overline{A} = C_0(X)$.*

Proof. Recall that every $f \in C_0(X)$ extends to $\widehat{f} \in C(X_\infty)$ with $\widehat{f}(\infty) = 0$. Then $B = \{\widehat{f} \mid f \in A\} + \mathbb{R}\mathbf{1}$ clearly is a unital subalgebra of $C(X_\infty)$. We claim that B separates the points of X_∞. This is obvious for $x, y \in X$, $x \neq y$ since already A does that. Now let $x \in X$. Since A vanishes at no point, there is $f \in A$ such that $f(x) \neq 0$. Let $\widehat{f} \in C(X)$ be the extension to X_∞ with $\widehat{f}(\infty) = 0$, cf. Exercise 7.8.83. In view of $\widehat{f}(x) = f(x) \neq 0$, we see that B also separates ∞ from the points of X, so that Theorem F.1.4 gives $\overline{B} = C(X)$. In view of $\overline{B} = \overline{A} + \mathbb{R}\mathbf{1}$ and $C(X) \upharpoonright X = C_0(X)$, we have $\overline{A} = \overline{B} \upharpoonright X = C_0(X)$. ■

Corollary F.1.9 *If X is locally compact Hausdorff and $A \subseteq C_0(X, \mathbb{C})$ is a self-adjoint subalgebra separating points and vanishing at no point then $\overline{A} = C_0(X, \mathbb{C})$.*

Proof. The proof just combines the ideas of the proofs of Corollaries F.1.6 and F.1.8. ■

F.1.3 Applications

We discuss some applications of Theorem F.1.4 and its corollaries.

Theorem F.1.10 (Weierstrass 1885) [1] *If $f \in C([0, 1], \mathbb{R})$ and $\varepsilon > 0$ then $\|f - p\| < \varepsilon$ for some real polynomial p, where $\|f\| = \sup_{x \in [0,1]} |f(x)|$.*

Proof. The real polynomials are an algebra, and they clearly separate the points. Now apply Theorem F.1.4. ■

Definition F.1.11 *Put $S^1 \equiv \{z \in \mathbb{C} \mid |z| = 1\}$. A trigonometric polynomial is a finite linear combination of the functions $S^1 \to \mathbb{C}$, $z \mapsto z^n$, where $n \in \mathbb{Z}$.*

Corollary F.1.12 *The trigonometric polynomials are dense in $C(S^1, \mathbb{C})$.*

Proof. The trigonometric polynomials evidently form a unital subalgebra $A \subseteq C(S^1, \mathbb{C})$. For $z \in S^1$ we have $(z^n)^* = \overline{z^n} = z^{-n}$, thus A is self-adjoint. Since A contains the identity map z^1, it separates the points of S^1. Now apply Corollary F.1.6. ■

Proposition F.1.13 *Let X be compact Hausdorff. Then the following are equivalent:*

(i) The metric space $(C(X), D)$ is second countable ($\Leftrightarrow$ separable).

[1] Karl Theodor Wilhelm Weierstrass (1815-1897). German mathematician.

(ii) X is second countable ($\Leftrightarrow$ metrizable).

Proof. (i)$\Rightarrow$(ii) Since $(C(X),D)$ is metric, second countability and separability are equivalent. Let $F \subseteq C(X)$ be a subset that is dense w.r.t. τ_D, i.e. uniformly. Let $C \subseteq X$ be closed and $x \in X\backslash C$. Since X is completely regular, there is an $f \in C(X)$ such that $f(x) \not\in \overline{f(C)}$. Thus $r = \mathrm{dist}(f(x), f(C)) > 0$. Since $F \subseteq C(X)$ is dense, we can find $f' \in F$ such that $\|f - f'\| < r/3$. Then it is immediate that $\mathrm{dist}(f'(x), f'(C)) > 0$, so that $f'(x) \not\in \overline{f'(C)}$. Thus F separates points from closed sets. This means that

$$\iota_F : X \to \prod_{f\in F}[\inf f, \sup f], \quad x \mapsto \prod_{f\in F} f(x)$$

is an embedding. Thus if F is countable then $X \cong \iota_F(X) \subseteq \prod_{f\in F}[\inf f, \sup f]$ is second countable and metrizable.

(ii)$\Rightarrow$(i) As in the proof of Urysohn's Metrization Theorem 8.2.33, we construct a countable family $F_1 \subseteq C(X,[0,1])$ separating points from closed sets. Let F_2 denote the set of all finite products of elements of F_1, which is clearly countable. Interpreting the empty product as the function $\mathbf{1}$, we have $\mathbf{1} \in F_2$. Then also the set F_3 of finite linear combinations of elements of F_2 with $\mathbb{Q}$-coefficients is countable. Since $A = \overline{F_3}$ contains the finite linear combinations of elements of F_2 with coefficients in $\mathbb{R}$, it is a unital $\mathbb{R}$-algebra. Since already F_1 separates the points of X, the same holds for A. Thus the Stone-Weierstrass theorem gives $C(X) = A = \overline{F_3}$. Thus $C(X)$ has F_3 as countable dense subset. ■

Corollary F.1.14 *If X is locally compact Hausdorff then the spaces $X, X_\infty, (C(X_\infty), D), (C_0(X), D)$ are either all second countable or they all are not.*

Proof. Second countablility of X_∞ obviously implies that of X, and the converse was proven in Exercise 7.8.44(ii). Second countability of X_∞ and $(C(X_\infty), D)$ are equivalent by the preceding result. Second countability of $C(X_\infty) \cong C_0(X) \oplus \mathbb{R}$ implies second countability of $C_0(X)$. Finally, If $F \subseteq C_0(X)$ is dense then $\{\widehat{f} \mid f \in F\} + \mathbb{Q}\mathbf{1} \subseteq C(X_\infty)$ is dense. ■

Remark F.1.15 Results analogous to the above hold for the algebras of complex-valued functions. The only change in the proofs consists in replacing $\mathbb{Q}$ by $\mathbb{Q} + i\mathbb{Q}$. □

F.2 Weak Gelfand duality: Characters of $C_0(X,\mathbb{F})$

If (X,τ) is compact Hausdorff and $\mathbb{F} \in \{\mathbb{R},\mathbb{C}\}$ then $C(X,\mathbb{F})$ is a commutative algebra. In this appendix we will show that (X,τ) can be recovered from $C(X,\mathbb{F})$. We begin by observing some properties of $C(X,\mathbb{F})$.

If $\mathbb{F} = \mathbb{C}$, the algebra $C(X,\mathbb{C})$ has an involution $f \mapsto \overline{f}$ given by pointwise complex conjugation. Instead of $\overline{f}$ we also write f^*. It is clear that $(f+g)^* = f^* + g^*, (fg)^* = f^*g^*$ and $(cf)^* = \overline{c}f^*$ for all $f,g \in C(X,\mathbb{C}), c \in \mathbb{C}$.

We know that $C(X,\mathbb{F})$ is complete w.r.t. the norm $\|f\| = \sup_x |f(x)|$, thus a Banach space. Furthermore, one has

$$\|fg\| = \sup_{x\in X} |f(x)g(x)| \le \sup_{x\in X} |f(x)| \cdot \sup_{x\in X} |g(x)| = \|f\| \cdot \|g\|,$$

$$\|f^*f\| = \sup_{x\in X} |f(x)|^2 = \left(\sup_{x\in X} |f(x)|\right)^2 = \|f\|^2.$$

A Banach space with algebra structure such that $\|fg\| \leq \|f\| \cdot \|g\|$ is called a Banach algebra. And a complex Banach algebra with an involution satisfying $\|f^f f\| = \|f\|^2$ is called a C^*-algebra. Thus: $C(X, \mathbb{F})$ are commutative unital C^*-algebras. (For $\mathbb{F} = \mathbb{R}$ take $f^* = f$.)

Definition F.2.1 *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. A character on an an $\mathbb{F}$-algebra A is a an algebra homomorphism $\phi : A \to \mathbb{F}$ (i.e. a linear map that is also multiplicative $\phi(ab) = \phi(a)\phi(b)$).*

Definition F.2.2 *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. An ideal in an $\mathbb{F}$-algebra A is a linear subspace $I \subseteq A$ such that $x \in I,\ y \in A \Rightarrow xy, yx \in I$. An ideal $I \subseteq A$ is maximal if $I \neq A$ and every ideal containing I equals either I or A.*

Lemma F.2.3 *Let (X, τ) be compact Hausdorff and $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$.*

(i) If $x \in X$ then $\phi_x : C(X, \mathbb{F}) \to \mathbb{F},\ f \mapsto f(x)$ is a non-zero character.

(ii) If $\phi : C(X, \mathbb{F}) \to \mathbb{F}$ is a non-zero character then $I_\phi = \ker(\phi) = \{f \in C(X, \mathbb{F}) \mid \phi(f) = 0\}$ is a maximal ideal.

Proof. (i) Follows from $\phi_x(f+g) = (f+g)(x) = f(x)+g(x) = \phi_x(f)+\phi_x(g)$ and $\phi_x(fg) = (fg)(x) = f(x)g(x) = \phi_x(f)\phi_x(g)$. That ϕ_x is non-zero follows from $\phi_x(\mathbf{1}) = 1$, where $\mathbf{1}$ is the constant function $x \mapsto 1$.

(ii) Defining $I_\phi = \{f \in C(X, \mathbb{F}) \mid \phi(f) = 0\}$, it is immediate that I_ϕ is an ideal. Since $\phi \not\equiv 0$, there is an f such that $\phi(f) \neq 0$. Then $\phi(\mathbb{F}f) = \mathbb{F}$, thus $\phi : A \to \mathbb{F}$ is a surjective ring homomorphism. From ring theory, one has that $\mathbb{F} \cong A/I_\phi$, and the fact that $\mathbb{F}$ is a field implies that I_ϕ is maximal. ■

Proposition F.2.4 *Let (X, τ) be compact Hausdorff and $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. For $x \in X$ define $M_x := \ker(\phi_x)$. Then for every maximal ideal $M \subseteq C(X, \mathbb{F})$ there is unique $x_M \in X$ such that $M = M_{x_M}$.*

Proof. Uniqueness: If $x \neq y$, Urysohn's lemma provides an $f \in C(X, \mathbb{F})$ such that $f(x) = 1,\ f(y) = 0$. Then $f \in \ker(\phi_y)$, but $f \notin \ker(\phi_x)$. Thus $\ker(\phi_x) \neq \ker(\phi_y)$, so that M can equal ϕ_x for at most one $x \in X$.

Existence: Assume $M \neq M_x$ for all $x \in X$. Since the ideals M, M_x are maximal, this implies $M \not\subseteq M_x$ (and $M_x \not\subseteq M$) for all $x \in X$. Thus for each $x \in X$ there exists $f_x \in M \backslash M_x$, thus $f_x(x) \neq 0$. Defining $U_x = \{y \in X \mid f_x(y) \neq 0\}$, we have $x \in U_x\ \forall x$ so that $\{U_x\}_{x \in X}$ is an open cover of X. By compactness, there are $x_1, \ldots, x_n$ such that $X = \bigcup_i U_{x_i}$. Defining $f(x) = \sum_{i=1}^n f_{x_i}(x)\overline{f_{x_i}(x)}$, we have $f \in M$ and $f(x) \neq 0\ \forall x$ (since each x is contained in some U_{x_i} and $|f| \geq |f_{x_i}|^2 > 0$ on U_{x_i}). Thus $x \mapsto 1/f(x)$ is in $C(X, \mathbb{F})$, so that $\mathbf{1} = f\frac{1}{f} \in M$. But an ideal containing $\mathbf{1}$ equals $C(X, \mathbb{F})$, which the maximal ideal M cannot do. This contradiction proves $M = M_x$ for some $x \in X$. ■

Theorem F.2.5 *Let X be compact Hausdorff space and $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. Then there are mutually consistent bijections between*

- *points of $x \in X$,*
- *non-zero characters $\phi : C(X, \mathbb{F}) \to \mathbb{F}$,*
- *maximal ideals $M \subseteq C(X, \mathbb{F})$.*

Proof. We have the maps $x \mapsto \phi_x$, $\phi \mapsto \ker(\phi)$ from Lemma F.2.3 and $M \mapsto x_M$ from Proposition F.2.4. If $M = M_x = \ker(\phi_x)$ then the uniqueness in the proposition implies $x_M = x$. Thus the circle closes and the bijections are mutually consistent. ■

Remark F.2.6 Let $\mathbb{F}=\mathbb{C}$. For every $x\in X$ we have $\phi_x(f^*)=f^*(x)=\overline{f(x)}=\phi_x(f)^*$, thus the characters $\phi_x : f\mapsto f(x)$ by definition are $*$-homomorphisms, i.e. satisfy $\phi_x(f^*)=\phi_x(f)^*$ for all f. Now Proposition F.2.4 implies that every character on $C(X,\mathbb{C})$ automatically is a $*$-homomorphism, which is not quite obvious a priori. □

Remark F.2.7 For a completely regular space X we can define the Stone-Čech compactification βX. By Theorem F.2.5, the points of βX are in bijective correspondence with the characters of $C(\beta X,\mathbb{R})$. Combining this with the algebra isomorphism $C_b(X,\mathbb{R})\to C(\beta X,\mathbb{R})$ obtained in Corollary 8.3.38, we see that the points of βX are in bijective correspondence to the characters on $C_b(X)$. This can indeed be taken as the definition of βX, provided one equips the set of characters of $C_b(X,\mathbb{R})$ with a suitable topology. □

Remark F.2.8 If X is locally compact Hausdorff, we have a compact Hausdorff topology on $X_\infty = X\cup\{\infty\}$ such that X_∞ is a compactification on X. In Exercise 7.8.83(iv) we have seen that there is an algebra isomorphism $C_0(X,\mathbb{R})\oplus\mathbb{R}\to C(X_\infty,\mathbb{R}),\ (f,c)\mapsto\widehat{f}+c\mathbf{1}$, where $\widehat{f}$ is the extension via $\widehat{f}(\infty)=0$. By Theorem F.2.5, there is a bijection between the points of X_∞ and the characters of $C(X_\infty,\mathbb{R})$. Thus we can interpret X_∞ as the (suitably topologized) character space of the minimal unitization $C_0(X,\mathbb{R})\oplus\mathbb{R}$ of $C_0(X,\mathbb{R})$. (But taking this as the definition of X_∞ would be quite ridiculuous.) □

Remark F.2.9 In this section we have shown that a compact Hausdorff space X can be recovered from the (commutative unital) C^*-algebra $C(X,\mathbb{F})$ by $X\cong\Omega(C(X,\mathbb{F}))$. One can in fact show much more: For every commutative unital C^*-algebra A over $\mathbb{F}$ there is a unique (up to homeomorphism) compact Hausdorff space X such that $A\cong C(X,\mathbb{F})$. Furthermore, this extends to a contravariant equivalence of the categories of compact Hausdorff spaces and commutative unital C^*-algebras. We cannot go further into this since it would lead too deep into the theory of C^*-algebras, but cf. e.g. [181]. □

Appendix G

Normed spaces: Applications of topology

G.1 Linear maps: Boundedness and continuity

Definition G.1.1 *Let E, F be normed spaces and $A : E \to F$ a linear map. Then the norm $\|A\| \in [0, \infty]$ is defined by*

$$\|A\| = \sup_{0 \neq e \in E} \frac{\|Ae\|}{\|e\|}.$$

If $\|A\| < \infty$ then A is called <u>*bounded*</u>*. The set of bounded linear maps from E to F is denoted $B(E, F)$.*

Remark G.1.2 It is easy to see that $A \mapsto \|A\|$ is a norm in the sense of Definition 2.1.10. Furthermore, by Proposition 3.1.8, $B(E, F)$ is a Banach space whenever F is a Banach space. □

Lemma G.1.3 *Let E, F be normed spaces and $A : E \to F$ a linear map. Then the following are equivalent:*

(i) A is bounded.

(ii) A is continuous (w.r.t. the norm topologies).

(iii) A is continuous at $0 \in E$.

Proof. (i)⇒(ii) For $x, y \in E$ we have $\|Ax - Ay\| = \|A(x - y)\| \leq \|A\| \, \|x - y\|$. Since $\|A\| < \infty$, continuity of E follows. (ii)⇒(iii) This is obvious.

(iii)⇒(i) By continuity at 0, there is $C > 0$ such that $A(B^E(0, C)) \subseteq B^F(0, 1)$. By linearity of A and the properties of the norm, this is equivalent to $A(B^E(0, 1)) \subseteq B^F(0, D)$, where $D = 1/C$. If $0 \neq x \in E$ then

$$Ax = 2\|x\| \, A\left(\frac{x}{2\|x\|}\right),$$

thus $\|Ax\| \leq 2\|x\|\|A(x/2\|x\|)\| < 2\|x\|D$, and A is bounded. ■

G.2 Applications of Baire's theorem

G.2.1 Uniform boundedness

Definition G.2.1 *Let E, F be normed spaces and $\mathcal{A} \subseteq B(E, F)$ a family of bounded linear maps.*

(i) $\mathcal{A}$ *is called* <u>*pointwise bounded*</u> *if* $\sup_{A\in\mathcal{A}} \|Ax\| < \infty$ *for each* $x \in E$.

(ii) $\mathcal{A}$ *is called* <u>*uniformly bounded*</u> *if* $\sup_{A\in\mathcal{A}} \|A\| < \infty$.

It is obvious that uniform boundedness implies pointwise boundedness. In fact, much more is true:

Theorem G.2.2 (Banach-Steinhaus Uniform Boundedness Theorem) [1] *Let E be a Banach space, F a normed space and $\mathcal{A} \subseteq B(E,F)$. Then exactly one of the following alternatives holds:*

(i) $\mathcal{A}$ *is uniformly bounded.*

(ii) There is a dense G_δ-set $X \subseteq E$ such that $\sup_{A\in\mathcal{A}} \|Ax\| = \infty$ for every $x \in X$.

Proof. It is clear that (i) and (ii) exclude each other. Thus we must show that at least one of the two statements holds.

The map $F \to \mathbb{R}_{\geq 0},\ x \mapsto \|x\|$ is continuous and each $A \in \mathcal{A}$ is bounded, thus continuous. Therefore the map $f_A : E \to \mathbb{R}_{\geq 0},\ x \mapsto \|Ax\|$ is continuous for every $A \in \mathcal{A}$. Defining

$$V_n = \{x \in E \mid \sup_{A\in\mathcal{A}} \|Ax\| > n\},$$

the definition of sup implies

$$V_n = \{x \in E \mid \exists A \in \mathcal{A} : \|Ax\| > n\} = \bigcup_{A\in\mathcal{A}} \{x \in E \mid \|Ax\| > n\} = \bigcup_{A\in\mathcal{A}} f_A^{-1}((n,\infty)),$$

which is open by continuity of the f_A. If each V_n is dense in E then Baire's Theorem 3.3.1 implies that the G_δ-set $X = \bigcap_n V_n$ is dense. In view of the definition of V_n it is obvious that $\sup_{A\in\mathcal{A}} \|Ax\| = \infty$ for each $x \in X$, thus (ii) holds.

Now assume that V_n is non-dense for some n. Thus there exists $x_0 \in E$ and $r > 0$ such that $B(x_0, r) \cap V_n = \emptyset$. This means $\sup_{A\in\mathcal{A}} \|A(x_0+x)\| \leq n$ for all x with $\|x\| < r$. With $x = (x_0+x) - x_0$ and the triangle inequality we have

$$\|Ax\| \leq \|A(x_0+x)\| + \|Ax_0\| \leq 2n \quad \forall A \in \mathcal{A}, x \in B(0,r).$$

This implies $\|A\| \leq 2n/r\ \forall A \in \mathcal{A}$, thus uniform boundedness (i). ■

The Banach-Steinhaus theorem is usually stated, proved and used in the following weaker form, which is an obvious corollary of Theorem G.2.2:

Corollary G.2.3 *Let E be a Banach space, F a normed space and $\mathcal{A} \subseteq B(E,F)$. If $\mathcal{A}$ is pointwise bounded then $\mathcal{A}$ is uniformly bounded.*

Occasionally, it is claimed that the above proof of the Uniform Boundedness Theorem is "non-elementary" due to its reliance on Baire's theorem. This claim does not make much sense since (a) Baire's theorem is about complete metric spaces, hardly a deeper concept than that of Banach spaces, and (b) its proof involves nothing but countable dependent choice, which also the supposedly elementary proofs cannot avoid. Nevertheless, it may be interesting to see that the weaker Corollary G.2.3 admits a fairly short proof:

[1] Proven in 1927 by the Polish mathematicians Stefan Banach (1892-1945) and Hugo Steinhaus (1887-1972).

Lemma G.2.4 *Let E, F be normed spaces and $A : E \to F$ a bounded linear map. Then for any $x \in E$ and $r > 0$ one has:*

$$\sup_{y \in B(x,r)} \|Ay\| \geq r\,\|A\|.$$

Proof. Obviously, $Ay = \frac{1}{2}(A(x+y) - A(x-y))$. Using the triangle inequality gives

$$\|Ay\| = \frac{1}{2}\|(A(x+y) - A(x-y))\| \leq \frac{1}{2}(\|A(x+y)\| + \|A(x-y)\|) \leq \max(\|A(x+y)\|, \|A(x-y)\|).$$

Now the claim follows by taking the supremum over $y \in B(0,r)$ and observing

$$\sup_{y \in B(0,r)} \|Ay\| = r\,\|A\| \qquad \text{and} \qquad \sup_{y \in B(0,r)} \max(\|A(x+y)\|, \|A(x-y)\|) = \sup_{y \in B(x,r)} \|Ay\|.$$

■

Alternative proof of Corollary G.2.3. If $\mathcal{A}$ is not uniformly bounded, we can find a sequence $\{A_n\}$ in $\mathcal{A}$ such that $\|A_n\| \geq 4^n$. Put $x_0 = 0$, and for $n \geq 1$ use the lemma to choose $x_n \in E$ inductively in such a way that

$$\|x_n - x_{n-1}\| \leq 3^{-n} \qquad \text{and} \qquad \|A_n x_n\| \geq \frac{2}{3} 3^{-n} \|A_n\|.$$

Then the sequence $\{x_n\}$ is Cauchy, thus converges to some $x \in E$ by completeness. More precisely,

$$\|x - x_n\| = \|\sum_{m=n}^{\infty} x_{m+1} - x_m\| \leq \sum_{m=n}^{\infty} 3^{-(m+1)} = 3^{-(n+1)} \frac{1}{1 - \frac{1}{3}} = \frac{1}{2} 3^{-n}.$$

With the triangle inequality, we have

$$\|A_n x_n\| = \|A_n(x - x + x_n)\| \leq \|A_n x\| + \|A_n(x - x_n)\| \leq \|A_n x\| + \|A_n\|\|x - x_n\|,$$

thus

$$\|A_n x\| \geq \|A_n x_n\| - \|A_n\|\|x - x_n\| \geq \|A_n\| \left(\frac{2}{3} 3^{-n} - \frac{1}{2} 3^{-n}\right) = \frac{1}{6} 3^{-n} \|A_n\| \geq \frac{1}{6}(4/3)^n \to \infty,$$

contradicting the pointwise boundedness of $\mathcal{A}$. ■

Remark G.2.5 The above proof is from [212]. It is quite close in spirit to the original proof by Banach and Steinhaus, which used the 'gliding hump' method. The latter goes back the the methods that were used to produce continuous functions whose Fourier series diverges in some point. Cf. the end of this section.

It should be obvious that also the above proof used the Axiom of Countable Dependent Choice! Since the latter is strictly equivalent to Baire's theorem, the alternative proof has gained us nothing from a foundational point of view. □

G.2.2 Open mappings and closed graphs

The following is a restatement of Lemma 8.2.21:

Lemma G.2.6 *Let E be a Banach space, F a normed space (real or complex) and $T : E \to F$ a linear map. Assume also that there are $m > 0$ and $r \in (0,1)$ such that for every $y \in F$ there is an $x_0 \in E$ with $\|x_0\|_E \leq m\|y\|_F$ and $\|y - Tx_0\|_F \leq r\|y\|_F$. Then for every $y \in F$ there is an $x \in E$ such that $\|x\|_E \leq \frac{m}{1-r}\|y\|_F$ and $Tx = y$. In particular, T is surjective.*

Corollary G.2.7 *Let E, F be Banach spaces and $T \in B(E, F)$ such that $T(B^E(0, \alpha))$ is dense in $B^F(0, \beta)$ with $\alpha, \beta > 0$. Then $B^F(0, \beta') \subseteq T(B^E(0, \alpha))$ for each $\beta' \in (0, \beta)$.*

Remark G.2.8 Here "A is dense in B", where $A, B \subseteq X$, means $\overline{A} \supseteq B$. Note that in general this is not equivalent to $\overline{A \cap B} = B$ since the left hand side can easily be empty even if A is dense in X. But when B is open, as is the case here, the two notions are equivalent. Recall Lemma 2.7.10(i). □

Proof. By homogeneity of the norm (cf. Definition 2.1.10), $x \mapsto \lambda x$ is a homeomorphism for every $\lambda > 0$ so that it is sufficient to consider $\beta = 1$. If $T(B^E(0, \alpha))$ is dense in $B^F(0, 1)$ then $T(B^E(0, \alpha'))$ is dense in $B^F(0, 1)$ for every $\alpha' \geq \alpha$. This in turn is equivalent to the statement that for every $y \in F$ and $\varepsilon > 0$ there exists an $x \in E$ such that $\|y - Tx\| \leq \varepsilon\|y\|$ and $\|x\| \leq \alpha\|y\|$. Now the preceding lemma gives for every $y \in F$ an $x \in E$ with $Tx = y$ and $\|x\| \leq \|y\|\alpha/(1 - \varepsilon)$. Thus $B^F(0, 1)$ is contained in $T(B^E(0, \alpha'))$ for every $\alpha' > \alpha$ which, again by homogeneity implies the claim. ■

Theorem G.2.9 (Open Mapping Theorem) *(Banach-Schauder) Let E, F be Banach spaces and let $T \in B(E, F)$ be surjective. Then T is an open map. (I.e. $TX \subseteq F$ is open for every open $X \subseteq E$.)*

Proof. Since T is surjective, we have

$$F = T(E) = \bigcup_{n=1}^{\infty} \overline{T(B^E(0, n))}.$$

Since F is a complete metric, thus Baire, and trivially has non-empty interior, Proposition 3.3.5 implies that at least one of the closed sets $\overline{T(B^E(0, n))}$ has non-empty interior. Thus there are $n \in \mathbb{N}, y \in F, \varepsilon > 0$ such that $B^F(y, \varepsilon) \subseteq \overline{T(B^E(0, n))}$. In view of $B^F(y, \varepsilon) = y + B^F(0, \varepsilon)$, we have $2B^F(0, \varepsilon) \subseteq B^F(y, \varepsilon) - B^F(y, \varepsilon)$ and thus

$$B^F(0, \varepsilon) \subseteq \frac{1}{2}(B^F(y, \varepsilon) - B^F(y, \varepsilon)) \subseteq \frac{1}{2}(\overline{T(B^E(0, n))} - \overline{T(B^E(0, n))}) \subseteq \overline{T(B^E(0, n))}.$$

Now Corollary G.2.7 implies that $B^F(0, \varepsilon') \subseteq T(B^E(0, n))$ for every $\varepsilon' \in (0, \varepsilon)$. Appealing to homogeneity again, we see that the image of every open ball around $0 \in E$ contains an open ball around $0 \in F$, and using linearity we conclude that T is open. ■

Corollary G.2.10 *If E, F are Banach spaces and $T : E \to F$ is linear, bounded and bijective then also T^{-1} is bounded. (Thus T is a homeomorphism.)*

Let E, F be Banach spaces and $E \times F$ the (set-theoretic) product. (This can be considered as the direct sum $E \oplus F$, but we don't need the vector space structure.) Then $\|(x, y)\|_1 = \max(\|x\|, \|y\|)$ and $\|(x, y)\|_2 = \|x\| + \|y\|$ are equivalent norms since $\|(x, y)\|_2 \leq 2\|(x, y)\|_1 \leq 2\|(x, y)\|_2$.

Theorem G.2.11 (Closed Graph Theorem) *Let E, F be Banach spaces and $T : E \to F$ a linear map such that the graph*

$$\mathfrak{G}(T) = \{(x, Tx) \mid x \in E\} \;\subseteq\; E \times F$$

is closed (where $E \times F$ is topologized by one of the above equivalent norms). Then T is bounded.

Proof. By assumption, $\mathfrak{G}(T) \subseteq E \times F$ is closed, thus a Banach space. The projection $p_1 : \mathfrak{G}(T) \to E$ obviously is linear, bijective and norm-decreasing, thus bounded. Therefore by Corollary G.2.10 it has a bounded inverse $p_1^{-1} : E \to \mathfrak{G}(T)$. It is clear that $p_2 \circ p_1^{-1} = T$, and since also $p_2 : \mathfrak{G}(T) \to F$ is trivially bounded, we conclude that T is bounded. ■

Remark G.2.12 1. One can also deduce Corollary G.2.3, i.e. the weak form of the Uniform Boundedness theorem, from the Closed Graph Theorem, cf. [189]. But this seems rather indirect. Furthermore, it is unlikely that one can also obtain Theorem G.2.2 in this way.

2. There also is a (less well known) open mapping theorem for topological groups, cf. [116, Vol. I, Theorem 5.29]. Its proof again involves Baire's theorem. □

G.2.3 Continuous functions with divergent Fourier series

Let $f : \mathbb{R} \to \mathbb{C}$ be 2π-periodic, i.e. $f(x+2\pi) = f(x)\ \forall x$, and integrable over finite intervals. Define

$$c_n(f) = \frac{1}{2\pi} \int_0^{2\pi} f(x) e^{-inx}\, dx \tag{G.1}$$

and

$$S_n(f)(x) = \sum_{k=-n}^{n} c_k(f) e^{ikx}, \quad n \in \mathbb{N}. \tag{G.2}$$

The fundamental problem of the theory of Fourier series is to find conditions for the convergence $S_n(f)(x) \to f(x)$ as $n \to \infty$, where convergence can be understood as (possibly almost) everywhere pointwise or w.r.t. some norm, like the uniform one. Here we will discuss only continuous functions and we identify continuous 2π-periodic functions with continuous functions on S^1. It is not hard to show that $S_n(f)(x) \to f(x)$ if f is differentiable at x and that convergence is uniform when $f \in C^1(S^1)$. Our aim here is to show that continuity of f is not enough:

Theorem G.2.13 *There is a dense G_δ-subset $X \subseteq C(S^1)$ w.r.t. the norm $\|\cdot\|_\infty$ such that $\lim_n S_n(f)(0)$ does not exist for any $f \in X$.*

Proof. Inserting (G.1) into (G.2) we obtain

$$S_n(f)(x) = \frac{1}{2\pi} \sum_{k=-n}^{n} e^{ikx} \int_0^{2\pi} f(t) e^{-ikt}\, dx = \frac{1}{2\pi} \int_0^{2\pi} f(t) \left(\sum_{k=-n}^{n} e^{ik(x-t)} \right) dx = (D_n \star f)(x),$$

where $\star$ denotes convolution $[(f \star g)(x) = (2\pi)^{-1} \int_0^{2\pi} f(t) g(x-t) dt]$ and

$$D_n(x) := \sum_{k=-n}^{n} e^{ikx} = \frac{\sin(n+\frac{1}{2})x}{\sin\frac{x}{2}}$$

is the Dirichlet kernel. The quickest way to check the last identity is:

$$(e^{ix/2} - e^{-ix/2}) D_n(x) = \sum_{k=-n}^{n} e^{ix(k+1/2)} - \sum_{k=-n}^{n} e^{ix(k-1/2)} = e^{ix(n+1/2)} - e^{-ix(n+1/2)}.$$

Since $D_n(x)$ is an even function, we have

$$\varphi_n(f) := S_n(f)(0) = (2\pi)^{-1} \int_0^{2\pi} f(x) D_n(x) dx.$$

For $g(x) = \mathrm{sgn}(D_n(x))$ we have $\varphi_n(g)(0) = (2\pi)^{-1} \int_0^{2\pi} |D_n(x)| dx =: \|D_n\|_1$. It is now quite clear that the norm of the map $\varphi_n : C(S^1) \to \mathbb{C}$ is bounded above by $\|D_n\|_1$. If the above g was continuous, we

could conclude directly that $\|S_n(\cdot)(0)\| \geq \|D_n\|_1$. Now, the above g is not continuous, but we can find a sequence of continuous g_n such that $g_n \to g$ pointwise. Now Lebesgue's dominated convergence theorem implies $\|\varphi_n\| = \|D_n\|_1$. We will soon prove that $\|D_n\|_1 \to \infty$ as $n \to \infty$. Thus the family $\mathcal{A} = \{\varphi_n\} \subseteq B(C(S^1), \mathbb{C})$ is not uniformly bounded. Now Theorem G.2.2 implies that there is a dense G_δ-set $X \subseteq C(S^1)$ such that $\varphi_n(f) = S_n(f)(0)$ diverges as $n \to \infty$. ■

Lemma G.2.14 *We have*

$$\|D_n\|_1 = \frac{4}{\pi^2} \log n + O(1).$$

Proof. [144] ■

Alternatively, one may use the contraposition of Corollary G.2.3 with its simple proof to deduce that $\mathcal{A}$ is not pointwise bounded, thus that there is at least one continuous f such that $S_n(f)(0)$ diverges. This is one of the shortest and most transparent existence proofs for such a function, but see also the nice almost constructive proof in [144, Section 18].

Remark G.2.15 1. It is not too hard to show that for every set $X \subseteq S^1$ of measure zero there exists an $f \in C(S^1)$ such that $S_n(f)(x)$ diverges for all $x \in X$.

2. With considerably more work one proves that for $f \in C(S^1)$ (more generally: $f \in L^p([0, 2\pi])$ for some $p > 1$) one has $S_n(f)(x) \to f(x)$ for almost all x, i.e. all x outside some set of measure zero. (This was long considered one of the hardest results in analysis, but this begins to change in the light of recent simplifications.) □

G.3 Alaoglu's theorem

We now consider a result in functional analysis that is often construed as an application of Tychonov's theorem. As we will see, it should rather be seen as an application of the circle of ideas around the latter.

Definition G.3.1 *If V is a Banach space, the $\sigma(V^*, V)$-topology (or weak$-*$ topology) is the initial topology on the dual space V^* defined by the family $\{n_x : V^* \to [0, \infty),\ \phi \mapsto |\phi(x)|\}_{x \in V}$ of "seminorms". (Thus a net $\{\phi_\iota\}$ in V^* converges to $\psi \in V^*$ if and only if $\phi_\iota(x) \to \psi(x)$ for every $x \in V$.)*

Remark G.3.2 If V is infinite-dimensional, the weak $*$-topology τ^* does not arise from a norm, thus V^* is not a Banach space, but only a locally convex vector space. □

Theorem G.3.3 [2] *If $(V, \|\cdot\|)$ is a Banach space then $V_1^* = \{\phi \in V^* \mid \|\phi\| \leq 1\}$ is compact in the $\sigma(V^*, V)$-topology.*

Proof using Tychonov's theorem. Define

$$Z = \prod_{x \in V} \{z \in \mathbb{C} \mid |z| \leq \|x\|\}.$$

Since the closed discs in $\mathbb{C}$ are compact, Z is compact by Tychonov's theorem. Now consider the map

$$f : V_1^* \to Z, \quad \phi \mapsto \prod_{x \in V} \phi(x).$$

[2] Proven by Leonidas Alaoglu (1914-1981) in 1938 (PhD thesis)/1940 (paper).

Since each $\phi \mapsto \phi(x)$ is continuous for each x, f is continuous. It is trivial that V separates the points of V^*, thus f is injective. By definition, a net ϕ_ι in V_1^* converges in the $\sigma(V^*, V)$-topology if and only if $\phi_\iota(x)$ converges for all $x \in V$, and therefore if and only if $f(\phi_\iota)$ converges. Thus $f : V_1^* \to f(V_1^*) \subseteq Z$ is a homeomorphism (w.r.t. the weak topology on the left hand side).

Thus let $z \in \overline{f(V_1^*)} \subseteq Z$. Clearly, $|z_x| \leq \|x\|$, thus $x \mapsto z_x$ is a bounded map. Now, by Proposition 5.1.19 and injectivity of f, there is a net ϕ_ι in V_1^* such that $f(\phi_\iota) \to z$. This means that $\phi_\iota(\alpha x + \beta y) = \alpha\phi_\iota(x) + \beta\phi_\iota(y) \to \alpha z_x + \beta z_y$ and $\phi_\iota(\alpha x + \beta y) \to z_{\alpha x+\beta y}$, thus the map $x \mapsto z_x$ is bounded and linear, thus $z = f(\phi)$ for some $\phi \in V_1^*$. This proves that $f(V_1^*) \subseteq Z$ is closed, thus compact by Lemma 7.3.4. Now the result follows from $V_1^* \cong f(V_1^*)$. ■

Proof using universal nets. Let ϕ_ι be a universal net in V_1^*. If $x \in V$ then $\phi_\iota(x)$ is a universal net in $\{z \in \mathbb{C} \mid |z| \leq \|x\|\}$, which is compact. Thus $\phi_\iota(x)$ converges for every $x \in V$. Therefore, ϕ_ι converges in the $\sigma(V^*, V)$-topology, thus V_1^* is weakly compact. ■

Remark G.3.4 1. The second proof of Alaoglu's theorem, found e.g. in [189], is extremely short and elegant and seems to show the superiority of nets. But it should be noted that Alaoglu's theorem can be proved equally easily using filters.

2. While the proof of Alaoglu's theorem using nets shows that it is a distortion to view that theorem as a consequence of Tychonov's theorem, the first proof nevertheless has the advantage of making explicit that Tychonov's theorem is needed only for T_2-spaces. In fact, since $D^2 \cong I^2$, what is needed for Alaoglu's theorem is the compactness of the cube I^χ where $\chi = \#V$. (The latter also implies Tychonov theorem for T_2-spaces: We will see that every compact T_2-space is homeomorphic to a closed subspace of the cube $I^{C(X,I)}$. Thus by Exercise 6.5.10 a product $\prod_i X_i$ of compact T_2-spaces is homeomorphic to a closed subspace of a cube and thus is compact.) This shows: Tychonov for T_2-spaces $\Leftrightarrow$ Tychonov for cubes $\Rightarrow$ Alaoglu. The missing implication is also true, but we leave matters here. □

G.4 Schauder's fixed point theorem

In this section we consider generalizations of Brouwer's fixed point theorem, more precisely of Corollary 10.3.6, to infinite dimensional vector spaces.

Theorem G.4.1 (Schauder) [3] *Every non-empty compact convex subset K of a normed vector space has the fixed point property. (I.e. every continuous $f : K \to K$ has a fixed point.)*

Proof. Let $(V, \|\cdot\|)$ be a normed vector space, $K \subseteq V$ a non-empty compact convex subset and $f : K \to K$ continuous. Let $\varepsilon > 0$. Since K is compact, thus totally bounded, there are $x_1, \ldots, x_n \in K$ such that $K \subseteq \bigcup_{i=1}^n B(x_i, \varepsilon)$. Thus if we define $\alpha_i(x) \geq 0$ by

$$\alpha_i(x) = \begin{cases} 0 & \text{if } \|x - x_i\| \geq \varepsilon \\ \varepsilon - \|x - x_i\| & \text{if } \|x - x_i\| < \varepsilon \end{cases} \qquad \forall i = 1, \ldots, n, \tag{G.3}$$

we see that for each $x \in K$ there is at least one i such that $\alpha_i(x) > 0$. The functions α_i clearly are continuous. Thus also the map

$$P_\varepsilon : K \to K, \quad x \mapsto \frac{\sum_{i=1}^n \alpha_i(x) x_i}{\sum_{i=1}^n \alpha_i(x)}$$

[3] Juliusz Schauder (1899-1943). Born in Lwow/Lviv (Ukraine, then Lemberg in the Austrian empire) and killed by the nazis during WW2.

is continuous. Since $P_\varepsilon(x)$ is a convex combination of those x_i for which $\|x - x_i\| < \varepsilon$, we have $\|P_\varepsilon(x) - x\| < \varepsilon$ for all $x \in K$. The finite dimensional subspace $V_n = \mathrm{span}(x_1, \ldots, x_n) \subseteq V$ is isomorphic to some $\mathbb{R}^m$, and by Theorem 7.7.51 the restriction of the norm $\|\cdot\|$ to V_n is equivalent to the Euclidean norm on $\mathbb{R}^m$. Thus the convex hull $\mathrm{conv}(x_1, \ldots, x_n) \subseteq V_n$ into which P_ε maps is homeomorphic to a compact convex subset of $\mathbb{R}^m$ and thus has the fixed point property by Corollary 10.3.6. Thus if we define $f_\varepsilon = P_\varepsilon \circ f$ then f_ε maps $\mathrm{conv}(x_1, \ldots, x_n)$ into itself and thus has a fixed point $x' = f_\varepsilon(x')$. Now,

$$\|x' - f(x')\| \leq \|x' - f_\varepsilon(x')\| + \|f_\varepsilon(x') - f(x')\| = \|f_\varepsilon(x') - f(x')\| = \|P_\varepsilon(f(x')) - f(x')\| < \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, we find $\inf\{\|x - f(x)\| \mid x \in K\} = 0$. Since K is compact and $x \mapsto \|x - f(x)\|$ continuous, the infimum is assumed (Corollary 7.7.30), thus f has a fixed point in K. ■

Corollary G.4.2 *Let K be a non-empty closed convex subset of a Banach space, and let $f : K \to K$ be continuous with $\overline{f(K)}$ compact. Then f has a fixed point $x \in K$.*

Proof. The closed convex hull $\overline{\mathrm{conv}(f(K))}$ of $f(X)$ is compact and convex and mapped into itself by f. Now the theorem applies. ■

Remark G.4.3 1. The above results were proven by Juliusz Schauder in 1930. In 1934 Tychonov generalized them to locally convex vector spaces. (For an exposition, cf. e.g. [63, Section V.10].) The ultimate generalization to all topological vector spaces with T_0-property (thus $T_{3.5}$, cf. Section C.2.3) was only proven in 2001 by R. Cauty[4] [47].

2. As we saw, Schauder's fixed point theorem is an infinite dimensional version of Brouwer's fixed point theorem. The latter has many ramifications, in particular the Brouwer degree of a map. Also degree theory can be extended to infinite dimenensions, leading to the Leray-Schauder degree. The definition of the latter makes use of the Brouwer degree, whose definition requires either homology theory, cf. e.g. [28, 108] or differential topology [101, 117]. There does not seem to be a way to do this that is as elementary as the proof of the fixed point theorem given in Section 10. For this reason, we leave the subject here. Cf. e.g. [36, 56]. □

G.5 A quick look at locally convex vector spaces

We recall the definition from Section 8.5.5:

Definition G.5.1 *A topological vector space V (over $\mathbb{R}$ or $\mathbb{C}$) is called locally convex if it is T_0 and $0 \in V$ has a neighborhood base consisting of convex open sets.*

Theorem G.5.2 *A T_0 topological vector space is locally convex if and only if its topology arises from a family of seminorms separating the points.*

Proof. Let $\mathcal{F}$ be a family of seminorms on the vector space V. Assume that for every $0 \neq x \in V$ there is a seminorm $\rho \in \mathcal{F}$ such that $\rho(x) \neq 0$. Define

$$\mathcal{S} = \{\rho^{-1}([0, \varepsilon)) \mid \rho \in \mathcal{F},\ \varepsilon > 0\}.$$

[4] Robert Cauty (19??-2013), French mathematician.

It is clear that $\bigcup \mathcal{S} = V$, so that there is a topology τ on V having $\mathcal{S}$ as a subbase. The sets $\rho^{-1}([0, \varepsilon))$ are convex open neighborhoods of 0, and so are the finite intersections that form the base $\mathcal{B}$ associated to $\mathcal{S}$. Thus τ is a locally convex topology.

■

Proposition G.5.3 *The topology of a T_0-topological vector space is metrizable if and only if it arises from a countable family of seminorms.*

Such a space is called a Fréchet space. (Clearly every Banach space is Fréchet.)

Proof. Let $\{\| \cdot \|_n\}_{n \in \mathbb{N}}$ be a family of seminorms inducing the topology. Then

$$d(x, y) = \sum_{n=1}^{\infty} \frac{\|x - y\|_n}{1 + \|x - y\|_n}$$

defines a metric. As in Lemma 8.3.26(iii) one shows that it induces the given topology.

■

Remark G.5.4 Note that the metric defined in the above proof is translation invariant, i.e. $d(x, y) = d(x - y, 0)$. Using this, we have $d(x - z, 0) = d(x, z) \leq d(x, y) + d(y, z) = d(x - y, 0) + d(y - z, 0)$. Thus $x \mapsto d(x, 0)$ almost is a norm, but not quite since $d(cx, 0) = cd(x, 0)$, where $c \in \mathbb{F}$, does not hold. It is not hard to find examples of Fréchet spaces that are not Banach spaces: For $f \in C^\infty(\mathbb{R})$ define $\|f\|_n = \|f^{(n)}\|_\infty$ for all $n \in \mathbb{N}_0$. Now

$$\mathcal{S} = \{f \in C^\infty(\mathbb{R}) \mid \|f\|_n < \infty \; \forall n = 0, 1, \ldots\}$$

with the topology induced by the family of seminorms $\{\| \cdot \|_n\}$ is a Fréchet space (the space of test functions, important in distribution theory), but not Banach. □

We thus have, in increasing order of generality: Hilbert space ⇒ Banach space ⇒ Fréchet space ⇒ locally convex topological vector space ⇒ T_0 topological vector space.

Appendix H

Rough schedule of a lecture course (16 weeks)

1. basics, motivation, from metrics to topologies, metrizability. Closed sets, [Zariski top.], T_1, T_2,
2. interior, closure, boundary, density. Metric spaces: completeness
3. Completions. Baire's theorem. Appl.: nowhere diffable fct. Axioms AC, ACC and ACDC.
4. bases, 2nd countability, separability, subbases, nbhd bases, 1st countability.
5. sequences, nets, continuity, continuity in terms of nets. The category $\mathcal{TOP}$
6. homeomorphisms. open/closed maps. initial and final topol., subspaces, disjoint unions.
7. quotient spaces and applications. Pushout.
8. product topology, box topology, products of metric spaces, separate vs. joint continuity.
9. Definition of Lindelöf and compactness, first examples. relation to subsets and continuity.
10. recall Zorn's lemma, Alexander subbase lemma, Tychonov. Section 7.7: countable/sequential compactness. Metric spaces: total boundedness, Lebesgue property
11. equivalences for metric spaces. Compactness for subsets of $\mathbb{R}^n$, equivalence of norms, compact convex subsets of $\mathbb{R}^n$ homeom. to $I^m, m \leq n$. Section 7.8: 1-pt compactification, local compactness.
12. A bit more on local compactness, proper maps. T_3, T_4. compact $T_2 \Rightarrow T_4$. metric$\Rightarrow T_4$. (T_4 not hereditary $\rightsquigarrow T_5$. T_6-spaces.) Urysohn's Lemma. (Tietze extension thm., Partitions of unity)
13. families of functions separating points from closed sets. Urysohn metrization thm. $T_{3.5}$, locally compact $T_2 \Rightarrow T_{3.5}$. embeddings into cube. Stone-Čech compactification. functoriality.
14. connected spaces, fixpt.thm for I, $\mathbb{R} \not\cong \mathbb{R}^n$ for $n \geq 2$. conn. component. path-connected. locally (path) conn. totally disconnected spaces
15. totally separated and zero-dimensional spaces. Cantor set $C \cong \{0,1\}^{\mathbb{N}}$. cont. surjections $C \to \mathbb{R}^n, \mathbb{R}^{\mathbb{N}}$, Peano curves (via Tietze and more concretely), devils' staircase fct. Retractions, retraction onto closed subsets of C, cont. surj. $C \to X$ for X compact metrizable. Comments on Peano spaces (compact T_2, 2nd cnt, path-conn, loc.path conn)

16. homotopy of maps, homotopy equiv. of spaces. π_1 and its properties, incl. homotopy inv. $\pi_1(S^n) = 0$ for $n \geq 2$.

Bibliography

[1] J. F. Aarnes, P. R. Andenæs: On nets and filters. Math. Scand. **31**, 285-292 (1972).

[2] P. S. Aleksandrov, V. V. Fedorchuk (with the assistance of V. I. Zaitsev): The main aspects in the development of set-theoretical topology. Russ. Math. Surv. **33**, 1-53 (1978).

[3] S. Alinhac, P. Gérard: *Pseudo-differential operators and the Nash-Moser theorem.* American Mathematical Society, 2007. (Translation of *Opérateurs pseudo-différentiels et théorème de Nash-Moser.* InterEditions, 1991.

[4] L. Ambrosio, P. Tilli: *Topics in analysis in metric spaces.* Oxford University Press, 2004.

[5] T. Andreescu, C. Mortici, M. Tetiva: *Mathematical bridges.* Birkhäuser, 2017.

[6] F. G. Arenas: Alexandrov spaces. Acta Math. Univ. Comenianae **LXVIII**, 17-25 (1999).

[7] R. F. Arens: A Topology for Spaces of Transformations. Ann. Math. **47**, 480-495 (1946).

[8] A. Arhangel'skii, M. Tkachenko: *Topological groups and related structures.* Atlantic Press, 2008.

[9] M. Atsuji: Normality of product spaces I. *In:* K. Morita, J. Nagata (eds.): *Topics in general topology*, North-Holland, 1989.

[10] J.-P. Aubin, I. Ekeland: *Applied nonlinear analysis.* Wiley, 1984, Dover, 2006.

[11] C. E. Aull & R. Lowen (eds.): *Handbook of the history of general topology.* Vols. 1-3. Kluwer Academic Publishers, 1997, 1998, 2001.

[12] R. Ayala, M. J. Chaâvez, A. Quintero: A Hahn-Mazurkiewicz Theorem for generalized Peano continua. Arch. Math. **71**, 325-330 (1998).

[13] W. Ballmann: *Lectures on spaces of nonpositive curvature.* Birkhäuser, 1995.

[14] Balogh, Z. T.d: A small Dowker space in ZFC. Proc. Amer. Math. Soc. **124**, 2555- 2560 (1996).

[15] R. G. Bartle: Nets and filters in topology. Amer. Math. Monthly **62**, 551-557 (1955) and **70**, 52-53 (1963) (erratum).

[16] H-J. Baues, A. Quintero: *Infinite homotopy theory.* Kluwer (now Springer), 2001.

[17] A. F. Beardon: *Limits. A new approach to real analysis.* Springer, 1997.

[18] R. Bhatia, K. M. Mukherjea: The space of unordered tuples of complex numbers. Lin. Alg. Applic. **52/53**, 765-768 (1983).

[19] A. Białynicki-Birula, W. Żelazko: On the multiplicative linear functionals on the cartesian product of algebras. Bull. Acad. Polon. Sci. Cl. III **5**, 589-593 (1957).

[20] C. E. Blair: The Baire category theorem implies the principle of dependent choices. Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom. Phys. **25**, 933-934 (1977).

[21] A. Blass: Existence of bases implies the Axiom of Choice. Contemp. Math. **31**, 31-33 (1984).

[22] L. M. Blumenthal: *Theory and Application of distance geometry*. Oxford University Press, 1953.

[23] F. Borceux, G. Jandelidze: *Galois theories*. Cambridge University Press, 2001.

[24] A. Borobia, U. R. Trias: A geometric proof of the Perron-Frobenius theorem. Rev. Mat. Complutense **5**, 57-63 (1992).

[25] N. Bourbaki: *General topology I*. Springer, 1989.

[26] N. Bourbaki: *Topologie Algébrique. Chapitres 1 à 4*. Springer, 2016.

[27] H. Brandsma: Alexander's subbase lemma and applications. Topology Q+A Board, Jan. 1, 2005.

[28] G. E. Bredon: *Topology and geometry*. Springer, 1993, 1997.

[29] H. Brézis: *Functional analysis, Sobolev spaces and partial differential equations*. Springer, 2011.

[30] H. Brézis, F. E. Browder: A general principle on ordered sets in nonlinear functional analysis. Adv. Math. **21**, 355-364 (1976).

[31] M. Bridson, A. Haefliger: *Metric spaces of non-positive curvature*. Springer, 1999.

[32] L. E. J. Brouwer: Beweis der Invarianz der Dimensionenzahl. Math. Ann. **70**, 161-165 (1911).

[33] L. E. J. Brouwer: Über Abbildung der Mannigfaltigkeiten. Math. Ann. **71**, 97-115 (1912).

[34] L. E. J. Brouwer: Über den natürlichen Dimensionsbegriff. Journ. für Math. **142**, 146-152 (1913).

[35] F. E. Browder: On a theorem of Caristi and Kirk. *In*: S. Swaminathan (ed.): *Fixed point theory and its applications*. Academic Press, 1976.

[36] R. F. Brown: *A topological introduction to nonlinear analysis*. Birkhäuser, 1993, 2004, 2014.

[37] R. Brown: *Topology and groupoids*. BookSurge, 2006.

[38] N. Brunner: Topologische Maximalprinzipien. Zeitschr. f. math. Logik u. Grdl. d. M. **33**, 135-139 (1987).

[39] D. Burago, Y. Burago, S. Ivanov: *A course in metric geometry*. American Mathematical Society, 2001.

[40] G. Buskes, A. C. M. van Rooij: *Topological spaces: From distance to neighborhood*. Springer, 1997.

[41] S. Buyalo, V. Schroeder: *Elements of asymptotic geometry*. European Mathematical Socitey, 2007.

[42] A. Carbery: The Brouwer fixed point theorem and the Borsuk-Ulam theorem. arXiv:1205.4540v1.

[43] J. Caristi: Fixed point theorems for mappings satisfying certain inwardness conditions. Trans. Amer. Math. Soc. **215**, 241-215 (1976).

[44] J. Scott Carter: *How surfaces intersect in space. An introduction to topology*. World Scientific, 1993.

[45] D. Cass, G. Wildenberg: Math Bite: A Novel Proof of the Infinitude of Primes, Revisited. Mathematics Magazine **76**, 203 (2003).

[46] F. S. Cater: A simple proof that a linearly ordered space is hereditarily and completely collectionwise normal. Rocky Mountain Journ. Math. **36**, 1149-1151 (2006).

[47] R. Cauty: Solution du problème de point fixe de Schauder. Fund. Math. **170**, 231-246 (2001).

[48] P. R. Chernoff: A simple proof of Tychonoff's theorem via nets. Amer. Math. Monthly **99**, 932-934 (1992).

[49] C. Chevalley, O. Frink: Bicompactness of cartesian products. Bull. Amer. Math. Soc. **47**, 612-614 (1941).

[50] C. O. Christenson, W. L. Voxman: *Aspects of topology*. 2nd edition, BCS Associates, 1998.

[51] K. Ciesielski: *Set theory for the working mathematician*. Cambridge University Press, 1997.

[52] J. B. Conway: *A course in point-set topology*. Springer, 2014.

[53] W. H. Cornish: On H. Priestley's dual of the category of bounded distributive lattices. Math. Vesnik **12**, 329-332 (1975).

[54] H. T. Croft: A question of limits. Eureka **20**, 11-13 (1957).

[55] J. P. Dalbec: When does the restricted continuity on continuous function graphs imply joint continuity? Proc. Amer. Math. Soc. **118**, 669-674 (1993).

[56] K. Deimling: *Nonlinear functional analysis*. Springer, 1985, Dover, 2010.

[57] J. Dieudonné: *A history of algebraic and differential topology 1900-1960*. Birkhäuser, 1989.

[58] D. N. Dikranjan, I. R. Prodanov, L. N. Stoyanov: *Topological groups. Characters, dualities, and minimal group topologies*. Marcel Dekker Inc., 1990.

[59] R. Douady, A. Douady: *Algèbre et théories galoisiennes*. Cedic, 1978/9, Cassini, 1999.

[60] J. Dugundji: *Topology*. Allyn and Bacon Inc., 1966.

[61] J. Dugundji, A. Granas: *Fixed point theory*. Springer, 2003.

[62] J. J. Duistermaat: Self-similarity of "Riemann's nondifferentiable function". Nieuw Arch. Wisk. (4) **9**, no. 3, 303-337 (1991).

[63] N. Dunford, J. T. Schwartz: *Linear operators. I. General theory.* Interscience Publishers, 1958, John Wiley & Sons, 1988.

[64] H.-D. Ebbinghaus, H. Hermes, E. Hirzebruch, M. Koecher, K. Mainzer, J. Neukirch, A. Prestel, R. Remmert: *Zahlen*. 3rd.ed. Springer, 1993.

[65] M. Edelstein: On fixed and periodic points under contractive mappings. J. Lond. Math. Soc. **37**, 74-79 (1962).

[66] G. Edgar: *Measure, topology, and fractal geometry*. Springer, 2008.

[67] S. Eilenberg, S. Mac Lane: General theory of natural equivalences. Trans. Amer. Math. Soc. **58**, 231-294 (1945).

[68] S. Eilenberg, E. Otto: Quelques propiétés caractéristiques de la dimension. Fund. Math. **31**, 149-153 (1938).

[69] S. Eilenberg, N. Steenrod: *Foundations of algebraic topology*. Princeton University Press, 1952.

[70] I. Ekeland: On the variational principle. J. Math. Anal. Appl. **47**, 324-353 (1974).

[71] I. Ekeland: Nonconvex minimization problems. Bull. Amer. Math. Soc. **1**, 443-474 (1979).

[72] I. Ekeland: An inverse function theorem in Fréchet spaces. Ann. I. H. Poincaré **AN 28**, 91-105 (2011).

[73] I. Ekeland, E. Séré: An implicit function theorem for non-smooth maps between Frchet spaces. arXiv:1502.01561.pdf.

[74] P. C. Eklof: Whitehead's problem is undecidable. Amer. Math. Monthly **83**, 775-788 (1976).

[75] R. Engelking: *General topology*. Heldermann Verlag, 1989.

[76] R. Engelking: *Theory of Dimensions: Finite and Infinite*. Heldermann Verlag, 1995.

[77] R. Engelking, K. Siekluchi: *Topology. A geometric approach*. Heldermann Verlag, 1992.

[78] K. Falconer: *Fractal Geometry: Mathematical Foundations and Applications*. Wiley, 1989, 2003.

[79] I. Fleisher: Priestley's duality from Stone's. Adv. Appl. Math. **25**, 233-238 (2000).

[80] G. B. Folland: A tale of topology. Amer. Math. Monthly **117**, 663-672 (2010).

[81] S. P. Franklin, G. V. Krishnarao: On the topological characterization of the real line. J. Lond. Math. Soc. (2) **2**, 589-591 (1970) and **3**, 392 (1971).

[82] M. Fréchet: *Sur quelques points du calcul fonctionnel*. PhD thesis, 1906.

[83] H. Freudenthal: Über die topologischen Räume und Gruppen. Math. Zeit. **33**, 692-713 (1931).

[84] H. Freudenthal: Neuaufbau der Endentheorie. Ann. Math. **43**, 261-279 (1942).

[85] W. Fulton: *Algebraic topology. A first course*. Springer, 1995.

[86] S. A. Gaal: *Point set topology*. Academic Press, 1964, Dover, 2009.

[87] T. W. Gamelin, R. E. Greene: *Introduction to topology.* Saunders College Publishing, 1983, Dover, 1999.

[88] D. J. H. Garling: *A course in mathematical analysis.* Vol. I-III. Cambridge University Press, 2013-2014.

[89] J. Gerver: The differentiability of the Riemann function at certain rational multiples of π. Amer. Math. Monthly **92**, 33-55 (1970). More on the differentiability of the Riemann function. Amer. J. Math. **93**, 33-41 (1971).

[90] L. Gillman, M. Jerison: *Rings of continuous functions.* Van Nostrand, 1960.

[91] E. Giusti: *Direct methods in the calculus of variations.* World Scientific, 2003.

[92] C. Godbillon: *Éléments de topologie algébrique.* Hermann, 1971.

[93] K. Goebel, W. A. Kirk: *Topics in metric fixed point theory.* Cambridge University Press, 1990.

[94] R. Goldblatt: On the role of the Baire category theorem and dependent choice in the foundations of logic. J. Symb. Logic **50**, 412-422 (1985).

[95] S. Grabiner: The Tietze extension theorem and the open mapping theorem. Amer. Math. Monthly **93**, 190-191 (1986).

[96] B. Gray: *Homotopy theory. An introduction to algebraic topology.* Academic Press, 1975.

[97] G. L. M. Groenewegen; A. C. M. van Rooij: *Spaces of continuous functions.* Atlantis Press, 2016.

[98] M. Gromov, J. Lafontaine, P. Pansu: *Metric structures for Riemannian and non-Riemannian spaces.* Birkhäuser, 1999.

[99] K. Gruenberg: Profinite groups. *In:* J. W. S. Cassels, A. Fröhlich (eds.) : *Algebraic Number Theory.* Academic Press, 1967.

[100] D. Guido, L. Tuset: Representations of the direct product of matrix algebras. Fund. Math. **169**, 145-160 (2001).

[101] V. Guillemin, A. Pollack: *Differential topology.* Prentice Hall, 1974, AMS Chelsea 2010.

[102] M. J. Hadamard: Sur quelques applications de l'indice de Kronecker. pp. 875-915 *in:* J. Tannery (ed.): *Introduction à la théorie des fonctions d'une variable*, vol. 2. Hermann, 1910.

[103] P. R. Halmos: *Naive set theory.* Van Nostrand, 1960.

[104] R. S. Hamilton: The inverse function theorem of Nash and Moser. Bull. Amer. Math. Soc. **7**, 65-222.

[105] H. Hanche-Olsen, H. Holden: The Kolmogorov-Riesz compactness theorem. Expos. Math. **28**, 385-394 (2010).

[106] K. P. Hart, J. Nagata, J. E. Vaughan (eds.): *Encyclopedia of General Topology.* Elsevier, 2004.

[107] R. Hartshorne: *Algebraic geometry.* Springer, 1977.

[108] A. Hatcher: *Algebraic topology*. Cambridge University Press, 2002.

[109] F. Hausdorff: *Grundzüge der Mengenlehre*. Veit & Comp., 1914.

[110] J. Heinonen: *Lectures on Analysis on Metric Spaces*. Springer, 2001.

[111] E. Hemingsen: Some theorems in dimension theory for normal Hausdorff spaces. Duke Math. Journ. **13**, 495-504 (1946).

[112] J. Hendrikson: Completeness and total boundedness of the Hausdorff metric. MIT Undergrad. Journ. (1999).

[113] G. T. Herman: On topology as applied to image analysis. Comp. Vision, Graphics & Image Proc. **52**, 409-415 (1990).

[114] H. Herrlich: *Axiom of choice*. Springer, 2006.

[115] E. Hewitt: The rôle of compactness in analysis. Amer. Math. Monthly **67**, 499-516 (1960).

[116] E. Hewitt, K. A. Ross: *Abstract harmonic analysis*. 2 volumes. Springer, 1962, 1970.

[117] M. W. Hirsch: *Differential topology*. Springer 1976, 1994.

[118] M. Hochster: Prime ideal structure in commutative rings. Trans. Amer. Math. Soc. **142**, 43-60 (1969).

[119] J. G. Hocking, G. S. Young: *Topology*, Addison-Wesley, 1961. Dover, 1988.

[120] W. Hodges: Krull implies Zorn. J. Lond. Math. Soc. **19**, 285-7 (1979).

[121] R.-E. Hoffmann: On weak Hausdorff spaces. Arch. Math. **32**, 487-504 (1979).

[122] J. Hubbard, B. Burke Hubbard: *Vector Calculus, Linear Algebra, And Differential Forms. A Unified Approach*. Prentice Hall, 1998, 2001.

[123] W. Hurewitz, H. Wallmann: *Dimension theory*. Princeton University Press, 1941.

[124] A. Idzik, W. Kulpa, P. Maćkowiak: Equivalent forms of the Brouwer fixed point theorem. Topol. Meth. Nonlin. Anal. **44**, 263-276 (2014).

[125] A. Idzik, W. Kulpa, P. Maćkowiak: On a size of a map. Preprint, 2015.

[126] J. R. Isbell: *Uniform spaces*. American Mathematical Society, 1964.

[127] N. V. Ivanov: Sperner's lemma, the Brouwer fixed-point theorem, and cohomology. arXiv:0906.5193

[128] N. V. Ivanov: A topologist's view of the Dunford-Schwartz proof of the Brouwer fixed-point theorem. Math. Intell. **22**, no. 3, 55-57 (2000).

[129] N. V. Ivanov: A differential forms perspective on the Lax proof of the change of variables formula. Amer. Math. Monthly **112**, 799-806 (2005).

[130] I. M. James: *Topological and uniform spaces*. Springer, 1987.

[131] I. M. James (ed.): *Handbook of algebraic topology.* North-Holland, 1995.

[132] I. M. James (ed.): *History of topology.* North-Holland, 1999.

[133] I. M. James: Combinatorial topology versus point set topology. *In:* [11, Vol. 3].

[134] K. Jänich: *Topologie*. Springer, 1980, ..., 2008.

[135] P. T. Johnstone: *Stone spaces*. Cambridge University Press 1982, 1986.

[136] S. H. Jones: Applications of the Baire category theorem. Real Anal. Exchange **23**, no. 2, 363-394 (1997/98).

[137] J. L. Kelley: The Tychonoff product theorem implies the axiom of choice. Fund. Math. **37**, 75-76 (1950).

[138] J. L. Kelley: Convergence in topology. Duke Math. J. **17**, 277-283 (1950).

[139] J. L. Kelley: *General topology.* Van Nostrand 1955, Springer 1975.

[140] M. A. Khamsi, W. A. Kirk: *An introduction to metrix spaces and fixed point theory.* Wiley, 2001.

[141] A. Khovanskii: *Galois Theory, Coverings, and Riemann Surfaces*. Springer, 2013.

[142] B. Knaster, C. Kuratowski, S. Mazurkiewicz: Ein Beweis des Fixpunktsatzes für n-dimensionale Simplexe. Fund. Math. **14**, 132-137 (1929).

[143] R. J. Knill: The Seifert and van Kampen theorem via regular covering spaces. Pac. J. Math. **49**, 149-160 (1973).

[144] T. W. Körner: *Fourier analysis*. Cambridge University Press, 1988.

[145] S. G. Krantz, H. R. Parks: *The implicit function theorem. History, theory, and applications.* Birkhäuser, 2002.

[146] M. Krom: A linearly ordered topological space that is not normal. Notre Dame Journ. Formal Logic **27**, 12-13 (1986).

[147] H. W. Kuhn: Some combinatorial lemmas in topology. IBM Journ. Res. Devel. **4**, 518-524 (1960).

[148] W. Kulpa: The Poincaré-Miranda theorem. Amer. Math. Monthly **104**, 545-550 (1997).

[149] W. Kulpa: Poincaré and domain invariance theorem. Acta Univ. Carolinae. Math. et Phys. **39**, 127-136 (1998).

[150] S. Kundu, Tanvi Jain: Atsuji spaces: Equivalent conditions. Topol. Proc. **30**, 301-325 (2006).

[151] K. Kunen: *Set theory. An introduction to independence proofs*. North Holland, 1980.

[152] K. Kunen, J. E. Vaughan (eds.): *Handbook of set-theoretic topology.* North-Holland, 1984.

[153] K. Kuratowski: *Topology*, 2 volumes. Academic Press, 1966, 1968.

[154] G. Laures, M. Szymik: *Grundkurs Topologie*. Spektrum Akademischer Verlag, 2009.

[155] P. D. Lax: Change of variables in multiple integrals. Amer. Math. Monthly **106**, 497-501 (1999).

[156] J. Lewin: A simple proof of Zorn's Lemma. Amer. Math. Monthly **98**, 353-354 (1991).

[157] E. H. Lieb, M. Loss: *Analysis*. American Mathematical Society, 1997, 2001.

[158] S. Lipschutz: *General topology*. Schaum's outlines. McGraw-Hill, 1965.

[159] L. H. Loomis: *An introduction to abstract harmonic analysis*. Van Nostrand, 1953.

[160] S. Mac Lane: *Categories for the working mathematician*. 2nd ed. Springer, 1998.

[161] R. Maehara: The Jordan curve theorem via the Brouwer fixed point theorem. Amer. Math. Monthly **91**, 641-643 (1984).

[162] A. R. Magid: *The separable Galois theory of commutative rings*. Marcel Dekker Inc., 1974, Taylor & Francis, 2014.

[163] R. Mańka: Some forms of the axiom of choice. Jahrbuch der Kurt-Gödel-Gesellschaft 1988, pp. 24-34.

[164] J. H. Manheim: *The genesis of point set topology*. Pergamon Press, 1964.

[165] J. Matoušek: *Using the Borsuk-Ulam theorem*. Springer, 2003.

[166] M. C. McCord: Classifying spaces and infinite symmetric products. Trans. Amer. Math. Soc. **146**, 273-298 (1969).

[167] T. B. M. McMaster: The Smirnov compactification as a quotient space of the Stone-Čech compactification. Internat. J. Math. & Math. Sci. **11**, 503-506 (1988).

[168] I. D. Mercer: On Furstenberg's proof of the infinitude of primes. Amer. Math. Monthly **116**, 355-356 (2009).

[169] J. van Mill: *Infinite-dimensional topology*. North-Holland, 1989.

[170] C. Miranda: Un'osservazione su un teorema di Brouwer. Boll. Un. Mat. Ital. **3**, 5-7 (1940).

[171] I. Moerdijk, J. van Oosten: *Sets, models and proofs*. Lecture notes. Univ. Utrecht, 2000, 2011.

[172] G. H. Moore: The evolution of the concept of homeomorphism. Hist. Math. **34**, 333-343 (2007).

[173] G. H. Moore: The emergence of open sets, closed sets, and limit points in analysis and topology. Hist. Math. **35**, 220-241 (2008).

[174] P. Morandi: *Field and Galois theory*. Springer, 1996.

[175] M. Morayne: On differentiability of Peano type functions I-II. Colloq. Math. **53**, 129–132, 133–135 (1987).

[176] J. Morgan, G. Tian: *Ricci flow and the Poincaré conjecture*. American Mathematical Society, 2007.

[177] K. Morita, J. Nagata (eds.): *Topics in general topology*. Elsevier, 1989.

[178] M. Müger: A remark on the invariance of dimension. Math. Semesterber. **62**, 59-68 (2015).

[179] A. Mysior: A regular space which is not completely regular. Proc. Amer. Math. Soc. **81**, 652-653 (1981).

[180] J. R. Munkres: *Topology*. Second edition. Prentice-Hall, 2000.

[181] G. J. Murphy: *C^*-Algebras and operator theory*. Academic Press, 1990.

[182] J. Nagata: *Modern general topology*. North Holland, 1968, 1974, 1985.

[183] S. A. Naimpally, B. D. Warrack: *Proximity spaces*. Cambridge University Press, 1970.

[184] J. Nash: Non-cooperative games. Ann. Math. **54** 286-295 (1951).

[185] L. Nirenberg: *Topics in nonlinear functional analysis*. American Mathematical Society, 1974, 2001.

[186] P. W. Nowak, G. Yu: *Large scale geometry*. European Mathematical Socitey, 2012.

[187] J. C. Oxtoby: *Measure and category. A survey of the analogies between topological and measure spaces*. Springer, 1971.

[188] R. S. Palais: A simple proof of the Banach contraction principle. J. Fixed Point Theor. Appl. **2**, 221-223 (2007).

[189] G. Pedersen: *Analysis Now*. Springer, 1989.

[190] J.-P. Penot: Fixed point theorems without convexity. Mém. Soc. Math. France **60**, 129-152 (1979).

[191] J. C. Pont: *La topologie algébrique. Des origines à Poincaré*. Presses Universitaires de France, 1974.

[192] B. von Querenburg: *Mengentheoretische Topologie*. 3rd ed. Springer 2008.

[193] T.-L. T. Rădulescu, V. D. Rădulescu, T. Andreescu: *Problems in Real Analysis. Advanced Calculus on the Real Axis*. Springer, 2009.

[194] M. Raman Sundström: A pedagogical history of compactness. arXiv:1006.4131

[195] M. Reid: *Undergraduate commutative algebra*. Cambridge University Press, 1995.

[196] M. Reid: *Undergraduate algebraic geometry*. Cambridge University Press, 1988.

[197] W. Rinow: *Die innere Geometrie der metrischen Räume*. Springer, 1951.

[198] J. Roe: *Lectures on coarse geometry*. American Mathematical Society, 2003.

[199] M. E. Rudin: A new proof that metric spaces are paracompact. Proc. Amer. Math. Soc. **20**, 603 (1969).

[200] M. E. Rudin: $\aleph$-Dowker spaces. Czech. Math. Journ., **28**, 324-326 (1978).

[201] W. Rudin: *Principles of mathematical analysis*. McGraw-Hill, 1953, 1964, 1976.

[202] W. Rudin: *Real and complex analysis*. McGraw-Hill, 1966, 1974, 1986.

[203] W. Rudin: *Functional analysis*. McGraw-Hill, 1973, 1991.

[204] V. Runde: *A taste of topology*. Springer, 2005.

[205] H. Sagan: *Space-filling curves*. Springer, 1994.

[206] H. Samelson: On the Perron-Frobenius theorem. Michigan Math. J. **4**, 57-59 (1957).

[207] E. Schechter (ed.): *Handbook of analysis and its foundations*. Academic Press, 1997.

[208] C. van Schoubroeck: The van Kampen theorem. Bachelor Thesis. Nijmegen, 2014.

[209] J.-P. Serre: *Topics in Galois theory*. Jones and Bartlett Publishers, 1992.

[210] J. Siegel: A New Proof of Caristi's Fixed Point Theorem. Proc. Amer. Math. Soc. **66**, 54-56 (1977).

[211] L. Smith: $\pi_1(S^N) = 0$. Kyungpook Math. Journ. **8**, 13-14 (1968).

[212] A. Sokal: A really simple elementary proof of the uniform boundedness theorem. Amer. Math. Monthly **118**, 450-452 (2011).

[213] E. H. Spanier: *Algebraic topology*. McGraw-Hill, 1966. Springer, 1981.

[214] E. Sperner: Neuer Beweis für die Invarianz der Dimensionszahl und des Gebietes. Abh. Math. Sem. Univ. Hamb. **6**, 265-272 (1928).

[215] M. Spivak: *Calculus on manifolds*. Addison Wesley, 1965.

[216] L. A. Steen, J. A. Seebach: *Counterexamples in topology*. Holt, Rinehart and Winston Inc. 1970, Springer, 1978, Dover, 1995.

[217] E. M. Stein, R. Shakarchi: *Fourier analysis*. Princeton University Press, 2005.

[218] E. M. Stein, R. Shakarchi: *Real analysis*. Princeton University Press, 2005.

[219] A. H. Stone: Hereditarily compact spaces. Amer. Journ. Math. **82**, 900-916 (1960) and erratum **86**, 888 (1964).

[220] R. E. Stong: Finite topological spaces. Trans. Amer. Math. Soc. **123**,325-340 (1966).

[221] T. Szamuely: *Galois groups and fundamental groups*. Cambridge University Press, 2009.

[222] W. Takahashi: Existence theorems generalizing fixed point theorems for multivalued mappings. pp. 397-406 *in* M. A. Théra, J.-B. Baillon (eds.): *Fixed point theory and applications*. Longman Scientific and Technical, 1991.

[223] T. Tao: *Analysis 1*. Hindustan Book Agency, 2006, 2014.

[224] T. Tao: *Analysis 2*. Hindustan Book Agency, 2006.

[225] V. V. Tkachuk: *A C_p-Theory problem book*. Springer, 2011.

[226] T. tom Dieck: *Algebraic topology*. European Mathematical Society, 2008.

[227] O. Ya. Viro, O. A. Ivanov, N. Yu. Nesvetaev, V. M. Kharlamov: *Elementary topology. Problem Textbook*. American Mathematical Society, 2008.

[228] H. Völklein: *Groups as Galois groups*. Cambridge University Press, 1996.

[229] S. Waldmann: *Topology. An introduction*. Springer, 2014.

[230] R. C. Walker: *The Stone-Čech compactification*. Springer, 1974.

[231] A. H. Wallace: *Algebraic topology. Homology & Cohomology*. Benjamin, 1970, Dover, 2007.

[232] W. Walter: *Ordinary differential equations*. Springer, 1998.

[233] W. S. Watson: Pseudocompact metacompact spaces are compact. Proc. Amer. Math. Soc. **81**, 151- 152 (1981).

[234] A. Weil: *Basic number theory*. Springer, 1967, 1974.

[235] A. Weil: *L'intégration dans les groupes topologiques et ses applications*. Herman, 1940.

[236] H. Whitney: *Complex analytic varieties*. Addison-Wesley, 1972.

[237] S. Willard: *General topology*. Addison-Wesley 1970, Dover 2004.

[238] D. Wilson: This ain't no meager theorem. http://www.math.washington.edu/~morrow/336_09/papers/Dylan.pdf

[239] L. A. Wolsey: Cubical Sperner lemmas as applications of generalized complementary pivoting. J. Combin. Theor. (A) **23**, 78-87 (1977).

[240] V. A. Zorich: *Mathematical Analysis I*. Springer, 2004.

Index

$C(X)$, 85
$C(X, Y)$, 85
D^n, 102, 115
S^n, 102, 108, 162
$S^n_\pm$, 104
βX, 210
$\ell^p(S)$, 367, 368
$\ell^p(S)$, 47
γX, 266
$\pi_0(X)$, 289
$\pi_1(X, x_0)$, 315
$\pi_c(X)$, 260
$\mathbb{R}\mathrm{P}^n$, 104
σ-discrete family, 228
σ-locally finite family, 233
$\mathcal{C}(X)$, 157

AC, 338
ACC, 339
accumulation point
    ω-, 41
    complete, 41, 123
    of net, 80
    of sequence, 77
    of subset, 40, 139
ACDC, 339
adjoint functor, 349
Alexander's subbase lemma, 122, 131, 132, 137
algebra
    character, 377
    ideal, 377
Axiom
    of choice, 338
    of countable choice, 339
    of countable dependent choice, 55, 339

Banach space, 47
base, 63
Boolean algebra, 31, 36, 274, 346
    complete, 264
boundary, 35
bounded
    function, 22
    set in metric space, 22
box topology, 106, 132

Cantor (middle third) set, 269
categories
    equivalence, 322, 349
    isomorphism, 317, 348
category, 346
    opposite, 347
character of an algebra, 377
clopen, 31
closed ball, 27
closed set
    in metric space, 21
    in topological space, 30
closure, 33
cofinal, 80
compactification, 156
    Alexandrov, 161
    one-point, 161
    Stone-Čech, 210, 279
complete accumulation point, 41, 123
completely separated, 196
completeness
    of $C_b(X, Y)$, 48
completion, 50
component
    connected, 241
    path-, 289
condensation point, 41
connected
    component, 241
continuity
    at a point, 83
    joint, 112
    separate, 112
Continuum Hypothesis, 110, 218
contraction, 359
    principle, 359

weak, 360
contraposition, 331
convergence
of a function, 84
of filter, 82
of net, 78
of sequence, 75
convex set
in $\mathbb{R}^n$, 152, 243
in ordered space, 70
coproduct
of topological spaces, 95
coproduct of sets, 336
cover, 119
point-finite, 194
covering dimension, 258
Croft's lemma, 56
curve, 281
Jordan, 294
Peano, 281
cut-point, 244

dense, 39
derived set, 40
diagonal, 108, 201
diam(eter), 22
dimension
covering, 258
large inductive, 257
separation-, 255
small inductive, 257
direct product
of sets, 337
of topological spaces, 105
directed set, 77
discrete family, 228
disjoint union, 336
dist(A, B), 22
dist(x, A), 22
dyadic rationals, 196, 199, 298

embedding, 94, 127, 172
endomorphism (in a category), 315
epimorphism, 86, 115, 347
equivalence relation, 334
closed, 99
open, 99
smallest, 335
eventually, 78

extension of continuous functions
existence, 59, 129, 167, 173, 199, 234
uniqueness, 86
extrinsic property, 208

fiber product, 115
filter, 37, 81, 133
convergence of, 82
maximal, 133
finite intersection property, 138, 145
fixed point theorem
of Banach, 359
of Brouwer, 253
of Caristi, 361
of Schauder, 385
Fréchet space, 387
frequently, 80
Fun, 331, 347
function
bounded, 22
closed, 88
continuous, 22, 85
open, 88
proper, 168
semicontinuous, 89, 234, 361
uniformly continuous, 60
functor, 347
contravariant, 348

generic point, 42
graph of a function, 108, 131
groupoid, 314

hereditary property, 29, 65
higher
connectedness, 249
homotopy groups π_n, 319
Hilbert
cube, 370
space, 369
homeomorphism, 22, 87, 127
homotopy
equivalence, 308
groups π_n, 319
of functions, 307

ideal in an algebra, 377
idempotent, 217, 275
inductive dimension

large, 257
small, 257
interior, 33
intrinsic property, 208
inverse function theorem, 246
isolated point, 29
isometry, 22
isomorphism (in a category), 347

lattice, 345
bounded, 345
distributive, 345
Lebesgue number, 142
left inverse, 338
length of path, 299
lexicographic ordering, 70
limit, 21
locally finite family
of functions, 205
of subsets, 205

Martin's axiom, 110, 218
maximal element, 340
measure zero, 79, 285–287, 384
metric, 19
discrete, 29
discrete (standard), 20
equivalence of, 27
space, 20
uniformly discrete, 147
metric space
complete, 47
completely metrizable, 49, 164
geodesic, 297
geodesically complete, 304
length space, 303
metrically convex, 297
proper, 175
properly metrizable, 175
totally bounded, 141
with Lebesgue property, 142
midpoint
approximate, 303
strict, 297
Minkowski's inequality, 21
monoid, 315
monomorphism, 347

natural transformation, 349

neighborhood, 37
base, 70
filter, 37
open, 37
neighborhood filter, 82
net, 77
convergent, 78
universal, 135
norm, 20, 150
equivalence of, 28
nowhere dense, 39

object
initial, 215, 347
terminal, 170, 347
open ball, 25
open set
in metric space, 25
in topological space, 26
opposite category, 347
order
(partial), 340
linear, 340
partial, 334
total, 340

partition of unity, 204
path, 289
composite, 289
length of, 299
parametrized by arc-length, 302
rectifiable, 299
reversed, 289
path-component, 289
point
generic, 42
isolated, 29
of continuity, 83
point-finite, 194
preorder, 334
profinite
group, 273
space, 272
projective space, 104
proper
function, 168
metric space, 175
property
extrinsic, 208

hereditary, 29, 65
intrinsic, 208
topological, 87, 208
proximity, 221
pseudometric, 19, 43, 202, 212
pullback, 115
pushout, 114

quasi-component, 242, 262
quotient map, 97

rectifiable, 299
refinement, 225
regular closed, open, 36
relation, 334
equivalence, 334
order, 334
relatively compact, 125
reparametrization, 313
retraction, 245
of $X \to \partial X$ does not exist, 254
of Cantor set to closed subsets, 270
Riemann integral, 79
right inverse, 338

saturated, 98
saturation, 98
section, 338
semicontinuity, 89, 234
seminorm, 20
separated, 190
separation-dimension, 255
sequence, 21
Cauchy, 47
converges, 21, 75
sequence space, 367
compact subsets, 369
completeness, 367
separability, 368
set
F_σ, 55
G_δ, 59
G_δ, 55
bounded, 22
closed, 21, 30
convex, 243
cozero-, 208
dense, 39
directed, 77
functionally closed, 197
functionally open, 208
locally closed, 164
nowhere dense, 39
open, 25, 26
perfect, 40
regular closed, 36
regular open, 36
star-shaped, 290
zero-, 197
shrinking, 194
singleton, 29
Sorgenfrey line, 67, 70, 72, 120, 190, 191, 262
Sorgenfrey plane, 108, 191, 208
space
above another, 116
Banach, 47
below another, 114
Fréchet, 387
Hilbert, 369
locally convex topological vector, 234, 386
metric, 20
normed, 20
topological, 26
uniform, 52
specialization preorder, 42
spectrum
Gelfand — of commutative Banach algebra, 163
maximal — of commutative ring, 352
of a Boolean algebra, 276
prime — of commutative ring, 351
Sperner's lemma, 250
sphere, 102, 108, 162
subbase, 69
lemma of Alexander, 122, 131, 132, 137
subcategory, 347
full, 347
reflective, 350
subcover, 119
subnet, 80
subsequence, 77
subspace, 26
support, 204
supremum, 136
symmetric, 356

Theorem

Baire's, 53
Banach-Steinhaus, 380
Borsuk-Ulam, 311
Brouwer's fixed point, 253
Caristi's fixed point, 361
closed graph, 382
Dini's, 147
Dugundji's extension, 234
Ekeland's variational, 363
Fundamental – of Algebra, 151, 327
Hahn-Mazurkiewicz, 293
Heine-Borel, 149
Hopf-Rinow, 304
intermediate value, 245
Kolmogorov-Riesz-Fréchet, 370
Lavrentiev, 60
Menger, 298
Morse-Sard, 286
Nagata-Smirnov metrization, 233
open mapping, 382
Poincaré-Miranda, 252
Stone-Weierstrass, 374
Tietze-Urysohn extension, 199
Tychonov, 131, 134, 136
uniform boundedness, 380
Urysohn's metrization, 201
Weierstrass, 375
topological group, 162, 261, 355
topological property, 87, 208
topological space, 26
T_0, 42
T_1, 31
T_2, 31
T_3, 185
T_4, 185
T_5, 189
T_6, 197
$T_{2.5}$, 209
$T_{3.5}$, 207
σ-compact, 166, 167
Čech-complete, 218
Alexandrov, 43
Baire, 53
cocountable, 28, 42
cofinite, 28, 42
collectionwise normal, 232
compact, 119
compactly generated, 182
completely Hausdorff, 209
completely metrizable, 49
completely normal, 189
completely regular, 207
connected, 31, 85, 239, 243
contractible, 309
countably compact, 138
countably paracompact, 232
dense-in-itself, 40
discrete, 28
extremely disconnected, 263
first countable, 71, 76, 89
fixed point property, 245
Fréchet, 76, 98
generalized ordered, 70, 190
Hausdorff, 31
hemicompact, 166
hereditarily compact, 126
hereditarily disconnected, 259
hereditarily normal, 189
hyperconnected, 42
indiscrete, 28
irreducible, 41
k-, 182
Lindelöf, 119, 138
locally compact, 151, 159
locally compact, strongly, 163
locally connected, 241
locally Euclidean, 160, 292
locally path-connected, strongly, 292
locally path-connected, weakly, 292
metacompact, 226
metrizable, 26
Noetherian, 126
normal, 185
ordered, 69
paracompact, 225
paracompact M-, 235
path-connected, 290
Peano, 294
perfectly normal, 197
Polish, 220
profinite, 272
pseudocompact, 146, 201, 206
realcompact, 211
reducible, 41
regular, 185
rim-compact, 224

scattered, 40
second countable, 64, 89
semiregular, 68, 189
separable, 64
sequential, 76, 98, 182
sequentially compact, 140
simply connected, 316
smallest neighborhood, 43, 292
Stone, 263
Stonean, 263
strongly Hausdorff, 209
strongly zero-dimensional, 266
supercompact, 137
totally disconnected, 259
totally separated, 261
weakly countably compact, 140
weakly Hausdorff, 125
with Knaster property, 111
with Souslin property, 66, 110
zero-dimensional, 262
topological vector space, 355
locally convex, 234, 386
topology, 26
$\sigma(V^*, V)$-, 384
box, 106
coarser, 29, 88
compact-open, 178
coproduct, 95
direct sum, 95
Euclidean, 28
final, 91
finer, 29, 88
generated by family of sets, 69
induced, 26
initial, 92
order, 69
product, 105
quotient, 97
subspace, 26
usual, 28
weak$-*$, 384
Zariski — on Spec(R), 351
Zariski — on affine space, 30, 352
Torus, 102, 327
transfinite induction, 344
transfinite recursion, 344
trigonometric polynomial, 375

ultrafilter, 133, 279
principal, 279
ultranet, 135
uniform space, 52
unitization, 177
universal property
of βX, 215
of X_∞, 170
of direct product of sets, 337
of direct product of spaces, 105
of direct sum of spaces, 96
of disjoint union (coproduct) of sets, 337
of pullback of spaces, 116
of pushout of spaces, 114
of quotient spaces, 99
universally closed, 52, 211
Unordered sums, 79
upper bound, 340
Urysohn
"Lemma", 195
metrization theorem, 202

vector field, 254, 310

weak contraction, 360
weight, 64, 128, 211
well-ordered, 342, 343

zero-set, 197
Zorn's Lemma, 340