Publishers' page

Publishers' page

Publishers' page

Publishers' page

Preface

Philosophical questions about the nature of abstract objects and the concept of infinity have always been part of mathematics, even in antiquity. In the modern era, one has only to recall the 17th century controversy over the legitimacy of imaginary numbers, or Gauss's famous rejection of actual (as opposed to potential) infinity, to see how deep the history goes. But the foundational crisis at the beginning of the 20th century — principally provoked by the discovery of Russell's paradox — sharpened the debate and led to the formation of three main foundational schools: set-theoretic platonism, intuitionism, and formalism.

The purpose of this book is to develop an alternative point of view. In the philosophy of mathematics there are two fundamental axes: one which deals with the general nature of mathematical truth and existence, ranging through various forms of platonism and constructivism, and another which deals with the legitimacy of infinity, ranging from ultrafinitism at one extreme to acceptance of various large cardinals at the other. My goal here is to examine the consequences of adopting *truth constructivism*, the idea that truth values are, in some sense, only available in stages — or, what is morally equivalent (Section 4.2), every fact can in principle be known — together with *countablism*, the doctrine that all sets are countable. This combination, or some version of it, has been considered many times before under various names. The simple, descriptive term for it is "constructive countablism".

(Maybe I should say "the purpose of *these* books", because there are really two: Part 1: Philosophy, and Part 2: Mathematics. They have been written to be, as much as possible, readable independently. Feel free to start with Part 2 if you like.)

I will argue that constructive countablism is philosophically justified in

a way that Zermelo-Frankel set theory (ZFC) is not. The complaint that ZFC is conceptually dubious is often met with some version of "Who cares, as long as it's consistent?" But as I pointed out in [52], we should care because consistency does not entail soundness. It could be that ZFC is consistent but proves false statements of first order arithmetic. I do not know any convincing reason to feel confident ZFC is arithmetically sound, and I think there are legitimate reasons to suspect that it might not be (Section 7.4).

Having said that, I must emphasize that I am not demanding that anyone stop studying ZFC, or ZFC plus various large cardinal axioms, or any other formal system they think is interesting. I do feel it is important to clearly distinguish formal systems that have a solid philosophical justification from those that do not, no matter how strong the empirical evidence for their consistency might be. It's important to me, anyway.

Countablism is not as crippling as one might think. It is reasonable to assume that Bertrand Russell was right when he said that rejecting uncountable sets would mean that "the irrationals, and the real numbers generally, can no longer be adequately dealt with" ([40], p. xiv). But that is not so. Uncountable collections such as the real line still exist as proper classes and can be adequately dealt with as such. A rough rule of thumb is that in countablistic mathematics sets that are classically countable survive as sets, sets of cardinality at most $2^{\aleph_0}$ become classes, and classes become concepts (Section 7.7). Countablism does not stop us from doing normal mathematics; to the contrary, its excellent match with actual mathematical practice — both in validating the really essential things and in eliminating a vast swath of set-theoretic pathology (how often do you work with objects whose cardinality is strictly larger than $2^{\aleph_0}$?) — may be counted as a point in its favor (Section 7.5).

But the theoretical case for countablism does not hinge on its practical utility; rather, it is based on the premise that the key property that distinguishes sets from proper classes is *surveyability*: any set could, in principle, be surveyed to completion, whereas it is, in principle, impossible for an element-by-element survey of a proper class to ever terminate. An argument can then be made that we have a relatively clear conception of what it would be like to survey a countable collection, but no conception at all of what it would be like to survey an uncountable one. I try to make this

argument in Chapter 7.

It may be a difficult argument to make for an audience that is already firmly convinced the real line is a set. I suspect that many mathematicians learn about the real line before they learn about the set/class distinction, when "set" is still naively synonymous with "collection" and they therefore have no choice but to accept it as a set. If this idea has hardened by the time classes are introduced, then at that point the suggestion that the real line might be a proper class will probably seem absurd. But that is merely because one has learned the topics in the wrong order. Once one realizes this, I think there is a good case for countabilism.

The case for constructivism, I believe, goes beyond "good" to "decisive". I say that because I can resolve the classical semantic paradoxes such as the liar paradox ("this sentence is not true") and the semantic form of Russell's paradox (the concept of being a concept that does not fall under itself), and intuitionistic logic is an essential component of their resolutions. This makes sense because the characteristic feature of intuitionistic logic, the loss of the law of excluded middle ("p or not-p") as a global principle, reflects skepticism that every sentence necessarily has a well-defined truth value, skepticism which is quite appropriate in relation to the liar sentence and its relatives.

Now, the liar paradox is as old as time, and many, many attempts have been made to resolve it, but they all have fairly obvious flaws if one looks at them critically. There is a vast literature on the subject, and I think it is fair to say that no previous attempt has attracted widespread support. My solution is very different from others — in particular, it does not suffer from a "revenge problem", for those familiar with this term — and I am confident that it will be recognized as essentially definitive. See Chapter 6 for a summary of the solution, and Chapters 3 through 5 for a fuller explanation.

When my approach to the liar paradox is formalized, the resulting formal system, which uses intuitionistic logic, is provably consistent (Section 6.5). But if we add the law of excluded middle it becomes inconsistent. This is why I feel that the case for constructivism is decisive: philosophical positions can be debated endlessly, but here they are trumped by the brute fact of inconsistency.

Perhaps this is not a convincing argument for readers whose primary

interest is mathematics and who may regard the semantic paradoxes as not relevant to their main concerns. But I think this attitude is mistaken. Consider the notion of a proper class. By definition, the elements of a proper class cannot be explicitly listed out, not even in principle. Proper classes cannot be surveyed; they do not exist by virtue of any conceivable physical embodiment. Rather, they exist by virtue of our possession of a membership criterion — a verbal description which tells us, as new objects are encountered, whether they belong to the class. Thus they are inherently linguistic objects, and the class-theoretic version of Russell's paradox therefore properly belongs to the family of semantic paradoxes. So we cannot, as mathematicians, dismiss the semantic paradoxes as someone else's problem. They are our problem too. And this problem can only be resolved using intuitionistic logic. And so I think we really have no choice but to be constructivists.

According to the conception of sets as surveyable collections, we can still affirm the law of excluded middle for sentences that only quantify over sets: if we can decide whether any given element has some property, then in principle we could survey the set in its entirety to check whether they all do, which implies that this question must have a definite yes-or-no answer. So within the world of mathematics the switch to intuitionistic logic that I think is demanded as a proper response to the paradoxes only applies to sentences that quantify over proper classes. While such sentences do sometimes appear in ordinary mathematics ("every compact Hausdorff space is normal"), the consequences of losing excluded middle for them are, as far as I can tell, virtually negligible if everything else is left the same.

But countablists do not leave everything else the same. Since we consider so many ordinary mathematical objects to be proper classes, the need to use intuitionistic logic when quantifying over proper classes has significant implications for us. However, we can at least affirm the law of excluded middle for purely arithmetical assertions, since the truth value of any arithmetical sentence can be evaluated by a computation of countable length. Therefore a rough summary of my position is that we should be:

- platonists about number theory
- intuitionists about analysis
- formalists about axiomatic set theory

(but, to clarify, when I say "be intuitionists" I mean "reject unrestricted use of the law of excluded middle", and not "reject any kind of actual infinity", as intuitionists also do). Constructive countablism can be seen as a sort of synthesis of the three main foundational schools.

Although constructive countablism combines elements of platonism, intuitionism, and formalism, it is closest to yet another foundational school, predicativism. Though little remembered today, this line of thought had an illustrious beginning, originating in ideas of Poincaré and Russell, and then being taken up by such luminaries as Skolem and Weyl. But it quickly faded into obscurity, probably for two reasons: first, its apparent incompatibility with the bulk of ordinary mathematics, and second, its lack of a committed advocate, as the other schools had in Zermelo, Brouwer, and Hilbert. Russell, for instance, was probably more concerned about logicism than predicativism, which may be why he was willing to abandon predicativist principles in his *Principia Mathematica* [40] when they seemingly blocked him from his goal of deriving mathematics from logic. (I refer to his adoption of the impredicative "axiom of reducibility".) Likewise, Weyl "converted" to intuitionism not long after writing his landmark predicativist treatise *Das Kontinuum*; indeed, I will argue in Section 7.1 that he was at some level already an intuitionist when he wrote *Das Kontinuum*, he just didn't realize it. Poincaré and Skolem, on the other hand, never abandoned predicativism, but their remarks about it were brief and unsystematic. So there was no one giving a full-throated defense of this view.

The main concern of predicativism is the so-called "vicious circle principle". This principle maintains that no object can be defined by means of a condition that quantifies over any collection to which that object belongs. Such a definition is said to be "impredicative". While this prohibition could be paired with a variety of views about infinity, the predicativism of Russell et al. is nowadays broadly understood to specifically urge countablism. (This is the established interpretation of predicativism, but it is actually a significant retcon; of the four founding fathers, only Skolem clearly and consistently advocated countablism. Poincaré and Weyl explicitly rejected it (Section 7.1). And Russell's "adequately dealt with" comment quoted above also comes pretty close to an outright rejection.) So I take the core premises of predicativism as it is now understood to be the vicious circle principle and countablism, which is not quite the same as my stance of

truth constructivism plus countablism, but is certainly similar.

Indeed, I will argue that the vicious circle principle only makes sense in the context of truth constructivism (Section 1.4). So the move from predicativism to constructive countablism is a natural development. I felt it was appropriate to introduce a new term for what I do in this book because circularities that I think traditional predicativists would reject, such as the concept of being a concept, are central to my treatment of the paradoxes. But before I worked out my ideas about the paradoxes I considered myself a predicativist, and if anyone wants to say I still am, I would not really object.

Now let me summarize the contents of the book.

Part 1 is about the philosophical basis of constructive countablism. I begin this part with a brief introductory chapter about Russell's paradox, and then in Chapter 2 I move into a general discussion of the nature of "abstract objects" in mathematics. I am no philosopher (as will be obvious to the reader, I am sure), but that is not necessarily a disadvantage here. To the contrary, it is my mathematical background which, despite my philosophical ignorance, gave me the resources to understand mathematical objects in a way that, as far as I know, philosophers have not. For instance, my treatment of the "type/token" distinction in terms of identity criteria (Sections 2.2 and 2.3) should come as no surprise to students of constructivism: Bishop and Bridges teach us, right at the beginning of their classic text ([4], p. 5), that

> To define a set we prescribe, at least implicitly, what we (the constructing intelligence) must do in order to construct an element of the set, and *what we must do to show that two elements of the set are equal* [emphasis added].

So I expect that anyone familiar with constructivism would easily grasp that switching between the philosopher's "types" and "tokens" need involve nothing more mysterious than modifying the prescription of what we must do to show that two things are equal. On the other hand, my treatment of abstract objects such as numbers and sets in terms of the notion of a conservative extension is based on a comment made by Terry Tao on MathOverflow [46]. The question asked there was: given that the various constructions of the real line (Cauchy completion of the rationals, Dedekind cuts, etc.) do not produce literally the same object, what is the "true" real

line? And the illuminating answer was that we use any of these constructions to prove that a complete ordered field exists, and then we add a new constant symbol, "$\mathbb{R}$", to our language, together with axioms asserting that it is a complete ordered field. Since we already knew that complete ordered fields exist, this will "conservatively" extend whatever reasoning we had previously been using, not adding any true deductive power. The point of making this extension is to allow us to talk about the real line abstractly, in a manner divorced from the details of any particular construction. It did not take much originality for me to see how similar ideas applied to numbers or sets could demystify their inscrutably abstract nature.

As I already mentioned, Chapters 3 through 5 explain how to resolve the classical semantic paradoxes, and this solution is summarized in Chapter 6. Then in Chapter 7 I present some arguments in favor of countablism, and in Chapter 8 I discuss the concept of well-ordering as understood from a constructively countablist perspective.

The aim of Part 2 is to show how to do mathematics on a constructively countablist basis. Here my approach changes somewhat: while the "sets" and "classes" in Part 1 are conceptually distinct, in Part 2 I follow the usual mathematical practice of taking a set to be a special kind of class. Also, in Part 1 I treat sets and classes in relation to an unrestricted pool of objects, but in Part 2 it becomes convenient to reduce the universe of relevant objects. Though I present it in somewhat different terms, the mathematical development I offer is essentially a version of third order arithmetic based on natural numbers, sets of natural numbers, and classes of sets of natural numbers modulo some equivalence. Thus every class is effectively a subquotient of $\mathcal{P}(\mathbb{N})$. This may seem far too restrictive an environment in which to do serious mathematics, but don't judge that too quickly; there's a lot of room in $\mathcal{P}(\mathbb{N})$.

Chapter 9 explains our foundational framework and develops the fundamentals of abstract mathematics (functions and relations, products, quotients, the real line, partial and total orders). The remaining seven chapters cover what I would consider a strong undergraduate curriculum, with the intention of explicitly showing how it all can be done by constructive countablists. I noted above that in the context of countablism, the consequences of constructivism are substantial (though not as great as they are for intuitionists), but it turns out that we can still do rather a lot. Of course we have to alter many of the definitions, and some proofs have to be adjusted too, but I claim that essentially everything from a typical undergraduate curriculum, suitably modified, can be accomodated.

Versions of many standard theorems are presented here. In Chapter 10: the Cauchy-Schwarz inequality in $\mathbb{R}^n$; completeness of $\mathbb{R}$; Banach's fixed-point theorem; Cauchy completion; the Baire category theorem; ratio and root tests; radius of convergence of a power series; the extreme value theorem; Lebesgue number; the Bolzano-Weierstrass theorem; the Heine-Borel theorem; the gluing lemma; Urysohn's lemma; the Tietze extention theorem; universal property of product topologies; Tychonoff's theorem; Urysohn's metrization theorem; embedding in cubes; universality of the Cantor space among compact spaces.

Chapter 11: Cayley's theorem; universal property of products and quotients of groups and rings; Lagrange's theorem; the three isomorphism theorems for groups and rings; Bézout's identity; Euclid's lemma; Euler's theorem; Fermat's little theorem; existence of maximal ideals; the relations between Euclidean domains, PID's, and UFD's; Wilson's theorem; Fermat's sum-of-two-squares theorem; uniqueness of the countable atomless Boolean algebra; the prime ideal theorem; Stone duality.

Chapter 12: rank-nullity; determinants; the fundamental theorem of Galois theory; the fundamental theorem of algebra; existence and uniqueness of algebraic closures; the structure theorem for finitely generated modules over a PID; classification of finitely generated abelian groups; the existence of Jordan forms for complex matrices.

Chapter 13: Rolle's theorem; the mean value theorem; Taylor's theorem; the mean value inequality; multivariable chain rule; the inverse function theorem; the implicit function theorem; regularity of Lebesgue measure on $\mathbb{R}^n$; Hahn-Jordan decomposition; Lebesgue decomposition; the monotone convergence theorem; Fatou's lemma; the dominated convergence theorem; the fundamental theorem of calculus; Cavalieri's principle; Tonelli's theorem; Fubini's theorem; the Radon-Nikodym theorem; Cauchy's theorem; Cauchy's formula; power series representation; the identity theorem; Liouville's theorem; the maximum modulus principle; infinite differentiability of analytic functions; Morera's theorem; Picard's existence theorem.

Chapter 14: Simplicial approximation; the fundamental group; Brouwer's fixed point theorem for the disk; existence and uniqueness of universal covers; the Seifert-van Kampen theorem; fundamental groups of handlebodies and cross-cap surfaces; classification of compact surfaces; homology of a cone; topological invariance of simplicial homology; homology of balls, spheres, and $\mathbb{R}^n$; invariance of dimension; Brouwer's fixed point theorem for $\mathbb{D}^n$.

Chapter 15: the Stone-Weierstrass theorem; the Hahn-Banach theorem;

the approximation lemma; the open mapping theorem; the Banach isomorphism theorem; the closed graph theorem; the principle of uniform boundedness; the Banach-Steinhaus theorem; the Banach-Alaoglu theorem; the dual of $L^p(X)$; the Riesz-Markov theorem; the Cauchy-Schwarz inequality in Hilbert space; Bessel's inequality; the Riesz representation theorem; the spectral theorem in finite dimensions; the Riemann-Lebesgue lemma; the Fourier inversion theorem.

Chapter 16: smooth partition of unity; baby Whitney embedding; tensor fields and differential forms; the exterior derivative; Stokes' theorem; Green's theorem; the divergence theorem; Riemannian manifolds as metric spaces; the Levi-Civita connection; the exponential map; local existence of geodesics; the Hopf-Rinow theorem.

Nik Weaver
nik.weaver@gmail.com

Contents

Preface vii

Philosophy **1**

1. Russell's Paradox 3
   - 1.1 The paradox . . . 3
   - 1.2 ZFC . . . 5
   - 1.3 The iterative conception . . . 7
   - 1.4 The vicious circle principle . . . 12
   - 1.5 Indefinite extensibility . . . 15
2. Abstract objects 19
   - 2.1 Concrete objects . . . 19
   - 2.2 Identity criteria . . . 20
   - 2.3 Possible worlds . . . 26
   - 2.4 Proxies . . . 29
   - 2.5 Sets . . . 33
   - 2.6 Pure sets . . . 37
3. Truth 41
   - 3.1 Classes . . . 41
   - 3.2 The liar paradox . . . 43
   - 3.3 Sentences and predicates . . . 47
   - 3.4 The T-scheme . . . 49
   - 3.5 Deflationary truth . . . 53
   - 3.6 Truth values . . . 57

4. Constructivism 61

4.1 Assertibility 61
4.2 Capture and release 66
4.3 Theory of meaning 69
4.4 Intuitionistic logic 72
4.5 Axiomatizing assertibility 74

5. Essential Notions 79

5.1 Meaningfulness 79
5.2 Axiomatizing meaningfulness 83
5.3 Propositions, concepts, and relations 87
5.4 Classes 90

6. Interlude: The Liar 95

6.1 The two versions 95
6.2 The three ingredients 96
6.3 Meaningfulness and truth 98
6.4 Formalization 101
6.5 Consistency 105

7. Countablism 111

7.1 Predicativism 111
7.2 Finitism 114
7.3 Cantorianism 118
7.4 Consistency versus soundness 121
7.5 Normal mathematics 126
7.6 Skolem and the multiverse view 132
7.7 Countablistic mathematics 136

8. Well-ordering 141

8.1 Induction on $\mathbb{N}$ 141
8.2 Well-ordering and induction 143
8.3 Full well-ordering 146
8.4 Iterated classical truth 150
8.5 Reaching Γ_0 152
8.6 Going further 156

Mathematics **159**

9. Fundamentals 161

9.1 Constructive countablism 161
9.2 Intuitionistic logic 166
9.3 Classes and sets 171
9.4 Meaningfulness and assertibility 177
9.5 Reasoning about classes 181
9.6 Relations and functions 185
9.7 Products 187
9.8 Quotients 191
9.9 Number systems 194
9.10 Order 198

10. General Topology 203

10.1 Metric spaces 203
10.2 Infinite series 214
10.3 Compactness 217
10.4 Topological spaces 228
10.5 Coherence 244

11. Abstract Algebra 253

11.1 Groups 253
11.2 Number theory 265
11.3 Rings 267
11.4 Integral domains 273
11.5 Boolean algebras 281

12. Linear Algebra 293

12.1 Vector spaces 293
12.2 Dimension 300
12.3 Galois theory 306
12.4 Modules 319

13. Real and complex analysis 329

13.1 Differentiation 329
13.2 Measure theory 336
13.3 Integration 350
13.4 Complex analysis 362
13.5 Ordinary differential equations 370

14. Algebraic Topology 373

14.1 Simplicial complexes 373
14.2 The fundamental group 378
14.3 Calculating $\pi_1(X)$ 383
14.4 Surfaces 390
14.5 Simplicial homology 397

15. Functional Analysis 407

15.1 Banach spaces 407
15.2 Bounded linear transformations 416
15.3 Dual Banach spaces 421
15.4 The duals of $L^p(X)$ and $C(K)$ 428
15.5 Hilbert spaces 436
15.6 Fourier analysis 442

16. Geometry 449

16.1 Smooth manifolds 449
16.2 Differential forms 456
16.3 Integration on Manifolds 463
16.4 Riemannian manifolds 468
16.5 Affine connections 472

Bibliography 481

Notation 485

Index 489

PART 1
Philosophy

Chapter 1

Russell's Paradox

1.1 The paradox

Two forms of Russell's paradox. Comprehension axioms.

In 1902 Bertrand Russell wrote to Gottlob Frege informing him of an inconsistency Russell had found in the system of logic presented in Frege's *Grundgesetze der Arithmetik*. Russell's first statement[1] of the paradox that would bear his name actually presented it in two forms:

> Let w be the predicate: to be a predicate that cannot be predicated of itself. Can w be predicated of itself? From each answer its opposite follows. Therefore we must conclude that w is not a predicate. Likewise there is no class (as a totality) of those classes which, each taken as a totality, do not belong to themselves.[2]

The first is a *semantic* version of the paradox, and the second is a *set-theoretic* version. Nowadays the set-theoretic version is better known, but from Frege's point of view sets were secondary to language: they arose as the "extensions" of concepts expressed by predicates.[3] E.g., the predicate "is a natural number greater than 1 whose only divisors are 1 and itself" expresses the concept of primality, whose extension is the set of all prime

[1] Zermelo independently discovered the paradox around the same time as Russell, but evidently did not consider it important enough to publish. The earliest known documentation of the discovery was a private note of Husserl, dated April 1902, which recorded that Zermelo had communicated it to him [34].

[2] [35], p. 125.

[3] Russell too thought that every class must have a "defining characteristic": "The vital difference between [a class and a defining characteristic of it] consists in the fact that there is only one class having a given set of members, whereas there are always many different characteristics by which a given class may be defined." ([38], p. 13).

numbers.

In order to come to a complete resolution of Russell's paradox we will need to consider both its semantic and set-theoretic versions, but let us start with the latter. Let me first restate it slightly. Forget about "extensions" and just take the notion of a collection to be primitive.[4] Then it would seem that

> the collection of all those collections which are not members of themselves

could neither be a member of itself nor fail to be a member of itself. Either statement leads to a contradiction. This puts the paradox in its barest and most maddening form.

We will get to the question of whether it is even possible for a collection to be a member of itself later, in Chapter 2. (Spoiler alert: it isn't.) If one is willing to stipulate that this cannot happen, then the paradox can be simplified by just considering the collection of all collections, as this collection evidently would have to belong to itself. But this move invites the rejoinder that, well, maybe some collections actually can belong to themselves after all. Russell's paradox neatly evades this issue by specifically excluding all self-containing collections, if there could be any such things, from the discussion.

Russell's construction hinges on the existence, for a certain property ϕ, of the collection of all the things which have that property. In symbols, this is the claim that there is a set $\{x : \phi(x)\}$ of those and only those x for which ϕ holds. A statement of this type is known as a *comprehension axiom* — or *comprehension scheme*, if it is formulated as a distinct axiom for each property in some family. If there is no restriction on which properties could be used, it is called *full comprehension*.

Suspicion naturally falls on comprehension as the problematic element in the paradox, simply because when the contradiction is derived formally it appears to be the only substantive principle that comes into play. And yet, on the face of it, this principle seems completely reasonable. Is there anything more to identifying a collection than saying which things belong to it and which do not? And is that not precisely what ϕ, any ϕ at all, is supposed to do?

[4] As we will see, there is more than one notion of "collection", and understanding these distinctions is crucial to the resolution of the paradox. But that will come later.

1.2 ZFC

Zermelo's guiding principles. His axioms as instances of comprehension. The wisdom of pragmatism.

The situation is this. At a time (the early 20th century) when the notion of a set is coming to be seen as the foundational basis of all mathematics, the world has discovered that this notion, as it is naively understood, is inconsistent. (Gödel: "the amazing fact that our logical intuitions (i.e., intuitions concerning such notions as: truth, concept, being, class, etc.) are self-contradictory."[5]) The source of the trouble is not clear. What to do? Zermelo asked and answered this question about set theory as follows:

> At present, however, the very existence of this discipline seems to be threatened by certain contradictions, or "antinomies," that can be derived from its principles — principles necessarily governing our thinking, it seems — and to which no entirely satisfactory solution has yet been found. In particular, in view of the "Russell antinomy" ... of the set of all sets that do not contain themselves as elements, it no longer seems admissible today to assign to an arbitrary logically definable notion a set, or class, as its extension ... Under these circumstances there is at this point nothing left for us to do but to proceed in the opposite direction and, starting from set theory as it is historically given, to seek out the principles required for establishing the foundations of this mathematical discipline.[6]

Although a philosophical justification was developed later (see Section 1.3), it seems that at the time Zermelo first formulated his celebrated axiomatization of set theory his guiding principles were wholly, and explicitly, pragmatic. Full comprehension is apparently verboten. But obviously not all instances of comprehension are actually used in ordinary mathematical practice. Therefore we should identify which instances are truly needed and adopt only those, with a view to leaving the work of mathematicians unaffected while, hopefully, blocking any contradictions.

What are these essential instances of comprehension? For example, if u and v are sets then $\{u, v\}$ should be too. This is an instance of comprehension where $\phi(x)$ is "$x = u$ or $x = v$", and it is now known as the *pairing* axiom. Or: for any set u, the union of all the sets which appear as elements of u should be a set. This *union* axiom is an instance of comprehension

[5] [12], p. 131.
[6] [59], p. 200.

with $\phi(x)$ = "there exists $y \in u$ such that $x \in y$". Or: for any set u, there should be a set whose elements are all the subsets of u. Here we take $\phi(x)$ = "x is a subset of u" and thereby obtain the *power set* axiom. And so on. Of the nine axioms and axiom schemes of ZFC, the Zermelo-Frankel axiomatization of set theory including choice, six can be seen as special cases of full comprehension. Extensionality, foundation, and choice are the three exceptions.

Seen in this light, the now standard Zermelo-Frankel axioms may appear rather ad hoc, a hodgepodge of unrelated instances of comprehension. Von Neumann expressed this sort of opinion in surprisingly strong terms:

> There is, to be sure, a certain justification for the axioms in the fact that they go into evident propositions of naive set theory if in them we take the word "set" ... in the sense of Cantor. But what is omitted from naive set theory — and to circumvent the antinomies some omission is essential — is absolutely arbitrary.[7]

Still, I think Zermelo's approach is smart. Of course we would like a principled explanation of what error in reasoning caused the difficulties, one which would actually tell us why some instances of comprehension are legitimate while others are not. But philosophy is hard. Good ideas are often not obvious, and can be very counterintuitive. And it can be so difficult to tell when an idea is bad. In science, ideally, if you have a bad idea, an experiment will show you that it is wrong. In math, if you have a bad idea your proof won't work. But in philosophy the signs that an idea is bad are never that final. It is psychologically normal to want to rationalize away any difficulties, and usually not very hard to do so. It is too easy to feel satisfied with bad philosophical ideas. This tends to make philosophical analyses unreliable, at least in comparison with science and mathematics.

In a situation like this, looking to actual mathematical practice to show the way forward may well be more likely to succeed than delving into philosophical analysis. The only problem is that Zermelo was working at a time when the recently conceived theory of sets was dramatically reshaping the entire mathematical enterprise, opening up broad new vistas, and it would not have been clear yet which directions would turn out to be fruitful and which would not. Topics which seemed centrally important back then, like the continuum hypothesis (first on Hilbert's famous list of 23 problems), might turn out to be barely, if at all, relevant to ordinary mainstream mathematics. If we were to revisit the problem of axiomatizing set theory on

[7] [50], p. 396.

the basis of normal mathematical practice, we might today come up with a very different result. I will return to this issue in Chapter 7.

1.3 The iterative conception

Gödel's version and subsequent elaborations. The cumulative hierarchy. Problems with a constructivist interpretation. Objects constructivism. Metaphors. "Ontological priority". Poltergeist explanations. "Too big".

On the other hand, we do now have a widely accepted philosophical justification for ZFC (or at least most of ZFC; opinions differ), the so-called "iterative conception" of sets. It was first enunciated by Gödel in his brief suggestion that "a set is something obtainable from the integers (or some other well-defined objects) by iterated application of the operation 'set of' ..."[8] and more recent authors have fleshed out the idea:

> A set is any collection that is formed at some stage of the following process: Begin with individuals (if there are any) ... At stage zero ... form all possible collections of individuals ... At stage one, form all possible collections of individuals and sets formed at stage zero ... Keep on going in this way ...[9]

> The intuition behind the iterative conception is that the sets are just those collections which can be 'built up' in stages by the repeated application of certain 'set-building' operations ...[10]

> ... sets are 'formed', 'constructed', or 'collected' from their elements in a succession of stages.[11]

> In the metaphor of the iterative conception, the steps that build up sets are 'operations' of 'gathering together' sets to form 'new' sets.[12]

(The authors of these passages are Boolos, Menzel, Parsons, and Sharlow.)

In the context of pure set theory, we would start with the empty set, not the integers. Letting $V_0 = \emptyset$ and recursively defining $V_{\alpha+1}$ to be the power set of V_α for any ordinal α, taking $V_\alpha = \bigcup_{\beta<\alpha} V_\beta$ at limit ordinals, we obtain the *cumulative hierarchy* which, in ZFC, exhausts the set-theoretic

[8] [13], pp. 518-519.
[9] [6], pp. 220-221.
[10] [24], p. 40.
[11] [25], p. 338.
[12] [41], p. 637.

universe. This blocks Russell's paradox because the "Russell set" could not appear in any V_α, as there are always more sets to come, and therefore there is no Russell set according to ZFC.

I suppose the iterative conception does well enough as an informal description of the cumulative hierarchy. And surely *something* feels right about the image of a universe of sets being built up in stages but never reaching completion. But is it really a cogent justification?

The main question is how seriously we are supposed to take phrases like "building up", "forming", "gathering together", etc. Now, it is not so easy (to put it mildly) to imagine how one might actually *build* or *form* abstract objects that do not exist in space and time, but still, a minority of authors have taken the heroic course of affirming this language as literally meaningful, and have tried to give some kind of quasi-constructivist interpretation of it. Parsons, for example, while acknowledging that "the concept of collecting is highly problematic"[13], seems (in my reading) to endorse the idea that it could be an operation of the mind, or at least a "possible" operation of the mind, whatever either of those might mean.

But the problem for any interpretation along constructivist lines is to explain how it jibes with the "formation" of power sets. Imagine, if you can, being situated at a stage in the process at which all subsets of $\mathbb{N} = \{1, 2, 3, \ldots\}$ are available. How would you go about "gathering together", in any sense, all these subsets? It seems like it would take a while. Specifically, it would take a series of steps of cardinality $2^{\aleph_0}$. So we can convince ourselves that a set of that cardinality exists provided we assume that a construction of that length is possible. It clearly begs the question. One could just as well argue that a measurable cardinal must exist on the grounds that it could be formed via a construction of its own length.

Suppose we drop the idea of "gathering together" and agree that a mind can somehow "form" $\mathcal{P}(\mathbb{N})$, the power set of $\mathbb{N}$, in a single step, without having to sequentially collect every set of natural numbers. Perhaps simply grasping the concept of an arbitrary set of natural numbers would in some obscure way suffice to "form" $\mathcal{P}(\mathbb{N})$, once all sets of natural numbers are available. We still have a problem, in the constructivist interpretation, of understanding how that point could have been reached. If sets have to be formed, then presumably we cannot form any set without first forming its elements. In the case of $\mathcal{P}(\mathbb{N})$, getting to that point would have required either $2^{\aleph_0}$ distinct minds to form all sets of natural numbers, or a single

[13] [25], p. 339.

mind capable of performing $2^{\aleph_0}$ distinct acts of forming these sets. Either way the existence of uncountable cardinalities has been assumed, just as before.

The confusing image of a mind forming sets is not important for this criticism. I think it applies to any version of the iterative conception according to which the universe of sets is somehow "not there all at once" (or, as one author puts it, "in some perpetual, atemporal process of becoming"[14]). This view is a relatively mild form of constructivism usually called "potentialism"; but, as I will clarify in the next section, for my purposes "objects constructivism" is a better description of it, so that is the term I will use. My point here is that any such conception has a problem explaining how the jump from countable to uncountable is made. Any arena in which this transition could take place must have uncountability already baked in, so to speak. Of course, one could say something similar about the transition from finite to infinite, but iterative theorists generally acknowledge that they take the existence of actual infinity as an assumption. It seems that one really needs a separate assumption of this type for each infinite V_α. But then you can't say that the iterative conception justifies the power set axiom, if you need to assume the existence of power sets going in.

Note that I do not make the easier criticism that objects constructivism is simply incoherent. (Just what is an "atemporal process"? Think about that for a minute.) Its apparent nonsensicality notwithstanding, I think Russell's paradox forces something along these lines to be true. The problem is to make sense of it. In fact, I think we can do this, and I will explain how in Chapter 2.

Returning to the iterative conception: I have taken literal interpretations first, because they seem to me more honest and straightforwardly debatable. But most iterative theorists either explicitly disavow their own constructivist-sounding language, or tacitly indicate that they don't really mean it, usually by throwing in a lot of scare quotes. (Sets are not formed by gathering together their elements, they are "formed" by "gathering together" their elements, etc.) Boolos, for example — the author of the one passage displayed at the beginning of this section which does not use scare quotes — notes that "such expressions as 'stage', 'is formed at', 'earlier than', 'keep on going' ... must be exorcised from any formal theory of sets. From the rough description it sounds as if sets were continually being created, which is not the case."[15] But he does not then go on to give us any

[14] [22], p. 299.
[15] [6], p. 220.

description, rough or not, of what actually is the case. Instead, he proceeds directly to the task of formalizing set theory. This has always confused me: he frankly acknowledges the difficulty and then soldiers right on, with no further comment, as if it were not a problem at all. Perhaps Boolos believes the iterative conception as he presents it is not literally true, but the formal theory (ZFC) inspired by it is. In that case I would think the iterative conception is playing only a heuristic role. It can hardly be a valid justification if it is literally false. Strangely, Boolos' paper seems to be widely regarded as providing a firm philosophical justification of formal set theory on the basis of the iterative conception. It is as if readers fail to notice that he explicitly denies the actual correctness of this conception.

Like many other iterative theorists, Sharlow refers to the questionable language as a "metaphor". But a metaphor is supposed to be something which represents something else. What is the phrase "gathering together", for example, supposed to represent, in the nonconstructive interpretation? What could it be a metaphor for, if sets are not in any sense constructed?

Since sets are supposed to be abstract objects which do not exist in space and time, advocates of a non-constructivist interpretation usually take the language of temporal precedence as metaphorically representing some sort of abstract notion of "ontological priority". Thus sets at lower stages of the hierarchy are "prior" to sets at higher stages, not in time, but in some other, more opaque sense. As far as I can tell, nothing like "forming" or "gathering together" is any part of this picture; that part of the iterative conception simply disappears.

But this, I think, is the majority position among contemporary iterative theorists: there is nothing constructive about sets, they are merely abstract objects stratified according to some vague notion of "ontological priority". This interpretation of the iterative conception makes it an example of what I call a *poltergeist explanation*. As in: a poltergeist is a person who makes a lot of noise and throws things around, only without the person. Or, closer to home: a set is a bag containing some elements, but without the bag.[16] What is left, after the entire content of the initial explanation has been disavowed? Only some kind of conceptual afterimage, something that conveys an illusion of understanding where there is none.

Sets are built up in temporal stages, except they aren't literally built and there is no passage of time. Does that mean anything? If we could accept that sets were, somehow, literally laid down in order, then maybe

[16]I have actually read essentially this explanation of sets somewhere, many years ago. But I don't remember where.

we could understand why the universe is structured the way it is. But if all that talk about gathering and building is not meant to be taken seriously, then all we are left with is a bare assertion that the cumulative hierarchy is in fact well-ordered by rank, but no explanation of why it should be that way.

In particular, the "ontological priority" interpretation has little force against Russell's paradox. Again, if the universe of sets were somehow "not there all at once", if it literally had to be "built in stages", this would (if one understood what it meant) explain why there cannot be a set of all sets. But if it does exist all at once, if all the sets are simultaneously present, arranged in their hierarchy, in what sense of "collection" do they not constitute a collection? Allowing that a collection cannot exist "ontologically prior" to its elements, what prevents us from conceiving of the universe of sets as a collection which is ontologically posterior to all the sets? But then, why would this collection not itself be a set, why would it not already have been included? Understanding lapses.[17]

Menzel takes the view that the universe of sets is indeed a genuine collection and that the iterative conception merely provides a distinction between sets, which appear in the cumulative hierarchy, and another kind of collection, proper classes, which do not — because they are "too big".[18] By itself, this idea is unhelpful against Russell's paradox because it does not tell us why the paradox should not apply to classes. (Sets are "small" collections, proper classes such as the collection of all sets are "big" collections, and the collection of all classes is so big it's not even a collection at all?) One has to postulate that proper classes cannot be elements of other collections, which is in fact the standard line, but this seems problematic if one is not an objects constructivist. For then one would have to affirm that, for instance, the class of all sets is a well-defined object which exists in the same way as any other mathematical object but for which the set whose sole element is that object does not exist. Why wouldn't it?

The iterative conception involves a sleight of hand: one first resolves Russell's paradox by explaining that his paradoxical collection cannot be formed, and one then finesses the problems with a constructivist interpretation by disavowing the language about "formation". Thus "any conviction

[17] A similar point is made in [19].

[18] "More important (for philosophers at least), the iterative conception *explains* the paradoxes ... by providing the grounds for a distinction unavailable to the Fregean between sets on the one hand and collections, known as proper classes, which are 'too big' to be sets on the other [emphasis in original]" ([24], p. 37).

that the iterative conception may carry is made to depend on metaphorical details that are dismissed as inessential to it"[19] — except that they are not merely dismissed as inessential by most iterative theorists, they are, at least implicitly, actually repudiated.

The fundamental difficulty underlying this whole discussion is that we lack a clear idea of what it means to say that sets are abstract objects. If we do not even understand what kind of things sets are supposed to be, it is impossible to know how or whether we can make any sense of "forming" them. I will address this further in Chapter 2.

1.4 The vicious circle principle

Russell's formulation. Impredicativity. Ramsey's "tallest person in the room". Totalities. Truth constructivism. Predicative comprehension.

Russell took a different approach to dealing with his paradox. Following a suggestion of Poincaré, he proposed a "vicious circle principle" which would forbid a certain kind of circularity that appeared to be responsible for the contradiction. This principle is the historical basis of the foundational school now known as *predicativism*. In one formulation it states that

> No totality can contain members defined in terms of itself.[20]

In other places Russell described the principle in similar but not exactly identical ways, leading to some degree of ambiguity. But since the principle is informal to begin with, minor variations in language are perhaps not so critical.

The word "totality" is no longer commonly used in this way, so the first question to address is how we should understand this term. I will say a little more about this in a moment, but to start with let us just assume that it means something like "set" or "collection".

Definitions which violate the vicious circle principle are said to be *impredicative*, and according to predicativism are to be rejected. This is supposed to block Russell's paradox because "let R be the set of all sets which are not elements of themselves" is impredicative, in that it purports to define a set R by means of a condition which picks out its elements from among "all sets", where "all sets" would include R itself. At an intuitive

[19] [19], p. 374.
[20] [37], p. 237.

level, one does sense a possibly illegitimate circularity when a set is defined using a condition that quantifies over all sets.

But why should it be illegitimate? Ramsey famously pointed out that "the tallest person in the room" defines an individual (the tallest person) by means of a condition that quantifies over a collection (all the people in the room) which contains that individual.[21] It therefore violates the vicious circle principle, yet there is certainly nothing wrong with this definition.

One is tempted to say that in Ramsey's example the quantification is unproblematic because all the people in the room are already there. No vicious circle is involved in merely looking through this collection and selecting one of them. This suggests that the vicious circle principle should be understood as only applying to collections which are, in some sense that we still need to clarify, not already there. And the idea is that we should not be allowed to introduce a new individual into such a collection using a definition that depends on properties the collection will manifest at some unknown, or even fictional, future time when it has achieved completion. The vicious circle principle is then just a special case of the more general idea that the initial construction of a mathematical object cannot depend on the properties of any other object which has not yet been constructed, collections which do not yet exist being just one kind of example. I scarcely need to add that it is desperately unclear what any of this could possibly mean, given that we are talking about mathematical objects which are not supposed to exist in space and time. Still, one feels there is something to the idea that the set $\{1, 2, 3, \ldots, 100\}$ is manageably "all there" in some sense in which the universe of sets is not. This was also part of the iterative conception and probably the source of some of its appeal. I will return to this topic in Chapter 2.

In any case, if we take a *totality* to be a collection which is "there all at once", in whatever sense we eventually settle on, then the formulation of the vicious circle principle given above gets it exactly wrong: totalities (such as "all the people in the room") are precisely those collections which *can* be unproblematically quantified over, those collections for which no vicious circle is threatened because their composition is settled and there is no possibility of adding any new elements. It is the collections which are not totalities — in modern language, the proper classes — that we need to be wary about.

Or do we? This is a good place to draw the promised contrast between

[21] [33], p. 41.

objects constructivism (a.k.a. "potentialism") and what I will call "truth constructivism". According to the former, mathematical objects appear in stages, in some sense. Any set will appear *after* its elements have appeared. According to the latter, truth values also appear in stages. (I will give my interpretation of what this really means in Chapter 4, and I will also give an alternative interpretation in terms of the knowability of every fact.) These views can be contrasted with objects platonism, which considers all mathematical objects to simultaneously exist in some abstract sense, and truth platonism, which affirms that all cogent mathematical assertions simultaneously have definite truth values.

The distinction is important here because only objects constructivism is needed to defeat the set-theoretic version of Russell's paradox: if the universe were, somehow, only available in stages, then things like Russell's "set" could never appear, and that would be all we needed to say. The vicious circle principle, however, trades in truth constructivism — when building a set, not only are we not allowed to reach ahead and grab sets that have not yet been constructed, we are not even allowed to look ahead and get information about those sets. The practical implication is that we cannot assume that statements which quantify over a proper class/non-totality have definite truth values. And this is why it might not be possible to identify a set using a condition which quantifies over all sets: because such a condition might not have a definite truth value in every case.

To see how this matters, suppose for the sake of illustration that we accept potential infinity but not actual infinity. Thus, $\mathbb{N}$ can be realized in stages but is not a totality; it is not, and can never be, "all there". It seems consistent with this view to nonetheless regard the number sequence as determinate and to consider sentences which quantify over it to have well-defined truth values. For instance, one could still reasonably feel that "let $x = 1$ if there are infinitely many prime pairs, and 0 if there are not" is a valid definition, because it must definitely either be or not be the case that there are infinitely many prime pairs. This would be an example of truth platonism, and intuitionists would reject it. They would say that unless and until we have a proof or disproof of the twin primes conjecture, the statement of that conjecture cannot be assumed to have a definite truth value. Intuitionists are truth constructivists.

The situation is even more acute for predicativists who regard $\mathbb{N}$ as a totality, but not $\mathcal{P}(\mathbb{N})$. Consider the formula $\phi(n)$ = "algorithm number n computes a well-ordering of $\mathbb{N}$" (i.e., it outputs the ordered pairs which characterize such an ordering relation). Algorithms are embodied in finite

computer programs, so we can lexicographically order them and speak of "algorithm number n", but in order to express the concept of well-ordering we have to quantify over the predicatively proper class of all subsets of $\mathbb{N}$. This means that for predicativists, $\phi(n)$ might not have a definite truth value for all values of n. So $\{n \in \mathbb{N} : \phi(n)\}$ is not a predicatively valid definition of a set of numbers. This shows that we could have a definition of a set of numbers that truth platonists would accept but predicativists would not, their concern being that the set we are trying to construct would itself have been one of the sets of numbers which ϕ quantifies over. What makes this different from the intuitionistic setting is that this set could, conceivably, *only* be definable in an impredicative way, and therefore simply not be available to the predicativist. Intuitionists may reject some definitions of numbers that a truth platonist would accept, but there are no natural numbers which they in principle cannot access. Predicativists, in contrast, are seemingly barred from accessing some sets of numbers which those who consider $\mathcal{P}(\mathbb{N})$ to be a totality believe to exist.

To summarize, objects constructivism sanctions constructions like

$$\{x \in S : \phi(x)\}$$

provided the set S is already available, with no restrictions on the property ϕ, whereas truth constructivism also demands that ϕ not involve reference to any sets which are not yet available. This entails a significant limitation on which comprehension axioms can be considered valid.

I have suggested above that objects constructivism alone suffices to resolve the set-theoretic version of Russell's paradox. But as I will explain later, we need truth constructivism to resolve the semantic version.

1.5 Indefinite extensibility

> *Russell and Dummett's objects constructivism. It cannot explain why some collections are "totalities". Questions raised in this chapter.*

As far as I know, Russell never explicitly endorsed truth constructivism, despite its being, in my view, implicit in the vicious circle principle. However, I think he did effectively endorse objects constructivism when he wrote that the property of being a set is "self-reproductive" in the sense that "we can never collect *all* of the terms having the said property into a whole; because, whenever we hope we have them all, the collection which we have

immediately proceeds to generate a new term also having the said property".[22] Of course, this begs the question of what it means to "collect" all the (abstract) objects which have the property, or to "have them all", or for a collection to "generate" a new term.

Dummett called this sort of idea "indefinite extensibility".[23] His explanations differ slightly, but the basic idea is that a property is supposed to be indefinitely extensible if any "totality" all of whose elements have the property is properly contained in another, larger such totality. I think this, too, is an expression of objects constructivism.

At first sight the notion of indefinite extensibility seems like it might explain the difference between a collection which is "there all at once" and one which is not: to not be there all at once is to be indefinitely extensible. But it becomes clear that this is really not helpful when one realizes that the definition of indefinite extensibility presumes we can already distinguish totalities from non-totalities. For example, imagine trying to decide whether the property of being a natural number is indefinitely extensible. A finitist would say it is: every totality is finite, and therefore any totality all of whose elements are numbers can be extended to a strictly larger totality all of whose elements are numbers (by adding 1 to the largest element in the original totality, for example). But others would say that the set of all natural numbers is itself a totality of numbers which admits no further extension. Who is right? In order to decide whether $\mathbb{N}$ is indefinitely extensible, we must first determine whether it is a totality. This illustrates how the notion of indefinite extensibility does not, on its own, clear up the difference between collections which are totalities and those which are not. We already have to understand this difference before we can understand what it means to be indefinitely extensible.

Some basic questions about sets unanswered in this chapter:

1. What are sets, and how are they formed (or "formed")?
2. In what sense is the mathematical universe "not there all at once"?
3. What is the principled resolution of the set-theoretic version of Russell's paradox?

These questions will be answered at the end of Chapter 2. There were also some basic questions about classes:

[22] [36], p. 36.
[23] See, e.g., [9], p. 195 or [10], p. 441.

4. Is there anything more to identifying a collection than saying which things belong to it, and can't any well-defined property be used to do this?
5. Are all collections sets, and if not what makes some of them fail to be sets? What does "indefinitely extensible" really mean?
6. What makes a collection a "totality"?
7. Why can't proper classes be elements of sets, or other classes?

I will answer these at the end of Chapter 5.

Chapter 2

Abstract objects

2.1 Concrete objects

Problems about concrete objects. Definiteness. Mereological sums.

In the last chapter we encountered such questions as "How are sets 'formed'?" and "What makes a collection a 'totality'?" It is futile to try to answer questions like these without having any idea what sets are. Merely saying that they are not physical objects, that they do not exist in space and time, tells us nothing. We need to get some positive idea of what it means to be an abstract object, in what sense they can be said to exist, and how we, as physical beings, could know anything about them. A central thesis of this chapter is that these are important questions that can actually be answered in an informative way. Moreover, this must be done before we can get to the bottom of Russell's paradox and mathematical foundations generally.

I should start by acknowledging that even concrete physical objects present philosophical problems. Is a cloud a concrete object? Is a deck of cards one object or fifty-two objects? The ship of Theseus: if, over time, all the parts of a ship are replaced, at what point does it become a new ship? Early Wittgenstein championed the picture theory of meaning, whereas late Wittgenstein rejected it. Were there two Wittgensteins, then? More than two?

We can sidestep the first question by agreeing that we are only interested in the behavior of (more or less) sharply defined macroscopic physical objects — "moderate-sized dry goods", in Austin's pithy phrase.[1] Let us proceed on the assumption that we are initially equipped with some reper-

[1] [2], p. 8.

toire of physical objects that can be unambiguously recognized and told apart. I will say that these objects are *definite*, meaning that any two of them either definitely are or definitely are not equal. And yes, at some point later we are going to have to deal with indefinite objects.

From this starting point, we will consider ways of expanding our notion of "object" to include more abstract things. How this may be done is the principal topic of the present chapter.

Let us begin by noting that one basic way to identify new concrete objects is as amalgamations of previously available concrete objects. The philosophical term for such a thing is "mereological sum". For instance, a deck of cards is the mereological sum of the individual cards, and each card is the mereological sum of the atoms which comprise it.

Mereological sums are not sets. My body is the mereological sum of my organs, not the set of my organs. (My body is a physical object, not an abstract set.) A mereological sum with only one component — say, a single rock — is no different from that rock, whereas the set whose sole element is that rock is something quite different. Unlike sets, mereological sums such as "shoals, herds, packs, tribes, and the like, are located and can move around in physical space, just like their members".[2] This is important because we have yet to understand what sets are, so they had better not be sneaking into the discussion early. And it needs to be said because uncareful explanations of what sets are often cite mereological sums like these as examples. (Halmos: "A pack of wolves, a bunch of grapes, or a flock of pigeons are all examples of sets of things."[3] Black: "It ought then to make sense, at least sometimes, to speak of being pursued by a set, or eating a set, or putting a set to flight."[4] All things which can sensibly be said of mereological sums, at least sometimes, but not of sets.)

2.2 Identity criteria

Types and tokens. Changing the identity criterion. Not set-theoretic, mereological, or psychological. No single "true" identity criterion. An axiomatic description.

I plan to argue that sets (in the natural language sense) and classes (in the mathematical sense) are fundamentally linguistic objects. So our

[2] [45], p. 61.
[3] [14], p. 1.
[4] [5], p. 615.

first order of business is to clarify the nature of linguistic entities such as nouns and sentences. The main philosophical issue for us is the so-called "type-token distinction". Suppose the same sentence is written twice on a chalkboard. This presents a problem because the physical arrangements of chalk are clearly distinct, so how can they be the *same* sentence? The conclusion may be drawn that sentences must be abstract objects — "types" — not existing in space and time while their physical embodiments in chalk or ink are merely "sentence-tokens".

This way of thinking seems to me badly misguided. Once (true story!) when I took my daughter to the neighborhood pool there was, to her delight, another young girl there wearing the same swimsuit. The *same* swimsuit ... yet a single physical object cannot simultaneously exist in two distinct locations. So was the swimsuit actually an abstract object not existing in space and time, and were the girls merely wearing "swimsuit-tokens"? Do we really want to say this? Is it, strictly speaking, *wrong* to say that the girls were wearing the same swimsuit? You can't literally wear an abstract type, can you?

Let's back up. Is it really clear that the same object cannot simultaneously exist in distinct locations? If *Gravity's Rainbow* may be found in fine bookstores everywhere, evidently it *can* be in more than one place at once. The issue here is what we mean by "the same", what it means to say that those different copies of *Gravity's Rainbow* are all the same book. Are you the same person you were ten years ago? Is a caterpillar that spins a cocoon the same creature as the butterfly that emerges later? Could you and I have seen the same movie at the same time, in different theaters? In most contexts we would say that early Wittgenstein was the same individual as late Wittgenstein, but if that were unconditionally true then by the logical principle of substitution of equals we should be able to replace "early Wittgenstein" by "late Wittgenstein" in any true sentence in which it appears. To the contrary, there are a lot of things one can truly say about early Wittgenstein's views that are false about late Wittgenstein's views, and vice versa. The need to make this distinction is precisely why philosophers invented and use the terms "early Wittgenstein" and "late Wittgenstein".

When we are talking about the views Wittgenstein espoused, it becomes necessary to distinguish between early and late Wittgenstein. We can do this. It is called *changing the identity criterion*.

I suppose the strictest possible identity criterion would treat any two physical objects at different times as different things, and regard two phys-

ical objects at a given time as being equal only if they were composed of precisely the same material. But this is not a criterion we commonly use. I consider the pen that is now sitting on my desk to be the same object as the pen that was in my pocket ten minutes ago, even though the former is situated in a different spatiotemporal location and contains slightly less ink. (So the two are not made up of exactly the same material, by any standard.) This is normal and convenient. Sometimes I may want to tighten the identity criterion and regard the pen as a different object at different times. Sometimes I may want to loosen it still further, as I do when I complain that the local store no longer sells my favorite pen. "Of course it doesn't, you already own it!" you may reply. But obviously that isn't what I meant. When I say this I am taking this pen here on my desk to be "the same" as any pen of that brand and appearance.[5] It is quite common to use different identity criteria in different contexts. People can wear the same outfits, read the same books, eat the same food, and when something like this is said, we typically understand from context what is meant by "the same". The ship of Theseus problem — at what point is it no longer the same ship? — is at heart a question about what identity criterion to use. Indeed, it is hardly a problem at all once we recognize that we are free to choose whatever identity criterion we like. Or rather, it becomes merely a problem about which identity criterion is most appropriate or useful in that setting.

At some point during my youth the Apple II represented the pinnacle of home computer technology. But why "the" Apple II? Was there only one of them? Well, if we consider all Apple II's to be "the same", then yes, there was only one. The TRS-80 and Commodore 64 were two other popular home computers, back then.

The idea of weakening identity criteria is a version of a well-known explication of abstract objects. But that idea is usually expressed in occult terms, whereby the abstract object is seen as a non-physical thing whose true nature is quite mysterious. Thus there would be an abstract swimsuit-type that does not exist in space and time, and concrete swimsuit-tokens which in some obscure way represent or embody it. I would rather say that the abstract swimsuit exists at multiple places in space and time and the concrete swimsuits are *instances* of it, in much the same way that my pen at the present moment is an instance of my pen as it persists over time. This

[5]Are brands and appearances the real abstract objects here, then? No. Two pens are of the same brand if some company makes both of them, and they have the same appearance if they visually match. Nothing about this needs to be abstract.

is an odd way of speaking, but that is because identity criteria are adopted precisely in order not to have to distinguish between different instances. I could go further and say that in order to uphold the principle of substitution of equals, a given identity criterion can *only* be used sensibly in the context of a language fragment that cannot distinguish between different instances. There is nothing wrong with saying "Wittgenstein is widely regarded as one of the most important philosophers of the twentieth century", but the sentence "according to Wittgenstein, every meaningful proposition pictures a state of affairs" does not have a well-defined truth value. According to early Wittgenstein yes, according to late Wittgenstein no.

The point is that there is no separate non-physical swimsuit; there is only a question of whether we want to regard two swimsuits which visually match as being the same or different. It is quite analogous to the question of whether we should regard a concrete physical object as remaining the same object over time, or as being a different object at each moment. My daughter was not wearing a "swimsuit-token", she was wearing a swimsuit. Whether or not it was the same as the other girl's swimsuit depends on what identity criterion we are using.

The notion of abstraction may have acquired an unsavory taint because of its role in Frege's notorious "Basic Law V", which is widely blamed for causing the original version of Russell's paradox. This law is, in fact, faulty in that it employs an ungrammatical identity criterion, but that error is fixable and does not reveal any problem with the general idea of abstraction. See Section 5.3.

Abstraction is sometimes interpreted in set-theoretic terms, taking "the Fuji apple" to mean "the set of all Fuji apples", I suppose. This is wrong. The Fuji apple is edible; the set of all Fuji apples is not. Sets still have not crept into our analysis. Nor is using a looser identity criterion the same as referring to an equivalence class. The girls were not wearing the same equivalence class of swimsuits; that is not a more precise way of saying that they were wearing the same swimsuit, it is just nonsense. Nor are we talking about mereological sums: the Apple II weighed 11.5 pounds, but the mereological sum of all of its instances weighed around 60 million pounds.

Abstraction is also not psychological. We do not weaken the identity criterion by "concentrating our attention on commonalities" and "forgetting about differences". We weaken the identity criterion by adopting a weaker identity criterion. There is no plainer way of putting it.

So, when I speak of "the" Apple II computer and regard all Apple II

computers as the same, am I talking about a physical object or an abstract one? The answer is both. It is a physical object that exists at a variety of times and places, which should not bother us any more than the fact that ordinary concrete objects can exist at a variety of times and places. The only difference is that the Apple II exists at different places at the same time (which is exactly why it is considered "abstract" — although, now that we understand this, it seems an unhappy choice of words).

I suppose one could argue that there is a single, canonical, *true* identity criterion — something along the lines of "amalgamations of matter situated at two different times constitute the same object if and only if they are connected through spacetime by a continuous, or nearly continuous, timelike series of intermediate amalgamations of matter", maybe — and that any other identity criterion is merely an abuse of language. But this seems like a difficult position to defend. The ship of Theseus becomes a legitimate problem under this view, for instance. According to the criterion just given, the pile left after a rock had been gradually ground down to dust would apparently be the same object as the original rock. If that dust was then scattered to the winds, its mereological sum would still be the same rock. This is not what we would usually say. On the other hand, a jigsaw puzzle that is taken apart and then put together again generally is thought of as the same object. ("Isn't this the same puzzle that was here yesterday?" "No, because I took it apart and put it back together" — really? So it's a different puzzle now?) If I lose a button off my shirt and sew on a new one, or I notice a small hole and mend it, it is still the same shirt. My point is that the identity criterion offered a few sentences back usually works reasonably well, but not always. There is no single identity criterion that is appropriate in all settings. Indeed, it is for just this reason that the type/token distinction is a standard philosophical topic.

The "abuse of language" line is worth examining a little more closely. Maybe you want to say that "this object here on my desk is my favorite pen" is literally true, but "that store no longer sells my favorite pen" is an abuse of language that really means "that store no longer sells pens of the same brand and appearance as my favorite pen". But why must the second sentence be the abusive one? Why not say that "that store no longer sells my favorite pen" is literally true, and "this object here on my desk is my favorite pen" really means "this object here on my desk is one particular specimen of my favorite pen"? The answer is that there is no fundamental reason to prefer one over the other. We can use whatever identity criteria we want. The limitations are questions of convenience and utility, not of

logic.

On my view, abstract objects obtained via abstraction are nothing other than concrete objects with a loosened identity criterion. But, as I have stressed, even concrete objects are themselves not usually subjected to the tightest possible identity criterion of existing at a single moment and being made of identical material. So there are degrees of abstraction, just because there is a range of possible identity criteria, some weaker and some stronger, and concrete physical objects which persist over time are already a little "abstract". But there is nothing non-physical in this picture.

This abstraction process can be described a little more formally as follows. Starting with some initial repertoire of objects, assume we are given an identity criterion on them, symbolized "$\sim$", that is reflexive ($x \sim x$, for every x), symmetric (if $x \sim y$, then $y \sim x$), and transitive (if $x \sim y$ and $y \sim z$, then $x \sim z$).[6] Then introduce new variables α, β, ..., intended to range over the abstracted objects, together with a relation $x \rightsquigarrow \alpha$ ("x instantiates α"), and impose the conditions

- $(\forall\alpha)(\exists x)(x \rightsquigarrow \alpha)$
- $(\forall x)(\exists!\alpha)(x \rightsquigarrow \alpha)$
- $(\forall x, y)(x \sim y \leftrightarrow (\exists\alpha)(x \rightsquigarrow \alpha \wedge y \rightsquigarrow \alpha))$

(every abstract object is instantiated by some concrete object; every concrete object instantiates exactly one abstract object;[7] x and y instantiate the same abstract object if and only if $x \sim y$). The variables α, β, ... can now be taken to range over the newly introduced abstract objects. Any formula ϕ with one free variable, ranging over the original concrete objects, that satisfies $x \sim y \rightarrow (\phi(x) \leftrightarrow \phi(y))$ can be converted into a formula $\tilde{\phi}$ with one free variable, ranging over the new abstract objects, using the condition $x \rightsquigarrow \alpha \rightarrow (\tilde{\phi}(\alpha) \leftrightarrow \phi(x))$.

A meta-comment about what I am doing. Our naive intuition about sets is self-contradictory. So if we take sets as our starting point, we can expect to find ourselves blindly groping our way around the paradoxes, and

[6] If we do not want to abstract from all of the original objects, only some of them, the reflexivity requirement can be dropped. Whenever $x \sim y$, symmetry and transitivity entail that $x \sim x$ and $y \sim y$, so given any symmetric, transitive relation the original objects can be placed in two categories: those which satisfy $x \sim x$ and are to be identified with any y such that $x \sim y$, and those which do not satisfy $x \sim y$ for any y and are to be discarded.

[7] If reflexivity is dropped, as in the last footnote, this second condition should be replaced by the two conditions $(\forall x)(x \sim x \leftrightarrow (\exists\alpha)(x \rightsquigarrow \alpha))$ and $(\forall x)(\forall\alpha, \beta)((x \rightsquigarrow \alpha \wedge x \rightsquigarrow \beta) \rightarrow \alpha = \beta)$.

very likely never coming to any real understanding of them. Instead, the approach I am taking here is to start with concrete physical objects which are clearly free of any serious logical confusion. Then we can progressively consider more general and abstract objects. In the present section we have taken our first step in that direction. A little ways down the road we will pick up sets, and at that point we will be able to understand what they really are and how they really work.

Since I have been talking about identity criteria, I should say something about what they are defined *on*. When I say that two swimsuits are "the same" if they visually match, it is tempting to suppose that I am imposing an equivalence relation on the set of all swimsuits. But of course we do not want to say this, at least not yet.[8] We know what swimsuits are, so let's just say that I am changing the identity criterion for swimsuits and leave it at that. Sets don't have to come into it.

One more comment. In the last section I made a point of insisting that the concrete objects with which we began had to be definite — any two of them must be either definitely equal or definitely unequal. So let us note that this will remain true of any new objects introduced via abstraction, provided that the identity criterion we have used is itself definite, meaning that any two objects to which it applies definitely are or are not equated.

2.3 Possible worlds

Character strings via identity criteria. Modal nominalism. Being "not there all at once". Trading one abstraction for another. Compatibility with various views about infinity.

So what kinds of things are sentences?

Sentences are abstract character strings which meet some grammatical specification determined by the language we are using. This becomes simpler if one is dealing with the formal languages of logic and mathematics, where the syntactic conditions for grammaticality are typically quite straightforward. The question therefore broadens to: what kinds of things

[8] In fact, never: if identity criteria had to be imposed on sets of things, then we could not say what it means for two sets to be the same, since there is no set of all sets. If they had to be imposed on classes of things, then we could not talk about equality of classes for the same reason. What we will ultimately want to say is that we impose identity criteria on concepts, but I will not be able to explain this until Chapter 5, after I have dealt with the circularity issue involving the fact that there actually is a concept of being a concept.

are strings of characters?

We know that character strings are abstract objects because "the same" string can appear in multiple locations at once. The reader will anticipate my position: abstract strings are what we get when we loosen the identity criterion for concrete strings (the printed characters in some book, physical arrangements of chalk, and so on) to say that two concrete strings are "the same" if they contain the same characters in the same order. Of course, this definition hinges on our possession of some criteria for recognizing a string of characters and for deciding when two individual characters are "the same".

The main objection to this account is that if abstract strings reduce to their physically existing instances, this could, depending on the nature of our universe, impose absurd restrictions on abstract strings. It could well be that some particular abstract string might have no concrete instances, and therefore, apparently, not exist according to this account. If the universe were finite, then arbitrarily long concrete strings could not exist. There would be a finite upper bound on the possible length of a string.

We need to think about the arena in which concrete strings are to be found. "Our physical universe" is one simple answer, but it is hardly the only possibility. We could adopt the more stringent condition that concrete strings must lie here on Earth at the present moment, or on Earth sometime in the future, or within humanity's forward light cone, maybe. The most extreme, solopsistic option would be to only accept as real those concrete strings which are immediately available to one's consciousness at the present moment. In the other direction, if our universe is only one of many possible worlds, we could allow concrete strings which appear in any universe.

It seems to me that in mathematics we are not exclusively concerned with the properties of the universe in which we happen to live; that is the province of physics and the other sciences. In mathematics we are concerned with things that are true in any world, which is precisely why the character of mathematical truth is so different from the character of scientific truth. And therefore I feel that the proper definition of abstract strings, and abstract mathematical objects in general, should be based on concrete instances not just in our world, but in any possible world.

This position might be called "modal nominalism".[9] We demand that abstract types must ultimately be reducible to physical objects, but we

[9] A somewhat similar view is defended in [16]. If one accepts Lewis's modal realism [21] (as I suppose I do, but I will not argue the point here), then this modal nominalism just becomes ordinary nominalism.

loosen the notion of "physical object" to include things in any possible world. So we take away with one hand but give back with the other. If we accept the idea of possible worlds containing concrete character strings of arbitrary finite length, then we are free to talk about abstract strings of arbitrary finite length. (Conversely, the idea that there could "exist", in any sense, abstract strings which have no instances in any possible world is surely absurd.)

A natural complaint is that the modal part of this view restores everything taken away by the nominalistic part, rendering the whole proposal vacuous. Perhaps at first it does seem as though a possible world could be conceived to accomodate any desired abstract object. But thinking a little harder, this is not clear at all. What would a possible world look like in which a rock and the set whose sole element is that rock existed as distinct physical objects? More work needs to be done here. And there is one major advantage to working with possible worlds, particularly relevant to the previous chapter: even before we have said what collections really are, we can now see what should be meant when we say that the mathematical universe is "not there all at once" (or "indefinitely extensible"). It should just mean that no possible world contains instances of every mathematical object, that there is always a more complex world with more objects. This is easy to accept; indeed, it is implausible on its face that, to the contrary, there is some sort of maximal possible world which encompasses anything that can be found in any other world.

A related objection is that possible worlds are themselves abstract, so we have merely traded one kind of abstraction for another. Now, for modal realists, other possible worlds are just as real as our own, so this criticism is simply false. But even if one thinks that other possible worlds are somehow less real than ours (personally, I am not even sure what content there is in saying that some other possible world is or is not "real"), the explanation of abstract objects in terms of possible worlds does not lose its clarifying value. If character strings are occult metaphysical objects, one is at a loss to answer basic questions like "How could we know anything about them?" But if they are concrete objects in other possible worlds, we can think about them in the same way that we think about concrete objects in other galaxies in our own universe, say, or for that matter, about any object not currently present in our field of vision. As I just said, there is still more work to do, but the "possible worlds" framing is a good first step.

What, exactly, would count as a possible world is a philosophical question that I will not try to answer. I take it as obvious that our universe

might have been different from the way it is, and I say that these different ways it could have been are possible worlds. Just how different a possible world could be from ours probably has no definitive answer, but I will talk more about this later, in Chapter 7.

How do possible worlds relate to the nature of infinity? Let me introduce terms for three basic standpoints. According to *finitism*, there is no actual infinity, only potential infinity: all sets and constructions are finite. *Countablism*, a less common term, refers to the view that countably infinite sets and constructions are legitimate, but uncountable ones are not. And *Cantorianism* will be my word for the received view that infinite sets exist and every infinite set has a power set.

Modal nominalism does not force us to decide between these three positions. Once we understand what sets are, it will be seen that according to finitism, no possible world contains literally infinite sets; according to countablism, some do but none contain uncountable sets; and according to Cantorianism, there is a possible world that contains an infinite set, and for every set appearing in any possible world there is another possible world in which its power set appears. Strictly speaking, modal nominalism is agnostic about which of these views might be right. However, it could be used to inform arguments for or against any of them, which is something I will get to later.

2.4 Proxies

> *Sets seemingly not explainable as abstract "types". What is the direction of a line? Benacerraf's overspecificity problem. Abstract objects via conservative extensions. Abstract numbers. Going proxy. Solving the epistemological problem.*

Understanding the use of identity criteria allows us to do away with the dichotomization into "types" and "tokens" which fails to capture, or at best awkwardly captures, the fact that there can be multiple layers of increasing abstraction. (Pepsi is an orange tabby, the orange tabby is a house cat, the house cat is a mammal, etc.) The notion of modifying identity criteria does better here. But do identity criteria help us understand things like numbers and sets? It would seem not, because it is not clear what the right tokens would be. Abstract strings are just concrete strings with a weakened identity criterion, but it is harder to say that abstract sets are just concrete sets with a weakened identity criterion. What would a "concrete set" even

be? (That is not a rhetorical question; I will answer it shortly.)

To avoid distractions, let us focus on finite sets for the moment. It does not seem like we understand what, say, the set of moons of Mars is by first identifying some concrete instances of this set, whatever those could be, and then abstracting. The procedure we follow is quite different. First we grasp the general principle that in order to identify a set it is necessary and sufficient to specify what its elements are, and then we explain what the set of moons of Mars is by saying that its elements are the astronomical bodies Phobos and Deimos.

In effect, we understand sets by saying how they behave. This invites the criticism that just saying how something should behave is not enough to show that such a thing exists. And this is true. It is not enough to say how sets are supposed to behave; we must also convince ourselves that there actually are things which do behave in that way.

Frege used the direction of a line to illustrate this point.[10] It is easy enough to say that two lines have the same direction if and only if they are parallel. With a little more precision: every line has a direction, every direction is the direction of some line, and two lines have the same direction if and only if they are parallel. This tells us, operationally, everything we need to know about directions, but it does not tell us what directions *are*.

Frege's solution was, essentially, to say that the direction of a line is the set of all the lines parallel to it. But since this solution is framed in terms of sets, it cannot help us understand what sets are. We might rather say that the direction of a line is that line itself, but subject to the identity criterion that two lines are "the same" if they are parallel. However, neither Frege's idea nor this modification of it is very satisfying, because we do not think of directions as things that can contain elements or be instantiated by lines. Besides, lines are not the only things that have directions. A line segment can have the same direction as a line, so does that mean that the direction of a line segment is a set of lines? Or should the direction of a line really be a set of line segments? It seems arbitrary.

The problem is too much specificity. In an influential essay,[11] Benacerraf pointed out that someone who identified the natural numbers including zero with the finite von Neumann ordinals as $0 = \emptyset$, $1 = \{0\}$, $2 = \{0, 1\}$, etc., would have basic disagreements with someone who identified them with the finite Zermelo ordinals as $0 = \emptyset$, $1 = \{0\}$, $2 = \{1\}$, etc. For instance, the first person would affirm that 0 is an element of 2, while the second would

[10] [11], Paragraph 64.
[11] [3].

deny this. But both would agree that 1 is an element of 2. This seems ridiculous because it is not part of our conception of numbers that some of them are elements of others. I could make similar comments about someone who defined directions as sets of lines and someone who defined them as sets of line segments, and draw the same conclusion. Directions should be abstract objects because they should not have any properties other than those which follow from the way they are correlated with things like lines and line segments, and any identification of directions with something unabstract would give them additional unwanted properties. In the same way, it is not as if the finite von Neumann ordinals fail to have any property the natural numbers have; the problem is that they have a good many other properties besides.

But, assuming we already know how to talk about things like lines and line segments, we can proceed to talk about abstract directions using the following elegant procedure. First introduce some new "direction" variables and a new relation, "is the direction of", whose first argument is supposed to be a direction and whose second argument can be any geometric object. Then adopt axioms which state that every direction is the direction of some line, for every line there is a unique direction which is the direction of that line, and there is a single direction which is the direction of both of two lines if and only if those lines are parallel. Additional axioms can be added relating directions to line segments and other things, as needed.

Any of the definitions of directions proposed earlier would satisfy these axioms. Using a term from mathematical logic, what I have described is a "conservative extension" of our original setting in the sense that it does not allow us to prove any new statements of the original unaugmented language. It enlarges the language without increasing deductive strength. This shows that introducing directions as abstract objects is not a substantive move. Postulating the existence of abstract directions which satisfy the stated axioms tells us nothing really new, because we already knew that there were things which satisfied those axioms. The only reason for doing this is to introduce directions in a way which doesn't give them unwanted extraneous properties.

So directions are abstract objects, but not in any mysterious way. They are only "abstract" because they could be lines modulo parallelism, or line segments modulo parallelism, or maybe something else, and we are not saying which. A contrast can be made with the abstraction process discussed in Section 2.2. When I talk about the transmission problems of the 2018 Honda Odyssey, I am talking about the car currently parked in

my driveway, *and* the car I saw parked on the street several houses down last week, *and* the car that was ahead of me on the freeway a few days ago, etc. When I talk about the direction of a line, I am talking about the set of lines parallel to it, *or* the set of line segments parallel to it, *or* (avoiding sets) the unique line through the origin parallel to it, etc.

The problem of having to choose between different candidates for the natural numbers can be solved in the same way. Starting with a language which does not have any number words, but does have candidates for numbers (such as finite von Neumann ordinals or finite Zermelo ordinals), we introduce some new "number" variables together with, say, constant symbols "0" and "1", and operations $+$ and $\cdot$ which take two number arguments. Then we can add as axioms any statements that hold of the candidates we started with. (If those candidates disagree about which statements of this kind are true, then we have done something wrong.) Again, we have introduced new abstract objects, but without increasing deductive strength.

I will say that finite von Neumann ordinals and finite Zermelo ordinals *go proxy* for abstract natural numbers. These proxies are both set-theoretic, but it is easy enough to come up with non set-theoretic proxies for numbers. For instance, if the reader will forgive this vulgarity, actual numerals — character strings of decimal digits — would do just fine as proxies. Similarly, lines modulo parallelism (or just lines passing through the origin) can go proxy for abstract directions, as can line segments modulo parallelism. When we introduce a new kind of abstract object in this manner, the extension is conservative precisely because we already knew of proxies for them. The possible existence of more than one kind of proxy is what distinguishes this second abstraction process from the first one involving identity criteria.

We solve the epistemological problem for this kind of abstract object by saying that we learn about numbers and directions by studying their proxies. In other words, the proxies do all the work. And that is what makes this kind of abstract object nominalistically acceptable, or at least acceptable to a modal nominalist.

The point of introducing abstract numbers isn't necessarily to let us think about them in a manner divorced from any system of proxies. I'm not even sure we could do that. When I ask you to form a mental representation of the number 5, what actually comes into your mind? I imagine five black dots. On the other hand, my mental representation of the number 547 would probably just be the decimal string "547". I don't really see how one could form a direct mental representation of an abstract platonic number.

Maybe the point of introducing abstract numbers is not to eliminate the proxies but rather to give us the freedom to switch between different systems of proxies, to think about numbers in whatever way is convenient at the moment.

However, there may also be good reasons *not* to abstract in this way. If we want to base mathematics on set theory, for instance, then introducing a separate category of abstract natural numbers may be a complication that carries little real benefit. In practice I think what we usually do is to officially define "natural numbers" to be finite von Neumann ordinals, and then privately think of them as more abstract entities for which the finite von Neumann ordinals merely go proxy. There is little harm in doing this; when challenged, we can always retreat to the position that we have just been talking about von Neumann ordinals all along.

Actually, I think we informally apply the abstraction-by-conservative-extension technique virtually everywhere in mathematics. We think of the elements of a manifold as undifferentiated points, not as sets, which is how they are encoded in ZFC. This informal abstraction is reflected in the way we talk, where isomorphic objects are often conflated. Is the horizontal axis in the complex plane the same thing as the ordinary real line? When stating things rigorously, probably no, but in ordinary discourse, usually yes. In Part 2 of this book I will follow normal mathematical practice and leave abstraction from proxies out of the official development. I still wanted to discuss this kind of abstraction because I think it does explain a lot about how we think and talk about mathematical objects.

Relative to either of our abstraction techniques, saying that an abstract object "exists" is nothing more than saying that some instance or proxy exists. So we can truly assert that the new kind of objects exist, but this does not tell us anything we didn't already know. All of our abstract objects ultimately reduce to physical reality, or at least possible physical reality. Of course they do. What else could there be?

2.5 Sets

> *Lists. Definiteness. Abstract sets. Can a set be an element of itself? Infinite sets and surveyability. Sets coming from other languages.*

Our first explanation of abstract objects, which has them arise from previously available objects by weakening an identity criterion, was used in

Section 2.3 to explain the nature of abstract strings. The second, which has them arise by postulating the existence of objects having certain properties which some previously available objects were already known to have, was used in Section 2.4 to explain the nature of abstract natural numbers. I will now use both kinds of abstraction together to explain the nature of abstract sets.

We need to find actual objects that behave in the way we think sets behave. As a first attempt, we might think about lists of names, i.e., proper nouns — in a formal system this role would be played by constant symbols. A list is a syntactic object; for instance, we could take it to simply be a series of names separated by commas and enclosed in square brackets, say, something like

[the Taj Mahal, the Arc de Triomphe].

But the order of the names should not matter, and if we have two names that refer to the same object, switching one for the other should not matter, and if an item is repeated this also should not matter. So the desired identity criterion is: two lists are "the same" if any object named in either one is also named in the other. Imposing this identity criterion on syntactic lists of names and abstracting produces a new kind of object, a *set*. The *elements* of a set are the objects named in any list that instantiates it.

These "sets" behave just like we want them to. Admittedly, our "sets" have some unexpected properties, such as having instances that are syntactic lists. These could be eliminated by passing to a conservative extension and retaining nothing beyond the membership relation, as discussed in the last section, but as we also discussed there, for us this second step will only be taken informally.

I hinted earlier that we will eventually have to deal with "indefinite" objects for which the statement that some two are equal might not have a definite truth value. Let us head off any possible difficulties coming from this direction by declaring right away that sets can only contain definite objects. That is, only lists consisting of names of definite objects are to be used in the abstraction procedure. This declaration is merely definitional: it simply reflects what we want to mean by the word "set".

While we are on this topic, are sets themselves definite? At the end of Section 2.2 I expressed some possible reservations about the definiteness of objects obtained by modifying an identity criterion. But there is no problem here. Since a syntactic list can instantiate a set only if it is composed of names of definite objects, for two such lists any particular name in the first

list definitely either is or is not synonymous with any particular name in the second list. Quantifying over the two lists, we find that it either is or is not the case that any name in either list is synonymous with some name in the other list. This shows that the identity criterion which generates sets is definite, so sets are definite objects.

All that matters about a set is what things are elements of it. No physical arrangement of the objects themselves behaves like a set; x and $\{x\}$ are not the same thing, nor are a deck of cards and the set of cards in that deck. Things that behave like sets (in the natural language sense) have to be linguistic in nature, or at least symbolic in the sense of involving objects which are taken to stand for other objects.

We get to abstract sets by performing a double act of abstraction, starting from the already abstract notion of a list of names. First we loosen the identity criterion to make two lists "the same" if they name the same objects, and then (optionally, informally) we pass to abstract sets by making a conservative extension of our language. To answer a question posed at the beginning of Section 2.4: we can take "concrete sets" to be lists of names. We don't think of them that way because we informally abstract away from syntactic lists to get abstract objects which have no properties besides those involving only the membership relation.

In any case, in order to answer basic foundational questions about sets, we may have to go back to syntactic lists. For example, could a set ever be an element of itself? This is a legitimate question because it *is* possible for a syntactic list of names to include a name which refers to that very list. We are free to introduce new names at any point, to decide, for instance, to call the syntactic string

[Jeffrey]

by the name "Jeffrey". This string "[Jeffrey]" would then be recognized as a syntactic list containing a single name which refers to that very list. But it is not possible for a list to include a name that refers to the set which that very list instantiates. That is because we can only abstract to sets from syntactic strings that we already recognize as lists of names for known objects. So in order to initially introduce a set that contains itself as an element, we must have already been able to recognize a name for that set, which is impossible. Since this is a slightly subtle point, let me explain it more carefully; there is a circularity here that does not arise with ordinary syntactic lists, because in order to derive sets from lists we must know the correct identity criterion needed to do this. But that means we

must already know the identity criteria for all the objects referred to by those lists, so that we can know when two lists enumerate the same objects. Thus, if a set were to contain itself as an element, the identity criterion by which it is introduced could not be specified unless we already knew it. Abstract objects are not given to us from above. They must be introduced somehow, and there is no possible circumstance in which we would be in a position to introduce a set that was an element of itself, so there cannot be any such set.

What about infinite sets? Their proxies would have to be infinitely long strings. If you do not believe there are possible worlds containing infinitely long strings, then according to the account I am giving, you do not believe in infinite sets (unless you do accept some other equivalent system of proxies, but that seems unlikely). Let us agree that "character string" means a well-ordered sequence of characters in some possible world, i.e., it is a linearly arranged character sequence which has no infinite strictly decreasing sequence of entries. The possible infinite cardinalities of sets are then just the possible infinite cardinalities of character strings in possible worlds.

In previous work I have proposed that what distinguishes sets from proper classes is their "surveyability".[12] In my terminology this means that one can, in some possible world, perform to completion a computation with each element of the set corresponding to one step of the computation. This is a useful criterion in relation to the issue of quantifying over a set; under reasonable assumptions, it means that the truth values of sentences which quantify over a set are, in a broad sense, decidable. Surveyability is not explicitly a part of the account of sets I am giving here, but it should be a feature of it, provided we assume that for any spatial resources available in some possible world (allowing for the existence of a character string) there is another possible world with equivalent temporal resources (allowing for its surveyability). I see no reason to doubt this assumption, and I think it is pretty clearly true of any of the standard foundational positions. Finitists accept finite concrete structures and computations of finite length; countablists accept countable structures and countable computations; and Cantorians accept structures and computations of whatever cardinalities there are in their conception of the mathematical universe.

The idea that sets are ultimately linguistic or representational in nature invites some obvious objections. For instance, one might worry that requir-

[12]See [54], Section 6.3.

ing an explicit list of elements puts too large of a burden on our linguistic resources. Surely we should not have to individually name every fish in the ocean before we can talk about the collection of all the fish in the ocean! This is true, but as we will see in Section 3.1, collections of this sort are the natural language version of classes, not sets. They are "collections" in a different sense. Besides, it is clear that in principle there could be a system of names which gives a distinct name to each fish, and we needn't pretend that we don't know this. Nothing prevents us from thinking and talking about sets that are not explicitly listable in terms of the names we currently have available. I did just that a few paragraphs back when I argued that no set, currently available or not, can be a member of itself. This is an important point, and it also speaks to the objection that whatever sets are, they ought not to be dependent on language. Certainly the names we currently possess determine which sets we are currently able to explicitly list out. But the general set concept applies to any system of names and is not tied to any particular one. We might say that sets are dependent on language but not on any particular language.

2.6 Pure sets

Iterating the set construction. Naming pure sets. Set-trees. Equivalence of the temporal and non-temporal characterizations.

Starting with no notion of any kind of set, we have seen how to introduce the notion of a set of objects, but only objects for which we already had names. These newly introduced sets could not yet themselves have been assigned names. As sets cannot accomodate unnamed objects, all the sets immediately available to us are sets of urelements, i.e., things that are not sets. None of them can be members of any of the others.

But, of course, once sets of urelements are identified as legitimate objects, we are free to add names for them to our language, and to then introduce sets whose entries may include these new names. These would be sets whose elements can be either urelements or sets of urelements. The same procedure can then be iterated. Given a possible world with sufficient temporal resources, sets could be constructed in this fashion through a transfinite sequence of stages. It begins to look rather like the iterative conception from Section 1.3, minus the assumption that it would be possible at every infinite stage to create a single set which enumerates all the

sets whose elements have appeared at previous stages. At any rate, we now have a general method for producing sets of sets. The "pure" sets of primary interest in mathematics would be those which arise from the process just described, starting with no urelements (so that the initial step would produce exactly one set, the empty set).

This discussion can be made more explicit by adopting a uniform prescription for naming pure sets. Let a list with its bookending square brackets replaced by curly braces be a name for the set, if there is one, which is instantiated by that list. Thus "{the Taj Mahal, the Arc de Triomphe}" is a name for the set instantiated by the list "[the Taj Mahal, the Arc de Triomphe]" (and also by the list "[the Arc de Triomphe, the Taj Mahal]"), the set whose elements are those two entities. Now, this assignment of a name can only be done after the set in question has been identified as an object. But once we agree that "[ ]" is a syntactic list with no entries, and therefore instantiates a set with no elements, we can introduce "{ }" as a name for that set. Having done this, we can now also recognize "[{ }]" as a syntactic list consisting of a single name which refers to the empty set, and we can then introduce "{{ }}" as a name for the set that list instantiates, whose sole element is the empty set. Then we can introduce "{{ }, {{ }}}" as a name for the set whose two elements are the previous two sets, and so on.

This process can be continued into transfinite stages, limited only by the resources available in possible worlds. If we require that any such construction must be presented in a well-ordered sequence of stages (and remembering that we also require all character strings to be well-ordered), then we can inductively define the appropriate identity criterion at each stage in order to identify a new set as well as a name for it that can be used in subsequent stages. This gives us a pretty good idea of what names for arbitrary pure sets would look like, even though there is no possible world that could contain actual representatives of all of them, and thus they could never all be available in that sense. There are some technical issues that must be addressed; for instance, we must check that even in transfinitely long strings there is at most one sensible way to pair off all right and left braces, in order to ensure that any character string can name at most one set. This can be shown by induction on the right braces appearing in the string, starting with the right brace in leftmost position. I leave these details for the reader to fill in, or to just ignore, because there is a simpler system of proxies for pure sets that does not need this kind of elaboration.

Namely, we can abandon lists altogether and work instead with (some

sort of physical realization of, or some syntactic encoding of) rooted trees. A *rooted tree* is a directed graph with one distinguished vertex, the *root*, which has no outgoing edges, such that every other vertex has exactly one outgoing edge, and following these outgoing edges takes us from any vertex to the root in a finite number of steps. It is *well-founded* if it contains no infinite directed paths, and *rigid* if it has no nontrivial automorphisms (non-identity bijections with itself that respect the edge structure). Let me now introduce the term *set-tree* for a rooted tree that is well-founded and rigid.[13]

If there is a directed edge from v to w then v is a *child* of w. So each child of the root is itself the root of a rooted subtree consisting of itself and all of its *descendants*, those vertices whose path to the root passes through it. The idea is that these rooted subtrees associated to the children of the root are supposed to correspond to the elements of the set that corresponds to the original tree. Well-foundedness ensures that there is no infinite descending chain of elements $x_1 \ni x_2 \ni \cdots$, and rigidity makes the representation of a set by a tree unique. (Dropping the rigidity requirement would effectively allow elements to be listed more than once.) The appropriate identity criterion is simply isomorphism as directed graphs. Set-trees may then be used as proxies for arbitrary abstract pure sets.

So we can, if we want, regard talk about pure sets as being an abstracted version of talk about set-trees. There is no linguistic aspect to this approach, nor is there any temporal aspect. General sets can be accomodated by allowing some or all of the initial vertices in the tree to be tagged with names for urelements.

Are the temporal and non-temporal set proxies outlined above equivalent? Yes, provided we again assume that the spatial resources of any possible world are matched by the temporal resources of some other world, and vice versa, as discussed in the last section. So either recursively generated set names or rooted, well-founded, rigid trees, both modulo the appropriate identity criteria, can be proxies for pure sets.

We can now see how much is hidden in the facile characterization of sets as mere "collections of objects". This is a reasonable gloss for sets of urelements, but sets of sets are a more complicated matter. In order to understand what "collections of objects" there are, we first need to know what objects there are, so if these collections of objects are themselves supposed to be objects, there is a circularity issue. General sets can only

[13]After we become constructivists, this definition will have to be modified, but that will not be important until Chapter 8. See Section 8.3.

be understood as collections of objects in some recursive fashion.

Sets of urelements and pure sets are totally different kinds of animals, with only the empty set in common. Sets of urelements are straightforwardly identified as lists of names, suitably abstracted. But since the items in any such list have to be individually nameable, pure sets can be accomodated in this picture only by a somewhat complicated recursive definition. It is simpler to model pure sets using trees, again showing their essential dissimilarity from sets of urelements.

I can now answer the first three questions that were posed at the end of Chapter 1.

1. What are sets, and how are they formed (or "formed")? Sets are abstract objects which do not exist independently of their concrete instances or proxies. These instances are physical objects found in possible worlds and thus can be formed in those worlds in the same way that other physical objects can be formed. This is a sense in which sets can be formed.
2. In what sense is the mathematical universe "not there all at once"? I answered this in Section 2.3. It just means that no possible world contains an instance of every mathematical object. (And there is nothing temporal about this.)
3. What is the principled resolution of the set-theoretic version of Russell's paradox? It is that the sets which do not belong to themselves, which is to say all sets, do not constitute a set. There is no possible world in which they could all be listed or surveyed. Or: having completed a sequential survey of all sets is not a possible state of affairs.

Finally, in reference to the "central thesis" stated at the beginning of this chapter, I hope it is now clear that having a genuine understanding of the nature of abstract objects is essential to our being able to answer the above questions. If all we have are bromides like "a set is a collection" and "abstract objects do not exist in space in time", we cannot understand anything.

Chapter 3

Truth

3.1 Classes

Two ways of specifying a collection. A first attempt to define classes. Russell's paradox for classes. Holding.

I explained in Section 2.5 why no set can be an element of itself. This is easier to see in the setting of pure sets, given the recursive characterization where a pure set is constructed in a well-ordered series of stages, with each set at any given stage taking as members only sets from previous stages. (It is also not hard to see if we take set-trees as the model. Hint: well-foundedness.) In particular, any set of pure sets is a pure set which is not an element of itself. This is how I understand Russell's comment about the "self-reproductive" nature of sets quoted in Section 1.5. Any set of sets does not "generate" a "new" set which does not appear among its elements, it simply *is* such a set.

But that does not mean there cannot be a *collection* of all pure sets, in some other sense of the word "collection". We have a pretty clear idea of what constitutes a set, and as I suggested back in Section 1.1, it seems like any well-defined property — such as, one would imagine, the property of being a set — ought to describe a collection. Indeed, this comment itself suggests a way of characterizing general collections, including those which might not be listable and therefore not sets.

Every beginner learns that there are two ways of specifying a set, either explicitly, by listing (names of) its elements, or indirectly, by stating a criterion for membership. It is usually added that infinite sets can only be specified in the second way. I disagree: infinite sets can be specified explicitly by infinite lists, and, in my view, the elements of any set, infinite or not, have to be explicitly listable in some possible world. But why not

invent a broader notion of collection arising from arbitrary predicates — in formal languages, formulas with exactly one free variable — instead of lists of names?

In more detail, we would like to define a new kind of "collection", to be called *classes*, by first considering two predicates to be "the same" if they hold of the same objects and then abstracting in the manner of Section 2.2. This would give us two distinct notions of "collection": sets, which are instantiated by explicit lists of names, and classes, which are instantiated by predicates. The familiar distinction between identifying a collection by explicitly listing its elements and identifying a collection by setting down a criterion for membership (what Russell called a "defining characteristic") would just become the distinction between sets and classes. The set of all prime numbers would be instantiated by the syntactic list "$\{2, 3, 5, \ldots\}$" and the class of all sets would be instantiated by the predicate "is a set".

It isn't that simple, however. We can see that something must have gone wrong because at this point there is apparently no way to avoid Russell's paradox for classes. If we could carry out the procedure I just outlined, we would then be able to talk about classes, and the predicate "is a class that does not belong to itself" would therefore define a class, leading to an immediate contradiction.

Our solution to Russell's paradox for sets is no help here. Yes, there cannot be any Russell set because it is not possible, even in principle, to list out all the sets that aren't elements of themselves (which is to say, all sets): any list of sets is necessarily incomplete, because it would itself instantiate a set that did not appear in the list. But the whole point of introducing classes is that they don't have to be listed out. They just have to be describable, and once we understand what classes are, we will be hard pressed to deny that "is a class that does not belong to itself" is a legitimate description.

We have a non-temporal system of proxies for pure sets that come from set-trees, but we also understand that syntactic lists that instantiate pure sets really can be built up in temporal stages by successive acts, in some possible world or other, of expanding the repertoire of names. So the idea that "there is no stage at which a Russell set could appear" has a pretty clear meaning. But nothing like this is the case for classes. Their elements need not exist prior to themselves, in any sense. To the contrary, if a class is not listable then no possible world could contain (instances of or proxies for) all of its elements; in that sense, there would never be a situation in which all of its elements were available. Classes are supposed to be instantiated by

predicates which can be used to tell us, as new objects become available, whether they belong to the class. So, again, the resolution of Russell's paradox for sets does not help with Russell's paradox for classes.

Something has indeed gone wrong. The problematic ingredient in our attempted definition of classes is the deceptively straightforward notion of a predicate "holding" of an object. No one has understood the subtleties in it, or even recognized that there *are* any subtleties, as far as I can tell. We would like to say that a predicate holds of an object if combining a name for that object with that predicate produces a true sentence. But if that is the definition, then in order to say what it means for a predicate to hold of an object we have to know what it means for a sentence to be true. And truth is problematic. According to Tarski, there is no global notion of classical truth, only an assortment of local truth predicates which apply to various limited settings, which can't help us understand proper classes because proper classes don't appear in limited settings.

We need to understand truth better. Let us think on this.

3.2 The liar paradox

The paradox. Tarski's undefinability theorem. Non-solutions to the liar paradox. The revenge problem. The "quining" construction.

The reason Tarski thought there could be no single, universal notion of truth, only a variety of local truth predicates with limited applicability, was because of the difficulties surrounding the infamous "liar paradox", the paradox that arises when one tries to evaluate the truth of the "liar sentence"

> This sentence is not true.

Tarksi showed that if we only want to talk about the truth of some preexisting group of sentences, we can use his "T-scheme" (see below) to define a truth predicate which applies to them. Crucially, the newly introduced truth predicate could not have been referenced by any of its target sentences, so there is no danger of any sort of self-referential paradox like the liar. Thus, in this sort of localized setting the liar paradox is not a problem, and conversely, Tarski proved a theorem, known as his "undefinability theorem", which in its most general form states that no consistent formal system that meets a minimal standard of expressiveness can contain a truth

predicate that applies to its own sentences. If it did, we would be able to formulate a liar sentence and get a contradiction.[1]

That seems like a pretty ironclad demonstration that there can be no global, self-applicative notion of classical truth. But it isn't. There are some hidden assumptions in the undefinability theorem, and the correct lesson to draw from it is not that there is no global notion of truth, but rather that these assumptions are inappropriate when one is reasoning about global truth. For instance, if we are going to reason with a self-applicative truth predicate, we cannot take it for granted that every syntactically correct sentence within our purview will have a well-defined meaning, and we should neither hope nor expect to be able to affirm the T-scheme for meaningless sentences. It is therefore unreasonable to require, as the undefinability theorem does, that the T-scheme must apply to all sentences. We have to allow for the possibility that some sentences might be meaningless. That is not a full argument, just a hint about where we are headed.

(I am not claiming that the liar sentence is meaningless, by the way. To the contrary, it is not meaningless — I am being deliberately careful with my wording here. See Theorem 6.1 (b) below. The resolution is not that easy, and it is going to take me a couple of chapters to explain it.)

Let us agree that the liar paradox is the fundamental obstacle to our having a global notion of truth. But Russell's paradox should be mentioned in this connection too. It is a familiar observation that the liar sentence has a self-negating quality which also seems to be present when the Russell predicate "is a predicate that does not hold of itself" is applied to itself. Now, unlike Russell's paradox, there is no set-theoretic version of the liar paradox, only a semantic version. But I claim that it is indeed closely related to the semantic version of Russell's paradox. Essentially the same problem occurs in both cases, and the same resolution applies to both.

In order to show this, I will have to explain what that resolution is. The difficulty I face here is that, as Jacques Monod said about evolution, everybody thinks they understand the liar paradox. This circumstance arises because there happen to be a large number of more or less obvious plausible routes to a solution. So anyone who has thought superficially about the liar paradox is bound to have considered that the liar sentence is illegitimately self-referential; or that truth is not even a property of sentences, but rather of the abstract propositions which sentences express; or that the truth of a sentence can change depending on the situation to which it is applied;

[1] [47], p. 273.

or that some sentences are neither true nor false; or that the truth of a sentence can change over time — and imagined that this simple observation (whichever one of them) nullifies the paradox. That is by no means an exhaustive list, by the way.

None of these ideas succeeds, however. It always seems to be possible to modify the liar sentence in a way that defeats whatever solution is offered. If we decide that truth is a property of propositions, not sentences, then

> This sentence does not express a true proposition

revives the paradox. If truth can change depending on the situation, then

> This sentence is not true in any situation

does the trick. If the truth of a sentence can change over time, then we use

> There is a point in time after which this sentence is never true.

And so on. The fact that "strengthened liar sentences" stymie so many plausible approaches to the paradox is known in the liar literature as the "revenge problem".

The idea that the liar sentence is illegitimate because sentences ought not be allowed to refer to themselves can also be belied in such a way. The modified liar sentence used here is more interesting. Given any character string ϕ, consider the operation that replaces every occurence of the symbol "#" in ϕ with the full string ϕ itself, in quotes. Call this operation *quining*. For example, quining the string "chef's special # 9" yields the string "chef's special 'chef's special # 9' 9". Then let Λ_q be the sentence

> The string obtained by quining 'The string obtained by quining # is not a true sentence' is not a true sentence.

The string obtained by quining the internally quoted string is just Λ_q itself, so Λ_q asserts its own untruth just like the standard liar sentence does. But there is no overt self-reference here; it has been replaced by a simple operation on character strings. The sentence Λ_q has the seemingly unobjectionable form "[elementary syntactic construction] is not a true sentence". This shows that forbidding the phrase "this sentence" does not address the real underlying difficulty. As long as we have a basic ability to manipulate character strings, there is no way to prevent sentences from referring to themselves in a roundabout fashion. A similar construction was used by Quine to make the same point.[2]

[2] [30], p. 7.

The quining construction can also be used to refute another proposal, that the liar sentence does not express any proposition and therefore can be neither true nor false, nor can it be deduced from any premise. This is wrong: when we clarify what "propositions" really are near the end of Chapter 5, we will see that the liar sentence does not fail to express a proposition. (Again, my wording is deliberate.) But apart from that slightly subtle point, a simple way to see that this idea has to fail is to apply it to the sentence Λ_q introduced above. If the suggestion is that Λ_q does not express a proposition and hence cannot be true or false, then in particular it follows that Λ_q is not true. So according to the view under consideration, we can affirm the sentence "Λ_q is not a true sentence" (because Λ_q does not even express a proposition). Call this sentence Λ'_q. The idea is supposed to be that we cannot proceed to affirm the sentence Λ_q itself because Λ_q only appears to say the same thing as Λ'_q, but it actually doesn't say anything at all. If Λ_q is senseless then we cannot infer it from any premise.

But "Λ_q" is the name of a particular string, and that same string is also identified by the noun phrase

> The string obtained by quining "The string obtained by quining # is not a true sentence"

— so by substitution of equal terms, given that we have affirmed Λ'_q, i.e., "Λ_q is not a true sentence", we can substitute the preceding noun phrase for "Λ_q" and infer the sentence Λ_q itself. This recovers the paradox. The liar-sentence-does-not-express-a-proposition idea runs afoul of the principle of substitution of equal terms. It can only work if, in some cases, substitution of equal terms converts a sentence which expresses a proposition into one which does not.

Readers familiar with Gödel's incompleteness theorems will notice that the sentence Λ_q is a natural language version of the Gödel sentence which asserts its own unprovability. Gödel's formal version is crafted by first representing character strings numerically, and then using the language of arithmetic to define the quining operation on numerically coded strings. The details are a little complicated, but the idea is simple. The moral for us is that there is no good way to exclude self-reference, in either formal or informal settings. The problem with the liar sentence is not self-reference; it is the notion of truth.

3.3 Sentences and predicates

Sentences, predicates, and relational phrases. Truth and holding. Paradoxes of sentences, predicates, and relational phrases.

In the last section I suggested that there is a close connection between the liar paradox and the predicate version of Russell's paradox. Let us clarify that connection.

We know what a sentence is in natural language. In formal languages, the corresponding notion is a formula with no free variables, which in logicians' jargon is also called a "sentence", conveniently enough. We know what a proper noun is in natural language; in formal languages, the corresponding notion is a constant symbol.[3] And we know what a predicate is in natural language: something that can combine with a proper noun, or indeed any noun phrase, to make a sentence. The corresponding formal language notion is a formula with one free variable.

In formal settings there is an obvious generalization to formulas with more than one free variable, and this can also be considered in the natural language setting. For instance, the natural language analog of a formula with two free variables would be an expression which can combine with two proper nouns to make a sentence, as in "is an ancestor of" (or "x is an ancestor of y", if that is clearer); I will call such a thing a "relational phrase" since it can be thought of as affirming a relationship between two things. We could go on to discuss the natural language analogs of formulas with more than two free variables, but there is little need for specialized terminology for these cases.

The most important question about a sentence is whether it is true. We have an intuitive understanding of what this means, though making that meaning explicit runs into some unexpected subtleties, as I will explain in the next couple of sections. In any case, we have an informal understanding of what it means for a sentence to be true. The corresponding question for a predicate is: for which objects does this predicate hold? And the corresponding question for a relational phrase is: for which pairs of objects does this relational phrase hold?

This "holding" is not merely analogous to truth, it is a variant form of truth. A predicate holds of an object if it combines with any proper noun referring to that object to make a true sentence; a relational phrase holds

[3]Noun phrases are roughly analogous to terms, but the analogy is not exact. Some noun phrases ("the present king of France") notoriously fail to have referents, for instance.

of a pair of objects under an analogous criterion. So if we know which sentences are true, we know which predicates and relational phrases hold of which objects. Conversely, knowing which sentences are true is effectively a special case of knowing which predicates hold of which objects (or of knowing which relational phrases hold of which pairs of objects). This is easier to see in formal languages, where a sentence ϕ is true if and only if the formula with one free variable $\phi \wedge (x = x)$ holds of any x. But it is not particularly hard to make an analogous reduction in the natural language setting.[4]

We have paradoxical sentences (liar) and predicates (Russell). Are there paradoxical relational phrases as well? Yes, and we can even formulate some general patterns. To start, let ϕ be the string

> ϕ is not a true sentence.

This is a simple liar sentence. (Recall that self-referential labels can be avoided with quining, e.g., instead of ϕ use the quine of "the quine of # is not a true sentence". But it is simpler to use labels.) Next, define a *meld* of an object and a string to be that string with the first occurence of the variable "x" in it replaced by some proper noun referring to that object, and let ψ be the string

> No meld of x with ψ is a true sentence.

We can also call this string "$\psi(x)$" to make it explicit that "x" is a free variable. Perhaps this is not a predicate in the grammarian's sense, but it is in the logical sense of being something that can combine with a proper noun to make a sentence. It could be quinified too, of course.[5] This ψ is not quite the same as Russell's predicate: it is paradoxical when applied to *any* object. For instance, "no meld of snow with ψ is a true sentence" is itself a meld of snow with ψ. So if it is true then it is false. And if it is false then it remains false when "snow" is replaced by any other proper noun with the same referent, but that would make it true.

This has already gotten a little complicated, so I will not write out the relational phrase θ that follows the same pattern as ϕ and ψ, but I expect the reader can see how it would go. We would modify the definition of

[4] The admittedly roundabout construction I used in [54] was: the sentence ϕ is true if and only if the predicate formed by concatenating "equals 1 whenever" with ϕ does not hold of 0.

[5] "No meld of x with the quine of 'no meld of x with the quine of # is a true sentence' is a true sentence." This works becaused I defined melding to only operate on the first "x".

melding to make it place two proper nouns in another string, and then let $\theta(x, y)$ be similar to $\psi(x)$, with "x and y" in place of "x". The same pattern would work with expressions that combine with three or more proper nouns to make a sentence; it's paradoxes all the way down.

I gave the melding example first because it exhibits a clear pattern which generalizes to any number of free variables. A different, simpler predicate analog of the liar sentence is

$$\psi = \text{“}x \text{ is not a predicate which holds of } \psi\text{”}.$$

As usual, we could use quining to avoid the explicit self-reference. This predicate clearly holds of any object that is not a predicate, and it even straightforwardly does or does not hold of most predicates, probably. But applying it to itself yields a contradiction. This is still not quite the same as Russell's "does not hold of itself", but it is certainly similar. A relational version might be

$$\theta = \text{“}x \text{ is not a relational phrase which holds of } \theta \text{ and } y\text{”};$$

this yields a contradiction when applied to itself and anything. (Does "x does not hold of θ and y" hold of "x does not hold of θ and y" and snow? If yes, then "x does not hold of θ and y" does not hold of "x does not hold of θ and y" and snow, etc.) Quining could be applied here, too, although this gets even more complicated.

As I said, "does not hold of this very predicate" is a little different from "does not hold of itself". A relational analog of the latter might be "x does not hold of itself and y". But I am not sure the liar sentence works as an exact analog. So I would not insist that the liar paradox and Russell's paradox are exactly analogous, but they clearly fit into a network of related paradoxes, which can even be extended beyond the no-free-variables (liar) and one-free-variable (Russell) settings. Moreover, all these paradoxes can be framed in terms of truth rather than holding, as I intentionally did above in the melding example. The key point is that *truth is the problematic element in all of these paradoxes.*

Therefore, we can expect that any resolution of the liar paradox should be very close to a resolution of the predicate version of Russell's paradox.

3.4 The T-scheme

Tarski's biconditional. Variables not in noun position. Quasi-quotation.

What is truth?

Truth is a property that some sentences of a language may enjoy and others may not. But in order to say which sentences are true, it is not enough for us to have syntactic knowledge of which strings count as sentences. The sentences must also have some sort of meaning. Just what it is for a sentence to have a meaning, and how these meanings can be learned or conveyed, are interesting philosophical questions that I will not discuss further, except rather briefly (and negatively) in Section 4.3.[6] I will take it as given that there are such things as *meaningful sentences*, character strings that are not only syntactically correct sentences but have a well-defined meaning.

In mathematical logic there are two views on truth, a classical one and a constructive one. Constructively, truth is identified with provability — not in the mechanical sense of being derivable in some formal system, but rather in the conceptual sense of being clearly demonstrable. Thus, a sentence is constructively "true" if we are rationally compelled to affirm it. I will come back to constructive truth in Chapter 4.

Classical truth is deceptively simpler. The idea here is that asserting that a sentence is true should be equivalent to asserting that sentence; for example, saying that the sentence "snow is white" is true should amount to the same thing as saying that snow is white. Tarski formalized this condition in his *T-scheme* consisting of biconditionals of the form

"#" is true if and only if #,

where the symbol "#" can be replaced by any sentence in the language. (Thus "snow is white" is true if and only if snow is white, etc.) The analogous scheme for holding would be

For any x, "#" holds of x if and only if x #,

with the symbol "#" to be replaced by any predicate in the language. (Thus, for any x, "is white" holds of x if and only if x is white, etc.) We can call this the *H-scheme*.

This seems perfectly sensible, but there is a grammatical issue which might not be immediately apparent. The subtlety has to do with the fact

[6] If every variable available in the language ranges over a set and every atomic assertion has a definite truth value, then there is a standard explanation, due to Tarski, of "meaning" in terms of interpretation in a set-theoretic model. But this does not apply to languages which can talk about arbitrary sets, for example, let alone languages which can talk about arbitrary classes. Surely such languages can nevertheless carry meaning and some of their sentences can be true.

that the symbol "#" is not in noun position, and so does not act like an ordinary variable. This means that the displayed expressions are not statements *about* arbitrary sentences and predicates; rather, they are *templates* which for each meaningful sentence (or predicate that meaningfully applies to some object) can be used to produce a definition that gives the condition under which it is true (or holds of that object).

Let me explain. The sentence

> For every number n, either n is even or $n+1$ is even

specializes when $n = 17$ to "either 17 is even or $17+1$ is even". What has replaced the variable symbol "n" is a constant term, the decimal string "17". But

> For every predicate ψ, either 17 ψ or $17+1$ ψ

is not even false — it does not make any sense. When ψ is the predicate "is even", it specializes to

> Either 17 "is even" or $17+1$ "is even",

which is not a well-formed sentence. We want to replace the symbol "ψ" with the actual string "is even", not with a name for that string. But "when ψ equals is even" is ungrammatical, because "is even" is not a noun phrase. Equality is a relation and it needs to be combined with two noun phrases to produce a coherent statement.

We specialize a universal statement to some particular case by replacing the quantified variable with a constant term, and what has gone wrong in the above example is that "ψ" is not in noun position and cannot sensibly be replaced by a constant term.[7] We can only make sense of

> Either 17 ψ or $17+1$ ψ

as a template which can be converted into a meaningful statement by replacing the symbol "ψ" with any predicate. It describes a scheme — a family of statements, one for each predicate ψ — but it is not a single statement about all predicates ψ. The same is true of the T-scheme: it is a scheme, not a single statement.

Using the "#" symbol in the preceding way is a little awkward. The same goal can be achieved more elegantly using Quine's quasi-quotation

[7] "Variables eligible for quantification therefore do not belong in predicate positions. They belong in name positions." ([32], p. 67)

notation.[8] The idea is this. We want to insert the string ϕ at various places in some other expression. We show this by writing that other expression out, placing the literal symbol "ϕ" at every location where we want the actual string ϕ to appear, and then enclosing the whole thing in the *quasi-quotation* marks "$\ulcorner$" and "$\urcorner$". For example, instead of saying "replace every occurrence of '#' in

'#' is true if and only if #

with the string ϕ", we simply write

$\ulcorner$"ϕ" is true if and only if $\phi\urcorner$.

The quasi-quotes indicate that we are referring to the string obtained from the one they enclose by replacing each occurrence of "ϕ" with the string that symbol refers to. Again, the literal string " 'ϕ' is true if and only if ϕ" is not a grammatical statement because the second "ϕ" is a term that is not in noun position (and another problem is that the first "ϕ" in quotes does not refer to a sentence, it refers to the single character "ϕ"). Notice that "$\ulcorner\phi\urcorner$" simply refers to the string ϕ, so that "ϕ" and "$\ulcorner\phi\urcorner$" are interchangeable.

To give another example, the expression

$\phi \wedge \psi$

is a string composed of three characters, "ϕ", "$\wedge$", and "ψ". It is not a noun phrase and does not refer to anything, even if "ϕ" and "ψ" do. In fact, it does not mean anything at all, if "$\wedge$" is understood to stand for "and" and "ϕ" and "ψ" are understood as variable symbols ranging over sentences. If we put "snow is white" and "grass is green" in for ϕ and ψ we get

"snow is white" and "grass is green",

which is not a sentence. Variable symbols must go in noun positions, and the "ϕ" and "ψ" in "$\phi \wedge \psi$" are not in noun positions.

In contrast, the expression

"$\phi \wedge \psi$"

(quotes included) is a constant term, like

$2 + (3 \cdot 5)$;

[8] [31], p. 33.

it refers to the string composed of the three characters "ϕ", "$\wedge$", and "ψ", in that order. Whereas the expression

$$\ulcorner \phi \wedge \psi \urcorner$$

(quasi-quotes included) is a variable term, like

$$x + (y \cdot z).$$

For any choice of two strings ϕ and ψ it refers to the string obtained by concatenating ϕ, "$\wedge$", and ψ. (Just as, for any choice of three numbers x, y, and z the expression "$x + (y \cdot z)$" refers to the number obtained by adding x to the product of y and z.) Logicians write things like "$\phi \wedge \psi$", with the quasi-quotes omitted but implicit, all the time — I even did this sort of thing myself earlier, in Section 3.3, for instance (and probably will again somewhere later, though I am now trying not to) — and usually this is not a problem, but we now have to be pedantic and write "$\ulcorner \phi \wedge \psi \urcorner$". For our purposes this subtlety is crucial.[9]

3.5 Deflationary truth

Using truth to affirm a scheme. The Tarskian catastrophe. Deflationism. The conjunction law. Tarski's approach. Infinitary logic. Schematic proofs.

So: "for every predicate ψ, either 17 ψ or $17+1$ ψ" is not a meaningful statement. It is not even grammatical. We have to write

$$\ulcorner \text{Either } 17 \ \psi \text{ or } 17+1 \ \psi \urcorner$$

and understand this to be a scheme, not a single statement.

A scheme is not the sort of thing that can be asserted. Only sentences can be asserted. What we *can* do is to convert this scheme into a single statement by saying

> For every predicate ψ, either ψ holds of 17 or ψ holds of $17+1$,

or, using truth instead of holding, simply

> Every instance of the scheme $\ulcorner$Either 17 ψ or $17+1$ $\psi\urcorner$ is true.

[9] I suppose we could define "$\wedge$" to be a syntactic operator such that for any two strings ϕ and ψ, the expression "$\phi \wedge \psi$" becomes a noun phrase referring to the string $\ulcorner \phi \wedge \psi \urcorner$. But we have to be clear that this "$\wedge$" is not a logical connective (unlike the "$\wedge$" that appears in the strings it creates), and that is definitely an invitation to confusion.

We have used truth to convert a scheme into a single statement, and indeed a similar strategem can be used to convert any scheme into a single statement. This is just the sort of thing truth is good for. But now a problem emerges. In order to use the T-scheme to define what it means in general for a sentence to be true, we would apparently have to convert it into a single assertion about all sentences, by saying that every instance of the T-scheme is true. But this can only be done if we already know what it means for a sentence to be true. It therefore appears that we need to employ the notion of truth in order to make the T-scheme work as a definition of truth. That looks very circular. In previous work I called this problem the "Tarskian catastrophe".[10]

Now, given that a global notion of classical truth leads to contradictions (Russell and the liar), we might consider it a good thing that the T-scheme apparently cannot be used in a non-circular way to define truth. There is little to complain about if we run into difficulty when trying to define a notion which, if we could succeed in defining it, would be paradoxical.

The situation is not that bleak, however. As I have already suggested, the T- and H-schemes must be restricted to meaningful sentences, or meaningful combinations of objects and predicates, and when we do this carefully the paradoxes disappear. This will be explained in Chapter 5.

In fact, the liar paradox is unrelated to the Tarskian catastrophe. The latter is a problem about affirming all instances of the T-scheme at once, while the former is a problem with just a single instance of the T-scheme, when it is applied to the liar sentence.

The theory of truth I am advancing is similar to others that fall under the general heading of "deflationism". I agree that truth is a genuine property of (meaningful) sentences, but it is a pretty meager one. I accept the deflationists' position that there is nothing more to truth than what is given by the T-scheme.

But if we abandon the idea of simultaneously affirming all instances of the T-scheme and instead take it at face value as a schematic definition of truth, then we must deal with the fact that information that is presented schematically can only be used in a piecemeal fashion. Any finite chain of reasoning can only use finitely many instances of the T-scheme. This means that we cannot use the T-scheme to derive simple facts like "if the sentences ϕ and ψ are both true, then so is $\ulcorner \phi \wedge \psi \urcorner$", because any finitely many instances of the T-scheme are consistent with the existence of sentences for

[10] [54], p. 13.

which this *conjunction law* fails, and therefore the T-scheme as a whole is consistent with the existence of such sentences.

Even worse (for us), recall that our interest in truth arose in Section 3.1 when we realized that we need to use it in order to define classes as predicates modulo holding of the same objects. But if holding is introduced schematically by the H-scheme (or, equivalently, reduced to a schematically defined notion of truth by saying "for any x such that $\phi(\hat{x})$ is meaningful, ϕ holds of x if and only if $\phi(\hat{x})$ is true"[11]), then we apparently cannot use it to prove general facts about classes. We could not prove that for any classes X, Y, and Z, if X is contained in Y and Y is contained in Z then X is contained in Z. No finitely many instances of the H-scheme would be adequate to establish this.

Tarski's idea was to say that "$\mathbb{T}$" is a truth predicate for a given system if we can prove, in some metasystem, every instance of the T-scheme involving "$\mathbb{T}$" for the sentences of the target system. This does not address the conjunction law issue, because there is no global affirmation of the T-scheme, only its individual instances, and again, we cannot infer "$(\forall \phi, \psi)(\mathbb{T}[\phi] \wedge \mathbb{T}[\psi] \rightarrow \mathbb{T}[\ulcorner \phi \wedge \psi \urcorner])$" from finitely many instances of the T-scheme. So the conjunction law is not a feature of Tarski's conception of truth. Tarski's approach is also unsatisfying because different metasystems can disagree as to which sentences of the target system are true, so that not only do we lack a global notion of classical truth, we cannot even specify a notion of classical truth for some target system independently of a choice of metasystem.

For instance, consider the number-theoretic assertion "Con(PA)" which standardly expresses that Peano arithmetic (PA) is consistent. It is possible to describe a metasystem of PA which proves every instance of the T-scheme for a predicate which makes "Con(PA)" true, and another metasystem which proves every instance of the T-scheme for a predicate which makes "Con(PA)" false.[12] But "Con(PA)" has an absolute meaning — that Peano arithmetic is consistent — and it should have an absolute truth value relative to its intended interpretation. It may be objected that ZFC does settle the truth of "Con(PA)". The argument might be that PA on its own potentially describes either the standard natural numbers or something else, and we determine which it is by passing to ZFC. In other words, we get the "right" truth predicate for PA by choosing the "right" metasystem.

[11] Here "$\hat{x}$" is assumed to be a canonical name for x.

[12] Most simply, the system whose axioms are the Peano axioms together with every instance of the T-scheme for "$\mathbb{T}$", plus either "Con(PA)" or "$\neg$Con(PA)".

The problem with this idea is that the same difficulty now attaches to ZFC: some metasystems will assess "Con(ZFC)" as true, and others as false. So we then have to go to the "right" meta-metasystem to answer this question. The point is that we apparently have an open-ended ability to choose the "right" metasystems, or at least to recognize that some are "right" and others "wrong", and this is not consistent with the idea that truth is nothing more than a consequence of an arbitrary choice of metasystem.

Another idea that doesn't work for us is to allow ourselves to use infinitely long sentences. Here the suggestion is that we could define classical truth for any given set-sized language using a single, infinitely long sentence obtained by conjoining all instances of the T-scheme for that language. This could be understood as a definition by cases of a single truth predicate. We could then prove the conjunction law by first proving it separately for each pair of sentences and then, in one final transfinite step, inferring the general statement. (Tarski even suggested something like this himself.[13])

But, once again, a notion of truth that applies only to a set-sized language, which is the most we can get using infinitary logic, could only tell us what it means for a set-sized family of predicates to hold of a set-sized family of objects. It still would be of no help at all toward the goal of being able to say what proper classes are.

What about expanding our notion of "proof" to include schematic proofs? If we accept schematic definitions, why not accept schematic proofs too? For instance, it is clear that for any meaningful sentences ϕ and ψ the following is a valid argument:

(1) Suppose ϕ and ψ are both true.
(2) Infer ϕ from (1), using an instance of the T-scheme.
(3) Infer ψ from (1), using an instance of the T-scheme.
(4) Infer $\ulcorner \phi \wedge \psi \urcorner$ from (2) and (3).
(5) Infer that $\ulcorner \phi \wedge \psi \urcorner$ is true from (4), using an instance of the T-scheme.

And if we can recognize it as a valid argument for all ϕ and ψ simultaneously, shouldn't it count as a full proof of the conjunction law? If so, then we could consider the Tarskian catastrophe to be a symptom not of any problem with the T-scheme, but of the inability of ordinary first order logic to accomodate schematic proofs. If we accept schematic proofs then the conjunction law does seem to follow from the T-scheme, without any need to affirm the entire scheme in a single statement.

[13] [47], p. 259.

The difficulty now seems as if it may be merely technical. Is there a good way of dealing formally with schematic proofs? Yes, there is. I know how to do this, and I will give the explanation in Section 5.2. It can be done using the constructive notion of assertibility (which will also figure heavily in the resolution of the liar paradox).

3.6 Truth values

We cannot affirm their global existence. Hidden variables. Mechanical computation of truth values.

Resolving the liar paradox is going to require a constructivist approach, because although we will see that the paradox does not block us from having a global notion of classical truth, it does legitimately show that there cannot be any workable global notion of the "truth value" of a sentence. For suppose there was a global truth-value function τ, defined on all meaningful sentences and taking the values T and F, for which we could affirm every instance of the T-scheme $\ulcorner \phi \leftrightarrow (\tau(\text{“}\phi\text{”}) = T)\urcorner$. Then we could formulate the sentence "τ takes the value F on the quine of 'τ takes the value F on the quine of #'" and quickly derive a contradiction. Given the existence of τ, this sentence would itself be meaningful, and either possible truth value for it would lead to a contradiction. There is only one conclusion to be drawn: if every meaningful sentence had a truth value then we could prove a contradiction.[14]

This reminds me of the problem of hidden variables in quantum mechanics. (Bear with me here.) Given that we cannot simultaneously measure the position and momentum of a particle to arbitrary accuracy, may we nevertheless assume that these quantities have unknowable but definite, exact values? The obvious answer is "Why not?" Even if a quality cannot be measured, what harm could there be in assuming it nonetheless takes an unknown but definite value? But in fact we cannot do this. Bell's inequality shows that the mere assumption that hidden variables exist has statistical consequences which contradict the predictions of standard quantum mechanics and are experimentally falsifiable. We *cannot* assume that these unmeasurable quantities take definite values.[15]

[14] In fact, I will show informally at the end of Section 4.5, and rigorously in Theorem 6.1, that there are particular meaningful sentences for which the assumption that they have a definite truth value entails that we can prove a contradiction.

[15] Yes, I know, assuming local realism as well as mild hypotheses about the nature of measurement. That isn't the point here.

Something like this also happens with truth values. Even granting that we may have no way to determine the truth values of some meaningful sentences, merely assuming that they all *have* truth values is enough to generate a paradox.

Now, in sufficiently simple situations we can know for a fact that every sentence has a well-defined truth value. We may even be able to give an explicit algorithm which mechanically computes it. For example, consider the setting of arithmetic mod 17. It would not be hard to write a computer program which could evaluate the truth value of the sentence

$$(\forall z)(\exists x)(\exists y)(x^2 + y^2 \equiv z)$$

where "x", "y", and "z" are taken to range over the integers mod 17, or indeed the truth value of any sentence of that general sort in such a setting. We just run through all 17^3 possible values for the variables and check. If we allow algorithms to run transfinitely long and use infinite memory, it becomes possible to mechanically compute the truth values of all sentences of elementary number theory. We can describe an algorithm that takes as input any sentence of first-order arithmetic, and after a computation of length less than ω^2, returns a valid output of "true" or "false".

We would probably only regard transfinite computations as legitimate if we believed that they could be performed in some possible world. So there can be disagreement over the extent to which truth values can be mechanically evaluated in this kind of way. In this connection recall the discussion about surveyability in Section 2.5. It is reasonable to assume that for any collection whose elements can be exhaustively listed in some possible world, there is another possible world which can run computations with each entry of the list corresponding to a different computational step. This means that as long as the atomic sentences of our language are decidable and all the variables range over sets, then every sentence will be decidable, in the broad sense of "decidable in some possible world". This legitimizes the idea that every sentence of such a language has a definite truth value, indeed, one that in principle could be found by brute force.

But what about sentences that quantify over proper classes? Even though we cannot visit other possible worlds, we can talk about what goes on in them. We can make assertions which refer to what could happen in an arbitrary possible world. The statement that there is no set of all sets is such an assertion. We know this is true, but not because of any mechanical computation we have done, or even possibly could do. There is no world with sufficient resources to allow a brute force survey of all sets (to check

whether there is one which contains them all). Of necessity, we come to this knowledge through proof, not computation.

So we are going to have to talk about constructive truth, i.e., provability.

Chapter 4

Constructivism

4.1 Assertibility

Constructive truth. Radical formalism. Formal derivability. Axiomatizing the natural numbers.

The liar paradox and the semantic version of Russell's paradox can both be resolved using the black magic of constructive truth.

In constructive mathematics, a sentence is considered to be "true" if it is provable, in the broad conceptual sense of being clearly demonstrable, not the syntactic sense of being formally derivable in some axiomatic system. But I feel that the classical interpretation of truth has historical priority, so I will use the term "assertible" for the constructive notion. Thus, a sentence is *assertible* if we have an absolute right to affirm it, or in the formulation I prefer, we are rationally compelled to affirm it.

(Incidentally, classical truth still has meaning and value in a constructive context. Indeed, we will see in Section 5.1 that the general notion of classical truth *only* makes sense in a constructive context. So there is a real need for a clear terminological distinction between truth and assertibility.)

The first question to address is whether there even is such a thing as "the broad conceptual sense of being clearly demonstrable". Is this really a meaningful concept?

From a naive perspective, it certainly seems meaningful. We know there are infinitely many primes because we have a compelling proof of this fact, not because we have actually looked through all the numbers to see. This seems to show that we do have a notion of rational compulsion. There is no question of deciding for ourselves whether we want to affirm the infinitude of the prime numbers; we must do so.

An opposing position, one that I really dislike, is that mathematics is

merely a formal game with symbols, and mathematical statements such as the claim that there are infinitely many primes have no genuine meaning. If this were so then there obviously could be no question of a mathematical statement being "assertible". We could never be rationally compelled to affirm a meaningless string of symbols, that is clear. But I find it telling that people who claim that math is only about derivability in formal systems generally do acknowledge that it matters whether the systems in question are consistent. Yet the statement that a given formal system is consistent is itself a mathematical assertion about whether a certain Turing machine halts. So I think they do admit, tacitly, that at least some mathematical statements actually mean something.

A more reasonable, superficially attractive idea is that the informal notion of rational compulsion can be identified with the rigorous notion of formal derivability in first order logic. After all, the clear moral of the completeness theorem for predicate calculi is that formal derivability does exactly capture the notion of logical implication in this setting. And this shows pretty indisputably that the statements we are rationally compelled to affirm as explicit consequences of some axioms are precisely those statements which can be formally derived from them. But the question then is, why were these axioms adopted in the first place?

Consider the natural numbers, for example. Can our knowledge of them be encapsulated axiomatically?

One point of view is that our axioms for the natural numbers may be chosen arbitrarily, that we simply define what we mean by "the natural numbers" by setting down axioms for them. I think that people who think this way are over-interpreting Gödel's incompleteness theorems. They are so impressed by the existence of nonstandard models that they begin to doubt the reality of the standard model. To the contrary, I feel that our conceptual understanding of $\mathbb{N}$ precedes any attempt to axiomatize it, and we should not let our inability to capture all arithmetical truths in a single recursively enumerable axiomatization trick us into thinking that we don't really know what the natural numbers are.

Incidentally, the view that the natural numbers are, merely by definition, fully characterized by whatever axioms we choose for them has some uncomfortable implications. For instance, suppose we took the view that $\mathbb{N}$ is simply anything that models some axiomatic system S (Peano arithmetic, perhaps). Could we then affirm that S definitely either is or is not consistent? Consider that unless S proves $\neg\mathrm{Con(S)}$, both $\mathrm{Con(S)}$ and $\neg\mathrm{Con(S)}$ are consistent with S. But $\mathrm{S}+\neg\mathrm{Con(S)}$ proves that there exists a (nonstan-

dard) natural number which is the Gödel code of a proof in S that $0 = 1$, so if we refine our notion of "natural number" by adding $\neg$Con(S) as a new axiom, then on this new, refined conception of natural numbers the proof that $0 = 1$ would have a legitimately finite length. But if we added Con(S) instead, it would not. So it seems that on the view we are considering, the question of whether S is consistent has no definite answer. On some refinements of the notion of a natural number yes, on others no.

No, the situation is more nuanced than that. We have a general ability to go beyond any recursively enumerable axiomatization of the natural numbers by affirming its consistency. We know from Gödel and Rosser that if the given axioms are consistent, they do not prove this fact about themselves, but we can always pass to a stronger formal system that does. I might even say that we are rationally compelled to accept the consistency of a formal system once we have grasped that its axioms correctly describe the natural numbers. If our concept of a natural number ultimately came from a single axiomatic description then there would be no such compulsion.

Having said that, let me invite the reader to question my position by asking how we could "really know what the natural numbers are" if we cannot *say* (through axioms) what they are. This is a serious objection.

In this connection I find it helpful to consider the notions of explicit and implicit knowledge. To a first approximation, our understanding of the natural numbers is captured by the first order Peano axioms (PA), but only to a first approximation. Once we grasp the correctness of these axioms, we are compelled to affirm their consistency. This consistency is *implicit* in our original acceptance of the axioms, but not *explicit* because both Con(PA) and $\neg$Con(PA) are consistent with PA.[1] However, once we recognize their correctness we can, indeed we are compelled to, affirm the consistency of $\mathrm{PA}' = \mathrm{PA} + \mathrm{Con(PA)}$. And then we can recognize the correctness of PA' and go on to affirm $\mathrm{PA}'' = \mathrm{PA} + \mathrm{Con}(\mathrm{PA}')$, and so on. The process can be iterated any finite number of times. Once we recognize that, we can jump ahead and add to PA a single axiom which states that PA, PA', PA'', ... are all consistent, and then go on even further from there. The whole process is open-ended; as long as we stay within the realm of recursively enumerable sets of axioms, there will never come a point when we must stop. It is not possible to leapfrog all the way to "the end" because wherever we land, there will always be an opportunity to go further.

This is a dynamic, temporal process. At any point in time our explicit

[1] I assume throughout this discussion that we have specified arithmetical formalizations of the statements Con(PA), Con(PA'), etc., that have the standard, expected properties.

knowledge of the natural numbers is captured by some particular axiomatic system, but we can also bring out implicit knowledge which enables us to strengthen the current system. In doing this, we make the implicit knowledge explicit. The best metaphor I can think of for this is developing a photographic negative. We have no say as to what objects are going to appear when the negative is developed, but those objects are not there to begin with and are only revealed over time.[2] In our setting the idea is that we do not get to arbitrarily choose whether some new axiom is true or false, but the axioms were not globally predetermined to be true or false ahead of time.

I believe that in some sense we do know exactly what the natural numbers are. But since any explicit characterization of them (via a recursively enumerable axiomatization) is necessarily incomplete, there will always be number-theoretic statements whose truth values are not explicitly determined by the current axioms. Their truth values are revealed over time, as we enrich our axiomatic description of $\mathbb{N}$. Until the axioms are strong enough to decide them, they *do not have* truth values, because we have not yet given a sufficiently precise characterization of the objects they are talking about.

Now, I have framed the preceding discussion from a finitary perspective. As I already mentioned in Section 3.6, if you believe in the cogency of computations of length ω^2 then you should believe that every sentence of first order arithmetic has a definite truth value. With a single computation of length ω^3, all of these truth values could be listed out. We would then be in a position to adopt as "axioms" for the natural numbers all true sentences of first order arithmetic. Here the process, with regard to first order arithmetic, really does terminate; we now have a complete set of axioms, albeit one which is not recursively enumerable.

In making this argument we have transcended the natural numbers and moved into the realm of transfinite set theory. At this point all questions of consistency are settled: the consistency of any countable set of axioms can be verified or falsified with a single infinite computation, so any statement of consistency has a definite truth value. But that does not mean our knowledge of the mathematical universe is complete. Now that transfinite well-ordered sets have come into the picture, we can look at all

[2]Dummett: "We do not *make* the objects but must accept them as we find them ... but they were not already there for our statements to be true or false of before we carried out the investigations which brought them into being [emphasis in original]" ([8], p. 348). (He adds "This is of course intended only as a picture".)

computable total orderings of $\mathbb{N}$ and ask which of them are well-ordered, and this leads into another open-ended process of expanding our repertoire of computable well-orderings of $\mathbb{N}$. Open-ended, that is, unless you believe[3] that we can leapfrog all the way to having a complete list of all computable well-orderings. But once we have done that, assuming we could, we would be in a position to amalgamate all computable well-orderings of $\mathbb{N}$ into a single uncomputable well-ordering of $\mathbb{N}$, recognize it as a well-ordering, and go on from there. Again, there is no final state of complete knowledge. Cantorians face the same phenomenon in relation to large cardinals. The general principle is that the achievement of complete knowledge within some particular domain can only be effected by passing to a larger domain in which one's knowledge remains incomplete.

If this all seems too mystical, that may be because one is used to reasoning classically, and that form of logic is not suited to reasoning about open-ended concepts such as assertibility. For me, the important question is whether the nature of assertibility is sufficiently clear that we can reason about it formally. And as we will see in Section 4.5, the answer is yes.

Returning to the question of truth values: once any set of axioms is accepted, we are rationally compelled to accept as true all logical consequences of those axioms. But any statement whose truth value is not explicitly determined by our current axioms can be consistently added to them, and we have a theorem, the "model existence theorem", which tells us that the resulting set of axioms still has a model. Thus any statement whose truth value is not explicitly determined by the axioms we have affirmed at the present moment is potentially true. At this point, it does not have a truth value. Yet again, one is tempted to try to leapfrog all the way to the end and get a complete list of all true sentences in the language of set theory, say. But once this is done we would be able to formulate a single infinitely long sentence that affirms the truth of all of those sentences simultaneously, and we could then go further to inquire about the truth of other infinitely long sentences. There is still no end state.[4] Thus: *truth values appear in stages*.

(There is an important qualification I have to make to the above: the model existence theorem only applies to consistent axiomatic systems.

[3]From the "constructive countablist" point of view I defend in this book, we do not believe this (Section 8.3).

[4]Dummett again: "the question whether a certain statement is provable cannot be given a mathematically definite formulation because we cannot forsee in advance all possible forms of argument that might be used in mathematics" ([8], p. 347).

There is always a danger, as we continue to strengthen our axiomatic description of $\mathbb{N}$, say, that at some point that description will become inconsistent. I do not believe that should ever happen. But if the only grounds for this belief is a vague idea that we somehow know what the natural numbers are, despite not being able to say exactly what they are, them maybe we should be modest and admit that inconsistency is a possible outcome somewhere along the way. This point will return with a vengeance in the next section.)

4.2 Capture and release

Is every fact knowable? Capture and release. $\perp$-release. The assertible liar paradox. Dialetheism. The release rule.

The appearance of truth values in stages relates to the question of whether every fact can be known. Could there be facts about the world that are, in principle, unknowable? According to truth constructivism, the answer is no: any fact can, in principle, be known. "If a statement is true, it must be in principle possible to know that it is true."[5] Writing "$\mathbb{A}[\phi]$" for "ϕ is assertible", we we can say this formally with the *capture* scheme

$$\ulcorner \phi \to \mathbb{A}[\text{“}\phi\text{”}] \urcorner$$

— that is, $\ulcorner$if ϕ, then "ϕ" is assertible$\urcorner$. Here the variable "ϕ" ranges over all meaningful sentences. The capture scheme expresses the idea that the notion of an unknowable fact is a metaphysical fiction.

In terms of the explicit/implicit knowledge distinction discussed in the last section, there is a simple reason why capture should be valid. Truth values appear in stages. A sentence becomes definitely true when the axioms that have been made explicit are strong enough to entail it. So any fact becomes a fact at the same stage at which it becomes provable. Ergo, we can say in advance that all facts are provable.

The converse can also be argued. The idea that truth values must be knowable strongly indicates that there can be no "end state" at which all truth values that will ever exist do exist. Because at such a state any sentence ϕ would have to be definitely either provable or not provable, but in the latter case, since "there does not exist a proof of ϕ" is true, this sentence must be provable. Thus, for any ϕ, either ϕ is provable or it is provable that ϕ is not provable, which is scarcely credible, if not an

[5] [10], p. 61.

outright contradiction, given that ϕ could quantify over a proper class.[6] So, morally, we can take either "truth values appear in stages" or "all facts are knowable" as our fundamental principle.

In contrast to capture, the reverse *release* scheme

$$\ulcorner \mathbb{A}[\text{“}\phi\text{”}] \rightarrow \phi \urcorner$$

is not a feature of truth constructivism. On its face, this law looks perfectly reasonable; it merely affirms the soundness of logical reasoning. And if every sentence had a truth value, we could ensure release by simply taking it as part of the definition of "validity" that every step of a valid argument must be true. Then the assertibility of ϕ would always entail the truth of ϕ. But this argument hinges on the universal existence of truth values, which we already know is paradoxical, so it is not a good argument. Indeed, if not all meaningful sentences have definite truth values, it is conceivable that we could have an argument that proceeds from true premises to false conclusions, but in which there is no definite step where it passes from truth to falsehood.

We can focus this discussion by looking at the particular case when ϕ is "$0 = 1$", which I will abbreviate as "$\bot$". This instance of the release scheme says that $\mathbb{A}[\text{“}\bot\text{”}] \rightarrow \bot$: if there is a proof of "$0 = 1$" then $0 = 1$. Let's call this statement "$\bot$-release".

We certainly know that "$\bot$" is false. It has this truth value, there is no question about that. But this does not obviously entail that we could never be rationally compelled to affirm "$\bot$". Of course, if we ever did reach a stage at which "$\bot$" was provable, then logical reasoning would be shown to be unreliable. The process would blow up; our axioms would be inconsistent, and they would not have any models. But we can say now, from our current perspective, that even this would not make zero equal one, because we already know it doesn't.

If we knew for sure that we could never be rationally compelled to affirm that $0 = 1$, then $\bot$-release would be vacuously true. And surely we do know this, do we not?

It certainly seems so. But we had better be absolutely certain about this, because if $\bot$-release is to be justified on the grounds that no argument which concludes that $0 = 1$ would, given enough time and intelligence, be recognized as valid, and *$\bot$-release is then accepted as a proof principle*

[6] A more careful way to say this is that we do not have any positive reason to affirm that for every ϕ, either ϕ or "ϕ is not provable" is provable, and therefore we have no positive reason to affirm that there is any state in which all truth values exist.

which can subsequently be used in arguments recognized as valid, then the justification of $\perp$-release hinges on the validity of proofs in which $\perp$-release itself is used. The law thus has a self-affirming quality such that adopting it would be something like, in a formal setting, adopting axioms which affirm their own consistency.

Could $\perp$-release actually be wrong? Is there any prospect of finding a rationally compelling demonstration that 0 equals 1, given that 0 in fact does not equal 1? Do we have any candidate for such a thing?

Yes. We have the liar paradox! If every meaningful sentence has a definite truth value, which is the view we all probably start out with, there is really no way to block the conclusion that $0 = 1$. Maybe the reason it is so hard to resolve the liar paradox is because there is no resolution. Maybe it actually is a valid proof of a contradiction.

To make matters worse, even if not all meaningful sentences have definite truth values, we must still confront the "assertible liar" paradox arising from the sentence "if this sentence is assertible then $0 = 1$". Call this sentence "Λ_a". If we assume that Λ_a is assertible, then release would let us infer Λ_a, which says that if Λ_a is assertible then $0 = 1$, and together with the initial assumption that Λ_a is assertible this yields that $0 = 1$. Thus, we have shown that if Λ_a is assertible then $0 = 1$, i.e., we have proven Λ_a. So Λ_a is assertible after all, and therefore $0 = 1$.

What is wrong with that argument? If there really is a meaningful notion of assertibility, which is at least plausible, and if we can affirm the release law, which is also plausible, then we can give a simple proof that $0 = 1$. Again we are led to contemplate the unthinkable, that there could be valid proofs of contradictions.

In fact, there is a school of thought, dialetheism, which maintains precisely this, that there are valid proofs of contradictions. I do not think it is right; I do not believe that the liar paradox, in either its classical or constructive forms, provides a rationally compelling proof of a contradiction. However, I do not claim to *know* this to be the case. Maybe I just haven't thought about it enough. Let us take the cautious course of granting that the dialetheists might possibly be right. If they were, if "$0 = 1$" had a rationally compelling proof despite the fact that $0 \neq 1$, then the assertibility of "$\perp$" clearly would not imply "$\perp$". So the release law would not be valid. This means that admitting the possibility that the dialetheists are right entails that we cannot affirm the release law. *And without release, the assertible liar sentence is no longer paradoxical*, or at least the derivation of a contradiction given above no longer works. (The classical liar paradox

is still a problem, though; I will resolve it in Section 5.1.)

This is the logical tightrope we have to walk. As long as we allow that the dialetheists could be right, we cannot agree as a general principle that $\ulcorner\mathbb{A}[\phi]\urcorner$ implies ϕ. And this is just what stops us from recognizing the assertible liar paradox as a valid proof that $0 = 1$ ... which would make the dialetheists right.

I have gone into some detail about the reasons for not adopting the release scheme. But it will typically be possible to adopt a deduction rule version of release which infers ϕ from $\ulcorner\mathbb{A}[\phi]\urcorner$. That is because we expect that any compelling proof that ϕ is assertible will in fact constructively provide us with a compelling proof of ϕ. We cannot mandate this, because we must leave open the possibility that there is a proof that $0 = 1$, and if there were then we could infer the assertibility of any sentence without constructing anything. However, if we can see that any proof of $\ulcorner\mathbb{A}[\phi]\urcorner$ within some trusted formal system must be constructive, then we will be justified in adding a release deduction rule to that system.

In a word: we have a capture *scheme*, and in typical settings a release *deduction rule*.

4.3 Theory of meaning

Meaning and proof. Proving $\forall$ statements. Constructive existence. Reasonable expectations. Explaining implication. Redefining negation.

Traditional constructivists typically do accept not only capture, but also release, on the grounds of a theory of meaning according to which the meaning of any sentence is embodied in the description of what would count as a proof of it. Of course, this is a very natural idea once one agrees that every fact is knowable: if a sentence can only be the case if it is provable, then maybe being the case is the same thing as being provable. But I think this is wrong. I have explained above why all facts are knowable, because they become knowable by virtue of the same axioms which make them true. But that does not entail that truth and knowability are the same thing.[7]

According to one common formulation of this idea, when we assert a sentence we are *really* asserting that we can prove it. (Dummett: "an assertion ... is to be construed, not as a claim that it is true, but as a claim

[7]I should add that classical truth-based theories of meaning fail just as badly as intuitionistic proof-based theories of meaning. See Section 4.2 of [54].

that a proof of it exists or can be constructed".[8] Heyting: "Voici donc l'affirmation brouwerienne de p: On sait démontrer p".[9] Prawitz: "An assertion is on this view simply to be understood as a claim to the effect that a proof of the asserted sentence is known".[10]) Under this theory there is literally no difference between the content of ϕ and the content of $\ulcorner\mathbb{A}[\phi]\urcorner$, so we clearly get both capture and release schemes.

Again, I disagree with this. "17 is prime" and "it is provable that 17 is prime" do not express the same thought. One is an assertion about a number, the other an assertion about a sentence. An affirmation of the sentence "17 is prime" is not to be construed as a claim that a proof of "17 is prime" exists or can be constructed — nor is it to be construed as a claim that "17 is prime" is a true sentence. It is to be construed as a claim that 17 is prime. Preserving this distinction is absolutely essential if we are to avoid the assertible liar paradox.

Besides the fact that unrestricted release is paradoxical, there are several other serious problems with the traditional constructivist theory of meaning. The first is the obvious infinite regress: if any sentence ϕ should really be understood as an affirmation that ϕ is assertible, then that affirmation that ϕ is assertible should really be understood as an affirmation of the assertibility of the statement that ϕ is assertible, and so on.

But this objection only applies to the crude formulation of the idea that says $\ulcorner\mathbb{A}[\phi]\urcorner$ is what ϕ *really means*. It does not refute the more careful formulation that the meaning of ϕ is explained by saying what would count as a proof of ϕ. The failure of that formulation only becomes apparent when we try to explain the meaning of a sentence which universally quantifies over a proper class. We are supposed to say what $\ulcorner(\forall x)\phi(x)\urcorner$ means by saying what would count as a proof of it. The standard constructivist gloss is something like: a proof of $\ulcorner(\forall x)\phi(x)\urcorner$ is a construction which, for each value of "x", produces a proof of $\phi(\hat{x})$. But how could you possibly recognize such a thing? You would have to determine that some given construction produces a proof of $\phi(\hat{x})$ *for each value of* "x". How could you determine this, if "x" ranges over a proper class? Certainly not by checking each value of "x" separately. If "x" ranges over a proper class, the only way to know that some construction produces a proof of $\phi(\hat{x})$ for each value of "x" is by proving that it does this, which means that in order to recognize the validity of a proof of $\ulcorner(\forall x)\phi(x)\urcorner$ you would first have to

[8] [10], p. 70
[9] [17], p. 959
[10] [27], p. 46

identify a secondary construction which, for every value of "x", produces a proof that the first construction produces a proof of $\phi(\hat{x})$. Thus we enter another infinite regress.

We cannot explain what "$\forall$" means in this way. "For all" is a primitive logical operator that cannot be reduced to anything simpler, and there is no sense in trying to somehow explain it in terms of proofs and constructions. I would say the same thing about "and", "or", and "there exists": they cannot be reduced to anything simpler and have to be taken at face value.

I gather that interpreting "there exists" to mean "we can construct" is seen by most constructivists as their fundamental principle. I don't think that's right. If there were a proof that $0 = 1$, which we cannot say is impossible, then we would be able to infer any existence statement, indeed any statement at all, without showing how to construct anything. So we either have to admit the possibility, however remote, that there could be valid existence proofs that don't provide a construction, or we have to reject the principle that anything follows from a falsehood. The core principle of constructivism is not about existence, it is about the appearance of truth values in stages. Formally, the core principle of constructivism is the capture law.

(That "however remote" deserves emphasis, though. We may very reasonably expect that any valid existence proof will in some sense show us how to find the desired object, in the same way that we very reasonably expect that we will never discover a valid proof of "$\bot$". And we may expect that any valid proof of a universally quantified statement will show us, in a uniform way, how to prove each instance. Moreover, once we are working with some particular formal system we may be able to recognize that no formal proof available within that system proves "$\bot$", that all existence proofs available within the system are constructive, and that all universal proofs available within the system are uniform. So it might be reasonable to expect that any formal system we accept will have the properties the constructivists want; we just cannot mandate this. This was the same point I was making about the release rule at the end of the last section.)

Coming back to the idea of reducing logical symbols to something simpler, for "$\forall$", "$\exists$", "$\wedge$", and "$\vee$" there is really nothing more to say. We all understand what these symbols mean. Implication, though, may be less clear. The traditional constructivist explanation of implication is that a proof of $\ulcorner \phi \to \psi \urcorner$ is a construction that converts any proof of ϕ into a proof of ψ. But knowing that there is such a construction does not tell us $\ulcorner$if ϕ, then $\psi \urcorner$, it tells us $\ulcorner$if there is a proof that ϕ, then there is a proof that $\psi \urcorner$.

That is, proofs of $\ulcorner \phi \to \psi \urcorner$ are conflated with proofs of $\mathbb{A}[\phi] \to \mathbb{A}[\psi]$. So this idea only makes sense if ϕ is understood to be equivalent to $\mathbb{A}[\phi]$.

But once one understands that truth values appear in stages, the interpretation is straightforward. Any statement may, at any point in the process, be either true, false, or merely potentially true. The statement $\ulcorner \phi \to \psi \urcorner$ affirms that even if ϕ or ψ lacks a truth value at the moment, for any future stage at which ϕ is true there is another future stage at which ψ is true. If ϕ is known to be true, then $\ulcorner \phi \to \psi \urcorner$ is equivalent to ψ, and if ϕ is known to be false then it can never become true and $\ulcorner \phi \to \psi \urcorner$ is automatically true. But there could be cases where we do not know the truth value of ϕ, or even whether it has a definite truth value, but can still see that at any point at which it did become true, ψ would have to as well. (For instance, we can certainly affirm $\ulcorner \phi \to \phi \urcorner$ for any meaningful sentence ϕ.)

Negation is the other logical operator that can be explained in more primitive terms. The classical notion of negation only makes sense in settings where every sentence has a definite truth value and $\ulcorner \neg \phi \urcorner$ can be defined to be the assertion $\ulcorner$"ϕ" is false$\urcorner$. In settings where we do not have a truth value model, we cannot think of $\ulcorner \neg \phi \urcorner$ in this way. Our picture of truth is dynamic, such that any sentence which lacks a truth value at the moment has the potential to later emerge as true. So we can informally take $\ulcorner \neg \phi \urcorner$ to mean that ϕ will never emerge as true. Formally, we define it to be $\ulcorner \phi \to \bot \urcorner$. Since we know that "$\bot$" will never become true, we can see that $\ulcorner \phi \to \bot \urcorner$ is equivalent to $\ulcorner$"ϕ" can never become true$\urcorner$. If it can never become true, then it is not potentially either true or false. It is just false. However, negation now works a little differently than it does classically: ϕ and $\ulcorner \neg \phi \urcorner$ are no longer diametrically opposed. The negation of $\ulcorner \neg \phi \urcorner$ says that it will never become true that ϕ can never become true. That is not quite the same as what ϕ says. A reasonable gloss of $\ulcorner \neg \neg \phi \urcorner$ is that ϕ will always be possible.

How does all this affect logical reasoning? That is what we must discuss next.

4.4 Intuitionistic logic

The law of excluded middle. Natural deduction. Ex falso.

How do we reason about sentences that might not have truth values? Obviously we cannot rely on truth tables. But there is an elegant form of

logic specifically suited to reasoning without this assumption: intuitionistic logic.

The characteristic feature of intuitionistic logic is that we do not adopt the *law of excluded middle* which affirms $\ulcorner \phi \vee \neg\phi \urcorner$ for every sentence ϕ. If every sentence had one of two definite truth values, and the truth values of ϕ and $\ulcorner \neg\phi \urcorner$ had to be different, then one of them would have to be true, and so we could affirm their disjunction. But if ϕ might lack a definite truth value then this reasoning breaks down.

An additional subtlety is that if we ask whether a given statement can be proven in the broadest sense, we are asking whether *there exists* a valid proof of it, and this question itself quantifies over a proper class of proof candidates. (Surely any explicit list could be referenced by some argument, so arguments must be at least as numerous as sets, for example.) So we might not be able to assume that it has a definite yes-or-no answer either. This means that the three truth value model "provable, disprovable, neither" is not appropriate: it is not as if every sentence is definitely provable, definitely disprovable, or definitely neither. We must abandon the whole idea of truth values, or rather, we must understand that it is only valid in limited settings.

How difficult does logical deduction become when we abandon the truth value model? The happy answer is, not difficult at all. The easiest way to learn intuitionistic logic is in terms of *natural deduction*. Detailed accounts are widely available, so I will not give one here. (But see Section 9.2 for a slightly more expanded account.) Very briefly, each logical symbol has an "introduction rule" and an "elimination rule", e.g., if we have proven both ϕ and ψ then we can infer $\ulcorner \phi \wedge \psi \urcorner$ ($\wedge$-introduction) and if we have proven $\ulcorner \phi \wedge \psi \urcorner$ then we can infer either ϕ or ψ, or both ($\wedge$-elimination). If we assume ϕ and prove ψ, we can infer $\ulcorner \phi \to \psi \urcorner$ ($\to$-introduction), and if we have proven ϕ and $\ulcorner \phi \to \psi \urcorner$ we can infer ψ ($\to$-elimination). And so on. The idea is that one reasons "naturally", just as one normally would, modulo our new interpretation of "$\neg$". We no longer regard ϕ and $\ulcorner \neg\phi \urcorner$ as diametrically opposed, so $\ulcorner \phi \vee \neg\phi \urcorner$ is not self-evident. And if assuming ϕ leads to a contradiction, then we can infer $\ulcorner \neg\phi \urcorner$, but if assuming $\ulcorner \neg\phi \urcorner$ leads to a contradiction we can only infer $\ulcorner \neg\neg\phi \urcorner$, not ϕ. In intuitionistic logic we no longer have the law of *double negation elimination* which infers ϕ from $\neg\neg\phi$. At the end of the last section I explained $\ulcorner \neg\neg\phi \urcorner$ as "it will never become true that ϕ can never become true". This statement certainly follows from ϕ itself, but the reverse implication is not obvious.

We do still have the *ex falso* (or *explosion*) law which lets us infer

anything from "$\bot$", though. Reasoning informally, given that $0 = 1$ we can multiply both sides by n and infer that $0 = n$ for arbitrary n. This shows that all numbers are equal, and we can then infer all atomic number-theoretic sentences, and from this all sentences of elementary number theory. Moving beyond number theory, any two finite strings can be encoded as numbers, so the equality of all numbers implies the equality of all finite strings, so if any finite string is a true sentence then every finite string is a true sentence. So we can again infer anything.

It is instructive to verify that any formula ϕ does intuitionistically imply $\ulcorner\neg\neg\phi\urcorner$. Assume ϕ; this assumption will remain in effect throughout the proof. Our task is to prove $\ulcorner\neg\neg\phi\urcorner$. To this end, temporarily assume $\ulcorner\neg\phi\urcorner$. Recalling that $\ulcorner\neg\phi\urcorner$ abbreviates $\ulcorner\phi \to \bot\urcorner$, and that we are working under a global assumption of ϕ, infer "$\bot$". Assuming $\ulcorner\neg\phi\urcorner$ has led to a contradiction, so by the introduction rule for "$\to$" we can stop assuming $\ulcorner\neg\phi\urcorner$ and infer $\ulcorner\neg\phi \to \bot\urcorner$, i.e., $\ulcorner\neg\neg\phi\urcorner$. Thus, from the premise ϕ we have proven $\ulcorner\neg\neg\phi\urcorner$. Another good exercise is to show that $\ulcorner\neg\neg\neg\phi\urcorner$ intuitionistically implies $\ulcorner\neg\phi\urcorner$. I leave that one to the reader.

It is also important to recognize that since negation works differently in intuitionistic logic, we cannot reduce the logical symbols to each other in the usual way. E.g., (1) $\ulcorner\neg\phi\vee\neg\psi\urcorner$ is not equivalent to (2) $\ulcorner\neg(\phi\wedge\psi)\urcorner$. Given (1), assuming $\ulcorner\phi \wedge \psi\urcorner$ leads to a contradiction, showing that (1) implies (2). (This is also a good exercise.) But the reverse implication is not a feature of intuitionistic logic. Rather surprisingly, however, $\ulcorner\neg\phi \wedge \neg\psi\urcorner$ is intuitionistically equivalent to $\ulcorner\neg(\phi \vee \psi)\urcorner$.

4.5 Axiomatizing assertibility

Hilbert axioms for intuitionistic logic. Logical axioms. The **MP**, $\forall$**-Int**, *and* $\exists$**-Elim** *axioms. Capture and release. The assertible liar paradox.*

We can use intuitionistic logic to reason formally about assertibility. I will assume here that we are working with a formal language in which we are able to talk about and manipulate character strings; alternatively, one could work in an arithmetical setting using Gödel numbering, for example.

Recall that we have a predicate symbol "$\mathbb{A}$", read as "is assertible", which takes a character string as its argument. It is convenient to allow ourselves to talk about the assertibility of formulas which contain free variables, not just sentences; asserting such a formula should always be understood as asserting its universal closure, i.e., a sentence obtained from it

by universally quantifying all of its free variables.

I will assume in this section that all of the sentences we are working with are meaningful. For languages some of whose sentences might not be meaningful, some of the axioms I present here must be modified, and I will do that in Section 5.2.

The simplest way to explain the formal treatment of assertibility is in terms of Hilbert-style deduction where one reasons linearly from premises to conclusions, rather than natural deduction where it is possible to introduce temporary assumptions. In Hilbert-style deduction the logical axioms for intuitionistic logic are all formulas of one of the following forms, where ϕ, ψ, and θ are any meaningful formulas, t is a term, and we allow replacement of other variables for "x":

(1) $\ulcorner \phi \wedge \psi \to \phi \urcorner$
(2) $\ulcorner \phi \wedge \psi \to \psi \urcorner$
(3) $\ulcorner \phi \to (\psi \to \phi \wedge \psi) \urcorner$
(4) $\ulcorner \phi \to \phi \vee \psi \urcorner$
(5) $\ulcorner \psi \to \phi \vee \psi \urcorner$
(6) $\ulcorner (\phi \to \theta) \to ((\psi \to \theta) \to (\phi \vee \psi \to \theta)) \urcorner$
(7) $\ulcorner \phi \to (\psi \to \phi) \urcorner$
(8) $\ulcorner (\phi \to (\psi \to \theta)) \to ((\phi \to \psi) \to (\phi \to \theta)) \urcorner$
(9) $\ulcorner \bot \to \phi \urcorner$
(10) $\ulcorner (\forall x)\phi(x) \to \phi(t) \urcorner$
(11) $\ulcorner \phi(t) \to (\exists x)\phi(x) \urcorner$

with the qualification that in (10) and (11) the substitution of the term t for "x" (or whatever variable replaces it) must be free, i.e., no variables in t become quantified. The deduction rules are modus ponens (given ϕ and $\ulcorner \phi \to \psi \urcorner$, infer ψ), $\forall$-introduction (given $\ulcorner \phi \to \psi(x) \urcorner$, infer $\ulcorner \phi \to (\forall y)\psi(y) \urcorner$), and $\exists$-elimination (given $\ulcorner \psi(x) \to \phi \urcorner$, infer $\ulcorner (\exists y)\psi(y) \to \phi \urcorner$). In the last two rules we require that the variable used for "x" must not occur freely in ϕ and the variable used for "y" either equals the variable used for "x" or else does not occur freely in ψ.

Let "Log(ϕ)" be the statement that the formula ϕ is is an instance of one of the above Hilbert axioms for intuitionistic logic. Then the "logical" axiom for assertibility is

$$\text{Log}(\phi) \to \mathbb{A}[\phi], \qquad \textbf{(Logical)}$$

and deduction under "$\mathbb{A}$" is made possible by the next three axioms, which correspond to the three deduction rules in Hilbert-style intuitionistic rea-

soning. We have:

$$\mathbb{A}[\phi] \wedge \mathbb{A}[\ulcorner \phi \to \psi \urcorner] \to \mathbb{A}[\psi] \qquad \textbf{(MP)}$$

$$\mathbb{A}[\ulcorner \phi \to \psi(x) \urcorner] \to \mathbb{A}[\ulcorner \phi \to (\forall x)\psi(x) \urcorner] \qquad (\forall\textbf{-Int})$$

$$\mathbb{A}[\ulcorner \psi(x) \to \phi \urcorner] \to \mathbb{A}[\ulcorner (\exists x)\psi(x) \to \phi \urcorner]. \qquad (\exists\textbf{-Elim})$$

(and similarly with any other variable in place of "x", and with the same restrictions on the variables in (∀-Int) and (∃-Elim) that apply to the corresponding deduction rules).

The justification for these three axioms is simple: we get **MP** because any proof of ϕ can be combined with any proof of $\ulcorner \phi \to \psi \urcorner$ to yield a proof of ψ; we get **∀-Int** because any proof of $\ulcorner \phi \to \psi(x) \urcorner$ can be converted into a proof of $\ulcorner \phi \to (\forall x)\psi(x) \urcorner)$ by adding a single application of ∀-introduction; and we get **∃-Elim** for a similar reason. I would justify the **Logical** axiom by saying that every instance of each of the axiom schemes (1) — (11) is provable in natural deduction.

The real power of the preceding axioms comes from the fact that they are axioms, not schemes; each is a single statement about all meaningful formulas ϕ and ψ.

We have the following easy fact:[11]

Theorem 4.1. *If some sentence ψ can be derived in intuitionistic logic from some other sentences $\phi_1, \ldots, \phi_n$, then the above axioms yield $\ulcorner \mathbb{A}[\phi_1] \wedge \cdots \wedge \mathbb{A}[\phi_n] \to \mathbb{A}[\psi] \urcorner$.*

This is shown by simply reasoning under "$\mathbb{A}$" via normal Hilbert-style deduction.

As discussed in Section 4.2, we also have a capture scheme and, typically, a release rule. Previously I described these in the simple forms $\ulcorner \phi \to \mathbb{A}[\text{“}\phi\text{”}] \urcorner$ and "infer ϕ from $\ulcorner \mathbb{A}[\text{‘}\phi\text{’}] \urcorner$", where ϕ can be any meaningful sentence. When ϕ is a predicate we have the analogs

$$\ulcorner (\forall x)(\phi(x) \to \mathbb{A}[\text{“}\phi(\hat{x})\text{”}]) \urcorner$$

and

infer $\ulcorner (\exists x)\phi(x) \urcorner$ from $(\exists x)\mathbb{A}[\phi(\hat{x})]$,

[11] [54], Theorem 4.2.

and these can be generalized to formulas with any number of free variables. The capture scheme is valid for predicates because for any value of x, if $\phi(x)$ is true then is must be provable. (Every fact can be known.) We agreed that the release rule for sentences is valid provided we are reasoning within a system which we know proves any statement of the form $\mathbb{A}[\phi]$ constructively, in a way which actually exhibits a proof of ϕ; the predicate version requires the further premise that we know the system also proves existence statements constructively, i.e., any proof of $(\exists x)\mathbb{A}[\phi(\hat{x})]$ would actually exhibit a value of x and a proof of $\phi(\hat{x})$. So, again, the release rule cannot be globally affirmed but will be available in typical settings.

If the variable x ranges over a set then we can affirm the **∀-Capture** scheme

$$(\forall x)\mathbb{A}[\phi(\hat{x})] \rightarrow \mathbb{A}[(\forall x)\phi(x)].$$

The idea is that if there is a proof of $\phi(\hat{x})$ for each x then these proofs can be amalgamated into a single proof of $(\forall x)\phi(x)$. Together with the release rule for sentences, this implies a release rule in the form "infer $\ulcorner(\forall x)\phi(x)\urcorner$ from $\ulcorner(\forall x)\mathbb{A}[\text{`}\phi(\hat{x})\text{'}]\urcorner$". But only if x ranges over a set.

So these are the basic axioms of assertibility:

- the **Logical** axiom "$\text{Log}(\phi) \rightarrow \mathbb{A}[\phi]$"
- the **MP**, **∀-Int**, and **∃-Elim** axioms
- the generalized **Capture** scheme and the **∀-Capture** scheme.

In addition, we cannot globally mandate but will typically be able to accept

- the generalized **Release** rule.

We can now see in more detail how the setup described above defeats the assertible liar paradox from Section 4.2. Assume the assertible liar sentence Λ_a, which can be compactly formulated using a constructive "not" as "this sentence is not assertible", so that Λ_a is equivalent to "$\neg\mathbb{A}[\Lambda_a]$". Then, since we have assumed Λ_a, we get "$\neg\mathbb{A}[\Lambda_a]$". But also, applying capture to Λ_a yields "$\mathbb{A}[\Lambda_a]$". So assuming Λ_a leads to a contradiction, and we conclude its negation, i.e., that Λ_a is not not assertible. But now we are stuck: in intuitionistic logic we cannot go on to infer Λ_a, and no further progress toward a contradiction is possible. Or we could start by assuming that Λ_a is assertible. This means that "$\neg\mathbb{A}[\Lambda_a]$" is assertible, and we can also apply capture to "$\mathbb{A}[\Lambda_a]$", yielding that both "$\mathbb{A}[\Lambda_a]$" and its negation are assertible. We conclude that "$\perp$" is assertible. Thus, assertibility of Λ_a entails assertibility of "$\perp$".

That is what we get: Λ_a is not not assertible, and if Λ_a is assertible then so is "$\bot$".[12] We have learned something interesting and informative about the assertible liar sentence, without running into any paradoxes. Indeed, it can be shown that no contradiction can be derived from the axioms described above even if the assertibility predicate is allowed to be self-applicative, as in this case; see Theorem 6.2.[13] (This is not a claim that no rationally compelling proof of a contradiction is possible, only that there is no derivation of a contradiction which uses only the axioms stated above.)

If, for some sentence ϕ, we have both $\ulcorner \phi \to \mathbb{A}[\text{“}\bot\text{”}]\urcorner$ and $\ulcorner \neg\neg\phi\urcorner$, I will say that ϕ is *anomalous*. So we have just seen that $\mathbb{A}[\Lambda_a]$ is anomalous. Anomalous sentences come very close to being paradoxical — they are definitely not false and almost certainly not true. But they are not literally paradoxical. We will see in the next chapter that the classical liar paradox and Russell's paradox for predicates also involve anomalous sentences.

It is essential to the assertible liar paradox that we be able to invoke the "$\mathbb{A}[\text{`}\bot\text{'}] \to \bot$" instance of the release scheme; otherwise we can only derive that the assertible liar sentence is not not assertible, which does not lead to a contradiction. And as long as we allow the possibility that there could actually be a valid derivation of "$\bot$", we are not entitled to assume release as an axiom scheme. The irony here is that determining that there is no valid derivation of "$\bot$" would let us affirm the unwanted instance of release, and there would then be no way to avoid the paradox. That is, establishing that "$\bot$" is not assertible would immediately lead to a demonstration that "$\bot$" is assertible.

The use of intuitionistic logic is also essential here. If we try to reason about assertibility classically, then we can affirm the release law and get a contradiction. Even if we affirm capture but not release — that is, use exactly the same nonlogical axioms as above — we can still say that "$\mathbb{A}[\Lambda_a] \vee \neg\mathbb{A}[\Lambda_a]$" implies "$\mathbb{A}[\bot] \vee \bot$", which implies "$\mathbb{A}[\bot]$". This is a more rigorous version of the argument I made in Section 3.6, that assuming every meaningful sentence has a definite truth value leads to a proof of "$\bot$". We now see that assuming the specific sentence Λ_a has a definite truth value is enough to get the provability of "$\bot$".

[12] Note that Λ_a *is* a meaningful sentence. Because "$\mathbb{A}[\Lambda_a]$" is meaningful regardless of whether Λ_a is meaningful, and therefore "$\neg\mathbb{A}[\Lambda_a]$", which is to say Λ_a itself, is meaningful.

[13] This remains the case in various stronger systems, e.g., Peano arithmetic augmented by an assertibility predicate [54].

Chapter 5

Essential Notions

5.1 Meaningfulness

Three conditions on meaningfulness. Two paradoxical arguments. Resolution of the liar paradox. Resolution of Russell's paradox for predicates.

In this chapter I will analyze and explain classical truth, classical and constructive holding, meaningfulness, propositions, concepts, and classes. Let us begin with classical truth.

The liar paradox is the key obstruction to a satisfying theory of classical truth. In order to explain how to resolve it, I first need to say a few things about meaningfulness. This is relevant because it is not a priori clear whether the liar sentence is a meaningful assertion.

I will make three dogmatic demands. The first is that any complex sentence can be considered meaningful if and only if its constituent parts are meaningful. Thus $\ulcorner \phi \wedge \psi \urcorner$ is meaningful if and only if both ϕ and ψ are meaningful, and so on. We may say that meaningfulness is *compositional*. If any part of a sentence is not meaningful, then the sentence is not meaningful. I will discuss quantifiers in the next section.

Sometimes it is possible to evaluate the truth of a sentence without bothering to evaluate the truth of some of its constituent parts; e.g., $\ulcorner \phi \vee (0 = 0) \urcorner$ is true regardless of whether ϕ is true. One might consider regarding such a sentence as meaningful even if the irrelevant part is not. But I insist that even sentences of this kind are not meaningful unless all of their parts are meaningful.

My second dogmatic demand is that valid arguments must be composed only of sentences that are known to be meaningful. If some step is not known to be meaningful, it is not a valid argument, or at least we cannot affirm it as such. And my third dogmatic demand is that ascriptions of

meaningfulness are always themselves meaningful, regardless of whether the object sentence is meaningful. I would say this about assertibility, too: it always means something to say that some sentence is assertible, whether that sentence is meaningful or not.

These may seem rather mild, as dogmatic demands go. But before you accept them, consider what they mean for implication. I am insisting that $\ulcorner \phi \to \psi \urcorner$ is meaningful if and only if ϕ and ψ are both meaningful. If we don't know whether ψ is meaningful, for instance, we cannot use this implication as a step in a valid argument. Even if we know that (1) ϕ is meaningful and (2) ϕ entails that ψ is meaningful, we still cannot accept the implication as meaningful. In this situation, as with the sentence $\ulcorner \phi \vee (0 = 0) \urcorner$, one feels some impulse to accept $\ulcorner \phi \to \psi \urcorner$ as meaningful, because its truth value only depends on ψ if ϕ is true, and we know that ψ is meaningful in that case. But no, we cannot affirm the meaningfulness of $\ulcorner \phi \to \psi \urcorner$ unless we can separately affirm the meaningfulness of both ϕ and ψ.

One has to be especially careful about this when reasoning using natural deduction. In the situation I just described, when we know that ϕ is meaningful and that ϕ entails ψ is meaningful, it is tempting to introduce ϕ as a temporary premise and then reason about ψ under this premise. But arguing under the temporary premise ϕ is really making a series of statements of the form "ϕ implies [something]". So the prohibition against reasoning about $\ulcorner \phi \to \psi \urcorner$ if we do not already know that ψ is meaningful translates, in the natural deduction setting, to the demand that even under temporary premises we can only make statements that we know to be meaningful without qualification, i.e., outside of the temporary premise. In particular, we cannot say "suppose ψ is meaningful" and then, under this assumption, make an argument some of whose steps are only meaningful if ψ is.

Now I can resolve the classical liar paradox. First, a partial solution.

Let Λ be the classical liar sentence, $\Lambda =$ "$\neg\mathbb{T}[\Lambda]$" ("Λ is not true"), and consider the following paradoxical argument. Assume Λ. Then Λ is true according to the T-scheme, but we can also infer $\neg\mathbb{T}[\Lambda]$ from the definition of Λ. This is a contradiction, so our initial assumption of Λ must be false. That is, we have proven $\neg\Lambda$, from which follows, by the T-scheme again, $\neg\mathbb{T}[\Lambda]$. So we have finally proven both Λ and $\neg\Lambda$, a contradiction. This is the *first paradoxical argument*.

The partial solution is simply to observe that unless we know in advance that Λ is meaningful, no part of this argument can be made. The very first

step is "assume Λ", but if Λ were meaningless then we could not assume it. Thus the derivation of a contradiction was not valid.

We might try to get around that obstacle by assuming "$\mathbb{T}[\Lambda]$" instead of Λ in the first step. But the T-scheme introduced in Section 3.4 only defines $\mathbb{T}[\phi]$ when ϕ is a meaningful sentence. So if Λ were not meaningful, then we would not have assigned a meaning to "$\mathbb{T}[\Lambda]$", and the proof again grinds to a halt.

Here one may object that the T-scheme characterization of truth is incomplete. It only says what $\mathbb{T}[\phi]$ means when ϕ is a meaningful sentence, but we should add that if ϕ is not meaningful then it cannot be true. The complete definition is therefore, schematically,

$$\ulcorner \mathbb{T}[\text{“}\phi\text{”}] \leftrightarrow \phi \urcorner$$

when ϕ is meaningful and

$$\mathbb{T}[\phi] \leftrightarrow (0 = 1)$$

when ϕ is not meaningful.

This recovers the paradox because if Λ were not meaningful then it could not be true, i.e., we could positively affirm "$\neg\mathbb{T}[\Lambda]$", which is to say, Λ itself, leading directly back to a contradiction. This is the *second paradoxical argument*.

There is nothing wrong with this new argument, but observe that it is totally different from the first one, the one made under the assumption that Λ was meaningful. So we really need both arguments. If Λ is meaningful then we can make the first paradoxical argument and derive a contradiction, and if Λ is not meaningful then we can make the second paradoxical argument and derive a contradiction.

And here is the answer. Meaningfulness is an open-ended concept that must be reasoned about constructively. We cannot assume that we have excluded middle here, that Λ is definitely meaningful or definitely not meaningful. If Λ is meaningful then we know what "$\mathbb{T}[\Lambda]$" means; it is equivalent to Λ. And if Λ is not meaningful then we also know what "$\mathbb{T}[\Lambda]$" means; it is equivalent to "$0 = 1$". But if we do not know whether Λ is meaningful or not, we cannot say which prong of the definition to use, so we cannot affirm that "$\mathbb{T}[\Lambda]$" has a meaning, and therefore we cannot affirm that "$\neg\mathbb{T}[\Lambda]$", i.e., Λ itself, is meaningful. Nor do we have any immediate justification for affirming that Λ is definitely not meaningful; indeed, if it were not meaningful then $\mathbb{T}[\Lambda]$ would be meaningful, which would entail that Λ is meaningful.

Note that we cannot enhance the definition of "$\mathbb{T}[\Lambda]$" by trying to say what it means when Λ is neither meaningful nor meaningless. Any assumption that an excluded middle fails yields a contradiction. This is a general feature of intuitionistic logic; we can prove $\ulcorner\neg\neg(\phi \vee \neg\phi)\urcorner$ for any meaningful sentence ϕ, and as ascriptions of meaningfulness are always themselves meaningful, we can in particular prove this when ϕ is "Λ is meaningful". So there is no point in trying to define what "$\mathbb{T}[\Lambda]$" could mean when Λ is neither meaningful nor meaningless, because we know that the statement "Λ is neither meaningful nor meaningless" is false. (But that does not tell us that Λ definitely is either meaningful or meaningless; again, in intuitionistic logic the double negative is not removable.)

That is the complete resolution of the classical liar paradox. We cannot make the first paradoxical argument without assuming Λ is meaningful, we cannot make the second paradoxical argument without assuming Λ is not meaningful, and we cannot a priori affirm that Λ is definitely either meaningful or not.

Maybe it doesn't feel like a complete resolution, because can we not say that the first paradoxical argument showed that assuming Λ is meaningful yields a contradiction, infer from this that Λ cannot be meaningful, and then apply the second paradoxical argument?

No, this is not correct. The second argument actually did show that assuming Λ is not meaningful yields a contradiction, so we do know that Λ is not not meaningful. But the first argument did not show that assuming Λ is meaningful yields a contradiction. At first glance, it didn't show anything, because according to my dogmatic demands, it is not possible to assume Λ is meaningful and then, under this assumption, make an argument some of whose steps are only meaningful if Λ is meaningful. We agreed on this. What we *can* say here is that if Λ were meaningful then the first paradoxical argument could be legitimately made. As that argument proves a contradiction, we can therefore recognize that if Λ were meaningful then "$\bot$" would be assertible. But, once again, without $\bot$-release we cannot conclude from this that Λ is not meaningful.

That's as far as we can go. If the liar sentence is meaningful then "$\bot$" is assertible, and in addition the liar sentence is not not meaningful. In the terminology I introduced in Section 4.5, we can say that the statement that Λ is meaningful is anomalous. There really is no contradiction here. I can say that with confidence because I can formalize the above suggestions for reasoning about truth and meaningfulness and prove that the resulting system is consistent. I will do that in the next chapter; see Theorem 6.2.

We now see that the liar sentence is an example of a sentence which we almost certainly can neither affirm to be meaningful nor meaningless (because in either case, "$\bot$" would be assertible).

The notion of meaningfulness is nothing new, nor is intuitionistic logic. The key to the resolution of the liar paradox is our refusal to adopt the release law; otherwise, the fact that we could derive a contradiction if Λ were meaningful would entail that Λ is not meaningful, and the second paradoxical argument could not then be avoided. We saw in Section 4.5 that the release law is the culprit in the assertible liar paradox, and we have just learned that it is the culprit in the classical liar paradox too.

As we suspected back in Section 3.3, the resolution of Russell's paradox for predicates goes along virtually identical lines. First, of course, we must augment our definition of "holding", as we did with truth, to include the case when $\phi(\hat{x})$ is not meaningful. Thus, writing "$\mathbb{H}[\phi, x]$" for "ϕ holds of x", we affirm $\ulcorner \mathbb{H}[\text{“}\phi\text{”}, x] \leftrightarrow \phi(\hat{x})\urcorner$ when $\phi(\hat{x})$ is meaningful, and $\mathbb{H}[\phi, x] \leftrightarrow (0 = 1)$ when $\phi(\hat{x})$ is not meaningful.

Now consider the Russell predicate $\mathcal{R}$ = "does not hold of itself". The relevant question here is whether this predicate can be meaningfully applied to itself, whether the sentence $\mathcal{R}(\mathcal{R})$ is meaningful. And when we ask whether $\mathcal{R}(\mathcal{R})$ is meaningful, we find that it is not not meaningful (if it were not meaningful then $\mathcal{R}$ could not hold of itself, and hence $\mathcal{R}(\mathcal{R})$ would be true and therefore meaningful, contradiction) and if it were meaningful then "$\bot$" would be assertible (if it were meaningful, one could easily infer the equivalence of "$\mathcal{R}$ holds of itself" and "$\mathcal{R}$ does not hold of itself", and then deduce "$\bot$".). That is, the statement that $\mathcal{R}(\mathcal{R})$ is meaningful is anomalous.

We can also define a constructive version of "holding", where ϕ constructively holds of x if $\phi(\hat{x})$ is assertible. The constructive version of Russell's paradox arising from the predicate $\mathcal{R}_a$ = "is not assertible of itself" is easily handled; we get that $\mathbb{A}[\mathcal{R}_a(\mathcal{R}_a)]$ is anomalous, just like what we found with the assertible liar paradox.

5.2 Axiomatizing meaningfulness

Meaningfulness of quantifiers. Axioms for meaningfulness. Schematic definitions, schematic proofs, and the Tarskian catastrophe. The conjunction law.

We still need to talk about meaningfulness for quantified statements.

The obvious criterion for $\ulcorner(\forall x)\phi(x)\urcorner$ and $\ulcorner(\exists x)\phi(x)\urcorner$ to be meaningful is that $\phi(\hat{x})$ should be meaningful for all values of x. But we need to say something about the range of x. If we intend to quantify not over all the objects there are, but only over those which satisfy some predicate ρ, then we can try to make sense of expressions like $\ulcorner(\forall x : \rho(x))\phi(x)\urcorner$ and $\ulcorner(\exists x : \rho(x))\phi(x)\urcorner$.

These expressions can be naturally interpreted as $\ulcorner(\forall x)(\rho(x) \to \phi(x))\urcorner$ and $\ulcorner(\exists x)(\rho(x) \wedge \phi(x))\urcorner$, where now x ranges over all objects without restriction. But here the meaningfulness criterion which asks for $\ulcorner\rho(\hat{x}) \to \phi(\hat{x})\urcorner$ and $\ulcorner\rho(\hat{x}) \wedge \phi(\hat{x})\urcorner$ to be meaningful for all values of x would require $\phi(\hat{x})$ to be meaningful for all x, and that is clearly too strong. However, this is one of those cases where the meaning of $\phi(\hat{x})$ for values of x not satisfying ρ are irrelevant. Since it doesn't matter what $\phi(\hat{x})$ means for values of x outside the intended range, we can simply define $\phi(\hat{x}) \leftrightarrow (0 = 1)$ for those x such that $\rho(\hat{x})$ is false.

As we saw in the last section, though, this kind of definition by cases will only be complete if $\rho(\hat{x})$ is definitely true or false for all values of x. If there are values of x for which we cannot affirm $\ulcorner\rho(\hat{x}) \vee \neg\rho(\hat{x})\urcorner$, then we cannot say what $\phi(\hat{x})$ is supposed to mean for those values, and then the expressions $\ulcorner(\forall x)(\rho(x) \to \phi(x))\urcorner$ and $\ulcorner(\exists x)(\rho(x) \wedge \phi(x))\urcorner$ cannot be assumed to be meaningful for all values of x. Conclusion: $\ulcorner(\forall x : \rho(x))\phi(x)\urcorner$ and $\ulcorner(\exists x : \rho(x))\phi(x)\urcorner$ are meaningful if (1) ρ is *universally meaningful* and *definite*, i.e., $\rho(\hat{x})$ is meaningful for all x without restriction and we have $\ulcorner\rho(\hat{x}) \vee \neg\rho(\hat{x})\urcorner$ for all x, and (2) $\phi(\hat{x})$ is meaningful for all x of which ρ holds.

A little more generally, we can allow ρ to be indefinite provided ϕ is meaningfully defined on a definite range that includes all those x of which ρ holds. For instance, if $\phi(\hat{x})$ is meaningful for all x without restriction, then we can make sense of $\ulcorner(\forall x : \rho(x))\phi(x)\urcorner$ and $\ulcorner(\exists x : \rho(x))\phi(x)\urcorner$ for any universally meaningful predicate ρ, definite or not. (But there is no real gain in generality here, because if ϕ is universally meaningful then we can always just rewrite $\ulcorner(\forall x : \rho(x))\phi(x)\urcorner$ and $\ulcorner(\exists x : \rho(x))\phi(x)\urcorner$ as $\ulcorner(\forall x)(\rho(x) \to \phi(x))\urcorner$ and $\ulcorner(\exists x)(\rho(x) \wedge \phi(x))\urcorner$, and these will be meaningful for any universally meaningful ρ, definite or not.)

In the last section I made a fuss about not accepting $\ulcorner\phi \to \psi\urcorner$ as meaningful if we only know that ψ is meaningful whenever ϕ is true. This had to do with not allowing ourselves to reason about the liar sentence under the premise that it is meaningful. Indeed, if we did accept such implications as meaningful then the liar paradox would reappear in the

guise of "if this sentence is meaningful, then it is not true". A simple way to see the necessity of being similarly strict with quantification is to consider the sentence "for all x such that this sentence is meaningful, it is not true".

I can now set down some formal axioms for meaningfulness. Let $\mathbb{M}[\phi]$ symbolize "ϕ is meaningful". Then compositionality is encapsulated in the axioms

(1) $\mathbb{M}[\ulcorner \phi \wedge \psi \urcorner] \leftrightarrow \mathbb{M}[\ulcorner \phi \vee \psi \urcorner] \leftrightarrow \mathbb{M}[\ulcorner \phi \rightarrow \psi \urcorner] \leftrightarrow (\mathbb{M}[\phi] \wedge \mathbb{M}[\psi])$

(i.e., each of these four statements implies the others) and, for any definite, universally meaningful predicate ρ,

(2) $\mathbb{M}[\ulcorner (\forall x : \rho(x))\phi(x) \urcorner] \leftrightarrow \mathbb{M}[\ulcorner (\exists x : \rho(x))\phi(x) \urcorner] \leftrightarrow (\forall x : \rho(x))\mathbb{M}[\phi(\hat{x})]$.

Note that (1) is a single axiom, not a scheme, with an implied universal quantification over "ϕ" and "ψ", which range over all sentences. In (2) "ρ" is fixed and "ϕ" and "ψ" range over all predicates. So (2) is schematic in ρ, but we can again universally quantify over ϕ and ψ.

We can also affirm that ascriptions of meaningfulness or assertibility are always themselves meaningful:

(3) $\mathbb{M}[\ulcorner \mathbb{M}[\text{“}\phi\text{”}] \urcorner]$
(4) $\mathbb{M}[\ulcorner \mathbb{A}[\text{“}\phi\text{”}] \urcorner]$,

with ϕ ranging over all sentences, and that anything assertible is meaningful:

(5) $\mathbb{A}[\phi] \rightarrow \mathbb{M}[\phi]$.

But for truth we cannot make these kind of blanket statements; we can only say

(6) $(\mathbb{M}[\phi] \vee \neg\mathbb{M}[\phi]) \rightarrow \mathbb{M}[\ulcorner \mathbb{T}[\text{“}\phi\text{”}] \urcorner]$.

This follows from the definition of truth, which is expressed in the two axioms

(7) $\mathbb{M}[\phi] \rightarrow \mathbb{A}[\ulcorner \mathbb{T}[\text{“}\phi\text{”}] \leftrightarrow \phi \urcorner]$
(8) $\neg\mathbb{M}[\phi] \rightarrow \mathbb{A}[\ulcorner \neg\mathbb{T}[\text{“}\phi\text{”}] \urcorner]$.

Note here that if ϕ is known to be meaningful then the first prong of the definition yields simply $\mathbb{A}[\ulcorner\mathbb{T}[\text{“}\phi\text{”}] \leftrightarrow \phi\urcorner]$, and an application of the release rule would then recover that instance of the T-scheme.

In axiom (7) we have used assertibility to encode a schematic definition in a single statement. This is the solution of the Tarskian catastrophe discussed in Section 3.5. We do not use truth to globally affirm all instances of the T-scheme, we use assertibility. Any individual sentence which is known to be meaningful can be substituted into (7), and we can then use the release rule to infer the relevant instance of the T-scheme.

Let us also revisit the axioms for assertibility presented in Section 4.5. Those axioms were stated for ϕ ranging over meaningful sentences. But if sentences that might not be meaningful are in play, then all of these axioms besides the logical axiom and the capture scheme still make sense. The logical axiom should be replaced by

(9) $(\mathbb{M}[\phi] \wedge \mathrm{Log}(\phi)) \rightarrow \mathbb{A}[\phi]$

and the capture scheme should be replaced by the single axiom

(10) $\mathbb{M}[\phi] \rightarrow \mathbb{A}[\ulcorner\phi \rightarrow \mathbb{A}[\text{“}\phi\text{”}]\urcorner]$.

For any particular ϕ, once we have proven that it is meaningful we can use the release rule to recover the relevant instance of the original capture scheme. We have again used assertibility to encapsulate a scheme in a single statement.

We can also use assertibility to reason schematically, an issue that arose in Section 3.5 in relation to the conjunction law $(\mathbb{T}[\phi] \wedge \mathbb{T}[\psi]) \rightarrow \mathbb{T}[\ulcorner\phi \wedge \psi\urcorner]$. We can formalize that kind of reasoning as follows. First, from axiom (7) we have

$$(\forall\phi : \mathbb{M}[\phi])\mathbb{A}[\ulcorner\mathbb{T}[\text{“}\phi\text{”}] \leftrightarrow \phi\urcorner],$$

$$(\forall\psi : \mathbb{M}[\psi])\mathbb{A}[\ulcorner\mathbb{T}[\text{“}\psi\text{”}] \leftrightarrow \psi\urcorner],$$

$$(\forall\phi, \psi : \mathbb{M}[\ulcorner\phi \wedge \psi\urcorner])\mathbb{A}[\ulcorner\mathbb{T}[\text{“}\phi \wedge \psi\text{”}] \leftrightarrow (\phi \wedge \psi)\urcorner].$$

Reasoning under $\mathbb{A}$, the first two of these entail

$$(\forall\phi, \psi : \mathbb{M}[\ulcorner\phi \wedge \psi\urcorner])\mathbb{A}[\ulcorner(\mathbb{T}[\text{“}\phi\text{”}] \wedge \mathbb{T}[\text{“}\psi\text{”}]) \leftrightarrow (\phi \wedge \psi)\urcorner],$$

and combining this with the third yields

$$(\forall\phi, \psi : \mathbb{M}[\ulcorner\phi \wedge \psi\urcorner])\mathbb{A}[\ulcorner(\mathbb{T}[\text{“}\phi\text{”}] \wedge \mathbb{T}[\text{“}\psi\text{”}]) \leftrightarrow \mathbb{T}[\text{“}\phi \wedge \psi\text{”}]\urcorner].$$

For any particular ϕ and ψ that are known to be meaningful, we can invoke the release rule to infer $(\mathbb{T}[\phi] \wedge \mathbb{T}[\psi]) \leftrightarrow \mathbb{T}[\ulcorner \phi \wedge \psi \urcorner]$. Indeed, for any set-sized language $\mathcal{L}$ all of whose sentences are known to be meaningful, we can use $\forall$-capture followed by the release rule to infer

$$(\forall \phi, \psi \in \mathcal{L}^s)((\mathbb{T}[\phi] \wedge \mathbb{T}[\psi]) \leftrightarrow \mathbb{T}[\ulcorner \phi \wedge \psi \urcorner])$$

where $\mathcal{L}^s$ is the set of sentences of $\mathcal{L}$. We have made a schematic argument by working under an assertibility predicate.

We can make arguments involving potentially meaningless statements by working under an assertibility predicate. That is, we cannot argue "suppose ϕ is meaningful, then [some chain of reasoning involving statements that are only meaningful if ϕ is], and therefore ψ", but we might still be able to recognize that if ϕ were meaningful then the subsequent chain of reasoning would be valid. In that case we cannot say $\ulcorner$if ϕ is meaningful, then $\psi \urcorner$, but we can say $\ulcorner$if ϕ is meaningful, then ψ is assertible$\urcorner$. Potentially meaningless statements can be reasoned with under an assertibility predicate. That is what we did in the last section in order to say that if Λ is meaningful then "$\perp$" is assertible.

5.3 Propositions, concepts, and relations

Propositions. Propositions as objects. Concepts and relations. Russell's paradox for concepts.

In this section I will define propositions, concepts, and relations. Propositions are an abstract version of sentences, concepts are an abstract version of predicates, and relations are an abstract version of relational phrases (Section 3.3). The rough idea is that two sentences ϕ and ϕ' should instantiate the same proposition if they are equivalent, i.e., we have $\ulcorner \phi \leftrightarrow \phi' \urcorner$; two predicates ψ and ψ' should instantiate the same concept if they hold of the same objects, i.e., we have $\ulcorner (\forall x)(\psi(x) \leftrightarrow \psi'(x)) \urcorner$; and so on.

However, just writing "$(\phi \sim \phi') \leftrightarrow (\phi \leftrightarrow \phi')$" is ungrammatical because the the first "ϕ" is in noun position and the second is not. A better attempt is

$$(\phi \sim \phi') \leftrightarrow \mathbb{T}[\ulcorner \phi \leftrightarrow \phi' \urcorner],$$

but this runs into trouble if we quantify over all sentences, including some which might be neither definitely meaningful nor definitely meaningless. We discussed this sort of issue in the last section. The correct approach is

to treat "$(\phi \sim \phi') \leftrightarrow \mathbb{T}[\ulcorner \phi \leftrightarrow \phi' \urcorner]$" as a schematic definition, to be used for any meaningful ϕ and ϕ'. This can be expressed using assertibility as

$$(\forall \phi, \phi')(\mathbb{M}[\phi \wedge \phi'] \rightarrow \mathbb{A}[(\phi \sim \phi') \leftrightarrow \mathbb{T}[\ulcorner \phi \leftrightarrow \phi' \urcorner]]).$$

We now have a legitimate identity criterion and we can use it to talk about propositions as "sentences modulo equivalence".

That is how we define classical propositions. We could also use constructive truth, i.e., assertibility, for the identity criterion and get a constructive version of propositions. That is, we could define

$$(\phi \sim \phi') \leftrightarrow \mathbb{A}[\ulcorner \phi \leftrightarrow \phi' \urcorner]$$

for any sentences ϕ and ϕ'. This is not a schematic definition: it makes sense for any sentences, meaningful or not. However, if ϕ is not meaningful then neither is $\ulcorner \phi \leftrightarrow \phi \urcorner$ and hence the latter is not assertible. So ϕ only instantiates a proposition if it is meaningful.

Coming back to something I said in Section 3.2, we have seen that the classical liar sentence is not not meaningful (or, as I put it there, it is "not meaningless"). Thus it does not fail to express a proposition. But of course that does not tell us that it does express a proposition. Another comment I can make here is that the "proposition" version of the liar paradox arising from the sentence "this sentence does not express a true proposition" is resolved in the same way as the standard liar paradox. The assertion that this sentence expresses any proposition at all, i.e., that it is meaningful, is anomalous.

The preceding explanation of propositions shows that they are themselves a type of object. Back in Section 2.1 I warned that there would come a time when we would have to deal with indefinite objects, and we are now at that point. If new objects are introduced by abstraction using an indefinite identity criterion, then we cannot affirm that any two of them definitely are or are not equal. There could be indefinite objects. In particular, some propositions could be indefinite objects. That is because any meaningful sentence that is not known to have a definite truth value is potentially true, so the proposition expessed by such a sentence is potentially the same as the proposition expressed by "$0 = 0$". But without knowing whether this sentence will ultimately turn out to have a truth value, we cannot affirm that the proposition it expresses definitely is or is not the same as the proposition "$0 = 0$" expresses. Propositions can be indefinite objects!

In fact, as long as there is one meaningful sentence which might be either true or false, *all* propositions will be indefinite objects. For if ϕ is such a

sentence and ψ is any meaningful sentence, then $\ulcorner(\phi \wedge \psi) \vee (\neg\phi \wedge \neg\psi)\urcorner$ will express the same proposition as ψ if ϕ is true, but not if ϕ is false. So all propositions are indefinite objects. This will also be true of the concepts, relations, and classes I am about to define, for similar reasons.

Now let us say what "concepts" are. Again, there are two versions, a classical one and a constructive one. The classical version arises from predicates via the schematic identity criterion

$$(\psi \sim \psi') \leftrightarrow \mathbb{T}[\ulcorner(\forall x)(\psi(x) \leftrightarrow \psi'(x))\urcorner],$$

defined for all universally meaningful predicates ψ and ψ'. This can be expressed in a single statement using an assertibility predicate, as we did for propositions. More generally, we can restrict to a definite range and consider all ψ and ψ' which are meaningful on the same definite range of objects, i.e., such that $(\forall x)(\mathbb{M}[\psi(\hat{x})] \leftrightarrow \mathbb{M}[\psi'(\hat{x})])$ and $(\forall x)(\mathbb{M}[\psi(\hat{x})] \vee \neg\mathbb{M}[\psi(\hat{x})])$. The range must be definite so that the quantification in $\ulcorner(\forall x)(\psi(x) \leftrightarrow \psi'(x))\urcorner$ is meaningful, as we discussed in the last section.

As with propositions, the constructive version of concepts is not schematic. We simply say that ψ is equivalent to ψ' if they are meaningful on the same objects, and for any x on which they are meaningful $\ulcorner\psi(\hat{x}) \leftrightarrow \psi'(\hat{x})\urcorner$ is assertible. Classical concepts come from predicates which are meaningful on a definite range of objects, but there is no restriction on the predicates which give rise to constructive concepts.

We say that an object x classically (respectively, constructively) *falls under* the classical (constructive) concept instantiated by ψ if $\psi(\hat{x})$ is true (assertible).

We can abstract from relational phrases to get abstract relations in a similar way. This establishes propositions, concepts, and relations as different kinds of objects.

Now that concepts are a kind of object, is there a concept of being a concept? Yes. This is immediate in the constructive setting, where every predicate instantiates a concept, in particular the predicate "is a constructive concept". In the classical setting, we just have to note that the predicate "is a classical concept" is universally meaningful: it always means something to say of an object that it is a classical concept, because it always means something to say of an object that it is a predicate that definitely is or is not meaningful when applied to any object. However, the predicate "is a classical concept that does not classically fall under itself" does not appear to instantiate a classical concept, because it is not meaningful on a

definite range of objects. This may be confusing if one thinks of concepts as occult non-physical objects; we must remember that concepts are just predicates with a weaker identity criterion. It is presumably not the case that for every predicate ϕ

$$\ulcorner(\forall x)(\mathbb{M}[\phi(\hat{x})] \vee \neg\mathbb{M}[\phi(\hat{x})]) \wedge \neg\mathbb{T}[\phi(\text{“}\phi\text{”})]\urcorner$$

definitely is or is not meaningful. The first part is always meaningful, but the second part is problematic. So, under my definition of classical concepts there is no classical Russell concept, or more precisely, we have no reason to affirm there is such a thing.

In constrast, the predicate “is a constructive concept that does not constructively fall under itself” does instantiate a constructive concept. It is paradoxical in the same way that “is a predicate that does not constructively hold of itself” is paradoxical, which is to say that the assertion that it constructively falls under itself is anamolous.

5.4 Classes

Classes. Russell's paradox for classes. The inadequacy of a platonic interpretation of classes. Is there a class of all pure sets?

We are finally in a position to be able to say what classes are! I started trying to do this in Section 3.1, and then got sidetracked for two chapters developing the technology of assertibility, but now we are ready. The definition I want to use is similar to the definition of a classical concept given in the last section, but here we impose two additional requirements. I will say that classes are instantiated by predicates ϕ which are universally meaningful, i.e.,

$$(\forall x)\mathbb{M}[\phi(\hat{x})],$$

with x ranging over all objects; definite, i.e.,

$$(\forall x)(\mathbb{H}[\phi, x] \vee \neg\mathbb{H}[\phi, x]);$$

and *definitely applicable*, i.e.,

$$(\forall x)(\phi(x) \to (\forall y)(x = y \vee x \neq y)).$$

A *class* is an object instantiated by such a predicate, under the identity criterion which says that two predicates are equivalent if they (classically) hold of the same objects. We write $\{x : \phi(x)\}$ for the class instantiated by ϕ and we say that x *belongs to* or *is an element of* that class if ϕ holds of

x. Thus, for any class C, every object definitely does or does not belong to C and any object that does is definite.

This definition could be modified in any number of ways. It is merely a question of terminology, of what we want the word "class" to mean.

I noted earlier that all propositions are (presumably) indefinite objects. The same is true of classes: they are not definite, or more carefully, we do not affirm them to be definite. To see why not, pretend you are an intuitionist and let ϕ be the predicate "is an even number greater than 2 which is not a sum of two primes". This predicate only holds of natural numbers, if it holds of anything. If we take "natural number" to mean "finite decimal string", say, then it should be the case that every object definitely is or is not a natural number. So ϕ is definitely applicable. Moreover, a finite computation will tell us whether any given number is a sum of two primes, so even intuitionists will agree that ϕ is definite. Thus it does instantiate a class. But according to the intuitionistic version of constructivism which does not admit infinite structures or computations, the question of whether there exists an even number greater than 2 which is not a sum of two primes might not be something we can either affirm or deny. If so, it would be indefinite whether ϕ instantiates the empty class. The question whether ϕ and the predicate "is not equal to itself" instantiate the same class would not have a definite truth value.

That shows how the empty class could be indefinite. In general, let ϕ be any predicate that instantiates a class, let ψ be any predicate that instantiates a class which might or might not be empty, and consider the "exclusive or" of ϕ and ψ, i.e., the predicate "either $\phi(x) \wedge \neg\psi(x)$ or $\neg\phi(x) \wedge \psi(x)$". This predicate is also definite and definitely applicable, and it holds of the same objects as ϕ if and only if ψ instantiates the empty class. So as long as there is any predicate which neither definitely does nor definitely does not instantiate the empty class, *every* class will be indefinite.

Under my definitions of "set" and "class", it follows that classes cannot appear as elements of sets or classes. We have agreed that sets and classes can only have definite objects as elements, and classes are (apparently[1]) never definite. In Section 1.3, I asked why a proper class could never be an element of a set, if it exists in the same way as any other object. We now see that classes, as I define them here, do *not* exist in the same way as most ordinary mathematical objects. The identity criterion by which they

[1]I have to keep making this hedge because we do not rule out the possibility, however remote, that all sentences (and thus also all objects) might turn out to be definite. The point is that as of now we have no reason to positively affirm this.

are defined is indefinite. And thus, proper or not, they are not candidate elements of any set or class. Again, this statement is merely definitional: they could fall under various concepts, but they cannot belong to any set or class simply because of the way I have defined sets and classes.

According to the definitions I am using here, sets are definite objects and can therefore be elements of sets and classes, but classes are indefinite and cannot. That goes for all classes, even "set-sized" classes. There is an empty set and there is an empty class, and they have the same elements — we may say that they are *coextensive* — but they really are completely different kinds of objects (one is a syntactic list modulo equivalence, the other is a predicate modulo equivalence), one of which is definite and the other of which is not. Every set definitely is or is not empty, but there could be classes whose equality with the empty class is indefinite.

It is now easy to see why the class version of Russell's paradox fails. The predicate "is a class that does not belong to itself" does not instantiate a class. Indeed, it is neither (better: we cannot affirm it to be) definite nor definitely applicable. Because of the way I have defined classes, this version of the paradox is rather boring. The interesting versions are the ones involving predicates and concepts.

Thinking of classes as mysterious supernatural entities completely blocks any understanding of this. As I have tried to explain, classes are introduced into our language via abstraction, such that talking about classes is nothing more than a tidied up way of talking about definite, definitely applicable predicates modulo a natural identity criterion. So asking whether classes are definite objects is the same as asking whether this identity criterion is definite.

The value of this observation is that it turns an inscrutable metaphysical conundrum — could there be abstract classes whose equality is indefinite? — into the still subtle, but fundamentally tractable question of whether there could be two predicates for which it is indefinite whether they hold of the same objects. I must emphasize the hopeless inadequacy of a platonic interpretation of classes in confronting Russell's paradox (to the extent that it even counts as an "interpretation", if it only tells us that abstract classes are not physical objects and do not exist in space and time, without giving any positive account of their actual nature). It is hard to imagine a platonic abstract object that might be neither definitely the empty class nor definitely not the empty class — well, it is hard to imagine a platonic abstract object, period — but maybe not so hard to imagine a predicate that we are able neither to affirm nor deny holds of any objects. At any

rate, it should not be hard after one has become comfortable with truth constructivism.

One other point. There is no class of all classes, but is there a class of all pure sets? In Section 2.6 we discussed two ways of representing pure sets, either as well-ordered strings composed of the characters "{", "}", and ",", or as well-founded, rigid, rooted trees. More needs to be said about how either of these models is supposed to be physically realized, but there should not be any serious difficulties here aside from the issue of determining whether a potential model is well-ordered or well-founded. To answer this question we must quantify over all nonempty subsets, to check whether there is one with no smallest element. This should be fine for finitists and Cantorians, but countablists will face the problem that any countably infinite string or tree has uncountably many nonempty subsets, so the question of whether there is a nonempty subset which has no smallest element might not have a definite answer. Thus, finitists and Cantorians should be able to affirm that there is a class of all pure sets, but countablists cannot. For what it is worth.

Here are the answers to the last four questions raised in Chapter 1:

4. Is there anything more to identifying a collection than saying which things belong to it, and doesn't any well-defined property do this? No, if by "collection" we mean "class", there is nothing more to identifying one than saying which things belong to it. But in the constructivist world not all predicates can be used for this purpose.
5. Are all collections sets, and if not what makes some of them fail to be sets? What does "indefinitely extensible" really mean? The "class" notion of collection is more general than the "set" notion. Some predicates represent classes which fail to be coextensive with any set because there is no possible world containing an explicit list of every object of which the predicate holds: given any explicit list in any possible world, there will always be another possible world which contains a more comprehensive list. That is what it means to say that a class is indefinitely extensible. Another way to put this is that having completed an element-by-element survey of such a class is not a possible state of affairs.
6. What makes a collection a "totality"? I take "totality" to be synonymous with "set", and again, what makes a class coextensive with a set is the existence, in some possible world, of an explicit list of all of its elements — a list which can be surveyed to completion.
7. Why can't proper classes be elements of sets, or other classes? We require

of both sets and classes that their elements must be definite objects. But classes, as I have defined them — proper or not — are (apparently) not definite objects, and are therefore forbidden from being elements of sets or classes.

Chapter 6

Interlude: The Liar

6.1 The two versions

Informal derivations of the classical and assertible liar paradoxes.

In Sections 4.5 and 5.1 I explained how to resolve the constructive and classical versions of the liar paradox. Since the problem is so old and so basic, I am dedicating the present short chapter to a more detailed explanation. The issue is fundamentally significant not because the liar sentences are themselves important, but because our failure to understand what has gone wrong with them indicates some sort of basic lack of understanding of truth or logic. Many attempts to resolve one or the other version of the paradox have been advanced, but they have all failed (cf. Section 3.1).

"This sentence is not true" is the classical liar sentence; call it Λ. If we reinterpret "truth" as constructive truth, i.e., assertibility — provability in a general semantic sense — and also take the "not" to be constructive, we get the assertible liar sentence Λ_a = "if this sentence is assertible then $0 = 1$". Thus, writing "$\mathbb{T}$" and "$\mathbb{A}$" for truth and assertibility, we have

$$\Lambda := \text{“}\neg\mathbb{T}[\Lambda]\text{”} \qquad \text{and} \qquad \Lambda_a := \text{“}\mathbb{A}[\Lambda_a] \to 0 = 1\text{”},$$

by definition. The classical liar paradox can then be derived as follows:

(1) Assume Λ.
(2) Infer both $\neg\mathbb{T}[\Lambda]$ (definition of Λ) and $\mathbb{T}[\Lambda]$ (since we have $\phi \leftrightarrow \mathbb{T}[\phi]$ for any sentence ϕ).[1]
(3) Assuming Λ has led to a contradiction, so we conclude $\neg\Lambda$.

[1] This is more accurately expressed as $\ulcorner \phi \leftrightarrow \mathbb{T}[\text{“}\phi\text{”}] \urcorner$ using the quasi-quotation marks $\ulcorner$ and $\urcorner$ introduced in Section 3.4. But in order to make this chapter independent I will omit all quasi-quotation marks and follow the normal convention of being a little sloppy about this sort of thing. Everything in Sections 6.4 and 6.5 will be strictly rigorous.

(4) Infer $\neg\mathbb{T}[\Lambda]$ (since, again, $\Lambda \leftrightarrow \mathbb{T}[\Lambda]$).
(5) Infer Λ (definition of Λ).
(6) We have proven both Λ and $\neg\Lambda$, contradiction.

The assertible liar paradox goes like this:

(1) Assume $\mathbb{A}[\Lambda_a]$.
(2) Infer Λ_a (since Λ_a is assertible).
(3) Infer $\mathbb{A}[\Lambda_a] \to 0 = 1$ (definition of Λ_a).
(4) Combining (1) and (3) yields $0 = 1$, so we have shown that $\mathbb{A}[\Lambda_a]$ entails that $0 = 1$.
(5) Infer Λ_a (definition of Λ_a).
(6) We have now proven Λ_a; therefore Λ_a is assertible, i.e., we have $\mathbb{A}[\Lambda_a]$.
(7) Combining (4) and (6) yields that $0 = 1$.

What is wrong with these arguments? I will explain.

6.2 The three ingredients

Intuitionistic logic. Proving falsehoods. The release law and release rule. Capture.

The resolutions of the classical and assertible liar paradoxes rest on three ingredients.

The first is intuitionistic logic: we have to use it when reasoning about open-ended qualities such as assertibility or meaningfulness. The essential point is that we do not assume the law of excluded middle $\phi \vee \neg\phi$ for every sentence ϕ. This idea alone is not enough, as neither of the arguments given in the last section uses the law of excluded middle, and hence both of them can be formalized in intuitionistic logic. But it is an essential feature of the full solutions (and the fact that it is provides a simple demonstration that we are indeed bound to use intuitionistic logic in these settings).

The second ingredient is that we must accept the *possibility* that there could be logically compelling proofs of falsehoods. Of course, that is completely outrageous because it would mean that logical reasoning is bankrupt; once any falsehood is provable, everything is provable, even "$0 = 1$". But "outrageous" does not entail "false".

When we reason classically, if the conclusion of some argument is false and the premises are all true, then there must be some definite step where a false statement is derived from true ones, and hence there must be one

particular inference that is definitely faulty. But we cannot say this in a constructivist setting. Lacking excluded middle, we might not be able to affirm that each statement in the argument is definitely true or false, and hence might not be able to affirm that there must be one definite step where a false statement is inferred from true ones. This creates a conceivable possibility that we could have an argument that concludes that $0 = 1$, in which we are not able to identify any faulty deductive step. When we reason constructively, we must simply take it on faith that this could never happen.

In case you cannot imagine what such an argument might look like, I direct your attention to the two paradoxical arguments given in Section 6.1. On the face of it, every step of both of these deductions appears correct. In both cases we seem to be rationally compelled to accept the paradoxical conclusion. Of course, no one believes that the liar paradox really is valid ... actually no, some people, called "dialetheists", do believe this. But most people assume that there must be a flaw somewhere.

I am not claiming that the dialetheists are right, only that we must not rule out the possibility that they could be, that there could be a rationally compelling proof of a falsehood.

The significance of this admission is that if there really were a logically compelling proof of some falsehood ϕ, then "ϕ can be proven" would be true but ϕ would be false. This means that the general principle "if ϕ can be proven, then ϕ" would fail. That is, admitting the possibility that logical reasoning might be unreliable entails that we cannot affirm that for every sentence ϕ, the assertibility of ϕ entails ϕ — in symbols, we cannot affirm the general principle $\mathbb{A}[\phi] \to \phi$. We lose this *release law*.

(Arguably one can justify a deduction rule version of release that infers ϕ from $\mathbb{A}[\phi]$, provided no temporary assumptions are in play; see Section 4.2. I will use this rule below, but as I will explain, only for cosmetic purposes; it is not essential here. Another parenthetical comment is that in constructivism the reverse *capture* principle $\phi \to \mathbb{A}[\phi]$ is valid; see Section 4.2 for more on this also. This law can be used to prove interesting things about the classical and assertible liar sentences, and it does not lead to any paradox, as shown in Theorems 6.1 and 6.2 below.)

In the derivation of the assertible liar paradox, we use release in step (2). As we are, at that point in the argument, working under the temporary assumption that Λ_a is assertible, we cannot make this inference, and that nips the paradoxical derivation in the bud. In an expression of perfect irony, the alleged proof of "$0 = 1$" is rendered invalid precisely by our accepting

the possibility that there could be a valid proof of "0 = 1". To put that differently: if we could be sure that there is no compelling proof of any falsehood, then we would have to accept the release law and would then be forced to acknowledge the assertible liar argument as a compelling proof of a falsehood. Cazart!

But wait. As we have just seen, from the premise that "0 = 1" is not assertible we can deduce that "0 = 1" is assertible, which contradicts the premise. Does this not show that the premise is false, that "0 = 1" actually is assertible?

Yes and no. Yes, the premise is false, but no, that does not tell us that "0 = 1" is assertible. It tells us that "0 = 1" is not not assertible. At this point constructivism saves us, because in intuitionistic logic the double negative is not removable. We can informally read "$\neg\neg\mathbb{A}$['0 = 1']" as "it will always be possible that '0=1' is assertible", which is the position I was arguing for earlier.

These two ingredients, using intuitionistic logic and not endorsing the release law, suffice to resolve the assertible liar paradox. The classical liar paradox additionally turns on a third ingredient, the notion of meaningfulness.

6.3 Meaningfulness and truth

Compositionality. Are truth ascriptions always meaningful? Defining truth schematically via deduction rules and globally via assertibility.

Meaningfulness is the key issue in the classical liar paradox. I am not endorsing the simplistic suggestion that Λ is meaningless; if it were meaningless then it could not be true, and from this we can infer Λ itself, leading straight back into a contradiction. (And attempting to block this last inference is bound to fail; see the "quining" construction in Section 3.2.) No, the situation is more subtle than that.

Notice that this little argument shows that the derivation of the liar paradox given in Section 6.1 is incomplete. It only works if Λ is meaningful — we need to know this in order to take the very first step, where we assume Λ. You can't assume something that has no meaning. So the paradox actually requires two separate arguments; if Λ is meaningful then we make the argument given in Section 6.1, and if Λ is not meaningful we make a completely different argument, the one I gave just now.

How do we reason about meaningfulness?

The fundamental principle I require is that if any component of a sentence is meaningless, then the sentence as a whole is also meaningless. Thus even such a simple statement as "$\phi \to \phi$" has no meaning if ϕ has no meaning. In some settings it may be possible to reason sensibly using sentences that contain meaningless bits, but for us this restriction is absolutely crucial.

We can express this principle by saying that the statements

- ϕ and ψ are both meaningful
- $\phi \wedge \psi$ is meaningful
- $\phi \vee \psi$ is meaningful
- $\phi \to \psi$ is meaningful

are all equivalent. (Statements with quantifiers need a separate condition; see Section 5.2. But we do not need to worry about that in this chapter because the sentences we will be evaluating for meaningfulness here do not contain any quantifiers.)

Let me now introduce the symbol "$\mathbb{M}$" for meaningfulness. Thus $(\mathbb{M}[\phi] \wedge \mathbb{M}[\psi]) \leftrightarrow \mathbb{M}[\phi \to \psi]$, etc.

Adhering strictly to this condition has a subtle consequence that will be important for us. If we know that ϕ is meaningful but not that ψ is meaningful then we do not know that $\phi \to \psi$ is meaningful, and therefore we cannot affirm $\phi \to \psi$ — a sentence can only be affirmed if it is known to be meaningful. But suppose we knew that ϕ entails the meaningfulness of ψ, i.e., suppose we had $\phi \to \mathbb{M}[\psi]$. Could we then regard $\phi \to \psi$ as a sensible statment? It is plausible to think that we can, but unless we already know that ϕ is the case, we cannot. If we know that ϕ is the case then we know that ψ is meaningful, and therefore we know that $\phi \to \psi$ is meaningful (and equivalent to ψ). But if we do not know that ϕ is the case, and we do not have any other reason to be sure ψ is meaningful, then we cannot be sure about the meaningfulness of the implication $\phi \to \psi$. And therefore it is not something we can affirm.

Let us now consider the relationship between truth and meaningfulness. The statement that a given sentence is meaningful is itself always meaningful, as is the statement that a given sentence is assertible. But truth is a defined notion, unlike meaningfulness and assertibility which are both primitive, so we must look to the formal definition of "ϕ is true" to determine in which cases this is a meaningful statement.

Truth is usually defined by Tarski's biconditional $\mathbb{T}[\phi] \leftrightarrow \phi$, but this

can only be done if ϕ is known to be meaningful. But we at least know that whenever ϕ is meaningful, so is "$\mathbb{T}[\phi]$" — it is equivalent to ϕ. Also, as I suggested above, it is part of our naive conception of truth that meaningless sentences can never be true, so we can add a separate clause to the definition of truth which states that "$\mathbb{T}[\phi]$" is equivalent to "$0 = 1$" whenever ϕ is not meaningful. Thus, if ϕ is meaningful then its truth is defined á la Tarski, and if ϕ is not meaningful, it simply cannot be true. But, lacking the law of excluded middle, we cannot say that every sentence definitely is or is not meaningful. And this means that we cannot make a blanket statement that "$\mathbb{T}[\phi]$" is always meaningful. If ϕ is not definitely known to be either meaningful or not, then we cannot say what "$\mathbb{T}[\phi]$" means because we do not know which (if either) clause of the definition of truth to use.

This already shows where the flaw in the classical liar paradox lies. The very first step of the derivation is "assume Λ". This would only be legitimate if we had previously proven that Λ is meaningful, but we do not have such a proof. We can only assume sentences that mean something. Now, we saw at the beginning of this section that it is possible to derive a contradiction from the assumption that Λ is not meaningful (cf. Theorem 6.1 (b) below), but without excluded middle this does not entail that Λ definitely is meaningful. It only shows that Λ is not not meaningful. Constructivism to the rescue again!

Truth is defined schematically. If we know that ϕ is definitely meaningful, or definitely not meaningful, then we can apply one of the two prongs in the definition of truth. Thus we have deduction rules

- given $\mathbb{M}[\phi]$, infer $\mathbb{T}[\phi] \leftrightarrow \phi$
- given $\neg\mathbb{M}[\phi]$, infer $\neg\mathbb{T}[\phi]$.

But these only work as deduction rules, not implications, because if we do not know that ϕ is definitely meaningful or not meaningful then we do not know whether $\mathbb{T}[\phi]$ is meaningful and therefore cannot affirm either $\mathbb{M}[\phi] \to (\mathbb{T}[\phi] \leftrightarrow \phi)$ or $\neg\mathbb{M}[\phi] \to \neg\mathbb{T}[\phi]$. This is the point I was making a few paragraphs back.

What we can say, for arbitrary ϕ, is that the meaningfulness of ϕ entails the assertibility of $\mathbb{T}[\phi] \leftrightarrow \phi$. That is, we can affirm $\mathbb{M}[\phi] \to \mathbb{A}[\mathbb{T}[\phi] \leftrightarrow \phi]$. Similarly, we can affirm that $\neg\mathbb{M}[\phi] \to \mathbb{A}[\neg\mathbb{T}(\phi)]$. This is how we can make statements about the truth of sentences which might not be meaningful.

In summary, we have both deduction rules for truth that do not involve assertibility, and implications which do. But let me note that the release

rule briefly mentioned earlier (given $\mathbb{A}[\phi]$, infer ϕ) obviates the need for separately stating these deduction rules. Given any implication of the form $\phi \to \mathbb{A}[\psi]$, once we have proven ϕ we can infer $\mathbb{A}[\psi]$ and then use the release rule to infer ψ. Thus, the release rule allows us to convert axioms of the form "$\phi \to \mathbb{A}[\psi]$" into deduction rules of the form "given ϕ, infer ψ". This shows that in the presence of the release rule, the deduction rules for truth follow from the assertibility implications for truth. This observation will enable a small simplification in the formal system to be described in the next section: I will include the assertibility implications for truth but omit the corresponding deduction rules in favor of the release rule. If one has doubts about this rule one can simply replace it with the deduction rules it is used to generate without affecting the validity of anything in this chapter.

6.4 Formalization

A formal system ATM for reasoning about assertibility, truth, and meaningfulness. Groundedness. The classical liar is not not meaningful, and if it is meaningful then falsehoods can be proven. The assertible liar is not not assertible, and if it is assertible then falsehoods can be proven.

Let us do all this formally. I will describe a minimalistic formal system ATM which allows treatment of sentences that can talk about each other's (or their own) assertibility, truth and meaningfulness.[2]

The language of ATM is specified as follows. First, we introduce the constant symbol $\dot{\bot}$ and a sequence of constant symbols $L_1, L_2, \ldots$. Each constant symbol is a *term*, and whenever s and t are terms so are $(s \mathbin{\dot{\wedge}} t)$, $(s \mathbin{\dot{\vee}} t)$, $(s \mathbin{\dot{\to}} t)$, $\dot{\mathbb{A}}[t]$, $\dot{\mathbb{T}}[t]$, and $\dot{\mathbb{M}}[t]$. I will sometimes omit parentheses, and I will also write "$\dot{\neg} t$" as an informal abbreviation of "$t \mathbin{\dot{\to}} \dot{\bot}$" and "$s \mathbin{\dot{\leftrightarrow}} t$" as an informal abbreviation of "$(s \mathbin{\dot{\to}} t) \dot{\wedge} (t \mathbin{\dot{\to}} s)$". All terms are built up from the constant symbols in the above manner; there are no variables in the language.

The atomic sentences of the language are $\bot$, which is interpreted as "$0 = 1$", and $\mathbb{A}[t]$, $\mathbb{T}[t]$, and $\mathbb{M}[t]$ for any term t. Whenever ϕ and ψ are sentences so are $(\phi \wedge \psi)$, $(\phi \vee \psi)$, and $(\phi \to \psi)$. As with terms, I will write "$\neg\phi$" as an abbreviation of "$\phi \to \bot$" and "$\phi \leftrightarrow \psi$" as an abbreviation of "$(\phi \to \psi) \wedge (\psi \to \phi)$". This is a propositional language, with no quantification.

For every sentence ϕ there is a corresponding term $\dot{\phi}$ simply obtained by putting dots over every $\bot$, $\wedge$, $\vee$, $\to$, $\mathbb{A}$, $\mathbb{T}$, and $\mathbb{M}$ symbol in ϕ. Conversely,

[2]It is a more elaborate version of the system discussed in Section 5.1 of [54].

every term t can be "evaluated" to a sentence $\hat{t}$ such that $\dot{\phi}$ always evaluates back to ϕ. This is done by letting $\dot{\bot}$ evaluate to $\bot$; letting $\dot{\mathbb{A}}[t]$, $\dot{\mathbb{T}}[t]$, and $\dot{\mathbb{M}}[t]$ evaluate to $\mathbb{A}[t]$, $\mathbb{T}[t]$, and $\mathbb{M}[t]$, respectively, for any term t (note that t is left unchanged); and letting $s \mathbin{\dot{\wedge}} t$, $s \mathbin{\dot{\vee}} t$, and $s \mathbin{\dot{\to}} t$ evaluate to $\hat{s} \wedge \hat{t}$, $\hat{s} \vee \hat{t}$, and $\hat{s} \to \hat{t}$. We also declare that each L_i evaluates to a specific sentence θ_i. The θ_i can be chosen in any manner whatever, but for the sake of concreteness let θ_1 and θ_2 be the classical and assertible liar sentences "$\neg\mathbb{T}[L_1]$" and "$\neg\mathbb{A}[L_2]$", respectively. Thus θ_1 states that θ_1 is not true (since $\hat{L}_1 = \theta_1$), and θ_2 states that θ_2 is not assertible (since $\hat{L}_2 = \theta_2$).

We could also formulate in this language a "liar pair" of sentences that affirm each other's truth and untruth, a variant assertible liar that affirms the assertibility of its negation, a variant classical liar that says of itself that if it is meaningful then it is not true, etc.

The logical axioms of ATM consist of all sentences of the form

$$(\mathbb{M}[s] \wedge \mathbb{M}[t] \wedge \mathbb{M}[u]) \to \mathbb{A}[\dot{A}(s,t,u)],$$

for any terms s, t, and u, where A is one of the Hilbert-style axiom schemes (1) through (9) for intuitionistic logic given in Section 4.5, and we insert s, t, and u in place of ϕ, ψ, and θ in the term $\dot{A}$.

These logical axioms all include a meaningfulness premise. But we will adopt nonlogical axioms which state that every sentence that contains truth in a "grounded" way is meaningful. Given these axioms, we can infer the assertibility of all the Hilbert axioms for any grounded sentences, and then, using the release rule, infer the Hilbert axioms themselves for grounded sentences. (In particular, we may use the Hilbert axioms when reasoning with any of our logical axioms, as they are all grounded.) Thus we reason as usual about grounded sentences, but only with some qualification about ungrounded sentences. Nonetheless, we will still be able to prove interesting things about ungrounded sentences.

Groundedness of sentences is defined according to the following rules.

- $\bot$ is grounded
- $\mathbb{A}[t]$ and $\mathbb{M}[t]$ are grounded, for any term t
- if ϕ and ψ are grounded then so are $\phi \wedge \psi$, $\phi \vee \psi$, and $\phi \to \psi$
- if $\hat{t}$ is grounded then $\mathbb{T}[t]$ is grounded.

The grounded sentences constitute the smallest set of sentences satisfying these conditions.[3] The idea is that sentences are ungrounded if they refer-

[3]This is similar to, but not quite the same as, Kripke's notion of groundedness [20].

ence truth in a circular way.

The nonlogical axioms of ATM consist of the following schemes (1) through (9). They are all schemes because s and t could be any terms.

(1) $\mathbb{M}[t]$, for any term t such that $\hat{t}$ is grounded
(2) $(\mathbb{M}[s] \wedge \mathbb{M}[t]) \leftrightarrow \mathbb{M}[s \mathbin{\dot{\wedge}} t] \leftrightarrow \mathbb{M}[s \mathbin{\dot{\vee}} t] \leftrightarrow \mathbb{M}[s \mathbin{\dot{\rightarrow}} t]$
(3) $\mathbb{A}[t] \rightarrow \mathbb{M}[t]$
(4) $(\mathbb{A}[s] \wedge \mathbb{A}[t]) \rightarrow \mathbb{A}[s \mathbin{\dot{\wedge}} t]$
(5) $(\mathbb{A}[s] \wedge \mathbb{A}[s \mathbin{\dot{\rightarrow}} t]) \rightarrow \mathbb{A}[t]$
(6) $\mathbb{M}[t] \rightarrow \mathbb{A}[t \mathbin{\dot{\rightarrow}} \dot{\mathbb{A}}[t]]$
(7) $\mathbb{M}[t] \rightarrow \mathbb{A}[t \mathbin{\dot{\leftrightarrow}} \dot{\mathbb{T}}[t]]$
(8) $\neg\mathbb{M}[t] \rightarrow \mathbb{A}[\dot{\neg}\dot{\mathbb{T}}[t]]$
(9) $\mathbb{M}[t] \rightarrow \mathbb{A}[t \mathbin{\dot{\leftrightarrow}} t']$, for any terms t and t' with $\hat{t} = \hat{t}'$.

Scheme (1) affirms the meaningfulness of all grounded sentences; scheme (2) says that a complex sentence is meaningful if and only if its components are; scheme (3) affirms that anything assertible is meaningful. Schemes (4) and (5) describe two basic inferences involving assertibility, and (6) is the capture scheme $\hat{t} \rightarrow \mathbb{A}[t]$ mentioned in Section 6.2 (and discussed in detail in Section 4.2), modulated with $\mathbb{M}$ and $\mathbb{A}$ to allow for the possibility that $\hat{t}$ is meaningless. Schemes (7) and (8) concern what one can say about $\mathbb{T}[t]$ if $\hat{t}$ is definitely either meaningful or not meaningful, and scheme (9) expresses the idea that any two terms which evaluate to the same sentence are interchangeable.

ATM has three rules of inference:

(1) given ϕ and ψ, infer $\phi \wedge \psi$ (conjunction)
(2) given ϕ and $\phi \rightarrow \psi$, infer ψ (modus ponens)
(3) given $\mathbb{A}[t]$, infer $\hat{t}$ (release).

As I mentioned earlier, we will only use the release rule for cosmetic purposes. Namely, we will use it to get deduction rule versions of the logical axiom schemes and the nonlogical schemes (6) through (9) (e.g.: given $\mathbb{M}[t]$, infer $\hat{t} \rightarrow \mathbb{A}[t]$, etc.). So alternatively we could omit the release rule and include deduction rule versions of the logical schemes and the nonlogical schemes (6) through (9).

I also mentioned earlier that we can only reason about ungrounded sentences with some qualification. Specifically, working with sentences of the form $(\mathbb{M}[t_1] \wedge \cdots \wedge \mathbb{M}[t_n]) \rightarrow \mathbb{A}[\dot{A}(t_1, \ldots, t_n)]$ and using scheme (5), we can prove $(\mathbb{M}[t_1] \wedge \cdots \wedge \mathbb{M}[t_n]) \rightarrow \mathbb{A}[\dot{B}(t_1, \ldots, t_n)]$ for any theorem

$B(\hat{t}_1, \ldots, \hat{t}_n)$ of the intuitionistic propositional calculus. I will say that *given meaningfulness, we can reason under* $\mathbb{A}$.

ATM dispenses a wealth of information. With regard to the specific sentences θ_1 and θ_2 defined above (the classical and assertible liar sentences), we can say the following.

Theorem 6.1. *ATM proves*
(a) $\mathbb{M}[L_1] \to \mathbb{A}[\dot{\bot}]$
(b) $\neg\neg\mathbb{M}[L_1]$
(c) $\mathbb{A}[L_2] \to \mathbb{A}[\dot{\bot}]$
(d) $\neg\neg\mathbb{A}[L_2]$.

Proof. All the sentences appearing in this proof are grounded and therefore meaningful (though some terms appearing in these sentences, e.g., the "L_1" in $\neg\neg\mathbb{M}[L_1]$, evaluate to ungrounded sentences). This means that the usual Hilbert-style axioms for intuitionistic logic can be used. For if $\hat{s}$, $\hat{t}$, and $\hat{u}$ are grounded then we get $\mathbb{M}[s]$, $\mathbb{M}[t]$, and $\mathbb{M}[u]$ from scheme (1), then $\mathbb{M}[s] \wedge \mathbb{M}[t] \wedge \mathbb{M}[u]$ from the conjunction rule, and from this we can infer $\mathbb{A}[\dot{A}(s,t,u)]$ from any of the logical axioms by modus ponens. The Hilbert-style axiom $A(\hat{s}, \hat{t}, \hat{u})$ then follows using the release rule. So we can reason normally in intuitionistic logic when all the sentences in play are grounded.

To prove part (a), assume $\mathbb{M}[L_1]$. Infer $\mathbb{A}[L_1 \dot{\leftrightarrow} \dot{\mathbb{T}}[L_1]]$ from (7), then infer $\mathbb{M}[L_1 \dot{\leftrightarrow} \dot{\mathbb{T}}[L_1]]$ from (3) and use (2) to get first $\mathbb{M}[\dot{\mathbb{T}}[L_1]]$ and then $\mathbb{M}[\dot{\neg}\dot{\mathbb{T}}[L_1]]$, i.e., $\mathbb{M}[\dot{\theta}_1]$. Now use (9) with $t = L_1$ and $t' = \dot{\theta}_1$ and reason under $\mathbb{A}$ (see the note just preceding this theorem) to get $\mathbb{A}[\dot{\theta}_1 \dot{\leftrightarrow} \dot{\mathbb{T}}[L_1]]$, i.e., $\mathbb{A}[\dot{\neg}\dot{\mathbb{T}}[L_1] \dot{\leftrightarrow} \dot{\mathbb{T}}[L_1]]$, and from this $\mathbb{A}[\dot{\bot}]$. We have shown that $\mathbb{M}[L_1] \to \mathbb{A}[\dot{\bot}]$.

For part (b), assume $\neg\mathbb{M}[L_1]$. Use (8) to get $\mathbb{A}[\dot{\neg}\dot{\mathbb{T}}[L_1]]$, i.e., $\mathbb{A}[\dot{\theta}_1]$, then (3) to get $\mathbb{M}[\dot{\theta}_1]$. This yields $\mathbb{A}[\dot{\theta}_1 \dot{\leftrightarrow} L_1]$ by (9), then $\mathbb{M}[\dot{\theta}_1 \dot{\leftrightarrow} L_1]$ by (3), and then $\mathbb{M}[L_1]$ by (2). Combining this with the initial $\neg\mathbb{M}[L_1]$ yields $\bot$ (modus ponens), so we have proven $\neg\mathbb{M}[L_1] \to \bot$, i.e., $\neg\neg\mathbb{M}[L_1]$.

For part (c), assume $\mathbb{A}[L_2]$. Infer $\mathbb{M}[L_2]$ from (3), then $\mathbb{A}[L_2 \leftrightarrow \dot{\theta}_2]$ from (9), and then use schemes (3) and (2) to infer $\mathbb{M}[\dot{\theta}_2]$ as well. We are now in a position to assume meaningfulness and reason under $\mathbb{A}$ to get $\mathbb{A}[L_2 \to \dot{\theta}_2]$. (Since $(\phi \leftrightarrow \psi) \to (\phi \to \psi)$ is a theorem of the intuitionistic propositional calculus, and using (5).) Combining this with the assumption of $\mathbb{A}[L_2]$, again using (5), then yields $\mathbb{A}[\dot{\theta}_2]$, i.e., $\mathbb{A}[\dot{\neg}\dot{\mathbb{A}}[L_2]]$. Finally, (6) with $t = L_2$ yields $\mathbb{A}[L_2 \dot{\to} \dot{\mathbb{A}}[L_2]]$, so we get $\mathbb{A}[\dot{\mathbb{A}}[L_2]]$ from (5). Combining that with $\mathbb{A}[\dot{\neg}\dot{\mathbb{A}}[L_2]]$, i.e., $\mathbb{A}[\dot{\mathbb{A}}[L_2] \dot{\to} \dot{\bot}]$, using (5) yet again, finally yields $\mathbb{A}[\dot{\bot}]$. Assuming $\mathbb{A}[L_2]$, we proved $\mathbb{A}[\dot{\bot}]$; thus, we have shown that $\mathbb{A}[L_2] \to \mathbb{A}[\dot{\bot}]$.

For part (d), start by noting that $\neg\mathbb{A}[L_2]$ is grounded and therefore meaningful, so (6) plus the release rule yields $\neg\mathbb{A}[L_2] \to \mathbb{A}[\dot{\neg}\dot{\mathbb{A}}[L_2]]$. Now, assuming $\neg\mathbb{A}[L_2]$, deduce $\mathbb{A}[\dot{\neg}\dot{\mathbb{A}}[L_2]]$, i.e., $\mathbb{A}[\dot{\theta}_2]$, and then use (9) to get $\mathbb{A}[L_2]$, in the same way that we went from $\mathbb{A}[L_2]$ to $\mathbb{A}[\dot{\theta}_2]$ in the proof of part (c). Combining this with the initial $\neg\mathbb{A}[L_2]$ yields $\bot$ (modus ponens), so we have proven $\neg\mathbb{A}[L_2] \to \bot$, i.e., $\neg\neg\mathbb{A}[L_2]$. □

Is the classical liar sentence true or not? This sounds like a meaningful question, but until we have said what it means for a sentence to be true, it is not. We do know what it means for a sentence to be true if the sentence is meaningful, and we also know what it means if the sentence is not meaningful — but these two definitions are completely separate, so we have to decide which category to place the liar sentence in before we can make sense of the question of its truth. Asking whether the liar sentence is true is only cogent if the liar sentence is definitely meaningful or not meaningful, as otherwise we do not know what the assertion that it is true means. So is the classical liar sentence meaningful? Well, it is not not meaningful, and if it is meaningful then "$0 = 1$" is assertible.

There is no contradiction here. We avoid contradiction by remaining in a state of uncertainty about the meaningfulness of the liar sentence.

Indeed, we can infer that if the classical liar sentence is either meaningful or not meaningful then "$0 = 1$" is assertible, and draw the same conclusion if the assertible liar sentence is either assertible or not assertible. Both of these statements follow from Theorem 6.1, using capture to get $\neg\mathbb{M}[L_1] \to \mathbb{A}[\dot{\bot}]$ and $\neg\mathbb{A}[L_2] \to \mathbb{A}[\dot{\bot}]$ from parts (b) and (d).

Another inference we can make is that $\neg\mathbb{A}[\dot{\bot}]$ entails a contradiction — because assuming $\neg\mathbb{A}[\dot{\bot}]$ converts parts (a) and (c) of Theorem 6.1 to $\neg\mathbb{M}[L_1]$ and $\neg\mathbb{A}[L_2]$, contradicting parts (b) and (d). So $\bot$ is not not assertible. This will sound absurd if one is thinking classically, because it would mean that $\bot$ is assertible, but not if one is thinking constructively. Constructively, there is no simplification beyond "if the assertibility of $\bot$ entails $\bot$, then $\bot$", and this formulation is valid and informative. A reasonable informal gloss of "$\neg\neg\mathbb{A}[\dot{\bot}]$" is that $\mathbb{A}[\dot{\bot}]$ will always be possible (see Section 4.3).

6.5 Consistency

The system ATM from Section 6.4 is consistent.

The nonexistence of paradoxes in ATM is a provable theorem.

Theorem 6.2. *ATM is consistent.*

Proof. We will construct a set $\mathcal{F}$ of sentences that contains $\bot$, does not contain any of the axioms, and whose complement is stable under the deduction rules. This shows that all the theorems of ATM lie in the complement of $\mathcal{F}$, and therefore $\bot$ cannot be a theorem.

The desired set is constructed in an infinite series of *levels*, each of which consists of an infinite series of *stages*, each of which consists of an infinite series of *steps*. At each step new sentences are added to $\mathcal{F}$; nothing is ever removed.

Let S be any set of implications. We will use the following rules:

- ($\wedge$ rule) if $\phi \in \mathcal{F}$ and ψ is any sentence, place $\phi \wedge \psi$ and $\psi \wedge \phi$ in $\mathcal{F}$
- ($\vee$ rule) if $\phi, \psi \in \mathcal{F}$, place $\phi \vee \psi$ in $\mathcal{F}$
- ($\mathbb{T}$ rule) if $\phi \in \mathcal{F}$ then place $\mathbb{T}[t]$ in $\mathcal{F}$ for every term t that evaluates to ϕ.
- ($\rightarrow$ rule) for any implication $\phi \rightarrow \psi$ in S such that $\phi \in \mathcal{F}^c$ and $\psi \in \mathcal{F}$, place $\phi \rightarrow \psi$ in $\mathcal{F}$

Define a function i ("implication complexity") from the set of sentences to ω by letting $i(\phi)$ be the number of appearances of the symbol "$\rightarrow$" in ϕ, and define a function n ("nesting") from the set of sentences to $\omega + 1$ by setting $n(\phi) = 0$ when ϕ is $\bot$, $\mathbb{M}[t]$, or $\mathbb{A}[t]$, for any term t; $n(\phi \wedge \psi) = n(\phi \vee \psi) = n(\phi \rightarrow \psi) = \max(n(\phi), n(\psi))$; $n(\mathbb{T}[t]) = n(\hat{t}) + 1$; and $n(\phi) = \omega$ for any ungrounded sentence ϕ. These functions will be used to construct $\mathcal{F}$.

When we are at a given step of the construction I will write $\mathcal{F}_-$ for the state of $\mathcal{F}$ going into that step.

The construction goes through ω levels labelled $p = 0, 1, \ldots$ ("principal steps"). Every level has $\omega+1$ stages labelled $n = 0, 1, \ldots, \omega$ (corresponding to nesting). Stage 0 on any level has ω steps labelled $i = 0, 1, \ldots$ and every other stage has ω steps labelled $i = 1, 2, \ldots$ (corresponding to implication complexity).

At step 0, stage 0, level 0, when $\mathcal{F}_- = \emptyset$, we put $\bot$ in $\mathcal{F}$ and then close up under the $\wedge$, $\vee$, and $\mathbb{T}$ rules. That is, at the end of this step $\mathcal{F}$ is the smallest set that contains $\bot$ and is stable under the $\wedge$, $\vee$, and $\mathbb{T}$ rules.

At step 0, stage 0, level 1, we place $\mathbb{M}[t]$ and $\mathbb{A}[t]$ in $\mathcal{F}$ for every term t whose evaluation $\hat{t}$ is ungrounded, and also every $\mathbb{A}[t]$ such that $\hat{t} \in \mathcal{F}_-$.

Then we close up under the $\wedge$, $\vee$, and $\mathbb{T}$ rules.

At step 0, stage 0 of any other level, we place every $\mathbb{A}[t]$ such that $\hat{t} \in \mathcal{F}_-$ in $\mathcal{F}$, then close up under the $\wedge$, $\vee$, and $\mathbb{T}$ rules.

At step $i \geq 1$, stage n of any level (including $n = \omega$), let S be the set of implications $\phi \to \psi$ with $n(\phi \to \psi) = n$ and $i(\phi \to \psi) = i$, and apply the $\to$ rule to $\mathcal{F}_-$, then close up under the $\wedge$, $\vee$, and $\mathbb{T}$ rules. (We do not attempt to close up under the $\to$ rule.)

That completes the description of $\mathcal{F}$. Now we have to verify that its complement, once the construction is complete, contains all the axioms stated in Section 6.4 and is stable under all of the deduction rules given there. Three helpful observations are

- if the sentence θ is added to $\mathcal{F}$ at step i of stage n on any level, then either $n(\theta) > n$ or else $n(\theta) = n$ and $i(\theta) \geq i$
- if $\phi \to \psi$ is added in to $\mathcal{F}$ anywhere on a given level, then at the end of that level ϕ will not (yet) have been added to $\mathcal{F}$
- at step 0, stage 0 of any level, if $\phi \notin \mathcal{F}_-$ and $\psi \in \mathcal{F}_-$ then $\phi \to \psi \in \mathcal{F}_-$ (i.e., at these points in the construction $\mathcal{F}_-$ is stable under the $\to$ rule with $S =$ all implications).

The first observation is trivial when $i = n = 0$, and at any step $i \geq 1$, if $\phi \to \psi$ is added in to $\mathcal{F}$ then $n(\phi \to \psi) = n$ and $i(\phi \to \psi) = i$. Then we close up under the $\wedge$, $\vee$, and $\mathbb{T}$ rules, but the set of sentences θ with $n(\theta) > n$ or $n(\theta) = n$, $i(\theta) \geq i$ is stable under these rules. The second observation can be seen by noting that when $\phi \to \psi$ is added in at stage n and step i, at that point ϕ does not belong to $\mathcal{F}$, and we also have either $n(\phi) < n(\phi \to \psi) = n$ or else $n(\phi) = n(\phi \to \psi) = n$ and $i(\phi) < i(\phi \to \psi) = i$. So by the first observation, ϕ has already missed any opportunity to be added in during the current level. The third observation holds because $n(\psi) \leq n(\phi \to \psi)$ and $i(\psi) < i(\phi \to \psi)$, so if ψ appears at any point in some level and ϕ does not, the step where $\phi \to \psi$ is potentially added in must come later in that level, and $\phi \to \psi$ *will* get added in at this later step.

We verify that $\mathcal{F}^c$ is stable under the three deduction rules. First, the only way $\phi \wedge \psi$ could be added is at one of the many points where we close up under the $\wedge$ rule. This means that if $\phi \wedge \psi$ appears in $\mathcal{F}$, then either ϕ or ψ must also appear, showing that the complement of $\mathcal{F}$ is stable under the conjunction deduction rule: if neither ϕ nor ψ lies in $\mathcal{F}$ then $\phi \wedge \psi$ cannot either.

For modus ponens, suppose ψ belongs to $\mathcal{F}$ and let ϕ be arbitrary. We want to show that either ϕ or $\phi \to \psi$ will also belong to $\mathcal{F}$. This follows from the "second observation" made above: at whatever level ψ first appears on, if ϕ has not appeared by the end of that level then $\phi \to \psi$ has. So if both ϕ and $\phi \to \psi$ are in $\mathcal{F}^c$, then so is ψ.

The release rule holds because if ψ appears in $\mathcal{F}$ at any level, then $\mathbb{A}[t]$ will appear in step 0, stage 0 of the following level, for any term t that evaluates to ψ. Thus $\mathbb{A}[t] \in \mathcal{F}^c$ implies $\hat{t} \in \mathcal{F}^c$.

Next, let us show that none of the nonlogical axioms belongs to $\mathcal{F}$. The first comment is that no instance of scheme (1) can appear at any point because the only place sentences of the form $\mathbb{M}[t]$ are placed in $\mathcal{F}$ is step 0, stage 0, level 1, and this is only done if $\hat{t}$ is ungrounded. Nor can any instance of scheme (2) be placed in $\mathcal{F}$ at any point, because if either $\hat{s}$ or $\hat{t}$ is ungrounded then all of $\mathbb{M}[s] \wedge \mathbb{M}[t]$, $\mathbb{M}[s\dot{\wedge}t]$, $\mathbb{M}[s\dot{\vee}t]$, and $\mathbb{M}[s\dot{\to}t]$ are placed $\mathcal{F}$ simultaneously, at step 0, stage 0, level 1; and if $\hat{s}$ and $\hat{t}$ are both grounded, none of them is ever placed in $\mathcal{F}$.

Closing up under the $\wedge$, $\vee$, and $\mathbb{T}$ rules never introduces any of the axioms from schemes (3) through (9), because these are all implications. Nor are any of these axioms introduced at step 0 of any stage and level, because no implications are introduced in those steps. Thus we must check that no instance of any of the schemes (3) — (9) can be added to $\mathcal{F}$ using the $\to$ rule at some step $i \geq 1$. Taking scheme (3), we see that the $\to$ rule will never add in something of the form $\mathbb{A}[t] \to \mathbb{M}[t]$, because if $\hat{t}$ is ungrounded then $\mathbb{A}[t]$ and $\mathbb{M}[t]$ both appear in $\mathcal{F}$ at the same time, and if $\hat{t}$ is grounded then $\mathbb{M}[t]$ never appears. So there is never a point at which $\mathbb{M}[t]$ has appeared and $\mathbb{A}[t]$ has not, which is a requirement to add $\mathbb{A}[t] \to \mathbb{M}[t]$ with the $\to$ rule. For scheme (4) observe that if either $\hat{s}$ or $\hat{t}$ is ungrounded then both sides of the implication appear in $\mathcal{F}$ at the same time, at step 0 of stage 0, level 1. If both are grounded then $\mathbb{A}[s\dot{\wedge}t]$ can only be added on step 0, stage 0 of some level such that $\hat{s} \wedge \hat{t}$ first appeared in $\mathcal{F}$ somewhere on the previous level. But $\hat{s} \wedge \hat{t}$ can never be added before both $\hat{s}$ and $\hat{t}$, so at least one of $\mathbb{A}[s]$ and $\mathbb{A}[t]$ must be added at the same time $\mathbb{A}[s\dot{\wedge}t]$ is added, and thus $\mathbb{A}[s] \wedge \mathbb{A}[t]$ must also appear at the same step.

For scheme (5), the "second observation" made earlier ensures that either $\hat{s}$ or $\hat{s} \to \hat{t}$ will appear at the same level as $\hat{t}$, or earlier, so at least one of $\mathbb{A}[s]$ and $\mathbb{A}[s\dot{\to}t]$ (and in either case, their conjunction) would appear at the same step as $\mathbb{A}[t]$, at the latest.

The left side of scheme (8) is placed in $\mathcal{F}$ at step 1, stage 0, level 0 (regardless of whether $\hat{t}$ is grounded), whereas the right side could not be

placed in $\mathcal{F}$ at any point in level 0.

In schemes (6), (7), and (9), if $\hat{t}$ is ungrounded then both sides are placed in $\mathcal{F}$ at the same time, at step 0, stage 0, level 1. So assume $\hat{t}$ is grounded. Every instance of these three schemes is an implication with a conclusion of the form $\mathbb{A}[u]$. Moreover, groundedness of $\hat{t}$ entails that $\hat{u}$ is grounded in every case. So $\mathbb{A}[u]$ could only appear subsequent to $\hat{u}$ appearing, and thus it will suffice to show that $\hat{u}$ never appears. In scheme (6) we get this because groundedness of $\hat{t}$ entails that $\mathbb{A}[t]$ cannot appear before $\hat{t}$; in scheme (7) we never get either $\hat{t} \to \mathbb{T}[t]$ or $\mathbb{T}[t] \to \hat{t}$, regardless of whether $\hat{t}$ is grounded, because the $\mathbb{T}$ rule is always invoked just after $\hat{t}$ is included, before we get a chance to apply the $\to$ rule to them. And scheme (9) is easy because $\mathbb{A}[t \dot{\leftrightarrow} t']$ can only be added after $\hat{t} \leftrightarrow \hat{t}'$ was added, and the latter can never happen, since $\hat{t} = \hat{t}'$.

Finally, we must show that none of the logical axioms can appear in $\mathcal{F}$. As these are all implications, we only have to show that none of them can be added at any step using the $\to$ rule. This clearly cannot happen if any of s, t, or u evaluates to an ungrounded sentence, because in that case the premise of the implication would have appeared at step 0, stage 0, level 1, and its conclusion could appear at that point at the earliest. If they all evaluate to grounded sentences, then we show that $\mathbb{A}[\dot{A}(s,t,u)]$ can never appear by checking that $A(\hat{s},\hat{t},\hat{u})$ never appears.

This can be seen using the three observations made earlier in the proof. For example, take the logical axiom scheme corresponding to (8) from Section 4.5, where $A(\hat{s},\hat{t},\hat{u})$ is the sentence $(\hat{s} \to (\hat{t} \to \hat{u})) \to ((\hat{s} \to \hat{t}) \to (\hat{s} \to \hat{u}))$. (This and (6) are the most complicated logical schemes.) If a sentence of this form appeared in $\mathcal{F}$ at any point, the second observation made above tells us that there would have to be a level at which $(\hat{s} \to \hat{t}) \to (\hat{s} \to \hat{u})$ is placed in $\mathcal{F}$ but $\hat{s} \to (\hat{t} \to \hat{u})$ is not. Then $\hat{s} \to \hat{u}$ must be placed in $\mathcal{F}$ in this level, as if it had appeared at an earlier level then $(\hat{s} \to \hat{t}) \to (\hat{s} \to \hat{u})$ would have too. So at the end of this level $\hat{u}$ is in $\mathcal{F}$ and $\hat{s}$ is not, and also $\hat{s} \to \hat{t}$ is not, so that since $\hat{s}$ is not in $\mathcal{F}$, $\hat{t}$ cannot be either. Thus at the end of this level we will have $\hat{u}$ in $\mathcal{F}$, but not $\hat{s}$ or $\hat{t}$. The second observation now tells us first that $\hat{t} \to \hat{u}$ must be in $\mathcal{F}$ at this point, and second that $\hat{s} \to (\hat{t} \to \hat{u})$ must be in $\mathcal{F}$ at this point. But this contradicts the earlier statement that $\hat{s} \to (\hat{t} \to \hat{u})$ does not appear in $\mathcal{F}$ at this level. We conclude that there is no level where $(\hat{s} \to (\hat{t} \to \hat{u})) \to ((\hat{s} \to \hat{t}) \to (\hat{s} \to \hat{u}))$ could be added. The other logical axioms are handled similarly. □

In more general settings we could also adopt deduction rules which,

given $\hat{t}$, let us infer $\mathbb{M}[t]$ and $\mathbb{A}[t]$. (Any theorem of a system we trust should be meaningful and assertible.) In the ATM setting, however, these rules are superfluous. This is a consequence of the next result, which follows easily from the proof of Theorem 6.2.

Corollary 6.3. *Every theorem of ATM is grounded.*

Proof. Go by induction on the length of a proof in ATM. All of the axioms are obviously grounded, and neither the conjunction rule nor modus ponens can derive an ungrounded sentence from grounded ones. The release rule could potentially do this, but we saw in the proof of Theorem 6.2 that $\mathbb{A}[t]$ cannot be a theorem if $\hat{t}$ is ungrounded, because every $\mathbb{A}[t]$ with $\hat{t}$ ungrounded is placed in $\mathcal{F}$ at step 0, stage 0, level 1. Thus the release rule also cannot introduce any ungrounded sentences. □

Axiom scheme (1) says that every grounded sentence is meaningful, so it follows from Corollary 6.3 that every theorem of ATM is meaningful, provably in ATM. If $\hat{t}$ is any theorem of ATM, scheme (6) then yields $\mathbb{A}[t \dot{\rightarrow} \dot{\mathbb{A}}[t]]$, and applying the release rule and modus ponens shows that $\mathbb{A}[t]$ is also a theorem of ATM.

Chapter 7

Countablism

7.1 Predicativism

Skolem's paradox. Actual infinity. Quantifying over $\mathbb{N}$ *and* $\mathbb{R}$*. Russell's "medical impossibility".*

This book is about constructive countablism, a foundational philosophy that rests on the two principles of truth constructivism and countablism. Chapters 3 through 5 gave an extended argument that we have little choice but to accept truth constructivism. We need it, and specifically the constructive notion of assertibility, to

- have a global self-applicative notion of classical truth
- say what classes are
- reason globally about propositions and concepts
- resolve the liar paradox and the class-theoretic and semantic versions of Russell's paradox.

These are things we cannot do without.

In contrast, I do not think there is any knock-down argument for countablism. However, I will argue that it is the most reasonable position, a middle ground between the skepticism of the finitists and the credulity of the Cantorians. The goal of this chapter is to make that case.

But first, it is interesting to look at the role of countablism in predicativism, the historical precursor to constructive countablism which I briefly mentioned in Section 1.4. Predicativism is nowadays understood to be based on the vicious circle principle together with countablism, so the similarity to constructive countablism is evident; we simply replace the vicious circle principle with truth constructivism. This is a natural move, since, as I explained in Section 1.4, the vicious circle principle really only makes

sense in the context of truth constructivism. However, we have also seen in the last few chapters how truth constructivism enables us to make sense of circular constructions such as the concept of being a concept, which classical predicativists would reject.

It is worth looking more closely at the relationship between countablism and predicativism, because that relationship is more complicated than one might think. The principal early exponents of predicativism, Poincaré, Russell, Skolem, and Weyl, all affirmed the vicious circle principle, but they did not all agree about countablism. In fact, of the four, only Skolem clearly and persistently advocated countablism, with statements like

> Since within any axiomatization of set theory or any formal logical system one reasons in such a way that the absolutely uncountable does not exist, the statement that uncountable sets exist can only be considered as a play on words, hence this absolutely uncountable is just a fiction.[1]

He especially emphasized what is now known as "Skolem's paradox": if ZFC has any models, then (by the Löwenheim-Skolem theorem) it has a countable model, yet ZFC proves the existence of uncountable sets and hence these must appear in any model. Most commentators have not seen this as a serious difficulty, as a countable set can easily appear to be uncountable within a model if the model contains no bijection between that set and $\mathbb{N}$. Indeed, Skolem pointed this out himself.[2] But his critique of uncountability was more subtle than this, as I will explain in Section 7.6. At any rate, we have here an affirmation of the idea that countably infinite collections exist, but not uncountably infinite collections.

Poincaré and Weyl express an unequivocal commitment to finitism in several places. (Poincaré: "There is no actual infinite; the Cantorians forgot that, and they fell into contradiction."[3] Weyl: "The notion that an infinite set is a 'gathering' brought together by infinitely many individual arbitrary acts of selection, assembled and then surveyed as a whole by consciousness, is nonsensical; 'inexhaustibility' is essential to the infinite."[4])

Poincaré and Weyl (at the time of writing his book *Das Kontinuum*) did both accept the definiteness of sentences which quantify over all natural

[1] [43], p. 468.

[2] "[T]here is no contradiction at all if a set M of the domain B is nondenumerable in the sense of the axiomatization; for this means merely that *within* B there occurs no one-to-one mapping Φ of M onto Z_0 (Zermelo's number sequence) [emphasis in original]" ([42], p. 295).

[3] [26], p. 316

[4] [56], p. 23 in the Pollard/Bole translation.

numbers, but not of sentences which quantify over all real numbers. Weyl again: " 'Fermat's last theorem,' for example, is intinsically meaningful and either true or false. But I cannot rule on its truth or falsity by employing a systematic procedure for sequentially inserting all numbers in both sides of Fermat's equation. Even though, viewed in this light, the task is infinite, it will be reduced to a finite one by the mathematical proof (which, of course, in this notorious case, still eludes us)."[5]

But if they both rejected actual infinity, why did they accept the definiteness of quantification over $\mathbb{N}$ (but not the real line, $\mathbb{R}$)? Weyl gives a full explanation of his views in *Das Kontinuum*[6] (and Poincaré's views, though less systematically expressed, were roughly similar). Since there is no actual infinity, real numbers cannot exist as literally infinite decimal expansions, or anything of that sort. They exist only by virtue of our having finite descriptions of them — things like an algorithm for computing the successive digits of π, for instance. But Richard's paradox[7] shows that the notion of a "finite description" of a real number is inherently vague. Since real numbers only truly exist as finite descriptions, it follows that the concept "real number" is also inherently vague. So it is senseless to quantify over it. In contrast, "the intuition of iteration assures us that the concept 'natural number' is existentially definite"[8], so quantifying over $\mathbb{N}$ is allowed.

(Granting the premise that all infinite collections are effectively proper classes, I think Weyl is absolutely right that real numbers have to be understood in this essentially linguistic way. Regardless of one's views on the legitimacy of infinity, the general principle is that proper classes by definition cannot be physically realized. They can only exist as linguistic objects. But that does not mean we are forbidden from quantifying over them, only that we have to be careful about it. We discussed this in Section 5.2.)

Weyl gives a second reason for rejecting quantification over $\mathbb{R}$, having to do with the idea that the real line must be constructed in stages. Stage 1 consists of all real numbers that can be defined only by quantifying over $\mathbb{N}$; stage 2 consists of all real numbers that can be defined by quantifying over $\mathbb{N}$ and all stage 1 reals; and so on. Since this process (allegedly) never

[5] [56], p. 49.
[6] See [56], pp. 29 and 111, and again in [57], p. 49.
[7] Enumerate all finitely long strings which, in English, unambiguously define real numbers. We can then diagonalize to get a new real number not in the list, but this diagonalization process can itself be described by a finite English string, so the new number should have appeared in the original list.
[8] [56], p. 110

terminates, there is no stage at which one can accept a definition of a real number which involves quantification over all real numbers.

What is surprising about this second reason is that it seems to be a clear expression of truth constructivism, so one wonders why, if there is no actual infinity, it would not apply equally well to quantification over $\mathbb{N}$. Evidently the inherent vagueness of $\mathbb{R}$ as opposed to $\mathbb{N}$ has something to do with this. But it is interesting that not long after stating these views Weyl abandoned predicativism in favor of intuitionism, the core principles of which I take to be rejection of actual infinity together with truth constructivism. So at the time of writing *Das Kontinuum* he may already have accepted these principles, but just not have worked them out to their logical conclusions.

Russell defended the cogency of countable constructions, for instance in his famous comment that running through all the digits of π is merely "medically impossible".[9] He goes on: "Might not [in principle] a man's skill increase so fast that he performed each operation in half the time required for its predecessor? In that case, the whole infinite series would take only twice as long as the first operation."[10] This comment is notable in that it can be seen as justifying countable constructions, but not uncountable ones. But, as far as I know, this is the closest he came to advocating countablism.

Skolem was the only one of the founding fathers of predicativism who clearly endorsed countablism. Perhaps one could argue that Weyl's comment about the existential definiteness of $\mathbb{N}$, when paired with his reasons for rejecting quantification over $\mathbb{R}$, effectively made him a countablist. But given his later switch to intuitionism and his explicit denial of actual infinity, this seems like a stretch. The association of predicativism as it is understood today with countablism has to do not with explicit endorsements by its originators, but with the line they drew between quantifying over $\mathbb{N}$ and quantifying over $\mathbb{R}$. It is this line which is generally considered to be characteristic of predicativism, even if the reasons for it have changed.

7.2 Finitism

Ultrafinitism. Thomson's lamp. Malament-Hogarth spacetimes and Davies machines.

Countablism faces challenges from two directions: from finitism, which urges stronger restrictions, and Cantorianism, which urges weaker ones.

[9] [39], p. 143.
[10] [39], p. 144.

Let us start with the finitists.

I admit that the idea that infinity exists only in potential and not in actuality has a basic intuitive appeal. Indeed, I have the impression that finitists tend to regard the idea as self-evident. Finite computations are possible and infinite computations impossible! Who could deny this? Notably, this attitude is embedded in the language of computability theory, where, e.g., "decidable" means "decidable in a finite number of steps".

But what is meant by "possible" and "impossible" here? It certainly seems unlikely that literally infinite computations could be performed in our universe (though I do not think we know this for a fact). Actually, under our current understanding of physics, it does not even look as though arbitrarily long finite computations are possible. We expect that our universe will eventually die a heat death, putting an upper bound on the physically possible lengths of computations. So if "possible" means "possible in our universe", we should go beyond finitism to *ultrafinitism*, the view that only "feasibly" long, not arbitrarily finitely long, computations are meaningful. There is a spectrum of skepticism, ultimately terminating in the solipsistic position that the only numbers that exist are those that are immediately available to one's consciousness at the present moment.

Indeed, anything that goes beyond extreme solipsism has to involve some degree of idealization. The question is then "what degree of idealization is legitimate?" But there need not be any definitive answer to this question. Framing the matter in terms of possible worlds, we should ask whether there are possible worlds in which computations of any finite length can be performed, and beyond that, whether literally infinite computations can be performed in any possible world.

There is no dearth of speculation on this topic. One of the most famous thought experiments along these lines, known as "Thomson's lamp", imagines a lamp which is switched on at some point in time, then switched off half a second later, then back on a quarter second after that, and so on. After one second it will have been switched on and off infinitely many times. What state would it be in then? According to Thomson,

> It cannot be on, because I did not ever turn it on without at once turning it off. It cannot be off, because I did in the first place turn it on, and thereafter I never turned it off without at once turning it on. But the lamp must be either on or off. This is a contradiction.[11]

And this is supposed to show that procedures involving infinitely many

[11] [48], p. 5.

steps are inherently impossible. But I am not very impressed by this reasoning. Why "cannot" the lamp be on? What is the rule of logic that prevents the lamp from, say, defaulting to an "on" position after infinitely many switches? I say "rule of logic" because if we are talking about what is physically possible in our universe then the whole thought experiment is ill-conceived. One might just as well declare that infinite tasks are impossible in our universe and end the discussion there. But if we are talking about what could happen in some possible world, then flat declarations that the lamp "cannot" be on or off seem quite senseless.[12]

(For that matter, if we are specifically talking about how our universe works, it isn't even true that the lamp "cannot" be in the "on" position a moment after it is switched off, or vice versa. Quantum mechanics tells us that in any time interval there is a minute but nonzero probability of it spontaneously transitioning from one state to the other.)

So much for Thomson's lamp. If someone says that they simply cannot conceive of a world in which infinite computations are possible, I accept this (though I may privately question their lack of imagination). But if they say that the idea of infinite computations is inherently contradictory, it is not well-taken.

Incidentally, some physicists have seriously proposed that infinite computations might in fact be possible in our universe. The literature on this subject uses the term "supertask". There have actually been a variety of suggestions as to how supertasks might be physically possible; the best-known is called "Malament-Hogarth spacetime".[13] It is a fact that there are models of general relativity in which some events are preceded by worldlines with infinite proper time. For example, if one falls into a rotating (Kerr) black hole, one should see time in the external universe accelerating, such that an infinite amount of external time will have passed just as one reaches the event horizon. The idea is that one could program a computer to perform an infinite computation, set it in orbit around the black hole, then jump in the black hole and await news of the results (to be contemplated in the very limited remainder of one's life after crossing the event horizon).

[12]It does not help Thomson's case when he later adds that "mathematicians do say that [the sequence $+1, -1, +1, \ldots$] has a sum; they say that its sum is $\frac{1}{2}$. And this answer does not help us, since we attach no sense to saying that the lamp is half-on" ([48], p. 6). This comment seems to show some lack of familiarity with both mathematics (we say that this sum is divergent, though it does have a Cesàro mean of $\frac{1}{2}$) and physics (in fact, we can attach good sense to saying that the lamp is in a quantum superposition of its on and off states, for example).

[13]See, e.g., [23].

For instance, the computer could systematically check whether every even number greater than 2 is a sum of two primes, with instructions to send a signal if any counterexample is found.

Personally, I am skeptical of ideas like this.[14] For instance, there is an attenuation problem having to do with one's ability to detect increasingly faint signals, the black hole should eventually evaporate, etc. I believe most physicists would share this skepticism. However, for my purposes here, any conclusion less final than "this is literally impossible" counts as a positive answer. My standard is what could happen in some possible world, not just in our universe. This means that as long as a supertask could be performed in some possible world, it counts as possible in principle.

Another, more intuitive approach to supertasks has been described by E. B. Davies.[15] It builds on Russell's idea, mentioned above, of a man with accelerating "skill". In Davies' formulation we imagine a Newtonian universe in which space, time, and matter are infinitely divisible, and we postulate the existence of a machine that is capable of building an exact replica of itself which is half as large and runs twice as fast. (The two conditions are not independent; actually, one might expect a machine half the size to run twice as fast, if its components move at the same speed but need to cover only half the distance. Maybe not if it had rotating parts, though.) The initial machine is designed to accept a natural number n as input, after which it is to perform the nth step of some computation, and then build a smaller, faster version of itself and feed it the input $n+1$. So if we build such a machine and feed it the number 1 as input, it will perform the first step of the computation, then proceed to build a smaller machine which will perform the second step and then build a third machine, and so on. If each computation takes half as long as the previous one, then an infinite amount of computation could take place in a finite length of time. The machines could be programmed to relay information back up the chain (stating that a counterexample of some sort had been found, for instance) so that the initial machine would register, after a finite amount of time, a definite result, including a negative result if no counterexample is signalled in time. In the latter case we would have the conclusion of a literally infinite computation.

[14] For the record, I do not think we know enough about high energy physics to definitively rule out the physical possibility of supertasks. I would say only that they seem unlikely to be possible in our universe (with greater skepticism toward specific proposals such as infinite computations via Malament-Hogarth spacetimes).

[15] [7].

Davies is quite clear that, according to our current understanding of physics, nothing like this should be possible in our universe. It is a thought experiment only. But again, my previous point stands that if we are asking about what could happen in some possible world, then plausible thought experiments have real force.

(I should mention that the attenuation problem for Malament-Hogarth spacetimes has an analog for Davies machines, having to do with the idea of relaying the result of a computation back up the chain of machines. If arbitrarily small machines could send signals that are dectectable on the scale of the original machine, and that all arrive after a finite time interval, then time evolution could not be continuous in this world. Arbitrarily small changes in the initial conditions could materially alter the outcome of the experiment, which is just what it means for time evolution to fail to be continuous. On the other hand, there is no sacrosanct principle that time evolution has to be continuous. For instance, it is a feature of the Copenhagen interpretation of quantum mechanics that systems experience a discontinuous jump when they are observed. There is plenty to be said against this interpretation, but I have never seen anyone take the alleged discontinuity of time evolution as a reason for disbelief.)

The reason for talking about Malament-Hogarth spacetimes and Davies machines is that they help build our intuition for worlds in which supertasks can be performed. The more plausible one finds this thought, the less willing one should be to accept finitism. In short, my argument against finitism is that one already has to go to possible worlds to get beyond "feasible" finite numbers, and our ability to imagine worlds which allow literally infinite computations seems just as robust as our ability to imagine worlds which allow arbitrarily long finite computations.

7.3 Cantorianism

Psychological basis. Reversing the order of topics. Conceivability and Löwenheim-Skolem. Nonstandard Davies machines.

What about Cantorianism, which says that the collection of all the subsets of an infinite set is always itself a set?

Our job here is rather different from the one we faced with finitism. There we were disputing the idea that infinite structures and computations are literally inconceivable, and our strategy was to paint a picture (actually, two pictures: Malament-Hogarth spacetimes and Davies machines) with the

intention of making such things seem plausible. But now we are trying to disabuse our interlocutors of the notion that uncountable structures and computations *are* conceivable. We have to somehow convince them that they really cannot imagine something they already think they can.

As with finitists, Cantorians are likely to regard their position as practically self-evident. This is seen in explanations of the iterative conception of sets, for instance (see Section 1.3). These explanations usually pass over without comment, as if it were obvious, the proposition that one could, at each successor stage, in some unexplained sense "gather together" *all* the sets whose elements were available at the previous stage. This idea is apparently seen as so straightforward and unremarkable as not to require any special justification.

Where does this attitude come from? We aren't born with it. What is it about the way set theory is taught and learned that makes the existence of infinite power sets as *sets*, as explicitly listable, seem reasonable? Starting with the power set of the natural numbers, $\mathcal{P}(\mathbb{N})$.

I suspect the answer may have to do with the order in which various topics are taught. The typical sequence of events could be something like this: we start out by learning that sets are collections; then we learn about various constructions such as power set, cartesian product, etc.; then we learn about infinite cardinalities and how they go on and on without limit; and then — only then — we learn about the paradoxes associated with taking the universe of sets to itself be a set, which are handled by some vague explanation about proper classes being collections which are "too big" to be sets, or something. (This is also more or less the historical order in which these topics were developed.) At this point one would naturally think that $\mathcal{P}(\mathbb{N})$ is not really so big, as there are so many vastly larger cardinalities. So the suggestion that it might be "too big" to be a set would sound obviously wrong.

The point is that if $\mathcal{P}(\mathbb{N})$ is brought in before we understand the set/class distinction, when "set" is still synonymous with "collection", then we have no alternative but to accept it as a set; and then we can build up an intuition for a universe with limitless infinite cardinalities which is likely to make the suggestion that $\mathcal{P}(\mathbb{N})$ might be a proper class seem absurd, when proper classes are finally introduced as massive collections mysteriously "in some perpetual, atemporal process of becoming" (Section 1.3).

What I am suggesting is that a more sensible educational sequence would begin with an understanding of there being two different kinds of collections, those described by explicitly listing their elements and those de-

scribed by a membership criterion, followed perhaps by a Davies machines type story to convey an intuition for the possibility of infinite explicit lists. Then it could be explained that, even allowing for infinite lists, the collection of all explicitly listable collections cannot itself be explicitly listable, showing that the two notions of "collection" do not coincide.

The next step for a Cantorian would be to affirm that, in principle, whenever it is possible to explicitly list all the elements of some collection, it will also be possible to explicitly list all of its subcollections. But now the burden is squarely on the Cantorian: why should this be true? On the face of it, $\mathcal{P}(\mathbb{N})$ does *not* seem explicitly listable. Recall the mathematics community's initial reaction of skepticism towards Zermelo's well-ordering theorem.

Thus, if the subject is learned in a different order, the Cantorian hypothesis may become quite dubious. For one struggles to imagine a world in which all the subsets of $\mathbb{N}$ can be explicitly listed. We don't have anything like Malament-Hogarth spacetimes or Davies machines to help us conceive of such a world. What would it be like to run a computation with one step corresponding to each set of natural numbers? How would that work?

It is hard to form a concrete picture of such a situation. Moreover, there is good reason to expect that *no* story we could tell could give us any real intuition for such a world. That is because any story you or I could tell would be finite in length. After being appropriately formalized, this story might have a model (if it were internally consistent), but by the Löwenheim-Skolem theorem it would then have a countable model. In other words, any finite description of an uncountable world will also describe some countable world. This argument is informal, of course, and it demands some clarification of what counts as a "description", but nonetheless it seems to me to strongly suggest that we literally *cannot*, in any concrete way, conceive of a world in which $\mathcal{P}(\mathbb{N})$, all of it, is a set. The world we are imagining might actually contain only a small part of $\mathcal{P}(\mathbb{N})$.

Incidentally, a similar argument can be deployed against Davies machines, and it is instructive to compare the two cases. No matter how well it is fleshed out, the Davies machines story must also have nonstandard models if it has any models. The infinite computation might go on too long, passing into infinitesimal Davies machines! But it seems to me that any such nonstandard model would have a standard part — just throw out all the infinitesimal stuff — and that standard part would by itself be a standard model. In other words, if there are nonstandard models then there should be a standard model, too. This goes back to the discussion we

had in Section 4.1 about not letting the existence of nonstandard models of PA fool us into thinking that we don't really know what the natural numbers are. The case of $\mathcal{P}(\mathbb{N})$ is different because here we would be looking for a supermodel with as many sets of numbers as possible, rather than a submodel with as few natural numbers as possible. I would still agree that we do understand what it means for a list to include all the subsets of $\mathbb{N}$; it means that there is no possible world which contains any subsets other than those in the list. But the point is that the existence of some countable model clearly does not help us to demonstrate the existence of a world containing such a list.

Of course, none of this shows that there could not be a possible world in which $\mathcal{P}(\mathbb{N})$ is a set, in some sense of "possible world". But perhaps it does show the difficulty of motivating a positive claim in this direction. In the next two sections I will discuss various attempts to do this.

7.4 Consistency versus soundness

Consistency and soundness. Recognizing arithmetical soundness. Do set theorists have an unexplained intuition for consistency? Consistency normal, soundness abnormal. Two forms of indispensability.

Once we understand the conceptual difference between sets and classes, if we want to go beyond finitism then we need to give a reason to think that $\mathbb{N}$ is a set. We cannot just say that $\mathbb{N}$ is a collection: in order for it to count as a set the natural numbers must be, in principle, explicitly listable. I tried to argue that it is in Section 7.2. Since our universe could be finite, getting past ultrafinitism to ordinary finitism already requires us to idealize beyond what we know can physically happen in our universe to what could in principle happen in some possible world. And the argument then is that it is not so hard to imagine possible worlds in which $\mathbb{N}$ is explicitly listable.

This takes us up to the world of countablism. Cantorians will wish to go further, to $\mathcal{P}(\mathbb{N})$ and beyond, and so the burden is now on them to convince us that there are possible worlds in which $\mathcal{P}(\mathbb{N})$ is explicitly listable. In the last section I speculated about why, psychologically, this move may seem unexceptionable. But that is wrong: it is a very substantial move and it demands a serious justification. I also pointed out that not only is it hard to imagine what it would be like to explicitly list all the subsets of $\mathbb{N}$, but the Löwenheim-Skolem theorem gives us good reason to think that such an activity cannot be concretely described, and thus is in a real sense literally

inconceivable. However, there are other ways to argue for the existence of possible worlds besides saying that they can be imagined or described. So we must now consider them too.

I will take ZFC as the primary representative of Cantorianism. As I mentioned in the last section, the most popular justification of ZFC, the iterative conception, has little to say about the listability of $\mathcal{P}(\mathbb{N})$. In explanations of the iterative conception it is typically just taken for granted that once a set is available, one should be able to somehow "gather together" all of its subsets. So there is no real argument here to be countered.

Sometimes people try to finesse the question of the correctness of Zermelo-Frankel set theory by noting that ZFC is most likely consistent, and declaring that this is all that matters. In other words, as long as ZFC, or even ZFC augmented by this or that large cardinal axiom, is consistent, no further justification is needed. We needn't try to show that there really are possible worlds containing uncountable lists, indeed it doesn't even matter whether such worlds exist. What matters is the axiomatic system itself, and if that system is consistent then there is nothing left to say and no need for a conceptual justification.

Of course, we do not know for a fact that ZFC really is consistent, and those of us who think it lacks a clear philosophical justification might well consider this an open question. But ZFC has been around long enough, and probed deeply enough, that I am willing to grant that it is most likely consistent.

But is this really all that matters? People who think this way perhaps have not considered that consistency does not entail soundness. ZFC could be arithmetically unsound, i.e., it could prove false sentences of first order number theory, yet still be consistent. For instance, ZFC might prove that some Turing machine halts even though it actually does not. I happen to think this is a real possibility, for reasons I will explain below. I take it we can agree that arithmetical soundness is something we care about.

Soundness is very different from consistency. We can easily probe ZFC for an inconsistency, simply by trying to derive a contradiction. The longer we go without succeeding, the greater our confidence should be in its consistency. But probing ZFC for arithmetical unsoundness is a different matter. Suppose there is a ZFC proof that some Turing machine halts even though it actually does not. How could we discover the truth of the matter, without running the machine forever to find out? It seems that we would need to prove in some other system that the machine really does not halt. (If we could also prove this in ZFC then ZFC would simply be inconsistent.) We

would then be pitting ZFC against some other system and debating which system we trusted more, and it seems unlikely that this question would have a definitive answer.

The bottom line is that if ZFC is unsound there might very well be no finite proof of this fact, and thus no way for us to know this. But if we care about soundness, we should not trust systems merely because they have not been shown to be inconstent, we should only trust systems that we can feel confident are sound because they have a clear philosophical justification.

No, the mere consistency of ZFC is not all we should care about! Soundness is what is needed, or at a minimum, I would think, arithmetical soundness. What reasons are there for thinking that ZFC is sound, or even just arithmetically sound?

One argument holds up the probable consistency of ZFC, as well as strengthenings of it by various large cardinal axioms, as indicators of their correctness. Here one acknowledges the difference between consistency and soundness but argues that the former is evidence for the latter. It is sometimes said that finitists and countablists have a problem explaining the success set theorists have had in formulating these axiomatic systems. How can they account for the fact that all these systems are consistent, or at any rate that no inconsistencies in them have been found, except by admitting that set theorists must have some genuine intuition for the actual nature of the universe of sets?

This is not a good argument. I mention it only because I have personally heard it made several times. The obvious response is that no inconsistencies have been discovered in the axiomatic systems around today mainly because those systems in which inconsistencies were discovered are no longer around. Let me remind the reader that the history of mathematical foundations is littered with inconsistent systems, starting with the very first, in Frege's *Grundgesetze*. Reinhardt cardinals are the best-known examples of large cardinals whose existence has been shown to be inconsistent with ZFC. I suppose we could expect that formal systems appearing in the literature are unlikely to have any really obvious inconsistencies, just because their authors ought to have thought about this question before publishing. But sometimes inconsistencies get through, and when they do, those systems are not going to be developed any further. We should not be impressed that the systems of greatest interest today are not known to be inconsistent — that is virtually tautological.

But still, haven't set theorists been, if not completely, at least largely successful in introducing systems which have withstood the test of time?

And is this not attributable to them having some special connection with the true universe of sets?

Again no, for two reasons. I mentioned the first a moment ago: systems which are obviously inconsistent would likely not have been introduced in the first place, so the pool of formal systems in play is already biased toward consistency. But going beyond that trivial observation, we must ask how remarkable consistency is, anyway?

Soundness is harder to achieve than consistency. Morally speaking, if we randomly chose n sentences of number theory as axioms, the chance that they would all be true is 2^{-n}. Otherwise this system would be unsound. In contrast, while I don't know how to calculate the chance that this system would be inconsistent (I suppose it would depend on how we interpret "randomly choosing"), I would expect that probability to be rather low. There has to be some special interaction among the axioms creating a contradiction. Specifically, the negation of one of the axioms would have to be a provable consequence of the others. So the general idea is that in some sense consistency is normal while soundness is abnormal.

Conclusion: the probable consistency of ZFC says little about its probable soundness.

I have taken the weakest argument first. Cantorians have other ways to argue that ZFC is sound. I just suggested that, broadly speaking, consistency is expected while soundness is unexpected. Indeed, since ZFC is not finitely axiomatizable, a naive estimate of its probable soundness is zero. But this kind of estimate only applies to axioms that are chosen "randomly", in some sense; if we specifically chose axioms that we thought were sound, then soundness obviously would be expected. Now, as we discussed in Section 1.2, the Zermelo-Frankel axioms were not originally chosen because they were thought to be sound, they were chosen on pragmatic grounds. And on the face of it they do mostly appear to be unrelated special cases of comprehension, a principle which is false in general, for reasons we discussed at length in earlier chapters.

On the other hand, even if the Zermelo-Frankel axioms appear to be a potpourri of unrelated instances of comprehension, the fact is that they were chosen for a reason, namely to accomodate normal mathematical practice. And concordance with normal mathematical practice, if true, seems like good evidence in favor of soundness (though not decisive evidence; let us not fetishize normal mathematical practice). But is it true?

This claim comes in two forms, a strong form and a weak form. According to the strong version, the mathematics used in experimental science

and engineering is based on set theory as formalized in ZFC. Consequently, every time an experimental prediction is verified, every time a bridge does not collapse, this is evidence for the correctness of ZFC. This is known as the "Quine-Putnam indispensability argument".[16]

Over the years, the indispensability argument has generated a good deal of discussion among philosophers of mathematics, but nowadays it seems to have fallen out of favor. At least I hope it has, because it should be obvious to everyone by now that it doesn't work. If one is out to justify some axioms on the basis of some applications, it might be a good idea to ask which of those axioms are actually needed for those applications. And thanks to the "reverse mathematics" school it is now widely recognized that virtually all, maybe literally all, of the mathematics that is needed in experimental science and engineering does not require any substantial set theory. It can be formalized in very weak systems, indeed in straightforwardly countablistic systems. Thus scientific and engineering successes provide no particular support for ZFC. I must emphasize this point. We are asking whether the collection of all subsets of the natural numbers is itself a set. The usefulness of mathematics which does not rely on this assumption is *totally irrelevant*. And seemingly all of the mathematics which has real, practical use does in fact fall into that category.

This statement sounds like an invitation to hunt for a counterexample. But a single counterexample would not carry much weight. We already know that *at least* almost all of the math used in physics and engineering is countablistic; barring some massive recalibration of what we know about the physical world, that ship has sailed. Given this, the most likely explanation for any isolated counterexample that one might find in the physics literature is that the uncountablistic aspect is not physically relevant. What is really needed is for the uncountablistic aspect of some analysis to have a clear physical meaning. But not only is this highly unlikely, even if it did happen it would not tell us anything much about ZFC. The relevant question at that point would merely be what additional axiom was actually needed to enable the given application — the kind of question that reverse mathematicians are good at answering.

There is a weaker, and less obviously wrong, form of the argument from concordance with mathematical practice. This argument cites mainstream mathematics in general, not just its scientifically applicable bits, in support of ZFC. The idea is that even if the uncountablistic aspect of ZFC has

[16] [29, 28].

no practical use, it is still indispensible for a lot of compelling mathematics, and this is supposed to confer some legitimacy. The problem here is that, even disregarding scientific applicability, it is still the case that the overwhelming bulk of mainstream mathematics is countablistic. Uncountablistic mathematics tends to be rather peripheral. Obviously I am going to have to back up such a provocative statement, and I will do that in the next section (and at much greater length in Part 2 of this book). It is important here because ordinary mathematics is constrained only by the interests of ordinary mathematicians, so if little of it is uncountablistic, this suggests that the extra territory provided by ZFC is not so interesting. And the reason it is not interesting, as I will discuss in a moment, is because the principal extra feature we get in ZFC that we do not get in countablistic systems is *a raft of set-theoretic pathology*. This seems to me to be meaningful evidence against the idea that ZFC is sound. From the point of view of the ordinary working mathematician, the uncountablistic part of the world envisioned in ZFC is quite pathological, and together with the overall rarity of soundness (an argument I made earlier) and the absence of a clear philosophical justification, I think this gives us a legitimate reason to suspect that ZFC could be arithmetically unsound.

This is a bold claim, and the reader will rightly demand that I explain myself.

7.5 Normal mathematics

What makes a space countablistic. Countablism and formalism. Most mainstream mathematical subjects are almost completely countablistic. Set-theoretic topology and independence phenomena. Sequences of dual Banach spaces. Multiplicity theory.

The vast majority of ordinary, everyday mathematics is essentially countablistic in nature and can be formalized in purely number-theoretic systems. This has been established in detail by the reverse mathematics school, but there is actually a long history of such work, going all the way back to Weyl's *Das Kontinuum*.

I will explain a little about how this works in Section 7.7, and then go into detail in Part 2. For now, let me just briefly say that in countablistic mathematics things like $\mathbb{N}$ and $\mathbb{Q}$ are sets while things like $\mathcal{P}(\mathbb{N})$ and $\mathbb{R}$ are proper classes. In general, the nature of finite and countable structures doesn't change under countablism, while things like finite dimensional man-

ifolds and separable Banach spaces become proper classes, but this does not stop us from being able to work with them. The fundamental informal principle is that a classical set is countablistically realized as a class if every point in it can be specified by a countable amount of information, which in practice generally means that classical sets of cardinality at most $2^{\aleph_0}$ are countablistic. So this includes virtually all objects of interest in mainstream mathematics. (Whether we can prove the same theorems about these objects is a separate question. That must be answered on a case-by-case basis; we will see in Part 2 that the answer is "pretty much all of the main theorems, although sometimes their statements change a little".)

There is a third category of objects, too; classical proper classes (the class of all finite dimensional manifolds, the class of all Banach spaces) are typically no longer classes in countablistic mathematics. They become concepts, which, according to my definitions in Chapter 5, are essentially just a less well-behaved version of classes.

We have no clear standards for judging exactly how much uncountablistic mathematics should count as "ordinary" or "mainstream". But these terms are meaningful and can still be talked about, even if they are not sharply defined. For instance, how many Fields Medals have been awarded for work that was inherently uncountablistic? I don't see any. That would seem to rather strongly support the idea that uncountablistic mathematics is somehow peripheral. Maybe someone who looks through the list of awards more carefully can find something.

Incidentally, the one Fields Medal in pure set theory, given to Paul Cohen for his proof of the independence of the continuum hypothesis from ZFC, was awarded for work which was purely countablistic: Cohen's relative consistency results are provable in Peano arithmetic (and in even weaker systems).[17] And this illustrates an important point: countablism does not seek to bar us from thinking about uncountablistic formal systems. Rather, it urges us to be formalists about such systems. You're still "allowed" to prove theorems in ZFC, or in any other formal system you like.

Uncountablistic systems such as ZFC may very well be consistent, and if they are consistent then they have models. But from the countablistic point of view, there is no compelling reason to think there should be one favored canonical model of ZFC. Thus, while a theorem proven in Peano arithmetic can be understood as telling us something about the natural numbers, a theorem proven in ZFC should be understood only as telling us

[17]This is the approach taken in [53], for instance.

something about what will be the case in models of ZFC.

With that in mind, let me return to my provocative suggestion that uncountablistic mathematics tends to be "pathological" by first looking at specific areas of mathematics which are, in their current state, almost wholly countablistic or almost wholly uncountablistic.

There are plenty of areas of mathematics that are almost completely countablistic: algebraic topology, number theory, differential geometry, PDEs, etc., etc. All uncontroversially mainstream. People working in any of these fields will employ uncountablistic reasoning rarely, if ever. (How often does one see uncountably long constructions, or objects whose cardinality is greater than $2^{\aleph_0}$, in any of these areas, for instance?) And there are also areas that, aside from a countablistic core that was essentially settled long ago, are now almost completely uncountablistic. I will cite set-theoretic topology as an example of such an area. Is it mainstream? We can debate this, but it is more to the point to ask a different question, namely, whether our confidence in the canonical nature of set-theoretic topology is such that we can see it as providing evidence for the canonical nature of ZFC. Is there any difference between how we feel about being formalists about, say, algebraic topology, versus being formalists about set-theoretic topology?

The relevance of this question is that countablism affirms that algebraic topology tells us true facts about standard mathematical objects. So according to countablism we should *not* be formalists about algebraic topology. In contrast, as I just explained, countablism tells us that we should be formalists about set-theoretic topology.

I am inclined to think that most mathematicians do not want to be formalists about algebraic topology but might be sympathetic to the idea of being formalists about set-theoretic topology. Is it really the case that every continuous function mapping the closed unit disc to itself has a fixed point? Yes, this statement means something and is true. We can see why it is the case, we can feel it in our bones. And then: is it really the case that every element of the Stone-Čech remainder $\beta\mathbb{N}^* = \beta\mathbb{N} \setminus \mathbb{N}$ has a countable family of neighborhoods whose intersection is not a neighborhood? Given the ineffable nature of $\beta\mathbb{N}$,[18] it is easier to imagine that there is no underlying fact of the matter here — especially when one learns that the answer is independent of ZFC (assuming ZFC is consistent, of course). In some models of ZFC it is true, in others it is false. That does not mean that there

[18]Gert Pedersen: "No one has ever seen a nonprincipal ultrafilter."

could not be a canonical "true" model in which the question has a definite answer, but the real issue here is whether we have sufficient confidence in the determinacy of this question to cite it as evidence for believing that the Zermelo-Frankel axioms truly describe a canonical universe of sets. It seems to me that exactly the opposite is the case: whatever confidence we have that the question is determinate comes from already believing in a canonical universe of sets which admits things like $\mathcal{P}(\mathcal{P}(\mathbb{N}))$.

Remember that we are asking whether the concordance of ZFC with ordinary mathematics is evidence for the correctness of ZFC. My point here is that not only does the cogency of countablistic mathematics say nothing about the correctness of the uncountablistic aspect of ZFC, neither does the cogency of uncountablistic mathematics, if that cogency only becomes convincing after one has accepted $\mathcal{P}(\mathbb{N})$ as a set. And $\beta\mathbb{N}^*$, like other uncountablistic spaces, does fall into this category, in my opinion.

I must emphasize the ubiquity of independence results outside the world of countablism. The question about neighborhood intersections in $\beta\mathbb{N}^*$ mentioned above is hardly an isolated example. In fact, I don't think it's much of an exaggeration to say that most nontrivial properties of $\beta\mathbb{N}^*$ are independent of ZFC. Is every first countable compact space a continuous image of $\beta\mathbb{N}^*$? Is every self-homeomorphism of $\beta\mathbb{N}^*$ implemented by a bijection between two cofinite subsets of $\mathbb{N}$? And so on. In some models of ZFC yes, in others no. Nor is this phenomenon special to $\beta\mathbb{N}^*$; to the contrary, it is broadly typical of uncountablistic mathematics. Independence results are relatively rare in the countablistic realm, but common in the uncountablistic realm.

The contrast is especially striking because under constructive countablism Gödel's axiom $V = L$ becomes natural, and this axiom settles practically all of the natural questions which are classically independent (Naimark's problem,[19] Whitehead's problem, the existence of outer automorphisms of the Calkin algebra, etc., etc.).[20]

This does not rule out the possibility that some other compelling new axiom could be added to ZFC which would settle the same problems. But

[19] Actually, it has not been shown that Naimark's problem is independent, only that the existence of a counterexample follows from $V = L$ [1]. But it is most likely independent.

[20] $V = L$ settles all these problems classically too, but $V = L$ is classically unacceptable because classically it is hard to deny that there does exist a set of all the sentences in the language of set theory that are true in L, and the existence of this set (known as "$0^{\#}$") contradicts $V = L$. Whereas in constructive countablism some sentences may be neither definitely true nor definitely false in L, so there is no reason to assume that $0^{\#}$ exists as a set.

that doesn't seem likely. Indeed, Hamkins has made a strong argument that identifying such an axiom at this point is practically out of the question, due to the fact that set theorists have so much experience with models of ZFC in which, for example, either the continuum hypothesis or its negation holds. Because "we have an informed, deep understanding of the CH and ¬CH worlds and of how to build them from each other ... if you were to present a principle Φ and prove that it implies ¬CH, say, then we can no longer see Φ as obviously true, since to do so would negate our experiences in the set-theoretic worlds having CH."[21] This argument also suggests that the actual experience of set theorists fits better with the idea of a family of countable models of ZFC, none having any special ontological supremacy, than with the idea of a canonical universe of sets; see Section 7.6.

We can also gauge the ease and naturalness of uncountablistic mathematics by looking at subjects that have both countablistic and uncountablistic regimes. Perhaps the best example of such a subject is functional analysis — broadly speaking, the study of infinite dimensional objects with both algebraic and metric/topological structure. Consider Banach spaces. In this context separable spaces and their duals are the typical examples of countablistic objects. There is a notable phenomenon here. If we start with a countablistic Banach space E, then either all the spaces in the sequence of duals E^*, E^{**}, ... will be countablistic (typically because $E \cong E^{**}$), or else at some point we will pass from countablistic to uncountablistic. What is striking is that if the initial space is one of the usual well-known spaces belonging to the standard repertoire, typically its dual will also be well-known, and so will its second dual, and so on ... until we get to the first uncountablistic space in the sequence. That one will be an obscure, pathological space that is not part of the standard repertoire. If we take "has its own special name" as a stand-in for belonging to the standard repertoire, then the situation is that the first uncountablistic space in the sequence will also be the first one that doesn't have a special name. This illustrates in a rather dramatic way the standard nature of countablistic spaces and the exotic nature of uncountablistic ones.

For example, take $E = C[0,1]$, the space of continuous functions on the unit interval. It is countablistic, as is its dual, the space $M[0,1]$ of Borel measures on $[0,1]$, but its second dual is not. And its second dual is an exotic, pathological Banach space that does not have any standard special name other than "the dual of $M[0,1]$" or "the second dual of $C[0,1]$". Or

[21] [15], p. 431.

take $E = K(H)$, the space of compact operators on a separable Hilbert space H. It is countablistic; so is its dual, the space of trace-class operators $TC(H)$, and its second dual, the space of bounded operators $B(H)$. But its third dual is uncountablistic, and this third dual is an exotic, pathological Banach space that has no special name other than "the dual of $B(H)$". Or if E is the Hardy space H^1, then its dual is the space BMO of functions of bounded mean oscillation, and both of these spaces are countablistic. But the dual of BMO is uncountablistic, and it is an exotic, pathological Banach space that has no special name. And so on. The first uncountablistic space in the sequence is always also the first space which is not part of the standard repertoire.[22]

In all of these cases we have separable Banach spaces up to the final countablistic space, which is nonseparable yet still countablistic by virtue of being the dual of a separable space. We pass to an uncountablistic space in the next step, when we take the dual of this nonseparable space.

Recall the informal criterion I mentioned above: a space is countablistically a class if every point in it can be identified by a countable amount of information. Clearly this is true of any separable Banach space, as any point in the space is the limit of a sequence of elements of some countable dense subset. (That is, unless no countable dense subset can be specified with a countable amount of information, but it is hard to think of any reasonable example of that). It is also true of the dual of any separable space, modulo the same caveat, because we can characterize any element of the dual by the values it takes on a countable dense subset of the original space. But it is usually not true of the duals of nonseparable spaces: we usually need an uncountable amount of information to specify a bounded linear functional on a nonseparable space (though not always).

Another example of nonseparable pathology in functional analysis is seen in multiplicity theory for self-adjoint operators on Hilbert space. Usually one simply assumes separability, but there does actually exist a theory of multiplicity for operators on nonseparable Hilbert spaces — one which I consider to be pretty pathological. But you can look it up and decide for yourself. As a C*-algebraist, I could go on to list many other examples of

[22] $E = c_0$ is arguably a counterexample, but the case for this is not very strong. Here c_0, its dual l^1, and its second dual l^∞ are all countablistic, while the dual of l^∞ is uncountablistic; yet the dual of l^∞ actually does have a special name — we can identify it with the space $M(\beta\mathbb{N})$ of regular Borel measures on $\beta\mathbb{N}$. Yes, I guess this space has a special name, but I don't agree that it belongs to the standard repertoire. I certainly think it's fair to say that $M(\beta\mathbb{N})$ is exotic and pathological.

nonseparable pathology in my subject, but perhaps I have made my point.

May I now recall for the reader my earlier comments about the strikingly exact fit of countablism with normal mathematics, and the elimination of "a raft of nonseparable pathology"? As well as the view I expressed back in Section 1.2 about the value of basing philosophical foundations on actual mathematical practice. Well, one hundred years on from Zermelo's axiomatization, we now have a much clearer idea of how much set theory is needed for normal mathematics, and the answer is, with remarkable precision, countablistic set theory.

Speaking from the perspective of someone who works in a field that straddles the countablistic and uncountablistic worlds, I think it is fair to call the uncountablistic part "pathological". I accept that others may disagree, but I will stand on my position that the uncountablistic mathematics supported by ZFC does not provide strong evidence for the correctness of ZFC.

7.6 Skolem and the multiverse view

Skolem and "absolute" uncountability. Hamkins' multiverse view. The multiverse view entails that ZFC is not a well-defined system.

In the last section I touched on the idea that if ZFC is consistent, then it must have a variety of countable models, but for a countablist that is all there is, a family of countable models. This is not a new suggestion: it goes back to Skolem, who called it a "relativity of set-theoretic notions",[23] implying that set-theoretic assertions only have meaning relative to a choice of (countable) model. More recently, Hamkins has generated substantial interest in his own version of this idea, which he calls "the multiverse view".[24] I would now like to compare my position with theirs.

Actually, I am not sure there is any real difference between Skolem's position, as I understand it, and mine. What needs explaining, perhaps, is his citation of the existence of countable models of ZFC as somehow implying the unreality of "absolutely" uncountable sets (as opposed to sets which only appear to be uncountable within some model, but are revealed to be countable when one steps outside the model). On its face, this inference seems like a complete non-sequitur, and casual commentators have generally regarded it as such. Fortunately for us, Jané has reconstructed

[23] [42], p. 296
[24] [15].

from Skolem's extant writings what his reasoning must have been:

> Skolem raises doubts about the existence of uncountable collections because he has been given no argument for their existence — either by means of a formalism or otherwise. On the other hand, he has been offered axiomatized set theory as the only reliable theory of sets, thus the one theory that could dispel his doubts, and he just makes it clear that, unaided, such a theory is unable to account for uncountable totalities.[25]

Skolem evidently took the legitimacy of countable infinity to be intuitively obvious. In contrast, the legitimacy of uncountable infinity is not obvious at all (as we can see in the mathematical community's initial reaction of shock when Cantor unveiled his diagonal argument). We do not simply intuit uncountability; we have to be brought to it by considering things like the power set of $\mathbb{N}$ or the set of all countable ordinals.

But are there even such sets? Naive set theory assures us that there are, but then, it also assures us that there is a set of all ordinals and a set of all sets that are not elements of themselves, so we might not want to trust it on this point.

In a word, Skolem's argument now is that Zermelo's axioms cannot show us that $\mathcal{P}(\mathbb{N})$ and $\aleph_1$ exist either. It says they exist, but for all we know it could be telling us about what happens in some countable model (which, until we have determined that uncountable sets exist, is all we have). In that case, the allegedly uncountable sets present in the model would be recognized as countable from an external perspective. Skolem:

> With a suitable axiomatic basis, therefore, the theorems of set theory can be made to hold in a merely *verbal* sense, on the assumption, of course, that the axiomatization is consistent; but this rests merely upon the fact that the use of the *word* "set" has been regulated in a suitable way.

Indeed, no formal system available to us can possibly establish the existence of genuinely uncountable sets, as opposed to sets that only appear to be countable within some restricted model:

> In order to obtain something absolutely nondenumerable, we would have to have either an absolutely nondenumerably infinite number of axioms or an axiom that would yield an absolutely nondenumerable number of first-order propositions. But this would in all cases lead to a circular introduction of the higher infinities; that is, *on an axiomatic basis higher infinities exist only in a relative sense.*[26]

[25] [18], p. 144.
[26] [42], p. 296.

The only formal systems which could do this must have uncountability already baked in (cf. Section 1.3).

It seems to me that as long as ZFC was merely a pragmatically constructed formal system with no conceptual justification (Section 1.2), Skolem's inference from "if ZFC has any model then it has a countable model" to "ZFC gives us no justification for the existence of absolutely uncountable sets" was exactly on point. So I find myself in complete agreement with Skolem's countablism.

As for Hamkins, his views and mine are certainly similar in important ways, but there are also serious divergences. As his "multiverse view", only introduced relatively recently, has already gained significant traction (no doubt aided by Hamkins' deep knowledge of set theory and the excellence of his exposition), maybe it is worthwhile for me to indicate what I think are the important ways in which it differs from my countablism.

Hamkins summarizes his position as follows:

> I shall argue for a contrary position, the *multiverse view*, which holds that there are diverse distinct concepts of set, each instantiated in a corresponding set-theoretic universe, which exhibit diverse set-theoretic truths. Each such universe exists independently in the same Platonic sense that proponents of the universe view regard their universe to exist. Many of these universes have been already named and intensely studied in set theory ... By adopting a particular concept of set, we in effect adopt that universe as our current mathematical universe; we jump inside and explore the nature of set theory offered by that universe.[27]

We know that if ZFC is consistent then there is a diverse array of countable models of it. And Hamkins and I are in agreement that these are the only models: "Every universe V is countable from the perspective of another, better universe W."[28] He also affirms, as I do, that none of these models has any special ontological supremacy over the others. These are the main similarities between Hamkins' position and mine (which I take to be essentially the same as Skolem's). Where we start to see a difference is in his idea that these models correspond to "distinct concepts of set". On my view, there is only one set concept: a set is a listable collection.

This may seem like it could be merely a disagreement over terminology, but it is more than that. ZFC, if consistent, has a variety of countable models, but I do not think that is because the set concept itself is somehow multifarious. To the contrary, *none* of these models can be the true universe

[27] [15], p. 417.
[28] [15], p. 440.

of sets, for two reasons: (1) because any countable model of ZFC is itself a set, and there is no set of all sets, and (2) because ZFC proves falsehoods, such as that the proper class $\mathcal{P}(\mathbb{N})$ is a set, which will be reflected in any model. In contrast, Hamkins thinks that *all* models of ZFC are valid, maybe just by definition; it sounds as though for him the word "set" is synonymous with "something in a model of ZFC". He goes on to say

> ... the multiverse view does not undermine the claim that set theory serves as an ontological foundation for mathematics, since one expects to find all the familiar classical mathematical objects and structures inside any one of the universes in the multiverse ...[29]

Whereas I do not think ZFC should serve as an ontological foundation for mathematics. I can quote Skolem on my side on this point:

> ... the initial foundations [should] be something immediately clear, natural, and not open to question. This condition is satisfied by the notion of integer and by inductive inferences, but it is decidedly not satisfied by set-theoretic axioms of the type of Zermelo's or anything else of that kind; if we were to accept the reduction of the former notions to the latter, the set-theoretic notions would have to be simpler than mathematical induction, and reasoning with them less open to question, but this runs entirely counter to the actual state of affairs.[30]

This matters because some models of ZFC, possibly all of them, contain nonstandard versions of $\mathbb{N}$, and Hamkins seems to believe that these different versions of $\mathbb{N}$ are all equally valid: what the natural numbers are is something that can only be decided within a model of ZFC, and hence has no absolute answer. ("Although we conventionally describe the natural numbers as 1, 2, 3, ..., and so on, why are we so confident that this ellipsis is meaningful as an absolute characterization?"[31] "Even though we prove in ZFC that $\mathbb{N}$ is unique, the situation is that not all models of ZFC have the same $\mathbb{N}$ or even the same arithmetic truths."[32])

I will recall what I said in Section 4.1, that our understanding of $\mathbb{N}$ precedes any attempt to axiomatize it, and we should not let the existence of nonstandard models of PA (or ZFC) make us doubt that the concept of a natural number has a clear, absolute meaning. As I noted there, if $\mathbb{N}$ really could be a model of either PA + Con(PA) or PA + $\neg$Con(PA) then there would be no absolute fact of the matter as to whether PA is

[29] [15], p. 417
[30] [42], p. 299.
[31] [15], p. 428.
[32] [15], p. 429.

consistent. The same could be said of ZFC: if ZFC is consistent then so is ZFC + ¬Con(ZFC), and the version of ℕ in any model of this theory would make ZFC inconsistent.

As I understand Hamkins' position, arithmetical truths are not determinate; they can vary between models, and all are valid within their model but none is universally true besides those that are provable in ZFC. This means that ¬Con(ZFC) is a potentially valid arithmetical truth and thus Con(ZFC) cannot be universally true. People who live in a model of ZFC + ¬Con(ZFC) would think that there are *no* models of ZFC, and they would apparently be just as right as people living in other models of ZFC who think that there do exist models of ZFC. This is confusing for me.

It gets worse. In models of ZFC with nonstandard natural numbers, our arithmetical encoding of the language of set theory will include formulas of nonstandard length, and thus the separation and replacement schemes of ZFC will include nonstandard axioms. *Inhabitants of different universes cannot even agree on what the axioms of ZFC are.* This means that, according to the multiverse view, the assertion that something is a model of ZFC has no absolute meaning. In this way it seems to me that the multiverse view undermines its own conceptual framework.

7.7 Countablistic mathematics

ℕ is a set. Peano arithmetic is valid. ℝ is a proper class. The least upper bound principle. Topology. Continuous functions.

Let me now lay the groundwork for constructively countablistic mathematics. First, I claim that natural numbers are definite objects. Since character strings are definite objects (Section 2.3), we can ensure this by taking as proxies for the natural numbers all finite strings of decimal digits, for example. (If you prefer some other system of definite proxies, be my guest. It doesn't matter because when we abstract from proxies we discard all of their inessential features.)

According to countablism ℕ is explicitly listable, as $\{1, 2, \ldots\}$. Since natural numbers are definite objects, this means that ℕ is a set. And therefore we are entitled, even as constructivists, to affirm that any assertion that only quantifies over ℕ has a definite truth value. So we have the law of excluded middle for arithmetical statements. In the next chapter we will see that countablists can affirm induction on ℕ for arbitrary meaningful properties, and therefore Peano arithmetic, with classical logic, is valid for

them.

It is also a feature of countablism that there are no uncountable sets. This means that $\mathbb{R}$ is *not* a set. But is it a class? We must check that the predicate "is a real number" is definite. We can start by noting that $\mathbb{Q}$ is a set, for much the same reason that $\mathbb{N}$ is. And then the predicate "is a set of rational numbers" is definite because, given any syntactic list, it either is or is not the case that any given entry in the list names a rational number (rational numbers are definite objects), and quantifying over the set of entries of the list, it definitely either is or is not the case that every entry names a rational number. Modelling real numbers as Dedekind cuts, more countable quantification shows that it definitely either is or is not the case that any given set of rational numbers is a cut. That is, the set is neither $\emptyset$ nor $\mathbb{Q}$, it has no largest element, and whenever it contains some rational number it contains all smaller rational numbers.

Taking Dedekind cuts as proxies for real numbers, we conclude that "is a real number" is definite. So $\mathbb{R}$ is a class. But it is a proper class, so as truth constructivists we cannot affirm excluded middle for arbitrary assertions that quantify over $\mathbb{R}$. One could say that when we do number theory we act like platonists, but when we pass to real analysis we act like intuitionists.

Since any (classical) set of cardinality at most $2^{\aleph_0}$ can be embedded in $\mathbb{R}$, the fact that $\mathbb{R}$ is a class entails that in countablistic mathematics such sets generally become classes. (We do need the definitions of both the set and of the embedding to be countablistically valid, and we need every real number to definitely belong or not belong to the embedded set.) Classical sets and classes larger than this typically can no longer be considered classes, merely concepts.

The least upper bound principle provides a good illustration of the way mathematical facts change under constructive countablism. Classically, any bounded set of real numbers has a least upper bound, and if real numbers are modelled as Dedekind cuts this least upper bound is simply a union. Constructive countablism tells us the same thing: any bounded set of real numbers (= Dedekind cuts) has a least upper bound, realized as their union. We can do this because a countable union of sets is also a set.

The difference is that the real line is classically itself a set, so the restriction to subsets of $\mathbb{R}$ is no restriction at all. But under countablism $\mathbb{R}$ is a proper class, and we only get the principle for bounded *subsets* of $\mathbb{R}$, not bounded subclasses. Constructive countablists would find an obstacle in defining a real number by means of a condition that quantifies over an

arbitrary family of real numbers (a rational number x belongs to the union of cuts $\bigcup A_\alpha$, i.e., is an element of the least upper bound, if there exists α such that $x \in A_\alpha$). If the condition "there exists α such that $x \in A_\alpha$" quantifies over a proper class of indices then it cannot be assumed to have a definite truth value. Thus the condition that x belong to $\bigcup A_\alpha$ need not be definite, and hence need not define a set.

On the other hand, since $\mathbb{R}$ is separable, for most purposes an ability to take the least upper bound of a countable set of real numbers suffices. In fact, classically every subset of $\mathbb{R}$ is separable, so any bounded subset itself contains a countable subset with the same least upper bound.

Open subclasses of $\mathbb{R}$ under constructive countablism are exactly the same as classical open subsets. A subclass I of $\mathbb{R}$ is a *bounded open interval* if there exist real numbers x and z such that $y \in \mathbb{R}$ belongs to I if and only if $x < y < z$. A subclass U of $\mathbb{R}$ is *open* if there is a pair of sequences (x_n) and (z_n) such that $y \in \mathbb{R}$ belongs to U if and only if $x_n < y < z_n$ for some n. Since we do not require these intervals to be disjoint, there is no loss of generality if we require every x_n and z_n to be rational.

(For every such pair of sequences (x_n) and (z_n), there does exist a corresponding open class, because the condition "$x_n < y < z_n$ for some n" is definite — it only quantifies over a set.)

A *topology* on a class X is an indexed family of subclasses that is stable under countable unions and finite intersections and includes $\emptyset$ and X. Here an "indexed family" means a class of ordered pairs $\mathcal{T} \subseteq X \times I$ for some class I. For each index $\lambda \in I$, we define U_λ to be the class of all $x \in X$ such that $(x, \lambda) \in \mathcal{T}$. Stability under countable unions means that for every sequence (λ_n) in I there exists $\lambda \in I$ such that $x \in U_\lambda$ if and only if $x \in U_{\lambda_n}$ for some n, and so on. The standard topology on $\mathbb{R}$ could be defined by letting I consist of all sequences $\lambda = (x_n, z_n)$ of pairs of rational numbers with $x_n < z_n$ for all n, with U_λ consisting of those $y \in \mathbb{R}$ such that $x_n < y < z_n$ for some n. (Plus one sequence, say with $x_n = z_n = 0$ for all n, to index the empty class.)

Our condition that a topology must satisfy with respect to unions is, informally, that the union of any set of open sets is an open set. This is the same as the classical definition, except that in the classical setting there can be uncountable sets of open sets and in our setting there cannot. It is similar to the way the least upper bound principle changes, and has similarly little effect on non-set-theoretic math.

A *function* $f : X \to Y$ is a subclass of $X \times Y$ satisfying the usual conditions. It is *continuous* if the inverse image of any open subclass of the

range is an open subclass of the domain.

(For any function $f : X \to Y$ and any subclass $Y_0 \subseteq Y$, the inverse image $f^{-1}(Y_0)$ does exist as a subclass of X. That is because for each $x \in X$ there is exactly one $y \in Y$ such that $f(x) = y$, and definiteness of Y_0 means that this y definitely does or does not belong to Y_0. Thus the condition "$x \in f^{-1}(Y_0)$" is definite.)

That extremely brief account should at least give some idea how countablists can do mathematics.

Chapter 8

Well-ordering

8.1 Induction on $\mathbb{N}$

Induction for countablists. Excluded middle not essential. Excluded middle and well-ordering.

Induction on the natural numbers is a — some might even say *the* — fundamental principle of mathematics. Schematically, for any predicate ϕ that meaningfully applies to all the natural numbers, this is the statement

$$\ulcorner \phi(1) \text{ and } (\forall n)(\phi(n) \to \phi(n+1)) \text{ imply } (\forall n)\phi(n)\urcorner.$$

What is the justification? For each value of n, we can see how to prove $\phi(\hat{n})$: we are given $\phi(1)$, then we use the second part of the premise to successively infer $\phi(2)$, $\phi(3)$, ... Thus, if $\phi(\hat{n})$ is meaningful for all n, then we can see how to prove $\phi(\hat{n})$ for any value of n. Writing $\ulcorner \text{Ind}(\phi)\urcorner$ for $\ulcorner \phi(1)$ and $(\forall n)(\phi(n) \to \phi(n+1))\urcorner$ ("ϕ is inductive"), we therefore have

$$\mathbb{M}[\ulcorner(\forall n)\phi(n)\urcorner] \to (\forall n)\mathbb{A}[\ulcorner \text{Ind}(\phi) \to \phi(\hat{n})\urcorner]$$

for any predicate ϕ.

Countablists can bring the quantifier inside using the $\forall$-capture scheme mentioned in Section 4.5, to get

$$(\forall \phi : \mathbb{M}[\ulcorner(\forall n)\phi(n)\urcorner])\mathbb{A}[\ulcorner \text{Ind}(\phi) \to (\forall n)\phi(n)\urcorner].$$

This is the principle of induction on $\mathbb{N}$ in its most general form. For any particular predicate ϕ that is known to be meaningful on all natural numbers, we can apply the release rule to get simply $\ulcorner \text{Ind}(\phi) \to (\forall n)\phi(n)\urcorner$.

The above argument uses two special principles, $\forall$-capture and the release rule. Again, the justification for $\forall$-capture here is that, as countablists, we can amalgamate the separate proofs of $\phi(1)$, $\phi(2)$, ... into a single, infinitely long proof of $\phi(\hat{n})$ for all n. The release rule is legitimate because

for any ϕ that is recognized as meaningful, we have an actual (infinitely long) proof of $\ulcorner \text{Ind}(\phi) \to (\forall n)\phi(n)\urcorner$, so we can affirm this implication itself and not its mere assertibility.

In particular, as countablists we can affirm induction for any predicate expressible in first order arithmetic. If ϕ has other free variables besides "n", we can make another countable amalgamation argument (over each choice of values of the other free variables) to get induction for ϕ. So countablists can affirm the full induction scheme of Peano arithmetic.

Notice that when justifying the principle of induction we did not have to assume at any point that ϕ verifies the law of excluded middle. This is important because under truth constructivism we generally will not be able to affirm that a given predicate definitely either does or does not hold of every natural number. Specifically, this becomes an issue for constructive countablists when ϕ involves quantification over things like $\mathbb{R}$ or $\mathcal{P}(\mathbb{N})$, as according to countablism these are proper classes, and according to truth constructivism, statements which quantify over a proper class might not have definite truth values.

But this is not a problem for us here. When we make the induction argument it turns out that, in fact, ϕ definitely does hold of every number. However, this definiteness is a consequence of the argument and did not have to be assumed separately — only the definite truth of $\phi(1)$ was initially assumed. Constructive countablists can affirm the principle of induction on the natural numbers for any meaningful predicate, definite or not.

It does become an issue when one turns to the related concept of well-ordering, though. The well-ordering property of $\mathbb{N}$, in its usual form, states that every nonempty subset A of $\mathbb{N}$ has a smallest element. If we take $\phi(n)$ to be the predicate "$k \notin \hat{A}$ for all $k < n$", then the well-ordering property for A is easily seen to both imply and be implied by induction for ϕ. A has no smallest element if and only if $(\forall n)(\phi(n) \to \phi(n+1))$ is true, and A is empty if and only if $(\forall n)\phi(n)$ is true; since $\phi(1)$ is vacuously true, induction for ϕ therefore says that if A has no smallest element then it is empty.

Let me pause to note that replacing "$n \notin A$" with "$k \notin A$ for all $k < n$" in effect converts ordinary induction into strong induction, i.e., the principle that

$$\ulcorner (\forall n)\left((\forall k < n)\phi(k) \to \phi(n)\right) \to (\forall n)\phi(n)\urcorner$$

(given that whenever ϕ holds for all numbers less than n it also holds for n, it follows that ϕ holds for all numbers). This version of induction generalizes better to other orderings of $\mathbb{N}$, as we will see in the next section.

Careful inspection of the argument given above for the equivalence between well-ordering and a special case of induction shows that it does use excluded middle. That is okay, because according to our conception of sets, for any n it definitely is or is not the case that n belongs to A (cf. Section 2.7). Even statements like "A has a least element" have definite truth values, because there is no quantification over a proper class. But it becomes a problem when we try to generalize the well-ordering property to predicates. For an arbitrary predicate ϕ we may well have trouble affirming that if ϕ holds of at least one number, then there is a smallest number of which it holds. For instance, suppose we could affirm $\phi(2)$ but did not see how to affirm $\phi(1) \vee \neg\phi(1)$: could we then be sure there is a definite value of n such that ϕ holds of n but not of any smaller number? Maybe not.

8.2 Well-ordering and induction

Progressivity for predicates and sets. Setwise well-ordering. Can we infer predicate induction from set induction?

Let us now consider induction on an arbitrary set X equipped with a total ordering $\preceq$. Here we must deal with the possibility that there could be elements of X with predecessors but no immediate predecessor, so that the hypothesis $\ulcorner\phi(n) \to \phi(n+1)\urcorner$ is no longer appropriate. We must now assume a version the strong induction hypothesis $\ulcorner(\forall k < n)\phi(k) \to \phi(n)\urcorner$.

Thus, the premises for our induction arguments will have the form

$$\ulcorner(\forall x)((\forall y \prec x)\phi(y) \to \phi(x))\urcorner$$

— for any $x \in X$, if ϕ holds of all $y \prec x$ then ϕ holds of x. We express this condition by saying that ϕ is *progressive*, or *progressive with respect to* $\preceq$ if we want to make the ordering explicit, and we write $\mathrm{Prog}\,\phi$ or $\mathrm{Prog}_{\preceq}\,\phi$.

Similarly, say that a subset A of X is *progressive*, and write $\mathrm{Prog}\,A$ or $\mathrm{Prog}_{\preceq}\,A$, to mean that for every $x \in X$, if every element preceding x belongs to A then so does x. That is, A is progressive if

$$(\forall x)((\forall y \prec x)(y \in A) \to x \in A).$$

This is the special case of predicate progressivity when $\phi(x)$ is the predicate "$x \in \hat{A}$".

Say that $\preceq$ is a *setwise* well-ordering if we can make induction arguments for subsets, i.e., if for every subset A of X, progressivity of A implies that $A = X$. So $\preceq$ is a setwise well-ordering if

$$\mathrm{Prog}_{\preceq}\,A \to A = X$$

for all $A \subseteq X$. This condition is equivalent to the familiar definition of well-ordering which states that every nonempty subset has a smallest element. For it is easy to see that $B \subseteq X$ lacks a smallest element if and only if $A = \{y \in X : y \prec x \text{ for all } x \in B\}$ is progressive; thus the existence of a nonempty subset with no smallest element is equivalent to the existence of a progressive subset that is not all of X. (And since X is a set, every subset definitely is or is not all of X.)

As with $\mathbb{N}$, this setwise well-ordering property is equivalent to the inductive condition

$$\ulcorner \mathrm{Prog}\, \phi \to (\forall x \in X)\phi(x) \urcorner$$

for $\phi(x) =$ "$x \in \hat{A}$". But of course we would like to have this condition for any predicate that meaningfully applies to all the elements of X. Does predicate induction follow from set induction?

No, this is not clear. The usual derivation of predicate induction from set induction goes like this: given any progressive predicate ϕ, the set $A = \{x \in X : \phi(x)\}$ is progressive; so set induction yields $(\forall x \in X)(x \in A)$, and hence we have $(\forall x \in X)\phi(x)$. But this argument only works if there is such a set A, and we only know that if ϕ is definite. Forming A from ϕ requires a comprehension axiom, and comprehension axioms are generally not available to us for predicates that quantify over uncountable collections. Thus we cannot use the classical argument to infer induction for arbitrary predicates from set induction, only induction for definite predicates.

Maybe there is some other way to affirm induction for indefinite predicates? In the case of $\mathbb{N}$, we did not deduce predicate induction from set induction; rather, we directly grasped the validity of proof trees of height ω. This would still work if, for instance, the set X consisted of two copies of $\mathbb{N}$, one laid out after the other. Given that ϕ is progressive, we can argue just as before for $\phi(n)$ for all n in the first copy of $\mathbb{N}$, and once we have this, we can then make the same induction argument for $\phi(n)$ for all n in the second copy of $\mathbb{N}$. Thus we can affirm induction for any predicate on a set of order-type $\omega + \omega$. This suggests that perhaps we could grasp the validity of proof trees of any setwise well-ordered height, and then use them to prove predicate induction. But why should we believe this? What is clear is that validity is progressive in proof trees: if the deduction rules are valid, then assertibility of the current step follows from the assertibility of all preceding steps. But inferring the validity of the entire tree from this would itself require an induction argument, so it cannot be used to justify predicate induction.

We can affirm predicate induction on $\mathbb{N}$ because we can directly grasp its structure. But for us setwise well-ordering is not even a definite predicate (see Section 5.4), so there is no question of directly grasping arbitrary setwise well-orderings in a similar way.

Another idea is to bring in some temporal intuition. Assume $\preceq$ is a setwise well-ordering. If ϕ is progressive, do we not have an intuition for working our way up the set, sequentially proving $\phi(x)$ for each element x, one at a time? Maybe we do, but there are a lot of things wrong with this idea. Most importantly, it assumes the validity of the proof that is being constructed, all the way up the set, so we have again cheated by taking for granted the validity of proofs built up in a setwise well-ordered series of steps, which itself requires an induction argument.

Perhaps there is some other, constructively legitimate, way of inferring general predicate induction from set induction. But I do not know of any such argument.

Why is this conclusion so shocking? (Evidently it is, based on the reactions I have gotten.) I suppose because we are so used to using comprehension that we have internalized the deduction given above. The most common reply I have encountered is that the inference from set induction to predicate induction is intuitively obvious: it needs no special justification because it is immediately evident, like the basic laws of logic. But if you think something is intuitively obvious, and when asked for an explanation you can produce a simple argument showing why it is so, then most likely the reason you thought it was obvious is because you have internalized that simple argument. The reason it seems obvious is because it is comfortable and familiar, not because it immediately presents itself to your intuition as a bedrock truth. When one is reasoning classically one does not stop to think about whether the inference is legitimate, one simply invokes it without a second thought. But the inference from sets to predicates is, classically, true for a reason, and that reason is the one I gave above, and it crucially involves comprehension. You can't just take an argument that is clearly constructively illegitimate and say "But the conclusion is so obvious that constructivists can accept it anyway". No, they can't. Not unless you can come up with some other, constructively legitimate, justification.

In Section 7.3 I suggested that it may seem obvious that $\mathcal{P}(\mathbb{N})$ is a set, not a proper class, because we learn it is a set before learning that there are such things as proper classes, so when that distinction is clarified $\mathcal{P}(\mathbb{N})$ has already been placed in the wrong category. Constructive countablists face a similar danger. There are all kinds of classically legitimate proofs that

are not constructively valid. We really have to be careful about what things we might think are "obviously" true only because we have internalized the classical justification, which is no longer operative.

8.3 Full well-ordering

Full well-ordering. Well-foundedness for ordered sets and for rooted trees. Definiteness of set-trees. Definable ordinals. KP^i.

Set induction is not very useful in itself. Its value is that it entails induction for arbitrary definite predicates. If ϕ is definite, then we really can form the set $\{x : \phi(x)\}$, and progressivity of ϕ implies progressivity of this set, and so on. But there is no way, or at least no obvious way, of inferring induction for indefinite predicates from set induction.

What we want is to make induction arguments for arbitrary predicates, and we saw in Section 8.1 how this works for $\mathbb{N}$. The desired condition, for a total order $\preceq$ on a set X, is

$$(\forall \phi : \mathbb{M}[\ulcorner (\forall x \in X)\phi(x) \urcorner])\mathbb{A}[\ulcorner \mathrm{Prog}_{\preceq}\, \phi \to (\forall x \in X)\phi(x) \urcorner].$$

I will say that $\preceq$ is a *full well-ordering* if it satisfies this condition. This is the useful condition, not setwise well-ordering, since it is what allows us to make general induction arguments. If ϕ is a predicate which we know meaningfully applies to all the elements of X, and $\preceq$ is a full well-ordering on X, then we can use the release law to infer $\ulcorner \mathrm{Prog}_{\preceq}\, \phi \to (\forall x \in X)\phi(x) \urcorner$.

For non-total orderings we can make analogous definitions of setwise and full well-foundedness. We can also do something similar in the context of rooted trees, for the related notion of well-foundedness of a rooted tree T. Here progressivity of a predicate ϕ is defined schematically by the condition $\ulcorner$for all v, if $\phi(w)$ for every descendant w of v, then $\phi(v)\urcorner$. The conditions that define setwise and full well-foundedness are then formally the same as those used to define setwise and full well-ordering, namely,

$$(\forall A \subseteq T)(\mathrm{Prog}\, A \to A = T)$$

and

$$(\forall \phi : \mathbb{M}[\ulcorner (\forall v \in T)\phi(v) \urcorner])\mathbb{A}[\ulcorner \mathrm{Prog}\, \phi \to (\forall v \in T)\phi(v) \urcorner].$$

Alternatively, since every non-initial vertex in a rooted tree does have children, we can formulate a version of ordinary (not strong) induction by using the hypothesis $\ulcorner$for all v, if $\phi(w)$ for every child w of v, then $\phi(v)\urcorner$. In either

case, setwise well-foundedness is equivalent to the nonexistence of an infinite directed path — our original, classical definition of well-foundedness of a rooted tree.

In Sections 2.5 and 2.6 I explained how we could introduce "sets" as abstract objects in a way that would give them the properties we want them to have. We found there that rigid, well-founded, rooted trees — what I called "set-trees" — are one elegant system of proxies for pure sets. Every child v of the root vertex determines a subtree T_v consisting of it and all of its descendants, and this subtree T_v is itself a set-tree that plays the role of an element of the original set. But now we have to ask what version of "well-founded" should be used here. The classical definition, read literally, is setwise well-foundedness, but in a constructive context this condition is too weak to be very useful. So I think full well-foundedness is the right version of well-foundedness to use in the definition of set-trees.

A new issue to consider now is whether set-trees are even definite objects. This is not obvious because deciding whether two set-trees represent the same pure set requires us to determine whether there is an isomorphism between them, and countablistically the bijections between two infinite sets always constitute a proper class, so quantifying over all bijections to find out whether an isomorphism exists is not guaranteed to produce a definite result. Nonetheless, we can argue as follows. First, observe that the predicate $\phi(T)$ = "for any rooted tree T', either T and T' are not isomorphic or there is exactly one isomorphism between them" is progressive: if we know this for T_v for all children v of the root vertex of T, then we can first check whether any two of these T_v's are isomorphic (in which case T has more than one isomorphism with itself), and if not, we can then check, for any rooted tree T', whether each T_v is isomorphic to exactly one T'_w and, if so, whether these isomorphisms use up all of the children of the root vertex of T'. So ϕ is progressive and full well-foundedness then yields that for any set-tree T, it is assertible that T definitely either is or is not isomorphic to any other rooted tree T'. Given any two set-trees, we can use the release rule to infer that they definitely are or are not isomorphic. So set-trees are indeed definite objects in that sense.

(However, that does not entail that there is a class of all set-trees, if we also require of classes, as I do, that they be instantiated by definite predicates. As we discussed at the end of Section 5.4, even setwise well-foundedness is apparently not definite, and the situation is, if anything, even worse for full well-foundedness.)

Which ordinals are available to a constructive countablist? Framed from

a Cantorian perspective, the question is, informally, how far up the class of ordinals constructive countablists can get. The easy version of this question asks about definability: which well-orderings of $\mathbb{N}$ can be given a countablistically valid definition? There is a simple answer to this question. First, every computable well-ordering of $\mathbb{N}$ is clearly countablistically available. It can be constructed by a computation of length ω. But, remarkably, any well-ordering of $\mathbb{N}$ which is constructed by a transfinite computation along a well-ordered computable ordering of $\mathbb{N}$ will still be computable. (The fancy way to say this is that all the ordinals that appear in the hyperarithmetical hierarchy are computable.) So uncomputable well-orderings are hopelessly out of reach of any form of countablism. The countablistically definable ordinals are precisely the computable ordinals. This was first suggested in reference to predicativism by Hao Wang.[1]

The hard version of the question is: which well-orderings of $\mathbb{N}$ can be recognized as well-orderings by a constructive countablist? Every computable total ordering of $\mathbb{N}$ is available to us, but that does not tell us whether, given the limitations of constructive countablism, there is any way for us to know exactly which of them are well-orderings. Even taking "well-ordering" to mean "setwise well-ordering", we cannot quantify over uncountably many nonempty subsets, to check whether they all have minimal elements, and be assured of a definite result.

A simple way to approach this question is to take some formal system whose "proof-theoretic ordinal" is known and ask whether that system can be justified on constructively countablistic grounds. For example, consider intuitionistic Kripke-Platek set theory (KP^i). This is a version of set theory which includes axioms of extensionality, empty set, pairing, and union, as well as schemes of induction, separation, and collection. It is sometimes referred to as a "predicative" system since it appears that all the axioms are consistent with the idea of a universe that is built up from below, and therefore entertains no vicious circles, and also since it does not include a power set axiom which would entail the existence of uncountable sets. As any predicatively valid intuitionistic system will also be valid under constructive countablism, the predicative legitimacy of KP^i would tell us that anything less than the proof-theoretic ordinal of KP^i is available on constructively countablistic grounds. That ordinal, known as the "Bachmann-Howard ordinal", is rather large, so this would tell us a lot about the strength of constructively countablistic reasoning.

[1]See [51], p. 643.

However, KP^i is not justifiable on any obvious constructively countablist, let alone predicativist, grounds. The problem lies in its induction scheme, which consists of the axioms

$$\ulcorner(\forall x)\,((\forall y \in x)\phi(y) \to \phi(x)) \to (\forall x)\phi(x)\urcorner,$$

where ϕ ranges over all the formulas of the language of set theory. In words, if ϕ is progressive in the sense that, for any x, $\phi(y)$ for all $y \in x$ implies $\phi(x)$, then ϕ must be true of all sets. The intuitive justification is that the universe of sets is thought of as being built up in a well-ordered series of stages, and so, inductively, every set satisfies ϕ because all of its elements do.

The problem, of course, is that we have to clarify what version of "well-ordered" is to be used here. If the set-theoretic universe is built up in a set-wise well-ordered series of stages, we could only affirm the induction scheme for formulas for which comprehension was available. This will be okay for Δ_0-formulas, formulas which only involve bounded quantifiers ($\forall y \in x$ and $\exists y \in x$). But weakening full induction to Δ_0-induction drastically lowers the proof-theoretic strength of KP^i.

In order for the justification to work, we must take "well-ordered" to mean fully well-ordered, in the sense of Section 8.3. We can dispense with the temporal intuition of building things up in stages. Instead, we can model pure sets with set-trees, where we have agreed to take the well-foundedness condition to mean full well-foundedness. The problem then is that we only know, for any given ϕ, that for any set x it is assertible that progressivity of ϕ implies $\phi(x)$; that is,

$$(\forall x)\mathbb{A}[\ulcorner(\forall y)((\forall z \in y)\phi(z) \to \phi(y)) \to \phi(\hat{x})\urcorner].$$

It is clear that any formula in the language of set theory is meaningful when its free variables are evaluated as sets and its bound variables are interpreted as ranging over all sets, so meaningfulness is not a problem. We can also, for any particular set x, apply the release rule to infer $\phi(\hat{x})$. But we only have $\forall$-capture when the variable ranges over a set (Section 4.5). When x ranges over all sets we cannot bring the universal quantifier under the assertibility predicate and then apply the release rule. So no, there is no straightforward constructively countablist justification of KP^i.

Among the formal systems with known proof-theoretic ordinals, those with a clear constructively countablistic justifications are proof-theoretically very weak. Another strategy for determining which ordinals are recognizable as ordinals to a constructive countablist is to simply try to

work up the chain of ordinals and see how far we get. That is the approach I will take in the next two sections.

8.4 Iterated classical truth

Grounded truth predicates. Veblen functions.

We can use classical truth to explore the range of ordinals available to constructive countablists. The notion of groundedness used in Section 6.4 is helpful here. If we start with some language all of whose sentences are meaningful, we can expand the language by introducing sentences that refer to the truth of the original sentences, and then expand it further to sentences that refer to the truth of the ones introduced in the last step, and so on. As long as the sentences introduced at each stage only refer to the truth of sentences brought in at previous stages, everything will still be meaningful. We can then introduce sentences that refer to the truth of sentences at arbitrary finite stages, and go on from there. That is, we can move into transfinite stages, and this gives us a tool to investigate the ordinals that are available to us as constructive countablists.

Ordinal analysis is a technical subject, even at the relatively low levels we will consider here, so I will just try to give the general idea.[2] We start with a language $\mathcal{L}$ in which we are able to talk about the meaningfulness and assertibility, but not the truth, of arbitrary sentences. All of the sentences of $\mathcal{L}$ should be meaningful.

Next, we fix a total order $\preceq$ on a set X. This is supposed to be an ordinal notation, but we do not assume that it is well-ordered. We do need to know that it has a least element, that every element has an immediate successor, and so on. There must be computable operations of addition, multiplication, and exponentiation which correspond to the usual ordinal operations. One can prove in Peano arithmetic that any of the standard countable ordinal notation systems has all the properties we need. I will write 0 for the $\preceq$-minimal element of X, 1 for its $\preceq$ successor, and so on.

For each $a \in X$ we require a predicate symbol $\mathbb{T}_a$, and we let $\mathcal{L}_a$ be the original language $\mathcal{L}$ augmented by the predicate symbols $\mathbb{T}_b$ for $b \prec a$. The intuition is supposed to be that $\mathbb{T}_a$ is the global classical truth predicate restricted to $\mathcal{L}_a$: we set $\ulcorner \mathbb{T}_a[\text{“}\phi\text{”}] \leftrightarrow \phi \urcorner$ for any sentence ϕ of $\mathcal{L}_a$ and we set $\mathbb{T}_a[\phi] \leftrightarrow 0 = 1$ for all other ϕ. Of course, this definition only works if we know that all the sentences of $\mathcal{L}_a$ are meaningful. Since we have assumed

[2]More details can be found in [55], from which the material that follows is adapted.

that every sentence of $\mathcal{L}_0 = \mathcal{L}$ is meaningful, we do at least know that $\mathbb{T}_0$ is universally meaningful. And then we can infer that $\mathbb{T}_1$ is too, and so on. The global fact here is that the predicate $M(a) =$ "all the sentences of $\mathcal{L}_a$ are meaningful" is progressive with respect to $\preceq$; that is, we have $(\forall b \prec a)M(b) \to M(a)$, for all $a \in X$. Although the $\mathbb{T}_a$ are not predicates of the base language $\mathcal{L}$, we can assume $\mathcal{L}$ is able to reference the languages $\mathcal{L}_a$ and thus this M is expressible as a predicate in $\mathcal{L}$. The definition of $\mathbb{T}_a$ can be converted from a scheme to a single statement in the usual way, modulated by meaningfulness and assertibility:

$$(\forall a)(M(a) \to (\forall \phi \in \mathcal{L}_a^s)\mathbb{A}[\ulcorner \phi \leftrightarrow \mathbb{T}_a[\text{“}\phi\text{”}]\urcorner])$$

where $\mathcal{L}_a^s$ is the set of sentences in $\mathcal{L}_a$. Again, we assume that the original language $\mathcal{L}$ is expressive enough to accomodate "is a sentence of $\mathcal{L}_a$" as a predicate. So the above axiom is a sentence of $\mathcal{L}$. We can then use the usual axioms of assertibility to prove things like: given $M(a)$, for any predicate ϕ in $\mathcal{L}_a$ whose free variable ranges over a set the sentence

$$(\forall x)\mathbb{T}_a[\phi(\hat{x})] \leftrightarrow \mathbb{T}_a[\ulcorner(\forall x)\phi(x)\urcorner]$$

is assertible. The argument goes along the same lines as the proof of the conjunction law given at the end of Section 5.2. It uses $\forall$-capture from Section 4.5.

Full well-ordering up to $a \in X$ is expressed in $\mathcal{L}$ as

$$W(a) = \text{“}(\forall \phi)(\mathbb{M}[\ulcorner(\forall b)\phi(b)\urcorner] \to \mathbb{A}[\ulcorner \mathrm{Prog}_{\preceq}\phi \to \phi(\hat{a})\urcorner])\text{”},$$

with the quantification going over arbitrary predicates ϕ whose free variable ranges over X. The key point here is that $W(\cdot)$ is itself progressive; this is easy to see by reasoning under the assertibility predicate, again using $\forall$-capture.

We can analyze full well-ordering in terms of the Veblen hierarchy. The *Veblen functions* v_a map ordinals into ordinals, and are defined by induction on the ordinal a by setting $v_0(b) = \omega^b$ and, for all $a > 0$, letting v_a enumerate the common fixed points of $\{v_{a'} : a' < a\}$. At a successor stage $a + 1$ this just means that v_{a+1} enumerates the fixed points of v_a.

We define $\gamma_0 = 0$, $\gamma_1 = v_0(0) = 1$, $\gamma_2 = v_1(0) = \varepsilon_0$, etc., with $\gamma_{n+1} = v_{\gamma_n}(0)$. The ordinal Γ_0 is defined as the limit $\Gamma_0 = \sup_n \gamma_n$. It is the first fixed point of the sequence $(v_a(0))$. We assume that Γ_0 is represented in the ordering $\preceq$.

Let a *γ_n-predicate* be a predicate whose free variable can take values in $\{a \in X : a \prec \gamma_n\}$. Say that a γ_n-predicate ϕ is *progressive up to* γ_n if it satisfies $(\forall b \prec a)\phi(b) \to \phi(a)$ for all $a \prec \gamma_n$. In the next section I will show

how we can use the truth predicate $\mathbb{T}_{\gamma_n}$ to prove that every predicate in $\mathcal{L}_{\gamma_n}$ that is progressive up to γ_{n+1} must hold of all $a \prec \gamma_{n+1}$. The catch is that we can only make this argument if we know that $\mathbb{T}_{\gamma_n}$ is meaningful, i.e., we have $M(\gamma_n)$. But since $M(\cdot)$ is itself progressive up to γ_{n+1}, once we know $M(\gamma_n)$ we can make the argument using $\mathbb{T}_{\gamma_n}$ and then infer $M(a)$ for all $a \prec \gamma_{n+1}$, and from this, $M(\gamma_{n+1})$. This makes reasoning using $\mathbb{T}_{\gamma_{n+1}}$ available and we can then proceed to the next step. The result is an infinitely long proof that ultimately establishes that every predicate in $\mathcal{L}_{\Gamma_0}$ that is progressive up to Γ_0 must hold of every $a \prec \Gamma_0$. Finally, applying this conclusion to $W(\cdot)$ will yield that Γ_0 is fully well-ordered.

8.5 Reaching Γ_0

Structure of the proof. Progressivity of $C(\cdot)$.

Throughout this section fix $n \in \mathbb{N}$ and assume we know $W(\gamma_n)$ and $M(\gamma_n)$ (and hence, by progressivity, also $W(\gamma_n + 1)$ and $M(\gamma_n + 1)$). For each $a, c \prec \gamma_n$ define a γ_{n+1}-predicate B^c_a in the language $\mathcal{L}_{\omega^a \cdot c + 1}$ by

$$B^c_a(b) = \text{“}(\forall \phi \in \mathcal{L}^p_{\omega^a \cdot c})\mathbb{T}_{\omega^a \cdot c}(\ulcorner \mathrm{Prog}\, \phi \to \phi(v_a(b))\urcorner)\text{”}$$

where $\mathcal{L}^p_a$ denotes the set of γ_{n+1}-predicates in $\mathcal{L}_a$ and here we write “$\mathrm{Prog}\, \phi$” to mean that ϕ is progressive up to γ_{n+1}. For fixed a, the predicates B^c_a say essentially the same thing as c varies: that any progressive γ_{n+1}-predicate holds of $v_a(b)$. The only difference is the languages to which they are applied. This is necessary because we want to apply some B^c_a's to others. Note that $a, c \prec \gamma_n$ implies $\omega^a \cdot c \prec \gamma_n$.

It would be a little more natural to define B^c_a to require $\mathrm{Prog}\, \phi \to (\forall b' \preceq b)\phi(v_a(b'))$. However, for any γ_{n+1}-predicate ϕ we have a γ_{n+1}-predicate $\tilde{\phi}(b) =$ “$(\forall b' \preceq b)\phi(b')$” which is progressive up to γ_{n+1} if and only if ϕ is, and applying B^c_a to $\tilde{\phi}$ then yields the desired conclusion about ϕ. So the two definitions are effectively equivalent, which is why I chose the one which is slightly simpler to write.

Let C be the γ_n-predicate in the language $\mathcal{L}_{\gamma_n+1}$

$$C(a) = \text{“}(\forall c \prec \gamma_n)\mathbb{T}_{\gamma_n}(\ulcorner \mathrm{Prog}\, B^c_a\urcorner)\text{”}.$$

Recalling that B^c_a and $B^{c'}_a$ differ only in the languages to which they are applied, the crude intuition for $C(a)$ is that it says the assertion that anything progressive holds of $v_a(b)$ is itself progressive in b. The key theorem we have to prove is that $C(\cdot)$ is itself progressive.

This proof is a little long because each a and each b can be zero, a successor, or a limit, leading to nine separate cases which are nearly all genuinely different. (For limit values of b, the argument is essentially the same regardless of what a is.) For the sake of readability I have split up the proof into three lemmas. The crucial case appears in Lemma 8.3, when a and b are both successors. (It is only here that the "$\omega^a \cdot c$" expression in B^c_a becomes important. The point is that the index $\omega^{a+1} \cdot c$ will always be a limit of indices of the form $\omega^a \cdot c'$.)

Lemma 8.1. *B^c_0 is progressive up to γ_{n+1}, for all $c \prec \gamma_n$. That is, we have $C(0)$.*

Proof. Fix $c \prec \gamma_n$; we must prove $\mathrm{Prog}\, B^c_0$, i.e., we must prove progressivity in b of the statement

$$(\forall \phi \in \mathcal{L}^p_c)\mathbb{T}_c(\ulcorner \mathrm{Prog}\, \phi \to \phi(\omega^b)\urcorner).$$

For $b = 0$ this is trivial; it just says that for any $\phi \in \mathcal{L}^p_c$ it is true (according to $\mathbb{T}_c$) that progressivity of ϕ implies that ϕ holds of $\omega^0 = 1$.

For limit values of b, assume we are given $(\forall b' \prec b)\mathbb{T}_c(B^c_0(b'))$; that is, assume

$$\mathbb{T}_c(\ulcorner \mathrm{Prog}\, \phi \to \phi(\omega^{b'})\urcorner)$$

for all $\phi \in \mathcal{L}^p_c$ and all $b' \prec b$. Bringing the quantifier "$\forall b' \prec b$" under $\mathbb{T}_c$, we get that for every $\phi \in \mathcal{L}^p_c$ it is true (according to $\mathbb{T}_c$) that progressivity of ϕ implies $\phi(\omega^{b'})$ for all $b' \prec b$. Since $\omega^b = \sup_{b' \prec b} \omega^{b'}$, we infer from this (for arbitrary ϕ, under $\mathbb{T}_c$) that progressivity of ϕ implies $\phi(\omega^b)$.

Finally, if $b = b' + 1$ is a successor, assume that for every $\phi \in \mathcal{L}^p_c$ it is true (according to $\mathbb{T}_c$) that progressivity of ϕ implies that ϕ holds of $\omega^{b'}$. As we noted above, this also implies that any progressive ϕ holds of everything less than $\omega^{b'}$, so $\phi^{\to \omega^{b'}}(b) = \phi(\omega^{b'} + b)$ is also progressive, and therefore ϕ holds of $\omega^{b'} + \omega^{b'}$. ontinuing inductively, we get the truth for all $n \prec \omega$ of $\phi(\omega^{b'} \cdot n)$, and progressivity of ϕ then implies that ϕ holds of $\omega^{b'} \cdot \omega = \omega^{b'+1} = \omega^b$. We have shown that $B^c_0(b')$ implies $B^c_0(b' + 1)$. This completes the proof that $(\forall \phi \in \mathcal{L}^p_c)\mathbb{T}_c(\ulcorner \mathrm{Prog}\, \phi \to \phi(\omega^b)\urcorner)$ is progressive in b and establishes $C(0)$. □

Lemma 8.2. *For any limit $a \prec \gamma_n$ we have $(\forall a' \prec a)C(a') \to C(a)$.*

Proof. Fix a limit $a \prec \gamma_n$ and assume $(\forall a' \prec a)C(a')$. This means that we are given the truth, according to $\mathbb{T}_{\gamma_n}$, of $\mathrm{Prog}\, B^{c'}_{a'}$ for every $a' \prec a$ and

$c' \prec \gamma_n$, and we must show that for every $c \prec \gamma_n$ it is true according to $\mathbb{T}_{\gamma_n}$ that B^c_a is progressive.

Fix $c \prec \gamma_n$. First we check the case $b = 0$, where we have to verify $B^c_a(0)$. Since we are assuming, for every $c' \prec \gamma_n$ and $a' \prec a$, that it is true (according to $\mathbb{T}_{\gamma_n}$) that $B^{c'}_{a'}$ is progressive, in particular we have for all such c' and a' the truth of $B^{c'}_{a'}(0)$. Thus, for every ϕ in some $\mathcal{L}^p_{\omega^{a'}\cdot c'}$ with $a' \prec a$ and $c' \prec \gamma_n$ — which is to say, every ϕ in $\mathcal{L}^p_{\gamma_n}$ — we have that progressivity of ϕ implies the truth of $\phi(v_{a'}(0))$ for all $a' \prec a$. In particular, for any $\phi \in \mathcal{L}^p_{\omega^a\cdot c}$ it is true according to $\mathbb{T}_{\gamma_n}$, and hence according to $\mathbb{T}_{\omega^a\cdot c}$, that progressivity of ϕ implies that A holds of $\sup_{a'\prec a} v_{a'}(0) = v_a(0)$. This verifies $B^c_a(0)$. That was the $b = 0$ case.

Next we will show, reasoning under $\mathbb{T}_{\gamma_n}$, that if b is a limit and we have $B^c_a(b')$ for all $b' \prec b$, then we also have $B^c_a(b)$. At this point we have two induction hypotheses going, one for $a' \prec a$ and one for $b' \prec b$. To verify $B^c_a(b)$, let $\phi \in \mathcal{L}^p_{\omega^a\cdot c}$. Then $B^c_a(b')$ for all $b' \prec b$ tells us that according to $\mathbb{T}_{\omega^a\cdot c}$, progressivity of ϕ implies that ϕ holds of $v_a(b')$ for all $b' \prec b$. So according to $\mathbb{T}_{\omega^a\cdot c}$, progressivity of ϕ implies that ϕ holds of $\sup_{b'\prec b} v_a(b') = v_a(b)$, as desired. We did not even need the induction hypothesis on a in this part.

The final case is for successor values of b. Say $b = b' + 1$. Reasoning under $\mathbb{T}_{\gamma_n}$, we have to verify that $B^c_a(b')$ implies $B^c_a(b'+1)$. For this we must use both (1) the induction hypothesis on a (for every $c' \prec \gamma_n$ and $a' \prec a$ the γ_{n+1}-predicate $B^{c'}_{a'}$ is progressive) and (2) the induction hypothesis on b (we have $B^c_a(b')$). Now $B^c_a(b')$ says that every progressive $\phi \in \mathcal{L}^p_{\omega^a\cdot c}$ holds of $v_a(b')$, and $v_a(b')$ is a fixed point of every $v_{a'}$ with $a' \prec a$, so we can also say that every progressive $\phi \in \mathcal{L}^p_{\omega^a\cdot c}$ holds of $v_{a'}(v_a(b'))$ for all $a' \prec a$. This means that $B^{c'}_{a'}(v_a(b'))$ holds for all $a' \prec a$ and $c' \prec \gamma_n$ such that $\omega^{a'}\cdot c' \preceq \omega^a \cdot c$.

But for any $a' \prec a$, we can find a'' such that $a' + a'' = a$, and then by putting $c' = \omega^{a''} \cdot c$, we get $\omega^{a'} \cdot c' = \omega^a \cdot c$. So by the conclusion reached in the last paragraph, for any $a' \prec a$ there exists $c' \prec \gamma_n$ such that $\omega^{a'} \cdot c' = \omega^a \cdot c$ and $B^{c'}_{a'}(v_a(b'))$ holds. And since every such $B^{c'}_{a'}$ is progressive, by (1), it follows that $B^{c'}_{a'}(v_a(b') + 1)$ also holds.

We are now in a position to affirm that for every $a' \prec a$, every progressive $\phi \in \mathcal{L}^p_{\omega^a\cdot c}$ lies within the scope of some $B^{c'}_{a'}$ which satisfies $B^{c'}_{a'}(v_a(b') + 1)$, so that ϕ holds of $v_{a'}(v_a(b') + 1)$. If a' is a successor, $a' = a'' + 1$, then this means that ϕ holds of the next fixed point of $v_{a''}$ after $v_a(b')$. Thus, by progressivity, ϕ holds of the supremum of these fixed points over all $a'' \prec a$, which equals $v_a(b' + 1) = v_a(b)$. That verifies $B^c_a(b)$

and completes the proof of $C(a)$. □

Lemma 8.3. *For any $a \prec \gamma_n$, $C(a)$ implies $C(a+1)$.*

Proof. Fix $a \prec \gamma_n$ and assume $C(a)$, and on the way to proving $C(a+1)$, fix $c \prec \gamma_n$. Reasoning under $\mathbb{T}_{\gamma_n}$, we must prove that B^c_{a+1} is progressive, i.e., that for every b, if $B^c_{a+1}(b')$ holds for all $b' \prec b$ then $B^c_{a+1}(b)$ also holds.

If $b = 0$ then we simply have to show that $B^c_{a+1}(0)$ holds, i.e., that any progressive $\phi \in \mathcal{L}^p_{\omega^{a+1}\cdot c}$ holds of $v_{a+1}(0)$. Since we have assumed $C(a)$, we know that for any $c' \prec \gamma_n$, according to $\mathbb{T}_{\gamma_n}$ the γ_{n+1}-predicate $B^{c'}_a$ is progressive. In particular, for each $m \prec \omega$ the γ_{n+1}-predicate B^{c+m}_a is progressive. Thus B^{c+m}_a holds of 0, so every progressive predicate in $\mathcal{L}^p_{\omega^a(c+m)}$ holds of $v_a(0)$, and hence of $v_a(0)+1$. In particular, B^{c+m-1}_a holds of $v_a(0)+1$. So every progressive predicate in $\mathcal{L}^p_{\omega^a(c+m-1)}$ holds of $v_a(v_a(0)+1)$. Inductively, every progressive predicate in $\mathcal{L}^p_{\omega^a(c+m-k)}$ holds of $v^k_a(v_a(0)+1)$, which means (taking $k=m$) that every progressive predicate in $\mathcal{L}^p_{\omega^a\cdot c}$ holds of $v^m_a(v_a(0)+1)$. Since m was arbitrary, it follows that every progressive predicate in $\mathcal{L}^p_{\omega^a\cdot c}$ holds of $\sup_{m\prec\omega} v^m_a(v_a(0)+1) = v_{a+1}(0)$. This finishes the $b=0$ case.

Next, assuming that for some limit b we have $B^c_{a+1}(b')$ for all $b' \prec b$, we must verify $B^c_{a+1}(b)$. This is easy, because we are given that any progressive predicate in $\mathcal{L}^p_{\omega^{a+1}\cdot c}$ holds of $v_{a+1}(b')$ for all $b' \prec b$, and so by progressivity it holds of $\sup_{b'\prec b} v_{a+1}(b') = v_{a+1}(b)$. The corresponding cases in Lemmas 8.1 and 8.2 were handled in the same way.

Finally, letting b be arbitrary, we must make the inference from $B^c_{a+1}(b)$ to $B^c_{a+1}(b+1)$. So assume $B^c_{a+1}(b)$ and fix $\phi \in \mathcal{L}^p_{\omega^{a+1}\cdot c}$. We can write $\omega^{a+1}\cdot c = \omega^a \cdot \omega c$, and since ωc is a limit, this shows that ϕ must belong to $\mathcal{L}^p_{\omega^a\cdot c'}$ for some $c' \prec \omega c$. Moreover, since ωc is a limit, we have $c'+m \prec \omega c$ for all $m \prec \omega$.

Now for each $m \prec \omega$ the γ_{n+1}-predicate $B^{c'+m}_a$ belongs to $\mathcal{L}^p_{\omega^{a+1}\cdot c}$, and is progressive since we are assuming $C(a)$, so the hypothesis of $B^c_{a+1}(b)$ yields $B^{c'+m}_a(v_{a+1}(b))$, and then by progressivity we get $B^{c'+m}_a(v_{a+1}(b)+1)$. Applying this to $B^{c'+m-1}_a$, which lies in $\mathcal{L}^p_{\omega^a(c'+m-1)+1} \subseteq \mathcal{L}_{\omega^a(c+m)}$, then yields $B^{c'+m-1}_a(v_a(v_{a+1}(b)+1))$, and inductively we finally get $B^{c'}_a(v^m_a(v_{a+1}(b)+1))$. Since $\phi \in \mathcal{L}^p_{\omega^a\cdot c'}$, if ϕ is progressive this entails $\phi(v^{m+1}_a(v_{a+1}(b)+1))$, and as m was arbitrary $\phi(v_{a+1}(b+1))$ follows. We have established $B^c_{a+1}(b)$, as desired. □

Putting these lemmas together yields the following theorem.

Theorem 8.4. *C is progressive up to γ_n.*

Finally, since we initially assumed that we knew $W(\gamma_n)$, we can now use the release rule to infer $C(a)$ for all $a \prec \gamma_n$; in particular, we have $B^0_a(0)$ for all $a \prec \gamma_n$. This will let us infer $W(v_a(0))$ and $M(v_a(0))$ for all $a \prec \gamma_n$, and then, again by progressivity, $W(\gamma_{n+1})$ and $M(\gamma_{n+1})$. We can now follow the strategy described at the end of the last section to work our way up to arbitrarily large values of n, and thereby prove $W(\gamma_n)$ for arbitrary n. Amalgamating these proofs yields a single infinitely long proof of $(\forall n)W(\gamma_n)$, and then by progressivity of $W(\cdot)$ we can infer $W(\Gamma_0)$.

8.6 Going further

The infinitary release rule. A new paradox. Larger ordinals.

In the last section we saw how to prove $W(\Gamma_0)$. The proof proceeds in a series of stages: at stage n, we use the truth predicate $\mathbb{T}_{\gamma_n}$ to prove $W(\gamma_{n+1})$; in particular, this implies $M(\gamma_{n+1})$, which legitimates the use of $\mathbb{T}_{\gamma_{n+1}}$ in the next step. Amalgamating the entire series of stages yields an infinitely long proof of $(\forall a \prec \Gamma_0)W(a)$, and one final inference using progressivity of $W(\cdot)$ yields $W(\Gamma_0)$.

As countablists, we can (and must) agree to accept the legitimacy of infinitely long proofs. As we saw in Section 8.1, ordinary induction on $\mathbb{N}$ already uses this. What we can prove in a finitistic manner is $(\forall n)\mathbb{A}[\ulcorner \mathrm{Ind}(\phi) \to \phi(\hat{n})\urcorner]$, for any ϕ which meaningfully applies to all natural numbers. Then, as countablists, we bring the $\forall n$ quantifier inside the assertibility predicate on the grounds that an infinite series of proofs can be amalgamated into a single proof. For any particular ϕ which satisfies the meaningfulness premise, we can then remove the assertibility predicate using the release rule.

The well-ordering proof for Γ_0 is a little different because we do not literally prove that $W(\gamma_n)$ implies $W(\gamma_{n+1})$. Rather, we use the implication $W(\gamma_n) \to M(\gamma_n)$ to legitimate the proof that infers $W(\gamma_{n+1})$. So what we are actually proving is $W(\gamma_n) \to \mathbb{A}[W(\gamma_{n+1})]$. A straight induction argument would then yield $(\forall n)\mathbb{A}^n[W(\gamma_n)]$. The final inference to $(\forall n)W(\gamma_n)$ could then be made using an infinitary version of the release rule which states "given $(\forall n)\mathbb{A}^n[\phi(\hat{n})]$, infer $(\forall n)\phi(n)$". The informal justification of this rule is simply that for each value of n we have $\mathbb{A}^{\hat{n}}[\phi(\hat{n})]$, and n applications of the ordinary release rule lets us infer $\phi(\hat{n})$ for that value of n. Thus the rule codifies an infinite series of inferences as a single inference, a sort of process that should be very familiar by now.

The release rule is subtle, however. Here I want to emphasize that we do not adopt "infer ϕ from $\mathbb{A}[\phi]$" as a global principle, only as one which is valid relative to formal systems which are trusted to constructively provide, along with any proof of $\mathbb{A}[\phi]$, an actual proof of ϕ. Indeed, taken as a global principle, the release rule becomes paradoxical. If we could agree, globally, that any proof of $\mathbb{A}[\phi]$ legitimates a subsequent inference of ϕ, then we could affirm the implication $\mathbb{A}[\mathbb{A}[\phi]] \to \mathbb{A}[\phi]$ as a global principle. And this would lead to a paradox in relation to the sentence "if this sentence is assertible, then $\bot$ is assertible". (Call this sentence "Λ'_a". If Λ'_a is assertible, then both $\mathbb{A}[\Lambda'_a]$ and $\mathbb{A}[\Lambda'_a] \to \mathbb{A}[\bot]$ are assertible, and therefore $\mathbb{A}[\bot]$ is assertible. That is, we have $\mathbb{A}[\Lambda'_a] \to \mathbb{A}[\mathbb{A}[\bot]]$. If we had "$\mathbb{A}[\bot]$-release", i.e., the implication $\mathbb{A}[\mathbb{A}[\bot]] \to \mathbb{A}[\bot]$, then we could infer $\mathbb{A}[\Lambda'_a] \to \mathbb{A}[\bot]$, i.e., Λ'_a itself. From this we could infer that Λ'_a is assertible, and then that $\bot$ is assertible.)

Any legitimate use of the release rule should, in principle, be removable. I made this point in Chapter 6 in relation to the formal system ATM, expressed there by saying that all uses of the release rule were cosmetic. In fact, this is why I chose to present the proof of $W(\Gamma_0)$ in the last section without using the infinitary release rule mentioned above: I wanted it to be clear that one could explicitly give an infinite series of proofs of the sentences $W(\gamma_n)$, showing that the use of an infinitary release rule would be merely cosmetic.

We did use the ordinary release rule in making the inference from $W(\gamma_n)$ to $C(a)$ for all $a \prec \gamma_n$, so maybe it is worth explaining how even this could have been avoided. We just need to recognize that the argument that proves $\mathrm{Prog}\,\phi \to \phi(\gamma_n)$ for $\phi \in \mathcal{L}^p$ could be applied to any $\mathcal{L}_a$ in place of $\mathcal{L}$, once we have $M(a+1)$ and can recognize $\mathcal{L}_a$ as meaningful. So if we have proven $M(\gamma_n+1)$ and can reason in the language $\mathcal{L}_{\gamma_n+1}$, we can prove progressivity of C up to γ_n as in the last section, then make the argument that lets us infer $C(a)$ for all $a \prec \gamma_n$. In this way we can avoid any reference to $W(\cdot)$. The cost of doing this is merely that it adds some recursion to the proof, making it a little more complicated.

The delicacy of the release rule has implications for going further beyond Γ_0. Now that we have proven $W(\Gamma_0)$, we can essentially repeat the argument to prove $W(\Gamma_1)$, then $W(\Gamma_2)$, and so on. We can amalgamate the resulting infinite series of proofs into a single proof of $W(\Gamma_\omega)$. But systematizing this process is not straightforward because the general inference is not from $W(\Gamma_a)$ to $W(\Gamma_{a+1})$, but rather from $W(\Gamma_a)$ to $\mathbb{A}[W(\Gamma_{a+1})]$. One can deal with this issue using the infinitary release rule, but one must al-

ways remember that any use of this rule should, in principle, be removable. So getting up to substantially larger countable ordinals may be difficult. I leave this as a problem for others more skilled than I.

PART 2
Mathematics

Chapter 9

Fundamentals

9.1 Constructive countablism

The main goal of Part 1 of this book was to explain and justify our two core principles: truth constructivism (Chapter 4) and countablism (Chapter 7). In Part 2 I will show how to do normal mathematics on this basis. I will write as if to someone learning abstract mathematics for the first time, though, out of necessity, in a very condensed fashion. I do not think the treatment given here is the best way to actually teach this material to a true beginner; that is a pedagogical problem that I am not trying to solve.

Countablism is encapsulated in the following two statements:

(1) if each of s_1, s_2, s_3, ... is a well-defined computation (or construction, logical derivation, etc.) then they can be amalgamated sequentially into a single, well-defined, infinitely long computation (construction, etc.) $s_1 s_2 \ldots$;
(2) there are no legitimate infinite processes or constructions other than those sanctioned by recursive applications of (1).

For instance, it is not currently known whether every even number greater than 2 can be written as a sum of two prime numbers (Goldbach's conjecture). But it is easy enough to answer this question for any particular even number n. Just check, for each $k < n$, whether k and $n - k$ are both prime. If there is such a k, the answer is yes, otherwise it is no. Since we can do this separately for each even number greater than 2, the statement (1) tells us that there is a single computation which sequentially checks it for *all* even numbers greater than 2. The result is an infinitely long computation which determines whether Goldbach's conjecture is true.

If it helps the intuition, one can imagine that the first step of such a

computation takes one second, the second step takes half a second, the third step takes one fourth of a second, and so on. Then the entire infinite sequence of steps would be finished two seconds after it started.

The real power of (1) lies in the fact that it can be applied recursively. Any infinitely long computation it legitimizes can be fed back into itself as a single step of some new computation. We can run computations with infinitely many steps, where each of these steps is itself an infinitely long computation. To see how this might work, suppose we wanted to know whether there exist natural numbers x, y, and z satisfying $x^3 + y^3 = z^3$. A single infinite computation could answer this question: for each value of z, we can run through all $x, y < z$ and check for a solution, and putting all these computations together, for all values of z, yields an infinite computation that will answer the original question. So far, this is not too different from the previous example. But suppose we also want to know whether there exist x, y, and z satisfying $x^4 + y^4 = z^4$. Another infinite computation, similar to the first, could answer this question. And likewise for $x^5 + y^5 = z^5$, for $x^6 + y^6 = z^6$, and so on. We see that for each exponent there is a well-defined infinite computation that determines whether there are any solutions with that exponent, and according to (1) these can be amalgamated into a single infinite series of infinite computations that determines whether there are any solutions to $x^n + y^n = z^n$ for any $n \geq 3$.

In fact, we know that there are no solutions, for any $n \geq 3$; this is Fermat's last theorem. But of course mathematicians didn't find this out by actually running an infinitely long computation. We know it's the case because we have a proof. In real life we can't perform infinitely long computations, or even prohibitively long finite computations, for that matter. The value of (1) is theoretical: it tells us that infinite computations are meaningful, and can be assumed to produce well-defined results — results that we might be able to discover without doing the brute-force computation, if we are clever enough.

Computations are not the only sort of thing that can be strung together in this way. Constructions can too. The "middle thirds Cantor space" is an example of an infinitely long construction. Start with the unit interval $[0, 1]$; then remove the open middle third interval $(\frac{1}{3}, \frac{2}{3})$, leaving the two closed intervals $[0, \frac{1}{3}]$ and $[\frac{2}{3}, 1]$; then remove the middle thirds of each of those, leaving the four closed intervals $[0, \frac{1}{9}]$, $[\frac{2}{9}, \frac{1}{3}]$, $[\frac{2}{3}, \frac{7}{9}]$, and $[\frac{8}{9}, 1]$; and continue on in this way. What is left, after an infinitely long process, is the Cantor space.

Logical derivations can be amalgamated as well. One basic example

here is the principle of induction. Let ϕ be some property, and assume both that the number 1 has this property, and that whenever n has this property so does $n + 1$. The principle of induction draws the conclusion that every natural number has the property ϕ, and we can show why this must be so with an infinitely long proof. In the first step of the proof we note that we are given $\phi(1)$ as a hypothesis. Next, putting $n = 1$ in the other hypothesis yields that $\phi(1)$ implies $\phi(2)$, so we can infer $\phi(2)$. Then putting $n = 2$ in the second hypothesis yields that $\phi(2)$ implies $\phi(3)$, so we can infer $\phi(3)$. And so on — the result is an infinitely long derivation which ultimately proves $\phi(n)$ for all n.

What about statement (2)? It tells us that ideas like "search through all the real numbers, and check whether they all have the property ψ" are senseless. There is no such thing as an infinite computation with one step for each real number. What would such a computation look like? One is at a complete loss to imagine such a thing.

We know for a fact that no computation sanctioned only by recursive applications of (1) can run through all the real numbers because the steps of any such computation can be *enumerated* — this means that they can be labelled with natural numbers in such a way that each step gets a distinct label. (Not all natural numbers need be used.) Another way to say this is that there are *countably many* steps in any computation sanctioned only by recursive applications of (1). To see this, observe that any labelling with natural numbers can be converted into a labelling with only odd numbers, by replacing the label n with $2n - 1$. (Instead of 1, 2, 3, ... use 1, 3, 5, ...) So if the steps that comprise s_1 are countable, as are the steps that comprise s_2, and so on, then we can relabel each of these groups using only odd numbers. Then leave the steps that comprise s_1 alone; multiply the labels of the steps that comprise s_2 by two; multiply the labels of the steps that comprise s_3 by four; and so on, running through all powers of two. Since every natural number can be written in only one way as a power of two times an odd number, we now have distinct labels for all the steps in the entire sequence $s_1 s_2 \ldots$. Conclusion: if the steps that comprise each s_i are countable, then the steps that comprise the entire series $s_1 s_2 \ldots$ are also countable. Recursive applications of (1) never take us out of the realm of countable computations. But the real numbers are not countable, they cannot be enumerated (Theorem 9.34). That is why we know we cannot run a computation with a separate step for each real number.

Certainly, it is possible that we might be able to prove every real number has some property in an indirect way, without actually checking them all,

one by one; the point is just that the idea of performing a brute force search of all real numbers is illegitimate. For example, consider the assertion that whenever a and b are irrational numbers, so is a^b. We reject as nonsensical the idea of checking this claim by exhaustively searching through all pairs of irrational numbers. But that does not mean that this question cannot be answered; see below.

That sums up the meaning of countablism. Our other core premise, *truth constructivism*, is the doctrine that

(3) every fact can, in principle, be known.

There is no such thing as an unknowable fact. This seeminly mild proposition means that even if some sentence is perfectly meaningful and precise we cannot necessarily assume it is definitely true or false, because there might be no way, even in principle, to either verify (prove) or falsify (disprove) it, which is required for it to be true or false.

The really subtle aspect of this idea is that the claim that some sentence is definitely either true or false might, conceivably, itself be something we can neither prove or disprove. So we cannot say of every sentence that it is either definitely true, or definitely false, or definitely neither true nor false. In fact, we do not expect to ever be able to prove that some statement is definitely neither true nor false. We should regard any statement that is not currently known to have a truth value (true or false) as potentially having one. Once we have proven something we can affirm it; once we have disproven something we can affirm its negation; but until one of those two things is done, or at least until we know that one of them *can* be done, we must hold our tongues. "Whereof one cannot speak, thereof one must be silent."[1]

The good news is that countablism gives us substantial resources to establish that statements are definitely true or false — let us just say *definite*. For instance, based on what we've already discussed, we can agree that Goldbach's conjecture is definite, because we know that there is, in principle, a way of either verifying or falsifying it. Just run through all the even numbers greater than 2 and check it; if at the end of the day we have found no counterexamples it is true, and if we have found at least one counterexample it is false. However, statements about arbitrary real numbers typically cannot be mechanically checked, not even in principle, according to countablism. So these statements cannot be assumed to always be def-

[1] [58], "Proposition 7".

initely true or false. (But let me emphasize that we never expect to be in a position to affirm that some statement is *not* definitely true or false, because statements that appear indefinite at the moment always have the potential to be revealed as definite as our knowledge increases.)

The practical implication is that we have to use *intuitionistic logic*, whose essential feature is that it demands extra care when reasoning about negations. If the statement ϕ is not assumed to have a truth value, we can no longer interpret the negation of ϕ to be a statement with opposite truth value to ϕ. Instead, we interpret it as "ϕ entails a contradiction". And we can no longer globally affirm the *law of excluded middle*, which states "either ϕ or not-ϕ". If we do not know how to prove either ϕ or that ϕ entails a contradiction, we cannot assume that ϕ satisfies excluded middle. This is just a restatement of what I was saying above about not being able to assume every sentence is definitely true or false. The absence of this law is what distinguishes intuitionistic logic from classical logic. (We can, of course, affirm excluded middle in many cases. For instance, if we happen to know that ϕ is true, then we can certainly infer that either ϕ or not-ϕ is true.)

I can illustrate the significance of excluded middle using the question I asked above about the irrationality of a^b. We may argue that there must be irrational numbers a and b such that a^b is rational, in the following way. First, consider $\sqrt{2}^{\sqrt{2}}$. If it is rational, we are done: $a = b = \sqrt{2}$ is an example. On the other hand, if it is irrational, then

$$\left(\sqrt{2}^{\sqrt{2}}\right)^{\sqrt{2}}$$

would be an irrational number raised to an irrational power, and we can calculate

$$\left(\sqrt{2}^{\sqrt{2}}\right)^{\sqrt{2}} = \sqrt{2}^{\sqrt{2}\cdot\sqrt{2}} = \sqrt{2}^{2} = 2.$$

So in this case $a = \sqrt{2}^{\sqrt{2}}$, $b = \sqrt{2}$ is the example.

This argument relies on the dichotomy "either $\sqrt{2}^{\sqrt{2}}$ is rational or it is not rational", an example of excluded middle. So unless we have some special reason to affirm this dichotomy, we cannot accept the preceding argument as valid. Now, in this case, countablism comes to the rescue, because a countable computation could determine whether or not $\sqrt{2}^{\sqrt{2}}$ is rational (by checking its decimal expansion for eventual periodicity, for example), so this must, in fact, definitely either be or not be the case. Therefore we can affirm the excluded middle and proceed with the argument.

This example illustrates both the significance of the law of excluded middle and the power of countablism in relation to it. Intuitionists, who combine truth constructivism with a rejection of any kind of infinitely long computation, do not accept the argument I gave above as a valid proof that an irrational number taken to an irrational power can be rational. (There are other simple proofs of this fact that they do accept.)

9.2 Intuitionistic logic

Here is a brief introduction to intuitionistic logic.

First, some notation: we use the symbol $\perp$ to denote "absurdity"; for the sake of concreteness, we can take it to stand for "$0 = 1$". For any statements ϕ and ψ,

$\phi \wedge \psi$	means	"ϕ and ψ" (conjunction)
$\phi \vee \psi$	means	"ϕ or ψ" (disjunction)
$\phi \to \psi$	means	"if ϕ, then ψ" (implication).

We can define other logical symbols in terms of these: we let $\neg\phi$ — in words, "not ϕ" — abbreviate "$\phi \to \perp$" (negation) and we let $\phi \leftrightarrow \psi$ — in words, "ϕ if and only if ψ" or "ϕ iff ψ" — abbreviate "$(\phi \to \psi) \wedge (\psi \to \phi)$" (biconditional).

If $\phi(x)$ is a different statement for each value of x (as in, say, "x is a prime number") then we say that x is a *free variable* in ϕ, and

$(\forall x)\phi(x)$	means	"for all x, $\phi(x)$" (universal quantification)
$(\exists x)\phi(x)$	means	"there exists an x such that $\phi(x)$" (existential quantification).

Once the variable x is prefaced by a "quantifier", $\forall$ or $\exists$, it is no longer free and is said to be *bound* or *quantified*.

We start with some pool of primitive assertions, called *atomic formulas*, whose meaning is understood; a general *formula* is then any grammatical statement that can be built up from the atomic formulas using the logical symbols. A formula in which all variables are bound is a *sentence*, and a formula with exactly one free variable (though it could appear more than once) is a *predicate*.

Intuitionistic reasoning is most easily understood in terms of *natural deduction*, which involves "introduction" and "elimination" rules. The rules for $\wedge$, $\vee$, and $\to$ are

- given ϕ and ψ, infer $\phi \wedge \psi$
- given $\phi \wedge \psi$, infer either ϕ or ψ (or both)
- given either ϕ or ψ, infer $\phi \vee \psi$
- given $\phi \vee \psi$, $\phi \to \theta$, and $\psi \to \theta$, infer θ
- given a proof of ψ from the premise ϕ, infer $\phi \to \psi$
- given ϕ and $\phi \to \psi$, infer ψ.

The last rule is called *modus ponens*. The rules for quantifiers involve the notion of a *term*, which can be any expression that may be substituted for one of the free variables. Thus

- given $\phi(x)$, infer $(\forall x)\phi(x)$
- given $(\forall x)\phi(x)$, infer $\phi(t)$ for any term t
- given $\phi(t)$ for any term t, infer $(\exists x)\phi(x)$
- given $(\exists x)\phi(x)$ and $\phi(x) \to \psi$, infer ψ.

And finally, we have one special rule for $\perp$:

- given $\perp$, infer anything.

This rule is called *ex falso*, and it should not be hard to accept. If $0 = 1$ then adding 1 to each side yields $1 = 2$; this implies that any two distinct sentences must be one and the same, so if any sentence is true then they all are.

The above is just meant to be a brief, informal description of natural deduction; more careful explanations can be found elsewhere. My goal is merely to give the reader an idea of how it works. We really do just reason "naturally", according to the way we intuitively understand the meanings of the logical symbols, but we don't have any rule that says $\phi \vee \neg\phi$, i.e., we lack excluded middle.

One important general consequence of not having the law of excluded middle is that we no longer have the ability to deduce ϕ from not-not-ϕ, a rule known as *double negation elimination*. And without double negation elimination, we lack the ability to give proofs "by contradiction", which is the technique whereby we deduce a contradiction from the assumption that some statement is false, and conclude that that statement must have been true. Under intuitionistic logic, if we assume ϕ and deduce a contradiction, we can conclude that ϕ entails a contradiction, which means that we have proven not-ϕ. But if we assume not-ϕ and deduce a contradiction, we cannot conclude ϕ, only not-not-ϕ. Not-not-ϕ says that ϕ is not false, but

it does not say that ϕ is true, only, if you like, that it is potentially true but not potentially false.

Now let's practice proving a few simple facts. The first proposition will show how excluded middle and double negation elimination are related.

In the next few results, ϕ and ψ can be any two formulas, unless otherwise specified.

Lemma 9.1. *We have $\neg(\phi \wedge \neg\phi)$, or equivalently $(\phi \wedge \neg\phi) \to \bot$. Indeed, $(\phi \wedge \neg\phi) \to \psi$ for arbitrary ψ.*

Proof. Assume $\phi \wedge \neg\phi$; we must derive a contradiction. That will establish $\neg(\phi \wedge \neg\phi)$.

To do this, write out $\neg\phi$ as $\phi \to \bot$. Thus, we have assumed both ϕ and $\phi \to \bot$. By modus ponens, we infer $\bot$. That's the whole proof; assuming $\phi \wedge \neg\phi$ we deduced $\bot$, which means that we proved $\neg(\phi \wedge \neg\phi)$.

For the second part, assume $\phi \wedge \neg\phi$. We just saw how to infer $\bot$ from this premise. In one further step, using ex falso, we can infer ψ, for any ψ at all. Thus, we have shown that $\phi \wedge \neg\phi$ implies ψ. □

Proposition 9.2. *ϕ implies $\neg\neg\phi$. If we have $\phi \vee \neg\phi$, we get the reverse implication from $\neg\neg\phi$ to ϕ.*

Proof. Assume ϕ; we want to prove $\neg\neg\phi$. That is, we want to show $\neg\phi \to \bot$.

Introduce $\neg\phi$, that is to say, $\phi \to \bot$, as a secondary assumption. Since we had already assumed ϕ we can now use modus ponens to infer $\bot$.

Under the assumption of ϕ, we just saw that $\neg\phi$ entails a contradiction. So we have shown that ϕ entails $\neg\phi \to \bot$, i.e., ϕ entails $\neg\neg\phi$. This proves the first part.

For the second part, assume $\phi \vee \neg\phi$; we now want to show that $\neg\neg\phi$ implies ϕ.

To do this, introduce $\neg\neg\phi$ as a secondary assumption and try to infer ϕ. Since we already had $\phi \vee \neg\phi$ as our main assumption, we know that either ϕ or $\neg\phi$ is true. If the former, we don't have to do anything more to infer ϕ, we already have it; if the latter, we can use the second part of Lemma 9.1 with $\neg\phi$ in place of ϕ and ϕ in place of ψ to deduce ϕ. So in either case we were able to infer ϕ, and this means that given $\phi \vee \neg\phi$ we are able to deduce ϕ from $\neg\neg\phi$. □

Exercise. Prove that $\neg\neg\neg\phi$ is equivalent to $\neg\phi$, without assuming any excluded middle.

Proposition 9.3. *$\neg\phi \wedge \neg\psi$ is equivalent to $\neg(\phi \vee \psi)$.*

Proof. To prove the forward implication, assume $\neg\phi\wedge\neg\psi$; then introduce $\phi\vee\psi$ as a secondary assumption, with the goal of deriving a contradiction.

So, we have both $\neg\phi$ and $\neg\psi$, i.e., $\phi \to \bot$ and $\psi \to \bot$, plus either ϕ or ψ. If the former, modus ponens with $\phi \to \bot$ yields $\bot$, and if the latter, modus ponens with $\psi \to \bot$ yields $\bot$. So regardless of whether ϕ or ψ holds, we get a contradiction. This shows that, under the assumption of $\neg\phi\wedge\neg\psi$, we have $(\phi \vee \psi) \to \bot$, i.e., $\neg(\phi \vee \psi)$.

To prove the reverse implication, assume $\neg(\phi \vee \psi)$. Then introduce ϕ as a secondary assumption. Now ϕ certainly implies $\phi \vee \psi$, and combining this with $\neg(\phi \vee \psi)$ in the usual way yields $\bot$. So, under the assumption of $\neg(\phi\vee\psi)$, we have proven that ϕ entails $\bot$, i.e., we have proven $\neg\phi$. We can prove $\neg\psi$ by an identical argument. So we conclude that $\neg(\phi \vee \psi)$ implies both $\neg\phi$ and $\neg\psi$, which is to say it implies $\neg\phi \wedge \neg\psi$. □

Surprisingly, the analogous statement with $\wedge$ and $\vee$ interchanged only works in one direction:

Exercise. Prove that $\neg\phi \vee \neg\psi$ implies $\neg(\phi \wedge \psi)$. Convince yourself that without using excluded middle there is no obvious way to deduce $\neg\phi \vee \neg\psi$ from $\neg(\phi \wedge \psi)$.

The next result shows that a weak form of excluded middle does always hold. Combining it with Proposition 9.2, we deduce that in the context of intuitionistic logic excluded middle is equivalent to double negation elimination.

Proposition 9.4. *We have $\neg\neg(\phi \vee \neg\phi)$.*

Proof. Assume $\neg(\phi\vee\neg\phi)$; we must derive a contradiction. But by Proposition 9.3 we can infer $\neg\phi\wedge\neg\neg\phi$, and then by Lemma 9.1, with $\neg\phi$ in place of ϕ, we can infer $\bot$. We have shown $\neg(\phi \vee \neg\phi) \to \bot$, i.e., $\neg\neg(\phi \vee \neg\phi)$. □

In the next proposition and exercise, assume that $\phi(x)$ contains a free variable x.

Proposition 9.5. *$(\forall x)(\neg\phi(x))$ is equivalent to $\neg(\exists x)\phi(x)$.*

Proof. Start by assuming $(\forall x)(\neg\phi(x))$. Then introduce $(\exists x)\phi(x)$ as a secondary assumption, with the goal of deriving a contradiction.

Infer $\neg\phi(x)$, i.e., $\phi(x) \to \bot$, from $(\forall x)(\neg\phi(x))$, and then combine this with $(\exists x)\phi(x)$ to infer $\bot$, as desired. We can now discharge the assumption

of $(\exists x)\phi(x)$ and conclude $((\exists x)\phi(x)) \to \bot$, i.e., $\neg(\exists x)\phi(x)$. This proves the forward implication.

To prove the reverse implication, assume $\neg(\exists x)\phi(x)$. Then introduce $\phi(x)$ as a secondary assumption, infer $(\exists x)\phi(x)$, and combine this with $\neg(\exists x)\phi(x)$ to get $\bot$. We can now discharge the assumption of $\phi(x)$ and conclude $\neg\phi(x)$. Finally, infer $(\forall x)(\neg\phi(x))$ from $\neg\phi(x)$. □

Here too, the analogous statement with $\forall$ and $\exists$ interchanged only works in one direction:

Exercise. Prove that $(\exists x)(\neg\phi(x))$ implies $\neg(\forall x)\phi(x)$.

And here are three more little exercises.

Exercise. Prove that $\phi \vee \psi$ and $\neg\phi$ together imply ψ.

Exercise. Prove that $\phi \to \psi$ and $\psi \to \theta$ together imply $\phi \to \theta$.

Exercise. Given $\phi \to \psi$, prove $\neg\psi \to \neg\phi$. If we have $\psi \vee \neg\psi$, we get the reverse implication from $\neg\psi \to \neg\phi$ to $\phi \to \psi$.

That concludes our brief introduction to intuitionistic logic. But let me also mention one very useful principle that is available to us as countablists, the *limited principle of omniscience (LPO)*.

(**LPO**) Let $\phi(n)$ and $\phi'(n)$ be predicates whose free variable ranges over the natural numbers. If $\phi(n) \vee \phi'(n)$ is true for all n, then either $\phi(n)$ is true for all n or $\phi'(n)$ is true for some n.

If $\phi'(n)$ is $\neg\phi(n)$, then the premise of the LPO is excluded middle for every $\phi(n)$, and the conclusion is that $\phi(n)$ is either true for all n or false for some n. This is probably the most common way the LPO is used.

We can informally justify the LPO in the following way. Suppose $\phi(n) \vee \phi'(n)$ is true for all n. For each n, there then exists a_n such that either $a_n = 0$ and $\phi(n)$ is true, or $a_n = 1$ and $\phi'(n)$ is true. (If $\phi(n)$ is true then such an a_n exists, and if $\phi'(n)$ is true then such an a_n exists. So the existence of such an a_n follows from $\phi(n) \vee \phi'(n)$.) We can then make the single infinitely long assertion that there exist a_1, a_2, ... with these properties. For countablists this is no more problematic than the inference of $(\exists a_1, \ldots, a_n)(\psi_1(a_1) \wedge \cdots \wedge \psi_n(a_n))$ from $(\exists a_1)\psi_1(a_1) \wedge \cdots \wedge (\exists a_n)\psi_n(a_n)$. An infinitely long search will then tell us whether $a_n = 0$ for all n or $a_n = 1$ for at least one n; in the former case ϕ_n is true for all n, and in the latter case ϕ'_n is true for some n.

We also have two "choice" principles with similar countablistic justifications; see the end of the next section.

9.3 Classes and sets

In standard mathematical foundations ordinary mathematical structures such as the natural numbers, the real line, and so on, are "encoded" as sets. Sets are the fundamental objects. For some finitists (famously, Kronecker), the fundamental objects are the natural numbers, and everything else is built up from them. But we are countablists. For us it is natural to build up the mathematical universe from infinite bitstrings — infinite strings each of whose entries, or *bits*, is either "on" or "off". I will just call them *bitstrings*. The entries of a bitstring are indexed by the positive natural numbers 1, 2, ... I assume that we come to the table with a basic ability to reason about natural numbers, but, in a sense, they are not foundational objects in our universe.

Sometimes we will want to "identify" different bitstrings and treat identified bitstrings as a single object. Such an object could be called "a bitstring modulo equivalence". Every such object is *instantiated* by at least one, but perhaps several or even infinitely many, bitstrings.

A *class*, for us, will be a collection of bitstrings modulo some notion of equivalence. Not all bitstrings need be included. Thus, we specify a class by saying which bitstrings are to be included and which pairs of them are to be identified. This can be done in a single step using an *identity criterion* given by a formula θ with exactly two free variables, both of which range over all bitstrings. It must be *symmetric* ($\theta(\alpha, \beta)$ implies $\theta(\beta, \alpha)$) and *transitive* ($\theta(\alpha, \beta)$ and $\theta(\beta, \gamma)$ together imply $\theta(\alpha, \gamma)$). A bitstring α instantiates an element of the associated class if $\theta(\alpha, \alpha)$ holds, and two bitstrings α and β instantiate the same element if $\theta(\alpha, \beta)$ holds.

We will also require that θ must be *definite*, meaning that it definitely does or does not hold of any given bitstrings α and β. This has to be said because, in constructivism, we do not assume that every statement has a definite truth value. But it will just be part of our definition of classes that the identity criteria that give rise to them must satisfy excluded middle. Thus, an identity criterion is a formula with two free variables that range over all bitstrings and which is symmetric, transitive, and definite. Although we reason intuitionistically, including definiteness in this definition gives us a lot of nice excluded middle. For example, any two elements of a class definitely are or are not equal, because the relevant identity criterion definitely does or does not hold of any pair of bitstrings which instantiate them.

If X is a class and x is a bitstring modulo some equivalence, I will write

$x \in X$ and say that x *is an element of* X or *belongs to* X when there is a bitstring α satisfying $\theta(\alpha, \alpha)$, such that x is instantiated by precisely those bitstrings β for which $\theta(\alpha, \beta)$ holds, where θ is an identity criterion for X.

Let X and Y be classes. We say that Y is a *subclass of* X, or Y is *contained in* X, and write $Y \subseteq X$, if every element of Y is also an element of X. The containment is *proper*, written $Y \subset X$, if there exists an element x of X that is not an element of Y, written $x \notin Y$. In terms of identity criteria, if θ_X and θ_Y are identity criteria for X and Y then the condition for containment is that whenever $\theta_Y(\alpha, \alpha)$ holds for some bitstring α, we have $\theta_X(\alpha, \beta) \leftrightarrow \theta_Y(\alpha, \beta)$ for all bitstrings β. Thus, if α instantiates an element of Y then (taking $\beta = \alpha$) it also instantiates an element of X, and the bitstrings that instantiate the same element of X are precisely those bitstrings that instantiate the same element of Y. It is easier to simply say "$x \in Y$ implies $x \in X$", and that is the kind of language that I will transition to after we get the basics out of the way. The point is that we only say that some $x \in Y$ belongs to X if there is an element of X that is instantiated by exactly the same bitstrings that instantiate x. Two classes are equal if each is contained in the other.

Crucially, if Y is a subclass of X then every element x of X definitely either does or does not belong to Y ($x \in Y$ or $x \notin Y$). This is because x is instantiated by a bitstring α and we have $\theta_Y(\alpha, \alpha) \vee \neg\theta_Y(\alpha, \alpha)$.

If θ_X is an identity criterion for X, one way to specify a subclass of X is to start with a predicate ϕ whose free variable ranges over all bitstrings and is compatible with θ_X in the sense that

$$\theta_X(\alpha, \beta) \rightarrow (\phi(\alpha) \leftrightarrow \phi(\beta)).$$

If in addition ϕ is definite, meaning that we have $\phi(\alpha) \vee \neg\phi(\alpha)$ for all bitstrings α, then

$$\theta_X(\alpha, \beta) \wedge \phi(\alpha)$$

is an identity criterion for a subclass of X. Indeed, every subclass Y of X can arise in this way; just take $\phi(\alpha) = \theta_Y(\alpha, \alpha)$ for some identity criterion θ_Y for Y.

In practice, we usually ensure compatibility with θ_X by working with predicates whose free variable ranges not over all bitstrings, but over the elements of X. Then all we need to check is that ϕ is well-defined and definite; given this, we can always form a subclass $\{x \in X : \phi(x)\}$ of X whose elements are just those that satisfy ϕ. This principle is known as *class comprehension*.

For instance, if X_1 and X_2 are subclasses of a class X, then their *difference* $X_1 \setminus X_2 = \{x \in X_1 : x \notin X_2\}$ is also a subclass of X. The condition "$x \notin X_2$" is well-defined and definite. If X_2 is contained in X_1, we may call $X_1 \setminus X_2$ the *relative complement* of X_2 in X_1, and when $X_1 = X$ we may call it the *complement* of X_2 and denote it X_2^c. An element of X belongs to X_2^c if and only if it does not belong to X_2.

The *union* of X_1 and X_2 is the class $X_1 \cup X_2 = \{x \in X : x \in X_1 \text{ or } x \in X_2\}$ and their *intersection* is the class $X_1 \cap X_2 = \{x \in X : x \in X_1 \text{ and } x \in X_2\}$. Again, the defining conditions are well-defined and definite. X_1 and X_2 are *disjoint* if their intersection is the empty class $\emptyset$.

The product of two classes is another basic construction. For any two bitstrings α and β, define $2^\alpha 3^\beta$ to be the bitstring whose 2^ath bit is on whenever the ath bit of α is on, and whose 3^bth bit is on whenever the bth bit of β is on, and whose other bits are all off. We then define the *product* of two classes X and Y to be the class $X \times Y$ whose elements are all the bitstrings $2^\alpha 3^\beta$ such that α instantiates an element of X and β instantiates an element of Y, setting $2^\alpha 3^\beta$ and $2^{\alpha'} 3^{\beta'}$ equivalent if α and α' instantiate the same element of X and β and β' instantiate the same element of Y. Thus the identity criterion is: θ holds of two bitstrings if they have the forms $2^\alpha 3^\beta$ and $2^{\alpha'} 3^{\beta'}$ for some bitstrings α, β, α', β' which satisfy $\theta_X(\alpha, \alpha')$ and $\theta_Y(\beta, \beta')$. The LPO guarantees that every bitstring definitely does or does not have the form $2^\alpha 3^\beta$ for some α and β. For any elements $x \in X$ and $y \in Y$, there is an element (x, y) of $X \times Y$, an *ordered pair*, which is instantiated by exactly those bitstrings of the form $2^\alpha 3^\beta$ with α instantiating x and β instantiating Y. Conversely, every element of $X \times Y$ has this form for exactly one $x \in X$ and $y \in Y$.

Now let us say when a class is a set. First, for any bitstring α and any prime p, let α_p be the bitstring whose ath bit is on if and only if the p^ath bit of α is on. Also, let p_n be the nth prime (so $p_1 = 2$, $p_2 = 3$, and so on). If X is a class with identity criterion θ, we then define the *sequence space* X^ω to be the class whose identity criterion holds of two bitstrings α and β if and only if (1) the only on bits in α and β are prime powers, and (2) for each n we have $\theta(\alpha_{p_n}, \beta_{p_n})$. These conditions are definite by the LPO. If α instantiates an element of X^ω then for each n the bitstring α_{p_n} instantiates an element of X, and α and β instantiate the same element of X^ω if and only if α_{p_n} and β_{p_n} instantiate the same element of X for all n. We call elements of X^ω *sequences in* X and write (x_n) or $\vec{x}$ for the sequence instantiated by α, where, for each n, x_n is the element of X instantiated by α_{p_n}.

The intuition is that an element of X^ω encodes a sequence of elements of X, with the nth one living on the powers of the nth prime. However, X^ω is not a new kind of object; it is a class, just like X.

We say that X is a *set* if it is either empty or there is an *exhaustive* sequence in X, a sequence (x_n) in X with the property that every element of X equals x_n for some n. We also say that X is *countable* or *enumerable*. The elements of any nonempty set can be enumerated, possibly with repetition, in an infinite list as x_1, x_2, ... Since any two elements definitely are or are not equal, we can ensure that there is no repetition by removing any duplicates; this could, of course, reduce the enumeration to a finite list. However, except where otherwise noted, we will allow enumerations to include duplicate elements.

As our first example of a set, let $\mathbb{N}$ be the class whose elements are all the bitstrings that have exactly one on bit, with no identification of distinct bitstrings. We have already been using the natural numbers informally, but now they officially belong to our universe. The natural number n has been "encoded" as the bitstring whose nth bit is on and all of whose other bits are off. It is clear that $\mathbb{N}$ is a set.

(Formally, what we are developing in this chapter is a version of third order arithmetic, with natural numbers as first order objects, bitstrings = sets of natural numbers as second order objects, and classes as third order objects. From this perspective the natural numbers *are* fundamental for us. However, the first order natural numbers will soon completely disappear from the exposition; all of the mathematical content in subsequent chapters will be couched in terms of classes of bitstrings modulo equivalence. That is, once the foundational material is out of the way, we will only be talking about third order objects as collections of second order objects. What I have just done is to encode the first order natural numbers as second order bitstrings, so that the collection of all of them will be a third order object. There is a basic terminological discrepancy between the two parts of this book: in Part 1, "sets" and "classes" are conceptually distinct, but in Part 2 a set is a special kind of class. The "sets" of Part 1 correspond to second order objects in the setup used here.)

A class X is *proper* or *uncountable* if for every sequence (x_n) in X there is an element of X which is not equal to x_n for any n. For example, let the *universal class* be the class Ω whose elements are all bitstrings, with no identification of distinct bitstrings.

Theorem 9.6. *Ω is a proper class.*

Proof. Given any sequence (α_n) in Ω, define a new bitstring α whose nth bit is on if and only if the nth bit of α_n is off, for all n. Then α cannot equal any α_n. We conclude that Ω is a proper class. □

To a large degree, classical logic can be used when reasoning about sets. For instance, any two sets contained in the same class definitely are or are not equal. That is a consequence of the LPO together with the fact that any two elements of a class definitely are or are not equal. More generally, if A and Y are subclasses of a class X and A is a set, then A definitely is or is not contained in Y. In contrast, we cannot always say that two subclasses of a class definitely are or are not equal, or that one definitely is or is not contained in the other, because they could be uncountable and the LPO would not apply. Another example is that it is built into the definition of a set that every set definitely is or is not empty, whereas we cannot always say this of a given class — determining nonemptiness might require searching through all bitstrings, but as we saw just above, there are uncountably many of them — nor can we always say that a given class definitely is or is not a set.

Since every set definitely is or is not empty, for sets "nonempty" is the same as "there exists at least one element", but for classes these are not the same. We say that a class is *inhabited* if it has at least one element. To say that a class is nonempty is to say that it is not not inhabited.

If a subclass of a class is a set, then we can also call it a *subset*. One situation where this will have to happen is when the containing class is itself a set, because then the subclass will automatically be a set too.

Proposition 9.7. *Let A be a set and let B be a subclass of A. Then B is a set.*

Proof. Since each element of A definitely is or is not an element of B, it follows from the LPO that B definitely is or is not empty. If B is empty there is nothing to prove; otherwise fix $y \in B$ and let (x_n) be an exhaustive sequence in A. We can then define an exhaustive sequence (y_n) in B by setting $y_n = x_n$ when $x_n \in B$ and $y_n = y$ when $x_n \notin B$. This shows that B is a set. □

We can generalize the union and intersection of two subclasses to countably many subclasses. An easy way to talk about infinitely many subclasses of a class X is with an *indexed family of subclasses* of X. This is simply a subclass $\mathcal{F}$ of $X \times I$ for some *index class* I. Each $\lambda \in I$ determines a

subclass $X_\lambda = \{x \in X : (x, \lambda) \in \mathcal{F}\}$ called the λ-*slice* of $\mathcal{F}$. We can also call λ a *code* for X_λ.

This "indexing" interpretation of products has far-ranging consequences for us. In our setup, classes cannot be elements of other classes. The elements of a class are always bitstrings modulo equivalence. However, when $\mathcal{F} \subseteq X \times I$ is an indexed family of classes we can effectively quantify over all the slices X_λ simply by quantifying over I, so that, in relation to $\mathcal{F}$, the class I to some extent functions as a collection of classes. We can harness this intuition by referring to I as a *pseudoclass* and using the language of elementhood — saying that X_λ "belongs to" I and writing "$X_\lambda \,\varepsilon\, I$" with the variant "$\varepsilon$" elementhood symbol. The literal meaning is just "$\lambda \in I$", but the point is to talk directly about the slices without mentioning the codes. I'll come back to this in Section 9.5.

If the slices are distinct, i.e., $X_\lambda \neq X_{\lambda'}$ whenever $\lambda \neq \lambda'$, then we can call I an *exact pseudoclass*. The exact pseudoclasses I will use in this book are: $\mathrm{fn}(A, X)$ and $\mathcal{P}(X)$ (Section 9.8); $C(X, Y)$ and $\mathrm{Homeo}(X)$ (Sections 10.3 and 10.5); $\hom(B, \{0, 1\})$ (Section 11.5); $\mathcal{L}(V, W)$ and $\mathrm{Sub}(V)$ (Section 12.2); $\mathcal{M}_0(X)$ (Section 15.1); $B(E, F)$, $B(E)$, and E^* (Section 15.3).

Returning to unions and intersections, if $\mathcal{F} \subseteq X \times I$ is a countably indexed family of subclasses of X, meaning that I is a nonempty set, with slices X_λ, then we write $\mathcal{F} = (X_\lambda)$ and define the *union* and *intersection* of the X_λ by

$$\bigcup X_\lambda = \{x \in X : x \in X_\lambda \text{ for some } \lambda\}$$

and

$$\bigcap X_\lambda = \{x \in X : x \in X_\lambda \text{ for all } \lambda\}.$$

These exist as subclasses of X by class comprehension. The conditions "$x \in X_\lambda$ for some λ" and "$x \in X_\lambda$ for all λ" are definite by the LPO; that is why we need I to be countable.

When each X_λ is a set we can say something more.

Theorem 9.8. *The union and intersection of a countably indexed family of subsets of a class are both sets.*

Proof. Let X be a class and let $\mathcal{F} \subseteq X \times I$ be a countably indexed family of subsets of X. Let (λ_n) be an exhaustive sequence in I and for each n let A_n be the λ_n-slice.

Since A_1 is a set, the intersection of the A_n is a class contained in a set and is therefore itself a set by Proposition 9.7. For the union, observe

first that either every A_n is empty or at least one is nonempty, by the LPO. If they are all empty then we are done; otherwise, let A be one of the nonempty ones and for each n let $A'_n = A_n$ if A_n is nonempty, $A'_n = A$ if A_n is empty. This does not affect the union and reduces us to the case where every A_n is nonempty.

Fix an exhaustive sequence in each A'_n. Then we can exhibit an exhaustive sequence in the union by initially listing the first entry in the sequence for A'_1, following that by the first entry for A'_2 and the second for A'_1, then the first for A'_3, second for A'_2, and third for A'_1, and so on. The result will be an exhaustive sequence in $\bigcup A'_n = \bigcup A_n$. So the union is also a set. □

In this proof we work with a sequence of sequences (one exhaustive sequence for each A_n), which can be realized as an element of $(X^\omega)^\omega$. We know that each A_n has an exhaustive sequence, and we can then, as countablists, infer the existence of a sequence of exhaustive sequences of the A_n. This is an example of *countable choice*.

(**Countable choice**) Let X be a class and let $\phi(n, x)$ be a formula with two free variables, the first ranging over the natural numbers and the second ranging over X. If for every n there exists $x \in X$ such that $\phi(n, x)$ is true, then there exists a sequence (x_n) in X such that $\phi(n, x_n)$ is true for all n.

The countablistic justification here is quite direct; there is an obvious infinitely long derivation that produces the desired sequence (x_n).

We also have a stronger choice principle called *dependent choice*.

(**Dependent choice**) Let X be a class and let $\phi(n, x, y)$ be a formula with three free variables, the first ranging over the natural numbers and the other two ranging over X. If for every natural number n and every $x \in X$ there exists $y \in X$ such that $\phi(n, x, y)$ is true, then every $x_1 \in X$ is the initial term of a sequence (x_n) in X with the property that $\phi(n, x_n, x_{n+1})$ is true for all n.

9.4 Meaningfulness and assertibility

Now we need to talk about what kinds of objects classes are, because this significantly affects how we can reason about them.

Bitstrings can be conceived in very concrete terms. We can think of a bitstring as an infinite procession of light bulbs, say, each one of which may be either on or off. The fact that bitstrings are infinitely long is not

a problem for us as countablists. If there are any philosophical problems with thinking of bitstrings in this kind of way, we can leave them to the philosophers. They are not mathematical problems.

Classes cannot be understood in such a concrete way, however. If a class is uncountable, we do not accept that there is any possible way to explicitly list out its elements. The only way to specify an uncountable class is through some kind of verbal description of it, by saying which pairs of bitstrings satisfy the identity criterion. Effectively, a class *is* an identity criterion modulo holding of the same pairs of bitstrings. *Classes are linguistic objects.*

(This issue is not particular to countabilists. One of the main themes of the first part of this book is that whatever one's views on infinity are, there will be "proper classes", collections which can only be identified linguistically because it is, in principle, not possible to explicitly present all of their elements. See Chapter 3.)

This matters because linguistic objects can be tricky in ways that things like infinite bitstrings are not. They can be self-referential and paradoxical. This demands extra care when reasoning about them.

Linguistic paradoxes often hinge on questions of meaningfulness. For instance, if we could affirm that the sentence "this sentence is not true" were meaningful then we could derive a contradiction, but affirming that it has no meaning also leads to a contradiction because meaningless sentences cannot be true. Constructivism is essential to handling this kind of problem: we evade disaster only by not affirming that the sentence definitely is or is not meaningful.

I discussed this issue in Chapter 6, and more thoroughly in Chapters 3 to 5. What matters here is that classes are represented by formulas with two free variables, and not every such formula is necessarily meaningful. But we cannot draw a sharp line between those formulas that are meaningful and those that are not. There are two ways to deal with this problem. One is to fix a particular language all of whose formulas are known to be meaningful and work only within it. For many purposes this is a perfectly adequate way to do things, but it lacks universality. In this approach we cannot prove anything about all classes, only about all classes represented by formulas from the chosen language. If we want to be able to say something as basic as "every class is contained in itself", and have this include any class that can be verbally described in any language, then we have to quantify over all formulas without restriction and accept that some of them might not be meaningful.

How do we reason about possibly meaningless statements? For example, consider that $\phi \vee \psi$ is true if either ϕ is true or ψ is true. Taking ψ to be a sentence which is known to be true — "$0 = 0$", say — it seems like we should therefore be able to affirm as true any sentence of the form $\phi \vee (0 = 0)$. But we cannot do this if ϕ might be meaningless. Suppose we are unsure whether "borogroves are mimsy" is a meaningful sentence, for instance. Then we cannot affirm

> either borogroves are mimsy or $0 = 0$,

because we do not even know that this is a meaningful assertion. In order for a sentence to be meaningful, all of its parts have to be meaningful. Not knowing whether "borogroves are mimsy" is meaningful, the best we can do is

> if "borogroves are mimsy" were a meaningful sentence, then we could affirm "either borogroves are mimsy or $0 = 0$".

We must reason in the subjunctive mood.

In more formal terms, the pattern we have to follow is

$$\mathbb{M}[\text{“borogroves are mimsy”}] \to \mathbb{A}[\text{“either borogroves are mimsy or } 0 = 0\text{”}]$$

where "$\mathbb{M}$" expresses "is meaningful" and "$\mathbb{A}$" expresses "is assertible".

Now, reasoning normally only involves sentences that we know are meaningful. But when we quantify over all sentences, we have to reason in the subjunctive mood because there is no way to neatly separate out exactly those which are meaningful. So if ϕ ranges over all sentences, including those which might not be meaningful, we have to say things like

$$(\forall \phi)(\mathbb{M}[\phi] \to \mathbb{A}[\phi \vee (0 = 0)]).$$

Or, a little more briefly (but with the same sense),

$$(\forall \phi : \mathbb{M}[\phi])\mathbb{A}[\phi \vee (0 = 0)].$$

Similarly, definitions that only make sense for meaningful sentences have to be made subjunctively. For example, Tarski showed us how to define the truth of any meaningful sentence: we simply follow the template

> "#" is true if and only if #,

e.g., "snow is white" is true if and only if show is white. This instance is definitely meaningful, because "snow is white" is a meaningful sentence, but the instance

> "borogroves are mimsy" is true if and only if borogroves are mimsy

only works as a definition if "borogroves are mimsy" means something. So Tarski's definition is properly formulated in full generality as

$$(\forall \phi : \mathbb{M}[\phi])\mathbb{A}[\mathbb{T}[\phi] \leftrightarrow \phi]$$

where ϕ ranges over all sentences, meaningful or not. Informally: whenever ϕ is a meaningful sentence, we can affirm the equivalence of ϕ with the affirmation that ϕ is true.

In Chapter 4 I discussed how to reason with the assertibility predicate $\mathbb{A}$. First, we can reason "under $\mathbb{A}$" in the same way that we normally reason. Any normal inference of a conclusion from some premises can also function as an inference that the conclusion is assertible from the assertibility of the premises (Theorem 4.1). The two principles that allow us to pass in and out of the assertibility predicate are the *capture scheme*

$$\phi \rightarrow \mathbb{A}[\phi]$$

and the *release rule*

$$\text{given } \mathbb{A}[\phi], \text{ infer } \phi,$$

where ϕ can be any meaningful sentence. The capture scheme is a restatement of the core principle of truth constructivism, that every fact can be known, and the release rule expresses the permissibility of affirming ϕ once we know how to prove it. (If we are reasoning using natural deduction, this step can only be taken if no temporary premises are in play. We can only infer ϕ if we actually have a proof of ϕ in hand. See Sections 4.2 and 8.6 for more about this.) The practical implications are that anything we can prove outside of the assertibility predicate can be brought under it, and once we are in a concrete situation where we know that something is assertible we can simply assert it. Thus for any particular sentence ϕ, the definition of truth given just above is $\mathbb{M}[\phi] \rightarrow \mathbb{A}[\mathbb{T}[\phi] \leftrightarrow \phi]$, and if we know that this sentence is meaningful we can infer $\mathbb{A}[\mathbb{T}[\phi] \leftrightarrow \phi]$, and then, using the release rule, get Tarki's biconditional $\mathbb{T}[\phi] \leftrightarrow \phi$.

A good deal of what I have already done in this chapter was subjunctive. In Section 9.1 I mentioned induction, the principle that $\phi(1)$ and $(\forall n)(\phi(n) \rightarrow \phi(n+1))$ together imply $(\forall n)\phi(n)$. But this is only for properties ϕ which meaningfully apply to all natural numbers. What we really should say is

$$(\forall \phi : \mathbb{M}_N[\phi])\mathbb{A}[(\phi(1) \wedge (\forall n)(\phi(n) \rightarrow \phi(n+1))) \rightarrow (\forall n)\phi(n)],$$

where the quantification is over all predicates ϕ and $\mathbb{M}_N$ means "is meaningful when applied to any natural number". For any particular predicate

ϕ which is known to meaningfully apply to all natural numbers, we can eliminate the premise $\mathbb{M}_N[\phi]$ and then apply the release rule to get simply

$$(\phi(1) \wedge (\forall n)(\phi(n) \to \phi(n+1))) \to (\forall n)\phi(n).$$

The LPO, countable choice, and dependent choice are also subjunctive when stated in full generality.

The rules of natural deduction can be seen in this light as well. E.g., if ϕ and ψ are both assertible then so is $\phi \wedge \psi$; if $\phi \vee \psi$, $\phi \to \theta$, and $\psi \to \theta$ are all assertible then so is θ; and so on. We do not need to add meaningfulness as a separate premise in these cases because assertibility entails meaningfulness.

If one is working with a specific target language in which everything is meaningful, then all of these things can be said directly, in the way I initially introduced them, without any reference to meaningfulness or assertibility. It is only when reasoning about *anything that could be said*, in any meaningful language — or when reasoning within a language not all of whose sentences are definitely meaningful, as is the case with the classical semantic paradoxes — that meaningfulness and assertibility need to be made explicit.

9.5 Reasoning about classes

Since classes are linguistic objects, we need to use assertibility to talk about arbitrary classes. Let θ be a formula with exactly two free variables. In Section 9.3 I said what it means for θ to be an identity criterion: it must be symmetric, transitive, and definite. But these conditions can only be stated if θ meaningfully applies to any two bitstrings. Let me also add here that two identity criteria instantiate the same class if and only if they hold of the same pairs of bitstrings. We can make these definitions under $\mathbb{A}$ as

$$(\forall \theta : \mathbb{M}_2[\theta])\mathbb{A}[\mathrm{Class}(\theta) \leftrightarrow (\mathrm{Sym}(\theta) \wedge \mathrm{Trans}(\theta) \wedge \mathrm{Def}(\theta))]$$

and

$$(\forall \theta, \theta' : \mathbb{M}_2[\theta] \wedge \mathbb{M}_2[\theta'])\mathbb{A}[(\theta \sim \theta') \leftrightarrow (\forall \alpha, \beta)(\mathbb{T}[\theta(\alpha, \beta)] \leftrightarrow \mathbb{T}[\theta'(\alpha, \beta)])]$$

where θ and θ' range over all formulas with two free variables and $\mathbb{M}_2[\theta]$ means "θ meaningfully applies to any two bitstrings". $\mathrm{Class}(\theta)$ means "θ instantiates a class", $\mathrm{Sym}(\theta)$ means "for all bitstrings α and β, θ holds of α and β (i.e., $\theta(\alpha, \beta)$ is true) if and only if θ holds of β and α, and $\mathrm{Trans}(\theta)$

and $\mathrm{Def}(\theta)$ are defined similarly. We need to use a truth predicate here for reasons explained in Sections 3.4 and 3.5.

When I say that two identity criteria instantiate the same class, this expresses an intention to treat them as a single object. Notationally, I will write θ, θ', etc. for identity criteria where the notion of equality is syntactic identity and X, Y, etc. for the classes they instantiate. At a formal level, however, nothing needs to happen. We do not need to introduce a new category of objects into our language; we can keep talking about identity criteria and simply reinterpret "equality" as the condition $(\forall\alpha,\beta)(\mathbb{T}[\theta(\alpha,\beta)] \leftrightarrow \mathbb{T}[\theta'(\alpha,\beta)])$. A similar comment can be made about bitstrings versus bitstrings modulo some identity criterion. When we are working with an identity criterion θ on bitstrings, all we need to do formally is to interpret quantification as "for all α such that $\theta(\alpha,\alpha)$ is true" and "there exists α such that $\theta(\alpha,\alpha)$ is true" and interpret equality as "$\theta(\alpha,\beta)$ is true". The informal language refers to new objects, but these do not need to be introduced formally (though they could be, if we wanted to).

The way we make universal statements about all classes, formally, is with sentences of the form

$$(\forall\theta : \mathbb{M}_2[\theta])\mathbb{A}[\mathrm{Class}(\theta) \to \phi].$$

Informally, we just say "for every class X, ϕ". In this informal expression the meaningfulness hypothesis is implicit, as is the fact that ϕ is being dealt with subjunctively (i.e., under $\mathbb{A}$). Informal universal quantification over multiple classes, as in "for any classes X and Y, ϕ", can be formally rendered as

$$(\forall\theta,\theta' : \mathbb{M}_2[\theta] \wedge \mathbb{M}_2[\theta'])\mathbb{A}[(\mathrm{Class}(\theta) \wedge \mathrm{Class}(\theta')) \to \phi],$$

and so on.

When we quantify over all classes, we are really quantifying over all formulas with two free variables and then stating what we could say if they instantiated classes. This subjunctive reasoning is formally expressed using an assertibility predicate. But one has to be careful about mixing ordinary and subjunctive reasoning. For instance, $\mathbb{A}[\phi \to \psi] \wedge \mathbb{A}[\phi]$ implies $\mathbb{A}[\psi]$ — if we could affirm both ϕ and $\phi \to \psi$, then we could affirm ψ — and this shows that $\mathbb{A}[\phi \to \psi]$ implies $\mathbb{A}[\phi] \to \mathbb{A}[\psi]$. But the reverse inference, from $\mathbb{A}[\phi] \to \mathbb{A}[\psi]$ to $\mathbb{A}[\phi \to \psi]$, is not legal. Likewise, $\mathbb{A}[\phi] \vee \mathbb{A}[\psi]$ implies $\mathbb{A}[\phi \vee \psi]$, but we do not have the reverse implication. See Section 4.5. Since we have to follow any universal quantification over all classes by an assertibility

predicate, there is a potential issue of dealing with statements involving multiple, possibly even nested, assertibility predicates. The simple solution for us will be to only work with universal statements about classes of the type described just above ("for any classes X and Y, ϕ"). It is helpful here to distinguish between "unbounded" quantifiers ($\forall X$) which quantify over all classes and "bounded" quantifiers ($\forall x \in X$) which quantify over all the elements of some class. An unbounded quantifier $\forall X$ must be followed by an assertibility predicate, under which $\forall x \in X$ can then appear anywhere. Once we are working under $\mathbb{A}$, we can reason using bounded quantifiers however we like.

It is a remarkable fact that everything we do mathematically throughout the rest of this book can be framed in the form of a string of universal unbounded quantifiers followed by a statement that only involves bounded quantifiers sitting inside a single assertibility predicate. The general proof strategy will be to fix one or more arbitrarily chosen classes and then reason normally under $\mathbb{A}$. Reasoning intuitionistically, of course, although the definiteness requirement for identity criteria means that excluded middle will often be available. The idea is that when we execute this normal reasoning we never need to go beyond the fixed, arbitrarily chosen classes and other classes that can be constructed from them.

(The one exception is when we are only talking about some specific class or classes and there are no unbounded universal quantifiers at all. Then no assertibility predicate is needed.)

In Section 9.3 I mentioned the possibility, given an indexed family of subclasses $\mathcal{F} \subseteq X \times I$ of some class X, of effectively quantifying over all the subclasses X_λ in the family simply by quantifying over I. I can now make the point that this will be a bounded quantifier, even if I is uncountable. So, anticipating the next chapter, we will be able to reason under $\mathbb{A}$ about all the open subclasses of an arbitrary topological space, for instance. The open subclasses of a topological space will be specified by an indexed family, so that once an arbitrary topological space is fixed, we will have a pseudoclass of all of its open subclasses and will be able deal with them using bounded quantifiers.

What about unbounded existential quantifiers? Sometimes these can be eliminated by actually exhibiting a class with the desired property instead of saying that such a class merely exists. And that is the only kind of unbounded existential quantification we will really need. Thus whenever I say "there exists a class" it will always be a positive statement of the form "under such and such conditions there exists a class with such and such

properties — namely, the class described by the following explicit formula ...". For instance, as I already noted, for any class X and any subclass Y of X, there exists a complementary subclass Y^c of X such that for all $x \in X$ we have $x \in Y^c \leftrightarrow x \notin Y$. This complementary subclass is described by the explicit formula $\{x \in X : x \notin Y\}$, so the word "exists" is not needed in this statement. We could just as well say that for any class X and any subclass Y of X, the predicate "$x \notin Y$" is definite and therefore instantiates a subclass Y^c of X with the stated properties. (More formally: start with "$\forall \theta, \theta' : \mathbb{M}_2[\theta] \wedge \mathbb{M}_2[\theta']$" and then, under $\mathbb{A}$, say "if θ instantiates a class X and θ' instantiates a subclass Y of X, then $Y^c = \{x \in X : x \notin Y\}$ is a subclass of X, etc.") Now, in practice we often do not care so much about the explicit description of a class with some property, and it may be easier to just say "there exists a class with such and such property". But there should always be an explicit description at the ready which could make the unbounded existential quantifier formally superfluous.

When reasoning under $\mathbb{A}$, there is no restriction on making unbounded existence statements that are backed up by explicit descriptions. They can appear anywhere. When we want an unbounded existential quantifier to appear outside of $\mathbb{A}$, in the initial string of quantifiers, as in "$(\forall X)(\exists Y)(\forall Z)\phi(X, Y, Z)$", we can formally eliminate the $\exists Y$ by writing something like

$$(\forall \theta, \theta' : \mathbb{M}_2[\theta] \wedge \mathbb{M}_2[\theta'])\mathbb{A}[(\mathrm{Class}(\theta) \wedge \mathrm{Class}(\theta')) \rightarrow (\mathrm{Class}(\theta_Y) \wedge \phi(\theta, \theta_Y, \theta'))]$$

where θ_Y is the explicit description of the desired class Y (which, in this case, can depend on θ but not on θ').

(Another comment that might be worth making is that any statement of the form $(\exists x)\phi(x) \rightarrow \psi$ is intuitionistically equivalent to $(\forall x)(\phi(x) \rightarrow \psi)$. So these kinds of existential quantifiers can be removed without doing any work at all.)

In summary: every theorem for the rest of this book can be formally rendered either as a statement involving only bounded quantifiers ranging over some explicitly identified classes, or as a string of unbounded universal quantifiers followed by a statement with only bounded quantifiers lying under an assertibility predicate. Every unbounded existential quantifer can be removed in favor of an explicit formula instantiating the desired class, and the proof of every theorem will proceed with normal inuitionistic reasoning under $\mathbb{A}$.

9.6 Relations and functions

Definition 9.9. Let X and Y be classes. A *relation on X and Y* is a subclass of $X \times Y$. If $X = Y$ we simply call it a *relation on X*.

The subclass requirement implies that for any $x \in X$ and $y \in Y$ it definitely either is or is not the case that the ordered pair (x, y) is an element of the given relation.

When R is a relation we can write xRy to indicate that R contains the ordered pair (x, y). If $X' \subseteq X$ and $Y' \subseteq Y$ are subclasses of X and Y, then the intersection $R \cap (X' \times Y')$ is the *restriction* of R to X' and Y'.

Definition 9.10. A relation R on a class X is *reflexive* if xRx holds for all $x \in X$, *symmetric* if xRy implies yRx, *antisymmetric* if xRy and yRx together imply $x = y$, and *transitive* if xRy and yRz together imply xRz.

Definition 9.11. Let X and Y be classes. A *function* from X to Y is a relation on X and Y with the property that for every $x \in X$ there is a unique ordered pair in the relation whose first element is x. X is its *domain* and Y is its *range*.

This definition employs the quantifier "there exists a unique", which can be written $\exists!$. Thus $(\exists! x)\phi(x)$ means "there is a unique x such that $\phi(x)$". In intuitionistic logic it is cashed out as $(\exists x)(\phi(x) \wedge (\forall y)(\phi(y) \to y = x)))$, but since (as noted earlier) any two elements of a class definitely are or are not equal, we are free to use the classically equivalent formulation $(\exists x)(\phi(x) \wedge (\forall y)(y \neq x \to \neg\phi(y)))$.

When f is a function we write $f(x) = y$ instead of xfy. We may also write $x \mapsto f(x)$. Here are two basic facts about functions:

Proposition 9.12. *Let X, Y, and Z be classes and let $f : X \to Y$ and $g : Y \to Z$ be functions. Then there is a function $g \circ f : X \to Z$ satisfying $(g \circ f)(x) = g(f(x))$ for all $x \in X$. It is defined by setting $(g \circ f)(x) = z$ if there exists $y \in Y$ such that $f(x) = y$ and $g(y) = z$.*

The condition defining $g \circ f$ is definite because for each $x \in X$ there exists exactly one $y \in Y$ such $f(x) = y$, and then for each $z \in Z$ it definitely will or will not be the case that $g(y) = z$. (This reasoning hinges on the fact that f is a function. The composition of two relations need not always be a well-defined relation: if we have to search through uncountably many values of y looking for one with (x, y) in the first relation and (y, z) in the

second, we may not be able to say that such a y definitely does or does not exist.)

Proposition 9.13. *Let X and Y be be classes, let $f : X \to Y$ be a function, and let X_0 be a subclass of X. Then the elements of f whose first entries lie in X_0 constitute a function from X_0 into Y.*

The function described in Proposition 9.13 is the *restriction* of f to X_0. We denote it $f|_{X_0}$. More simply, $f|_{X_0}$ is the intersection of f with $X_0 \times Y \subseteq X \times Y$.

Definition 9.14. A function $f : X \to Y$ between classes is *1-1* or *injective* if $f(x) = f(y)$ implies $x = y$. (Again, we can equivalently say "$x \neq y$ implies $f(x) \neq f(y)$" since we have excluded middle for $x = y$.) It is *onto* or *surjective* if for every $y \in Y$ there exists $x \in X$ with $f(x) = y$, and *subjective* if there exists $y \in Y$ such that no $x \in X$ satisfies $f(x) = y$ (equivalently, every $x \in X$ satisfies $f(x) \neq y$). It is a *1-1 correspondence* or *bijection* if it is both 1-1 and onto.

Classically, "surjective" and "subjective" are each other's negations, but intuitionistically neither is exactly the negation of the other. Each does imply the other's negation.

If $f : X \to Y$ is a bijection we may write $f : X \cong Y$. We also write f^{-1} for the bijection from Y to X obtained by switching the elements in each ordered pair in f.

Next we will discuss images and preimages. First, preimages.

Proposition 9.15. *Let X and Y be classes, let $f : X \to Y$ be a function, and let Y_0 be a subclass of Y. Then there is a subclass $X_0 \subseteq X$ with the property that $x \in X_0$ if and only if $f(x) \in Y_0$.*

Proof. For every $x \in X$, there is a unique $y \in Y$ with $f(x) = y$, and this y definitely either does or does not belong to Y_0. So $\{x \in X : f(x) \in Y_0\}$ is the desired subclass of X. □

The preimage of a set need not be a set, however. There could be multiple elements in the domain which map to the same element in the range, so enumerability of the range does not entail enumerability of its preimage.

Images are less well-behaved.

Proposition 9.16. *Let X and Y be classes, let $f : X \to Y$ be a function,*

and let X_0 be a subset of X. Then there is a subset Y_0 of Y with the property that $y \in Y_0$ if and only if there exists $x \in X_0$ with $f(x) = y$.

Proof. We can assume X_0 is nonempty. Let $\phi(y)$ be the predicate "there exists $x \in X_0$ with $f(x) = y$". Since X_0 is a set, the LPO ensures that this predicate is definite. So there is a subclass of Y consisting of all the elements $f(x)$ for $x \in X_0$. It can then be enumerated as $(f(x_n))$ where (x_n) is an enumeration of X_0, so it is a set. □

The main thing to notice here is that X_0 is assumed to be a set. If it were a proper class, then in order to identify the desired subclass Y_0 of Y we would need the condition "there exists $x \in X_0$ such that $f(x) = y$" to be definite. But this condition quantifies over X_0 — determining whether it holds of some specific y might require us to search through all of X_0, which we cannot do if X_0 is uncountable. So in general we cannot assume it has a definite truth value.

Definition 9.17. Let X and Y be classes, let $f : X \to Y$ be a function, let X_0 be a subset of X, and let Y_0 be a subclass of Y. Then the *image* of X_0 under f is the set $f(X_0) = \{f(x) \in Y : x \in X_0\}$ and the *preimage* of Y_0 is the class $f^{-1}(Y_0) = \{x \in X : f(x) \in Y_0\}$.

Here is one last result illustrating the good behavior of preimages.

Proposition 9.18. *Let $f : X \to Y$ be a function between classes, let Z be a subclass of Y, and let (Z_λ) be a countably indexed family of subclasses of Y. Then $f^{-1}\left(\bigcap Z_\lambda\right) = \bigcap f^{-1}(Z_\lambda)$, $f^{-1}\left(\bigcup Z_\lambda\right) = \bigcup f^{-1}(Z_\lambda)$, and $f^{-1}(Z^c) = f^{-1}(Z)^c$, taking the complements in Y and X, respectively.*

9.7 Products

In Section 9.3 I defined the product of two classes using bitstrings of the form $2^\alpha 3^\beta$. That construction straightforwardly generalizes to the product of n classes for any given value of n using bitstrings of the form $p_1^{\alpha_1} \cdots p_n^{\alpha_n}$, whose meaning is defined in a similar way. However, we want to be able to make statements about the product of n classes for variable n, or even about infinite sequences of classes. The next definition shows how to do that.

For any bitstring α, let $\alpha_\rightarrow$ be the bitstring whose first bit is on and whose $(a+1)$st bit is on if and only if the ath bit of α is on.

Definition 9.19. A *sequence of classes* is a class whose identity criterion θ satisfies the condition: whenever $\theta(\alpha, \beta)$ holds, there is a prime number p such that the pth bits of α and β are both on and the only bits in which α or β are on are powers of p.

Given a sequence of classes, for each n we can define an identity criterion θ_n by letting it hold of α and β if and only if θ holds of $p_n^{\alpha\rightarrow}$ and $p_n^{\beta\rightarrow}$. This instantiates a class X_n, and we write $\langle X_1, X_2, \ldots\rangle$ or $\langle X_n\rangle_{n=1}^{\infty}$, or just $\langle X_n\rangle$, for the sequence of classes instantiated by θ.

A *finite sequence of classes* is a sequence of classes for which $X_i = \emptyset$ for all i greater than some n; that is, it is a sequence of classes whose identity criterion only holds of bitstrings that have p_i on for some $i \leq n$. We denote it $\langle X_1, \ldots, X_n\rangle$ or $\langle X_i\rangle_{i=1}^{n}$, or again just $\langle X_i\rangle$. We can also call this a *disjoint union of n classes* and write it as $X_1 \sqcup \cdots \sqcup X_n$.

When I just say "sequence of classes" it will always mean "infinite sequence of classes".

We already had a notion of an indexed family of subclasses of a class; what is new in Definition 9.19 is that the X_n are arbitrary and need not be subclasses of some ambient class.

The intuition behind the definition is that we can disjointly embed a sequence of classes into a single class by letting the nth class in the sequence live on the powers of the nth prime. We turn the pth bit on to indicate that an element lives on the powers of p. Two sequences of classes $\langle X_n\rangle$ and $\langle Y_n\rangle$ are equal if and only if $X_n = Y_n$ for all n.

We can also say, for any n, that if X_1, ..., X_n are any classes then there exists a disjoint union of n classes $X_1 \sqcup \cdots \sqcup X_n$. This is a separate statement for each n.

Now let us define infinite products of classes. Recall that α_p is the bitstring whose ath bit is on if and only if the p^ath bit of α is on.

Definition 9.20. The *product* of an infinite sequence of classes $\langle X_n\rangle$ is the class $\prod_{n=1}^{\infty} X_n$ (or just $\prod X_n$) whose identity criterion holds of two bitstrings α and β if and only if (1) the only on bits in α and β are prime powers, and (2) for each n the identity criterion θ_n for X_n holds of α_{p_n} and β_{p_n}. If α instantiates an element of $\prod X_n$ then for each n the bitstring α_{p_n} instantiates an element of X_n, and α and β instantiate the same element of $\prod X_n$ if and only if α_{p_n} and β_{p_n} instantiate the same element of X_n for all n. We call elements of the product *sequences* and write $\alpha = (x_n)$ or $\vec{x}$ where x_n is the element of X_n instantiated by α_{p_n}; x_n is the *nth coordinate* of the sequence.

The product of a finite sequence of classes $\langle X_1, \ldots, X_n \rangle$ is defined by the obvious modification of the above where we only consult θ_i for $i \leq n$. It is written $\prod_{i=1}^{n} X_i$ (or just $\prod X_i$). We call elements of the product *n-tuples*, or *ordered pairs* or *ordered triples* for $n = 2$ or 3, and denote them $(x_1, \ldots, x_n)$ or $\vec{x}$.

I may sometimes refer to a tuple as a "finite sequence".

For $n = 2$, this definition agrees with our definition of a product of two classes in Section 9.3, and if each $X_i = X$ then $\prod_{i=1}^{\infty} X_i$ is the sequence space X^ω defined there. We can also write X^n for the product $\prod_{i=1}^{n} X_i$ when each $X_i = X$. Both X^ω and X^n for any n are *powers* of X.

If each X_i is inhabited then so is $\prod X_i$. This is a consequence of countable choice.

Given a finite sequence of classes $\langle X_1, \ldots, X_n \rangle$, we can say that for any $x_1 \in X_1$, ..., $x_n \in X_n$ there is a corresponding n-tuple $(x_1, \ldots, x_n)$, and every n-tuple arises in this way. This is a separate statement for each n.

The elements of X^n correspond to functions from $\{1, \ldots, n\} \subset \mathbb{N}$ into X, and the elements of X^ω correspond to functions from $\mathbb{N}$ into X. That is, we have an indexed family of classes $\mathcal{F} \subset (\{1, \ldots, n\} \times X) \times X^n$ or $(\mathbb{N} \times X) \times X^\omega$ such that every function from $\{1, \ldots, n\}$ or $\mathbb{N}$ into X appears exactly once as a slice — namely, the $\vec{x}$-slice is the function that sends i to x_i for all i. Again, in the framework we are developing there is no such thing as "the class of all functions from $\mathbb{N}$ into X" because a function is a certain kind of class, in this case a subclass of $\mathbb{N} \times X$, and the elements of a class cannot be classes, they must be bitstrings modulo equivalence.

A sequence in X is exhaustive if the function from $\mathbb{N}$ to X that it codes is surjective, and it excludes an element of X if the function it codes is subjective. Thus we can say that X is a nonempty set if there exists a surjective function from $\mathbb{N}$ onto X, and X is a proper class if every function from $\mathbb{N}$ to X is subjective, but these statements are formally expressed by putting the quantifiers on the indexing class X^ω.

Any product of finitely many sets is a set.

Proposition 9.21. *Let A_1, ..., A_n be sets. Then $A_1 \times \cdots \times A_n$ is a set.*

Proof. If any of the A_i is empty then the product is empty and we are done. So we can assume the A_i are all nonempty.

We know that A_1 is a set. Assume we have proven that $A_1 \times \cdots \times A_i$ is a set, for some $1 \leq i < n$. For each $x \in A_{i+1}$, all the $(i+1)$-tuples in $A_1 \times \cdots \times A_{i+1}$ whose last coordinate is x can be enumerated in the

same way that $A_1 \times \cdots \times A_i$ is enumerated. Thus these $(i+1)$-tuples constitute a set, and taking the union of these sets over all $x \in A_{i+1}$ yields all of $A_1 \times \cdots \times A_{i+1}$. Since any union of countably many subsets is a set (Theorem 9.8), it follows that the $(i+1)$-fold product is a set. By induction we conclude that $A_1 \times \cdots \times A_n$ is a set. □

The informal "let A_1, ..., A_n be sets" in the statement of this proposition should be formalized as a single unbounded universal quantifier which, under $\mathbb{A}$, is then qualified by saying that it instantiates a finite sequence of sets.

The product has a universal property. Let $\pi_{i_0} : \prod X_i \to X_{i_0}$ be the *coordinate projection* onto the i_0th coordinate, defined by putting $f(\vec{x}) = x_{i_0}$. This is a function because for any $\vec{x} \in \prod X_i$ and $y \in X_{i_0}$ it definitely either is or is not the case that $y = x_{i_0}$.

Theorem 9.22. *(Universal property of products) Let X and Y_i ($1 \leq i \leq n$ or $i \in \mathbb{N}$) be classes. For any functions $f_i : X \to Y_i$, there is a unique function $f : X \to \prod Y_i$ satisfying $f_i = \pi_i \circ f$ for all i.*

Proof. Define $f : X \to \prod Y_i$ by putting $f(x) = (f_i(x))$. This ensures that $f_i = \pi_i \circ f$ for all i. It is a function because for any $x \in X$ and any $\vec{y} \in \prod Y_i$, it definitely either is or is not the case that $f_i(x) = y_i$ for all i (LPO). Uniqueness is clear. □

The uniqueness of f in the statement of this theorem can be formalized by making a separate statement quantifying over any X, Y_i, and f_i, and any f and g, which then says, under $\mathbb{A}$, that if f and g both have the stated property then they are equal.

We can use products to define inverse limits.

Definition 9.23. Let X_i ($i \in \mathbb{N}$) be classes and let $f_i : X_{i+1} \to X_i$ be functions. The *inverse limit* of this system is the subclass $\varprojlim X_i$ of $\prod X_i$ consisting of those sequences $\vec{x}$ with the property that $x_i = f_i(x_{i+1})$ for all i. We have natural maps $g_{i_0} : \varprojlim X_i \to X_{i_0}$ simply defined as the restrictions of the coordinate projections on $\prod X_i$. They satisfy the compatibility condition that $f_i \circ g_{i+1} = g_i$ for all i.

It is easy enough to check that the preceding does identify a subclass of $\prod X_i$. For any sequence $\vec{x}$ and any value of i it definitely is or is not the case that $f_i(x_{i+1}) = x_i$. By the LPO, it definitely is or is not the case that this holds for all i.

We need to use quotients to define direct limits, so that will come at the end of the next section.

9.8 Quotients

An *equivalence relation* is a relation that is reflexive, symmetric, and transitive. Given an equivalence relation $\sim$ on a class X, we can decompose X into *blocks* of $\sim$, such that two elements are related if and only if they belong to the same block. Each block is a subclass of X.

We can also define a new identity criterion $\theta_\sim$ by letting it hold of bitstrings α and β if they instantiate elements of X which are related by $\sim$. This generates a new class $X/\sim$. All the bitstrings which instantiate a given element x of X will also instantiate the same element $[x]$ of $X/\sim$; we can also say that x instantiates $[x]$. We therefore have a *quotient map* $\pi : X \to X/\sim$ which takes $x \in X$ to $[x] \in X/\sim$.

Subclasses of X are related to subclasses of $X/\sim$ in the following way.

Proposition 9.24. *Let X be a class, let $\sim$ be an equivalence relation on X, and let Y be a subclass of X with the property that $x \in Y$ and $x \sim y$ imply $y \in Y$. Then $\{[x] : x \in Y\}$ is a subclass of $X/\sim$. Every subclass of $X/\sim$ arises in this way.*

Proof. To see that $\{[x] : x \in Y\}$ is a subclass of $X/\sim$, let $[x] \in X/\sim$; we must show that it definitely is or is not the case that $[x] = [y]$ for some $y \in Y$. But the hypothesis on Y says that this happens if and only if $x \in Y$, and since Y is assumed to be a subclass of X, this condition is indeed definite.

Conversely, given any subclass Y' of $X/\sim$, we know from Proposition 9.15 that $\pi^{-1}(Y')$ is a subclass of X, and it is clear both that $x \in \pi^{-1}(Y')$ and $x \sim y$ imply $y \in \pi^{-1}(Y')$ and that $Y' = \{[x] : x \in \pi^{-1}(Y')\}$. □

I noted earlier that the image of a subclass of the domain need not exist as a subclass of the range. However, in some cases it does. For instance, Proposition 9.24 shows that if $Y \subseteq X$ is compatible with $\sim$ in the obvious sense then $\pi(Y) = \{[x] : x \in Y\}$ is a legitimate subclass of $X/\sim$.

Like products, quotients have a universal property.

Theorem 9.25. *(Universal property of quotients) Let X and Y be classes and let $\sim$ be an equivalence relation on X. Suppose $f : X \to Y$ is a*

function satisfying $x \sim y \to f(x) = f(y)$. Then there is a unique function $g : X/\sim \to Y$ satisfying $f = g \circ \pi$.

Proof. We are given that for each x and x' in X, if $[x] = [x']$ then $f(x) = f(x')$. Thus, for each element $[x]$ of $X/\sim$ there is a unique element y of Y such that $f(x') = y$ for all $x' \sim x$. Define $g([x]) = y$. This is a function because for each $[x] \in X/\sim$ and $y \in Y$, it definitely either is or is not the case that $f(x) = y$. The condition $f = g \circ \pi$ and uniqueness are clear. □

We put this theorem, minus the uniqueness statement, in the form described in Section 9.5 as follows: for all θ_1, θ_2, θ_3, θ_4 that meaningfully apply to any two bitstrings, it is assertible that if θ_1 and θ_2 instantiate classes X and Y, θ_3 instantiates an equivalence relation $\sim$ on X, and θ_4 instantiates a function f from X to Y, then compatibility of f with $\sim$ implies that $g([x]) = y$ defines a function from X to Y that satisfies $f = g \circ \pi$. The uniqueness part can be handled by formalizing "for any g_1 and g_2, $g_1 \circ \pi = g_2 \circ \pi$ implies $g_1 = g_2$".

Proposition 9.26. *Any quotient of a set is a set.*

This follows from Proposition 9.16 since $A/\sim$ is the image of A under the quotient map.

Here is why Ω, the class of all bitstrings with no equivalence, is "universal".

Theorem 9.27. *Every class is naturally in bijection with a quotient of a subclass of Ω.*

Proof. Let X be a class instantiated by an identity criterion θ. Then $\{\alpha : \theta(\alpha, \alpha)\}$ is a subclass Y of Ω, and $\alpha \sim \beta \leftrightarrow \theta(\alpha, \beta)$ defines an equivalence relation on Y such that the natural map from $Y/\sim$ to X, taking the equivalence class of α to the element of X it instantiates, is a bijection. □

The point is that identity criteria, in the sense of Section 9.3, naturally correspond to equivalence relations on subclasses of Ω.

In the last section we saw how to realize all of the functions from $\{1, \ldots, n\}$ or $\mathbb{N}$ into a class X as a family of subclasses of $\{1, \ldots, n\} \times X$ or $\mathbb{N} \times X$ indexed by X^n or X^ω. This allows us to effectively quantify over all such functions using bounded quantifiers over X^n or X^ω. Now, since every nonempty set can be put in bijection either with $\mathbb{N}$ or with $\{1, \ldots, n\}$ for some n, we can handle the functions from any nonempty set into X in a

similar way. However, the choice of bijection is not canonical. The quotient space construction can be used to remedy this.

Definition 9.28. Given a nonempty set A and a class X, let Y be the subclass of $A^\omega \times X^\omega$ consisting of all pairs of sequences $(\vec{a}, \vec{x})$ with the properties that (1) $\vec{a}$ is exhaustive, i.e., every element of A equals a_i for some i, and (2) $\vec{a}$ and $\vec{x}$ are compatible, i.e., whenever $a_i = a_j$ we also have $x_i = x_j$. These conditions are definite (LPO) so they do define a subclass. We get an equivalence relation on Y by setting $(\vec{a}, \vec{x}) \sim (\vec{b}, \vec{y})$ if whenever $a_i = b_j$ we also have $x_i = y_j$. We define the class $\mathrm{fn}(A, X)$ to be the quotient $Y/\!\sim$.

It should now be clear how to define an indexed family of subclasses of $A \times X$, with indexing class $\mathrm{fn}(A, X)$, such that every slice is a function from A to X and every such function appears exactly once as a slice, i.e., $\mathrm{fn}(A, X)$ is an exact index. The slice indexed by $[(\vec{a}, \vec{x})]$ takes a_i to x_i for all i.

As I suggested in Section 9.3, we can call $\mathrm{fn}(A, X)$ a "pseudoclass" and write $f \,\varepsilon\, \mathrm{fn}(A, X)$ to say that f is a slice of the indexed family, which is just to say that it is a function from A to X.

Power classes can be handled in the same kind of way.

Definition 9.29. Given a class X, let Y be the class X^ω augmented by one additional bitstring whose first bit is on and all of whose other bits are off (i.e., the element 1 of $\mathbb{N}$). Then let the *power class* of X, denoted $\mathcal{P}(X)$, be the quotient of Y by the equivalence relation which sets 1 equivalent only to itself and sets $\vec{x}$ equivalent to $\vec{y}$ if every x_i equals some y_j and every y_j equals some x_i.

Here there is a family of subclasses of X, with indexing class $\mathcal{P}(X)$, such that every slice is a subset of X and every subset of X appears exactly once as a slice, i.e., we again have exact codes. The slice indexed by 1 is $\emptyset$, and the slice indexed by $[\vec{x}]$ is $\{x \in X : x = x_i \text{ for some } i\}$. So $\mathcal{P}(X)$ can be regarded as a pseudoclass of all the subsets of X.

Finally, here is the promised definition of direct limits.

Definition 9.30. Let X_i ($i \in \mathbb{N}$) be inhabited classes and let $f_i : X_i \to X_{i+1}$ be functions. Define X to be the subclass of $\prod X_i$ consisting of those sequences $\vec{x}$ with the property that there exists N such that $x_{i+1} = f_i(x_i)$ for all $i \geq N$. Then define an equivalence relation on X by putting $\vec{x} \sim \vec{y}$ if there exists N such that $x_i = y_i$ for all $i \geq N$. The quotient $X/\!\sim$ is

the *direct limit* of the system, denoted $\varinjlim X_i$. Fix $a_i \in X_i$ for all i. We can then define maps $g_{i_0} : X_{i_0} \to \varinjlim X_i$ by letting $g_{i_0}(z)$ be the element of $\varinjlim X_i$ instantiated by the sequence $\vec{x}$ where $x_i = a_i$ for $i < i_0$, $x_{i_0} = z$, and $x_{i+1} = f_i(x_i)$ for $i \geq i_0$ (dependent choice). These maps satisfy the compatibility condition that $g_{i+1} \circ f_i = g_i$ for all i.

This construction is does not depend on the choice of the a_i's.

9.9 Number systems

The integers and the rationals can be constructed from $\mathbb{N}$ in the usual way.

Definition 9.31. The set of integers $\mathbb{Z}$ is the quotient set obtained by factoring out the equivalence relation

$$(a, b) \sim (c, d) \quad \leftrightarrow \quad a + d = b + c$$

on $\mathbb{N} \times \mathbb{N}$. The set of rational numbers $\mathbb{Q}$ is the quotient set obtained by factoring out the equivalence relation

$$(p, q) \sim (r, s) \quad \leftrightarrow \quad ps = qr$$

on $\mathbb{Z} \times \mathbb{N}$.

In the construction of $\mathbb{Z}$, the ordered pair (a, b) plays the role of $a - b$. If $a > b$ then we "identify" the element $[(a, b)]$ of $\mathbb{Z}$ instantiated by (a, b) with the natural number $a-b$, and write it the same way (e.g., as a decimal number); if $a = b$ then we write "0" for $[(a, a)]$, and if $a < b$ then we write "$-(b - a)$" for $[(a, b)]$. Thus every integer is either 0, or n for some $n \in \mathbb{N}$, or $-n$ for some $n \in \mathbb{N}$.

In order to define $+$ and $\cdot$ on $\mathbb{Z}$, we put $(a, b) + (c, d) = (a + c, b + d)$ and $(a, b) \cdot (c, d) = (ac + bd, ad + bc)$; then we have to check that these operations are compatible with the equivalence relation described above. We should also check that $+$ and $\cdot$, so defined, agree with the usual operations on $\mathbb{N}$ when the latter is identified with a subset of $\mathbb{Z}$. These are all straightforward calculations. We also define $(a, b) \leq (c, d)$ if $a + d \leq b + c$; this induces a relation on $\mathbb{Z}$ which agrees with the usual order relation on $\mathbb{N}$.

In the construction of $\mathbb{Q}$, the ordered pair (p, q) plays the role of $\frac{p}{q}$, and we denote the element of $\mathbb{Q}$ it instantiates in this way. We define $+$, $\cdot$, and $\leq$ on $\mathbb{Z} \times \mathbb{N}$ by putting $(p, q) + (r, s) = (ps + rq, qs)$ and $(p, q) \cdot (r, s) = (pr, qs)$, and letting $(p, q) \leq (r, s)$ if $ps \leq qr$. We "identify" $[(p, 1)]$ with the integer p. The well-definedness of $+$, $\cdot$, and $\leq$ on $\mathbb{Q}$, and the fact that they agree

with $+$, $\cdot$, and $\leq$ on $\mathbb{Z}$ when the latter is identified with a subset of $\mathbb{Q}$, are also straightforward calculations.

We know that $\mathbb{N} \times \mathbb{N}$ is a set by Proposition 9.21, and then $\mathbb{Z}$ is a set by Proposition 9.26. Similar reasoning shows that $\mathbb{Q}$ is a set.

We write $\mathbb{Z}_+ \subset \mathbb{Z}$ for the nonnegative integers 0, 1, 2, ...

A minor irritation here is that, strictly speaking, $\mathbb{N}$ is not contained in $\mathbb{Z}$ and $\mathbb{Z}$ is not contained in $\mathbb{Q}$. The elements of $\mathbb{Z}$ are ordered pairs modulo an equivalence relation and the elements of $\mathbb{Q}$ are ordered pairs modulo a different equivalence relation. Nonetheless, we have a natural map from $\mathbb{N}$ into $\mathbb{Z}$ which takes $n \in \mathbb{N}$ to the element of $\mathbb{Z}$ instantiated by $(n+1, 1)$. This map preserves the algebraic structure of $\mathbb{N}$. Similarly, we have a natural map from $\mathbb{Z}$ into $\mathbb{Q}$ taking $p \in \mathbb{Z}$ to the element of $\mathbb{Q}$ instantiated by $(p, 1)$. This map, too, preserves algebraic structure, as one easily checks. So it is customary to be sloppy and "identify" $\mathbb{N}$ with its embedded copy in $\mathbb{Z}$, and $\mathbb{Z}$ with its embedded copy in $\mathbb{Q}$. This causes no real issues, but we should be aware that we are doing it.

Since $\mathbb{Q}$ is a set, there is an exhaustive sequence (x_n) in $\mathbb{Q}$, and since it is infinite, we can assume that all repeated entries in the sequence have been deleted, so that (x_n) represents a bijection from $\mathbb{N}$ to $\mathbb{Q}$. There are several rather elegant methods of making such a bijection explicit, but we do not need to do this. All we need is the existence of an exhaustive sequence in $\mathbb{Q}$ without any repetitions, so that we can transfer the addition and multiplication operations on the rationals over to the natural numbers by defining $m +_Q n$ and $m \times_Q n$ in such a way that

$$x_m + x_n = x_{m+_Q n} \qquad \text{and} \qquad x_m \times x_n = x_{m \times_Q n}$$

for any natural numbers m and n. We also set $m \leq_Q n$ if $x_m \leq x_n$ and $m <_Q n$ if $x_m < x_n$, i.e., $x_m \leq x_n$ and $x_m \neq x_n$.

We can now define the real line using "Dedekind cuts". "Cauchy completion" is another approach that results in an object with identical structure to the one obtained via Dedekind cuts, but it fits more naturally in later material (Section 10.1), so I will return to it then.

Definition 9.32. A bitstring is a *cut* if there is at least one on bit and at least one off bit, and whenever the nth bit is on the mth bit is also on for all $m <_Q n$ and for some $m >_Q n$. Define the *real line* $\mathbb{R}$ to be the class whose elements are the cuts, with no identification of distinct cuts.

The predicate "is a cut" is definite (LPO, several times), so $\mathbb{R}$ is a well-defined class. We define a relation $\leq$ on $\mathbb{R}$ by setting $\alpha \leq \beta$ if every on

bit in α is also on in β; this is also definite by the LPO. We write $\alpha < \beta$ if $\alpha \leq \beta$ and $\alpha \neq \beta$. For any $\alpha, \beta \in \mathbb{R}$ exactly one of $\alpha < \beta$, $\alpha = \beta$, or $\beta < \alpha$ holds, a fact known as "trichotomy". We also define $\alpha + \beta$ to be the cut whose on bits are precisely those of the form $m +_Q n$ where m is on in α and n is on in β. Multiplication of cuts is defined similarly, with the usual irritating complication when multiplying negative numbers. We use the interval notations (α, β), $[\alpha, \infty)$, etc. with their usual meanings. Every nonzero real number has a multiplicative inverse.

The definition of $\mathbb{R}$ can be modified to allow the all-on and all-off bitstrings to count as cuts. This introduces two new elements, denoted $-\infty$ and ∞, and the resulting class is called the *extended real line* and denoted $\overline{\mathbb{R}}$.

We cannot, strictly speaking, say that $\mathbb{Q}$ is contained in $\mathbb{R}$. Rather, there is a natural embedding which takes a rational number x_n (in the enumeration used above) to the cut whose on bits are all $m <_Q n$.

Let the *join* of a sequence of cuts (α_i) be the bitstring whose nth bit is on if the nth bit of α_i is on for some i.

Lemma 9.33. *The join of any sequence of cuts is either a cut or the all-on bitstring.*

Theorem 9.34. $\mathbb{R}$ *is a proper class.*

Proof. Let (α_i) be a sequence in $\mathbb{R}$. We must find an element of $\mathbb{R}$ that is not in this list.

Put $n_1 = 1$ and $n_2 = 2$. Without loss of generality we can assume $\alpha_1 \leq \alpha_2$. Then let n_3 be the smallest index such that α_{n_3} is nested strictly between $\alpha_{n_1} = \alpha_1$ and $\alpha_{n_2} = \alpha_2$, and continue in the same way, for each k letting n_{k+2} be the smallest index such that $\alpha_{n_{k+2}}$ is nested strictly between α_{n_k} and $\alpha_{n_{k+1}}$. If at any point no α_n lies strictly between α_{n_k} and $\alpha_{n_{k+1}}$ then $(\alpha_{n_k} + \alpha_{n_{k+1}})/2$ is a real number that does not equal α_n for any n. Otherwise (and such an α_n always definitely does or does not exist, LPO) the construction can continue indefinitely, and we get $\alpha_{n_1} \leq \alpha_{n_3} \leq \alpha_{n_5} \leq \cdots$ and $\alpha_{n_2} \geq \alpha_{n_4} \geq \alpha_{n_6} \geq \cdots$. The join α of the sequence $(\alpha_{n_{2k-1}})$ is a cut by Lemma 9.33.

This cut α is nested strictly between every successive pair of α_{n_k} and $\alpha_{n_{k+1}}$. It therefore cannot equal α_n for any n, because the sequence (n_k) increases to infinity, and if we had $\alpha = \alpha_n$ for some n then any $n_k > n$ could not be the smallest index of a cut nested between the previous two terms. We have shown that there is a cut that does not appear in the

original sequence of cuts. □

We can also pick up the following fundamental property of $\mathbb{R}$.

Theorem 9.35. *(Least upper bound principle) Any nonempty subset of* $\mathbb{R}$ *that is bounded above has a least upper bound.*

Proof. Any nonempty subset of $\mathbb{R}$ is enumerated by a sequence, and by Lemma 9.33 the join of that sequence is either a real number or ∞. If the subset is bounded above then the second case is excluded, so the join will be a real number, and this is clearly its least upper bound. □

Note that I do not assert this of subclasses of $\mathbb{R}$, only of subsets. (More generally, it can be affirmed of "separable" subclasses of $\mathbb{R}$, in the sense of Definition 10.9.)

We call the least upper bound of a set $A \subset \mathbb{R}$ that is bounded above the *supremum* of A and denote it $\sup A$. Multiplying by -1 lets us infer the existence of a greatest lower bound, or *infimum*, of any set $B \subset \mathbb{R}$ that is bounded below. This is denoted $\inf B$.

From now on I will denote real numbers by x, y, etc. We can forget that these are literally bitstrings. Write $\min(a, b)$ (respectively, $\max(a, b)$) for the smaller (respectively, larger) of a and b.

Theorem 9.36. *Let* x *be a positive real number and let* $n \in \mathbb{N}$*. Then there is a unique positive real number* y *with* $y^n = x$*.*

Proof. Let A be the set of all positive rational numbers t such that $t^n < x$. This set is nonempty because it contains every positive rational less than $\min(x, 1)$, and bounded above because it does not contain anything larger than $\max(x, 1)$. Thus by the least upper bound principle (Theorem 9.35) the set A has a least upper bound y.

Exactly one of $y^n < x$, $y^n = x$, and $y^n > x$ must hold. If $y^n < x$ then let $h = \min(\frac{x-y^n}{n(y+1)^{n-1}}, 1)$; using the inequality $a^n - b^n = (a-b)(a^{n-1} + a^{n-2}b + \cdots + b^{n-1}) < na^{n-1}(a-b)$ for $a > b \geq 0$, this yields

$$(y+h)^n - y^n < nh(y+h)^{n-1} \leq nh(y+1)^{n-1} \leq x - y^n,$$

showing that $(y+h)^n < x$. This implies that any positive rational number less than $y + h$ belongs to A, which contradicts the definition of y.

If $y^n > x$, let $k = \min(\frac{y^n - x}{ny^{n-1}}, y)$. Then

$$y^n - (y-k)^n < nky^{n-1} \leq y^n - x,$$

showing that no rational number greater than $y - k$ belongs to A. This also contradicts the definition of y. The only remaining possibility is that $y^n = x$.

If $s < t$ then $s^n < t^n$ for all n; this establishes uniqueness. □

Thus, every positive real number has a unique positive nth root, for all $n \in \mathbb{N}$.

Let $\mathbb{C} = \mathbb{R}^2$ and write "$a + bi$" for the ordered pair (a, b). We call a and b the *real and imaginary parts* of z, and write $\mathrm{Re}(z) = a$ and $\mathrm{Im}(z) = b$. We define a sum and a product on $\mathbb{C}$ by putting $(a + bi) + (c + di) = (a + c) + (b + d)i$ and $(a + bi)(c + di) = (ac - bd) + (bc + ad)i$. This is the *complex plane*. Letting the *modulus* of $z = a + bi$ be the real number $|z| = \sqrt{a^2 + b^2}$ and letting its *complex conjugate* be $\overline{z} = a - bi$, we get that $z\overline{z} = |z|^2$ is real and nonnegative. If z is not zero then it is invertible with inverse $\frac{\overline{z}}{|z|^2}$. So we can also divide complex numbers.

Once again, I make the comment that according to our definitions $\mathbb{R}$ is not literally contained in $\mathbb{C}$. Here the natural embedding takes the real number a to the complex number $(a, 0)$, and one is usually sloppy and identifies the two (as we have already implied with the notation "$a + bi$"). This allows us to define $|z|$ to be a real number but also say things like "$z\overline{z} = |z|^2$" in $\mathbb{C}$.

The modulus is multiplicative, i.e., we have $|wz| = |w| \cdot |z|$ for all $w, z \in \mathbb{C}$; this is a straightforward calculation. Define $\cos\theta$ and $\sin\theta$ by letting $\cos\theta + i\sin\theta$ be the point on the unit circle in $\mathbb{C}$ such that the ray from the origin to this point makes the angle θ with the x-axis. (A more rigorous definition will come in Section 13.4.) For any nonzero $z \in \mathbb{C}$, the point $\frac{z}{|z|}$ lies on the unit circle, and hence has the form $\cos\theta + i\sin\theta$ for some $\theta \in [0, 2\pi)$. Thus $z = |z|(\cos\theta + i\sin\theta)$; this is the *polar representation* of z. If $w = |w|(\cos\phi + i\sin\phi)$ then elementary trigonometric identities yield that $zw = |z||w|(\cos(\theta + \phi) + i\sin(\theta + \phi))$. We will also derive this identity from scratch in Section 14.2. As a consequence, we have $z^k = |z|^k(\cos(k\theta) + i\sin(k\theta))$ for any $k \in \mathbb{N}$.

9.10 Order

Definition 9.37. A relation $\leq$ on a class X is a *partial order* or *partial ordering of* X if it is reflexive, antisymmetric, and transitive. We write "$x < y$" for "$x \leq y$ and $x \neq y$". A partial order $\leq$ is *total* if $x \leq y$ or $y \leq x$

for all $x, y \in X$. A subset A of X is *dense* relative to some total order $\leq$ on X if for all $x, z \in X$ with $x < z$ there exists $y \in A$ with $x < y < z$. If X is a set and $A = X$ then we can say that X *dense in itself*. A total order is *unbounded* if there is no least or greatest element.

When we have fixed a partial ordering $\leq$ of a class X, we often refer to X as an "ordered class" and suppress the "$\leq$" notation. Thus "let X be an ordered class" abbreviates "let X be a class and let $\leq$ be a partial ordering of X".

If X and Y are ordered classes, an *order-isomorphism* between them is a bijection $f : X \cong Y$ with the property that $x \leq y$ iff $f(x) \leq f(y)$, for all $x, y \in X$. X and Y are *order-isomorphic* if there exists an order-isomorphism between them. This is an unbounded existential quantifier and any affirmative use of it ("in such and such a situation, there does exist an order-isomorphism") should always be supported by an explicit construction.

Theorem 9.38. *Any two nonempty, unbounded, totally ordered sets which are dense in themselves are order-isomorphic.*

Proof. Let A and B be nonempty, unbounded, totally ordered sets which are dense in themselves. Since they are sets we can enumerate their elements as a_1, a_2, ... and b_1, b_2, ..., without repetition. (Note that the unboundedness condition forces both A and B to be infinite.) We will define finite subsets $A_1 \subset A_2 \subset \cdots$ of A and $B_1 \subset B_2 \subset \cdots$ of B, together with bijections $f_i : A_i \cong B_i$ such that (1) $\bigcup A_i = A$ and $\bigcup B_i = B$, (2) each f_i is an order-isomorphism, giving A_i and B_i the orders inherited from A and B, and (3) f_{i+1} extends f_i, for every i. Given this, the union of the f_i is the desired order-isomorphism between A and B.

The construction goes as follows. Starting with $i = 1$, let $A_1 = \{a_1\}$ and $B_1 = \{b_1\}$ and define $f_1(a_1) = b_1$. Given A_1, ..., A_n and B_1, ..., B_n and $f_n : A_n \cong B_n$, let j be the smallest index such that $a_j \notin A_n$, and let k be the smallest index such that b_k is situated relative to the elements of B_n in the same way a_j is situated relative to the corresponding elements of A_n. Then let k' be the smallest index such that $b_{k'} \notin B_n \cup \{b_k\}$ and let j' be the smallest index such that $a_{j'}$ is situated relative to the elements of $A_n \cup \{a_j\}$ in the same say $b_{k'}$ is situated relative to the corresponding elements of $B_n \cup \{b_{k'}\}$. Let $A_{n+1} = A_n \cup \{a_j, a_{j'}\}$ and $B_{n+1} = B_n \cup \{b_k, b_{k'}\}$ and extend $f_n : A_n \to B_n$ to $f_{n+1} : A_{n+1} \to B_{n+1}$ by defining $f_{n+1}(a_j) = b_k$ and $f_{n+1}(a_{j'}) = b_{k'}$.

We can see inductively that j, the smallest index such that $a_j \notin A_n$, satisfies $j > n$. Thus $a_n \in A_n$ for all n, and similarly for the b_n's. This establishes (1), and the other two conditions are easy. □

The promised order-isomorphism between A and B is indeed explicitly given in the above proof. The statement of Theorem 9.38 uses unbounded universal quantification over A, B, and their total orders; all quantification in the proof on subsets of A and B and functions between these subsets can be formalized as quantification over the classes $\mathcal{P}(A)$, $\mathcal{P}(B)$, and $\mathrm{fn}(A_i, B_i)$.

Since $\mathbb{Q}$ meets the hypotheses of Theorem 9.38, the following corollary is immediate.

Corollary 9.39. *Any unbounded totally ordered set that is dense in itself is order-isomorphic to* $\mathbb{Q}$.

This also yields an abstract characterization of $\mathbb{R}$.

Theorem 9.40. *Let X be a class equipped with an unbounded total order with respect to which it contains a dense subset. Suppose every bounded subset of X has a least upper bound. Then X is order-isomorphic to* $\mathbb{R}$.

Proof. Let A be a dense subset of X. Density in X entails both that A is unbounded and that it is dense in itself. So there is an order-isomorphism $f : A \cong \mathbb{Q}$ by Corollary 9.39. Now define a function $g : X \to \mathbb{R}$ by letting $g(x)$ be the supremum of the set $f(\{y \in A : y < x\})$. This set is bounded above by $f(z)$ for any $z \in A$ greater than x, so its supremum in $\mathbb{R}$ does exist. It is clear that g preserves order, i.e., $x \leq y$ in X implies $g(x) \leq g(y)$. Injectivity of g follows from this plus the fact that it is injective on A. It is surjective because for any $r \in \mathbb{R}$ the preimage under f of $\{s \in \mathbb{Q} : s < r\}$ is a bounded subset of A, and its least upper bound is an element of X that maps to r. Thus g is a bijection; since both orders are total, this plus the fact that it preserves order implies that it is an order-isomorphism. □

Here are two more general results about ordered sets. First, let a *tree* be a partially ordered set A which is a union of disjoint unordered sets A_i ($1 \leq i \leq n$ or $i \in \mathbb{N}$), called *levels*, such that A_1 contains exactly one element (the *root* of the tree), and for all $i \geq 2$, every element of A_i lies above exactly one element of A_j for each $1 \leq j < i$ and below no element of A_j for any $1 \leq j \leq i$. The elements of A_i which lie above a given element x of A_{i-1} (if there are any) are called *successors* of x.

Proposition 9.41. *(König's infinity lemma) Let A be an infinite tree. If*

every level is finite then there exists a sequence (x_n) *in* A *such that* x_1 *is the root and* x_{n+1} *is a successor of* x_n*, for all* n*.*

Proof. Enumerate the elements of A as z_1, z_2, ... For each $x \in A$ and $n \in \mathbb{N}$, it definitely is or is not the case that $x < z_m$ for some $m > n$ (LPO), and therefore for any $x \in A$ either there exists $n \in \mathbb{N}$ such that $x \not< z_m$ for all $m > n$ or there does not (LPO). Conclusion: for any x it definitely is or is not the case that x lies below infinitely many other elements.

Let x_1 be the root of A. It clearly lies below infinitely many other elements. Since x_1 has only finitely many successors, it is impossible that each of them lies under only finitely many other elements. By the last paragraph, this implies that at least one of these successors must also lie below infinitely many other elements; call it x_2. Continuing in this way, we get a sequence $x_1 < x_2 < \cdots$. □

Say that one partial order $\leq^*$ *extends* another partial order $\leq$ defined on the same set if $x \leq y \to x \leq^* y$, for all x and y. Two elements of a partially ordered class are *incomparable* if neither is less than the other.

Lemma 9.42. *Let* $\leq$ *be a partial ordering of a set* A *and suppose* $x_0, y_0 \in A$ *are incomparable. Then there is a partial ordering* $\leq^*$ *of* A *with* $x_0 <^* y_0$ *which extends* $\leq$ *and which is minimal among all extensions of* $\leq$ *which make* x_0 *less than* y_0*.*

Proof. Put $x \leq^* y$ if either (1) $x \leq y$ or (2) $x \leq x_0$ and $y_0 \leq y$. Verifying that $\leq^*$ is a partial order is routine, and minimality is clear. □

Theorem 9.43. *Let* $\leq$ *be a partial ordering of a set* A*. Then there is a total ordering* $\leq^*$ *of* A *which extends* $\leq$*.*

Proof. Assume A is infinite and enumerate its elements as x_1, x_2, ... (The case where A is finite can be dealt with using an identical argument to what follows, or can be reduced to the infinite case by adding infinitely many "dummy" elements which are removed after the construction is over.) Find the smallest index i such that x_i is incomparable to some x_j with $j < i$. Then apply Lemma 9.42 with x_i and x_j in place of x and y, and call the resulting partial order $\leq_1$. Then find the smallest index i such that x_i is incomparable according to $\leq_1$ to some x_j with $j < i$, apply Lemma 9.42 to this x_i and x_j, and so on. If we reach a stage n where $\leq_n$ is total then we are done; otherwise the union of the relations $\leq_n$ is the desired total extension of $\leq$. □

Chapter 10

General Topology

10.1 Metric spaces

Definition 10.1. A *distance function* or *metric* on a class X is a function $d : X \times X \to [0, \infty)$ that satisfies

(1) $d(x, y) = 0$ if and only if $x = y$, for all $x, y \in X$
(2) (symmetry) $d(x, y) = d(y, x)$ for all $x, y \in X$
(3) (triangle inequality) $d(x, z) \leq d(x, y) + d(y, z)$ for all $x, y, z \in X$.

When we have fixed a metric on X, we say that X is a *metric space*. Thus "let X be a metric space" will be our typical shorthand for "let X be a class and let d be a metric on X". I will generically use "d" for the distance function on any metric space, even when more than one space is in play.

A *subspace* of a metric space X is a subclass $C \subseteq X$, equipped with the restriction of the metric on X. A function $f : X \to Y$ between metric spaces is an *isometry* if $d(f(x), f(y)) = d(x, y)$ for all $x, y \in X$; it is *nonexpansive* if $d(f(x), f(y)) \leq d(x, y)$ for all $x, y \in X$; and for any $c \geq 0$ it is *c-Lipschitz* if $d(f(x), f(y)) \leq c \cdot d(x, y)$ for all $x, y \in X$. It is *Lipschitz* if it is c-Lipschitz for some $c \geq 0$.

Example 10.2. $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ are all metric spaces relative to the metric $d(x, y) = |x - y|$.

There are multiple natural metrics on $\mathbb{R}^n$, for any $n \geq 2$. One very natural metric, for example, is

$$d(\vec{a}, \vec{b}) = \max_{1 \leq i \leq n} |a_i - b_i|.$$

This is the l^∞ *metric*. Another very natural metric on $\mathbb{R}^n$ is the *Euclidean metric*

$$d(\vec{a},\vec{b}) = \sqrt{(a_1-b_1)^2+\cdots+(a_n-b_n)^2},$$

or more simply $d(\vec{a},\vec{b}) = \|\vec{a}-\vec{b}\|$ where we take the difference coordinatewise and the *Euclidean norm* $\|\cdot\|$ is defined on $\mathbb{R}^n$ by $\|\vec{c}\| = \sqrt{c_1^2+\cdots+c_n^2}$. The fact that this is a metric requires a short argument.

Lemma 10.3. *(Cauchy-Schwarz inequality in $\mathbb{R}^n$) For any $\vec{a},\vec{b} \in \mathbb{R}^n$ we have*

$$\left|\sum_{i=1}^n a_i b_i\right| \leq \|\vec{a}\|\|\vec{b}\|.$$

Proof. Consider the quadratic polynomial $(a_1x+b_1)^2+\cdots+(a_nx+b_n)^2$. It is positive for all x, so collecting coefficients yields $Ax^2+Bx+C \geq 0$ for all x, where $A = \|\vec{a}\|^2$, $B = 2\left(\sum_{i=1}^n a_i b_i\right)$, and $C = \|\vec{b}\|^2$. From the quadratic formula we get that $B^2-4AC \leq 0$, i.e.,

$$4\left(\sum_{i=1}^n a_i b_i\right)^2 \leq 4\|\vec{a}\|^2\|\vec{b}\|^2.$$

Dividing by 4 and taking square roots yields the desired inequality. □

Proposition 10.4. *The Euclidean metric on $\mathbb{R}^n$ is a metric.*

Proof. The first two properties of a metric are easily checked. The interesting condition is the triangle inequality. We must show that for any $\vec{a},\vec{b},\vec{c} \in \mathbb{R}^n$ we have

$$\|\vec{a}-\vec{c}\| \leq \|\vec{a}-\vec{b}\| + \|\vec{b}-\vec{c}\|.$$

Setting $\vec{u} = \vec{a}-\vec{b}$ and $\vec{v} = \vec{b}-\vec{c}$, this becomes

$$\|\vec{u}+\vec{v}\| \leq \|\vec{u}\| + \|\vec{v}\|.$$

We will show that

$$\|\vec{u}+\vec{v}\|^2 \leq (\|\vec{u}\|+\|\vec{v}\|)^2 = \|\vec{u}\|^2 + 2\|\vec{u}\|\|\vec{v}\| + \|\vec{v}\|^2.$$

This is true because

$$\|\vec{u}+\vec{v}\|^2 = (u_1+v_1)^2+\cdots+(u_n+v_n)^2 = \|\vec{u}\|^2 + 2\sum_{i=1}^n u_i v_i + \|\vec{v}\|^2$$

and

$$2\sum_{i=1}^n u_i v_i \leq 2\|\vec{u}\|\|\vec{v}\|$$

by the Cauchy-Schwarz inequality (Lemma 10.3). □

One of the most fundamental concepts in a metric space is convergence.

Definition 10.5. Let X be a metric space. A sequence (x_n) in X *converges* to $x \in X$ if for every $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that $d(x_n, x) < \epsilon$ for all $n \geq N$. If (x_n) converges to some point then it is a *convergent* sequence and the point it converges to is the *limit* of the sequence. We write $x_n \to x$ or $\lim x_n = x$, and we say of the sequence (x_n) that its limit exists. A sequence (x_n) is *Cauchy* if for every $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that $d(x_m, x_n) < \epsilon$ for all $m, n \geq N$. X is *complete* if every Cauchy sequence in X converges.

Relative to the usual distance function, $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{R}$, and $\mathbb{C}$ are all complete metric spaces, but $\mathbb{Q}$ is not. Here is the argument for $\mathbb{R}$:

Proposition 10.6. *$\mathbb{R}$ is complete for the distance function $d(x, y) = |x-y|$.*

Proof. Let (x_n) be a Cauchy sequence in $\mathbb{R}$. Then $\{x_n : n \in \mathbb{N}\}$ is bounded above and below. For each $j \in \mathbb{N}$ let $a_j = \sup\{x_n : n \geq j\}$, so that $a_1 \geq a_2 \geq \cdots$, and then let $a = \inf\{a_j : j \in \mathbb{N}\}$. I claim that (x_n) converges to a. To see this, let $\epsilon > 0$ and $N \in \mathbb{N}$. Then there exists $j \geq N$ such that $a \leq a_j < a + \epsilon$, and $n \geq j$ such that $a_j \geq x_n > a_j - \epsilon$. We then have $|a - x_n| < \epsilon$. Since (x_n) is Cauchy and there are arbitrarily large values of n for which x_n is arbitrarily close to a, it follows that $x_n \to a$. □

It follows easily that $\mathbb{C}$ is also complete, relative to the distance function $d(z, w) = |z - w|$.

Now here is an important result about complete metric spaces. A *fixed point* of a function $f : X \to X$ is an element $x \in X$ satisfying $f(x) = x$.

Theorem 10.7. *(Banach fixed-point theorem) Let X be an inhabited complete metric space and suppose $f : X \to X$ is c-Lipschitz for some $c < 1$. Then f has a fixed point.*

Proof. Fix $y \in X$, and for all $n \in \mathbb{N}$ define $x_n = f^n(y)$. Setting $r = d(x_1, y)$, we have $d(x_1, x_2) \leq cr$, $d(x_2, x_3) \leq c^2 r$, etc., with $d(x_n, x_{n+1}) \leq c^n r$. It follows from the triangle inequality that for any $N \in \mathbb{N}$ we have

$$d(x_m, x_n) \leq (c^n + \cdots + c^{m-1}) \cdot r = \frac{c^n - c^m}{1 - c} r \leq \frac{c^N}{1 - c} r$$

for all $m > n \geq N$. Since c^N goes to 0 as N goes to ∞, this shows that the sequence (x_n) is Cauchy; let x be its limit. Then $d(x_n, x) \leq \frac{c^n}{1-c} r$ for all n, and applying f to both x_n and x yields that $d(x_{n+1}, f(x)) \leq \frac{c^{n+1}}{1-c} r$ for all n. So $f(x) = \lim x_n = x$ and x is a fixed point. □

A function $f : X \to Y$ is a *dense embedding* if it is an isometry and for every $y \in Y$ and $\epsilon > 0$ there exists $x \in X$ with $d(f(x), y) < \epsilon$.

Theorem 10.8. *(Cauchy completion) Every metric space can be densely embedded in a complete metric space. This space is unique up to an isometry that respects the embedding.*

Proof. Given a metric space X, let Y be the class of all Cauchy sequences in X. This is a subclass of the infinite power X^ω. Then let $\widetilde{X} = Y/\!\sim$ where we set $\vec{x} \sim \vec{y}$ if $d(x_n, y_n) \to 0$. Define a metric on $\widetilde{X}$ by letting $d([\vec{x}], [\vec{y}]) = \lim d(x_n, y_n)$. In order for this to make sense we need to show that this limit exists and is unaffected if $\vec{x}$ and $\vec{y}$ are replaced by equivalent sequences. We must also show that $d([\vec{x}], [\vec{y}])$ satisfies the axioms of a metric. This is all routine.

To see that $\widetilde{X}$ is complete, let (s_k) be a Cauchy sequence in $\widetilde{X}$. Each s_k is instantiated by a sequence in X. For each n, let x_n be a term in the instantiating sequence for s_n whose distance to any later term is at most $\frac{1}{n}$. Then the sequence (x_n) is Cauchy because if $d(s_i, s_j) \leq \frac{1}{n}$ and $i, j \geq n$ then $d(x_i, x_j) \leq \frac{3}{n}$. And (s_k) converges to $[\vec{x}]$ because if $k \geq n$ and $d(s_i, s_j) \leq \frac{1}{n}$ for all $i, j \geq k$ then the distance from s_k to $[\vec{x}]$ is at most $\frac{4}{n}$. This shows that $\widetilde{X}$ is complete.

Next, define a function $\iota : X \to \widetilde{X}$ by letting $\iota(x)$ be the element of $\widetilde{X}$ instantiated by the sequence $\overline{x}$ that takes the constant value x. It is clear that ι is an isometry, and it is a dense embedding because if $[\vec{x}]$ is any element of $\widetilde{X}$ then $[\overline{x}_n] \to [\vec{x}]$.

Finally, for uniqueness, suppose $\widetilde{X}$ and $\widetilde{X}'$ are complete metric spaces and $\iota : X \to \widetilde{X}$ and $\iota' : X \to \widetilde{X}'$ are dense embeddings. For any $x \in \widetilde{X}$, we can find a sequence (x_n) in X such that $\iota(x_n) \to x$. Since ι is an isometry, that means (x_n) is Cauchy, and hence so is the sequence $(\iota'(x_n))$, so by completeness of $\widetilde{X}'$ there exists $x' \in \widetilde{X}'$ with $\iota'(x_n) \to x'$. We can then define $f(x) = x'$ to get a function from $\widetilde{X}$ to $\widetilde{X}'$. Of course, we need to check that the choice of (x_n) doesn't matter, but that is easy because if (y_n) was another sequence in X with $\iota(y_n) \to x$ then (again using that ι is an isometry) we must have $d(x_n, y_n) \to 0$. So $d(\iota'(x_n), \iota'(y_n)) \to 0$, and this shows that the sequences $(\iota'(x_n))$ and $(\iota'(y_n))$ converge to the same point. So f is well-defined. It is clear that $f \circ \iota = \iota'$. We must also check that f is an isometry; that is again routine. The density of ι' entails that f is surjective. □

I do not claim that the image of X exists as a subclass of $\widetilde{X}$, though

of course it does if X is countable. (Recall that images of classes cannot always be assumed to exist; cf. Proposition 9.16.) That is why we had to be slightly careful with our definition of "dense embedding".

Because of the uniqueness property of the Cauchy completion, we typically just refer to *the* completion of a metric space.

The technique of Theorem 10.8 can be used to give another construction of $\mathbb{R}$, as the completion of $\mathbb{Q}$. We don't need to assume any knowledge of $\mathbb{R}$ to do this: just define a Cauchy sequence in $\mathbb{Q}$ to be a sequence of rational numbers (q_n) with the property that for every $n \in \mathbb{N}$ there exists $N \in \mathbb{N}$ such that $|q_i - q_j| < \frac{1}{n}$ for all $i, j \geq N$. Then define equivalence between Cauchy sequences and factor out that equivalence to get (another version of) $\mathbb{R}$.

Now we are ready to discuss metric topology.

Definition 10.9. Let X be a metric space. An *open ball* in X is a subclass of the form $B_r(x) = \{y \in X : d(x, y) < r\}$, for some $x \in X$ and $r > 0$; a *closed ball* is a subclass of the form $\overline{B}_r(x) = \{y \in X : d(x, y) \leq r\}$. Both balls are said to be *centered at* x and to have *radius* r. A subclass Y of X is *dense* in X if every open ball in X meets (has inhabited intersection with) Y. X is *separable* if there is a dense subset — not just subclass — $\mathring{X} \subseteq X$. If X is separable, then we say that a subclass $U \subseteq X$ is *open* if it is a union of countably many open balls, and that $C \subseteq X$ is *closed* if its complement is open. If $C \subseteq X$ contains the limits of all sequences in C that converge, then is it *sequentially closed*. (Here X need not be separable.)

We adopt the convention that a union of no classes is the empty class $\emptyset$. So $\emptyset$ is open in any separable metric space, because it is the union of no balls.

$\mathring{X}$ will be my generic notation for a dense subset of a metric space X.

We can realize the open balls in a metric space X as a family of subclasses indexed by $X \times (0, \infty)$; this allows us to effectively quantify over all open balls. The same class can be used to index the closed balls in X. If X is separable, so that we have a notion of openness in X, then the power class $\mathcal{P}(X \times (0, \infty))$, which previously was used to index all the subsets of $X \times (0, \infty)$, can now be used to index all the open subclasses of X. Each subset of $X \times (0, \infty)$ specifies countably many open balls, whose union will be the corresponding slice. So we can also effectively quantify over all open subclasses of a separable metric space (or, by taking complements, over all closed subclasses).

I have only defined the notions of openness and closedness in separable

spaces because analogous notions in nonseparable spaces do not behave well. For instance, in nonseparable spaces the intersection of two open balls might not be a countable union of open balls. Although most of the metric spaces we care about are separable, we will be interested in some nonseparable metric spaces, and they typically will in fact come with natural notions of openness and closedness, but these notions will not arise directly from the metric. The notion of sequential closedness, however, can still be useful in nonseparable metric spaces.

Now, about the terminology:

Proposition 10.10. *In any separable metric space, every open ball is open and every closed ball is closed.*

Proof. The first statement is trivial; for the second, consider a closed ball $\overline{B}_s(y)$ in a separable metric space X. Let $\mathring{X}$ be a dense subset of X, and consider all the balls $B_r(x)$ with $x \in \mathring{X}$, $d(x,y) > s$, and $r = d(x,y) - s$. There are countably many such balls and so their union exists and is open. One easily checks that this union is the complement of $\overline{B}_s(y)$, so $\overline{B}_s(y)$ is closed. □

Here is a simple observation about balls with rational radii centered at elements of a dense subset.

Lemma 10.11. *Let X be a separable metric space and let $\mathring{X} \subset X$ be a dense subset. Then there are countably many balls $B_s(y)$ with $y \in \mathring{X}$ and $s > 0$ rational, and every open class is a union of such balls.*

Proof. It will suffice to show that every open ball is a countable union of balls of this form. Let $B_r(x)$ be any open ball. We can find a sequence (x_n) of elements of $\mathring{X}$ that converge to x and a sequence of rational radii (r_n) that converge to r, such that $B_{r_n}(x_n) \subseteq B_r(x)$ for all n. Then $\bigcup B_{r_n}(x_n) = B_r(x)$. □

Proposition 10.12. *Let X be a separable metric space. Then*

(a) $\emptyset$ and X are open
(b) any union of countably many open classes is open
(c) any intersection of finitely many open classes is open.

Proof. (a) $\emptyset$ is open by definition. Since X is separable, it is definitely either empty or inhabited, as this is true of any set, in particular any dense subset of X. In the latter case, for any $x \in X$ we have $X = \bigcup_{n\in\mathbb{N}} B_n(x)$. So X is a countable union of open balls and hence is open.

(b) A countable union of countable unions of balls is a countable union of balls (cf. Theorem 9.8).

(c) Let $\mathring{X} \subseteq X$ be a dense subset. It will suffice to show that any intersection of two open classes is open, and since each open class is the union of a sequence of open balls, the equality

$$\left(\bigcup_i S_i\right) \cap \left(\bigcup_j T_j\right) = \bigcup_{i,j} (S_i \cap T_j)$$

reduces this to showing that the intersection of two open balls is always open.

Let $B_r(x)$ and $B_{r'}(x')$ be two open balls. For every $y \in \mathring{X} \cap B_r(x) \cap B_{r'}(x')$ let $s = \min(r - d(x,y), r' - d(x',y))$. Then $B_s(y) \subseteq B_r(x) \cap B_{r'}(x')$, so we get countably many open balls contained in $B_r(x) \cap B_{r'}(x')$. Their union equals $B_r(x) \cap B_{r'}(x')$ because any point z in that intersection belongs to $B_s(y)$ for $s = \frac{1}{2} \cdot \min(r - d(x,z), r' - d(x',z))$ and any $y \in \mathring{X}$ with $d(y,z) < s$. $B_s(y)$ is contained in the ball about y that contributes to the union described above, and as z was arbitrary, we conclude that this union is all of $B_r(x) \cap B_{r'}(x')$. □

Let X be a metric space and let Y be a subspace of X. Then the intersection of a ball in X with Y need not be a ball in Y, but, for X and Y separable, the intersection of any open class in X with Y is open in Y.

Corollary 10.13. *Let X be a separable metric space and let Y be a separable subspace of X. Then a subclass V of Y is open in Y if and only if there is an open class U of X with $V = U \cap Y$.*

Proof. The forward direction is easy, because any ball in Y is the intersection with Y of a ball in X. For the reverse direction, let $B_r^X(x)$ be an open ball in X and let $V = B_r^X(x) \cap Y$. Let $\mathring{Y}$ be a dense subset of Y. Then for any $y \in \mathring{Y}$ such that $d(x,y) < r$ let $s = r - d(x,y)$, so that $B_s^Y(y)$ is an open ball in Y which is contained in V. Moreover, the union of all such balls exactly equals V because any $z \in V$ belongs to $B_s(y)$ for any $y \in \mathring{Y}$ that satisfies $d(y,z) < \frac{1}{2}(r - d(x,z))$. Thus the intersection of any open ball in X with Y is open in Y, and taking countable unions (Proposition 10.12 (b)) yields the same result for any open $U \subseteq X$. □

Corollary 10.14. *In any separable metric space, every open class is separable and definitely either is or is not inhabited, and any two open classes definitely do or do not meet.*

Proof. Let $\mathring{X}$ be a dense subset of a separable metric space X and let $U \subseteq X$ be open. Then $\mathring{X} \cap U$ is dense in U because every element of U is the limit of a sequence in $\mathring{X}$, but since U is open all but finitely many terms of the sequence must belong to $\mathring{X} \cap U$. Now $\mathring{X} \cap U$ is a set (Proposition 9.7), so it definitely is or is not empty, and this implies that U definitely is or is not inhabited.

According to Proposition 10.12, the intersection of any two open classes is open, and therefore definitely is or is not inhabited by the first part of the corollary. □

Corollary 10.15. *Let X be a separable metric space. Then*

(a) $\emptyset$ and X are closed
(b) any intersection of countably many closed classes is closed
(c) any union of finitely many closed classes is closed.

The *sequential closure* of a subset A of a metric space X is the class $\overline{A}$ of all $x \in X$ which are the limits of sequences in A. This is a class because we can decide whether x belongs to $\overline{A}$ by checking, for each $n \in \mathbb{N}$, whether there exists $y \in A$ with $d(x, y) \leq \frac{1}{n}$. Since A is countable this question has a definite answer (LPO), so the question whether x belongs to $\overline{A}$ is definite.

More generally, if Y is a separable subspace of a metric space X, with dense subset $\mathring{Y}$, then we define the sequential closure $\overline{Y}$ of Y to simply equal the sequential closure of $\mathring{Y}$. Here we must check that the definition does not depend on the choice of dense subset, but that is easy. If $\mathring{Y}'$ is another dense subset of Y then for every Cauchy sequence (x_n) in $\mathring{Y}$ there is a Cauchy sequence (x'_n) in $\mathring{Y}'$ such that $d(x_n, x'_n) \to 0$, and vice versa. Thus the limits of Cauchy sequences in $\mathring{Y}$ are the same as the limits of Cauchy sequences in $\mathring{Y}'$.

In particular, Y is dense in X if and only if its sequential closure is X.

Proposition 10.16. *In any metric space, the sequential closure of any set is sequentially closed. In a separable metric space, any closed subspace is sequentially closed, and any separable sequentially closed subspace is closed.*

Proof. Let A be a set in some metric space X and let (x_i) be a sequence in the sequential closure $\overline{A}$ that converges to $x \in X$. For each $n \in \mathbb{N}$ there exists i such that $d(x, x_i) \leq \frac{1}{n}$, and then there exists $y_n \in A$ such that $d(x_i, y_n) \leq \frac{1}{n}$. We now have a sequence (y_n) in A that converges to x. We have shown that sequential closures are sequentially closed.

Now suppose X is separable and $C \subseteq X$ is closed, and let (x_n) be a sequence in C that converges to $x \in X$. Then x definitely either does or does not belong to C, and if it does not then it must belong to an open ball that is disjoint from C, contradicting the assumption that it is the limit of a sequence in C. Thus the limit of any convergent sequence of elements of C is in C.

Finally, suppose $D \subseteq X$ is separable and sequentially closed. Let $\mathring{D}$ be a dense subset of D and let $\mathring{X}$ be a dense subset of X. We must show that D is closed. To do this, observe that each $x \in \mathring{X}$ that satisfies $r = \inf\{d(x,y) : y \in \mathring{D}\} > 0$ is contained in the open ball $B_r(x)$, which does not meet $\mathring{D}$. The union U of all these balls is open. The complement U^c of U is therefore closed; we will show that $U^c = D$. Clearly $D \subseteq U^c$, because no element of any of the balls that make up U can be the limit of a sequence in $\mathring{D}$, showing that the limit of any convergent sequence in $\mathring{D}$ belongs to U^c. Conversely, let $y \in U^c$. I claim that any ball $B_s(y)$ centered at y must meet $\mathring{D}$. It definitely does or does not meet $\mathring{D}$ by the LPO, but if it did not then any $x \in \mathring{X}$ with $d(x,y) < \frac{s}{2}$ would be the center of a ball that contains y and contributes to U. This contradicts the choice of $y \in U^c$ and proves the claim. But now we can choose, for each $n \in \mathbb{N}$, a point $y_n \in \mathring{D}$ with $d(y, y_n) \leq \frac{1}{n}$. Then y is the limit of a sequence in $\mathring{D}$ and hence belongs to D. □

Corollary 10.17. *In any separable metric space, the sequential closure of any separable subspace is closed.*

Because of this result, the sequential closure of a separable subspace of a separable metric space may be simply referred to as its *closure*.

We can now prove the Baire category theorem.

Theorem 10.18. *(Baire category) Let X be a separable complete metric space and let U_n, $n \in \mathbb{N}$, be dense open classes in X. Then $\bigcap U_n$ is dense in X.*

Proof. Let $x \in X$. It will suffice to show that for each $m \in \mathbb{N}$ there exists $y \in \bigcap U_n$ with $d(x,y) \leq \frac{2}{m}$. Fix m.

By density there exists $x_1 \in U_1$ with $d(x, x_1) \leq \frac{1}{m}$, and since U_1 is open there exists $r_1 > 0$ such that $B_{2r_1}(x_1) \subseteq U_1$. We can assume $r_1 \leq \frac{1}{m}$. Let $V_1 = B_{r_1}(x_1)$.

Next, since $U_2 \cap V_1$ is open (Proposition 10.12 (c)) and inhabited (by density of U_2), we can find $x_2 \in U_2 \cap V_1$ and $r_2 > 0$ such that $B_{2r_2}(x_2) \subseteq U_2 \cap V_1$. We can assume $r_2 \leq \frac{1}{2m}$. Let $V_2 = B_{r_2}(x_2)$.

In the general step, given x_n, r_n, and V_n, find $x_{n+1} \in U_{n+1} \cap V_n$ and $r_{n+1} < \frac{1}{(n+1)m}$ such that $B_{2r_{n+1}}(x_{n+1}) \subseteq U_{n+1} \cap V_n$, and let $V_{n+1} = B_{r_{n+1}}(x_{n+1})$.

The sequence (x_n) is Cauchy because any two points in V_n are at most $\frac{2}{nm}$ apart, and x_k belongs to V_n for all $k \geq n$. The limit $y = \lim x_n$ belongs to every U_n because it is a limit of elements of $V_n = B_{r_n}(x_n)$ while $B_{2r_n}(x_n) \subseteq U_n$. We also have $d(x, y) \leq \frac{2}{m}$ because y is a limit of elements of V_1. This completes the proof. □

The next definition is important enough to be formally numbered.

Definition 10.19. Let X and Y be separable metric spaces. A function $f : X \to Y$ is *continuous* if the inverse image of any open class in Y is an open class in X. It is a *homeomorphism* if it is bijective and its inverse is also continuous. X and Y are *homeomorphic*, written $X \cong Y$, if there exists a homeomorphism between them.

Equivalently, $f : X \to Y$ is continuous if the inverse image of any open ball in Y is an open class in X. The assertion that X and Y are homeomorphic involves an unbounded existential quantifier and any affirmation of homeomorphism should always be supported by an explicit construction (but see the comments following Corollary 10.49 and Theorem 10.93). We can also express ideas like "if two metric spaces are homeomorphic and one is separable then so is the other" using the general equivalence of $(\exists x)\phi(x) \to \psi$ and $(\forall x)(\phi(x) \to \psi)$. Thus we can take $\phi(f) =$ "f is a homeomorphism between X and Y" and say "for any X, Y, and f, if X and Y are metric spaces and f is a homeomorphism between them then X is separable if and only if Y is separable."

I only define continuity for separable spaces because these are the spaces for which we have a good notion of metric openness. A variant notion of continuity, *sequential continuity*, says that $x_n \to x$ in X implies $f(x_n) \to f(x)$ in Y; this applies to functions between any two metric spaces, with no assumption of separability. But in fact it is equivalent to ordinary continuity in the separable setting:

Theorem 10.20. *Let $f : X \to Y$ be a function between separable metric spaces. Then the following are equivalent:*

(a) f is continuous
(b) f is sequentially continuous

(c) for every $x \in X$ and $\epsilon > 0$, there exists $\delta > 0$ such that $d(x,y) < \delta$ implies $d(f(x), f(y)) < \epsilon$.

Proof. Let $\mathring{X}$ be a dense subset of X.

Assume f is continuous and $x_n \to x$ in X. Then it definitely is or is not the case that for every $m \in \mathbb{N}$ there exists $N \in \mathbb{N}$ such that $d(f(x_n), f(x)) < \frac{1}{m}$ for all $n \geq N$ (LPO). If this fails for some m then $U = f^{-1}(B_{1/m}(f(x)))$ is an open class which contains x and excludes infinitely many x_n; letting $s > 0$ be small enough that $B_s(x) \subseteq U$, we then contradict $x_n \to x$. Therefore the condition does not fail for any m, and this shows that $f(x_n) \to f(x)$. We have proven (a) $\to$ (b).

Next, suppose f is sequentially continuous. We will prove (c). Fix $x \in X$ and $\epsilon > 0$; it will suffice to find $\delta > 0$ such that $f(y) \in B_\epsilon(f(x))$ whenever $y \in B_\delta(x) \cap \mathring{X}$, because any $y \in B_\delta(x)$ is the limit of a sequence (y_n) in $B_\delta(x) \cap \mathring{X}$, and $d(f(y_n), f(x)) < \epsilon$ for all n plus $f(y) = \lim f(y_n)$ entails $d(f(y), f(x)) \leq \epsilon$.

Now for each n there definitely either does or does not exist $y \in \mathring{X}$ with $d(x,y) < \frac{1}{n}$ and $d(f(x), f(y)) \geq \epsilon$ (LPO). Thus there either exists such a y for all n or there does not exist such a y for some n (LPO). In the former case, for each $n \in \mathbb{N}$ let y_n be such a point; then $y_n \to x$ but $f(y_n) \not\to f(x)$, a contradiction. Therefore for some n there does not exist such a y, which means that $\delta = \frac{1}{n}$ works. We have proven (b) $\to$ (c).

We complete the proof by showing that (c) $\to$ (a). Assume (c) holds and let $z \in Y$ and $s > 0$; we must show that $U = f^{-1}(B_s(z))$ is open. Index the balls $B_r(x)$ with $x \in \mathring{X}$, $r > 0$ rational, and $\sup_{y \in B_r(x) \cap \mathring{X}} d(f(y), z) < s$ as B_i. We will show that $U = \bigcup B_i$.

For any such ball $B_r(x)$ and any $y \in B_r(x)$, I claim that $d(f(y), z) < s$. If this failed for some $B_r(x)$ and y, then for sufficiently small $\epsilon > 0$ the ball $B_\epsilon(f(y))$ would be disjoint from $f(B_r(x) \cap \mathring{X})$. (Note that the image of $B_r(x) \cap \mathring{X}$ exists because it is a set.) But then there would exist elements of $B_r(x) \cap \mathring{X}$ which are arbitrarily close to y, none of whose images are within ϵ of $f(y)$, contradicting condition (c). Therefore $d(f(y), z) < s$, and this shows that $y \in U$. So each B_i is contained in U. Conversely, for any $y \in U$ we have $d(f(y), z) < s$, and letting $\epsilon < s - d(f(y), z)$, we can find $\delta > 0$ such that f maps everything in $B_\delta(y)$ into $B_\epsilon(f(y))$. Then we can find $x \in \mathring{X}$ and $r > 0$ rational such that $y \in B_r(x) \subseteq B_\delta(y)$. Thus y belongs to some B_i, and we conclude that U is contained in $\bigcup B_i$. Thus $U = \bigcup B_i$, as desired. □

Corollary 10.21. *Any Lipschitz function between metric spaces is sequen-*

tially continuous. If the spaces are separable, any Lipschitz function between them is continuous.

(The first statement is easy, and the second then follows Theorem 10.20.)

For any metric space X and any separable subspace $Y \subseteq X$, let $\mathring{Y}$ be a dense subset of Y and define $d_Y : X \to \mathbb{R}$ by letting $d_Y(x) = \inf\{d(x,y) : y \in \mathring{Y}\}$. This is the *distance to* Y. One easily checks that it is nonexpansive (and therefore sequentially continuous by Corollary 10.21) and that it does not depend on the choice of $\mathring{Y}$. The class $d_Y^{-1}(0)$ is precisely the sequential closure of Y.

Proposition 10.22. *(Urysohn's lemma for metric spaces) Let C and D be disjoint separable closed classes in a separable metric space X. Then there is a continuous function $f : X \to [0,1]$ with $C \subseteq f^{-1}(0)$ and $D \subseteq f^{-1}(1)$.*

Proof. Define

$$f(x) = \frac{d_C(x)}{d_C(x) + d_D(x)}.$$

It is clear that f is constantly zero on C and constantly one on D. Continuity of f follows from continuity of the functions d_C and d_D. □

10.2 Infinite series

Let (a_n) be a bounded sequence of complex numbers. Then we write $\lim a_n = a$ if for every $\epsilon > 0$ there exists $N \in \mathbb{N}$ with $|a - a_n| < \epsilon$ for all $n \geq N$. This is just the specialization of Definition 10.5 to the complex plane. If the a_n are all real, then we can also define $\limsup a_n$ to be the limit of the sequence $b_n = \sup_{k \geq n} a_k$, and $\liminf a_n$ to be the limit of the sequence $c_n = \inf_{k \geq n} a_k$. Since the b_n's decrease and the c_n's increase, and both sequences are bounded by hypothesis, they always converge.

Definition 10.23. Let (a_i) be a sequence of complex numbers. The *partial sums* of the series $\sum a_i$ are the sums $s_n = a_1 + \cdots + a_n$. We say that $\sum a_i$ *converges* if $a = \lim s_n \in \mathbb{C}$ exists. In that case we write $\sum a_i = a$ and say that a is the *sum* of the series. The series $\sum a_i$ converges *absolutely* if $\sum |a_i|$ converges. A series *diverges* if it does not converge.

For series of positive terms, we may write "$\sum a_n < \infty$" for "$\sum a_n$ converges".

Proposition 10.24. *The series $\sum a_n$ converges if and only if for every $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that $m > n \geq N$ implies $|s_m - s_n| = |\sum_{k=n+1}^{m} a_k| \leq \epsilon$.*

This is an immediate consequence of the completeness of $\mathbb{C}$. It follows that every absolutely convergent series is convergent, and also that if $\sum a_i$ converges then the sequence (a_i) must converge to 0.

Proposition 10.25. *Let $-1 < r < 1$. Then $\sum_{n=0}^{\infty} r^n = \frac{1}{1-r}$.*

Proof. We have

$$\frac{1}{1-r} - (1 + r + \cdots + r^n) = \frac{1}{1-r}(1 - (1 - r^{n+1})) = \frac{r^{n+1}}{1-r},$$

and $\frac{r^{n+1}}{1-r}$ converges to zero. □

Proposition 10.26. *(Ratio test) Let $\sum a_n$ be a series of nonzero complex numbers. If $\limsup |\frac{a_{n+1}}{a_n}| < r < 1$ then $\sum a_n$ converges absolutely. If $|a_{n+1}| \geq |a_n| \neq 0$ for all but finitely many n, then $\sum a_n$ diverges.*

Proof. Removing finitely many terms from a series does not affect whether it converges. So for the first part, we can remove finitely many terms and assume that $|\frac{a_{n+1}}{a_n}| < r < 1$ for all n. Then inductively $|a_{n+1}| < r^n |a_1|$. Since $\sum r^n$ converges, it follows that $\sum a_n$ converges absolutely. If $|a_{n+1}| \geq |a_n|$ for all but finitely many n, then (a_n) cannot converge to 0 and hence $\sum a_n$ cannot converge. □

Proposition 10.27. *(Root test) Let $\sum a_n$ be a series of complex numbers. If $\limsup \sqrt[n]{|a_n|} < r < 1$ then $\sum a_n$ converges absolutely. If $\sqrt[n]{|a_n|} \geq 1$ for infinitely many n, then $\sum a_n$ diverges.*

Proof. As in the proof of the last proposition, for the first part we can remove finitely many terms and assume that $\sqrt[n]{|a_n|} < r < 1$ for all n. Then $|a_n| < r^n$ for all n, and since $\sum r^n$ converges this implies that $\sum a_n$ converges absolutely. If $\sqrt[n]{|a_n|} \geq 1$ for infinitely many n, then $|a_n| \geq 1$ for infinitely many n and hence $\sum a_n$ cannot converge. □

Proposition 10.28. *(Radius of convergence) Let (a_n) be a sequence of complex numbers. Then either $\sum a_n z^n$ converges absolutely for all $z \in \mathbb{C}$, or it converges only for $z = 0$, or else there exists $R > 0$ such that $\sum a_n z^n$ converges absolutely if $|z| < R$ and diverges if $|z| > R$. The value of R is given by the formula $\frac{1}{R} = \limsup \sqrt[n]{|a_n|}$.*

Proof. For each k let $b_k = \sup\{\sqrt[n]{|a_n|} : n \geq k\}$. If some b_k is infinite, then $\sum a_n z^n$ converges only for $z = 0$ by the root test (Proposition 10.27). Otherwise the sequence (b_k) is nonincreasing and bounded below by 0, and hence it converges to $r = \limsup \sqrt[n]{|a_n|} \geq 0$. If $r = 0$ then $\sqrt[n]{|a_n z^n|} = |z| \sqrt[n]{|a_n|}$ converges to 0 for any z and hence by the root test the series $\sum a_n z^n$ converges absolutely. Otherwise let $R = \frac{1}{r}$. If $|z| < s < R$ then $\sqrt[n]{|a_n|} < \frac{1}{s}$ entails $\sqrt[n]{|a_n z^n|} \leq \frac{|z|}{s} < 1$, for all but finitely many n. By the root test, $\sum a_n z^n$ converges for all such z. If $|z| > R$ then $\sqrt[n]{|a_n|} > \frac{1}{|z|}$ and hence $|a_n z^n| > 1$ for infinitely many n. So $\sum a_n z^n$ diverges. □

The value of R in Proposition 10.28 is the *radius of convergence* of $\sum a_n z^n$.

Proposition 10.29. *Let $\sum a_n$ and $\sum b_n$ be absolutely converging series, summing to A and B, respectively. Define $c_n = a_1 b_n + a_2 b_{n-1} + \cdots + a_n b_1$ for each n. Then $\sum c_n$ converges absolutely to AB.*

Proof. Set $A_n = a_1 + \cdots + a_n$, $B_n = b_1 + \cdots + b_n$, and $C_n = c_1 + \cdots + c_n$, for all n. Setting $\beta_n = B_n - B$, we then have $C_n = a_1\beta_n + \cdots + a_n\beta_1 + A_n B$. Also let $\alpha = \sum |a_n|$. We can assume $\alpha > 0$.

Since $B_n \to B$ and $A_n \to A$, given any $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that $|\beta_n| \leq \frac{\epsilon}{\alpha}$ and $|A_n - A| \leq \frac{\epsilon}{|B|}$ for all $n \geq N$. There then also exists $M \in \mathbb{N}$ such that

$$|a_n| \leq \frac{\epsilon}{|\beta_1| + \cdots + |\beta_N|}$$

for all $n \geq M$.

Let $n \geq M + N$. We estimate

$$\begin{aligned} |C_n - AB| &= \left| \sum_{i=1}^{n} a_{n+1-i}\beta_i + (A_n - A)B \right| \\ &\leq \sum_{i=1}^{N} |a_{n+1-i}||\beta_i| + \sum_{i=N+1}^{n} |a_{n+1-i}||\beta_i| + |A_n - A||B| \end{aligned}$$

and the previous inequalities show that each of the three terms on the right is at most ϵ. Thus $C_n \to AB$. □

We can also sum infinite series of real- or complex-valued functions, by letting $\sum f_n$ be the function whose value at x is $\sum f_n(x)$, if it exists. This is the *pointwise sum* of the functions f_n. Formally, we handle infinitely many functions $f_i : X \to \mathbb{R}$ as an indexed family of subclasses of $X \times \mathbb{R}$ with indexing class $\mathbb{N}$.

If X is a metric space and f_1, f_2, ... are complex-valued functions on X, then we say that the f_n converge *uniformly* to a function f if for every $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that $n \geq N$ implies $|f(x) - f_n(x)| \leq \epsilon$ for all $x \in X$.

Proposition 10.30. *If f_1, f_2, ... are continuous complex-valued functions on a separable metric space X and they converge uniformly to a function f, then f is also continuous.*

This result is perhaps most easily seen as a consequence of Theorem 10.20 (b).

Corollary 10.31. *Let f_1, f_2, ... be continuous complex-valued functions on a separable metric space X, let $\mathring{X}$ be a dense subset of X, and define $\|f_n\|_\infty = \sup_{x \in \mathring{X}} |f_n(x)|$. Suppose $\sum \|f_n\|_\infty < \infty$. Then $\sum f_n$ is summable at every point and it converges uniformly to a continuous function.*

We also have a similar result for sums of Lipschitz functions.

Proposition 10.32. *Let f_1, f_2, ... be complex-valued functions on a metric space and suppose f_n is c_n-Lipchitz for all n, with $c = \sum c_n < \infty$. If the pointwise sum $\sum f_n$ exists everywhere then it is c-Lipschitz.*

10.3 Compactness

Compactness is another property of metric spaces of basic importance.

Definition 10.33. Let X be a metric space. An *ϵ-net* in X is a set $S \subseteq X$ such that $\bigcup_{x \in S} B_\epsilon(x) = X$. X is *totally bounded* if there is a finite ϵ-net for every $\epsilon > 0$. X is *compact* if it is separable and whenever C_i ($i \in \mathbb{N}$) are closed classes such that $\bigcap_1^n C_i$ is inhabited for all n, then $\bigcap_1^\infty C_i$ is inhabited. It is *boundedly compact* if every closed ball is compact, *locally compact* if it is the union of a sequence of open classes each of whose closures is compact, and *σ-compact* if it is the union of a sequence of compact subspaces. Open subclasses U_i ($1 \leq i \leq n$ or $i \in \mathbb{N}$) of X are said to *cover* X if their union equals X.

Note that boundedly compact implies locally compact implies σ-compact. Also note that separability is part of all of these conditions.

Lemma 10.34. *Every separable metric space definitely is or is not totally bounded. Any dense subset is totally bounded if and only if the whole space*

is totally bounded. If the space fails to be totally bounded then there exists $\epsilon > 0$ *and a sequence* (x_n) *with* $d(x_m, x_n) \geq \epsilon$ *for all distinct* m *and* n.

Proof. Let $\mathring{X} = \{x_1, x_2, \ldots\}$ be a dense subset of a separable metric space X. For each $n \in \mathbb{N}$ it definitely either is or is not the case that $\mathring{X}$ is contained in the union $B_{1/n}(x_1) \cup \cdots \cup B_{1/n}(x_k)$ for some $k \in \mathbb{N}$ (LPO). If this is the case then X is contained in the union $B_{2/n}(x_1) \cup \cdots \cup B_{2/n}(x_k)$. By the LPO again, either the statement is true for all n, in which case X is totally bounded, or there is some n for which it is false, in which case there must be an infinite subset of $\mathring{X}$ any two elements of which are at least $\frac{1}{n}$ apart, showing that X is not totally bounded.

We have established that X is totally bounded if and only if $\mathring{X}$ is totally bounded, and that if $\mathring{X}$ is not totally bounded then for some $n \in \mathbb{N}$ it contains an infinite sequence any two of whose elements are at least $\frac{1}{n}$ apart. □

Theorem 10.35. *Let* X *be a separable metric space. The following are equivalent:*

(a) X *is compact*
(b) every sequence in X *has a convergent subsequence*
(c) X *is complete and totally bounded.*

Proof. Let $\mathring{X} = \{x_1, x_2, \ldots\}$ be a dense subset of X.

Suppose X is compact and let (y_n) be a sequence in X. For each n let $\overline{B}_n$ be the closure of $B_n = \{y_n, y_{n+1}, \ldots\}$. Each $\overline{B}_n$ is closed by Corollary 10.17, and clearly any finite intersection of $\overline{B}_n$'s is inhabited, so by compactness there exists $y \in \bigcap \overline{B}_n$. We can now choose $n_1 \in \mathbb{N}$ such that $d(y_{n_1}, y) \leq 1$, then $n_2 > n_1$ such that $d(y_{n_2}, y) \leq \frac{1}{2}$, etc. This yields a subsequence (y_{n_k}) of (y_n) which converges to y. We have proven (a) $\rightarrow$ (b).

Next, suppose every sequence in X has a convergent subsequence. This clearly implies that X is complete. Suppose X is not totally bounded. According to Lemma 10.34, there exists $\epsilon > 0$ and a sequence (x_n) in X any two of whose elements are at least ϵ apart. Such a sequence clearly cannot have a convergent subsequence, a contradiction. By Lemma 10.34 it definitely is or is not the case that X is totally bounded, so we conclude that X is totally bounded. This proves (b) $\rightarrow$ (c).

We now prove (c) $\rightarrow$ (b). Suppose X is complete and totally bounded and let (y_n) be a sequence in X. By total boundedness there are finitely

many balls of radius 1 which cover X, and then there must be an infinite set of indices $I_1 \subseteq \mathbb{N}$ such that $\{y_n : n \in I_1\}$ is contained in a single ball of radius 1. Let n_1 be the smallest element of I_1. Then, by similar reasoning, there is an infinite set of indices $I_2 \subseteq I_1 \setminus n_1$ such that $\{y_n : n \in I_2\}$ is contained in a single ball of radius $\frac{1}{2}$; let n_2 be the smallest element of I_2. Continuing in this way, using radius $\frac{1}{k}$ at the kth stage, we get a subsequence (y_{n_k}) of (y_n) with the property that $d(y_{n_i}, y_{n_j}) < \frac{2}{N}$ whenever $i, j \geq N$, for all $N \in \mathbb{N}$. Thus the subsequence (y_{n_k}) is Cauchy, and it converges by completeness. So (y_n) has a convergent subsequence.

Finally, we prove (b) $\to$ (a). Assume every sequence has a convergent subsequence and let C_i ($i \in \mathbb{N}$) be closed classes such that $\bigcap_1^n C_i$ is inhabited for all n. For each n let x_n belong to this intersection. Then the sequence (x_n) has a convergent subsequence, and the limit of that subsequence must belong to every C_n because $x_m \in C_n$ for all $m \geq n$. We conclude that $\bigcap C_n$ is inhabited. □

The equivalence of these conditions has a number of consequences.

Theorem 10.36. *(Extreme value theorem) Let $f : X \to \mathbb{R}$ be a continuous function on a compact space. Then f is bounded above and below and has a supremum and an infimum, both of which it attains.*

Proof. Let $\mathring{X}$ be a dense subset of X. It definitely either is or is not the case that $f(\mathring{X})$ is bounded above and below. Suppose it is not bounded above. Then we can find a sequence (x_n) in $\mathring{X}$ such that $(f(x_n))$ increases without bound. By compactness, we can find a convergent subsequence $x_{n_k} \to x$, and then by continuity $f(x_{n_k}) \to f(x)$. But this is impossible because $(f(x_{n_k}))$ increases without bound. Thus $f(\mathring{X})$ is bounded above, and it is bounded below by similar reasoning.

Any point in X is the limit of a sequence in $\mathring{X}$, so continuity now implies that f is bounded above and below on all of X. This argument also shows that $\sup_{x \in \mathring{X}} f(x)$ is an upper bound for $f(y)$, for all $y \in X$.

Finally, let $a = \sup_{x \in \mathring{X}} f(x)$. By the comment just made, we have $f(x) \leq a$ for all $x \in X$. Then we can find a sequence (x_n) in $\mathring{X}$ such that $f(x_n) \to a$, and pass to a convergent subsequence $x_{n_k} \to x$, which by continuity implies $f(x) = a$. So f attains its supremum, and it attains its infimum similarly. □

Theorem 10.37. *(Lebesgue number) Let U_i ($i \in \mathbb{N}$) be open classes in a compact metric space X which cover X. Then there exists $\delta > 0$ such that for every $x \in X$ the ball $B_\delta(x)$ is wholly contained in some U_i.*

Proof. Since every open class is a countable union of open balls, we can assume the U_i are all balls, say $U_i = B_{r_i}(y_i)$.

Let $\mathring{X} = \{x_1, x_2, \ldots\}$ be a dense subset of X. By the LPO, it definitely is or is not the case that there exists k such that every x_n has distance at most $r_i - \frac{1}{k}$ to y_i, for some i. If this fails, then for each k we can find n_k such that $d(x_{n_k}, y_i) > r_i - \frac{1}{k}$ for all i. Let x be the limit of some convergent subsequence of (x_{n_k}); this point must belong to some U_i, but at the same time its distance to y_i is at least r_i for all i, a contradiction. It follows that there must exist k such that every x_n has distance at most $r_i - \frac{1}{k}$ to y_i, for some i. Then each ball $B_{1/k}(x_n)$ is wholly contained in some U_i, and this implies that every ball $B_{1/2k}(x)$, for any $x \in X$, is wholly contained in some U_i. □

"Every open cover has a finite subcover" means that whenever U_i ($i \in \mathbb{N}$) are open classes which cover X, it is also the case that $U_1, \ldots, U_n$ cover X for some n. Here is another equivalent formulation of compactness.

Corollary 10.38. *A complete separable metric space is compact if and only if every open cover has a finite subcover.*

Proof. Let X be a compact metric space. Then it is complete by Theorem 10.35 (a) → (c). Also, if U_i ($i \in \mathbb{N}$) cover X then by the existence of Lebesgue number (Theorem 10.37) there exists δ such that every open ball of radius δ is contained in some U_i. By Theorem 10.35 (c) we can find a finite δ-net, and the balls of radius δ about the points in this set cover X (since it was a δ-net), and each one is contained in some U_i. So X is the union of finitely many U_i's.

Conversely, let X be a complete separable metric space and suppose that every open cover of X has a finite subcover. Letting $\mathring{X} = \{x_1, x_2, \ldots\}$ be a dense subset, for any $\epsilon > 0$ the open balls $B_\epsilon(x_k)$ cover X, so by assumption we have $X = B_\epsilon(x_1) \cup \cdots \cup B_\epsilon(x_n)$ for some n, which means that $\{x_1, \ldots, x_n\}$ is an ϵ-net. We have shown that X is totally bounded, so it is compact by Theorem 10.35 (c) → (a). □

Theorem 10.39. *(Bolzano-Weierstrass) Every sequence in the unit interval* $[0, 1]$ *in* $\mathbb{R}$ *has a convergent subsequence. The unit interval in* $\mathbb{R}$ *is compact.*

Proof. The first statement follows from the second via Theorem 10.35. The second statement follows from Theorem 10.35 plus the fact that $[0, 1]$

is complete and totally bounded. Total boundedness is trivial, and completeness of $[0,1]$ follows from completeness of $\mathbb{R}$ (Proposition 10.6). □

The *zero class* of any function $f : X \to \mathbb{R}$ is the class $f^{-1}(0)$. The next result is another important consequence of Theorem 10.35.

Theorem 10.40. *The zero class of any continuous real-valued function on a compact metric space is compact.*

Proof. Let X be a compact metric space and suppose $f : X \to \mathbb{R}$ is continuous. Then $f^{-1}(0)$ is a closed class in a compact space and it is trivial that $f^{-1}(0)$ satisfies, e.g., condition (b) from Theorem 10.35. But we must also show that $f^{-1}(0)$ is separable, which is a requirement in our definition of compactness.

Let $\mathring{X} = \{x_1, x_2, \ldots\}$ be a dense subset of X. For each rational $r > 0$ and each $n \in \mathbb{N}$, it definitely is or is not the case that $\inf_{x \in \mathring{X} \cap B_r(x_n)} |f(x)| = 0$. If this is the case, then we can find a sequence (y_k) in $\mathring{X} \cap B_r(x_n)$ with $f(y_k) \to 0$. Then by compactness, extract a convergent subsequence and conclude that $\overline{B}_r(x_n) \cap f^{-1}(0)$ is inhabited.

Now for each r and n such that $\inf_{x \in \mathring{X} \cap B_r(x_n)} |f(x)| = 0$, choose $y_{n,r} \in \overline{B}_r(x_n)$ with $f(y_{n,r}) = 0$. Let A be the set of all these points $y_{n,r}$. It is clearly contained in the zero class of f, and it is dense in the zero class because for any $z \in f^{-1}(0)$ and any rational $\epsilon > 0$ there exists $x \in \mathring{X}$ such that $z \in B_\epsilon(x)$, which means that some $y \in A$ will lie in $\overline{B}_\epsilon(x)$ and hence satisfy $d(y,z) < 2\epsilon$. □

It is easy to see that the union of finitely many compact subspaces of a metric space will always be compact. The same is true of infinite intersections; again, separability must also be addressed.

Corollary 10.41. *If K_i $(i \in \mathbb{N})$ are compact subspaces of a metric space, then $\bigcap K_n$ is compact.*

Proof. For each $n \in \mathbb{N}$ and $x \in K_1$ let $f_n(x) = \frac{1}{2^n} \cdot \min(d_{K_n}(x), 1)$ and define $f(x) = \sum f_n(x)$. Since all the terms of this sum are positive, $f(x)$ is simply the supremum of the partial sums $\sum_1^n f_i(x)$. These partial sums are bounded by $\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots = 1$, so the supremum exists for all x.

Each $f_n : K_1 \to \mathbb{R}$ is 2^{-n}-Lipschitz, and then f is 1-Lipschitz by Proposition 10.32. So f is continuous by Corollary 10.21. The zero class of f is exactly $\bigcap K_n$, so we conclude from Theorem 10.40 that this intersection is compact. □

Every open class in any separable metric space X is separable (Corollary 10.14). Remarkably, in σ-compact spaces every closed class C is also separable. This is another consequence of Theorem 10.40. The obvious idea — use countable choice to select one element from each inhabited intersection $B_r(x) \cap C$ with x in some dense subset and $r > 0$ rational — is no good because it is not clear how to diagnose which of these intersections are inhabited. In fact the condition "$B_r(x) \cap C$ is inhabited" actually is definite when C is a closed class in a σ-compact metric space, but I do not see how to prove that without already knowing that C is separable.

Say that a metric space is *hereditarily separable* if every closed class in it is separable.

Corollary 10.42. *Every σ-compact metric space is hereditarily separable.*

Proof. Let C be a closed class in a compact metric space. Then its complement is open and therefore the union of a sequence of open balls $B_{r_n}(x_n)$. We can assume each r_n is at most 1. For each n define

$$f_n(x) = \frac{1}{2^n} \cdot \max(r_n - d(x, x_n), 0).$$

This function peaks at x_n, then steadily goes down as we move away from x_n, but stops decreasing and stays zero outside of $B_{r_n}(x_n)$.

As in the proof of Corollary 10.41, we can define $f = \sum f_n$ and show that it is nonexpansive and therefore continuous. The zero class of f is exactly C, so separability of C follows from Theorem 10.40. Any closed class in a σ-compact metric space is a union of countably many closed classes in compact metric spaces (Corollary 10.13), so we then get separability of any closed class in any σ-compact metric space. □

We will see below that $\mathbb{R}^n$ is σ-compact for either of the metrics discussed in Section 10.1 (the Heine-Borel theorem, Corollary 10.46), so in particular, every closed subclass of $\mathbb{R}^n$ is separable.

Corollary 10.43. *In any σ-compact metric space, a subclass is closed if and only if it is separable and sequentially closed.*

This follows from Proposition 10.16.

Next, here is a very useful result about compact spaces.

Theorem 10.44. *Let $f : X \to Y$ be a continuous function between separable metric spaces and suppose X is σ-compact. Then the image $f(X)$ exists as a subclass of Y. If X is compact then so is $f(X)$. If X is compact and f is injective, then f is a homeomorphism between X and $f(X)$.*

Proof. It will suffice to prove the first statement when X is compact; the σ-compact case follows immediately.

Suppose X is compact, let $\mathring{X}$ be a dense subset of X, and let C be the closure of $f(\mathring{X})$ in Y. I claim that C is precisely the image of X. First, every $x \in X$ is the limit of a sequence in $\mathring{X}$, so by continuity $f(x)$ is the limit of a sequence in $f(\mathring{X})$ and hence belongs to C. Conversely, let $y \in C$ and find a sequence $(f(x_n))$ in $f(\mathring{X})$ that converges to it. Then (x_n) is a sequence in X, so by compactness we can pass to a convergent subsequence. Then the limit of that subsequence maps to y by continuity. We conclude that C is the image of X.

The proof that $f(X)$ is compact is similar. It is separable because, as we have just seen, it is the closure of the image of a dense subset of X. Given any sequence $(f(x_n))$ in $f(X)$, we can find a subsequence (x_{n_k}) of (x_n) that converges, say to $x \in X$. Then $f(x_{n_k})$ converges to $f(x)$ by continuity. Thus every sequence in $f(X)$ has a convergent subsequence. By Theorem 10.35 that shows $f(X)$ is compact.

Finally, suppose X is compact and f is injective. Then it is a continuous bijection between X and $f(X)$, and we must show that its inverse is also continuous. Let $(f(x_n))$ be a sequence in $f(X)$ that converges to some point $f(x)$. Then there is a subsequence (x_{n_k}) of (x_n) that converges in X, say to $y \in X$. But then continuity of f entails that $f(x_{n_k}) \to f(y)$, and we already knew that $f(x_{n_k}) \to f(x)$, so by injectivity we must have $y = x$. In fact, this shows that any convergent subsequence of (x_n) must converge to x. But then (x_n) must itself converge to x, as otherwise there would be infinitely many terms, and hence a subsequence, lying outside of some open ball about x, and no subsequence of this subsequence could converge to x, a contradiction. Since the condition "$x_n \to x$" is definite we conclude that $f(x_n) \to f(x)$ in $f(X)$ implies $x_n \to x$ in X, so that $f^{-1} : f(X) \to X$ is continuous. □

We will now examine the relation between being closed and being compact.

Proposition 10.45. *Let X be a separable metric space. Then every compact subspace of X is closed. If X is compact, then every closed subclass of X is compact.*

Proof. Let K be a compact subspace of X. Then K has a dense subset $\mathring{K}$; let us show that K equals the closure of $\mathring{K}$. Every element of K is a limit of elements of $\mathring{K}$, so K is contained in the closure of $\mathring{K}$, and conversely, by

compactness any sequence in $\mathring{K}$ that converges must converge to a limit in K, so the reverse inclusion also holds. Thus K is closed by Corollary 10.17.

If X is compact then every closed subclass of it is separable by Corollary 10.42, and then Theorem 10.35 (b) shows that every closed subclass of X is compact. □

Say that a subclass of a metric space is *bounded* if it is contained in some ball.

Corollary 10.46. *(Heine-Borel) A subspace of $\mathbb{R}^n$ is compact if and only if it is closed and bounded.*

Proof. The same sequences in $\mathbb{R}^n$ converge to the same points relative to either of the two metrics (l^∞ and Euclidean) mentioned at the beginning of this section — or, indeed, relative to any reasonable metric on $\mathbb{R}^n$. So let us work with the l^∞ metric.

Let $K \subseteq \mathbb{R}^n$ be compact. Then K is closed by Proposition 10.45. Letting $\mathring{K}$ be a dense subset of K, it definitely either is or is not the case that there are elements of $\mathring{K}$ arbitrarily far from the origin; if there are, then we can find a sequence (x_k) in $\mathring{K}$ satisfying $d(x_k, 0) \geq k$ for all k. It is clear that no subsequence of this sequence can converge, contradicting the assumption of compactness. We conclude that there exists $N \in \mathbb{N}$ such that $d(x, 0) \leq N$ for all $x \in \mathring{K}$, and this shows that K is bounded.

Conversely, let $K \subseteq \mathbb{R}^n$ be closed and bounded. Then it is contained in the cube $[-N, N]^n$ for some $N \in \mathbb{N}$. It will suffice to show that $[-N, N]^n$ is compact, as compactness of K will then follow from Proposition 10.45.

Let us show that $[-N, N]^n$ is complete and totally bounded. Total boundedness is easy. For completeness, let $(\vec{x}_k)$ be a Cauchy sequence in $[-N, N]^n$. Then each coordinate of the sequence is Cauchy and hence converges (the Bolzano-Weierstrass theorem, Theorem 10.39). Let $\vec{x}$ be the point whose ith coordinate is the limit of the ith coordinates of the $\vec{x}_k$'s; then $\vec{x}_k \to \vec{x}$, verifying completeness. We conclude that $[-N, N]^n$ is compact. □

Proposition 10.47. *Every locally compact metric space is homeomorphic to a boundedly compact metric space.*

Proof. Let X be a locally compact metric space and let U_i ($i \in \mathbb{N}$) be open classes each of whose closures is compact and whose union is X. Let $n_1 = 1$, and recursively for every k find $n_{k+1} > n_k$ such that the closure of

$U_1 \cup \cdots \cup U_{n_k}$ is contained in $U_1 \cup \cdots \cup U_{n_{k+1}}$. Such an n_{k+1} always exists by Corollary 10.38.

Now for each k use Urysohn's lemma for metric spaces (Proposition 10.22) to find a continuous function $f_k : X \to [0,1]$ which is constantly 0 on $U_1 \cup \cdots \cup U_{n_k}$ and constantly 1 on the complement of $U_1 \cup \cdots \cup U_{n_{k+1}}$. Then define $f : X \to [0,\infty)$ by $f(x) = \sum f_k(x)$. The sum exists because for each x, $f_k(x) \neq 0$ for only finitely many values of k. Define a new metric d' on X by the formula

$$d'(x,y) = \max(d(x,y), |f(x) - f(y)|).$$

I claim that the identity map is a homeomorphism from X with the metric d to X with the metric d'. It is clear that convergence with respect to d' implies convergence with respect to d; conversely, if $x_n \to x$ with respect to d then x belongs to some U_i and hence x_n belongs to U_i for all sufficiently large n. Then since f is continuous on U_i (it is a finite sum of continuous functions there), we get $f(x_n) \to f(x)$, and this shows that $x_n \to x$ with respect to d'. As this entails that X is separable with respect to d', the proof of homeomorphism is complete.

The proof is concluded by observing that every closed ball with respect to d' is compact. This is because every closed ball with respect to d' is contained in $U_1 \cup \cdots \cup U_{n_k}$ for some k (if $x \in U_1 \cup \cdots \cup U_{n_k}$ and $y \notin U_1 \cup \cdots \cup U_{n_{k+N}}$ then $f(y) \geq f(x) + N$), and hence is contained in a compact subspace of X. □

The unbounded existential quantifier (there exist a boundedly compact metric space and a homeomorphism) is legal because the constructions are explicitly described.

Our last topic related to metric compactness deals with function spaces. We want to realize the continuous functions from one metric space, X, to another, Y, as an indexed family. We do this, under some compactness assumptions, by characterizing exactly which functions from a given dense subset $\mathring{X}$ of X to Y extend continuously to X.

Let $f : X \to Y$ be a function between metric spaces. We say that f is *uniformly continuous* if for every $\epsilon > 0$ there exists $\delta > 0$ such that $d(f(x), f(y)) \leq \epsilon$ whenever $d(x,y) \leq \delta$. This differs from ordinary continuity in that here δ is independent of x.

Theorem 10.48. *Let X and Y be separable metric spaces and suppose X is compact and Y is complete. Then every continuous function $f : X \to Y$ is uniformly continuous. If $\mathring{X}$ is a dense subset of X, then every continuous function from X to Y restricts to a uniformly continuous function*

from $\mathring{X}$ to Y, and every uniformly continuous function from $\mathring{X}$ to Y is the restriction of exactly one continuous function from X to Y.

Proof. Let $f : X \to Y$ be continuous and let $\mathring{X}$ be a dense subset of X. Then the restriction of f to $\mathring{X}$ definitely either is or is not uniformly continuous (LPO). Suppose it is not; then there exist $\epsilon > 0$ and a pair of sequences (x_n) and (y_n) in A such that $d(x_n, y_n) \to 0$ and $d(f(x_n), f(y_n)) \geq \epsilon$ for all n. By passing to subsequences we can assume that (x_n) and (y_n) converge in X, and they clearly converge to the same point, but $(f(x_n))$ and $(f(y_n))$ clearly cannot converge to the same point, so we have contradicted continuity of f. This shows that $f|_{\mathring{X}}$ is uniformly continuous.

But we must show that f itself is uniformly continuous. For any $\epsilon > 0$, find $\delta > 0$ such that for any $x, y \in \mathring{X}$, $d(x, y) < \delta$ implies $d(f(x), f(y)) < \epsilon$. Now for any $x, y \in X$ with $d(x, y) < \delta$ we can find sequences (x_n) and (y_n) in $\mathring{X}$ that converge to x and y, respectively, and such that $d(x_n, y_n) < \delta$ for all n. Uniform continuity on $\mathring{X}$ entails that $d(f(x_n), f(y_n)) < \epsilon$ for all n, and taking limits, this shows that $d(f(x), f(y)) \leq \epsilon$. That suffices to verify uniform continuity of f. We have shown that every continuous function from X to Y is uniformly continuous.

We saw above that the restriction of any continuous function to a dense subset is uniformly continuous. Conversely, let $\mathring{f} : \mathring{X} \to Y$ be uniformly continuous. For each $x \in X$, let (x_n) be a sequence in $\mathring{X}$ that converges to x and define $f(x) = \lim \mathring{f}(x_n)$. (This is where we use completeness of Y.) There are two things to check: that this limit exists, and that all choices of (x_n) yield the same value, but these are both straightforward. The argument made just above about a continuous function on X being uniformly continuous if its restriction to $\mathring{X}$ is uniformly continuous applies here too, and shows that f is uniformly continuous. It is clear that f is the unique continuous extension of $\mathring{f}$. □

For X compact and Y separable and complete, any function from $\mathring{X}$ to Y definitely is or is not uniformly continuous and hence definitely does or does not extend to a continuous function from X to Y.

The compactness hypothesis in Theorem 10.48 can be relaxed to local compactness.

Corollary 10.49. *Let X and Y be separable metric spaces and assume X is locally compact and Y is complete. If $\mathring{X}$ is a dense subset of X and U_n ($n \in \mathbb{N}$) are open classes whose union is X and each of whose closures is compact, then every continuous function from X to Y restricts to a*

uniformly continuous function on each $\mathring{X} \cap U_n$, and every function from $\mathring{X}$ to Y that is uniformly continuous on each $\mathring{X} \cap U_n$ is the restriction of exactly one continuous function from X to Y.

Proof. If $f : X \to Y$ is continuous then its restriction to each $\overline{U}_n$ is continuous, and therefore uniformly continuous by Theorem 10.48. Uniform continuity of the restrictions to the sets $\mathring{X} \cap U_n$ follows trivially. Conversely, given a function $\mathring{f} : \mathring{X} \to Y$ that is uniformly continuous on each $\mathring{X} \cap U_n$, we know from Theorem 10.48 how to extend it to a continuous function on each U_n, and these extensions agree on any intersection $U_n \cap U_m$ by the uniqueness part of Theorem 10.48. Any function which is continuous on each U_n is continuous on X, because any convergent sequence in X eventually lies in some U_n, and continuity on U_n then entails that the image of the sequence converges to the right limit in Y. □

By Proposition 10.47, any locally compact metric space is homeomorphic to a complete metric space; it follows that for X and Y locally compact, any function from $\mathring{X}$ to Y definitely does or does not extend to a continuous function from X to Y. I now define $C(X, Y)$ as follows. First, let Z be the subclass of $X^\omega \times Y^\omega$ consisting of all pairs of sequences $((x_n), (y_n))$ such that the x_n are pairwise distinct and collectively dense in X, and the function that takes x_n to y_n for all n extends to a continuous function from X to Y. We have just seen that this is a definite condition and so does define a subclass of $X^\omega \times Y^\omega$. Now put an equivalence relation $\sim$ on Z by making two pairs equivalent if they have the same continuous extension. For any pair of sequences $((x_n), (y_n))$ and $((x'_n), (y'_n))$, it is enough to check that the continuous extension of the first takes the value y'_n on x'_n for each n, and vice versa. So this is also a definite condition. We put $C(X, Y) = Z/\!\sim$. Then there is a natural family of subclasses of $X \times Y$ indexed by $C(X, Y)$ such that each slice is a continuous function from X to Y and each continuous function from X to Y appears exactly once as a slice. In other words, $C(X, Y)$ is an exact pseudoclass of continuous functions from X to Y. We can informally talk about continuous functions "in" $C(X, Y)$ and write "$f \,\varepsilon\, C(X, Y)$" for an arbitrary continuous function $f : X \to Y$.

I pointed out after Definition 10.58 that saying two metric spaces are homeomorphic involves an unbounded existential quantifier (there exists a homeomorphism) and any affirmation that two spaces are homeomorphic requires that we possess an explicit construction. However, when we are working with locally compact spaces the existential quantifier can be taken to range over the pseudoclass $C(X, Y)$, meaning that it is bounded and can

be used freely.

We can even define a pseudoclass of all homeomorphisms between two locally compact metric spaces.

Corollary 10.50. *Let X and Y be locally compact metric spaces and let $\mathring{X}$ be a dense subset of X. Then every function from $\mathring{X}$ into Y definitely does or does not extend to a homeomorphism between X and Y.*

Proof. By Proposition 10.47 we can assume that X and Y are both boundedly compact, and hence complete. Let $\mathring{f} : \mathring{X} \to Y$ be a function and let U_i ($i \in \mathbb{N}$) be open classes whose union is X and each of whose closures is compact. The condition "for each i, the restriction of $\mathring{f}$ to $\mathring{X} \cap U_i$ is uniformly continuous" is definite, so by Corollary 10.49 $\mathring{f}$ definitely does or does not extend to a continuous function f from X into Y.

It is also definitely either the case or not the case that $\mathring{f}$ is injective. If not, then f is not a homeomorphism. Also, $\mathring{f}(\mathring{X})$ definitely either is or is not dense in Y — we just have to check whether every element of a dense subset of Y is approximated arbitrarily well by elements of $\mathring{f}(\mathring{X})$. If not, then again f is not a homeomorphism between X and Y. If $\mathring{f}$ is injective and $\mathring{f}(\mathring{X})$ is dense in Y, then $\mathring{f}^{-1}$ is a function from a dense subset $\mathring{f}(\mathring{X})$ of Y into X, and by the first part of this proof it definitely either does or does not extend to a continuous function g from Y into X. If it does not, then f is not a homeomorphism, and if it does, then $g \circ f$ is a continuous function from X into itself which fixes every element of a dense subset, so it must be the identity map, and $f \circ g$ is the identity map on Y by the same argument. So in this case f is a homeomorphism between X and Y. □

If X is a locally compact metric space, then we define a pseudoclass Homeo(X) in much the same way that we defined $C(X, Y)$, such that it exactly codes the self-homeomorphisms of X.

10.4 Topological spaces

Definition 10.51. A *topology* on a class X is an indexed family of subclasses $\mathcal{T} \subseteq X \times I$, for some index class I, such that

(a) $\emptyset$ and X are both slices of $\mathcal{T}$
(b) if $U_1, U_2, \ldots$ are all slices of $\mathcal{T}$ then so is $\bigcup U_i$
(c) if $U_1, \ldots, U_n$ are all slices of $\mathcal{T}$ then so is $U_1 \cap \cdots \cap U_n$.

The slices of a topology on X are the *open* classes in X and their complements are the *closed* classes.

When a topology on X has been specified, we call X a *topological space*. Thus "let X be a topological space" abbreviates "let X be a class and let $\mathcal{T}$ be a topology on X". We consider two topologies to be equivalent if they have the same open classes, even if their indexing classes differ.

Definition 10.52. Let $\mathcal{T} \subseteq X \times I$ be a topology on a class X. Any subclass X_0 of X is a *subspace* of X when equipped with the *subspace topology* $\mathcal{T} \cap (X_0 \times I)$.

Thus, the open classes for the subspace topology on X_0 are precisely the classes of the form $U \cap X_0$ for $U \subseteq X$ open (cf. Corollary 10.13).

For the sake of clarity, we can say that a class $U_0 \subseteq X_0$ is *relatively open* if it is open for the subspace topology. Thus, $U_0 \subseteq X_0$ is relatively open in X_0 if and only if $U_0 = U \cap X_0$ for some open $U \subseteq X$. Likewise, $C_0 \subseteq X_0$ is *relatively closed* if $C_0 = C \cap X_0$ for some closed $C \subseteq X$.

We say that a sequence (x_n) *converges* to a point x if every open class containing x also contains x_n for all but finitely many n. We write "$x_n \to x$". The notation "$\lim x_n = x$" is not so good if a sequence might have more than one limit; a little later we will present a condition that ensures this cannot happen. See Definition 10.65 and the comment that follows it. We say that $C \subseteq X$ is *sequentially closed* if whenever (x_n) is a sequence in C and $x_n \to x$, we also have $x \in C$.

A subclass Y of X is *dense* if every element of X is a limit of a sequence in Y, and X is *separable* if there exists a dense subset. If every closed subclass of X is separable, then X is *hereditarily separable*. X is *second countable* if there is a sequence of open classes U_i such that every open class is a union of some of the U_i's. We say that the U_i's are *basic* open classes and constitute a *basis* for the topology.

Separability is a powerful condition because to a significant extent it allows us to work with sets in place of classes. For example, every set definitely is or is not inhabited, and it follows that every separable class definitely is or is not inhabited. But note that there is no obvious implication from second countability to separability, because the obvious argument — given a basis U_1, U_2, ..., select one element from each inhabited U_n — requires that each U_n definitely be empty or inhabited. To be a little more concrete, fix a total ordering $\preceq$ on $\mathbb{N}$ and let $X \subseteq \mathcal{P}(\mathbb{N})$ be the class of (codes of) subsets of $\mathbb{N}$ with no $\preceq$-minimal element. If we don't know

whether $\preceq$ is a well-ordering (cf. Chapter 8) then we don't know whether X is inhabited, but X is a legitimate subclass of $\mathcal{P}(\mathbb{N})$ since every subset of $\mathbb{N}$ definitely does or does not have a $\preceq$-minimal element, by the LPO. We have a second countable topology on X whose open classes are just $\emptyset$ and X, but we cannot affirm separability if we cannot affirm that X definitely is or is not inhabited. Having said all that, if we know that the basic open classes U_i of some second countable space are all inhabited, then the simple argument mentioned above does show that the space is separable.

Now here are some easy facts about separable, hereditarily separable, and second countable spaces.

Proposition 10.53. *Let X be a second countable space. Then for any subset $A \subseteq X$ there is a minimal closed class $C \subseteq X$ containing A (i.e., if C' is any other closed class containing A then $C \subseteq C'$). An element of X belongs to C if and only if it is the limit of a sequence of elements of A.*

Proof. Let U_i ($i \in \mathbb{N}$) be a basis for the topology and let $A = \{x_1, x_2, \ldots\}$ be a subset of X. For each i and j it definitely either is or is not the case that $x_i \in U_j$. Then by the LPO, for each j it definitely either is or is not the case that some x_i belongs to U_j, i.e., that A meets U_j. The union of the basic open classes which do not meet A is then an open class whose complement is the desired minimal closed class C containing A.

Any limit of a sequence of elements of A has the property that every open class that contains it meets A. Thus no basic open class containing such a limit contributes to the complement of C, which means that the limit must belong to C. Conversely, let $x \in C$; we must show that it is the limit of a sequence of elements of A. We know that every open class containing x must meet A, as every basic open class containing x definitely does or does not meet A, and if any of them did not then x could not be an element of C. Let (U_{n_k}) be the sequence of basic open classes that contain x. For each k the class $U_{n_1} \cap \cdots \cap U_{n_k}$ is open and contains x, so by the previous comment it must contain an element of A; let x_k be such an element. Then (x_k) is a sequence of elements of A with the property that every basic open class containing x must contain x_k for all but finitely many k. Thus $x_k \to x$. □

We call C the *closure* of A and denote it $\overline{A}$. More generally, for any separable subclass U of X, we define the closure $\overline{U}$ of U to be the closure of a dense subset of U. This will be the smallest closed class containing U, and it does not depend on the choice of dense subset.

We can generalize the second part of Proposition 10.16.

Proposition 10.54. *Let X be a second countable space. Then any closed subspace of X is sequentially closed, and any separable sequentially closed subspace of X is closed.*

Proof. Suppose $C \subseteq X$ is closed and let (x_n) be a sequence of elements of C which converges to a point x. Then x definitely does or does not belong to C, but if it did not then C^c would be an open class that contains x but not any x_n, contradicting the assumption of convergence. Thus x must belong to C. This proves the first part of the proposition.

Conversely, suppose C is separable and sequentially closed. Let $\mathring{C}$ be a dense subset of C and let C' be the closure of $\mathring{C}$; we will show that $C = C'$. The containment $C' \subseteq C$ is easy, because every element of C' is a limit of a sequence of elements of $\mathring{C}$. The reverse inclusion $C \subseteq C'$ is even easier, an immediate consequence of the definition of density. We conclude that $C = C'$. □

The next result generalizes Corollary 10.14; its proof is more or less identical.

Proposition 10.55. *Let X be a separable space. Then any open class in X is separable and is definitely either empty or inhabited, and any two open classes definitely do or do not meet.*

Proposition 10.56. *Let X be a hereditarily separable space. Then any closed class in X is definitely either empty or inhabited, any two closed classes definitely do or do not meet, and any closed class definitely is or is not contained in any other closed class.*

Proof. Hereditary separability says that every closed class is separable. This implies that every closed class is definitely either empty or inhabited; consequently, since the intersection of any two closed classes is closed, any two closed classes definitely do or do not meet. Also, one closed class is contained in another if and only if a given dense subset of the first is contained in the second, and that condition is definite by the LPO. □

If X is a class then any sequence of subclasses V_i whose union is X generates a second countable topology $\mathcal{T}$ on X as follows. First, enumerate the finite intersections of the V_i's as U_j ($j \in \mathbb{N}$), with repetitions allowed. That is, every intersection of finitely many V_i's equals U_j for some j, and every U_j is such an intersection. Then for any $S \,\varepsilon\, \mathcal{P}(\mathbb{N})$ let the S-slice of

$\mathcal{T} \subseteq X \times \mathcal{P}(\mathbb{N})$ be the subclass $\bigcup_{n\in S} U_n$ of X. This is a topology and its open classes are precisely the classes obtained by taking countable unions of finite intersections of the V_i's. The V_i's constitute a *subbasis* for $\mathcal{T}$ and the finite intersections of the V_i's constitute a basis.

Example 10.57. As I mentioned just after Definition 10.9, we can index the open subclasses of any separable metric space X by $\mathcal{P}(X \times (0, \infty))$, and by Proposition 10.12 this is a topology. It is called the *metric topology* on X, and it is second countable by Lemma 10.11.

The topologies arising from separable metrics as in Example 10.57 are the most important ones, though we will make occasional contact with some others. If a topology arises from a metric in this way, we say it is *separably metrizable*. However, even if one is only interested in separably metrizable spaces, passing to the topological setting can still have clarifying value when the details of the metric don't matter. The rest of this section will illustrate this, particularly the material on products.

Continuity is, of course, of central importance. Here is the definition, followed by two quick propositions.

Definition 10.58. A function $f : X \to Y$ between topological spaces is *continuous* if the inverse image $f^{-1}(V)$ of any open class V in Y is open in X, or equivalently, the inverse image of any closed class in Y is closed in X. It is *sequentially continuous* if $x_n \to x$ in X implies $f(x_n) \to f(x)$ in Y. A *homeomorphism* is a continuous bijection whose inverse is also continuous, and two spaces are *homeomorphic* if there exists a homeomorphism between them.

Proposition 10.59. *(Gluing lemma) Let X and Y be topological spaces and let $C, D \subseteq X$ be closed classes with $X = C \cup D$. Then a function $f : X \to Y$ is continuous if and only if its restrictions to C and D are continuous.*

Proof. Suppose f is continuous. Then the inverse image $f^{-1}(B)$ of any closed class B in Y is closed in X. But $f|_C^{-1}(B) = f^{-1}(B) \cap C$ and $f|_D^{-1}(B) = f^{-1}(B) \cap D$, so the inverse images of any closed class in Y under $f|_C$ and $f|_D$ are relatively closed in C and D. This shows that $f|_C$ and $f|_D$ are continuous.

Conversely, suppose $f|_C$ and $f|_D$ are continuous and let $B \subseteq Y$ be closed. Then $f^{-1}(B) = f|_C^{-1}(B) \cup f|_D^{-1}(B)$ is the union of a relatively closed subclass C_0 of C with a relatively closed subclass D_0 of D. Here C_0

is the intersection with C of a closed class in X, but since C is closed that just means that C_0 is closed in X. D_0 is closed in X for the same reason, so $f^{-1}(B)$ is the union of two closed subclasses of X, and hence is closed. Thus f is continuous. □

Lemma 10.60. *Every continuous function between topological spaces is sequentially continuous.*

Proof. Let X and Y be topological spaces and let $f : X \to Y$ be a continuous function. Suppose $x_n \to x$ in X and let $U \subseteq Y$ be an open class that contains $f(x)$. Then $f^{-1}(U) \subseteq X$ is an open class that contains x, and therefore it contains x_n for all but finitely many n. Thus U contains $f(x_n)$ for all but finitely many n. We have shown that $f(x_n)$ converges to $f(x)$. □

Proposition 10.61. *Let X and Y be topological spaces and let $f : X \to Y$ be a surjective continuous function. If X is separable or hereditarily separable then so is Y.*

Proof. The lemma shows that if $\mathring{X}$ is a dense subset of X then $f(\mathring{X})$ will be a dense subset of Y. So if X is separable then so is Y.

If X is hereditarily separable then every closed subclass of X is separable, and every closed subclass of Y is the continuous image of its inverse image, which is closed in X. So the separability of every closed subclass of Y follows from the first part of the proposition. □

A subclass of a topological space is *clopen* if it is both open and closed. A topological space X is *connected* if every clopen subclass equals $\emptyset$ or X, and *disconnected* if there exists a clopen subclass such that both it and its complement are inhabited.

An *interval* in $\mathbb{R}$ is a class $X \subseteq \mathbb{R}$ with the property that if $x < y < z$ and $x, z \in X$, then $y \in X$ also.

Proposition 10.62. *The connected subclasses of $\mathbb{R}$ are precisely the intervals in $\mathbb{R}$.*

Proof. Let $X \subseteq \mathbb{R}$ be an interval and let $C \subseteq X$ be relatively clopen. We can assume X contains more than one point. (It definitely does or does not contain more than one point, because it contains more than one point if and only if it contains more than one rational number.) C is the intersection with X of an open class in $\mathbb{R}$, and, using the fact that X is an interval, intersecting a dense subset of that open class with X yields a dense subset

$\mathring{C}$ of C. As a set, $\mathring{C}$ is definitely either empty or inhabited. Also, $\mathring{C}$ is dense in X if and only if every rational interval (p, q) with $p, q \in \mathbb{Q} \cap X$ meets $\mathring{C}$, and this condition is definite as well. If $\mathring{C}$ is neither empty nor dense in X, then there exists a rational interval (p, q) in $\mathring{C}^c$ and there exists an element of $\mathring{C}$ which is either less than p or greater than q. In the former case $a = \sup\{x \in \mathring{C} : x \leq p\}$ belongs to C but is also the limit of a sequence of elements of C^c, contradicting clopenness of C. In the latter case we get a contradiction from $b = \inf\{x \in \mathring{C} : x \geq q\}$. We conclude that $\mathring{C}$ must be either empty or dense in X, and therefore C must be either empty or all of X. We have shown that X is connected.

Now let $Y \subseteq \mathbb{R}$ be connected and suppose $x, z \in Y$ and $x < y < z$. If $y \notin Y$ then $(-\infty, y) \cap Y = (-\infty, y] \cap Y$ is a relatively clopen subclass of Y which contains x but not z, contradicting connectedness of Y. Thus $y \in Y$, and we have shown that Y is an interval. □

In particular, $\mathbb{R}$ itself is connected.

A subclass K of $\mathbb{R}^n$ is *convex* if for every pair of elements of K the line segment joining them is contained in K.

Corollary 10.63. *Every hereditarily separable convex subclass of $\mathbb{R}^n$ is connected.*

Proof. Let $K \subseteq \mathbb{R}^n$ be hereditarily separable and convex and let $C \subseteq K$ be relatively clopen. Let $\mathring{C}$ be a dense subset of C. Then $\mathring{C}$ is definitely either empty or inhabited; I claim that it also definitely either is or is not dense in K. That is because every element x of a dense subset of K definitely does or does not satisfy $\inf\{d(x, y) : y \in \mathring{C}\} = 0$; if they all do, then $\mathring{C}$ is dense in K, otherwise it is not.

Now if $\mathring{C}$ is empty then so is C, and if $\mathring{C}$ is dense in K then $C = K$, so to show that K is connected we must derive a contradiction from the assumption that $\mathring{C}$ is inhabited but not dense in K. But in this case there exists $x \in \mathring{C}$ and $y \in K \setminus C$, and letting L be the line segment joining x and y, we get that $C \cap L$ is a nonempty relatively clopen subclass of L whose complement is also nonempty, contradicting Proposition 10.62. □

Proposition 10.64. *Let $f : X \to Y$ be a surjective continuous map between topological spaces. If X is connected then so is Y.*

Proof. Suppose X is connected and let $C \subseteq Y$ be clopen. Then $f^{-1}(C)$ is a clopen subclass of X and must therefore equal either $\emptyset$ or X. Thus C is either $\emptyset$ or Y, showing that Y is connected. □

Next, here are some "separation properties".

Definition 10.65. A topological space X is T_1 if every subclass containing exactly one point is closed. It is *Hausdorff* if whenever $x, y \in X$ are distinct, there exist disjoint open classes $U, V \subseteq X$ which contain x and y, respectively. It is *regular* if whenever $C \subseteq X$ is closed and $x \in C^c$, there exist disjoint open classes $U, V \subseteq X$ which contain x and C, respectively. It is *normal* if whenever $C, D \subseteq X$ are closed and disjoint, there exist disjoint open classes $U, V \subseteq X$ which contain C and D, respectively.

The Hausdorff property is the promised condition that ensures no sequence can converge to more than one point.

Proposition 10.66. *Let $f : X \to Y$ be a continuous function between second countable spaces. Also assume X is separable and Y is hereditarily separable and regular. Then f is continuous if and only if it is sequentially continuous.*

Proof. The forward implication was proven in Lemma 10.60. No assumptions about X and Y were needed.

Conversely, suppose f is sequentially continuous and let $V \subseteq Y$ be open. Let (U_i) and (V_j) be the basic open classes of X and Y, respectively, and let $\mathring{X}$ be a dense subset of X. The closure of each V_j exists by Propositions 10.53 and 10.55 and definitely is or is not contained in V by Proposition 10.56. ($\overline{V}_j$ and V^c definitely do or do not meet.) Let J be the set of natural numbers j such that $\overline{V}_j \subseteq V$.

For each i, it definitely is or is not the case that $f(\mathring{X} \cap U_i)$ is contained in V_j for some $j \in J$; let I be the set of natural numbers i for which this is the case. I claim that $f^{-1}(V) = \bigcup_{i \in I} U_i$. First, for each $i \in I$ we have $f(U_i) \subseteq \overline{V}_j \subseteq V$ for some $j \in J$, since every element of U_i is the limit of a sequence of elements of $\mathring{X} \cap U_i$ all of which map into V_j. This shows that $\bigcup_{i \in I} U_i \subseteq f^{-1}(V)$. Conversely, for any $x \in f^{-1}(V)$ we have $f(x) \in V_j$ for some $j \in J$ by regularity. We will show that there exists i such that $x \in U_i$ and $f(\mathring{X} \cap U_i) \subseteq V_j$. This definitely is or is not the case, and if it is not then for any basic open class U_i that contains x, and hence any open class U that contains x, the set $\mathring{X} \cap U$ does not map entirely into V_j. Letting U_{i_k} ($k \in \mathbb{N}$) be the basic open classes that contain x, for each k we can then find $x_k \in \mathring{X} \cap U_{i_1} \cap \cdots \cap U_{i_k}$ with $f(x_k) \notin V_j$. But then $x_k \to x$ and therefore $f(x) \notin V_j$, a contradiction. We conclude that x belongs to U_i for some $i \in I$, as desired. So $f^{-1}(V) = \bigcup_{i \in I} U_i$, showing that $f^{-1}(V)$ is open. □

Proposition 10.67. *Every second countable, hereditarily separable, regular space is normal.*

Proof. Suppose X is second countable, hereditarily separable, and regular, and let $C, D \subseteq X$ be disjoint closed classes.

By hereditary separability we can enumerate the basic open classes whose closures do not meet D as (U_n) (Proposition 10.56). Similarly, enumerate the basic open classes whose closures do not meet C as (V_n).

Now define $U'_n = U_n \setminus (\overline{V}_1 \cup \cdots \cup \overline{V}_n)$ and $V'_n = V_n \setminus (\overline{U}_1 \cup \cdots \cup \overline{U}_n)$ for all $n \in \mathbb{N}$, and let $U = \bigcup U'_n$ and $V = \bigcup V'_n$.

U and V are both countable unions of open classes, so both are open. Suppose there exists $x \in U \cap V$. Then $x \in U'_m \cap V'_n$ for some $m, n \in \mathbb{N}$; without loss of generality suppose $m \geq n$. We then have $x \in U_m \setminus (\overline{V}_1 \cup \cdots \cup \overline{V}_m)$, showing that $x \notin V_n$, which contradicts $x \in V'_n$. We conclude that there does not exist $x \in U \cap V$, i.e., U does not meet V.

Finally, to see that $C \subseteq U$ and $D \subseteq V$, let $x \in C$. By regularity there exists an open class which contains x and whose closure does not meet D. Then there must exist some basic open class with the same properties, which means that x belongs to some U_n. But also, since each $\overline{V}_i$ is disjoint from C, we have $x \notin \overline{V}_1 \cup \cdots \cup \overline{V}_n$, and this shows that $x \in U'_n$. Thus every element of C belongs to U, and every element of D belongs to V by similar reasoning. □

Corollary 10.68. *Every hereditarily separable metric space is normal.*

(This could also be inferred from Urysohn's lemma for metric spaces, Proposition 10.22.)

Theorem 10.69. *(Urysohn's lemma) Let C and D be disjoint closed classes in a normal topological space X. Then there is a continuous function $f : X \to [0,1]$ with $C \subseteq f^{-1}(0)$ and $D \subseteq f^{-1}(1)$.*

Proof. Let $K_0 = C$ and $U_1 = D^c$. By normality there exist open and closed classes $U_{1/2}$ and $K_{1/2}$ such that $K_0 \subseteq U_{1/2} \subseteq K_{1/2} \subseteq U_1$. Then there exist open and closed $U_{1/4}$ and $K_{1/4}$ with $K_0 \subseteq U_{1/4} \subseteq K_{1/4} \subseteq U_{1/2}$, and open and closed $U_{3/4}$ and $K_{3/4}$ with $K_{1/2} \subseteq U_{3/4} \subseteq K_{3/4} \subseteq U_1$. Continuing in this way, we get open and closed classes U_r and K_r for every dyadic rational r in $(0,1)$, satisfying $K_{k/2^n} \subseteq U_{(2k+1)/2^{n+1}} \subseteq K_{(2k+1)/2^{n+1}} \subseteq U_{(k+1)/2^n}$ for all $n \in \mathbb{N}$ and $0 \leq k < 2^n$. Finally, define $f : X \to [0,1]$ by letting $f(x) = \inf\{r : x \in U_r\}$, taking $f(x) = 1$ if x does not belong to U_1.

We clearly have $C \subseteq f^{-1}(0)$ and $D \subseteq f^{-1}(1)$. For continuity, it will suffice to show that the inverse images of intervals of the form $[0, a)$ and $(b, 1]$ are open. This is true because $f^{-1}([0, a)) = \bigcup_{r<a} U_r$ and $f^{-1}((b, 1]) = \bigcup_{r>b} K_r^c$. □

In this proof we use dependent choice to select open and closed classes U_r and K_r for all dyadic rationals r in $(0, 1)$; this can be done using bounded quantification over the class that indexes the topology on X. That is, we "quantify over all open (closed) subclasses" by quantifying over the index class, a technique that we have already used copiously. Then we use the U_r and K_r to given an explicit construction which justifies the unbounded existence statement "there is a continuous function ..."

However, in the proof of the next theorem we need to boundedly quantify over all continuous functions from X into $[-a, a]$, and to do this we need a pseudoclass of all such functions. Near the end of Section 10.3 we saw how to define a pseudoclass $C(X, Y)$ of all continuous functions from X into Y when X and Y are locally compact, and this will generalize to compactly coherent X and Y in the next section. But for the specific case of bounded real-valued functions we can use the proof of Urysohn's lemma to weaken the hypothesis on X to hereditary separability. To do this, let A be the set of dyadic rationals in $(0, 1)$. Then for any class X with hereditarily separable topology $\mathcal{T} \subseteq X \times I$, consider the pseudoclass of all families (U_r) and (K_r) of open and closed subclasses of X indexed by A — this can be modelled as a subclass of $\mathrm{fn}(A, I)^2$ — which satisfy $U_r \subseteq K_r$ for all r and $K_r \subseteq U_s$ for all $r < s$. Then factor out the equivalence relation that makes (U_r) and (K_r) equivalent to (U'_r) and (K'_r) if $\bigcup_{r<a} U_r = \bigcup_{r<a} U'_r$ and $\bigcap_{r>b} K_r = \bigcap_{r>b} K'_r$ for all $a, b \in A$. (This is where we need X to be hereditarily separable, in order to apply Proposition 10.56.) The result is a pseudoclass which exactly indexes the continuous functions from X into $[0, 1]$. Thus, assuming only that X is hereditarily separable, we can boundedly quantify over all continuous functions from X to $[0, 1]$, or, with easy modifications, continuous functions from X to $[-a, a]$ or bounded continuous functions from X to $\mathbb{R}$.

Lemma 10.70. *Let K be a closed subclass of a normal hereditarily separable topological space X and let $f : K \to [-a, a]$ be continuous. Then there is a continuous function $g : X \to \left[-\frac{a}{3}, \frac{a}{3}\right]$ satisfying $|f(x) - g(x)| \leq \frac{2}{3}a$ for all $x \in K$.*

Proof. Let $C = f^{-1}\left[-a, -\frac{a}{3}\right]$ and $D = f^{-1}\left[\frac{a}{3}, a\right]$. These are disjoint

closed subclasses of K and hence of X. By Urysohn's lemma (Theorem 10.69) there is a continuous function $g : X \to \left[-\frac{a}{3}, \frac{a}{3}\right]$ with $C \subseteq g^{-1}(-\frac{a}{3})$ and $D \subseteq g^{-1}(\frac{a}{3})$. For any $x \in C$ we then have $-a \leq f(x) \leq -\frac{a}{3} = g(x)$, for any $x \in D$ we have $g(x) = \frac{a}{3} \leq f(x) \leq a$, and for any $x \in K \setminus (C \cup D)$ we have $f(x), g(x) \in \left[-\frac{a}{3}, \frac{a}{3}\right]$. In all cases $|f(x) - g(x)| \leq \frac{2}{3}a$. □

Theorem 10.71. *(Tietze extension) Let K be a closed subclass of a normal hereditarily separable topological space X and let $f : K \to \mathbb{R}$ be bounded and continuous. Then there is a bounded, continuous extension $\widetilde{f} : X \to \mathbb{R}$ of f.*

Proof. Say $f : K \to [-a, a]$. Let $g_1 : X \to \left[-\frac{a}{3}, \frac{a}{3}\right]$ be the continuous function given by Lemma 10.70 satisfying $|f(x) - g_1(x)| \leq \frac{2}{3}a$ for all $x \in K$. Thus $f - g_1$ is a continuous function from K into $\left[-\frac{2}{3}a, \frac{2}{3}a\right]$.

Now let $g_2 : X \to \left[-\frac{2}{9}a, \frac{2}{9}a\right]$ be the continuous function given by Lemma 10.70 satisfying $|(f(x) - g_1(x)) - g_2(x)| \leq \frac{4}{9}a$ for all $x \in K$; continuing in this way, for each n we get a continuous function $g_n : X \to \left[-\frac{2^{n-1}}{3^n}a, \frac{2^{n-1}}{3^n}a\right]$ such that $|f(x) - (g_1(x) + \cdots + g_n(x))| \leq \left(\frac{2}{3}\right)^n a$ for all $x \in K$. (This construction recursively describes a sequence of functions as the slices of a family of subclasses of $X \times \mathbb{R}$ indexed by $\mathbb{N}$.)

Define $\widetilde{f}(x) = g_1(x) + g_2(x) + \cdots$. This series converges absolutely for all x because its nth term is at most $\frac{a}{2} \cdot (\frac{2}{3})^n$ in absolute value, and it agrees with f everywhere on K since the partial sums converge to f at each point of K. Finally, it is continuous by Corollary 10.31. □

The following development illustrates the value of working with topologies rather than metrics. I mentioned at the beginning of this chapter that there is more than one natural metric on $\mathbb{R}^n$, and the same is true of products of metric spaces generally. However, there is only one really natural topology on products of separable metric spaces.

Definition 10.72. Let X_i ($1 \leq i \leq n$ or $i \in \mathbb{N}$) be second countable spaces. The *product topology* on $\prod X_i$ is generated from the subbasis consisting of the classes $\pi_i^{-1}(U)$ with $1 \leq i \leq n$ or $i \in \mathbb{N}$ and U a basic open class in X_i. Here π_i is the projection onto the ith coordinate.

This definition only works for second countable spaces, but it has the benefit that we can describe a topology for which the class indexing the open subclasses of the product is fairly simple. If one does not want to assume second countability, a more complicated construction can generate an index

class that indexes all countable unions of finite intersections of classes of the form $\pi_i^{-1}(U)$ with $U \subseteq X_i$ open. We won't need this, however.

Product topologies have a universal property.

Theorem 10.73. *(Universal property of product topologies) Let X be a topological space and let X_i ($1 \leq i \leq n$ or $i \in \mathbb{N}$) be second countable topological spaces. For any continuous functions $f_i : X \to X_i$ there is a unique continuous function $f : X \to \prod X_i$ satisfying $f_i = \pi_i \circ f$ for all $x \in X$.*

Proof. The function f satisfying the desired equations exists and is unique by the universal property of products (Theorem 9.22). It is continuous because the inverse image $f^{-1}(\pi_i^{-1}(U))$ under f of any subbasic open class equals $f_i^{-1}(U)$ and is therefore open by continuity of f_i. Taking finite intersections and countable unions, we get that the inverse image of any open class in $\prod X_i$ is open in X (Proposition 9.18). □

Since sequential convergence is so important in metric spaces, let us note that it has a very simple characterization in product spaces.

Proposition 10.74. *Let X_i ($1 \leq i \leq n$ or $i \in \mathbb{N}$) be second countable topological spaces. Then a sequence $(\vec{x}_n)$ in $\prod X_i$ converges to an element $\vec{x} \in \prod X_i$ if and only if it converges to $\vec{x}$ on each coordinate.*

Proof. If $(\vec{x}_n)$ converges to $\vec{x}$ on each coordinate, then for any i and any open class $U \subseteq X_i$ containing x_i, all but finitely many of the $\vec{x}_n$'s belong to $\pi_i^{-1}(U)$, which shows that $\vec{x}_n \to \vec{x}$, as it suffices to check that the sequence eventually belongs to every subbasic open class containing $\vec{x}$. Conversely, if $\vec{x}_n \to \vec{x}$ in $\prod X_i$ then all but finitely many terms belong to any open class containing $\vec{x}$, in particular any open class of the form $\pi_i^{-1}(U)$ with i arbitrary, $U \subseteq X_i$ open and $x_i \in U$. And this shows convergence on each coordinate. □

The following result is also quite useful.

Proposition 10.75. *Any product of separably metrizable topological spaces is separably metrizable.*

Proof. The proof is easy for finite products. Let $X_1, \ldots, X_n$ be separable metric spaces with dense subsets $\mathring{X}_1, \ldots, \mathring{X}_n$. Then the open balls of radius r in $\prod X_i$ for the l^∞ metric

$$d(\vec{x}, \vec{y}) = \max(d(x_1, y_1), \ldots, d(x_n, y_n))$$

are precisely the classes of the form $\prod U_i$ where each U_i is an open ball of radius r in X_i. From this it is easy to check that the product topology on $\prod X_i$ is the metric topology for the l^∞ metric, and that $\prod \mathring{X}_i$ is a dense subset.

Now let X_n ($n \in \mathbb{N}$) be separable metric spaces with dense subsets $\mathring{X}_n$. By replacing the metric d on X_n with $\min(d(x,y), \frac{1}{n})$, we can assume that all distances in X_n are at most $\frac{1}{n}$. Note that this leaves intact the balls of radius less than $\frac{1}{n}$, and as every open class with respect to either the original metric or the new metric is a countable union of such balls, the two metrics give rise to the same topology. We will show that the metric topology on $\prod X_i$ for the l^∞ metric $d(\vec{x}, \vec{y}) = \sup d(x_i, y_i)$ is separable and agrees with the product topology.

Fix $y_n \in \mathring{X}_n$ ($n \in \mathbb{N}$). Let B be the set of all the sequences $\vec{z} \in \prod \mathring{X}_n$ for which $z_n = y_n$ for all but finitely many n. This is dense for the metric topology because for any $\vec{x} \in \prod X_n$ and any $r > 0$ we can find $m > \frac{1}{r}$ and $z_1 \in \mathring{X}_1$, …, $z_m \in \mathring{X}_m$ with $z_i \in B_r(x_i)$ for $1 \leq i < m$; since every element of X_i belongs to $B_r(y_i)$ for $i \geq m$, setting $z_i = y_i$ for $i \geq m$ yields an element $\vec{z}$ of $B \cap B_r(\vec{x})$. Thus B is dense for the metric topology, so the metric topology is separable.

Let $k \in \mathbb{N}$, let $r > 0$ be rational, and let $x \in \mathring{X}_k$. Let $n > \max(\frac{1}{r}, k)$. For each choice of $\vec{x} \in \prod \mathring{X}_i$ with $x_k = x$ and $x_m = y_m$ for all $m \geq n$ we have a ball $B_r(\vec{x})$ in $\prod X_i$; there are countably many such $\vec{x}$ and therefore countably many such balls, and their union is the class $\pi_k^{-1}(B_r(x))$. Thus every class of this form is metrically open, and taking countable unions we get that $\pi_k^{-1}(U)$ is metrically open for every open $U \subseteq X_k$. In other words, every subbasic open class for the product topology is metrically open. Then since finite intersections and countable unions of metrically open classes are metrically open (Proposition 10.12), we infer that every class that is open for the product topology is metrically open.

Conversely, we must show that every open ball is open for the product topology. Since every metrically open class is a countable union of balls, this will yield that every metrically open class is open for the product topology, completing the proof that the two topologies are equal. So let $\vec{x} \in \prod X_i$ and let $r > 0$. Let $n > \frac{1}{r}$. Then $B_r(\vec{x}) = \pi_1^{-1}(B_r(x_1)) \cap \cdots \cap \pi_n^{-1}(B_r(x_n))$, showing that $B_r(\vec{x})$ is open for the product topology, as desired. □

Products of compact metric spaces are especially nice. In Section 10.3 I included separability in the definition of compactness for metric spaces; in the topological setting I am going to upgrade that to separable metrizability.

Thus a topological space is *compact* if it is the metric topology for some compact metric space. We will not have any need for topological spaces which are compact in any wider sense. I will also say that a topological space is *locally compact* if it is the metric topology for some locally compact metric space. But I will not require metrizability of σ-compact topological spaces; let us agree that a topological space is *σ-compact* if it is the union of countably many compact subspaces (each of which must itself be separably metrizable, however).

The basic nice features of σ-compactness carry over from the metric setting nonetheless.

Proposition 10.76. *Every σ-compact space is hereditarily separable. If $f : X \to Y$ is a continuous function from a σ-compact space into a second countable space, then the image $f(X)$ exists as a subclass of Y.*

Proof. Let C and K be, respectively, closed and compact subspaces of some topological space. Then $C \cap K$ is relatively closed in K and therefore separable by Corollary 10.42. It easily follows that in any σ-compact space, every closed class is separable. The second part of the proposition differs from the metric setting (Theorem 10.44) only in that Y is not assumed to be metrizable; its proof is identical, now using Proposition 10.53 to get the existence of closures in Y. □

Theorem 10.77. *(Tychonoff product theorem) Let X_i ($1 \leq i \leq n$ or $i \in \mathbb{N}$) be compact spaces. Then $\prod X_i$ is compact for the product topology.*

Proof. We know from Proposition 10.75 that $\prod X_i$ is separably metrizable. So by Theorem 10.35 it will suffice to show that every sequence has a convergent subsequence.

Let $(\vec{x}_n)$ be a sequence in $\prod X_i$. If the product is finite then we can use compactness of X_1 to extract a subsequence of $(\vec{x}_n)$ whose first coordinates converge in X_1, then use compactness of X_2 to extract a subsequence of the first subsequence whose second coordinates converge in X_2, and so on. This process terminates after finitely many steps, and generates a subsequence of $(\vec{x}_n)$ that converges on every coordinate, and hence in $\prod X_i$.

In the case of infinite products, we start by extracting a subsequence of $(\vec{x}_n)$ whose first coordinates converge in X_1. Let $\vec{x}_{n_1}$ be the first term of that subsequence. Then extract a further subsequence whose second coordinates converge in X_2 and let $\vec{x}_{n_2}$ be the first term of that sequence with $n_2 > n_1$. Note that after this second extraction the sequence converges on both the first and second coordinates. Continuing in this way, we ultimately

get a subsequence $(\vec{x}_{n_k})$ of the original sequence which converges on every coordinate, and hence converges in $\prod X_i$. □

Corollary 10.78. *Let X_1, ..., X_n be locally compact spaces. Then $\prod X_i$ is locally compact for the product topology.*

Proof. Let U_1^1, U_2^1, ... be a basis for the topology on X_1 such that each $\overline{U}_k^1$ is compact. Then let U_1^2, U_2^2, ... be a basis for the topology on X_2 with the same property, and so on. Now for any $k_1, \ldots, k_n \in \mathbb{N}$ the class $U_{k_1}^1 \times \cdots \times U_{k_n}^n$ is open in $\prod X_i$, and its closure equals $\overline{U}_{k_1}^1 \times \cdots \times \overline{U}_{k_n}^n$, which is compact by Tychonoff's theorem. As the union of all such classes $U_{k_1}^1 \times \cdots \times U_{k_n}^n$ clearly equals $\prod X_i$, and we know from Proposition 10.75 that $\prod X_i$ is separably metrizable, we get that $\prod X_i$ is locally compact. □

Here is one basic application of product topologies.

Theorem 10.79. *(Urysohn metrization)*[1] *Every T_1, second countable, hereditarily separable, regular space is separably metrizable.*

Proof. Let X be such a space. Let U_i ($i \in \mathbb{N}$) be the basic open classes. Recall that since X is hereditarily separable, for any i and j the intersection $\overline{U}_i \cap U_j^c$ is definitely either empty or inhabited (Proposition 10.56). Thus $\overline{U}_i$ definitely either is or is not contained in U_j.

For all $i, j \in \mathbb{N}$ for which $\overline{U}_i \subseteq U_j$, use Proposition 10.67 and Urysohn's lemma (Theorem 10.69) to get a continuous function $f_{i,j} : X \to [0,1]$ with $\overline{U}_i \subseteq f^{-1}(0)$ and $U_j^c \subseteq f^{-1}(1)$. This yields countably many functions from X into $[0,1]$, and they can be amalgamated into a single continuous function $f : X \to [0,1]^\omega$ by the universal property of product topologies (Theorem 10.73).

Since X is T_1, for any distinct x and y we can find a basic open class U_j containing x but not y. Then by regularity we can find a basic open class U_i which contains x and whose closure is contained in U_j. For this i and j we have $f_{i,j}(x) = 0$ and $f_{i,j}(y) = 1$. So f is injective.

By Proposition 10.75, $[0,1]^\omega$ is separably metrizable. Since f is injective, we can then define a metric on X by setting $d(x,y) = d(f(x), f(y))$. We must show that the metric topology for this metric agrees with the original topology on X. First, every metrically open class is open for the original topology because f is continuous. Conversely, it will suffice to show that

[1] Ukrainian mathematician Pavel Urysohn (Urysohn's lemma, Urysohn metrization theorem, Urysohn universal space, Fréchet-Urysohn space, Menger-Urysohn dimension, Urysohn integral equation) died at twenty-six. What have you done today?

every U_j is metrically open. Fix j and let (U_{n_k}) be a sequence whose terms are precisely those U_i's whose closure is contained in U_j. For each k, then, there is a metrically open class $f^{-1}_{n_k,j}([0,1))$ which contains U_{n_k} and is contained in U_j. Taking the union of these classes yields a metrically open class which contains every U_{n_k} and is contained in U_j. But then it must equal U_j because, by regularity, every element of U_j belongs to a basic open class U_i whose closure is contained in U_j. We conclude that U_j is metrically open. As j was arbitrary, this shows that the metric topology on X induced from $[0,1]^\omega$ equals the original topology on X. □

The proof of Theorem 10.79 shows that we have a homeomorphic embedding of X into $[0,1]^\omega$, provided the image of X in $[0,1]^\omega$ exists. If X is σ-compact then we know this from Theorem 10.44; thus we have the following result.

Theorem 10.80. *(Embedding in cubes) Every metrizable σ-compact space is homeomorphic to a subspace of $[0,1]^\omega$.*

The unbounded existential quantifier (there exists a homeomorphism) is legitimate because we have an explicit description of the desired homeomorphism.

We can now identify a universal property of the *Cantor space* $\{0,1\}^\omega$. It is homeomorphic to the middle-thirds Cantor space mentioned at the beginning of Chapter 9; this follows from Theorem 10.44 using the map which takes a sequence of 0's and 1's in $\{0,1\}^{\mathbb{N}}$ to the number in $[0,1]$ whose base 3 expansion is the corresponding sequence of 0's and 2's.

Lemma 10.81. *There is a continuous map from $\{0,1\}^\omega$ onto $[0,1]$. For every inhabited closed subspace K of $\{0,1\}^\omega$ there is a continuous retraction from $\{0,1\}^\omega$ onto K, i.e., a continuous function from $\{0,1\}^\omega$ to K which fixes every element of K.*

Proof. Consider the map from $\{0,1\}^\omega$ to $[0,1]$ which takes a sequence of 0's and 1's to the number in $[0,1]$ which has that same sequence as a binary expansion. It is not injective but it is continuous and surjective.

Let $C \subset [0,1]$ be the middle-thirds Cantor space and let K be an inhabited closed subspace of it. We will find a continuous retraction from C onto K.

Every element of $C \setminus K$ is either less than everything in K, or greater than everything in K, or, since K is closed, belongs to an interval (a,b) with $a,b \in K$ but $(a,b) \cap K = \emptyset$. Define $f : C \to K$ by letting $f(x) = x$ for

all $x \in K$; letting $f(x)$ be the minimal element of K for every $x \in C$ that is less than every element of K; letting $f(x)$ be the maximal element of K for every $x \in C$ that is greater than every element of K; and, for each $a, b \in K$ with $a < b$ and $(a, b) \cap K = \emptyset$, choosing an element c of $(a, b) \setminus C$ and then sending every element of $(a, c) \cap C$ to a and every element of $(c, b) \cap C$ to b. One straightforwardly checks that this map is continuous. □

Theorem 10.82. *(Universal property of the Cantor space) Every compact space is a continuous image of the Cantor space.*

Proof. Let X be a compact space. By Theorem 10.80 we may take it to be a closed subspace of $[0, 1]^\omega$. From Lemma 10.81 we have a continuous function from $C = \{0, 1\}^\omega$ onto $[0, 1]$, and together with the obvious homeomorphism between $C^\omega = (\{0, 1\}^\omega)^\omega$ and C this induces a continuous function ϕ from C onto $[0, 1]^\omega$.

Now $\phi^{-1}(X)$ is a closed subspace of C, so by the second part of Lemma 10.81 we have a continuous retraction from C onto $\phi^{-1}(X)$; composing this retraction with ϕ then yields a continuous function from C onto X. □

The product topology construction is a special case of a more general construction. Let X be a class and let $f_i : X \to X_i$ ($1 \leq i \leq n$ or $i \in \mathbb{N}$) be functions from X into second countable spaces X_i. Fixing a basis for the topology on each X_i, there are countably many subclasses of X of the form $f_i^{-1}(U)$ with i arbitrary and U a basic open class in X_i; the topology for which they constitute a subbasis is the *weak* topology relative to the f_i's. This construction does not essentially depend on the choice of bases of the X_i's.

10.5 Coherence

Another kind of topological space that will be important for us is a *coherent* topological space.

A *filtration* of a class X is given by a sequence of subclasses $X_1 \subseteq X_2 \subseteq \cdots$ whose union is X.

Definition 10.83. Let X be a class with filtration $X_1 \subseteq X_2 \subseteq \cdots$. Suppose each X_i is equipped with a hereditarily separable topology $\mathcal{T}_i \subseteq X_i \times I_i$ such that $\mathcal{T}_i$ is the subspace topology on X_i coming from $\mathcal{T}_{i+1}$, for all i.

Given these ingredients, we define a *coherent* topology $\mathcal{T}$ on X as follows. Let $I = \prod I_i$ and let $J \subseteq I$ consist of those sequences $\vec{\lambda} \in I$ such that the

intersection of X_i with the λ_{i+1}-slice of $\mathcal{T}_{i+1}$ equals the λ_i-slice of $\mathcal{T}_i$, for all i. In other words, we consider sequences $\vec{\lambda}$ which index relatively open subclasses $U_1 \subseteq X_1$, $U_2 \subseteq X_2$, ... such that $U_i = U_{i+1} \cap X_i$ for all i. This condition is definite by the LPO, because two open subclasses of a hereditarily separable space are equal if and only if their complements are contained in each other, which is definite by Proposition 10.56. Then we let the $\vec{\lambda}$-slice of $\mathcal{T} \subseteq X \times J$ be the union over $i \in \mathbb{N}$ of the λ_i-slice of $\mathcal{T}_i$. In terms of the relatively open classes U_i, this slice of $\mathcal{T}$ is just $\bigcup U_i$.

It is straightforward to check that $\mathcal{T}$ is a topology on X. Here is a simpler characterization:

Proposition 10.84. *Let X be a coherent topological space relative to the filtration $X_1 \subseteq X_2 \subseteq \cdots$. Then a subclass of X is open if and only if its intersection with each X_i is relatively open in X_i, and a subclass of X is closed if and only if its intersection with each X_i is relatively closed in X_i. A function $f : X \to Y$ to another topological space Y is continuous if and only if $f|_{X_i}$ is continuous for all i.*

The most basic examples of coherent topologies are metric.

Example 10.85. Let X be a hereditarily separable metric space and let $x \in X$. Then the closed balls $\overline{B}_n(x)$ satisfy the hypotheses to define a coherent topology. The metric topology has the property that a subclass is closed if and only if its intersection with each $\overline{B}_n(x)$ is closed, so in this case the coherent topology is just the standard metric topology.

In Example 10.85, if X is boundedly compact then it is coherent relative to the compact subspaces $\overline{B}_n(x)$. It follows by Proposition 10.47 that every locally compact space is coherent relative to a filtration consisting of compact subspaces. Thus every locally compact space is *compactly coherent* — coherent relative to a *compact filtration* $X_1 \subseteq X_2 \subseteq \cdots$ in which each X_i is compact.

In algebraic settings local compactness is a strong assumption that has a lot of nice consequences. But for many purposes a compactly coherent topology is just as good. This is a little surprising, as coherent topologies are typically not even second countable. Compactly coherent spaces that are not locally compact appear in nature as, for example, coordinate rings and dual Banach spaces; see Sections 11.4 and 15.3.

In the rest of this section I will collect some results which demonstrate the tractability of compactly coherent spaces.

Proposition 10.86. *Let X be a class equipped with a compactly coherent topology relative to the compact filtration $X_1 \subseteq X_2 \subseteq \cdots$. Then*

(a) X is σ-compact, hereditarily separable, T_1, and normal
(b) the subspace topology on any closed subclass of X is compactly coherent
(c) every convergent sequence in X is contained in and converges to a limit in some X_i
(d) every compact subspace of X is contained in some X_i
(e) if $f : X \to Y$ is a continuous function into some other compactly coherent space and $K \subseteq X$ is compact, then $f(K)$ exists as a compact subclass of Y
(f) if $f : X \to Y$ is a continuous function into some other compactly coherent space then the image $f(X)$ exists as a subclass of Y.

Proof. X is obviously σ-compact. It is then hereditarily separable by Proposition 10.76. X is T_1 by Proposition 10.84.

Normality is built into our definition of compactness (cf. Urysohn's lemma for metric spaces, Proposition 10.22). So if $C, D \subseteq X$ are disjoint and closed then $C \cap X_1$ and $D \cap X_1$ are disjoint closed subclasses of a compact metric space, and therefore we can find relatively open subclasses $U_1, V_1 \subseteq X_1$ whose closures are disjoint and which contain $C \cap X_1$ and $D \cap X_1$, respectively. Then we can find relatively open subclasses $U_2, V_2 \subseteq X_2$ whose closures are disjoint and which contain $\overline{U}_1 \cup (C \cap X_2)$ and $\overline{V}_1 \cup (D \cap X_2)$, and so on indefinitely. Then let $U = \bigcup_{i=n}^{\infty} U_i$ and $V = \bigcup_{i=n}^{\infty} V_i$. These are disjoint classes whose intersections with the X_i are relatively open, so they are open, and by construction they contain C and D, respectively. This shows that X is normal. We have proven part (a).

Next, if $C \subseteq X$ is closed then its subspace topology is compactly coherent relative to the compact filtration $C \cap X_1 \subseteq C \cap X_2 \subseteq \cdots$. This is part (b).

For part (c), let $x_n \to x$ in X and suppose that for each i there exists n_i with $x_{n_i} \notin X_i$. Then $C_i = \{x_{n_i}, x_{n_{i+1}}, \ldots\}$ is closed for all i by Proposition 10.84, so x must belong to $\bigcap C_i = \emptyset$. As membership in classes is definite, this contradiction shows that there must exist i such that $x_n \in X_i$ for all n.

Any compact subspace definitely is or is not contained in some X_i (Proposition 10.56 plus the LPO). If some compact subspace K were not contained in any X_i, then we could find a sequence (x_i) in K with $x_i \notin X_i$

for all i, which would contradict compactness by part (c). So every compact subspace must be contained in some X_i. This proves part (d).

For part (e), let $\mathring{K}$ be a dense subset of K and suppose Y is compactly coherent relative to the filtration $Y_1 \subseteq Y_2 \subseteq \cdots$. Then it definitely is or is not the case that $f(\mathring{K})$ is contained in some Y_i. If not, then we can find a sequence (x_i) of elements of $\mathring{K}$ such that $f(x_i) \notin Y_i$ for all i. By passing to a subsequence we can assume that (x_i) converges in X, but this contradicts Lemma 10.60 plus part (c). We conclude that there exists i with $f(\mathring{K}) \subseteq Y_i$. By Lemma 10.60, it then follows that $f(K) \subseteq Y_i$, and $f(K)$ is then compact by Theorem 10.44.

Part (f) follows from part (e) because X is σ-compact. □

Corollary 10.87. *A compactly coherent space is locally compact if and only if it is second countable.*

Proof. The forward direction is just Lemma 10.11, since we require locally compact spaces to be separably metrizable. For the reverse direction, let X be a compactly coherent second countable space. By the Urysohn metrization theorem (Theorem 10.79) X is separably metrizable. Let $X_1 \subseteq X_2 \subseteq \cdots$ be the given compact filtration on X and let $\mathring{X} = \{x_1, x_2, \ldots\}$ be a dense subset of X. I claim that for every $x \in X$ there exists a rational $r > 0$ such that $B_r(x)$ is contained in some X_i. For any given x, this definitely is or is not the case by Proposition 10.55 (and the LPO), since an open class is contained in X_i if and only if it fails to meet X_i^c. If it were not the case then we could find a sequence (y_i) that converges to x with $y_i \notin X_i$ for all i. But this contradicts Proposition 10.86 (c). That proves the claim. It follows that for every $x \in X$ there exists $k \in \mathbb{N}$ and $r > 0$ rational such that $x \in B_r(x_k)$ and $B_r(x_k)$ is contained in some X_i.

Now let $U_1, U_2, \ldots$ enumerate all the open balls $B_r(x_k)$ for $k \in \mathbb{N}$ and $r > 0$ rational which are contained in some X_i. By the last paragraph their union is X, and the closure of each one is compact. This shows that X is locally compact. □

We can also consider "locally compactly coherent" spaces which are coherent relative to a filtration by locally compact subspaces, or even "compactly coherently coherent" spaces which are coherent relative to a filtration by compactly coherent spaces. But every such space is actually compactly coherent.

Proposition 10.88. *Let X be a topological space whose topology is coher-*

ent relative to a filtration $X_1 \subseteq X_2 \subseteq \cdots$ in which each X_i is compactly coherent. Then X is compactly coherent.

Proof. For each i let $X_1^i \subseteq X_2^i \subseteq \cdots$ be a compact filtration which gives rise to the given topology on X_i. Then define $Y_1 = X_1^1$, $Y_2 = X_2^1 \cup X_1^2$, $Y_3 = X_3^1 \cup X_2^2 \cup X_1^3$, etc. These are all compact subspaces of X. (The union of two compact subspaces of X is compact. Just as in Proposition 10.86 (d), any two compact subspaces of X must be contained in some X_i, and hence in some X_n^i. It is then clear that their union is compact.) Moreover, any subclass of X whose intersection with each X_i is closed will also have closed intersection with each Y_j (since X_j^i is closed in X_i for all j), and conversely, any subclass whose intersection with each Y_i is closed also has closed intersection with each X_j^i (since X_j^i is a closed subclass of Y_{i+j-1}) and hence with each X_i. So $Y_1 \subseteq Y_2 \subseteq \cdots$ is a compact filtration of X which gives rise to the given topology. □

Now here is a version of Proposition 10.86 (b) for open classes.

Corollary 10.89. *Every open subspace of a compactly coherent space is compactly coherent.*

Proof. Let X be the space, U the open subspace, and $X_1 \subseteq X_2 \subseteq \cdots$ the given compact filtration of X. Then $U \cap X_i$ is relatively open in X_i and hence locally compact, for each i. So U is coherent relative to the locally compact spaces $X_1 \cap U \subseteq X_2 \cap U \subseteq \cdots$, and by Proposition 10.88 this shows that U is compactly coherent. □

Here is an analog of Proposition 10.66 for compactly coherent spaces.

Proposition 10.90. *Let X and Y be compactly coherent spaces and let $f : X \to Y$ be a function. Then f is continuous if and only if it is sequentially continuous.*

Proof. The forward direction is Lemma 10.60. For the converse, suppose f is sequentially continuous and let X_i and Y_i be the approximating compact subspaces of X and Y. By Proposition 10.84, to show that f is continuous it will suffice to show that its restriction to each X_i is continuous.

Fix i. The proof of Proposition 10.86 (e) used only sequential continuity, so it applies here to show that $f(X_i) \subseteq Y_j$ for some j. The result now follows from Theorem 10.20. □

And here is a compactly coherent analog of Proposition 10.54.

Proposition 10.91. *Let X be a compactly coherent space and let $C \subseteq X$. Then C is closed if and only if it is separable and sequentially closed.*

Proof. If C is closed then its intersection with each approximating compact subspace X_i is closed, and this implies that it is sequentially closed by Proposition 10.86 (c). Conversely, suppose C is separable and sequentially closed and let $\mathring{C}$ be a dense subset of C. Then for each $i \in \mathbb{N}$ we have

$$X_i \cap C = X_i \cap \bigcup_{j>i} \overline{\mathring{C} \cap X_j} = \bigcup_{j>i} \left(X_i \cap \overline{\mathring{C} \cap X_j} \right)$$

(Proposition 10.86 (c) again). This shows that $X_i \cap C$ is separable, and therefore it is closed by Proposition 10.16. Thus C is closed. □

One has to be a little careful with the idea of "sequential closure" here. The following definition is sometimes useful:

Definition 10.92. Let A be a subset of a compactly coherent space X with approximating compact subspaces X_i. The *sequential pseudoclosure* $\breve{A}$ of A is then the union over i of $\overline{A \cap X_i}$.

By Proposition 10.86 (c), a point belongs to $\breve{A}$ if and only if it is the limit of a sequence of elements of A. However, $\breve{A}$ might not be closed (though in many cases it will be). (For a counterexample, take $\mathbb{R}[x]$ in Example 11.29 below. For each n let $A_n = \{\frac{1}{n} + \frac{1}{k}x^n : k \in \mathbb{N}\}$. Then the sequential pseudoclosure of $\bigcup A_n$ contains $\frac{1}{n}$ for all n but not 0.)

If X is compactly coherent, say that a subset $\mathring{X} \subseteq X$ is *coherently dense* if there is a compact filtration $X_1 \subseteq X_2 \subseteq \cdots$ that gives rise to the topology on X, such that $\mathring{X} \cap X_i$ is dense in X_i for all i.

I defined $C(X, Y)$ and $\mathrm{Homeo}(X)$ for locally compact spaces in Section 10.3. We can now generalize those definitions to compactly coherent spaces, using the following theorem.

Theorem 10.93. *Let X and Y be compactly coherent spaces and let $\mathring{X}$ be a coherently dense subset of X. Then every function from $\mathring{X}$ into Y definitely does or does not extend to (1) a continuous function from X to Y, or (2) a homeomorphism between X and Y.*

Proof. Let $\mathring{f} : \mathring{X} \to Y$ be a function. For (1), it definitely is or is not the case that for each i the image of $\mathring{X} \cap X_i$ is contained in some Y_j. If this fails for any i, then $\mathring{f}$ does not extend to a continuous function from X to Y.

Otherwise, we can use Theorem 10.48 to determine for each i whether the restriction of $\mathring{f}$ to $\mathring{X} \cap X_i$ extends continuously to X_i. If not, then again $\mathring{f}$ does not extend to a continuous function from X to Y; if so, then since the extensions are unique they must be compatible, and then $\mathring{f}$ does extend to a continuous function $f : X \to Y$ by Proposition 10.84.

For (2), first determine whether $\mathring{f}$ extends to a continuous function $f : X \to Y$. If yes, let $\mathring{Y}$ be a coherently dense subset of Y and check whether it is contained in $f(X)$. If not, f is not a homeomorphism. If so, check whether each $f|_{X_i}$ is a homeomorphism between X_i and $f(X_i)$ using Corollary 10.50; if not, then again f is not a homeomorphism. If so, then f must be injective, so we can define $\mathring{g} : \mathring{Y} \to X$ to be the inverse of the restriction of f to $f^{-1}(\mathring{Y})$, then use (1) to check whether $\mathring{g}$ extends to a continuous function from Y to X. As in the proof of Corollary 10.50, f is a homeomorphism if and only if this extension exists. □

Thus we define $C(X, Y)$ to be the subclass of $X^\omega \times Y^\omega$ consisting of pairs of sequences that generate continuous functions from X to Y, modulo equality of the generated functions, and we define $\mathrm{Homeo}(X)$ similarly. These are essentially the same as the constructions given at the end of Section 10.3, and they are both exact pseudoclasses. As I already noted for locally compact spaces, this means that if X and Y are compactly coherent then the statement that they are homeomorphic can be taken to involve a bounded existential quantifier.

Finally, products. I have only defined the product topology on $X \times Y$ if X and Y are both second countable, but if they are both coherent with approximating compact subspaces X_i and Y_i, we can define the *coherent product topology* on $X \times Y$ to be the coherent topology arising from the compact filtration $X_1 \times Y_1 \subseteq X_2 \times Y_2 \subseteq \cdots$, giving each $X_i \times Y_i$ its ordinary product topology. The first thing to check is that this construction does not depend on the choice of compact filtrations.

Proposition 10.94. *Let X be a compactly coherent space with compact filtrations $X_1 \subseteq X_2 \subseteq \cdots$ and $X'_1 \subseteq X'_2 \subseteq \cdots$, both of which give rise to the topology on X. Let Y, Y_i, and Y'_i have the same properties. Then the coherent product of X and Y relative to the first pair of filtrations equals the coherent product relative to the second pair.*

Proof. By Proposition 10.86 (d), for each k we can find n such that $X_k \subseteq X'_n$ and $Y_k \subseteq Y'_n$. So any subclass of $X \times Y$ whose intersection with $X'_n \times Y'_n$ is closed also has closed intersection with $X_k \times Y_k$. The reverse

is also true, and this shows that a class has closed intersection with every $X_i \times Y_i$ if and only if it has closed intersection with every $X_i' \times Y_i'$. We conclude the proof by appealing to Proposition 10.84. □

The coherent product of compactly coherent spaces has the desired properties of a product:

Proposition 10.95. *Let Z_1 and Z_2 be compactly coherent spaces and give $Z_1 \times Z_2$ the coherent product topology. Then $Z_1 \times Z_2$ is compactly coherent. The projections from $Z_1 \times Z_2$ to Z_1 and Z_2 are continuous, and if Z is also compactly coherent then a function $f : Z \to Z_1 \times Z_2$ is continuous if and only if both of its components $f_i : Z \to Z_i$ $(i = 1, 2)$ are continuous. A sequence in $Z_1 \times Z_2$ converges if and only if it converges on both coordinates.*

Proof. It is immediate from the definition of the coherent product topology that $Z_1 \times Z_2$ is compactly coherent. The projections π_i are continuous because, e.g., if U is any open class in Z_1 then $\pi_1^{-1}(U) = U \times Z_2$ is open for the coherent product topology. This implies that if $f : Z \to Z_1 \times Z_2$ is continuous then so are its components $f_i = \pi_i \circ f$. For the converse, by Proposition 10.84 we can reduce to the case where Z is compact; in that case, if f_1 and f_2 are continuous then $f_1(Z)$ is contained in one of the approximating compact spaces for Z_1 by Proposition 10.86 (c), and the same is true for $f_2(Z)$, so f maps Z into an ordinary product of two compact spaces and is continuous by the universal property of product topologies (Theorem 10.73).

Let (x_n) be a sequence in $Z_1 \times Z_2$. If it converges, then it converges on each coordinate by Lemma 10.60 because the projections are continuous, as we just showed. Conversely, if (x_n) converges on each coordinate then its projection in Z_1 is contained in one of the approximating compact spaces for Z_1 (Proposition 10.86 (c)), and similarly for its other projection. So (x_n) is contained in a product of two compact spaces with its usual product topology and its convergence follows from Proposition 10.74. □

It should not be surprising, then, that if X and Y are locally compact then the coherent product topology is just the standard product topology.

Proposition 10.96. *Let X and Y be locally compact spaces. Then the standard product topology on $X \times Y$ equals the coherent product topology on $X \times Y$.*

Proof. Let Z and Z' be $X \times Y$ with the standard product topology and the coherent product topology, respectively. Then Z is locally compact

(Corollary 10.78) and therefore compactly coherent, so continuity of the two coordinate projections on Z plus Proposition 10.95 implies that the identity map from Z to Z' is continuous. The continuity of the two coordinate projections on Z' (Proposition 10.95) then yields continuity of the identity map from Z' to Z by the universal property of product topologies (Theorem 10.73). □

Chapter 11

Abstract Algebra

11.1 Groups

Definition 11.1. A *group operation* on a class G is a function $*$ from $G \times G$ to G that satisfies

(1) (associativity) $(g * h) * k = g * (h * k)$ for all $g, h, k \in G$
(2) (identity) there exists an *identity element* $e \in G$ such that $g * e = e * g = g$ for all $g \in G$
(3) (inverses) for every $g \in G$ there exists an *inverse* element $g^{-1} \in G$ such that $g * g^{-1} = g^{-1} * g = e$.

G is a *group* relative to $*$. If we additionally have $g * h = h * g$ for all $g, h \in G$, then G is *abelian*.

"Let G be a group" will be our typical shorthand for "let G be a class and let $*$ be a group operation on G". For general groups I will usually write "e" for the identity element, even when more than one group is in play, unless G is abelian, in which case I will usually write "0". Also, if G is abelian I will usually denote the group operation "+" and refer to it as a "sum", and write "$-g$" for the inverse of g. If G is not abelian, or not assumed to be abelian, I will usually write gh or $g \cdot h$ for $g * h$ and refer to this as a "product".

The inverse operation is always a function from G to itself. As a relation it is the class of ordered pairs (g, h) with the property that $gh = hg = e$; this condition is definite, so it does identify a class.

Example 11.2. $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ are all examples of abelian groups, taking the group operation to be ordinary addition.

Example 11.3. If X is a compactly coherent topological space then there is a pseudoclass Homeo(X) of all self-homeomorphisms of X. Composition defines a group operation on Homeo(X).

If X is countable and equipped with the *discrete topology* according to which all subsets of X are open, then Homeo(X) is just the pseudoclass of all permutations of X (bijections of X with itself). In this case Homeo(X) is a nonabelian group, provided X has at least three elements.

Generally speaking, uncountable algebraic objects tend to be less well-behaved than countable ones. This issue is mitigated to some degree by the fact that uncountable algebraic objects typically come with a natural topology, which often restores good behavior when brought into play. For instance, the image of a proper class under a function need not exist as a class, but by Proposition 10.86 (f), if the domain and range are equipped with compactly coherent topologies with respect to which the function is continuous, the image does exist.

For this reason, the following notion is basic for us.

Definition 11.4. Let G be a class and let $*$ be a group operation on G. Then a compactly coherent topology on G is *compatible with the group structure* if $* : G \times G \to G$ and $^{-1} : G \to G$ are both continuous. A *compactly coherent group* is a class equipped with a group operation and a compactly coherent topology that is compatible with the group structure.

In this definition we use the coherent product topology on $G \times G$. Thus according to Propositions 10.95 and 10.90 the condition for a compactly coherent topology to be compatible with the group structure is that $g_i \to g$ and $h_i \to h$ should imply $g_i^{-1} \to g^{-1}$ and $g_i h_i \to gh$.

We can use terms like "compact group" and "locally compact group" as well. In the group theory setting we are mainly interested in locally compact groups, but the hypothesis of local compactness can often be weakened to compact coherence. We will see later in this chapter that there are rings of great interest that are compactly coherent but not locally compact. However, on a first reading of the present section one may want to assume local compactness everywhere.

(Much of the material in this section can also be adapted to the case of separably metrizable groups. However, the principal examples of separably metrizable groups that are not locally compact are infinite dimensional Banach spaces, which are more naturally dealt with as vector spaces rather than groups; see Section 15.1. So I will not develop the material of this

section in that direction.)

Every countable group G can be given the discrete topology. This topology is only compact if G is finite, but it is always locally compact. The discrete topology is compatible with any group structure on any set. Thus everything we say about compactly coherent groups specializes to purely algebraic (i.e., topology-free) statements about discrete countable groups.

The standard topologies on $\mathbb{Z}$, $\mathbb{R}$, and $\mathbb{C}$ are compatible with the standard group structures on these classes. The standard topology on $\mathbb{Q}$, however, is not compactly coherent (by Corollary 10.87, since it is second countable but not locally compact), so if one is dealing with $\mathbb{Q}$ as a stand-alone algebraic object one may want to use the discrete topology on it; in fact, I will assume throughout that every countable group is equipped with the discrete topology. (That is, unless it is considered as a subgroup of some larger compactly coherent group, e.g., $\mathbb{Q} \subseteq \mathbb{R}$.)

A *homomorphism* from one group G to another group H is a function $\phi : G \to H$ satisfying $\phi(gh) = \phi(g)\phi(h)$ for all $g, h \in G$. An *isomorphism* is a bijective homomorphism, written $\phi : G \cong H$. This was the same notation we used for a bijection between two sets, or for a homeomorphism between metric or topological spaces; in general, if A and B are classes with any sort of additional structure we may write $f : A \cong B$ to mean that f is a bijection that preserves all the relevant structure. Hopefully what constitutes "relevant structure" will always be clear.

As with homeomorphisms between metric or topological spaces, for two groups to be *isomorphic* means that there exists an isomorphism between them, and this is an unbounded existential quantifier. However, if both groups are compactly coherent and we are only interested in homeomorphic isomorphisms (or even continuous isomorphisms) then we can use a bounded existential quantifier over $C(G, H)$.

I omit the simple proof of the next proposition.

Proposition 11.5. *If $\phi : G \to H$ is a homomorphism of groups then it takes the identity element of G to the identity element of H and, for any $g \in G$, takes the inverse of g to the inverse of $\phi(g)$.*

A *subgroup* of a group G is a subclass $H \subseteq G$ which contains the identity element and is stable under products and inverses. The restriction to H of the group operation on G will then be a group operation on H.

Let G be a compactly coherent group. Then for each $g \in G$ we have a self-homeomorphism $\ell_g : h \mapsto gh$ of G. Define $\ell : G \to \mathrm{Homeo}(G)$ by $\ell : g \mapsto \ell_g$. In order for this to be a function (specifically, a subclass of

$G \times \mathrm{Homeo}(G)$), for any $g \in G$ and any self-homeomorphism $\phi : G \to G$ it must definitely be the case or not be the case that $\phi = \ell_g$. But since G is separable this equality can be tested on a dense subset of G, so definiteness follows from the LPO.

Theorem 11.6. *(Cayley's theorem) For any compactly coherent group G the map ℓ is an isomorphism between G and a subgroup of* $\mathrm{Homeo}(G)$.

Proof. It is clear that $\ell_g \circ \ell_h = \ell_{gh}$ for any $g, h \in G$, which shows that ℓ is a homomorphism. If $g \neq h$ then $\ell_g(e) = g \neq h = \ell_h(e)$, so that ℓ is injective.

It follows that the map ℓ is an isomorphism between G and $\ell(G)$, assuming the latter exists as a class. In fact it does, because $\phi \,\varepsilon\, \mathrm{Homeo}(G)$ equals ℓ_g for some $g \in G$ if and only if it equals ℓ_g for $g = \phi(e)$; so it suffices to check the equality $\phi(h) = \phi(e)h$ for all h belonging to some dense subset of G (LPO). □

As I mentioned above, any countable group G can be equipped with the discrete topology, with respect to which it is locally compact, and then $\mathrm{Homeo}(G)$ is just the group of (codes of) permutations of G. So Theorem 11.6 specializes in the countable setting to the statement that any countable group is isomorphic to a subgroup of the group of its permutations.

Indeed, the countable setting seems to be the main arena in which Cayley's theorem is helpful. It is somewhat unsatisfying in the uncountable setting because we have not topologized $\mathrm{Homeo}(G)$, so the embedding only respects algebraic structure. (For uncountable locally compact groups there is a different, more useful version of Cayley's theorem called the "left regular representation", which embeds G in a group of transformations in a manner which is both algebraically isomorphic and topologically homeomorphic.)

Next, we define product groups and note their universal property.

Definition 11.7. Let G_i ($1 \leq i \leq n$ or $i \in \mathbb{N}$) be groups. Then we define a group operation on $\prod G_i$ by putting $\vec{g} \cdot \vec{h} = (g_1 h_1, \ldots, g_n h_n)$ or $(g_i h_i)$. This is a function because for any $\vec{g}, \vec{h}, \vec{k} \in \prod G_i$ it definitely is or is not the case, for each i, that $g_i h_i = k_i$, and then by the LPO it definitely is or is not the case that $\vec{g} \cdot \vec{h} = \vec{k}$ (without the need for the LPO in the case of a finite product). And $\vec{g} \cdot \vec{h}$ is easily seen to be a group operation, with the identity and inverses defined coordinatewise.

If the G_i's are compactly coherent and the product is finite, then we equip it with the coherent product topology. If each G_i is compact and the product is finite or infinite, then we equip $\prod G_i$ with the usual product

topology. By Proposition 10.96 these two conventions agree in the case of a finite product of compact groups.

Theorem 11.8. *(Universal property of product groups) Let G and H_i ($1 \leq i \leq n$ or $i \in \mathbb{N}$) be groups. Then for any homomorphisms $\phi_i : G \to H_i$ there is a unique homomorphism $\phi : G \to \prod H_i$ satisfying $\phi_i = \pi_i \circ \phi$ for all i. If either G and all the H_i are compact, or the product is finite and G and all the H_i are compactly coherent, then ϕ is continuous if and only if each ϕ_i is continuous.*

This is a consequence of the universal property of products (Theorem 9.22), the universal property of product topologies (Theorem 10.73), and Proposition 10.95. Those results give us everything except the fact that ϕ is a homomophism, which is easy.

Now let us discuss quotients. Quotients of general topological spaces can be problematic, but quotients of compactly coherent groups are reasonably well-behaved.

Lemma 11.9. *Let U be a subclass of a group G and let $g \in G$. Then $gU = \{gh : h \in U\}$ and $Ug = \{hg : h \in U\}$ exist as subclasses of G.*

Proof. For each $h' \in G$, we need it to definitely either be the case or not be the case that $h' = gh$ for some $h \in U$. That is equivalent to asking whether $g^{-1}h'$ belongs to U, which is definite because there exists a unique element h of G such that $g^{-1}h' = h$, and this h definitely does or does not belong to U because U is a class. Thus gU is a class, and Ug is also a class by essentially the same argument. □

(Topology does not come into play in this lemma.)

Proposition 11.10. *Let G be a compactly coherent group. If $U \subseteq G$ is open and $V \subseteq G$ is separable then $UV = \{gh : g \in U, h \in V\}$ and $VU = \{hg : g \in U, h \in V\}$ both exist as open subclasses of G. If $C \subseteq G$ is compact and $D \subseteq G$ is closed, then CD and DC both exist as closed subclasses of G, compact if D is also compact. If C and D are both closed then CD exists as a separable subclass of G.*

Proof. Suppose U is open and V is separable with dense subset $\mathring{V}$. Then Uh exists as a subclass of G for all $h \in \mathring{V}$ by the lemma, and so $W = \bigcup_{h \in \mathring{V}} Uh$ exists as an open subclass of G. It is clear that every element of W has the form gh for some $g \in U$ and $h \in V$; we must show that every element of G of this form belongs to W.

Thus suppose $g \in U$ and $h \in V$. Let (h_i) be a sequence in $\mathring{V}$ that converges to h. Then $(gh h_i^{-1})$ converges to $g \in U$ since the group operation and the group inverse are both continuous. Since U is open, there exists $k \in \mathbb{N}$ such that $h' = gh h_k^{-1} \in U$. Then $gh = h' h_k$ belongs to W. We conclude that W consists of precisely the elements of G of the form gh for some $g \in U$ and $h \in V$. We have shown that UV exists as an open subclass of G; the argument for VU is identical.

Next, suppose C is compact and D is closed. Let $\mathring{C}$ and $\mathring{D}$ be dense subsets of C and D and let K be the sequential pseudoclosure of $\mathring{C}\mathring{D}$ (Definition 10.92). Then for any $g \in C$ and $h \in D$ we can find sequences (g_i) in $\mathring{C}$ and (h_i) in $\mathring{D}$ that converge to g and h, and $(g_i h_i)$ will then converge to gh. Since $g_i h_i \in \mathring{C}\mathring{D}$ for all n, this shows that gh belongs to K.

Conversely, let $h' \in K$; we must show that it has the form gh for some $g \in C$ and $h \in D$. This will establish that $CD = K$ exists as a subclass of G. Let (g_i) and (h_i) be sequences in $\mathring{C}$ and $\mathring{D}$, respectively, such that $g_i h_i \to h'$. By passing to a subsequence we can assume that (g_i) converges in C, say $g_i \to g$. But then $h_i = g_i^{-1} g_i h_i \to g^{-1} h'$, showing that $g^{-1} h' \in D$. That is sufficient because now $h' = g g^{-1} h' \in CD$.

We can now show that $CD = K$ is actually closed using Proposition 10.91; we just have to verify sequential closure. But if $(g_i h_i)$ is a sequence in CD that converges, just as above we can pass to a subsequence for which (g_i) converges, and this implies that (h_i) also converges, so that the limit of the sequence $(g_i h_i)$ also belongs to CD.

We have shown that the product of a compact class and a closed class always exists and is closed. The same argument applies to the product in the other order. If C and D are both compact, then CD is the image of the compact space $C \times D$ under the product map, so it is compact by Proposition 10.86 (e). Finally, if C and D are both merely closed, then C is a countable union of compact classes, so it follows from the compact/closed result that the product CD exists. Separability is clear. □

In Proposition 11.10, note that the product (in an abelian group, the sum) of two closed classes neither of which is compact need not be closed, not even if they are subgroups. Let me record a counterexample:

Example 11.11. $\mathbb{Z}$ and $\sqrt{2}\mathbb{Z}$ are both closed subgroups of $\mathbb{R}$. Their sum is dense in $\mathbb{R}$ but not equal to $\mathbb{R}$.

Given a group G and a subgroup H of G, define $g_1 \sim g_2$ if there exists

$h \in H$ such that $g_1 = g_2 h$. This is a relation, i.e., a subclass of $G \times G$, because for any $(g_1, g_2) \in G \times G$ it definitely is or is not the case that $g_2^{-1} g_1 \in H$, and therefore it definitely is or is not the case that $g_1 \sim g_2$. It is then an equivalence relation because (1) $g = g \cdot e$ for all $g \in G$; (2) if $g_1 = g_2 h$ for some $h \in H$ then $g_2 = g_1 h^{-1}$, and $h^{-1} \in H$ also; and (3) if $g_1 = g_2 h_1$ and $g_2 = g_3 h_2$ with $h_1, h_2 \in H$ then $g_1 = g_3 h_2 h_1$, and $h_2 h_1 \in H$ also. We write G/H for the quotient $G/\sim$, and $[g]$ or $[g]_H$ for the element of G/H instantiated by $g \in G$.

In this construction the blocks of $\sim$ are precisely the subclasses of G of the form $gH = \{gh : h \in H\}$, which I will call *left translations* of H.

Now suppose G is a compactly coherent group and H is a closed subgroup. For any open subclass $U \subseteq G$ the image $\pi(U) \subseteq G/H$ exists as a class, where $\pi : G \to G/H$ is the quotient map; it consists of those elements $[h]$ of G/H for which $h \in UH$. (And UH exists as a class by Proposition 11.10.) If $\mathcal{T} \subseteq G \times I$ is the given topology on G, define a *quotient* topology on G/H as the topology consisting of those pairs $([g], \lambda)$ in $(G/H) \times I$ such that $g \in U_\lambda H$, where U_λ is the λ-slice of $\mathcal{T}$. In this topology a class is open iff it is the image of an open class in G iff its inverse image under the quotient map is open in G.

Say that a subclass C of G is *H-invariant* if $g \in C$ implies $gh \in C$ for all $h \in H$, i.e., C contains every left translation of H that it meets.

Lemma 11.12. *Let G be a compactly coherent group, $H \subseteq G$ a closed subgroup, and $C, D \subseteq G$ compact with CH and DH disjoint. Then there exist disjoint open H-invariant classes $U, V \subseteq G$ with $C \subseteq U$ and $D \subseteq V$.*

Proof. Let $X_1 \subseteq X_2 \subseteq \cdots$ be a compact filtration for the topology on G. Then C and D are contained in some X_n by Proposition 10.86 (d); without loss of generality suppose they are contained in X_1. Put $C_1 = C$ and $D_1 = D$ and find a relatively open subclass U_1 of X_1 with $C_1 \subseteq U_1$ and $\overline{U}_1 \cap D_1 H = \emptyset$. This implies that $\overline{U}_1 H \cap D_1 H = \emptyset$. Then find a relatively open subclass V_1 of X_1 with $D_1 \subseteq V_1$ and $\overline{U}_1 H \cap \overline{V}_1 H = \emptyset$.

Inductively introduce classes U_n and V_n for all $n \in \mathbb{N}$ as follows. Given $U_n, V_n \subseteq X_n$, let $C_{n+1} = \overline{U}_n H \cap X_{n+1}$ and $D_{n+1} = \overline{V}_n H \cap X_{n+1}$ and as above find relatively open subclasses $U_{n+1}, V_{n+1} \subseteq X_{n+1}$ with $C_{n+1} \subseteq U_{n+1}$, $D_{n+1} \subseteq V_{n+1}$, and $\overline{U}_{n+1} H \cap \overline{V}_{n+1} H = \emptyset$. Finally, put $U = \bigcup U_n$ and $V = \bigcup V_n$; these are the desired disjoint H-invariant open subclasses containing C and D. (They are H-invariant because, e.g., $U_n H \cap X_{n+k} \subseteq U_{n+k}$ for all $n, k \in \mathbb{N}$.) $\square$

Theorem 11.13. *Let G be a compactly coherent group and let H be a closed subgroup of G. The projection $\pi(K)$ of any compact subspace $K \subseteq G$ exists and is compact in G/H. If G is compactly coherent relative to the compact filtration $X_1 \subseteq X_2 \subseteq \cdots$ then G/H is compactly coherent relative to the compact filtration $\pi(X_1) \subseteq \pi(X_2) \subseteq \cdots$. If G is locally compact or compact, then G/H has the same property.*

Proof. Fix a compact subspace K of G. Then $\pi(K) = \pi(KH)$ exists as a subclass of G/H by Proposition 9.24, where KH exists as a class by Proposition 11.10.

The quotient is T_1 because for any element $[g]$ of G/H the inverse image under π of the complement of the single point $\{[g]\}$ in G/H is the complement of gH in G, which is open since gH is closed. Thus the complement of $\{[g]\}$ is open and so $\{[g]\}$ is closed.

Since G is hereditarily separable (Proposition 10.86 (a)), G/H has the same property by Proposition 10.61. Observe also that $\pi(K) = \pi(KH) \subseteq G/H$ is closed since KH is closed (Proposition 11.10) and H-invariant, so $\pi(K)$ is also hereditarily separable.

Now suppose $C, D \subseteq \pi(K)$ are disjoint and relatively closed. Then applying Lemma 11.12 to $\pi^{-1}(C) \cap K$ and $\pi^{-1}(D) \cap K$ yields disjoint H-invariant open classes $U, V \subseteq G$ with $\pi^{-1}(C) \cap K \subseteq U$ and $\pi^{-1}(D) \cap K \subseteq V$. Since $\pi^{-1}(C) = (\pi^{-1}(C) \cap K)H$, it follows that $\pi^{-1}(C) \subseteq U$, and $\pi^{-1}(D) \subseteq V$ similarly. Then $\pi(U), \pi(V) \subseteq G/H$ are disjoint open classes which contain C and D, respectively. This shows that $\pi(K)$ is normal.

Next, let us show that $\pi(K)$ is second countable. Since K is compact and therefore second countable, fix a basis for the topology on K. Then for any basic relatively open classes $U_1, \ldots, U_k \subseteq K$ the class $G \setminus (\overline{U}_1 \cup \cdots \cup \overline{U}_k)H$ is open. Let V_j ($1 \leq j \leq n$ or $j \in \mathbb{N}$) enumerate these open classes. I claim that the classes $\pi(V_j) \cap \pi(K) = \pi(V_j \cap K)$ constitute a basis for the subspace topology on $\pi(K)$.

It is clear that each $\pi(V_j) \cap \pi(K)$ is relatively open in $\pi(K)$. We must show that every relatively open subclass of $\pi(K)$ is a union of some of the $\pi(V_j) \cap \pi(K)$. Thus let $W \subseteq G/H$ be open, so that $W \cap \pi(K)$ is an arbitrary relatively open subclass of $\pi(K)$; we want the latter to be a union of some of the $\pi(V_j) \cap \pi(K)$, namely all those $\pi(V_j) \cap \pi(K)$ that are contained in $W \cap \pi(K)$. (This condition is definite by Proposition 10.56.) To show this, for every $[g] \in W \cap \pi(K)$ we must find a V_j with $[g] \in \pi(V_j) \cap \pi(K) \subseteq W \cap \pi(K)$.

We can assume g belongs to K. Let $C = K \setminus \pi^{-1}(W)$. Lemma 11.12

gives us disjoint H-invariant open classes $V, V' \subseteq G$ satisfying $g \in V$ and $C \subseteq V'$. Now $V' \cap K$ is relatively open in K, and therefore is a union of some of the basic relatively open classes in K. Since $C \subseteq K$ is compact, there are finitely many basic relatively open classes $U_i \subseteq K$ whose union contains C and is contained in V'. Therefore the union of these classes is disjoint from V, so g belongs to the corresponding V_j. This V_j therefore has the desired property that $[g] \in \pi(V_j) \cap \pi(K) \subseteq W \cap \pi(K)$. This completes the proof that $\pi(K)$ is second countable.

We now know that $\pi(K)$ is T_1, second countable, hereditarily separable, and normal. It follows from the Urysohn metrization theorem (Theorem 10.79) that $\pi(K)$ is metrizable, and therefore compact, since it is a continuous image of K (Theorem 10.44).

The filtration $\pi(X_1) \subseteq \pi(X_2) \subseteq \cdots$ exhibits compact coherence of G/H, because if $U \cap \pi(X_i)$ is relatively open in $\pi(X_i)$ for all i then $\pi^{-1}(U) \cap X_i = \pi|_{X_i}^{-1}(U \cap \pi(X_i))$ is relatively open in X_i for all i, and hence $\pi^{-1}(U)$ is open in G. Thus $U = \pi(\pi^{-1}(U))$ is open in G/H. If G is locally compact then it is second countable, say with basic open classes U_i, and then the classes $\pi(U_i)$ constitute a basis for the quotient topology on G/H, so G/H is also second countable and therefore locally compact (Corollary 10.87). If G is compact then, as we just saw, G/H is locally compact, and therefore it is compact by Theorem 10.44. □

Whenever H is a subgroup of a group G and $g \in G$, the map $h \mapsto gh$ from H to gH is a bijection. In particular, if H is finite then all of its left translations have the same number of elements as H, a number we denote "$|H|$" and call the *order* of H. This yields the following result.

Theorem 11.14. *(Lagrange's theorem) If H is a subgroup of a finite group G then $|H|$ divides $|G|$.*

In particular, given a finite group G and $g \in G$, the least $n \in \mathbb{N}$ such that $g^n = e$, known as the *order* of g, must divide $|G|$. Because $\{e, g, \ldots, g^{n-1}\}$ is a subgroup of G. There exists such an n because finiteness implies that $g^i = g^j$ for some $i > j$, and then $n = i - j$ works.

A subgroup N of a group G is *normal* if gng^{-1} belongs to N for any $n \in N$ and any $g \in G$. The *kernel* of a group homomorphism $\phi : G \to H$ is the preimage $\ker(\phi) = \phi^{-1}(e) \subseteq G$ of the identity element of H.

Proposition 11.15. *Let G be a compactly coherent group and let N be a closed normal subgroup of G. Then there is a group operation on G/N with*

which the quotient topology is compatible, and which makes the quotient map $\pi : G \to G/N$ a continuous homomorphism whose kernel is N.

Proof. Define a group operation on G/N by putting $[g][h] = [gh]$ for any $g, h \in G$. In order for this to make sense we need to have $[g_1h_1] = [g_2h_2]$ whenever $[g_1] = [g_2]$ and $[h_1] = [h_2]$, i.e., whenever $g_2^{-1}g_1, h_2^{-1}h_1 \in N$. But if $g_2^{-1}g_1, h_2^{-1}h_1 \in N$ then

$$(g_2h_2)^{-1}(g_1h_1) = h_2^{-1}g_2^{-1}g_1h_1 = (h_2^{-1}g_2^{-1}g_1h_2)(h_2^{-1}h_1)$$

is a product of two elements of N and therefore belongs to N, as desired. It is now clear that π is a homomorphism whose kernel is N. It is obviously continuous.

Let X_k be the approximating compact subspaces of G; by Theorem 11.13 the quotient topology on G/H is compactly coherent with approximating compact subspaces $\pi(X_k)$. So to check that the quotient topology is compatible with the group operation we just defined on G/N, we must show that $[g_i] \to [g]$ and $[g'_i] \to [g']$ imply $[g_ig'_i] \to [gg']$ in G/H (Propositions 10.90 and 10.95). Fix convergent sequences $[g_i] \to [g]$ and $[g'_i] \to [g']$. By Proposition 10.86 (c) there is a value of k such that $[g]$, $[g']$, every $[g_i]$, and every $[g'_i]$ belong to $\pi(X_k)$, so we can assume $g, g', g_i, g'_i \in X_k$. Let (g_{i_j}) and (g'_{i_j}) be convergent subsequences of (g_i) and (g'_i); then $g_{i_j} \to gh$ and $g'_{i_j} \to g'h'$ for some $h, h' \in N$. Then by continuity of the product in G we have $g_{i_j}g'_{i_j} \to ghg'h'$, and therefore $[g_{i_j}][g'_{i_j}] \to [gg']$ in G/H (since $g \sim gh$ and $g' \sim g'h'$). We have shown that whenever $[g_i] \to [g]$ and $[g'_i] \to [g']$, there is a subsequence of $[g_ig'_i]$ that converges to $[gg']$.

That suffices to prove continuity of the group operation in G/H. A similar argument shows that the inverse operation in G/H is also continuous. We conclude that the quotient topology on G/H is compatible with its group structure. □

We call G/N a *quotient group*.

Theorem 11.16. *(Universal property of quotient groups) Let $\phi : G \to H$ be a continuous homomorphism between compactly coherent groups and let N be a closed normal subgroup of G that is contained in $\ker(\phi)$. Then there is a continuous homomorphism $\psi : G/N \to H$ such that $\phi = \psi \circ \pi$.*

Proof. The existence of a function $\psi : G/N \to H$ with $\phi = \psi \circ \pi$ follows from the universal property of quotients (Theorem 9.25): we just have to check that ϕ is compatible with the equivalence relation used to define G/N, and this is easy. It is also easy to check that ψ is a homomorphism, and

it is continuous because if $U \subseteq H$ is open then $\phi^{-1}(U)$ is an N-invariant open class in G, and therefore $\psi^{-1}(U) = \pi(\phi^{-1}(U))$ is open in G/N. □

Theorem 11.17. *(First isomorphism theorem for groups) Let $\phi : G \to H$ be a continuous homomorphism of compactly coherent groups. Then* $\ker(\phi)$ *is a closed normal subgroup of G, $\phi(G)$ is a subgroup of H, and there is a continuous isomorphism $\psi : G/\ker(\phi) \cong \phi(G) \subseteq H$ such that $\phi = \psi \circ \pi$.*

Proof. First, $\ker(\phi) = \phi^{-1}(e)$ is a closed subclass of G, and it is a subgroup of G because (1) $\phi(e) = e$; (2) if $m, n \in \ker(\phi)$ then $\phi(mn) = \phi(m)\phi(n) = e^2 = e$ and thus $mn \in \ker(\phi)$; and (3) if $n \in \ker(\phi)$ then $\phi(n^{-1}) = \phi(n)^{-1} = e^{-1} = e$, so $n^{-1} \in \ker(\phi)$. Moreover, for any $n \in \ker(\phi)$ and any $g \in G$ we have

$$\phi(gng^{-1}) = \phi(g)\phi(n)\phi(g^{-1}) = \phi(g)\phi(g^{-1}) = \phi(g)\phi(g)^{-1} = e.$$

So $gng^{-1} \in \ker(\phi)$, and thus we have shown that $\ker(\phi)$ is a closed normal subgroup of G.

The image $\phi(G)$ exists by Theorem 10.86 (f), and it is easy to see that $\phi(G)$ is a subgroup of H. The universal property of quotient groups (Theorem 11.16) now gives us a continuous homomorphism $\psi : G/\ker(\phi) \to \phi(G)$ such that $\phi = \psi \circ \pi$.

If $[g_1] \neq [g_2]$ in $G/\ker(\phi)$ then $g_2^{-1}g_1 \notin \ker(\phi)$ and therefore $\phi(g_2)^{-1}\phi(g_1) = \phi(g_2^{-1}g_1) \neq e$, i.e., $\phi(g_1) \neq \phi(g_2)$, so ψ is injective. Thus it is a continuous group isomorphism. □

The continuous isomorphism of Theorem 11.17 need not be a homeomorphism. For example, consider the homomorphism $\phi : \mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ defined by composing the function $n \mapsto \sqrt{2}n$ from $\mathbb{Z}$ to $\mathbb{R}$ with the quotient map from $\mathbb{R}$ to $\mathbb{R}/\mathbb{Z}$. Here $\mathbb{Z}$ is discrete, but $\phi(\mathbb{Z})$ is a dense subset of $\mathbb{R}/\mathbb{Z}$ and the topology it inherits from $\mathbb{R}/\mathbb{Z}$ is not discrete.

Let H_1 and H_2 be closed subgroups of a compactly coherent group G. If either one is normal then their product H_1H_2 is also a subgroup of G. For instance, if H_1 is normal then for any $n_1, n_2 \in H_1$ and $g_1, g_2 \in H_2$ we have

$$(n_1g_1)(n_2g_2) = n_1(g_1n_2g_1^{-1})(g_1g_2),$$

which belongs to H_1H_2 because $g_1n_2g_1^{-1} \in H_1$ by normality. (But recall that H_1H_2 need not be closed.)

Theorem 11.18. *(Second isomorphism theorem for groups) Let N and H be closed subgroups of a compactly coherent group G. Assume N is*

normal. Then $H \cap N$ is a closed normal subgroup of H, N is a closed normal subgroup of HN, and there is a continuous isomorphism from $H/(H \cap N)$ to $HN/N \subseteq G/N$ (giving HN/N the topology it inherits from G/N).

Proof. Normality of N in HN is trivial, since $HN \subseteq G$, and normality of $H \cap N$ in H is almost as easy. Define a function $\phi : H \to HN/N \subseteq G/N$ by putting $\phi(g) = [g]_N$. This is clearly a homomorphism. It is surjective because any element of HN/N has the form $[gn]_N$ for some $g \in H$ and $n \in N$, and $[gn]_N = [g]_N$. It is continuous because the identity map from H into $HN \subseteq G$ and the quotient map from G to G/N are both continuous.

The kernel of ϕ consists of those $g \in H$ with $[g]_N = [e]_N$, i.e., $g \in N$. That is to say, the kernel is just $H \cap N$. The proof is concluded by invoking the first isomorphism theorem for groups (Theorem 11.17). □

The continuous isomorphism of Theorem 11.18 also need not be a homeomorphism. Example 11.11 illustrates this. Also, HN need not be compactly coherent, though it is σ-compact. (Since H is closed in G, it is σ-compact, and the product of any compact subspace with N is closed in G and therefore again σ-compact.) HN/N is also σ-compact, by Proposition 10.86 (e).

Theorem 11.19. *(Third isomorphism theorem for groups) Let N and H be closed subgroups of a compactly coherent group G, and assume that N is normal and $N \subseteq H$. Then H/N is a closed subgroup of G/N, and every closed subgroup of G/N has this form. Moreover, H is normal in G if and only if H/N is normal in G/N, in which case $(G/N)/(H/N)$ is homeomorphically isomorphic to G/H.*

Proof. H/N is a subclass of G/N by Proposition 9.24. It is clear that the inclusion of H/N into G/N respects the group structure of H/N, so H/N is a subgroup of G/N. H/N is closed because $H^c \subset G$ is open and N-invariant, and therefore its image $(H/N)^c \subset (G/N)$ under the quotient map is open.

Given any closed subgroup H' of G/N, consider $\pi^{-1}(H')$ where $\pi : G \to G/N$ is the quotient map. This is a subgroup of G because (1) $\pi(e) = [e] \in H'$; (2) if $\pi(g) = [g] \in H'$ then $\pi(g^{-1}) = [g]^{-1} \in H'$; and (3) if $\pi(g), \pi(h) \in H'$ then $\pi(gh) = \pi(g)\pi(h) = [g][h] = [gh] \in H'$. Also $\pi^{-1}(H')$ is closed since π is continuous. It is clear that $\pi^{-1}(H')/N = H'$.

If H is normal then for any $h \in H$ and $g \in G$ we have $[g]^{-1}[h][g] = [g^{-1}hg] \in H/N$. This shows that H/N is normal in G/N. Conversely,

if H/N is normal then for any $h \in H$ and $g \in G$ we have $[g^{-1}hg] = [g]^{-1}[h][g] \in H/N$, so that $g^{-1}hg \in H$. Thus H is normal.

Finally, suppose H is normal in G (equivalently, H/N is normal in G/N) and consider the function $\phi : G \to (G/N)/(H/N)$ taking $g \in G$ to $[[g]_N]_{H/N} \in (G/N)/(H/N)$. This is a homomorphism and its kernel consists of those $g \in G$ such that $[[g]_N]_{H/N} = [[e]_N]_{H/N}$, i.e., $[g]_N \in H/N$. Since H contains N this shows that $\ker(\phi) = H$. Also, ϕ is the composition of two quotient maps, so it is continuous. As ϕ is clearly surjective, we get a continuous isomorphism from G/H to $(G/N)/(H/H)$ by the first isomorphism theorem for groups (Theorem 11.17).

We can also define a continuous homomorphism $\psi : G/N \to G/H$ by $\psi([g]_N) = [g]_H$. Its kernel is H/N, so it induces a continuous isomorphism from $(G/N)/(H/N)$ to G/H. This isomorphism is the inverse of the one from G/H to $(G/N)/(H/N)$, so it is a homeomorphic isomorphism. □

11.2 Number theory

We can use what we just learned about groups to pick up some basic number theory.

Example 11.20. In an abelian group, every subgroup is normal. Fix $n \in \mathbb{N}$ and take $G = \mathbb{Z}$ and $N = n\mathbb{Z} = \{nk : k \in \mathbb{Z}\}$; then N is a normal subgroup, and the quotient group $G/N = \mathbb{Z}/n\mathbb{Z}$ is an abelian group (under addition) with exactly n elements. It is the *cyclic group of order* n.

In fact, every subgroup of $\mathbb{Z}$ besides $\{0\}$ has the form $n\mathbb{Z}$. To see this, let N be any subgroup of $\mathbb{Z}$ which is not $\{0\}$. Then it contains a nonzero number a, and since it is stable under additive inverses we can assume a is positive. Now let n be the smallest natural number that belongs to N. Clearly $n\mathbb{Z} \subseteq N$; conversely, if b is any element of N we can divide by n and write $b = qn + r$ with $0 \leq r < n$. But since b and n both belong to N, this implies that r belongs to N, and minimality of n then implies that $r = 0$. So every element of N is divisible by n.

This observation leads to a quick proof of the next result.

Proposition 11.21. *(Bézout's identity) Let $a, b \in \mathbb{N}$ and let d be their greatest common divisor. Then the numbers $ax + by$ with $x, y \in \mathbb{Z}$ are precisely the integer multiples of d.*

Proof. Let $N = \{ax + by : x, y \in \mathbb{Z}\}$. It is easy to see that N is a

subgroup of $\mathbb{Z}$. It cannot be $\{0\}$, since a and b are nonzero. Thus, as we have just seen, it equals $N = n\mathbb{Z}$ for some $n \in \mathbb{N}$. Then d divides n, since it divides everything of the form $ax+by$. Conversely, write $n = ax_0+by_0$ and suppose a leaves the remainder r when divided by n, i.e., $a = nq+r$ where $0 \leq r < n$; then $r = a - nq = a - (ax_0 + by_0)q = (1 - x_0)a + (-qy_0)b \in N$, which implies that r is divisible by n. That can only happen if $r = 0$, so n divides a. Similarly, n divides b, and therefore it divides their gcd d. We have shown that d and n divide each other, so $d = n$, and this completes the proof. □

Write "$d|n$" for "d divides n".

Corollary 11.22. *(Euclid's lemma) Let $a, b \in \mathbb{N}$ and let $p \in \mathbb{N}$ be prime. If $p|ab$, then either $p|a$ or $p|b$.*

Proof. Either p divides a or p does not divide a. In the latter case, since p is prime, the greatest common divisor of a and p is 1, so Bézout's identity (Proposition 11.21) says there are integers x and y such that $ax + py = 1$. Then $abx + pby = b$, and p divides everything on the left side, so p must divide b. □

We can get a little more number theory by looking at the nonzero elements of $\mathbb{Z}/n\mathbb{Z}$ under multiplication, not addition. If $n = ab$ with $a, b < n$ then $[a][b] = [0]$ in $\mathbb{Z}/n\mathbb{Z}$, so we need n to be prime to ensure that the nonzero elements of $\mathbb{Z}/n\mathbb{Z}$ are stable under multiplication. The following lemma shows that this means the nonzero elements of $\mathbb{Z}/n\mathbb{Z}$ form a group under multiplication.

Lemma 11.23. *Let G be a finite set and let $* : G \times G \to G$ be a function which is associative. Assume G contains an element e such that $a * e = e * a = a$ for all $a \in G$, and assume that $a * b = a * c$ implies $b = c$, for all $a, b, c \in G$. Then $*$ is a group operation.*

Proof. We must show that inverses exist. Let $a \in G$ and consider its powers a, a^2, ... Since G is finite they cannot all be distinct, so find $i, j \in \mathbb{N}$ such that $a^{i+j} = a^i = a^i \cdot e$. By the cancellation property, we get $a^j = e$, so a^{j-1} is an inverse of a. □

Proposition 11.24. *For all $n \geq 2$, the set $\{[a] : 1 \leq a < n$ and $\gcd(a, n) = 1\} \subseteq \mathbb{Z}/n\mathbb{Z}$ is a multiplicative group. For any prime number p, the nonzero elements of $\mathbb{Z}/p\mathbb{Z}$ form a multiplicative group.*

Proof. Multiplication is well-defined in $\mathbb{Z}/n\mathbb{Z}$ because $(a+cn)(b+dn) = ab + (cb + ad + cdn)n$, so that $a \sim a'$, $b \sim b'$ implies $ab \sim a'b'$. And the elements $[a]$ with a coprime to n (i.e., $\gcd(a, n) = 1$) are stable under products because any prime that divides both ab and n must divide either a or b (Euclid's lemma, Corollary 11.22). To verify cancellation suppose $[ac] = [bc]$ with a, b, c all coprime to n. Then $n|(a-b)c$, but since no prime factor of n divides c we must have $n|(a-b)$, i.e., $[a] = [b]$. The result now follows from Lemma 11.23. □

For any $n \geq 2$ let "$\phi(n)$" denote the number of elements of the set $\{a \in \mathbb{N} : a < n \text{ and } a \text{ is coprime to } n\}$. This is *Euler's totient function*.

Theorem 11.25. *(Euler's theorem; Fermat's little theorem) Suppose $a, n \in \mathbb{N}$ are coprime. Then $a^{\phi(n)} - 1$ is divisible by n. If p is prime and $1 \leq a < p$ then $a^{p-1} - 1$ is divisible by p.*

The second statement follows from the first because every $a < p$ is coprime to p, and $\phi(p) = p - 1$. For the first statement, we have $[a]^{\phi(n)} = [1] \in \mathbb{Z}/n\mathbb{Z}$ for any a that is coprime to n because the order of any element must divide the order of the group.

Before moving on, let me mention one more fact about Euler's totient function, that we will use later.

Proposition 11.26. *Let $n \in \mathbb{N}$. Then $n = \sum_{d|n} \phi(d)$, where the sum is taken over all divisors d of n.*

Proof. For each divisor d of n, let $A(d) = \{k \leq n : \gcd(k, n) = d\}$. Then dividing every element of $A(d)$ by d yields the set $\{k \leq \frac{n}{d} : \gcd(k, \frac{n}{d}) = 1\}$. Thus $|A(d)| = \phi(\frac{n}{d})$. As the sets $A(d)$ are disjoint and their union is $\{1, \ldots, n\}$, this shows that $\sum_{d|n} \phi(\frac{n}{d}) = n$. But as d ranges over the divisors of n, so does $\frac{n}{d}$. So we get $\sum_{d|n} \phi(d) = n$. □

11.3 Rings

Definition 11.27. Let R be a class. *Ring operations* on R are functions $+, \cdot : R \times R \to R$ such that

(1) $+$ is an abelian group operation
(2) $\cdot$ is associative

(3) (distributivity) $(a+b)\cdot c = a\cdot c + b\cdot c$ and $c\cdot(a+b) = c\cdot a + c\cdot b$ for all $a, b, c \in R$.

A *ring* is a class equipped with ring operations. It is *commutative* if the product operation satisfies the abelian law $a \cdot b = b \cdot a$. It *is unital* if there is a multiplicative identity, an element "1" with the property that $a \cdot 1 = 1 \cdot a = a$ for all $a \in R$. We refer to 1 as the *identity element*, when it exists.

We refer to $+$ and $\cdot$ as "sum" and "product" and denote the additive inverse of a by "$-a$". We denote the additive identity by "0" and usually write "ab" for "$a \cdot b$".

Example 11.28. $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ are all examples of unital commutative rings, taking the ring operations to be ordinary addition and multiplication.

I will present a class of examples of noncommutative rings in Section 12.1.

A *compactly coherent ring* is a ring R equipped with a compactly coherent topology that makes addition, multiplication, and the additive inverse function continuous. As with groups, we take the topology on $R \times R$ to be the coherent product in this definition.

$\mathbb{Z}$ and $\mathbb{Q}$ with their discrete topologies are both compactly coherent rings, as are $\mathbb{R}$ and $\mathbb{C}$ with their usual topologies.

Example 11.29. Let R be a ring and let $n \in \mathbb{N}$. The *ring of polynomials in n commuting variables* over R is the pseudoclass $R[x_1, \ldots, x_n] \subseteq \mathrm{fn}(\mathbb{Z}_+^n, R)$ of all functions from $\mathbb{Z}_+^n$ to R which are zero on all but finitely many points of $\mathbb{Z}_+^n$. We write $\sum a_{i_1,\ldots,i_n} x_1^{i_1} \cdots x_n^{i_n}$ for the function which takes the value $a_{i_1,\ldots,i_n}$ on the n-tuple $(i_1, \ldots, i_n)$. This notation makes it clear what the sum and product operations are; the fact that they are ring operations is a straightforward exercise. The *degree* of $\sum a_{i_1,\ldots,i_n} x_1^{i_1} \cdots x_n^{i_n}$ is the maximum value of $i_1 + \cdots + i_n$ for which $a_{i_1,\ldots,i_n} \neq 0$.

Let $P_k \subset R[x_1, \ldots, x_n]$ be the subclass consisting of all polynomials of degree at most k. This class is in a natural bijection with R^m where m is the number of monomials in n variables of degree at most k. ("Monomial" in the sense of "product of variables with repetition".) Thus if R is compactly coherent then P_k carries a natural compactly coherent topology for each k (Proposition 10.95), and we can then give $R[x_1, \ldots, x_n]$ the coherent topology relative to the filtration $P_1 \subseteq P_2 \subseteq \cdots$. By Proposition 10.88, $R[x_1, \ldots, x_n]$ is compactly coherent.

This construction is one place where we really need to generalize "locally compact" to "compactly coherent". For instance, the topology on $\mathbb{R}[x_1, \ldots, x_n]$ is compactly coherent but not locally compact.

$R[x_1, \ldots, x_n]$ does not literally contain R, but it does contain a class of "constant polynomials", which, formally, are functions which vanish on every element of $\mathbb{Z}_+^n$ besides $(0, \ldots, 0)$. We usually use the same symbol for an element of R and the corresponding constant polynomial, a minor abuse of notation.

A *homomorphism* from one ring R to another R' is a function $\phi : R \to R'$ which satisfies $\phi(a+b) = \phi(a) + \phi(b)$ and $\phi(ab) = \phi(a)\phi(b)$ for all $a, b \in R$. If R and R' are both unital and ϕ takes the identity of R to the identity of R', then ϕ is a *unital* homomorphism. A bijective homomorphism is an *isomorphism*, and two rings are *isomorphic* if there is an isomorphism between them.

Proposition 11.30. *Fix $n \in \mathbb{N}$ and let $\phi : R \to R'$ be a homomorphism of commutative rings. For any $a_1, \ldots, a_n \in R'$ there is a unique homomorphism $\Phi : R[x_1, \ldots, x_n] \to R'$ satisfying $\Phi(a) = \phi(a)$ for all $a \in R$ and $\Phi(x_i) = a_i$ for $i = 1, \ldots, n$. If R and R' are compactly coherent and ϕ is continuous then Φ is also continuous.*

In particular, taking ϕ to be the identity map from R to itself, we can evaluate polynomials in $R[x_1, \ldots, x_n]$ on elements of R.

Definition 11.31. Let R_i ($1 \leq i \leq n$ or $i \in \mathbb{N}$) be rings. Then we define an additive operation on $\prod R_i$ by putting $\vec{a} + \vec{b} = (a_1 + b_1, \ldots, a_n + b_n)$ or $(a_i + b_i)$. This is a function because for any $\vec{a}, \vec{b}, \vec{c} \in \prod R_i$ it definitely is or is not the case, for each i, that $a_i + b_i = c_i$, and then by the LPO it definitely is or is not the case that $\vec{a} + \vec{b} = \vec{c}$ (without the need for the LPO if it is a finite product). We define multiplication similarly, coordinatewise. These are easily seen to be ring operations on $\prod R_i$. If the R_i's are compactly coherent and the product is finite, then we equip it with the coherent product topology; if they are compact then we equip it with the standard product topology.

Like product groups (Theorem 11.8), product rings have a universal property.

Theorem 11.32. *(Universal property of product rings) Let R and R_i ($1 \leq i \leq n$ or $i \in \mathbb{N}$) be rings. For any homomorphisms $\phi_i : R \to R_i$, there is a unique homomorphism $\phi : R \to \prod R_i$ satisfying $\phi_i = \pi_i \circ \phi$ for all i. If R*

and the R_i are compactly coherent and the product is finite, or R and the R_i are all compact, then ϕ is continuous if and only if each ϕ_i is continuous.

A *subring* of a ring R is a subclass R' which contains 0 and is stable under sums, products, and additive inverses. If R is unital then a *unital subring* is a subring which contains the identity element of R. A subring of R is an *ideal* if it contains ab and ba whenever it contains a, for all $b \in R$, and a *proper* ideal if it is properly contained in R. When R is unital, an ideal is proper if and only if it does not contain the identity element.

Let R be a commutative compactly coherent ring. For any $a \in R$, the map $b \mapsto ab$ is continuous, and therefore the image of R under this map exists as a subclass of R (Proposition 10.86 (f)). One straightforwardly checks that this image is an ideal. We denote it aR or (a), and call it a *principal ideal*. An element belongs to aR if and only if it has the form ab for some $b \in R$. Principal ideals are not necessarily closed, but in the rings in which we are interested they typically will be.

For any ideal I of a ring R, define an equivalence relation on R by putting $a \sim b$ if $a - b \in I$. It is easy to see that this is indeed an equivalence relation; in particular, it is a subclass of $R \times R$ because I is a subclass of R and therefore the condition "$a - b \in I$" is definite. We write R/I for the quotient $R/\sim$, and $[a]$ or $[a]_I$ for the element of R/I instantiated by a. If R is compactly coherent and I is closed, then we topologize R/I by treating R as an additive group and applying the group quotient construction.

The *kernel* of a ring homomorphism $\phi : R \to R'$ is the preimage $\ker(\phi) = \phi^{-1}(0) \subseteq R$ of the additive identity of R'.

The next series of results are proven in the same way as the corresponding results for groups.

Proposition 11.33. *Let R be a compactly coherent ring and let I be a closed ideal of R. Then there are ring operations on R/I with which the quotient topology is compatible, and which makes the quotient map $\pi : R \to R/I$ a continuous homomorphism whose kernel is I.*

R/I is a *quotient ring*. The key point here is that when we define $[a][b]$ to be $[ab]$, this works because if $[a] = [a']$ and $[b] = [b']$ then $a - a', b - b' \in I$ and therefore $ab - a'b' = a(b - b') + (a - a')b' \in I$.

Example 11.34. $\mathbb{Z}/n\mathbb{Z}$, the quotient of the ring $\mathbb{Z}$ by the ideal $n\mathbb{Z}$, is a ring for any $n \in \mathbb{N}$.

Given two closed subrings $R_1, R_2 \subseteq R$, if either one is an ideal then

$R_1 + R_2$ (which exists by Proposition 11.10) is a subring of R. For instance, if R_1 is an ideal then $(a_1+b_1)(a_2+b_2) = (a_1a_2+a_1b_2+b_1a_2)+b_1b_2 \in R_1+R_2$ for any $a_1, a_2 \in R_1$ and $b_1, b_2 \in R_2$.

Theorem 11.35. *(Universal property of quotient rings) Let $\phi : R \to R'$ be a continuous homomorphism between compactly coherent rings and let I be a closed ideal of R that is contained in $\ker(\phi)$. Then there is a continuous homomorphism $\psi : R/I \to R'$ such that $\phi = \psi \circ \pi$.*

Theorem 11.36. *(First isomorphism theorem for rings) Let $\phi : R \to R'$ be a continuous homomorphism of compactly coherent rings. Then $\ker(\phi)$ is a closed ideal of R, $\phi(R)$ is a subring of R', and there is a continuous isomorphism $\psi : R/\ker(\phi) \cong \phi(R)$ such that $\phi = \psi \circ \pi$.*

Theorem 11.37. *(Second isomorphism theorem for rings) Let I and R' be closed subrings of a compactly coherent ring R. Assume I is an ideal. Then $R' \cap I$ is a closed ideal of R', I is a closed ideal of $R' + I$, and there is a continuous isomorphism from $R'/(R' \cap I)$ to $(R' + I)/I \subseteq R/I$.*

Theorem 11.38. *(Third isomorphism theorem for rings) Let I and R' be closed subrings of a compactly coherent ring R, and assume that I is an ideal and $I \subseteq R'$. Then R'/I is a closed subring of R/I, and every closed subring of R/I has this form. Moreover, R' is an ideal of R if and only if R'/I is an ideal of R/I, in which case $(R/I)/(R'/I)$ is homeomorphically isomorphic to R/R'.*

Now let us discuss fields. A *field* is a nonzero unital commutative ring in which every nonzero element a is *invertible*, i.e., there exists b such that $ab = 1$. Equivalently, it is a nonzero unital commutative ring R with the property that $R^* = R \setminus \{0\}$ is a multiplicative group. $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ are examples of fields. Another example is $\mathbb{Z}/p\mathbb{Z}$ for any prime $p \in \mathbb{N}$ (Proposition 11.24).

Here is another class of examples of fields. We continue to use the notation $F^* = F \setminus \{0\}$.

Example 11.39. Let F be a compactly coherent field. Place an equivalence relation on $F[x_1, \ldots, x_n] \times F[x_1, \ldots, x_n]^*$ by putting $(f_1, f_2) \sim (g_1, g_2)$ if $f_1 g_2 = f_2 g_1$. The quotient of the product by this equivalence relation is the *field of rational functions in n variables* over F, denoted $F(x_1, \ldots, x_n)$. The ring operations are given by $(f_1, f_2)+(g_1, g_2) = (f_1g_2+f_2g_1, f_2g_2)$ and $(f_1, f_2) \cdot (g_1, g_2) = (f_1g_1, f_2g_2)$; one has to check that these operations are compatible with $\sim$.

For the sake of notational simplicity I will only explain how to topologize $F(x)$. The same idea works for $F(x_1, \ldots, x_n)$ for any n.

First, give $F[x]$ the topology from Example 11.29, and give $F[x] \times F[x]^*$ the coherent product topology. Let $Z_1 \subseteq Z_2 \subseteq \cdots$ be a compact filtration of F which gives rise to its given compactly coherent topology. Then for each n let $Y_n \subseteq F[x] \times F[x]^*$ consist of those ordered pairs (f, g) with $\deg(f), \deg(g) \leq n$ and with each coefficient of f and g belonging to Z_n. Then let X_n be the class of ordered pairs (f, g) which are equivalent to something in Y_n; this is a class because Y_n is compact, and if $\mathring{Y}_n$ is a dense subset of Y_n then $(f, g) \in X_n$ if and only if there are sequences (f_k) and (g_k) in $\mathring{Y}_n$ such that $fg_k - gf_k \to 0$. (This convergence would have to take place in some Y_m, which are all metrizable.) Now, letting $\widetilde{X}_n$ be the image of X_n under the quotient map (Proposition 9.24) yields a filtration of $F[x] \times F[x]^*$. We let a subclass of $\widetilde{X}_n$ be open if it is the image under the quotient map of a relatively open class $U \subseteq X_n$; this image exists as a class because the ordered pairs which are equivalent to something in U constitute a class, for the same reason X_n exists as a class (since U is σ-compact). Then each $\widetilde{X}_n$ is compact by an argument similar to the one used in the proof of Theorem 11.13, and their topologies are compatible, so this filtration determines a compactly coherent topology on $F(x)$. Checking that the ring operations on $F(x)$ are continuous is routine.

The first isomorphism theorem for rings (Theorem 11.36) can also be used to construct fields from rings. A proper ideal I of a unital commutative ring R is *maximal* if the only ideal that properly contains it is R. When R is countable, maximal ideals exist in abundance:

Proposition 11.40. *(Existence of maximal ideals) Let I be a proper ideal of a countable unital commutative ring R. Then I is contained in a maximal ideal.*

Proof. If R is finite the result is trivial. Otherwise, enumerate the elements of R as $b_1, b_2, \ldots$. Then recursively construct ideals $I_1 \subseteq I_2 \subseteq \cdots$ as follows. Start by putting $I_1 = I$. Given I_n, let $I_{n+1} = I_n + b_n R$, provided this is not all of R; if it is, put $I_{n+1} = I_n$. For each n, I_{n+1} is the smallest proper ideal that contains I_n and b_n, if there are any. Then $I' = \bigcup I_n$ is an ideal, and since each I_n is proper none of them contains 1, so I' also does not contain 1 and is therefore proper. But for any b_n that is not contained in this union we know that no proper ideal contains both I_n and b_n, so we can also say that no proper ideal contains both I' and b_n. This shows that

I' is maximal. □

If I is a closed ideal, then it is a *maximal closed* ideal if the only closed ideal that properly contains it is R.

Proposition 11.41. *Let I be a maximal closed ideal of a compactly coherent unital commutative ring R. Assume all principal ideals of R are closed. Then R/I is a field. R is a field if and only if $\{0\}$ is a maximal closed ideal of R.*

Proof. The first assertion follows from the second because if I is maximal then the third isomorphism theorem for rings (Theorem 11.38) says that $\{0\}$ and R/I are the only closed ideals of R/I.

Suppose R is a field and let J be an ideal of R which contains a nonzero element a. Since R is a field there exists $b \in R$ with $ab = 1$, so $1 \in J$ and therefore $J = R$. Thus R is the only ideal that properly contains the zero ideal.

Conversely, suppose $\{0\}$ is a maximal closed ideal. Let $a \in R$ be nonzero; we must find $b \in R$ such that $ab = 1$. But aR is a nonzero closed ideal of R, so it must equal R, showing that indeed $ab = 1$ for some $b \in R$. So a has a multiplicative inverse, and thus R is a field. □

11.4 Integral domains

A *zero divisor* is a nonzero element a of a ring for which there exists a nonzero element b with $ab = 0$. An *integral domain* is a unital commutative ring that has no zero divisors. In this section we will briefly study three special kinds of integral domains, and then extract some number theoretic consequences. In any integral domain, say that a is *prime* if whenever a divides bc it divides either b or c. Say that a is *irreducible* if it is not invertible and whenever it is expressed as a product $a = bc$ one of the factors is invertible. Say it is *reducible* if there is a factorization $a = bc$ for which neither of the factors is invertible. And say that a and b are *associates* if they differ by an invertible factor, i.e., $a = ub$ with u invertible.

Definition 11.42. Let R be a compactly coherent integral domain.

(1) R is a *Euclidean doman* if there are functions $\phi : R^* \to \mathbb{Z}_+$ and $q : R \times R^* \to R$ satisfying (a) for each fixed $b \in R^*$ the function $q(\cdot, b)$ is continuous from R to itself; (b) if b divides $a \neq 0$ then $\phi(b) \leq \phi(a)$;

and (c) for any $a, b \in R$, $b \neq 0$, we have $r(a,b) = 0$ or $\phi(r(a,b)) < \phi(b)$, where $r(a,b) = a - bq(a,b)$. We say that ϕ is a *Euclidean function*, q is a *quotient function*, and r is a *remainder function*.

(2) R is a *principal ideal domain* (PID) if every separable ideal is closed and principal.

(3) R is a *unique factorization domain* (UFD) if (a) every element that is neither zero nor invertible can be written as a product of irreducibles and (b) this factorization is unique up to rearranging the factors and replacing them with associates.

Note that in a Euclidean domain, continuity in a of the quotient function q implies continuity in a of the remainder function r. Also note that any principal ideal aR in a commutative compactly coherent ring R is separable, since it is the image of R under the continuous map $x \mapsto ax$. Thus in a PID every principal ideal is closed. Indeed, in a PID each of the three conditions (separable, closed, principal) implies the other two.

Let us begin by collecting some basic examples of Euclidean domains.

Example 11.43. Let $a, b \in \mathbb{Z}^*$. If b divides a then $|b| \leq |a|$, and otherwise dividing a by b yields $a = bq(a,b) + r$ where the remainder r satisfies $|r| < |b|$. Thus the absolute value function on $\mathbb{Z}^*$ is a Euclidean function. All the relevant functions are continuous because the topology is discrete.

Recall that $F[x]$ is the ring of polynomials over F in one variable, and that if F is compactly coherent then so is $F[x]$ (Example 11.29).

Proposition 11.44. *If F is a compactly coherent field then $F[x]$ is a Euclidean domain.*

Proof. For any nonzero polynomial $f \in F[x]$ let $\deg(f)$ be its degree, the highest power of x that appears in f with nonzero coefficient. We verify that this is a Euclidean function on the polynomial ring $F[x]$. First, if g divides f with $f \neq 0$ then clearly $\deg(g) \leq \deg(f)$. Next, we define the quotient function by the usual method of long division of polynomials. This gives us functions q and r satisfying $f = gq(f,g) + r(f,g)$, with $r(f,g) = 0$ or $\deg(r(f,g)) < \deg(g)$. We must show that q is continuous in f.

Fix $g \in F[x]$ nonzero and suppose $f_n \to f$. By Proposition 10.86 (c) the degrees of the f_n's are bounded above, say by $N > 0$, and there is a compact subclass $K \subseteq F$ such that all the coefficients of every f_n belong to K. We prove that $q(f_n, g) \to q(f,g)$ by induction on N. By passing to a subsequence we can assume that all the f_n's have the same degree.

If $N = 0$ then every f_n is a constant polynomial, and either $\deg(g) = 0$ so $q(f_n, g) = \frac{f_n}{g} \to \frac{f}{g} = q(f, g)$, or $\deg(g) > 0$ so $q(f_n, g) = q(f, g) = 0$ for all n. This settles the base case.

Now let $N > 0$ and suppose $q(\cdot, g)$ preserves the convergence of all sequences of polynomials whose degrees are all strictly less than N. Let (f_n) be a sequence of polynomials of degree N and suppose it converges to a limit f. If $\deg(g) > N$ then $q(f, g) = q(f_n, g) = 0$ for all n and we are done. Otherwise let $d = \deg(g)$. Then for each n there is a unique monomial $a_n x^{N-d}$ of degree $N - d$ for which $f_n - a_n x^{N-d} \cdot g$ has degree strictly less than N — namely, a_n must be the ratio of the leading coefficient of f_n by the leading coefficient of g. Also let ax^{N-d} be the monomial of degree $N - d$ for which $f - ax^{N-d}g$ has degree strictly less than N. Since the leading coefficient of f_n converges to the leading coefficient of f, we have $a_n \to a$ and therefore $a_n x^{N-d} g \to ax^{N-d} g$. Thus

$$f_n - a_n x^{N-d} g \to f - ax^{N-d} g,$$

where now the convergence is taking place among polynomials of degree $N - 1$. By the induction hypothesis, we get

$$q(f_n - a_n x^{N-d} g, g) \to q(f - ax^{N-d} g, g),$$

so

$$q(f_n) = a_n x^{N-d} + q(f_n - a_n x^{N-d} g, g) \to ax^{N-d} + q(f - ax^{N-d} g, g) = q(f, g).$$

This completes the proof that q is continuous. □

We can use this result to establish another way of constructing fields, but first let us consider the relationship between the three properties presented in Definition 11.42.

Lemma 11.45. *In a Euclidean domain R, any two nonzero elements a and b have a greatest common divisor d (i.e., d divides both a and b, and any d' that divides both a and b also divides d). We can write $d = ax + by$ for some $x, y \in R$.*

Proof. Put $a_1 = a$ and $a_2 = b$. Given a_i and a_{i+1}, write $a_i = q_i a_{i+1} + a_{i+2}$ with $a_{i+2} = 0$ or $\phi(a_{i+2}) < \phi(a_{i+1})$. Since $\phi(a_i)$ decreases at each step, we must eventually reach a point where $a_n = 0$. Then $a_{n-1} | a_{n-2}$, and so $a_{n-1} | q_{n-3} a_{n-2} + a_{n-1} = a_{n-3}$, and so on, so that ultimately a_{n-1} divides both $a = a_1$ and $b = a_2$. By induction on i, every a_i can be put in the form $ax + by$ for some $x, y \in R$, so this is true of a_{n-1}. Thus anything that divides both a and b must also divide a_{n-1}, showing that a_{n-1} is a greatest common divisor of a and b. □

Theorem 11.46. *Every Euclidean domain is a PID.*

Proof. In a Euclidean domain R every principal ideal is closed. To see this, let $b \in R^*$ and let (c_n) be a sequence in R such that (bc_n) converges, say to a. By continuity of the quotient function for division by b, we get $q(bc_n, b) \to q(a, b)$, but $q(bc_n, b) = c_n$ so $bc_n \to bq(a, b)$. This shows that $a = bq(a, b) \in bR$, and we conclude that bR is sequentially closed. Since, as noted earlier, every principal ideal in a commutative compactly coherent ring is separable, it follows from Proposition 10.91 that bR is closed.

Also, every separable ideal is principal. Let I be a separable ideal with dense subset $\mathring{I}$. Write $\mathring{I} = \{a_1, a_2, \ldots\}$, define $b_1 = a_1$, and successively let b_{i+1} be a greatest common divisor of b_i and a_{i+1}. Note that b_i belongs to I for all i by the lemma. Since $\phi(b_{i+1}) \leq \phi(b_i)$ for all i, we must eventually reach an index n such that $\phi(b_n) = \phi(b_{n+1}) = \phi(b_{n+2}) = \cdots$. But this means that $b_n | b_{n+1}$, as otherwise we could write $b_{n+1} = qb_n + r$ where $r \neq 0$ and $\phi(r) < \phi(b_n) = \phi(b_{n+1})$, which is impossible since $b_{n+1} | b_n$ implies $b_{n+1} | r$. Similarly b_n divides b_{n+2}, b_{n+3}, etc. We conclude that b_n divides every a_i, so that every a_i belongs to the ideal $b_n R$, and since we just showed that principal ideals are closed, this implies that $I \subseteq b_n R$ and therefore $I = b_n R$. We have shown that I is principal. □

Lemma 11.47. *In any compactly coherent integral domain in which every principal ideal is closed, every nonzero element definitely is or is not reducible.*

Proof. Let R be a compactly coherent integral domain in which every principal ideal is closed and let $a \in R$ be nonzero. Let $X_1 \subseteq X_2 \subseteq \cdots$ be a compact filtration of R and let $\mathring{R}$ be a coherently dense subset of R relative to this filtration. Fix a metric on each X_k. Suppose a were reducible, i.e., $a = bc$ with neither $a | b$ nor $a | c$. Then we could find sequences (x_n) and (y_n) in $\mathring{R}$ which converge to b and c, so that $x_n y_n \to a$. By Proposition 10.86 (c) there must exist k such that every x_n, y_n, and $x_n y_n$ belong to X_k. Since $R_k = aR \cap X_k$ is closed and therefore separable, there is a distance function $d_{R_k}(x) = \inf\{d(x, y) : y \in R_k\}$ defined on X_k. Now neither b nor c belongs to aR, so without loss of generality we can suppose that there exists $m \in \mathbb{N}$ such that $d_{R_k}(x_n), d_{R_k}(y_n) \geq \frac{1}{m}$ for all n. Thus: if a is reducible then there exist $k, m \in \mathbb{N}$ and elements $x, y \in \mathring{R} \cap X_k$ with $d_{R_k}(x), d_{R_k}(y) \geq \frac{1}{m}$ and $d(xy, a)$ arbitrarily small.

Conversely, suppose there exist k and m with this property. Then, using compactness, we can find convergent sequences $x_n \to x$ and $y_n \to y$ with

each x_n and y_n in $\mathring{R} \cap X_k$, and $d(x_n y_n, a) \to 0$, and $d_{R_k}(x_n), d_{R_k}(y_n) \geq \frac{1}{m}$ for all n. Thus $xy = a$ and $x, y \notin aR$, showing that a is reducible. We conclude that the condition stated above ("there exist $k, m \in \mathbb{N} \ldots$") is equivalent to the statement that a is reducible, and this condition is definite, so a definitely is or is not reducible. □

Lemma 11.48. *In a PID, every irreducible element is prime.*

Proof. Let p be an irreducible element of a PID R and suppose $p|ab$. We must show that $p|a$ or $p|b$. Consider the ideal $pR + aR$. (This exists by Proposition 11.10. Recall that in a PID every principal ideal is closed.) Since R is separable, so is $pR + aR$, and since R is a PID this ideal must be principal, say $pR + aR = cR$. Then $c|p$ and $c|a$. Since p is irreducible, this means that c must be either invertible or an associate of p. If it is an associate of p then $c|a$ implies $p|a$, and if it is invertible then $1 \in cR$ (in fact, this is equivalent to saying that c is invertible, which shows that invertibility of c is definite), so we have $px + ay = 1$ for some $x, y \in R$. Then $pbx + aby = b$, and since p divides the left side it must also divide the right side. □

Theorem 11.49. *Every PID is a UFD.*

Proof. Let R be a PID and suppose $a \in R$ is neither zero nor invertible. Form a tree tagged with elements of R as follows. Place a at the bottommost vertex. If a is reducible, write it as a product of two elements neither of which is invertible, and place them at the second level of the tree, as sucessors of a. Continue in this way, putting two vertices above each reducible element and labelling them with noninvertible factors. Vertices which are tagged with irreducible elements have no successors. By Lemma 11.47, every element of R definitely is either reducible or irreducible.

By König's infinity lemma (Proposition 9.41), if this tree contains infinitely many vertices there is an infinite path up it. That means we have a sequence of noninvertible elements $a = a_1, a_2, \ldots$ such that a_n can be written as a product of a_{n+1} and another noninvertible element, for each n. But then $\bigcup_n a_n R$ would be a nonprincipal ideal. If it were principal, generated by a', then a' would have to belong to $a_n R$ for some n and this would mean that a_n must divide a_{n+1}, which is impossible. But $\bigcup_n a_n R$ is separable, so we conclude that the tree can contain only finitely many vertices. Thus there are finitely many irreducible elements of R labelling the terminal, successorless vertices, and their product is a. This verifies the first condition to be a UFD.

For the second condition, suppose some element is expressed as a product of irreducibles in two different ways. By Lemma 11.48 each irreducible in the first list must divide, and hence be an associate of, some irreducble in the other list. Removing such a pair from the two lists, we make the same argument to find another pair, and so on. This shows that the two factorizations of a are the same up to reordering and replacing the factors with associates. This verifies the second condition to be a UFD. □

Now, as promised, here is another way of constructing fields, as quotients of polynomial rings.

Proposition 11.50. *Let F be a compactly coherent field and let $p \in F[x]$ be an irreducible polynomial. Then $F[x]/(p)$ is a compactly coherent field and there is a natural embedding ι of F in $F[x]/(p)$.*

Proof. We need every nonzero element of $F[x]/(p)$ to be invertible. So let $q \in F[x]$, $q \notin (p)$; we must show that $[q]$ is invertible in $F[x]/(p)$. We prove this by induction on the degree of q. We can assume that $\deg(q) < \deg(p)$, as Proposition 11.44 assures us every polynomial is something in (p) plus a polynomial of lower degree than p. So all elements of $F[x]/(p)$ are instantated by polynomials of lower degree than p.

If $q \neq 0$ has degree 0 then it is already invertible in F. Now suppose q has degree > 0 and every nonzero element of $F[x]/(p)$ instantiated by a polynomial of lower degree is invertible. Write $p = qr + s$ where $r, s \in F[x]$ and s is either zero or has lower degree than q. If s is zero then $p = qr$, and since we are assuming $q \notin (p)$, r cannot be invertible. But also q is not invertible in $F[x]$ because its degree is not zero. This contradicts irreducibility of p, so we conclude that s is not zero. Since s is a polynomial of lower degree than q, the induction hypothesis yields that $[s]$ is invertible in $F[x]/(p)$. Thus means that $[q] \cdot [r] = [qr] = [p - s] = -[s]$ is invertible, showing that $[q]$ is invertible. By induction, we conclude that $[q]$ is invertible in $F[x]/(p)$ for every $q \notin (p)$.

Regarding $F[x]$ as an abelian group, (p) is a closed subgroup by Proposition 11.44 and Theorem 11.46, so the quotient space $F[x]/(p)$ has a quotient topology which is compactly coherent (Theorem 11.13) and compatible with the ring structure (Proposition 11.33). So $F[x]/(p)$ is a compactly coherent field.

The map ι is defined by first taking every element of F to the corresponding constant polynomial in $F[x]$, then applying the quotient map from $F[x]$ to $F[x]/(p)$. This composition is clearly a homomorphism, and

it is injective because any unital homomorphism of fields is injective. □

In fact $F[x]/(p)$ is homeomorphic to the coherent power F^n where $n = \deg(p)$; see Proposition 12.26 in the next chapter.

Now I will illustrate the number-theoretic use of some of this basic ring theory in a series of results culminating in Fermat's "sum of two squares" theorem.

Proposition 11.51. *Let F be a finite field. Then F^* is a cyclic group under multiplication.*

Proof. F^* is clearly a multiplicative group; the point is that it is cyclic, i.e., it is isomorphic to $\mathbb{Z}/n\mathbb{Z}$ for some $n \in \mathbb{N}$. Let $n = |F| - 1$ be the order of F^*. Thus the order of any element of F^* divides n. (This was a comment after Lagrange's theorem, Theorem 11.14.) For each divisor d of n let $\psi(d)$ be the number of elements of order d. If, for a given d, there is any such element a, then $\{1, a, \ldots, a^{d-1}\}$ is a subgroup of F^* of order d, and every element b of this subgroup satisfies $b^d = 1$ (since the order of b divides d, the order of the subgroup). But Proposition 11.44 plus Theorems 11.46 and 11.49 show that $F[x]$ has unique factorization, so there can be at most d distinct roots of the polynomial $x^d - 1$; as 1, a, …, a^{d-1} are all roots of this equation, there must not be any others.

The elements of F^* whose order is exactly d must therefore all be powers of a; specifically, they are the powers a^k such that $0 \leq k < d$ is coprime to d. So there are exactly $\phi(d)$ such elements, where ϕ is Euler's totient function defined just before Euler's theorem/Fermat's little theorem (Theorem 11.25). Thus for each divisor d of n, either $\psi(d) = 0$ or $\psi(d) = \phi(d)$. But $\sum_{d|n} \psi(d) = n = \sum_{d|n} \phi(d)$ (Proposition 11.26), so we must have $\psi(d) = \phi(d)$ for all $d|n$. In particular, $\psi(n) = \phi(n) > 0$, so there is at least one element, in fact $\phi(n)$ many elements, of F^* of order n. This shows that $F^* \cong \mathbb{Z}/n\mathbb{Z}$. □

Corollary 11.52. *(Wilson's theorem) The number $(p-1)! + 1$ is divisible by p, for any prime number p.*

Proof. The result is trivial if $p = 2$, so we can assume p is odd.

Let $F = \mathbb{Z}/p\mathbb{Z}$. According to Propositions 11.24 and 11.51, F^* is a cyclic group under multiplication; its order is $p-1$. Let $[a]$ be an element of order $p-1$. Then $\{[a], [a]^2, \ldots, [a]^{p-2}\} = \{[2], [3], \ldots, [p-1]\}$ — in both cases the missing element of F^* is $[1]$. For any $1 \leq k \leq p-2$ we have $[a]^k \cdot [a]^{p-1-k} = [a]^{p-1} = [1]$, which shows that $[a]^k$ is distinct from its multiplicative inverse

for every value of k except $k = \frac{p-1}{2}$. That is, $[a]^{(p-1)/2}$ is the one element of F^* besides $[1]$ whose square is $[1]$, and therefore it must be $[-1] = [p-1]$. Thus, in F^*, all the terms of the product $[2]\cdots[p-1]$ cancel each other except for $[p-1]$, so $[(p-1)!] = [2]\cdots[p-1] = [p-1] = [-1]$. We have shown that $(p-1)!+1$ is divisible by p. □

The proof of Fermat's sum of two squares theorem uses the *Gaussian integers* $\mathbb{Z}[i] = \{a+bi : a,b \in \mathbb{Z}\}$. This is clearly a subring of $\mathbb{C}$ and an integral domain. Moreover, it is a Euclidean domain.

Proposition 11.53. *$\mathbb{Z}[i]$ is a Euclidean domain.*

Proof. Define the "norm" of $z = a+bi \in \mathbb{Z}[i] \subset \mathbb{C}$ to be the number $N(z) = |z|^2 = a^2+b^2$. We will show that this is a Euclidean function on $\mathbb{Z}[i]$. First, we must prove that if z divides $w \neq 0$ then $N(z) \leq N(w)$. This works because if $w = zu$ then $N(w) = |w|^2 = |zu|^2 = |z|^2|u|^2 = N(z)N(u)$, and $N(u) \geq 1$ for all nonzero $u \in \mathbb{Z}[i]$. The key identity $|zu|^2 = |z|^2|u|^2$ is easily verified by a direct calculation. Second, we must show that for any z and w with $z \neq 0$, there exist q and r such that $w = zq+r$ and r is either zero or else $N(r) < N(z)$. To do this, let $x+iy = \frac{w}{z} \in \mathbb{C}$. Here x and y need not be integers. However, there exist integers m and n such that $m-\frac{1}{2} \leq x \leq m+\frac{1}{2}$ and $n-\frac{1}{2} \leq y \leq n+\frac{1}{2}$. Thus $|(x-m)+i(y-n)|^2 = (x-m)^2+(y-n)^2 \leq \frac{1}{2}$. Now put $q = m+in$ and $r = z((x-m)+i(y-n)) = w-zq$; then $w = zq+r$ and $N(r) \leq \frac{1}{2}N(z)$ (using multiplicativity of $|\cdot|^2$). Continuity of the quotient function is not an issue because the topology is discrete and all functions are continuous. This completes the proof. □

Lemma 11.54. *Let p be a prime number of the form $p = 4n+1$ and put $x = \left(\frac{p-1}{2}\right)!$. Then x^2+1 is divisible by p.*

Proof. Since $p = 4n+1$, $\frac{p-1}{2} = 2n$ is even, so we also have $x = (-1)(-2)(-3)\cdots(-\frac{p-1}{2})$. (There are an even number of terms in this product.) Working in $\mathbb{Z}/p\mathbb{Z}$, we then have $[x] = [1][2]\cdots\left[\frac{p-1}{2}\right] = [p-1][p-2]\cdots\left[\frac{p+1}{2}\right]$ and so

$$[x]^2 = [1]\cdot[2]\cdots\left[\frac{p-1}{2}\right]\cdot\left[\frac{p+1}{2}\right]\cdots[p-1] = [(p-1)!].$$

Wilson's theorem (Corollary 11.52) yields that x^2+1 is divisible by p. □

Theorem 11.55. *(Fermat's sum of two squares theorem) Let p be a prime number of the form $p = 4n+1$. Then there exist $a,b \in \mathbb{N}$ with $a^2+b^2 = p$.*

Proof. By Proposition 11.53 and Theorems 11.46 and 11.49, $\mathbb{Z}[i]$ is a PID and a UFD. Find $x \in \mathbb{N}$ such that $x^2 + 1$ is divisible by p (Lemma 11.54); then $x^2 + 1$ factors as $(x + i)(x - i)$ in $\mathbb{Z}[i]$, but p does not divide either factor. (Any multiple of p in $\mathbb{Z}[i]$ has real and imaginary parts which are both divisible by p.) Therefore p is not prime in $\mathbb{Z}[i]$, and so it must factor (Lemma 11.48). But $N(p) = p^2$, and the only elements of $\mathbb{Z}[i]$ with norm 1 are the invertibles (namely, 1, i, -1, $-i$), so any nontrivial factorization of p has to have exactly two factors, each of which has norm p. But $N(a+bi) = p$ means $p = a^2 + b^2$. □

11.5 Boolean algebras

Definition 11.56. Let B be a class. *Lattice operations* on B are functions $\vee, \wedge : B \times B \to B$ ("join" and "meet"), satisfying

(1) (commutativity) $a \vee b = b \vee a$ and $a \wedge b = b \wedge a$
(2) (associativity) $a \vee (b \vee c) = (a \vee b) \vee c$ and $a \wedge (b \wedge c) = (a \wedge b) \wedge c$
(3) (idempotence) $a \vee a = a \wedge a = a$
(4) (absorption) $a \vee (a \wedge b) = a$ and $a \wedge (a \vee b) = a$

for all $a, b, c \in B$. A class equipped with lattice operations is a *lattice*. It is a *Boolean algebra* if we also require

(5) (distributivity) $a \vee (b \wedge c) = (a \vee b) \wedge (a \vee c)$ and $a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c)$ for all $a, b, c \in B$
(6) (0 and 1) there exist elements 0 and 1 such that $a \vee 0 = a \wedge 1 = a$ for all $a \in B$
(7) (complementation) for each $a \in B$ there exists a *complementary* element $\neg a \in B$ such that $a \vee \neg a = 1$ and $a \wedge \neg a = 0$.

With the exception of Example 11.57 below, we will only work with countable Boolean algebras, so there is no need to equip them with topologies (though Proposition 11.60 below shows how that would be done, if we wanted to).

Complements are unique: if $a \vee b = a \vee c = 1$ and $a \wedge b = a \wedge c = 0$, then $b = c$. This can be seen by the calculation

$$c = c \wedge 1 = c \wedge (a \vee b) = (c \wedge a) \vee (c \wedge b) = c \wedge b$$

plus the same calculation with b and c interchanged, which shows that $b = c \wedge b$ also. This implies that the complement of the complement of a is a, i.e., $\neg\neg a = a$.

We can use the uniqueness of complements to prove the de Morgan laws from set theory and logic: $\neg(a \vee b) = \neg a \wedge \neg b$ and $\neg(a \wedge b) = \neg a \vee \neg b$. For instance, to prove the first equality it will suffice to show that $(\neg a \wedge \neg b) \wedge (a \vee b) = 0$ and $(\neg a \wedge \neg b) \vee (a \vee b) = 1$, because this establishes that $\neg a \wedge \neg b$ is the complement of $a \vee b$. But

$$(\neg a \wedge \neg b) \wedge (a \vee b) = (\neg a \wedge \neg b \wedge a) \vee (\neg a \wedge \neg b \wedge b) = 0 \vee 0 = 0$$

and

$$(\neg a \wedge \neg b) \vee (a \vee b) = (\neg a \vee a \vee b) \wedge (\neg b \vee a \vee b) = 1 \vee 1 = 1.$$

The resemblance to set theory is not an accident.

Example 11.57. If X is any set then its power class $\mathcal{P}(X)$ is a Boolean algebra with join and meet corresponding to union and intersection, 0 and 1 to $\emptyset$ and X, and $\neg A$ to the set-theoretic complement A^c.

If X is finite with n elements, then $\mathcal{P}(X)$ has 2^n elements; if X is infinite then $\mathcal{P}(X)$ is uncountable. There are also countably infinite Boolean algebras. Here are two examples.

Example 11.58. Let B be the subset of $\mathcal{P}(\mathbb{N})$ consisting of all the (codes of) sets which are either finite or cofinite (i.e., their complement is finite). It is easy to see that B is stable under the Boolean operations, so it is a Boolean algebra.

Example 11.59. Let B be the subset of $\mathcal{P}(\mathbb{N})$ consisting of all the (codes of) periodic subsets of $\mathbb{N}$, i.e., those for which the corresponding bitstring is periodic. This is also stable under the Boolean operations, so it is a Boolean algebra too.

This last example is countable because for any n there are exactly 2^n bitstrings that repeat every n elements, so B is a countable union of finite sets.

Example 11.59 is also notable in that this Boolean algebra has no *atoms*, nonzero elements a for which $a \wedge b$ equals a or 0, for every b. Every periodic subset A of $\mathbb{N}$ besides $\emptyset$ contains a strictly smaller non-empty periodic subset B, and we have $A \cap B = B$, which is neither $\emptyset$ nor A.

Lattices can be understood in order-theoretic terms. If B is a lattice, we can define a partial order on it by putting $a \leq b$ if $a \vee b = b$, or equivalently

$a \wedge b = a$. Relative to this partial order $a \vee b$ is the least upper bound of a and b, and $a \wedge b$ is their greatest lower bound. If B is a set, then this establishes an exact correspondence between lattice structures on B and partial orders on B for which every two elements have a least upper bound and a greatest lower bound. However, for uncountable classes the order-theoretic definition is a little weaker because we require $\vee$ and $\wedge$ to be functions, so that, for instance, the statement "$a \vee b = c$" must have a definite truth value for all a, b, and c. Conceivably we could be presented with a partial order for which we know that any two elements have a least upper bound and a greatest lower bound, but for which the actual values of those operations are indefinite.

Since every Boolean algebra is a lattice, we can think of Boolean algebras in order-theoretic terms too. There is also an equivalent approach to Boolean algebra via ring theory. A *Boolean ring* is a unital ring which satisfies $a^2 = a$ for all a. In such rings every element satisfies $a = -a$. This is because $1 + a = (1 + a)^2 = 1 + 2a + a$, so that $a + a = 2a = 0$. Boolean rings are also commutative, by the calculation $a + b = (a + b)^2 = a + ab + ba + b$ plus the fact that $ba = -ba$.

Proposition 11.60. *Let B be a Boolean algebra. Then $a + b = (a \wedge \neg b) \vee (\neg a \wedge b)$ and $a \cdot b = a \wedge b$ are ring operations on B which make it a Boolean ring in which 0 is the zero element and 1 is the identity element. Conversely, if R is a Boolean ring then $a \vee b = a + b + ab$ and $a \wedge b = a \cdot b$ are lattice operations on R which make it a Boolean algebra in which 0 is the least element, 1 is the greatest element, and the complement of a is $1 + a$. The two constructions are inverse to each other.*

The proof is routine. One benefit of this natural correspondence between Boolean algebras and Boolean rings is that it tells us what the right notions of subobject, homomorphism, etc. are. If B is a Boolean algebra then a *Boolean subalgebra* of B is a subclass which contains 0 and 1 and is stable under join, meet, and complementation. In the corresponding Boolean ring structure on B, this corresponds to a unital subring. A *homomorphism* $\phi : B \to B'$ between Boolean algebras is a function which takes 0 to 0 and 1 to 1 and respects join, meet, and complementation; this corresponds to a unital ring homomorphism. The *kernel* of ϕ is the subclass $\phi^{-1}(0)$. It is an *ideal* of B, an inhabited subclass I that satisfies (1) if $a \in I$ and $b \leq a$ then $b \in I$ and (2) if $a, b \in I$ then $a \vee b \in I$. Boolean algebra ideals correspond exactly to Boolean ring ideals, showing that every kernel is an ideal and every ideal is a kernel. A bijective homomorphism of Boolean algebras is

an *isomorphism*, and two Boolean algebras are *isomorphic* if there is an isomorphism between them. We define products and quotients of Boolean algebras in the obvious way; these too correspond to the analogous Boolean ring constructions.

Also, given the way the Boolean operations and the ring operations can be defined in terms of each other, a compactly coherent topology will be compatible with the Boolean algebra structure in the sense that $\vee$, $\wedge$, and $\neg$ are continuous if and only if it is compatible with the Boolean ring structure.

Finite Boolean algebras have a particularly simple structure.

Proposition 11.61. *Every finite Boolean algebra is isomorphic to $\mathcal{P}(X)$ for some finite set X.*

Proof. Given a finite Boolean algebra B, let X be the set of its atoms. We will show that B is isomorphic to $\mathcal{P}(X)$ via the map $\phi : B \to \mathcal{P}(X)$ which takes each $a \in B$ to (the code of) the set of atoms it lies above. This map clearly takes 0 to $\emptyset$ and 1 to X. It preserves meets because an atom m lies under $a \wedge b$ if and only if it lies under both a and b. It preserves joins because an atom m lies under $a \vee b$ if and only if it lies under a or b: here the reverse implication is clear, and the forward implication holds because if m does not lie under either a or b — and it must definitely either lie or not lie under a or b — then we must have $m \wedge a = m \wedge b = 0$, which implies $m \wedge (a \vee b) = 0$. Finally, ϕ preserves complements because every atom lies under either a or $\neg a$, but not both. If it lay under both then it would lie under their meet, which is zero, an impossiblity; if it lay under neither then, by what we just saw, it would not lie under their join, which is one, also an impossibility.

Injectivity follows from the fact that every nonzero element of B is the join of the atoms it lies above. To see this, let $a \in B$ be nonzero, let m_1, $\ldots$, m_k be the atoms that lie under a, and let $b = m_1 \vee \cdots \vee m_k$. Since $m_i \leq a$ for all i and b is the least upper bound of the m_i's, it follows that $b \leq a$. Now suppose $b < a$. Then $a \wedge \neg b > 0$ since

$$(a \wedge \neg b) \vee b = (a \vee b) \wedge (\neg b \vee b) = a \vee b = a,$$

whereas if $a \wedge \neg b$ were 0 then $(a \wedge \neg b) \vee b$ would equal b. So we can find an atom m lying under $a \wedge \neg b$. It cannot lie under b because it lies under $\neg b$. But it does lie under a and hence must be among the m_i's, a contradiction. We conclude that $a = b$, so we have shown that every nonzero element of B is the join of the atoms it lies above.

That shows that ϕ is injective. To show it is surjective, Let $m_1, \ldots, m_k$ be atoms and let a be their join; we must show that $\phi(a) = \{m_1, \ldots, m_k\}$. To see this, let m be an atom that lies under a. Then

$$m = m \wedge (m_1 \vee \cdots \vee m_k) = (m \wedge m_1) \vee \cdots \vee (m \wedge m_k),$$

showing that m must equal one of the m_i's. Thus $\{m_1, \ldots, m_k\}$ is precisely the set of all the atoms which lie under a. □

Thus, for each $n \in \mathbb{Z}_+$ there is, up to isomorphism, just one Boolean algebra with exactly 2^n elements, and there are no other finite Boolean algebras.

There is more than one countably infinite Boolean algebra; for instance, the Boolean algebras in Examples 11.58 and 11.59 are both countable but the first one contains atoms and the second does not. However, up to isomorphism there is only one countably infinite Boolean algebra which is atomless.

Lemma 11.62. *Let F be a finite subset of a Boolean algebra which is stable under complementation. Then every element of the Boolean subalgebra it generates can be written as a join of meets of its elements.*

Proof. Every element a of the Boolean algebra generated by F can be obtained from F by taking joins, meets, and complements; that is, if $F = \{a_1, \ldots, a_n\}$ then we have $a = \phi(a_1, \ldots, a_n)$ for some formula ϕ built up from the a_i's using $\vee$, $\wedge$, and $\neg$. We must find an equivalent formula which has the stated form.

First, we can eliminate every $\neg$ from ϕ. We prove this by induction on the length of ϕ, which we define to be the total number of $\wedge$, $\vee$, and $\neg$ symbols appearing in it. If $\phi = a_i$ is one of the elements of F then it is already free of complementation. If ϕ is $\phi_1 \vee \phi_2$ or $\phi_1 \wedge \phi_2$ then the induction hypothesis tells us that there are complementation-free formulas ψ_1 and ψ_2 which evaluate to the same elements as ϕ_1 and ϕ_2, so that $\psi_1 \vee \psi_2$ or $\psi_1 \wedge \psi_2$ evaluates to the same element as ϕ. If ϕ is the complement of some shorter formula, then we break up into cases. For $\phi = \neg a_i$ where a_i is one of the generators, we have assumed that the complement of a_i appears as another generator a_j, so we can replace ϕ with a_j. For $\phi = \neg\neg\psi$ we can replace ϕ with ψ and then, by the induction hypothesis, replace ψ with a complementation-free formula. For $\phi = \neg(\phi_1 \vee \phi_2)$ or $\neg(\phi_1 \wedge \phi_2)$, we can use the de Morgan laws to replace ϕ with $\neg\phi_1 \wedge \neg\phi_2$ or $\neg\phi_1 \vee \neg\phi_2$, then apply the induction hypothesis to $\neg\phi_1$ and $\neg\phi_2$. Thus, in every case we can eliminate every $\neg$ from ϕ.

We now have to deal with elements arising from formulas built up from the a_i's using $\vee$ and $\wedge$. We must prove that every such element can be written in the desired form, and we again prove this by induction on the length of ϕ.

If no $\vee$ or $\wedge$ appears in ϕ then it must consist of a single a_i and we are done. Now suppose at least one $\vee$ or $\wedge$ appears in ϕ and assume the result is true for all formulas with shorter length. If $\phi = \psi_1 \vee \psi_2$ then the induction hypothesis ensures that both ψ_1 and ψ_2 can be written as joins of meets of elements of F, so their join can also be written in this form. If $\phi = \psi_1 \wedge \psi_2$ then we can again write ψ_1 and ψ_2 as joins of meets of elements of F. If no $\vee$'s appear in either ψ_1 or ψ_2 then each of them is a meet of elements of F, and hence so is their meet with each other. Otherwise, without loss of generality suppose $\psi_1 = \sigma_1 \vee \cdots \vee \sigma_k$ where each σ_j is a meet of elements of F and $k \geq 2$. Then

$$\phi = \psi_1 \wedge \psi_2 = (\sigma_1 \vee \cdots \vee \sigma_k) \wedge \psi_2 = (\sigma_1 \wedge \psi_2) \vee \cdots \vee (\sigma_k \wedge \psi_2),$$

and by the induction hypothesis each $\sigma_i \wedge \psi_2$ can be put in the desired form, which again puts ϕ in the desired form. □

Theorem 11.63. *(Uniqueness of the countable atomless Boolean algebra) Any two countable, nonzero, atomless Boolean algebras are isomorphic.*

Proof. Let B and B' be countable, nonzero, atomless Boolean algebras. Every finite nonzero Boolean algebra has atoms, so B and B' must be infinite. Enumerate their elements as $B = \{a_1, a_2, \ldots\}$ and $B' = \{a'_1, a'_2, \ldots\}$. We will construct two sequences (B_i) and (B'_i) of finite Boolean subalgebras of B and B' together with isomorphisms $\phi_i : B_i \cong B'_i$, such that (1) $B_1 \subseteq B_2 \subseteq \cdots$ and $B'_1 \subseteq B'_2 \subseteq \cdots$; (2) $B = \bigcup B_i$ and $B' = \bigcup B'_i$; and (3) ϕ_{i+1} extends ϕ_i for all i. Then the union of the ϕ_i's will be the desired isomorphism between B and B'.

Assume a_1 and a_2 are the 0 and 1 of B, and a'_1 and a'_2 are the 0 and 1 of B'. Letting $B_1 = \{a_1, a_2\}$ and $B'_1 = \{a'_1, a'_2\}$, it is clear what ϕ_1 has to be. Now, at the nth stage of the construction, $n \geq 2$, if n is even then we look for the first a_i which does not appear in B_{n-1}, let B_n be the Boolean subalgebra of B generated by B_{n-1} and a_i, and try to enlarge B'_{n-1} in such a way that the isomorphism $\phi_{n-1} : B_{n-1} \cong B'_{n-1}$ can be extended to an isomorphism $\phi_n : B_n \cong B'_n$. If n is odd, we look for the first a'_i which does not appear in B'_{n-1}, and follow the same procedure in the opposite direction. If such a construction can be achieved, then it automatically satisfies conditions (1) and (3), and it satisfies (2) since every element of

B or B' will eventually be picked up. (Specifically, a_n and a'_n must belong to B_{2n} and B'_{2n}, respectively.) Each B_n and each B'_n is finite because the Boolean subalgebra generated by a finite set is always finite by Lemma 11.62.

The remaining question is how, at an even numbered step, after letting B_n be the Boolean algebra generated by B_{n-1} and one new element, to find B'_n in a way that lets us extend ϕ_{n-1}. (Obviously the same procedure can be used in the opposite direction at odd numbered steps.) We know from Proposition 11.61 that B_{n-1} is isomorphic to the power class of its atoms, and so is B_n. Every atom of B_n lies under at least one atom of B_{n-1}, since every atom is less than 1 and 1 is the join of the atoms of B_{n-1}. No atom of B_n can lie under more than one atom of B_{n-1}, since the meet of any two atoms of B_{n-1} is 0, and every atom of B_{n-1} belongs to B_{n-1} and hence lies above at least one atom of B_n. In a word, every atom of B_{n-1} either remains an atom of B_n or splits into more than one atom of B_n.

Let m be an atom of B_{n-1} that splits into atoms $m_1, \ldots, m_k$ of B_n (possibly with $k = 1$). Then $m' = \phi_{n-1}(m)$ is an atom of B'_{n-1}, and since B' is nonatomic we can find elements $m'_1, \ldots, m'_k \in B'$ whose join is m' and whose meet is 0. (Since $m' > 0$ we can find $0 < m'_1 < m'$, write $m' = m'_1 \vee (m' \wedge \neg m'_1)$, and then repeat $k-2$ more times.) We now have, for each atom m of B_{n-1}, a finite set of elements of B whose join is m and whose meet is 0, namely, the atoms of B_n that lie under it, and for each atom m' of B'_{n-1} a corresponding finite set of elements of B'. Letting X and X' be the unions of all these finite sets in B and B', respectively, we can fix a bijection between X and X' such that any element of X that lies under the atom m of B_{n-1} corresponds to an element of X' that lies under the corresponding atom of B'_{n-1}. We also know that B_n is naturally isomorphic to $\mathcal{P}(X)$; if we can show that $B'_n =$ the Boolean subalgebra of B' generated by X' is isomorphic to $\mathcal{P}(X')$ in a corresponding way, then we can define $\phi_n : B_n \cong B'_n$ by taking a join of elements of X to the corresponding join of elements of X'. This extends ϕ_{n-1} because each atom of B_{n-1} lies over a set of elements of X which correspond to the elements of X' which lie under the corresponding atom of B'_{n-1}.

The final claim is easy because the meet of any two elements of X' is 0, and the complement of any element of X' is the join of the others. By Lemma 11.62, every element of B'_n can be written as a join of meets of elements of X', and hence as a join of elements of X'. Since we already knew B'_n was isomorphic to a power class, this shows that it is naturally isomorphic to $\mathcal{P}(X')$. □

In arbitrary countable Boolean algebras we have an analog of the existence of maximal ideals in commutative rings (Proposition 11.40). In fact, using the correspondence between Boolean algebras and Boolean rings, we can easily deduce the next result from that proposition. (Or it can be proven independently in exactly the same way.)

As with rings, we call a maximal proper ideal of a Boolean algebra a *maximal ideal*. These are also known as *prime ideals*.

Theorem 11.64. *(Prime ideal theorem) Every proper ideal of a countable Boolean algebra is contained in a maximal ideal.*

In the case of Boolean algebras, maximal ideals have a very special form. In the next proof we use the simple fact that in any Boolean algebra, the ideal generated by (i.e., the smallest ideal containing) a set X of elements consists of precisely those elements of the Boolean algebra which lie below some join of finitely many elements of X.

Proposition 11.65. *An ideal of a countable Boolean algebra B is maximal if and only if for every $a \in B$ it contains one of a and $\neg a$, but not both.*

Proof. It is clear that any ideal with this property must be maximal. Conversely, if I is an ideal and there exists $a \in B$ such that neither a nor $\neg a$ belongs to I, then the ideal generated by I and a consists precisely of those elements of B which lie below something of the form $b \vee a$ with $b \in I$. This would not include $\neg a$, because if $\neg a \leq b \vee a$ for some $b \in I$, then

$$\neg a = \neg a \wedge (b \vee a) = (\neg a \wedge b) \vee (\neg a \wedge a) = \neg a \wedge b,$$

and this implies $\neg a \leq b$, which would force $\neg a \in I$. So I is contained in a strictly larger proper ideal, contradicting maximality. Since, by countability, there definitely does or does not exist $a \in B$ such that neither a nor $\neg a$ belongs to I, we conclude that there is no such a, so for every $a \in B$ either a or $\neg a$ belongs to I. □

This means that we can separate points with homomorphisms into a two-element Boolean algebra, in the following sense.

Corollary 11.66. *If a and b are elements of a countable Boolean algebra B and $a \not\leq b$ then there is a homomorphism from B into the two-element Boolean algebra $\{0, 1\}$ which takes a to 1 and b to 0.*

Proof. I claim that $b \vee \neg a \neq 1$. Because if $b \vee \neg a = 1$ then

$$b \vee a = (b \vee a) \wedge (b \vee \neg a) = b \vee (a \wedge \neg a) = b,$$

which would imply $a \leq b$, which we have assumed is not the case. (And the condition "$a \leq b$" is definite, so we can make a proof by contradiction.)

Thus the ideal generated by $b \vee \neg a$ is proper and contains both b and $\neg a$. By the prime ideal theorem (Theorem 11.64) there is then a maximal ideal I of B which contains both b and $\neg a$, and by Proposition 11.65 the quotient B/I contains exactly two elements. Thus the quotient map takes a to 1 and b to 0. □

The last topic in this chapter will be Stone duality. A topological space is *totally disconnected* if for any two distinct elements there is a clopen class that contains one but not the other. A compact space that is totally disconnected is a *Stone space*.

Any finite set with the discrete topology is a Stone space. The Cantor space $\{0,1\}^\omega$ is one basic example of an infinite Stone space. Every countable compact space is also an example. (If X is countable and compact then, fixing a metric on X, for all $x \in X$ the set $d_x(X) \subseteq \mathbb{R}$ is countable, where $d_x(y) = d(x,y)$, so there exist arbitrarily small values of $\epsilon > 0$ not in this set, and $B_\epsilon(x)$ is clopen for any such ϵ.)

Let A be a nonempty set. For every finite subset $A_0 \subseteq A$ and every function $f : A_0 \to \{0,1\}$, let $U_{f_0} \subseteq \mathrm{fn}(A, \{0,1\})$ be the pseudoclass of all (codes of) functions from A to $\{0,1\}$ whose restriction to A_0 equals f_0. These classes form the basis of a topology on $\mathrm{fn}(A, \{0,1\})$. If A is finite this is the discrete topology, and if A is infinite then it is homeomorphic to the Cantor space; in either case, we get a Stone space. According to this topology, a sequence of functions $\phi_i : A \to \{0,1\}$ converges to $\phi : A \to \{0,1\}$ if $\phi_i(x) \to \phi(x)$ for all $x \in A$, i.e., for each $x \in A$ we have $\phi_i(x) = \phi(x)$ for all but finitely many values of i. We call this the *topology of pointwise convergence*.

For any countable Boolean algebra B, let $B^* = \hom(B, \{0,1\}) \subseteq \mathrm{fn}(B, \{0,1\})$ be the pseudoclass of all (codes of) homomorphisms from B to $\{0,1\}$, with the topology of pointwise convergence it inherits from $\mathrm{fn}(B, \{0,1\})$. It is easy to see that any pointwise limit of a sequence of homomorphisms must itself be a homomorphism, so B^* is a closed subspace of $\mathrm{fn}(B, \{0,1\})$ and therefore also a Stone space.

Conversely, given any Stone space X we have a pseudoclass $X^* = C(X, \{0,1\})$ of all (codes of, but I am going to stop saying this now) continuous functions from X to $\{0,1\}$. The elements of X^* correspond to clopen subclasses of X, via the correspondence between a function and the inverse image of 1. The usual class-theoretic operations of union, intersection,

and relative complement then become operations on X^* which, one easily checks, makes it a Boolean algebra. This Boolean algebra is countable because we can fix a countable basis for the topology on X, and then any clopen class is the union of a sequence of basic open classes, and by Corollary 10.38 there is a finite subcover. Thus every clopen class is the union of finitely many basic open classes, and therefore there are only countably many of them.

So every countable Boolean algebra determines a Stone space and every Stone space determines a countable Boolean algebra. Also, any homomorphism $\phi : B_1 \to B_2$ between two countable Boolean algebras induces a continuous function $\phi^* : B_2^* \to B_1^*$ which takes $\psi \,\varepsilon\, \hom(B_2, \{0,1\})$ to $\psi \circ \phi$. And any continuous function $f : X_1 \to X_2$ between two Stone spaces induces a homomorphism $f^* : X_2^* \to X_1^*$ which takes $g \,\varepsilon\, C(X_2, \{0,1\})$ to $g \circ f$. The passage betwen countable Boolean algebras and Stone spaces, together with the passage between homomorphisms and continuous functions, is known as "Stone duality".

For any countable Boolean algebra B we have a natural map $\sigma : B \to B^{**}$ which takes $a \in B$ to the element of $B^{**} = C(B^*, \{0,1\})$ that maps $\phi \,\varepsilon\, B^*$ to $\phi(a)$. For any Stone space X we have a natural map $\tau : X \to X^{**}$ which takes $x \in X$ to the element of $X^{**} = \hom(X^*, \{0,1\})$ that maps $f \,\varepsilon\, X^*$ to $f(x)$.

Theorem 11.67. *(Stone duality) Let σ and τ be as above. Then*

(a) for every countable Boolean algebra B the map σ is an isomorphism between B and B^{**}, and for any homomorphism $\phi : B_1 \to B_2$ we have $\phi^{**} = \sigma \circ \phi \circ \sigma^{-1} : B_1^{**} \to B_2^{**}$

(b) for every Stone space X the map τ is a homeomorphism between X and X^{**}, and for any continuous function $f : X_1 \to X_2$ we have $f^{**} = \tau \circ f \circ \tau^{-1} : X_1^{**} \to X_2^{**}$.

Proof. For part (a), let B be a countable Boolean algebra. It is routine to check that σ is a homomorphism from B into B^{**}. It is injective because if $a, b \in B$ are distinct then according to Corollary 11.66 there is a homomorphism $\phi : B \to \{0,1\}$ with $\phi(a) \neq \phi(b)$. Then $\sigma(a)(\phi) = \phi(a) \neq \phi(b) = \sigma(b)(\phi)$, showing that $\sigma(a) \neq \sigma(b)$. To see that σ is surjective, let $f : B^* \to \{0,1\}$ be a continuous function; we must show that there exists $a \in B$ such that $f = \sigma(a)$, i.e., $f(\phi) = \phi(a)$ for all $\phi \,\varepsilon\, B^*$.

Let $A = f^{-1}(0)$. This is a clopen subclass of B^*. Also, for every $a \in B$

define A_a to be $\{\phi \,\varepsilon\, B^* : \phi(a) = 0\}$, so we have a family of clopen subclasses of B^* indexed by B, and consider $I = \{a \in B : A_a \subseteq A\}$. The condition "$A_a \subseteq A$" is definite by Corollary 10.42 and Proposition 10.56. I claim that for every $\phi \,\varepsilon\, A$ there exists $a \in I$ with $\phi \,\varepsilon\, A_a$. So fix $\phi \,\varepsilon\, A$. There definitely does or does not exist $a \in I$ with $\phi \,\varepsilon\, A_a$, so suppose there does not exist such an a. Then enumerate the elements $b \in B$ such that $\phi(b) = 0$ as b_1, b_2, ..., and define $c_1 = b_1$, $c_2 = c_1 \vee b_2$, $c_3 = c_2 \vee b_3$, etc. For each i, since $\phi \,\varepsilon\, A_{c_i}$ we must have $c_i \notin I$, i.e., there exists $\psi_i \,\varepsilon\, A_{c_i} \setminus A$. Let ψ be the limit of some convergent subquence of (ψ_i). Since $\psi_i(b_j) = 0$ for all $i \geq j$, it follows that $\psi(b_j) = 0$ for all j, i.e., ψ vanishes on every element on which ϕ vanishes. By Proposition 11.65 it follows that $\phi = \psi$. But every ψ_i belongs to the complement of A, which is clopen, so $\phi = \psi$ must also belong to the complement of A, a contradiction. This proves the claim.

This shows that the clopen classes A_a with $a \in I$ cover A, so by compactness there are finitely many $a_1, \ldots, a_n \in I$ such that $A = A_{a_1} \cup \cdots \cup A_{a_n}$. But then $A = A_a$ with $a = a_1 \wedge \cdots \wedge a_n$. Thus $f = \sigma(a)$, as desired.

Now let $\phi : B_1 \to B_2$ be a homomorphism. We want to show that $\phi^{**} = \sigma \circ \phi \circ \sigma^{-1}$. To do this, fix $g \,\varepsilon\, B_1^{**}$. By the first part of the proof, we know that $g = \sigma(a)$ for some $a \in B_1$. Then for any $\psi \,\varepsilon\, B_2^*$

$$\phi^{**}(g)(\psi) = (g \circ \phi^*)(\psi) = g(\psi \circ \phi) = \psi(\phi(a))$$

and

$$(\sigma \circ \phi \circ \sigma^{-1})(g)(\psi) = \sigma(\phi(a))(\psi) = \psi(\phi(a)).$$

This proves the second assertion in part (a).

For part (b), let X be a Stone space. It is routine to check that τ is a continuous function from X to X^{**}. It is injective because if $x \neq y$ then there is a clopen subclass of X that contains x but not y, and thus there exists a continuous function $f : X \to \{0, 1\}$ with $f(x) = 0$ and $f(y) = 1$; then $\tau(x)(f) = f(x) \neq f(y) = \tau(y)(f)$, showing that $\tau(x) \neq \tau(y)$. For surjectivity, let $\phi : X^* \to \{0, 1\}$ be a homomorphism; we must show that there exists $x \in X$ such that $\tau(x) = \phi$, i.e., $\phi(f) = f(x)$ for all $f \,\varepsilon\, X^*$.

If $f, g \,\varepsilon\, X^*$ satisfy $\phi(f) = \phi(g) = 1$, then also $\phi(f \wedge g) = 1$, showing that $f \wedge g \neq 0$. So the intersection of all the clopen subclasses $C \subseteq X$ for which $\phi(1_C) = 1$ is inhabited, since X is compact. (Here 1_C is the indicator function which takes the value 1 on C and 0 on C^c.) Suppose x and y are distinct elements of that intersection; then there is a clopen class C that contains x but not y, and then either $\phi(1_C) = 0$ or $\phi(1_{C^c}) = 0$, which would mean that either x or y does not belong to the intersection. This

contradiction shows that there is exactly one element x in the intersection. Thus $\phi(1_C) = 1$ if and only if $x \in C$. That is, for any $f \,\varepsilon\, X^*$ we have $\phi(f) = f(x)$. The fact that τ is a homeomorphism now follows from Theorem 10.44.

Now let $f : X_1 \to X_2$ be continuous. We want to show that $f^{**} = \tau \circ f \circ \tau^{-1}$. To do this, fix $\psi \,\varepsilon\, X_1^{**}$. We know that $\psi = \tau(x)$ for some $x \in X_1$. So for any $g \,\varepsilon\, X_2^*$

$$f^{**}(\psi)(g) = (\psi \circ f^*)(g) = \psi(g \circ f) = g(f(x))$$

and

$$(\tau \circ f \circ \tau^{-1})(\psi)(g) = \tau(f(x))(g) = g(f(x)).$$

This proves the second assertion in part (b). □

Chapter 12

Linear Algebra

12.1 Vector spaces

Definition 12.1. Let V be a class and let F be a field. *F-vector space operations* on V are functions $+ : V \times V \to V$ and $\cdot : F \times V \to V$ such that $+$ is an abelian group operation and

(1) $a \cdot (x + y) = a \cdot x + a \cdot y$
(2) $(a + b) \cdot x = a \cdot x + b \cdot x$
(3) $(ab) \cdot x = a \cdot (b \cdot x)$
(4) $1 \cdot x = x$

for all $x, y \in V$ and $a, b \in F$. A class equipped with F-vector space operations is an *F-vector space* or a *vector space over F*.

As usual, when I say that V is a vector space over F, or just a vector space when F is understood, this means that V is a class and we have fixed F-vector space operations on it. F is the *scalar field* and elements of F are *scalars*. The elements of a vector space are referred to as *vectors*.

Example 12.2. For any field F and any $n \in \mathbb{N}$, the nth power F^n is a vector space with operations defined coordinatewise. As is F^ω, as well as the class $F^{<\omega}$ of all the sequences in F^ω all but finitely many of whose entries are zero.

In Chapter 11, we mostly assumed our groups and rings were equipped with a compactly coherent topology. But we left Boolean algebras untopologized because in that case we were willing to assume countability. In the case of vector spaces, we can either assume they are countably generated ("based", Definition 12.3), or that they have been topologized or metrized.

The latter cases will be considered in Chapter 15; in this chapter we will mostly assume our vector spaces are based.

If V and W are vector spaces over a field F then a *linear transformation* from V to W is a function $T : V \to W$ which satisfies $T(x+y) = T(x)+T(y)$ and $T(ax) = aT(x)$ for all $x, y \in V$ and $a \in F$. It is customary to omit parentheses and write "Tx" instead of "$T(x)$" when this will not cause confusion. This is the appropriate notion of "homomorphism" in the setting of vector spaces. An *isomorphism* between vector spaces is a bijective linear transformation, and two vector spaces are *isomorphic* if there exists an isomorphism between them (an unbounded existential quantifier). A *linear subspace* (or just *subspace*) of V is a subclass which contains 0 and is stable under addition and multiplication by scalars. For groups and rings, only certain subobjects can be the kernels of homomorphisms, namely, normal subgroups and ideals, but for vector spaces (as for abelian groups) the subobjects are the same as the kernels: the kernel of a linear map is always a subspace and every subspace is the kernel of a linear map. One easily sees this by considering the quotient V/W of a vector space by a subspace, defined in the natural way by factoring out the equivalence relation given by $x \sim y$ if $x - y \in W$.

The following notions are fundamental.

Definition 12.3. A *linear combination* of an n-tuple of vectors $(x_1, \ldots, x_n)$ in a vector space V is any vector that can be expressed as $a_1x_1 + \cdots + a_nx_n$ for some a_1, …, $a_n \in F$. A linear combination of an infinite sequence of vectors (x_i) is a linear combination of some finite initial segment of the sequence. In either case the vectors are *dependent* (*independent*) if 0 can (cannot) be expressed as a linear combination with at least one nonzero a_i.

A set $S \subseteq V$ is said to be dependent or independent if its elements, when listed in a finite or infinite sequence with distinct entries, are dependent or independent. S is a *spanning set* and *spans* V if every element of V is a linear combination of elements of S. A *basis* for V is a subset S which is both independent and spanning — equivalently, every element of V can be expressed as a linear combination of elements of S in exactly one way, up to reordering. V is *based* if it has a basis and *finite dimensional* if it has a finite basis.

We define independence and dependence for tuples and sequences because we want to be able to say, for example, that the vectors x and y are dependent when $x = y$.

For any field F and any $n \in \mathbb{N}$, let $e_i \in F^n$ $(1 \leq i \leq n)$ be the n-tuple

whose ith entry is 1 and whose other entries are all 0. Similarly, for each $i \in \mathbb{N}$ let $e_i \in F^{<\omega}$ be the sequence whose ith entry is 1 and whose other entries are all 0. (Any use of the notation "e_i" should occur in a context where it is clear what the ambient space is.) Each e_i is a *standard basis vector* and the set $\{e_1, \dots, e_n\}$ (respectively, $\{e_i : i \in \mathbb{N}\}$) is the *standard basis* for F^n (respectively, $F^{<\omega}$). Every element of F^n or $F^{<\omega}$ is a linear combination of standard basis vectors in a unique way. Thus the standard basis vectors do constitute a basis in both cases. F^n is finite dimensional.

Not all vector spaces have bases according to this definition. For instance, if $F = \mathbb{Z}/2\mathbb{Z}$ is the two-element field then F^ω cannot be spanned by countably many vectors, since this would imply that it is countable, whereas in fact F^ω is not a set (cf. Theorem 9.6). We will be very interested in some vector spaces that are not based, but the ones we care about typically come equipped with some metric or topological structure and our philosophy is that the purely algebraic notion of a basis is no longer a good one for them. (What we usually want is a "Schauder basis". I'll come back to this in Section 15.3.)

Every based nonzero vector space over a field F is isomorphic either to F^n for some $n \in \mathbb{N}$ or to $F^{<\omega}$. For instance, if V has a finite basis $\{x_1, \dots, x_n\}$ then there is an isomorphism $T : F^n \cong V$ defined by putting $T\vec{a} = a_1x_1 + \cdots + a_nx_n$. Surjectivity follows from the fact that the x_i's span V, and injectivity follows from their independence: if $\vec{a}$ is not the zero vector in F^n then $T\vec{a} = a_1x_1 + \cdots + a_nx_n$ cannot be zero because this would imply a linear dependence among the x_i's. So the kernel of T is $\{0\}$, and this implies that T is injective.

Similarly, if V has an infinite basis $\{x_1, x_2, \dots\}$ then there is an isomorphism $T : F^{<\omega} \cong V$ defined by putting $T\vec{a} = \sum a_i x_i$, where the sum makes sense because if $\vec{a} \in F^{<\omega}$ then only finitely many of its coordinates are nonzero. Thus we have the following result.

Proposition 12.4. *If V is a finite dimensional vector space with an n-element basis then it is isomorphic to F^n. If V has an infinite basis then it is isomorphic to $F^{<\omega}$.*

This simple result allows us to transfer information about the spaces F^n and $F^{<\omega}$ to arbitrary based vector spaces. For instance, it may not be clear in a general vector space over an uncountable field whether every set of vectors is definitely dependent or independent. We would have to quantify over all choices of coefficients, which is not guaranteed to produce a definite result if there are uncountably many possible coefficients. However,

we will soon see that in F^n and $F^{<\omega}$, every set of vectors definitely is either dependent or independent, even if F is uncountable. By Proposition 12.4 the same is true in any based vector space. (It also works for compactly coherent vector spaces; see Corollary 15.30.)

The key tool we need is the ability to determine whether a given linear transformation T has a nonzero kernel. We know the kernel $\ker(T) = T^{-1}(0)$ is a subspace of the domain, but if the scalar field is uncountable we might not a priori be able to affirm that it definitely does or does not contain any nonzero vectors, because there are uncountably many vectors that could potentially belong to it. That is the issue here.

But if $T : F^n \to F^m$ is a linear transformation, with $m, n \in \mathbb{N}$, we can use the standard bases of F^n and F^m to get a good handle on T. For each $1 \leq j \leq n$ write $Te_j = t_{1j}e_1 + \cdots + t_{mj}e_m$. Once we know the values t_{ij} for all i and j, we can calculate the image of any vector in F^n by expressing it as a linear combination of the standard basis vectors and invoking linearity. Thus the map T is completely determined by the $m \times n$ *matrix* with entries t_{ij}. This is the *matrix of* T, which I will denote $[T]$, and t_{ij} is the (i, j) *entry*. Informally, an $m \times n$ matrix over F is a rectangular array with i varying vertically and j varying horizontally; formally, it can be modelled as an element of F^{mn}, where we call the entry with index $(i-1)n+j$ the "(i, j)-entry". We therefore have vector space operations of entrywise addition and scalar multiplication on matrices, which correspond to pointwise addition and scalar multiplication of linear transformations, which we write as $T+T'$ or aT. Thus $[T + T'] = [T] + [T']$ and $[aT] = a[T]$. Also, when $m = n$ the composition of two linear transformations $T, T' : F^n \to F^n$ is linear, and we denote it TT'; it corresponds to the *matrix product* $[TT'] = [T][T']$ whose (i, j) entry is the sum $t_{i1}t'_{1j} + \cdots + t_{in}t'_{nj}$.

Definition 12.5. Let F be a field and let $n \in \mathbb{N}$. Then we define $M_n(F)$ to be the class of all $n \times n$ matrices over F, with addition, multiplication, and scalar product specified as above. For distinct $m, n \in \mathbb{N}$ we can still define $M_{m,n}(F)$ to be the class of all $m \times n$ matrices over F, equipped with addition and scalar product, but not matrix multiplication. (However, for any $k, m, n \in \mathbb{N}$ we can define the matrix product of a $k \times m$ matrix with an $m \times n$ matrix in a similar way, with the result being a $k \times n$ matrix.)

$M_n(F)$ is a ring. If $n \geq 2$ it is noncommutative.

The identity element of $M_n(F)$ is the $n \times n$ matrix $[I]$ for which every diagonal entry is the identity element of F, and all other entries are zero. The corresponding linear transformation is just the identity map I from

F^n to itself.

Let $T : F^n \to F^m$ be a linear transformation and let the matrix entries of $[T]$ be t_{ij}. We will use this matrix to determine whether T has a nonzero kernel. First, find i_1 and j_1 such that $t_{i_1 j_1} \neq 0$, if any such i and j exist. (If not, then $\ker(F) = F^n$ and we are done.) Then by subtracting multiplies of the j_1th column from the other columns, we can ensure that the entries of the i_1th row are all zero besides the j_1th one, $t_{i_1 j_1}$.

This operation can be effected by multiplying $[T]$ on the right by the $n \times n$ matrix $[U_1]$ whose diagonal entries are all 1, whose (j_1, j) entry is $-t_{i_1 j}$ for all $j \neq j_1$, and whose other entries are all zero. This matrix is invertible; in fact, its inverse is defined exactly the same way but with $t_{i_1 j}$ in the (j_1, j) entry rather than $-t_{i_1 j}$. The point is that the linear transformation U_1 represented by $[U_1]$ is an isomorphism, so that TU_1 has a nonzero kernel if and only if T has a nonzero kernel. This helps because the matrix $[T][U_1]$ is a little simpler than the matrix $[T]$ — it has only one nonzero entry in the i_1th row.

Next, if there exist $i_2 \neq i_1$ and $j_2 \neq j_1$ such that the (i_2, j_2) entry of $[T][U_1]$ is nonzero, then we can subtract multiples of the j_2th column from the other columns and ensure that the entries of the i_2th row are all zero besides the j_2th one. This does not affect the i_1th row at all, which is still zero outside of its j_1th entry. Continuing in this way until we cannot find the desired nonzero entry (i_{r+1}, j_{r+1}) with i_{r+1} different from all previous i values and j_{r+1} different from all previous j values, we eventually convert the original matrix $[T]$ into a matrix $[T'] = [T][U_1] \cdots [U_r]$ with the special form that for some r there are indices i_1, ..., i_r, all distinct, and indices j_1, ..., j_r, all distinct, such that for any $1 \leq k \leq r$ we have $t'_{i_k j_k} \neq 0$ and $t'_{i_k j} = 0$ for all $j \neq j_k$, where t'_{ij} is the (i, j) entry of $[T']$. In addition, $t_{ij} = 0$ for any i and any $j \notin \{j_1, \ldots, j_r\}$; this corresponds to our not being able to continue the construction any further. Putting $U = U_1 \cdots U_r$, we find that the kernel of TU consists precisely of the linear combinations of the standard basis vectors e_i with $i \notin \Lambda = \{i_1, \ldots, i_r\}$. Since U is an isomorphism, the kernel of T consists of precisely the linear combinations of the vectors Ue_i.

Lemma 12.6. *For any linear transformation $T : F^n \to F^m$ there is an isomorphism $U : F^n \cong F^n$ and a set of indices $\Lambda \subseteq \{1, \ldots, m\}$ such that the vectors Ue_i with $i \in \Lambda^c$ constitute a basis for* $\ker(T)$.

So, to decide whether T has a nonzero kernel, check whether $\Lambda^c = \{1, \ldots, n\} \setminus \Lambda$ is empty. This also solves our earlier problem of determining

whether a set of vectors is independent or dependent, because for any finite set of distinct vectors $\{\vec{a}_1, \ldots, \vec{a}_k\}$ in F^m we can define a linear map $T : F^k \to F^m$ by putting $Te_i = \vec{a}_i$ and extending linearly; and then S is independent if and only if $\ker(T) = \{0\}$, dependent if and only if $\ker(T) \neq \{0\}$. So any finite set of vectors in F^m definitely is or is not independent. We can say the same of any finite set of vectors in $F^{<\omega}$, since any such set belongs, for some $m \in \mathbb{N}$, to the subspace of $F^{<\omega}$ consisting of all vectors $\vec{a}$ which satisfy $a_i = 0$ for all $i > m$, and this subspace is naturally isomorphic to F^m. The same is even true of infinite sets of vectors, since the set $\{\vec{a}_i : i \in \mathbb{N}\}$ (with the $\vec{a}_i$'s distinct) is independent if and only if $\{\vec{a}_1, \ldots, \vec{a}_k\}$ is independent for all k (LPO). Recalling Proposition 12.4, we get that any subset of a based vector space definitely is either dependent or independent.

If S is a subset of a vector space V and the vectors in V which are linear combinations of elements of S constitute a subclass of V, then we call that subclass the *span* of S and denote it "span(S)". Every span is a linear subspace. We collect what we have learned about spans:

Proposition 12.7. *Every subset of a based vector space has a span, and definitely is either dependent or independent. A subspace of a based vector space V is based if and only if it is the span of a subset of V.*

Proof. The definiteness of dependence/independence was demonstrated just above. For the existence of spans, let S be a subset of a based vector space V. Then $x \in V$ belongs to the span of S if and only if there exists an independent finite subset S_0 of S such that $S_0 \cup \{x\}$ is dependent, and we now know that this condition is definite (LPO). Any such span is based because S can be reduced to a basis by listing its elements as x_1, ..., x_n or x_1, x_2, ... and removing each x_k which lies in the span of $\{x_1, \ldots, x_{k-1}\}$. This condition is definite so it does define a subset of S, and that subset is independent and hence a basis for span(S). Conversely, any based subspace is trivially the span of a subset, namely, its basis. □

We can now give a quick proof of the following result.

Theorem 12.8. *In any based vector space, every independent set can be extended to a basis and every spanning set can be reduced to a basis.*

Proof. Let S be an independent subset of a based vector space V and let $\{x_1, \ldots, x_n\}$ or $\{x_1, x_2, \ldots\}$ be a basis for V. Then S can be extended to a basis $S' = S \cup \{x_i : i \in \Lambda\}$ where Λ is the set of indices such that x_i is

not in the span of $S \cup \{x_1, \ldots, x_{i-1}\}$. This S' is clearly independent, and it spans V because any element of V is a linear combination of finitely many x_i's, and given such an expression, for any $i \notin \Lambda$ we can replace x_i with a linear combination of vectors in S'. Thus S' spans V.

For the second statement, we saw in the proof of Proposition 12.7 that any set can be reduced to an independent subset with the same span. If the original set spans V then this independent subset will be a basis. □

Here are some more fundamental facts about bases.

Theorem 12.9. *The kernel of any linear transformation between based vector spaces is a based subspace of the domain. The image of any linear transformation between based vector spaces exists as a based subspace of the range. The quotient of a based vector space by a based subspace is based. The intersection of two based subspaces of a based vector space is based. The sum of two based subspaces of a based vector space exists as a based subspace.*

Proof. Let $T : V \to W$ be a linear transformation between based vector spaces. V is either finite dimensional or the union of a sequence of finite dimensional subspaces V_n (Proposition 12.4). By Lemma 12.6, in the former case the kernel of T is a based subpace of V. In the latter case the kernel of T is the union of the kernels of its restrictions to the V_n. Each of these kernels has a basis, and the kernel of T is the span of these bases. Thus the kernel of T is based by Proposition 12.7. The second statement also follows from Proposition 12.7 because the image of a linear transformation equals the span of the image of a basis for the domain.

If V is a based vector space and W is a based subspace, then any basis S of W can be extended to a basis $S \cup S'$ of V (Theorem 12.8). Without loss of generality we can assume S and S' are disjoint. The image of S' in V/W under the quotient map is then a basis for V/W.

Next, let V_1 and V_2 be based subspaces of a based vector space V. Then the quotient map $\pi : V \to V/V_2$ is a linear transformation between based vector spaces, and so is its restriction $\pi|_{V_1}$. Since $V_1 \cap V_2$ is the kernel of $\pi|_{V_1}$, it follows that $V_1 \cap V_2$ is based.

Since V_1 and V_2 are based, they are spanned by subsets S_1 and S_2, respectively. Then the span of $S_1 \cup S_2$ exists as a based subspace of V by Proposition 12.7. The sum of V_1 and V_2 is this span. □

I noted earlier that the kernel of any linear transformation from F^n to F^m definitely is or is not zero. That statement can be generalized:

Proposition 12.10. *Let V be a vector space and let W_1 and W_2 be subspaces of V. Suppose W_1 is based. Then it definitely is or is not contained in W_2.*

Proof. Let S_1 be a basis for W_1. Then W_1 is contained in W_2 if and only if S_1 is contained in W_2, and this condition is definite because S_1 is countable and membership in subclasses is definite. □

The kernel of a linear transformation is based by Theorem 12.9, so Proposition 12.10 tells us that the kernel definitely is or is not contained in — hence equal to — the zero subspace.

The three isomorphism theorems for based vector spaces follow easily from what we have proven in this section.

Theorem 12.11. *(First isomorphism theorem for based vector spaces) Let $T : V \to W$ be a linear transformation between based vector spaces. Then $\ker(T)$ is a based subspace of V, the image $T(V)$ exists as a based subspace of W, and there is an isomorphism $\widetilde{T} : V/\ker(T) \cong T(V) \subseteq W$ such that $T = \widetilde{T} \circ \pi$.*

Theorem 12.12. *(Second isomorphism theorem for based vector spaces) Let V_1 and V_2 be based subspaces of a based vector space V. Then $V_1 \cap V_2$ is a based subspace of V_1, V_2 is a based subspace of $V_1 + V_2$, and there is an isomorphism between $V_1/(V_1 \cap V_2)$ and $(V_1 + V_2)/V_2$.*

Theorem 12.13. *(Third isomorphism theorem for based vector spaces) Let V_1 and V_2 be based subspaces of a based vector space V with $V_2 \subseteq V_1$. Then V_1/V_2 is a based subspace of V/V_2, and every based subspace of V/V_2 has this form. We have an isomorphism between $(V/V_2)/(V_1/V_2)$ and V/V_1.*

There are also isomorphism theorems for compactly coherent vector spaces which might not be based; see Section 15.3.

12.2 Dimension

Now that we know based vector spaces are well-behaved in so many ways, we are in a position to develop the notion of the dimension of a finite dimensional vector space.

Lemma 12.14. *An n-tuple $(x_1, \ldots, x_n)$ or sequence (x_i) of vectors in some based vector space is dependent if and only if for some k, x_k is a linear combination of the vectors x_1, …, x_{k-1}.*

Proof. Suppose the tuple or sequence is dependent; then for some k the tuple $(x_1, \ldots, x_k)$ is dependent. By Proposition 12.7, for each k the tuple $(x_1, \ldots, x_k)$ definitely is or is not dependent, so this means there is a minimal value of k for which it is dependent. For this k there is a linear combination $a_1x_1 + \cdots + a_kx_k = 0$ with at least one a_i nonzero. By minimality of k, it must be the case that a_k is nonzero, and so we have $x_k = -\frac{a_1}{a_k}x_1 - \cdots - \frac{a_{k-1}}{a_k}x_{k-1}$. We have shown that this x_k is a linear combination of the vectors x_1, ..., x_{k-1}.

Conversely, suppose some x_k is a linear combination of x_1, ..., x_{k-1}, i.e., we have $x_k = b_1x_1 + \cdots + b_{k-1}x_{k-1}$ for some coefficients b_1, ..., b_{k-1}. Then $b_1x_1 + \cdots + b_{k-1}x_{k-1} - x_k = 0$ and this shows that $(x_1, \ldots, x_k)$ is dependent, so that the entire tuple or sequence is dependent. □

We know from Theorem 12.8 that every independent set can be extended to a basis, but we do not yet know that all bases have the same size, so that does not help with the next lemma.

Lemma 12.15. *If some basis for a vector space V has exactly n elements, then every independent set has at most n elements. Every based subspace of a finite dimensional space is finite dimensional.*

Proof. Let $S = \{x_1, \ldots, x_n\}$ be a basis and suppose we had an independent set S' with $n + 1$ elements, $S' = \{y_1, \ldots, y_{n+1}\}$. (If there were an independent set with more than $n + 1$ elements, then any $n + 1$ element subset of it would do.) Consider the list

$$y_1, x_1, \ldots, x_n.$$

We know that these vectors span V and are dependent, since y_1 is a linear combination of the x_i's. By Lemma 12.14, we can find a value of k such that x_k is a linear combination of vectors preceding it in the list. Remove this x_k. The remaining vectors in the list clearly still span V.

Next, add y_2 to the beginning of the list. This ensures a new dependence since the vectors in the list spanned V before y_2 was added, so again we can find a vector in the list that is a linear combination of vectors that come before it. Remove that vector.

Continue in this way, at each step adding the next y_i to the beginning of the list and then removing one of the x_i's. It will always be one of the x_i's that is removed, because we knew at the start that the y_i's are independent, and each y_j is preceded only by other y_i's. We also know that, at each step, after we have removed our x_i the remaining vectors will still span V.

This process cannot end until we reach the list $y_n, \ldots, y_1$, so we know that these vectors span V, which means S' could not have been independent, a contradiction. We have shown that every independent set cannot have more than n elements. As every set definitely does or does not have more than n elements, it follows that every independent set has at most n elements.

For the last statement, just note that every based subspace of V has a basis which is an independent set in V and is therefore finite. □

Theorem 12.16. *If S and S' are bases of the same vector space, and S is finite, then S' is also finite and they have the same number of elements.*

Proof. We know from Lemma 12.15 that S' cannot contain more vectors than S. In particular, S' is finite, which means that we can now invoke Lemma 12.15 with the roles of S and S' reversed, to show that S cannot contain more vectors than S'. Thus the two bases have the same number of elements. □

Thus, if V is finite dimensional then all bases of V have the same number of elements. That number is the *dimension* of V, denoted "$\dim(V)$".

Corollary 12.17. *Suppose* $\dim(V) = n$ *and let S be a set of exactly n vectors in V. Then S is a basis if and only if it is independent if and only if it is spanning.*

Proof. By definition, every basis is independent and spanning. If S is independent then it can be extended to a basis by Theorem 12.8, and by Theorem 12.16 that basis must contain exactly n elements, so S must already have been a basis.

To show that S is a basis if it is spanning, note first that any set of strictly fewer than n vectors cannot span V. That is because it could be reduced to an independent spanning set, i.e., a basis (Theorem 12.8), but we already know that every basis has n elements. Now, supposing that S is a spanning set of exactly n vectors in V, there can be no linear dependence among these vectors, as that would mean we could remove a vector and still span. As we just saw, this is impossible. So if S spans V then it is a basis. □

The kernel $\ker(T)$ and image $\mathrm{im}(T)$ of a linear transformation T between finite dimensional vectors spaces are always based (Theorem 12.9) and hence finite dimensional (Lemma 12.15). These dimensions satisfy the

following relation. (The *rank* of T is the dimension of $\mathrm{im}(T)$ and its *nullity* is the dimension of $\ker(T)$.)

Theorem 12.18. *(Rank-nullity) Let $T : V \to W$ be a linear transformation between finite dimensional vector spaces. Then* $\dim(V) = \dim(\ker(T)) + \dim(\mathrm{im}(T))$.

Proof. We noted just above that $\ker(T)$ is finite dimensional. Thus it has a basis $\{x_1, \ldots, x_r\}$, where $r = \dim(\ker(T))$. By Theorems 12.8 and 12.16 we can find vectors x_{r+1}, ..., x_n where $n = \dim(V)$, such that $\{x_1, \ldots, x_n\}$ is a basis for V. I claim that the vectors Tx_{r+1}, ..., Tx_n constitute a basis for $\mathrm{im}(T)$. This entails that $\dim(V) = n$ equals $\dim(\ker(T)) = r$ plus $\dim(\mathrm{im}(T)) = n - r$.

To prove the claim, we first check that $\{Tx_{r+1}, \ldots, Tx_n\}$ spans $\mathrm{im}(T)$. This is easy, because $\mathrm{im}(T)$ is spanned by the vectors Tx_1, ..., Tx_n (since $\{x_1, \ldots, x_n\}$ is a basis for V), but $Tx_1 = \cdots = Tx_r = 0$ so in fact Tx_{r+1}, ..., Tx_n suffice. Next, we must check that these vectors are independent. To see this suppose $a_{r+1}Tx_{r+1} + \cdots + a_nTx_n = 0$; we must show that $a_{r+1} = \cdots = a_n = 0$. But applying T to the vector $a_{r+1}x_{r+1} + \cdots + a_nx_n \in V$ yields the linear combination which we have assumed is zero, so this vector must belong to $\ker(T)$. Since the span of $\{x_{r+1}, \ldots, x_n\}$ intersects $\ker(T)$ only in the zero vector (otherwise there would be a linear dependence among $\{x_1, \ldots, x_n\}$), the only way this can happen is if all the a_i's are zero. □

Theorem 12.19. *Let V_1 and V_2 be finite dimensional subspaces of a based vector space V. Then $V_1 \cap V_2$ and $V_1 + V_2$ are also finite dimensional subspaces of V, and we have* $\dim(V_1) + \dim(V_2) = \dim(V_1 \cap V_2) + \dim(V_1 + V_2)$.

Proof. $V_1 \cap V_2$ is a based subspace of V_1 by Theorem 12.9, and therefore finite dimensional by Lemma 12.15. $V_1 + V_2$ is finite dimensional because it is spanned by the union of a basis for V_1 and a basis for V_2.

Let S_0 be a basis for $V_1 \cap V_2$, and extend it to a basis S_1 for V_1 and a basis S_2 for V_2. Then $S_1 \cup S_2$ is a basis for $V_1 + V_2$, and it is the union of the disjoint sets S_1 and $S_2 \setminus S_0$. So the dimension of $V_1 + V_2$ is $\dim(V_1) + (\dim(V_2) - \dim(V_1 \cap V_2))$. □

We can also use bases to describe the linear transformations $T : V \to W$ from one finite dimensional vector space to another as an indexed family in a canonical way. We already know that V and W are isomorphic to F^n and F^m where $n = \dim(V)$ and $m = \dim(W)$, and the linear transformations

from F^n to F^m are naturally indexed by the matrix space $M_{m,n}(F)$, but the choice of isomorphisms with F^n and F^m depend on a choice of bases of V and W. As in our construction of $C(X,Y)$ at the end of Section 10.3, we can do this canonically by treating all bases of V simultaneously. Thus let Z be the subclass of $V^n \times W^n$ consisting of all $2n$-tuples $(v_1, \ldots, v_n, w_1, \ldots, w_n)$ such that $\{v_1, \ldots, v_n\}$ is a basis for V, and define an equivalence relation $\sim$ on Z by making two such $2n$-tuples equivalent if the maps $v_i \mapsto w_i$ and $v'_i \mapsto w'_i$ extend to the same linear transformation from V to W. We then define $\mathcal{L}(V,W)$ to be the quotient class $Z/\!\sim$, and let it act as a system of exact codes for linear transformations from V to W in the obvious way. When $V = W$ we write $\mathcal{L}(V) = \mathcal{L}(V,V)$.

Similarly, let two subsets of a based vector space V be equivalent if they have the same span (i.e., each element of one belongs to the span of the other, LPO), then let $\mathrm{Sub}(V)$ be $\mathcal{P}(V)$ modulo this equivalence. Then $\mathrm{Sub}(V)$ is a system of exact codes for the based subspaces of V.

We close this section with a brief discussion of determinants.

Let σ be a permutation of $\{1, \ldots, n\}$, i.e., a bijection between $\{1, \ldots, n\}$ and itself. Say that σ *crosses* i and j if $i < j$ but $\sigma(i) > \sigma(j)$. The *sign* of σ, denoted $\mathrm{sign}(\sigma)$, is 1 if the number of pairs (i,j) that σ crosses is even and -1 if it is odd. We say that the sign of σ measures the *parity* of the number of crossing pairs.

If σ' is obtained from σ by switching a single pair of elements (that is, $\sigma'(i) = \sigma(j)$ and $\sigma'(j) = \sigma(i)$ for some $i \neq j$, and otherwise σ' equals σ) then $\mathrm{sign}(\sigma') = -\mathrm{sign}(\sigma)$. This is because if $\sigma(k)$ lies between $\sigma(i)$ and $\sigma(j)$ then σ crosses i and k if and only if σ' does not cross i and k, and similarly for j and k, so the even/odd parity of the total number of crosses is not affected by these pairs, and at the same time, σ crosses i and j if and only if σ' does not cross i and j, so that does change the parity of the total number of crosses. Crossings of i and j with those k where $\sigma(k)$ is not between $\sigma(i)$ and $\sigma(j)$ are unaffected. Thus the parity of the number of pairs crossed by σ' is different from the parity of the number of pairs crossed by σ.

Any permutation can be decomposed into a product of some number of single pair transpositions. It follows from the last paragraph that the sign of the permutation measures the parity of this number (and thus all such decompositions must have the same parity).

Let $[T] \in M_n(F)$ be an $n \times n$ matrix over a field F, with entries t_{ij}.

The *determinant* of $[T]$ is the number

$$\det([T]) = \sum \operatorname{sign}(\sigma) t_{\sigma(1),1} \cdots t_{\sigma(n),n},$$

taking the sum over all permutations σ of $\{1, \ldots, n\}$.

We can also consider the determinant to be a function of the columns of $[T]$. That is, let $\vec{x}^j = Te_j = t_{1j}e_1 + \cdots + t_{nj}e_n \in F^n$ and define $\det(\vec{x}^1, \ldots, \vec{x}^n) = \det([T])$.

Lemma 12.20. *Let F be a field, let $n \in \mathbb{N}$, and let $[I]$ be the $n \times n$ identity matrix over F. Then $\det([I]) = 1$. If all but one of the vectors $\vec{x}^j$ is fixed, then $\det(\vec{x}^1, \ldots, \vec{x}^n)$ is a linear function of the one that is not fixed. If $[T] \in M_n(F)$ and $[T']$ is the matrix formed by switching two of the columns of $[T]$, then $\det([T']) = -\det([T])$. If $[T]$ has two equal columns then $\det([T]) = 0$, and $\det(\vec{x}^{\sigma(1)}, \ldots, \vec{x}^{\sigma(n)}) = \operatorname{sign}(\sigma) \cdot \det(\vec{x}^1, \ldots, \vec{x}^n)$ for any permutation σ of $\{1, \ldots, n\}$.*

I omit the simple proof of this lemma.

Theorem 12.21. *Let F be a field, let $n \in \mathbb{N}$, and let $[T], [T'] \in M_n(F)$. Then $\det([T][T']) = \det([T]) \cdot \det([T'])$.*

Proof. Let $\vec{x}^1, \ldots, \vec{x}^n$ be the columns of $[T']$ and define $\Delta_T(\vec{x}^1, \ldots, \vec{x}^n) = \Delta_T([T']) = \det([T][T'])$. The columns of $[T][T']$ are $[T]\vec{x}^1, \ldots, [T]\vec{x}^n$, so we can also say $\Delta_T(\vec{x}^1, \ldots, \vec{x}^n) = \det([T]\vec{x}^1, \ldots, [T]\vec{x}^n)$. From this we can see that Δ_T has the properties of the determinant given in Lemma 12.20. Thus

$$\Delta_T([T']) = \Delta_T\left(\sum t'_{i1} e_i, \vec{x}^2, \ldots, \vec{x}^n\right) = \sum t'_{i1} \Delta_T(e_i, \vec{x}^2, \ldots, \vec{x}^n);$$

doing the same thing successively for each entry, we finally obtain

$$\Delta_T([T']) = \sum t'_{i_1,1} \cdots t'_{i_n,n} \Delta_T(e_{i_1}, \ldots, e_{i_n}),$$

with the sum taken over all $1 \leq i_1, \ldots, i_n \leq n$. But Δ_T vanishes when applied to a matrix with two equal columns, so this sum can be written as $\sum t'_{\sigma(1),1} \cdots t'_{\sigma(n),n} \Delta_T(e_{\sigma(1)}, \ldots, e_{\sigma(n)})$, summing over all permutations σ of $\{1, \ldots, n\}$, and then using the last part of the lemma we get

$$\begin{aligned} \Delta_T([T']) &= \left(\sum \operatorname{sign}(\sigma) \cdot t'_{\sigma(1),1} \cdots t'_{\sigma(n),n}\right) \cdot \Delta_T(e_1, \ldots, e_n) \\ &= \left(\sum \operatorname{sign}(\sigma) \cdot t'_{\sigma(1),1} \cdots t'_{\sigma(n),n}\right) \cdot \Delta_T([I]) \\ &= \det([T']) \cdot \det([T]), \end{aligned}$$

as desired. □

Theorem 12.22. *Let F be a field, let $n \in \mathbb{N}$, and let $T : F^n \to F^n$ be linear. Then T is invertible if and only if* $\det([T]) \neq 0$.

Proof. T definitely is or is not invertible because the vectors Te_1, ..., Te_n definitely are or are not independent. And if T is invertible then

$$\det([T]) \cdot \det([T^{-1}]) = \det([T][T^{-1}]) = \det([I]) = 1,$$

so $\det([T])$ must be nonzero. Whereas if T is not invertible then there is a linear dependence among the columns of $[T]$, say $[T]e_k = a_1[T]e_1 + \cdots + a_{k-1}[T]e_{k-1}$, and we know from Lemma 12.20 that the determinant of the matrix obtained from $[T]$ by replacing the kth column with a copy of the jth column for some $j < k$ is zero, so by linearity in the kth entry we get that $\det([T]) = 0$. □

Define the *matrix of a permutation* σ of $\{1, \ldots, n\}$ to be the matrix $[I_\sigma]$ with columns $e_{\sigma(1)}$, ..., $e_{\sigma(n)}$. Thus I_σ sends e_i to $e_{\sigma(i)}$, and it follows that $I_\sigma I_{\sigma'} = I_{\sigma\circ\sigma'}$. We have $\operatorname{sign}(\sigma) = \det([I_\sigma])$ and

$$\operatorname{sign}(\sigma \circ \sigma') = \det([I_\sigma][I_{\sigma'}]) = \det([I_\sigma]) \cdot \det([I_{\sigma'}]) = \operatorname{sign}(\sigma) \cdot \operatorname{sign}(\sigma').$$

That is, sign is a homomorphism from the group of permutations of $\{1, \ldots, n\}$ to the multiplicative group $\{1, -1\}$.

12.3 Galois theory

In this section we will discuss the basics of Galois theory and algebraic closures. "Homomorphism" will mean "unital ring homomorphism" throughout.

Let F be a field. An *extension* of F is a field which contains F as a subfield. Every extension K of F is a vector space over F in an obvious way. If K is finite dimensional as a vector space over F then we say K is a *finite* extension of F and write $[K : F]$ for its dimension as an F-vector space.

Proposition 12.23. *Let $F \subseteq L \subseteq K$ be fields. Suppose L is a finite extension of F and K is a finite extension of L. Then K is a finite extension of F and* $[K : F] = [K : L][L : F]$.

Proof. Let $\{a_1, \ldots, a_n\}$ be a basis for L as a vector space over F and let $\{b_1, \ldots, b_m\}$ be a basis for K as a vector space over L. I claim that the vectors $a_i b_j$ for $1 \leq i \leq n$ and $1 \leq j \leq m$ constitute a basis for K

over F. They span because every element of K is a linear combination of the b_j's with coefficients in L, and those coefficients can all be expanded as linear combinations of the a_i's with coefficients in F. Replacing the coefficients of the first linear combination with their expansions yields an expression of the original element of K as a linear combination of the $a_i b_j$'s with coefficients in F. For independence, suppose we are given that some linear combination of the $a_i b_j$'s with coefficients in F is zero. Collecting terms, we can write this linear combination as $A_1 b_1 + \cdots + A_m b_n = 0$ where each A_i is a linear combination of the a_i's with coefficients in F and hence belongs to L. Since $\{b_1, \ldots, b_m\}$ is a basis for K over L, it follows that $A_1 = \cdots = A_m = 0$. Then since $\{a_1, \ldots, a_n\}$ is a basis for L over F, for each j all of the coefficients of A_j (a linear combination of the a_i's) must be zero. Thus all the coefficients of the original linear combination of the $a_i b_j$'s are zero, and we have shown that these vectors are independent. □

If K is a finite extension of F then it is isomorphic as a vector space to F^n for some $n \in \mathbb{N}$. Thus if F is equipped with a coherently compact topology then we can give K the coherent product topology. The isomorphism between K and F^n depends on a choice of basis, but the coherent product topology does not; that is a consequence of the following fact.

Proposition 12.24. *Let F be a compactly coherent field and give F^m and F^n ($m, n \in \mathbb{N}$) the coherent product topologies. Then any linear transformation from F^m to F^n is continuous, and any invertible linear transformation from F^n to itself is a homeomorphism.*

Proof. Given a linear transformation $T : F^m \to F^n$, write $T(\vec{x}) = (T_1(\vec{x}), \ldots, T_n(\vec{x}))$ where $T_1, \ldots, T_n$ are linear transformations from F^m to F obtained by composing the coordinate projections on F^n with T. It follows from the characterization of convergence in coherent products (Proposition 10.95) and the axioms for a compactly coherent ring that each T_i is continuous. Continuity of T then follows from Proposition 10.95. The second part of the proposition is an immediate consequence. □

Corollary 12.25. *Let F be a compactly coherent field and let W be an F-vector subspace of F^n, equipping the latter with the coherent product topology. Then W is based if and only if it is closed.*

Proof. Suppose W is based. Then any basis for W can be extended to a basis for F^n (Theorem 12.8), and this yields an isomorphism of F^n with

itself that takes W to the span of some set of standard basis vectors. The latter is clearly closed, so W must be closed by Proposition 12.24.

Now suppose W is closed. Then W definitely either does or does not equal $\{0\}$ because it definitely either is or is not contained in $\{0\}$ (Proposition 10.56). In the former case we are done; in the latter case let $v_1 \in W \setminus \{0\}$ and define $W_1 = \operatorname{span}\{v_1\}$. Then W definitely either is or is not contained in, hence equal to, W_1. In the former case we are done; in the latter case let $v_2 \in W \setminus W_1$ and define $W_2 = \operatorname{span}\{v_1, v_2\}$. Continue on in this way; since the v_i's are linearly independent, by Lemma 12.15 the process must terminate after at most n steps. At that point W is seen to equal a spanned, and hence based (Proposition 12.7) subspace of F^n. □

I will refer to the coherent product topology on K as the *standard* topology (relative to its being a finite extension of F). In fact, this terminology is superfluous because there is only one topology on K that is compatible with the topology on F as an F-vector space (in the sense of Section 15.3), but proving this would take us too far afield and we don't need it here.

Every countable field is compactly coherent relative to the discrete topology. $\mathbb{R}$ and $\mathbb{C}$ are two basic examples of uncountable compactly coherent (even locally compact) fields. And we know from Example 11.39 that if F is compactly coherent then so is the field $F(x_1, \ldots, x_n)$, and also from Theorem 11.13 that any quotient of a compactly coherent (respectively, locally compact or compact) ring by a closed ideal is compactly coherent (respectively, locally compact or compact). This will include fields of the form $F[x]/(p)$ as in Proposition 11.50; in fact, these fields come equipped with the standard topology.

Proposition 12.26. *Let F be a compactly coherent field and let $p \in F[x]$ be irreducible. Then $F[x]/(p)$ is linearly homeomorphic to the coherent power F^n where $n = \deg(p)$.*

Proof. We have a natural homeomorphic vector space isomorphism between F^n and P_{n-1}, the polynomials of degree at most $n-1$. This is just how the topology on P_{n-1} was defined. The quotient map from $F[x]$ to $F[x]/(p)$ restricts to a bijection ϕ between P_{n-1} and $F[x]/(p)$; we must show that this is a homeomorphism.

It is continuous because the quotient map is continuous. Conversely, for every relatively open subclass U of P_{n-1} we need $\phi(U)$ to be the image under the quotient map of a (p)-invariant open subclass of $F[x]$. It will suffice to show that $U + (p)$ is open in $F[x]$, or equivalently, that $(U + (p)) \cap$

$P_k = U + ((p) \cap P_k)$ is relatively open in P_k for all $k \geq n$.

Fix $k \geq n$ and consider the vector space product $((p) \cap P_k) \times P_{n-1}$. The map $(p, q) \mapsto p + q$ is a bijection between this product and P_k. Relative to a basis for P_k that is the union of a basis for P_{n-1} and a basis for $(p) \cap P_k$, then, $U + ((p) \cap P_k)$ has the form (open subclass of $P_{n-1} \times F^{k-(n-1)}$), which is clearly open in P_k for the standard topology. So $U + ((p) \cap P_k)$ is indeed open in P_k, as desired. □

Let K be an extension of F. An element $a \in K$ is *algebraic over* F if it is a root of a polynomial with coefficients in F. In this case the polynomials in $F[x]$ of which a is a root form an ideal, which by Proposition 11.44 and Theorem 11.46 has the form (p), the principal ideal generated by some $p \in F[x]$. This p is irreducible — otherwise a would be a root of one of its factors — and dividing by the leading coefficient (the coefficient of x^n where n is the degree of p), we can assume p is *monic*, meaning that its leading coefficient is 1. This is the *minimal polynomial of* a, or the *minimal polynomial of* a *over* F if that needs to be said. It is the unique monic polynomial which generates the ideal of all polynomials that vanish on a.

Assume F and K are compactly coherent, and for any $a \in K$ write $F[a]$ for the subring of K generated by F and a. That is, $F[a]$ is the class of elements of K that can be expressed as a linear combination of powers of a with coefficients in F. It is the image of the continuous map Φ from $F[x]$ to K which takes x to a from Proposition 11.30.

If a is algebraic with minimal polynomial p, then the kernel of Φ is the ideal (p), and by the first isomorphism theorem for rings (Theorem 11.36) this induces a continuous isomorphism between $F[x]/(p)$ and $\Phi(F[x]) = F[a]$. It takes the *residue* $\overline{x}$ of x, the element of $F[x]/(p)$ instantiated by x, to a. Since $F[x]/(p)$ is a field by Proposition 11.50, it follows that $F[a]$ is also a field.

The next proposition gives two more alternative characterizations of the property of being algebraic. It requires a lemma whose proof is similar to the proof of Lemma 11.47, so I will omit it. (If a is algebraic then there is a polynomial $p \in F[x]$ with $p(a) = 0$; then, letting $\mathring{F}$ be a coherently dense subset of F, there exist polynomials q_n of the same degree as p, with coefficients in $\mathring{F}$ and with $q_n \to p$ and $q_n(a) \to 0$, etc.)

The following terminology will be helpful in the sequel. Suppose $\iota : F \to R$ and $\iota' : F \to R'$ are continuous unital homomorphisms from a compactly coherent field F into two unital compactly coherent rings, R and R'. Note that ι and ι' are automatically injective since F is a field, so $\iota' \circ \iota^{-1}$ is

an isomorphism between $\iota(F)$ and $\iota'(F)$. In this situation we will call a continuous unital ring homomorphism $\phi : R \to R'$ which takes $\iota(a)$ to $\iota'(a)$, for all $a \in F$, an *F-homomorphism*.

Lemma 12.27. *Let K be a compactly coherent field and let F be a closed subfield of K. Then every element of K definitely is or is not algebraic over F.*

Proposition 12.28. *Let K be a compactly coherent field, let F be a closed subfield of K, and let $a \in K$. Then a is algebraic over F if and only if $F[a]$ is finite dimensional as a vector space over F if and only if $F[a]$ is a field.*

Proof. The equivalence is trivial if $a = 0$, so suppose $a \neq 0$ is algebraic over F and let $p \in F[x]$ be its minimal polynomial. Since $F[x]$ is a Euclidean domain (Proposition 11.44), for any $m \in \mathbb{N}$ we can write $x^m = qp + r$ where $r = 0$ or $\deg(r) < \deg(p)$. Evaluating both sides on a yields $a^m = r(a)$. It follows that all powers of a belong to the span of $\{1, a, \ldots, a^{n-1}\}$ where $n = \deg(p)$. This shows that $F[a]$ is finite dimensional.

Conversely, suppose $F[a]$ is finite dimensional. If its dimension is n then there must be a linear dependence among the $n + 1$ elements 1, a, …, a^n. But this means that there is a nonzero polynomial in $F[x]$ which has a as a root, i.e., a is algebraic over F.

Now we saw above that if a is algebraic over F then $F[a]$ is a field. Conversely, suppose a is not algebraic over F. Then the F-homomorphism from $F[x]$ to $F[a]$ which takes x to a has null kernel, and hence is an isomorphism. But $F[x]$ is not a field, and thus neither is $F[a]$. □

Corollary 12.29. *Let K be a compactly coherent field and let F be a closed subfield of K. Then the elements of K that are algebraic over F form a subfield of K.*

Proof. The algebraic elements constitute a subclass of K by Lemma 12.27. Now suppose $a, b \in K$ are both algebraic over F (and nonzero). It will suffice to show that there is a subfield L of K containing both a and b and such that $[L : F]$ is finite; then every element of L is algebraic over F by Proposition 12.28, in particular $a + b$, ab, $-a$, and a^{-1} are algebraic. This will show that sums, products, and multiplicative inverses of nonzero algebraic elements are algebraic.

We know that $F[a]$ is a finite extension of F by Proposition 12.28, and since b is algebraic over F it is also algebraic over $F[a]$, so that $F[a, b] = F[a][b]$ is a finite extension of $F[a]$ by Proposition 12.28 again. Proposition

12.23 then yields that $L = F[a, b]$ is a subfield of K containing both a and b and such that $[L : F] = [F[a, b] : F[a]][F[a] : F]$ is finite. □

If $\phi : F \to K$ is any homomorphism of fields, for any $p \in F[x]$ we write $\phi(p)$ for the polynomial in $K[x]$ obtained by applying ϕ to each coefficient of p.

Lemma 12.30. *Let F be a compactly coherent field and let $p \in F[x]$ be nonconstant. Then there is a compactly coherent field K and a continuous homomorphism $\phi : F \to K$ such that K is a finite extension of $\phi(F)$ and $\phi(p)$ is a product of linear factors in $K[x]$.*

Proof. We can assume that the unique factorization of p in $F[x]$ has no linear factors, since those could be divided out at the start. Next, if $\deg(p) = 1$ then we can just take $K = F$. Otherwise let q be an irreducible factor of p in $F[x]$, let $F' = F[x]/(q)$, and let ι be the natural map from F into F'. Now $\iota(p)$ has a root in F', namely the residue of x, and therefore it has a linear factor in $F'[x]$. Letting $\iota(p) = rs$ where $s \in F'[x]$ is the product of all the linear factors of $\iota(p)$ in $F'[x]$ (recall from Proposition 11.44 and Theorems 11.46 and 11.49 that $F'[x]$ is a UFD), we obtain a polynomial $r \in F'[x]$ with no linear factors and degree strictly less than $\deg(p)$. We can now apply the same argument to r to get a further finite extension in which $\phi(r)$ has linear factors. Continuing in this fashion, we eventually get a continuous homomorphism ϕ from f into a field K which is a finite extension of $\phi(F)$ by Proposition 12.23, such that $\phi(p)$ is a product of linear factors in $K[x]$. □

Now that those general preliminaries about algebraic elements are out of the way, we can proceed to Galois theory.

Lemma 12.31. *Let K and L be finite extensions of a compactly coherent field F, let $\phi : L \to K$ be an F-homomorphism, and let a be an algebraic element in some extension of L. Let $p \in L[x]$ be the minimal polynomial of a and let n be the number of roots of $\phi(p)$ in K. Then ϕ extends to exactly n homomorphisms from $L[a]$ into K, each of which takes a to a different root of $\phi(p)$.*

Proof. Since a is a root of p, the image of a under any F-homomorphism from $L[a]$ into K is a root of $\phi(p)$. Any such homomorphism is determined by its value on a, so there can be at most n homomorphisms from $L[a]$ into K which extend ϕ.

As we discussed earlier, for any root b of $\phi(p)$ in K there is an F-homomorphism from $L[x]/(p)$ into K which takes $\overline{x}$ to b. Composing with the natural F-isomorphism between $L[a]$ and $L[x]/(p)$ yields a homomorphism from $L[a]$ into K which takes a to b. So there are at least n extensions of ϕ. We conclude that there are exactly n extensions. □

In this proof the standard topologies on L and K are compactly coherent; we need this in order to use Proposition 11.44 and Theorem 11.46 to ensure that a has a minimal polynomial, and to ensure that every polynomial in $K[x]$ definitely does or does not have a root in K, so that n is well-defined. We can effectively quantify over all F-homomorphisms from L to K by quantifying over the class $\mathcal{L}(L, K)$ introduced near the end of Section 12.2.

Definition 12.32. Let K be a finite extension of F. It is a *normal* extension if every irreducible polynomial in $F[x]$ that has a root in K factors into a product of linear polynomials in $K[x]$. (If one root belongs to K, then they all do.) It is *Galois* if every irreducible polynomial in $F[x]$ that has a root in K factors into a product of linear polynomials in $K[x]$ with distinct roots.

In the next series of results we consider a compactly coherent field F, a finite extension K of F equipped with the standard topology, and closed intermediate fields L satisfying $F \subseteq L \subseteq K$. We can quantify over these closed intermediate fields by regarding them as based subspaces of K regarded as a based F-vector space. That is, every closed intermediate field is coded by an element of the pseudoclass $\mathrm{Sub}(K)$ introduced near the end of Section 12.1. Note that according to Corollary 12.25 and Proposition 12.28 a based subspace of K is a closed intermediate field if and only if it is a subring of K.

Lemma 12.33. *Let F be a compactly coherent field, let K be a finite extension of F equipped with the standard topology, and let L be a closed intermediate field. If K is a Galois extension of F then it is a Galois extension of L.*

Proof. Let p be an irreducible polynomial in $L[x]$ which has a root a in K. We may assume p is monic. By Proposition 12.28, every element of L is algebraic over F, so a has a minimal polynomial q in $F[x]$. Since K is a Galois extension of F, and q has the one root a in K, it follows that q factors into a product of linear polynomials in $K[x]$ with distinct

roots. But p is the minimal polynomial of a in $L[x]$, so p must divide q in $L[x]$. Thus, in $K[x]$, p divides a product of linear polynomials and it must therefore, by unique factorization, itself be a product of some of those linear polynomials in $K[x]$. This shows that p factors in $K[x]$ into a product of linear polynomials with distinct roots. □

I have been saying "F-homomorphism" for a homomorphism that respects embedded copies of some field F. If K is an extension of F, then an *F-automorphism* of K will be an isomorphism of K with itself that fixes every element of F.

Lemma 12.34. *Let F be a compactly coherent field and let K be a finite extension of F, equipped with the standard topology. Then any F-homomorphism from K to itself is a homeomorphic F-automorphism of K.*

Proof. Let $\phi : K \to K$ be an F-homomorphism. Since K is a field, ϕ is automatically injective. This implies that $\phi(K) \subseteq K$ has the same dimension over F as K, and therefore ϕ is also surjective. It is a homeomorphism by Proposition 12.24. □

We can leverage Lemma 12.31 to get a more general fact about Galois extensions.

Lemma 12.35. *Let F be a compactly coherent field, let K be a finite Galois extension of F equipped with the standard topology, and let L be a closed intermediate field. Then every F-homomorphism of L into K extends to exactly $[K : L]$ distinct F-automorphisms of K.*

Proof. We will show that every F-homomorphism from $\phi : L \to K$ extends to exactly $[K : L]$ homomorphisms from K into itself; this suffices by Lemma 12.34. Note that any homomorphism that extends an F-homomorphism is an F-homomorphism.

Let us prove this by induction on $[K : L]$. (Here we quantify over the closed intermediate subfields between F and K.) If $[K : L] = 1$ then $K = L$ and the result is trivial. If $[K : L] > 1$, let $a \in K \setminus L$ and let p and q be its minimal polynomials in $L[x]$ and $F[x]$, respectively, so that p divides q in $L[x]$. Thus the roots of p are among the roots of q, and these all belong to K and are distinct. Lemma 12.31 now assures us that there are exactly n extensions of ϕ to an F-homomorphism from $L[a]$ into K, where n is the degree of p, or equivalently, $n = [L[a] : L]$. Let $\psi_1, \ldots, \psi_n$ be these extensions. By the induction hypothesis we know that there are exactly

$[K : L[a]] = \frac{[K:L]}{n}$ distinct extensions of each ψ_i to a homomorphism from K into itself. Thus there are exactly $n \cdot \frac{[K:L]}{n} = [K : L]$ distinct extensions of ϕ to a homomorphism from K into itself. □

If K is a finite Galois extension of F then the set of F-automorphisms of K (literally, their codes in $\mathcal{L}(K)$) form a group under composition which is finite by Lemma 12.35; it is called the *Galois group* of K over F and denoted "Gal(K/F)". For any closed intermediate field L we define Gal(K/L) to be the subgroup of Gal(K/F) consisting of all L-automorphisms of K. Also, for any subgroup $H \subseteq \text{Gal}(K/F)$ let K^H be the class of elements of K which are fixed by everything in H. This is easily seen to be a closed subfield of K that contains F; it is called the *fixed field* of H.

Our final lemma is the "hard" part of the fundamental theorem of Galois theory (Theorem 12.37).

Lemma 12.36. *Let $F \subseteq K$ be compactly coherent fields and suppose K is a finite Galois extension of F. Let H be a subgroup of* Gal(K/F) *with fixed field K^H. Then $H = \text{Gal}(K/K^H)$.*

Proof. Put $H' = \text{Gal}(K/K^H)$. The inclusion $H \subseteq H'$ is clear. (The elements of H are certainly among the elements of Gal(K/F) that fix K^H pointwise.) Also, according to Lemma 12.35 the identity map on K^H extends to exactly $[K : K^H]$ distinct automorphisms of K, i.e., the group Gal(K/K^H) has order $[K : K^H]$. We will prove that $H = H'$ by showing that the order of H is at least $[K : K^H]$.

To do this, let $n = [K : K^H]$ and choose a basis $\{a_1, \ldots, a_n\}$ of K as a vector space over K^H. Also list the elements of H as ϕ_1, …, ϕ_m, with ϕ_1 being the identity automorphism. Suppose the n vectors $\vec{x}_i = (\phi_1(a_i), \ldots, \phi_m(a_i)) \in K^m$ (for $1 \leq i \leq n$) were linearly dependent over K. Let k be the smallest index for which $\{\vec{x}_1, \ldots, \vec{x}_k\}$ is dependent and find coefficients $c_i \in K$ such that $c_1x_1 + \cdots + c_kx_k = 0$ with at least one c_i nonzero. Actually, c_k has to be nonzero by minimality. Dividing through by this value, we can assume $c_k = 1$.

Looking just at the jth coordinate in K^m, $1 \leq j \leq m$, we then have

$$c_1\phi_j(a_1) + \cdots + c_{k-1}\phi_j(a_{k-1}) + \phi_j(a_k) = 0.$$

With $j = 1$ this becomes $c_1a_1 + \cdots + c_{k-1}a_{k-1} + a_k = 0$; put that aside. For now, let $1 \leq j' \leq m$ be arbitrary, substitute $\phi_{j'}^{-1} \circ \phi_j$ for ϕ_j (since $\phi_{j'}^{-1} \circ \phi_j$ belongs to H and hence appears somewhere in the list ϕ_1, …, ϕ_m), and then apply $\phi_{j'}$ to both sides to get

$$\phi_{j'}(c_1)\phi_j(a_1) + \cdots + \phi_{j'}(c_{k-1})\phi_j(a_{k-1}) + \phi_j(a_k) = 0.$$

Subtracting this from the original equation, we get

$$(c_1 - \phi_{j'}(c_1))\phi_j(a_1) + \cdots + (c_{k-1} - \phi_{j'}(c_{k-1}))\phi_j(a_{k-1}) = 0$$

since the last terms cancel. For fixed j' this means all the coordinates of the x_i's satisfy this relation, i.e.,

$$(c_1 - \phi_{j'}(c_1))x_1 + \cdots + (c_{k-1} - \phi_{j'}(c_{k-1}))x_{k-1} = 0.$$

Since k was chosen to be minimal, this must be a trivial linear combination with all zero coefficients, i.e.,

$$\phi_{j'}(c_j) = c_j,$$

and that is for any $1 \leq j \leq k-1$ and $1 \leq j' \leq m$. Thus each c_j is fixed by every element of H, and must therefore belong to K^H. But now the equation we put aside, $c_1a_1 + \cdots + c_{k-1}a_{k-1} + a_k = 0$, implies a linear dependence of the a_i's over K^H, which is impossible since they constituted a basis. Conclusion (Proposition 12.7): the vectors $\vec{x}_1, \ldots, \vec{x}_n \in K^m$ are linearly independent. Thus $m \geq n$, as we wanted to show. □

Theorem 12.37. *(Fundamental theorem of Galois theory) Let F be a compactly coherent field and let K be a finite Galois extension of F, equipped with the standard topology. Then the map $L \mapsto \mathrm{Gal}(K/L)$ is a bijection between the closed intermediate fields $F \subseteq L \subseteq K$ and the subgroups of $\mathrm{Gal}(K/F)$. Its inverse is the map $H \mapsto K^H$. A closed intermediate field L is a normal extension of F if and only if $\mathrm{Gal}(K/L)$ is a normal subgroup of $\mathrm{Gal}(K/F)$.*

Proof. Let L be a closed intermediate field between F and K. One map takes L to $\mathrm{Gal}(K/L)$ and the other takes $\mathrm{Gal}(K/L)$ back to $K^{\mathrm{Gal}(K/L)}$. Let $L' = K^{\mathrm{Gal}(K/L)}$; we must show that $L = L'$.

The inclusion $L \subseteq L'$ is clear. (The elements of L are certainly among the elements of K that are fixed by everything in $\mathrm{Gal}(K/L)$.) Now let a be an element of $K \setminus L$. The minimal polynomial p of a in $L[x]$ factors into distinct linear polynomials in $K[x]$ since K is a Galois extension of L (Lemma 12.33). By Lemma 12.31 there is an L-homomorphism from $L[a]$ into K which takes a to a different root of p, and by Lemma 12.35 this L-homomorphism extends to an L-automorphism of K, i.e., an element of $\mathrm{Gal}(K/L)$. So a is not fixed by everything in $\mathrm{Gal}(K/L)$ and therefore does not belong to L'. We have shown that $L = L'$.

Lemma 12.36 completes the proof that we have a bijection between the closed intermediate fields between F and K and the subgroups of $\mathrm{Gal}(K/F)$.

Now suppose L is a normal extension of F. For any $a \in L$, since L is normal all the roots of the minimal polynomial p of a in $F[x]$ belong to L. Then since every automorphism in $\mathrm{Gal}(K/F)$ fixes F pointwise, every such automorphism fixes the coefficients of p and hence must permute the roots of p. In particular, every such automorphism takes a to another element of L. We have shown that every F-automorphism of K takes L into itself. Now let $\phi \,\varepsilon\, \mathrm{Gal}(K/L)$ and $\psi \,\varepsilon\, \mathrm{Gal}(K/F)$. Then ψ^{-1} also belongs to $\mathrm{Gal}(K/F)$, so for any $a \in L$, as we have just seen, we have $\psi^{-1}(a) \in L$. Since ϕ fixes everything in L we therefore have $\psi \circ \phi \circ \psi^{-1}(a) = \psi \circ \psi^{-1}(a) = a$. This shows that $\psi \circ \phi \circ \psi^{-1}$ fixes L pointwise and so it belongs to $\mathrm{Gal}(K/L)$. Since ϕ and ψ were arbitrary, we see that $\mathrm{Gal}(K/L)$ is a normal subgroup of $\mathrm{Gal}(K/F)$.

Conversely, suppose $\mathrm{Gal}(K/L)$ is a normal subgroup of $\mathrm{Gal}(K/F)$ and let $a \in L$. Let p be the minimal polynomial of a in $F[x]$; then p factors in $K[x]$ into a product of linear polynomials with distinct roots. Suppose one of these roots a' did not belong to L. By Lemmas 12.31 and 12.35 there exists an F-automorphism ψ of K which takes a to a'. Then since $a' \notin L$, the first part of this theorem says that there exists an L-automorphism ϕ of K that does not fix a', i.e., $a'' \neq a'$ where $a'' = \phi(a')$. We therefore have

$$\psi^{-1} \circ \phi \circ \psi(a) = \psi^{-1} \circ \phi(a') = \psi^{-1}(a''),$$

and this cannot equal $a = \psi^{-1}(a')$ because $a' \neq a''$ and ψ is an automorphism. Thus $\psi^{-1} \circ \phi \circ \psi$ is not an L-automorphism of K, contradicting the hypothesis that $\mathrm{Gal}(K/L)$ is normal. Since every root of p definitely does or does not belong to L, we conclude that every root of p must belong to L. We have shown that L is a normal extension of F. □

Now let us discuss algebraic closures. A field K is *algebraically closed* if every nonconstant polynomial with coefficients in K has a root in K. This entails that every nonconstant polynomial in $K[x]$ has a linear factor, so every irreducible nonconstant polynomial must be linear, which shows that every nonconstant polynomial in $K[x]$ is a product of linear factors.

An *algebraic closure* of a field F is an algebraically closed field K together with a homomorphism $\phi : F \to K$ such that $\phi(F)$ exists and every element of K is algebraic over $\phi(F)$. Again, since all homomorphisms are assumed to be unital and F is a field, ϕ must be injective.

The most important example of an algebraically closed field is the complex plane. The proof of this is actually quite easy, modulo some estimates which are routine though tedious to write out explicitly. Recall from Section 9.9 that if $z = |z|(\cos\theta + i\sin\theta)$, then $z^k = |z|^k(\cos(k\theta) + i\sin(k\theta))$.

Thus, for fixed $|z|$, as θ varies from 0 to 2π the values of z lie on a circle of radius $|z|$ about the origin, and the values of z^k go k times around a circle of radius $|z|^k$ about the origin. (We will treat complex exponential functions rigorously in Section 13.4, and the proof of Theorem 12.38 may be deferred until then.)

Theorem 12.38. *(Fundamental theorem of algebra)* $\mathbb{C}$ *is algebraically closed.*

Proof. Let $p(z) = a_0 + a_1 z + \cdots + a_n z^n \in \mathbb{C}[x]$ be a complex polynomial with $n \geq 1$ and $a_n \neq 0$. We must show that it has a root. If $a_0 = 0$ then $z = 0$ is a root; otherwise, divide p by a_0 to make the constant term 1. This of course does not affect the existence of roots.

Consider the modulus $|p(z)|$ of $p(z)$. For large values of z, the leading term $a_n z^n$ dominates all other terms; in particular, there is some radius R such that $|a_n z^n| \geq 2 + |1 + a_1 z + \cdots + a_{n-1} z^{n-1}|$ for all $|z| \geq R$. Thus $|p(z)| \geq 2$ for all $|z| \geq R$. The closed disk $D_R = \{z : |z| \leq R\}$ is compact (the Heine-Borel theorem, Corollary 10.46), so according to the extreme value theorem (Theorem 10.36) there is a value $w \in D_R$ with $|p(w)| \leq |p(z)|$ for all $z \in D_R$. Since $|p(0)| = 1$ and $|p(z)| \geq 2$ when $|z| = R$, we must have $|w| < R$.

Now "expand p about w" and write $p(z) = b_0 + b_1(z-w) + \cdots + b_n(z-w)^n$. We want to show that w is a root of p, i.e., $b_0 = 0$. If not, let $k \geq 1$ be the smallest index with $b_k \neq 0$. (There must be some nonzero coefficient after b_0 since p was assumed nonconstant.) Now, for any $\epsilon > 0$, as z ranges over the circle $\{w + z_0 : |z_0| = \epsilon\}$ the value $b_0 + b_k(z-w)^k = b_0 + b_k z_0^k$ ranges over a circle about b_0 which wraps k times around b_0. But for sufficiently small $|z_0|$ the term $b_k(z-w)^k = b_k z_0^k$ dominates all higher order terms, so assuming $b_0 \neq 0$, for sufficiently small ϵ there exists z_0 with $|z_0| = \epsilon$ and $|p(w + z_0)| < |b_0| = |p(w)|$. This contradicts the minimality of $|p(w)|$, so we must have $b_0 = 0$. □

The following lemma shows that in the definition of algebraic closure, the condition that every polynomial with coefficients in K must have a root in K can be replaced by the condition that every polynomial with coefficients in $\phi(F) \subseteq K$ must be a product of linear factors in $K[x]$.

Lemma 12.39. *Let F and K be compactly coherent fields and let $\phi : F \to K$ be a continuous homomorphism such that every element of K is algebraic over $\phi(F)$. Suppose every polynomial with coefficients in $\phi(F)$ is a product of linear factors in $K[x]$. Then K is algebraically closed.*

Proof. We need every polynomial in $K[x]$ to have a root in K. We will show that every monic irreducible in $K[x]$ is linear. This suffices because linear polynomials always have roots, and every polynomial can be factored into irreducibles.

Fix a monic irreducible $p \in K[x]$ and let $F' = \phi(F)$. Since everything in K is algebraic over F', we can find a subfield $L \subseteq K$ which is a finite extension of F', such that all the coefficients of p lie in L; this was essentially done in the proof of Corollary 12.29. Thus $p \in L[x]$, and irreducibility in $K[x]$ clearly implies irreducibility in $L[x]$. Give L the standard topology. By Proposition 11.50, $L[x]/(p)$ is then a compactly coherent field and there is a natural continuous homomorphism $\iota : L \to L[x]/(p)$. The residue $\overline{x}$ of x is a root of $\iota(p)$.

Now $L[x]/(p)$ is a finite extension of $\iota(L)$ — the dimension of the extension equals the degree of p — so by Proposition 12.23 it is a finite extension of $\iota(F') \subseteq \iota(L)$. Thus every element of $L[x]/(p)$ is algebraic over $\iota(F')$. In particular, $\overline{x}$ has a minimal polynomial over $\iota(F')$, say $\iota(q)$ with $q \in F'[x]$, as well as a minimal polynomial over $\iota(L)$, namely $\iota(p)$. Since $F' \subseteq L$, this shows that p divides q in $L[x]$, and hence p divides q in $K[x]$. By hypothesis, q is a product of linear factors in $K[x]$, so by Lemma 11.48 the irreducibility of p in $K[x]$ entails that it is linear. □

Every countable field has an algebraic closure. In the following proof, we construct a sequence of countable fields F_i together with connecting homomorphisms. One simple way to formalize this would be to work with $\mathbb{N} \times \mathbb{N}$, treating each copy of the first factor as one of the F_i's. The field operations and the connecting homomorphisms can be categorized by separate families indexed by $\mathbb{N}$. Also, note that saying two countable fields K and K' are isomorphic only needs to quantify over the pseudoclass $\mathrm{fn}(K, K')$, i.e., it can be taken to involve a bounded existential quantifier.

Theorem 12.40. *(Existence and uniqueness of algebraic closures) Let F be a countable field. Then it has an algebraic closure. Every algebraic closure of F is countable, and any two algebraic closures of F are F-isomorphic.*

Proof. First we construct an algebraic closure. Enumerate the nonconstant polynomials in $F[x]$ as p_1, p_2, ... Put $F_1 = F$, and for each i use Lemma 12.30 to get a field F_{i+1} and a homomorphism $\phi_i : F_i \to F_{i+1}$ such that F_{i+1} is a finite extension of $\phi_i(F_i)$ and $\widetilde{\phi}_i(p_i)$ is a product of linear factors in $F_{i+1}[x]$, where $\widetilde{\phi}_i = \phi_i \circ \cdots \circ \phi_1$. The direct limit $K = \varinjlim F_i$ is then a field every element of which is algebraic over $\phi(F)$, where ϕ is the

natural map from F into K, and every polynomial in $\phi(F[x])$ is a product of linear factors in $K[x]$. It follows from Lemma 12.39 that K is an algebraic closure of F.

Clearly K is countable, so all we have left to show is that K is F-isomorphic to any algebraic closure K' of F, relative to a homomorphism $\phi' : F \to K'$. Enumerate the elements of K as $a_1, a_2, \ldots$. Since K' contains roots of everything in $\phi'(F[x])$, by Lemma 12.31 we can find a subfield K_1 of K containing $\phi(F)$ and the element a_1, together with an F-homomorphism $\sigma_1 : K_1 \to K'$. Then we can find a subfield $K_2 \subseteq K$ containing K_1 and the element a_2, and a K_1-homomorphism $\sigma_2 : K_2 \to K'$ which extends σ_1. Continuing in this way, we constuct an F-homomorphism $\sigma : K \to K'$. Since K is a field this map is automatically injective. For surjectivity, note that every element a of K' is algebraic over $\phi'(F) = \sigma(\phi(F))$ and has a minimal polynomial over $\sigma(\phi(F))$ which is a product of linear factors in $K'[x]$. Write this minimal polynomial as $\sigma(p)$, where p is a polynomial over $\phi(F)$. Then $p = (x - a_1) \cdots (x - a_n)$ where $a_1, \ldots, a_n \in K$ and $n = \deg(p)$, so $\sigma(p) = (x - \sigma(a_1)) \cdots (x - \sigma(a_n))$. This yields a complete factorization of $\sigma(p)$ in $K'[x]$, so we must have $a = a_i$ for some i. Thus a lies in $\sigma(K)$ and we have shown that σ is surjective. □

Because of this uniqueness up to isomorphism, we refer to "the" algebraic closure of a countable field.

12.4 Modules

Let R be a unital ring. *R-modules* are defined in exactly the same way as F-vector spaces (Definition 12.1), with R taking the place of the field F. We have notions of homomorphism, submodule, product module, and quotient module which are so similar to the corresponding vector space concepts that they needn't be spelled out. A *compactly coherent* module over a compactly coherent ring R is an R-module equipped with a compactly coherent topology with respect to which the module operations are continuous.

One important difference between modules and vector spaces is that rings can be noncommutative. In that case what we are defining is a *left* R-module.

Example 12.41. Any unital ring R is a module over itself. If R is commutative, the R-submodules of R (as an R-module) are precisely the ideals

of R (as a ring). If R is a PID then every nonzero separable submodule of R is isomorphic to R (by the map $x \mapsto ax$ from R to aR).

Example 12.42. Let M be a $\mathbb{Z}$-module. Then M is, by definition, an abelian group equipped with a map $\cdot : \mathbb{Z} \times M \to M$ which satisfies the conditions of Definition 12.1.

From distributivity and the law $1 \cdot x = x$ we get

$$2 \cdot x = (1 + 1) \cdot x = 1 \cdot x + 1 \cdot x = x + x.$$

This generalizes to the statement that $n \cdot x = x + \cdots + x$ (n terms), for any $n \in \mathbb{N}$. Also,

$$0 \cdot x = (0 + 0) \cdot x = 0 \cdot x + 0 \cdot x,$$

showing that $0 \cdot x = 0$; and for each $n \in \mathbb{N}$ we then have

$$n \cdot x + (-n) \cdot x = (n - n) \cdot x = 0 \cdot x = 0,$$

showing that $(-n) \cdot x = -(n \cdot x) = -x - \cdots - x$ (n terms). In other words, the module laws completely determine the action of $\mathbb{Z}$ on any $\mathbb{Z}$-module in terms of its additive abelian group structure.

Conversely, it is straightforward to check that every abelian group, equipped with the product $n \cdot x$ just described for any $n \in \mathbb{Z}$, is a $\mathbb{Z}$-module. Thus the notion of a $\mathbb{Z}$-module is functionally identical to the notion of an abelian group. Homomorphisms of $\mathbb{Z}$-modules are just homomorphisms of abelian groups, submodules are subgroups, and quotient modules are quotient groups.

We have a general structure theorem for modules over a PID. It involves the following notions.

Definition 12.43. A module M over a ring R is *cyclic* if there exists $x \in M$ such that every element of M equals ax for some $a \in R$. A *generating set* for M is a subset $S \subseteq M$ such that every element of M equals $a_1x_1 + \cdots + a_nx_n$ for some $a_1, \ldots, a_n \in R$ and $x_1, \ldots, x_n \in S$. M is *finitely generated* if it has a finite generating set.

If there are submodules M_1, ..., M_n of M such that every element of M can be written in exactly one way as $x_1 + \cdots + x_n$ with each $x_i \in M_i$, then the M_i constitute an *internal direct sum*. In this case there is a natural isomorphism between M and the product $M_1 \times \cdots \times M_n$.

For any $x \in M$, if there exists a nonzero element $a \in R$ with $ax = 0$ then x is a *torsion element*. If every element of M is a torsion element then M is a *torsion module*; if 0 is the only torsion element then M is

torsion-free. If M is isomorphic to R^n for some $n \in \mathbb{N}$ —equivalently, if there exist elements $x_1, \ldots, x_n \in M$ such that every element of M can be uniquely written as a linear combination of the x_i's — then it is a *finitely generated free module*.

On its face, the definition of finitely generated free modules involves an unbounded existential quantifier (there exists an isomorphism with some R^n); the reframing in terms of $x_1, \ldots, x_n \in M$ turns that into a bounded quantifier.

Recall that it is part of our definition of a PID (Definition 11.42) that it must be equipped with a compactly coherent topology. If R is a PID, M is a compactly coherent R-module, and $x \in M$ is a torsion element, then the class of $a \in R$ for which $ax = 0$ — the kernel of the continuous map $a \mapsto ax$ from R into M — is a nonzero closed (and hence principal) ideal, so there exists $b \in R$ such that $ax = 0$ if and only if $b|a$. This b is unique up to multiplication by an invertible; we call it the *order* of x. The elements of R whose product with every $y \in M$ is zero constitute a closed, hence principal, ideal, and if this ideal is nonzero its generator (up to multiplication by an invertible) is the *order* of M. If x is not torsion, or there are no nonzero elements of R whose product with every $y \in M$ is zero, then we will say the order of x is infinity, or the order of M is infinity, respectively.

We will show that every finitely generated compactly coherent module over a PID is an internal direct sum of cyclic submodules. We require several preliminary lemmas.

Lemma 12.44. *Every finitely generated submodule of a finitely generated free module over a PID is a finitely generated free module.*

Proof. Let R be a PID and let M be a finitely generated R-submodule of R^n. Then for each $0 \leq i \leq n$ let $N_i = \{(r_1, \ldots, r_i, 0, \ldots, 0) : r_1, \ldots, r_i \in R\}$. Giving R^n the coherent product topology, N_i is a closed submodule of R^n that is isomorphic to R^i, and $M \cap N_n = M \cap R^n = M$. Put $M_i = M \cap N_i$ for all i. Let $S = \{x_1, \ldots, x_k\}$ be a finite generating set for M and define a continuous function $\phi : R^k \to R^n$ by putting $\phi(a_1, \ldots, a_k) = a_1x_1 + \cdots + a_kx_k$. Let $\widetilde{M}_i = \phi^{-1}(N_i)$ for all i; then $\widetilde{M}_i$ is a closed submodule of R^k and ϕ takes $\widetilde{M}_i$ onto M_i.

We prove that M_i is free by induction on i. M_0 is the zero module, so the base case is trivial. Now suppose M_i is free. Letting π_{i+1} be the projection from R^n onto its $(i+1)$st factor, we get that $\pi_{i+1} \circ \phi$ takes $\widetilde{M}_{i+1}$

into R. By compact coherence, the image of $\widetilde{M}_{i+1}$ (which equals the image of M_{i+1} under π_{i+1}) exists as a separable submodule of R. But since R is a PID, this image must be a principal ideal. Thus M_{i+1}/M_i is generated by a single element, say the element $[x]$ instantiated by $x \in M_{i+1}$. If this element is zero then $M_{i+1} = M_i$, which is free by hypothesis. Otherwise I claim M_{i+1} is the internal direct sum of M_i and Rx, the submodule generated by x. This will suffice because R is isomorphic to Rx by the map $a \mapsto ax$.

Any $y \in M_{i+1}$ satisfies $[y] = b[x]$ for some $b \in R$, and then $bx \in Rx$ and $[y - bx] = 0$ so $y - bx \in M_i$. So y is the sum of something in M_i and something in Rx. And if $z + cx = 0$ for some $z \in M_i$ and $c \in R$ then $c[x] = [z + cx] = 0$ implies $c = 0$ (R^n is torsion-free), and that implies $z = 0$. This proves the claim. □

Lemma 12.45. *Every finitely generated torsion-free module over a PID is a finitely generated free module.*

Proof. Let M be a finitely generated torsion-free module over a PID R and let $\{x_1, \ldots, x_n\}$ be a generating set for M. For any subset $S \subseteq \{1, \ldots, n\}$, the vectors $\{x_i : i \in S\}$ definitely do or do not satisfy a nontrivial relation (a linear combination of the x_i's with at least one nonzero coefficient, which evaluates to 0). This can be seen using compact coherence of R by an argument that should be familiar. (If there were a nontrivial relation with coefficients a_i then there would be coefficients in a dense subset of R which converge to the a_i and are all contained within a compact subspace of R, such that the corresponding linear combinations converge to 0, etc.) Thus there exists a maximal subset S of $\{1, \ldots, n\}$ such that the elements x_i for $i \in S$ satisfy no nontrivial linear relation. Rearrange the indices so that $S = \{1, \ldots, k\}$ for some $k \leq n$.

We have an injective homomorphism $\phi : R^k \to M$ defined by $\phi(a_1, \ldots, a_k) = a_1x_1 + \cdots + a_kx_k$. Now for all i there exists a nonzero ring element $r_i \in R$ such that r_ix_i is a linear combination of x_1, …, x_k (with $r_i = 1$ for $i \leq k$). Let $r = r_1 \cdots r_n$ be their product. Then rx is a linear combination of x_1, …, x_k, for all $x \in M$. So we can define $\psi : M \to R^k$ by $\psi(x) = \phi^{-1}(rx)$. This map is injective and thus takes M isomorphically into R^k. And $\psi(M)$ exists as a subclass of R^k, namely the span of $\psi(x_1)$, …, $\psi(x_n)$ (which exists by a familiar compact coherence argument). By Lemma 12.44 $\psi(M)$, and therefore also M, is a finitely generated free module. □

Lemma 12.46. *Let M be a finitely generated submodule of a compactly*

coherent module over a PID R. Suppose the order of M is a power of some prime $p \in R$. Then M is an internal direct sum of cyclic submodules each of whose orders is a power of p.

Proof. Let $\{x_1, \ldots, x_n\}$ be a generating set and for each i let k_i be the smallest natural number for which $p^{k_i} x_i = 0$. Assume none of the x_i's is zero, and reorder so that $k_1 \leq \cdots \leq k_n$. We will prove that there exist z_1, ..., $z_n \in M$ such that (1) for each i the order of z_i is a power of p, at most p^{k_i}, and (2) M is the internal direct sum of Rz_1, ..., Rz_n. Here the Rz_i exist as submodules of M by compact coherence, plus the fact that the map $a \mapsto az_i$ is continuous from R into M.

The proof goes by induction on the sum $k = k_1 + \cdots + k_n$, which we call the *height* of the generating set $\{x_1, \ldots, x_n\}$. If $n = 1$ then the result is trivial. So let $n > 1$ and assume the result is true for all submodules of M with generating sets of height less than k. (This statement quantifies over all finite subsets of M, which is a bounded quantification in the sense of Section 9.5.) Let M' be the span of x_2, ..., x_n = the image of R^{n-1} under the map $\vec{a} \mapsto a_1 x_2 + \cdots + a_{n-1} x_n$. Now M' has a generating set with smaller height, namely $\{x_2, \ldots, x_n\}$, so by the induction hypothesis there exist $y_2, \ldots, y_n \in M'$ such that $p^{k_i} y = 0$ for all i and M' is the internal direct sum of Ry_2, ..., Ry_n. Thus the height of $\{y_2, \ldots, y_n\}$ is at most the height of $\{x_2, \ldots, x_n\}$. But if it were strictly smaller, then $\{x_1, y_2, \ldots, y_n\}$ would be a generating set of M with height strictly less than k, so we could draw the desired conclusion by the induction hypothesis. Therefore we can assume the height of $\{y_2, \ldots, y_n\}$ equals the height of $\{x_2, \ldots, x_n\}$, and thus p^{k_i} is precisely the order of y_i for $1 \leq i \leq n$.

Let j be the smallest natural number such that $p^j x_1 \in M'$. So $j \leq k_1$, and for any $a \in R$ we have $ax_1 \in M'$ if and only if $p^j | a$. Now if $p^j x_1 = a_2 y_2 + \cdots + a_n y_n$ then

$$0 = p^{k_2 - j} p^j x_1 = p^{k_2 - j}(a_2 y_2 + \cdots + a_n y_n)$$

and thus $p^{k_2 - j} a_2 y_2 = \cdots = p^{k_2 - j} a_n y_n = 0$ (using the fact that M' is the internal direct sum of Ry_2, ..., Ry_n). Thus $a_i = p^j b_i$ for some $b_2, \ldots, b_n \in R$, and $p^j x_1 = p^j(b_2 y_2 + \cdots + b_n y_n)$. Let $y' = b_2 y_2 + \cdots + b_n y_n$. We will complete the proof by showing that M is the internal direct sum of Rx' and M', where $x' = x_1 - y'$. First, anything in M can be written as a sum of the form $ax_1 + y$ where $a \in R$ and $y \in M'$. This can be rewritten as $ax' + (ay' + y) \in Rx' + M'$. So everything in M is the sum of something in Rx' and something in M'. Finally, suppose $ax' + y = 0$ for some $a \in R$ and $y \in M'$. Then $ax_1 = a(x' + y') = ay' - y \in M'$, so $p^j | a$.

But $p^j x' = p^j x_1 - p^j y' = 0$, so this shows that $ax' = 0$. Thus $y = 0$ as well and we have shown that M is the internal direct sum of Rx' and M'. □

Lemma 12.47. *Every finitely generated torsion submodule of a compactly coherent module over a PID is an internal direct sum of cyclic submodules M_i each of which has order $p_i^{k_i}$ for some prime p_i and natural number k_i.*

Proof. Let M be such a module, over a PID R. Let $\{x_1, \ldots, x_n\}$ be a finite set of generators and for each i find $a_i \in R$ such that $a_i x_i = 0$. Letting a be the product of all the a_i's, we see that $aM = 0$.

Since R is a PID, it is a UFD (Theorem 11.49). Thus we can write a as a product of irreducibles, which we also know are primes (Lemma 11.48). Collecting associate primes, we can write $a = up_1^{\alpha_1} \cdots p_k^{\alpha_k}$ where u is invertible and none of the p_i's is an associate of any of the others. We can assume $u = 1$. For each i let b_i be the product $p_1^{\alpha_1} \cdots p_k^{\alpha_k}$ with the $p_i^{\alpha_i}$ factor omitted, and put $M_i = b_i M$. This submodule of M exists by compact coherence; it is generated by the finite set $\{b_i x_1, \ldots, b_i x_n\}$.

I claim that M is the internal direct sum of the M_i's. The ideal of R generated by the b_i's has the form cR with $c|b_i$ for all i. By unique factorization, c must be invertible. Thus 1 belongs to the ideal, so we can find $c_1, \ldots, c_n \in R$ such that $b_1 c_1 + \cdots + b_n c_n = 1$. Any $x \in M$ satisfies

$$x = (b_1 c_1 + \cdots + b_n c_n)x,$$

showing that the sum $M_1 + \cdots + M_n$ exists and equals M. Finally, we must show that if $b_1 y_1 + \cdots + b_n y_n = 0$ for some $y_1, \ldots, y_n \in M$, then $b_1 y_1 = \cdots = b_n y_n = 0$. To do this, make an argument similar to the one given above, now looking at the ideal generated by p_1^α and b_1, to find $d, d' \in R$ such that $dp_1^\alpha + d'b_1 = 1$. Then

$$b_1 y_1 = (dp_1^\alpha + d'b_1)b_1 y_1 = ady_1 - d'b_1(b_2 y_2 + \cdots + b_n y_n) = 0$$

since $a|b_1 b_i$ for all $i \neq 1$. This shows that $b_1 y_1 = 0$, and the same argument can be used to show that $b_2 y_2$, …, $b_n y_n$ are all zero. This proves the claim.

We can now invoke Lemma 12.46 to get that each M_i is an internal direct sum of cyclic submodules whose orders are powers of p_i, and conclude that M is an internal direct sum of cyclic submodules whose orders are prime powers. □

Lemma 12.48. *Let M be a a finitely generated compactly coherent module over a PID. Then there are submodules $M_1, M_2 \subseteq M$ such that M_1 is finitely generated and torsion, M_2 is finitely generated and free, and M is the internal direct sum of M_1 and M_2.*

Proof. Let R be the PID and let M_1 be the class of torsion elements of M (compact coherence). It is easy to see that M_1 is a submodule of M. Also, M/M_1 is torsion-free because if $a \cdot [x] = 0$ with $a \neq 0$, where $[x]$ is the element of M/M_1 instantiated by $x \in M$, then $ax \in M_1$, hence there exists a nonzero $b \in R$ with $bax = 0$, which means that $x \in M_1$ and hence $[x]$ is the zero element of M_1. And M/M_1 is finitely generated because any quotient of a finitely generated module is finitely generated. M/M_1 is then a finitely generated free module by Lemma 12.45. Say $M/M_1 \cong R^k$, and for each $1 \leq i \leq k$ let $[x_i]$ be the element of M/M_1 corresponding to the element of R^k which has a 1 in the ith coordinate and 0's in all the other coordinates.

Let M_2 be the submodule of M generated by $x_1, \ldots, x_k$ (compact coherence). Then M_2 is the internal direct sum of $Rx_1, \ldots, Rx_k$ because $a_1x_1 + \cdots + a_kx_k = 0$ implies that $a_1[x_1] + \cdots + a_k[x_k] = 0$, which implies that $a_1 = \cdots = a_k = 0$. Each Rx_i is isomorphic to R since the x_i are not torsion elements. So M_2 is a finitely generated free module.

To see that M is the internal direct sum of M_1 and M_2, first observe that the quotient map takes M_2 isomorphically onto M/M_1, so $M_1 \cap M_2 = \{0\}$, and for any $x \in M$ there exists $y \in M_2$ with $[x] = [y]$ in M/M_1, i.e., $x - y \in M_1$. Then $x = (x - y) + y$, so x is a sum of something in M_1 and something in M_2. This establishes that M is the internal direct sum of M_1 and M_2. Finally, M_1 is finitely generated because $M_1 \cong M/M_2$ and any quotient of a finitely generated module is finitely generated. □

Combining Lemmas 12.48, 12.47, and 12.45 yields the following.

Theorem 12.49. *(Structure theorem for finitely generated modules over a PID) Let M be a finitely generated compactly coherent module over a PID. Then M is an internal direct sum of cylic submodules each of whose order is a prime power or infinity.*

The next result is an immediate consequence (cf. Example 12.42).

Corollary 12.50. *(Classification of finitely generated abelian groups) Every finitely generated abelian group is isomorphic to a finite product of groups, each of which is either $\mathbb{Z}$ or $\mathbb{Z}/p^k\mathbb{Z}$ for some prime power p^k.*

A standard class of examples of modules over $\mathbb{C}[x]$ arises as follows. Let $T : \mathbb{C}^n \to \mathbb{C}^n$ be a linear transformation. Then for any polynomial $p(x) = a_0 + a_1x + \cdots + a_kx^k \in \mathbb{C}[x]$ and any $\vec{v} \in \mathbb{C}^n$ define $p(x) \cdot \vec{v} =$

$a_0\vec{v} + a_1 T\vec{v} + \cdots + a_k T^k\vec{v}$. This makes $\mathbb{C}^n$ into a compactly coherent $\mathbb{C}[x]$-module.

Information about the structure of T can be obtained by applying the structure theorem for finitely generated modules over a PID (Theorem 12.49) to this module. However, in this case we can directly prove a structure theorem which is even better in that it not only decomposes $\mathbb{C}^n$ into an internal direct sum of cyclic submodules of prime power orders, but also gives us a basis of $\mathbb{C}^n$ with respect to which the matrix of T has a very simple form.

Fix T and let $\lambda \in \mathbb{C}$. A λ-*Jordan chain* of length m is a sequence of nonzero vectors $\vec{v}_1, \ldots, \vec{v}_m$ such that $T\vec{v}_1 = \lambda\vec{v}_1$ and $T\vec{v}_i = \lambda\vec{v}_i + \vec{v}_{i-1}$ for $2 \leq i \leq m$.

The matrix of T restricted to the span of $\{\vec{v}_1, \ldots, \vec{v}_m\}$ has the following form: all the diagonal entries are λ, all the entries of the first superdiagonal are 1, and all other entries are 0. This is a λ-*Jordan matrix*. If we can find a set of λ-Jordan chains, for varying values of λ, such that $\mathbb{C}^n$ is the internal direct sum of their spans, then their disjoint union is a basis for $\mathbb{C}^n$ and the matrix of T with respect to this basis is block diagonal, with each block having the form just described. Such a matrix is said to be in *Jordan form*. We will prove that this can be done for any linear transformation, not only from $\mathbb{C}^n$ to itself, but from F^n to itself, for any algebraically closed field F.

This result improves on the conclusion of Theorem 12.49 in that if $\vec{v}_1$, $\ldots$, $\vec{v}_m$ is a λ-Jordan chain then the subspace spanned by the $\vec{v}_i$'s is cyclic as an $F[x]$-module in the manner described above, with generator $\vec{v}_m$, of order $(x-\lambda)^n$ which is a prime power in $F[x]$. So we get a decomposition of the type described in Theorem 12.49.

Theorem 12.51. *(Jordan forms) Let F be an algebraically closed field, let $n \in \mathbb{N}$, and let $T : F^n \to F^n$ be a linear transformation. Then there is a set of disjoint λ-Jordan chains in F^n, for varying values of $\lambda \in F$, whose union constitutes a basis for F^n.*

Proof. The proof goes by induction on n. The result is trivial when $n = 1$. Now let $n > 1$ and assume the theorem is true in all smaller dimensions.

Let $[T]$ be the matrix of T with respect to the standard basis for F^n. The determinant $\det([T - xI])$ is then a polynomial in x of degree n with coefficients in F, and since F is algebraically closed this polynomial has a root λ in F. Then $\det([T-\lambda I]) = 0$, showing that the linear transformation $T' = T - \lambda I$ has nonzero kernel (Theorem 12.22). Any nonzero element v

of that kernel is an *eigenvector* of T corresponding to the *eigenvalue* λ; it is characterized by the property that $Tv = \lambda v$.

Let $V \subset F^n$ be the range of T'. Since T' has nonzero kernel, V has dimension strictly less than n (the rank-nullity theorem, Theorem 12.18), and clearly $T'(V) \subseteq V$. By the induction hypothesis, we can find a basis B for V which is a disjoint union of r distinct λ-Jordan chains relative to the restriction $T'|_V$, for varying values of λ. Every chain in this basis looks like $\vec{v}_1, \ldots, \vec{v}_k$ for some value of k, and since it lies in $V = \text{im}(T')$ there exists $\vec{v}_{k+1} \in F^n$ such that $T'\vec{v}_{k+1} = \vec{v}_k$. Also, the first element $\vec{v}_1$ of each chain lies in $\ker(T')$, so we can extend those first elements to a basis for $\ker(T')$ by adding $s = \dim(\ker(T')) - r$ vectors $\vec{w}_1, \ldots, \vec{w}_s$ (each of which constitutes a length 1 Jordan chain for T'). Now let B' be the union of B, the vectors $\vec{w}_j$, and the vectors $\vec{v}_{k+1}$ (one for each of the chains which collectively form a basis for B). This set has

$$\begin{aligned}\dim(V) + s + r &= \dim(\text{im}(T') + s + r \\ &= \dim(\text{im}(T')) + \dim(\ker(T')) = n\end{aligned}$$

elements (the rank-nullity theorem, Theorem 12.18).

Each Jordan chain contained in B can be extended to a one-element-longer Jordan chain by including the extra vector $\vec{v}_{k+1}$. These are Jordan chains relative to T', but that means they are also Jordan chains relative to T. So B' is a union of Jordan chains relative to T, and it contains n elements, so we just have to show that its span is F^n. Let $W = \text{span}(B')$. Then W contains the kernel of T', and the image of $T'|_W$ is V, the range of T', so $T'|_W$ has both the same kernel and the same image as T'. Thus $\dim(W) = \dim(\ker(T')) + \dim(\text{im}(T')) = n$, showing that B' does indeed span F^n. □

The induction hypothesis in this proof which says that the result is true for all linear transformations on lower dimensional vector spaces quantifies over $\mathcal{L}(F) \sqcup \cdots \sqcup \mathcal{L}(F^{n-1})$; this is a bounded quantification and therefore can be used in the induction argument.

Chapter 13

Real and complex analysis

13.1 Differentiation

If X and Y are separable metric spaces, $f : X \to Y$ is a function, and x and y are elements of X and Y, then we write $\lim_{u\to x} f(u) = y$ if for every $\epsilon > 0$ there exists $\delta > 0$ such that $0 \neq d(u, x) < \delta$ implies $d(f(u), y) < \epsilon$. We say that f is *continuous at* x if $\lim_{u\to x} f(u) = f(x)$. Thus f is continuous on X if and only if it is continuous at every point of X (Theorem 10.20 (c)).

Now suppose $A \subseteq \mathbb{R}$ contains an open interval about $x \in \mathbb{R}$ and let $f : A \to \mathbb{R}$ be a function. We say that f is *differentiable at* x if the limit

$$\lim_{h\to 0} \frac{f(x+h) - f(x)}{h} = \lim_{u\to x} \frac{f(u) - f(x)}{u - x}$$

exists, and we define the *derivative of* f *at* x, denoted $f'(x)$, to be this limit. If f is continuous on A, then we can restrict to rational h in this definition. Thus the derivative of a continuous function definitely either does or does not exist at each point. Discontinuous functions, on the other hand, do not have derivatives, i.e., if f is differentiable at x then it must be continuous at x. I omit the easy proof of the following proposition.

Proposition 13.1. *For any $n \in \mathbb{N}$, the derivative of the function x^n on $\mathbb{R}$ is nx^{n-1}. If f and g are both differentiable on (a, b), and $\alpha, \beta \in \mathbb{R}$, then $\alpha f + \beta g$ and fg are differentiable on (a, b), with derivatives $\alpha f' + \beta g'$ and $f'g + fg'$.*

Let $a, b \in \mathbb{R}$, $a < b$, and let $f : [a, b] \to \mathbb{R}$ be a continuous function. If f attains its maximum or minimum value at $c \in (a, b)$ and is also differentiable at c, then we must have $f'(c) = 0$. Any other value of the derivative would imply that there are points arbitrarily close to c on one side at which the value of f is strictly larger, and points arbitrarily close to c on the other

side at which the value of f is strictly smaller. This yields the following result.

Proposition 13.2. *(Rolle's theorem) Let $f : [a, b] \to \mathbb{R}$ be continuous and suppose f is differentiable on (a, b) and $f(a) = f(b)$. Then there exists $c \in (a, b)$ with $f'(c) = 0$.*

If f is constant on $[a, b]$ then any $c \in (a, b)$ would work; otherwise there is either a maximum value $f(c)$ strictly larger than, or a minumum value $f(c)$ strictly smaller than, $f(a) = f(b)$, at which the derivative of f must vanish. And a continuous function definitely is or is not constant, since we can check this on $[a, b] \cap \mathbb{Q}$.

Corollary 13.3. *(Mean value theorem) Let $f : [a, b] \to \mathbb{R}$ be continuous and suppose f is differentiable on (a, b). Then there exists $c \in (a, b)$ with $f'(c) = \frac{f(b)-f(a)}{b-a}$.*

This is proven by applying Rolle's theorem to the function $g(x) = f(x) - r(x - a)$ where $r = \frac{f(b)-f(a)}{b-a}$.

One simple consequence of the mean value theorem is that if f is differentiable on (a, b) and its derivative extends to a continuous function on $[a, b]$, then $\left|\frac{f(x)-f(y)}{x-y}\right| \leq \sup_{u\in(a,b)} |f'(u)| < \infty$ for any distinct $x, y \in [a, b]$. That is, f is Lipschitz on $[a, b]$.

The mean value theorem can be generalized to take higher order derivatives into account. The *nth order derivative* of a function f is the function $f^{(n)}$ obtained by taking the derivative n times, assuming that this is possible. If it is, we say that f is *n times differentiable*. The 0th order derivative of f is just f itself, and the first order derivative is f'.

Theorem 13.4. *(Taylor's theorem) Suppose that $f : [a, b] \to \mathbb{R}$ is continuous and n times continuously differentiable on (a, b), and that $f^{(1)}$, ..., $f^{(n-1)}$ can be continuously extended to $[a, b]$. Let $\alpha, \beta \in [a, b]$ be distinct and define*

$$P(x) = \sum_{k=0}^{n-1} \frac{f^{(k)}(\alpha)}{k!}(x - \alpha)^k$$

on $\mathbb{R}$. Then there exists a point γ between α and β with

$$f(\beta) = P(\beta) + \frac{f^{(n)}(\gamma)}{n!}(\beta - \alpha)^n.$$

Proof. Let $M = \frac{f(\beta)-P(\beta)}{(\beta-\alpha)^n}$, so that $f(\beta) = P(\beta) + M(\beta-\alpha)^n$. We need to find a point γ between α and β with $M = \frac{f^{(n)}(\gamma)}{n!}$.

Define $g(x) = f(x) - P(x) - M(x-\alpha)^n$ for $t \in [a,b]$. P is a polynomial of degree at most $n-1$ and its nth derivative is zero, whereas the nth derivative of $(x-\alpha)^n$ is the constant $n!$. So taking the nth derivative of g yields $g^{(n)}(x) = f^{(n)}(x) - n!M$. Thus it will suffice to find a point γ between α and β with $g^{(n)}(\gamma) = 0$.

By inspection we have $g(\alpha) = g(\beta) = 0$, so by the mean value theorem (Corollary 13.3) there exists a point γ_1 strictly between α and β with $g'(\gamma_1) = 0$. If $n > 1$ then, by inspection, $g'(\alpha) = 0$, so applying the mean value theorem again yields a point γ_2 between α and y_1 with $g''(\gamma_2) = 0$. Continuing in this way, we eventually reach a point γ_n between α and γ_{n-1}, and hence also between α and β, with $g^{(n)}(\gamma_n) = 0$, as desired. □

Now let us move to higher dimensions. We will use the Euclidean norm on $\mathbb{R}^n$. Suppose we are given an open class $A \subseteq \mathbb{R}^n$ and a function $f : A \to \mathbb{R}^m$. Then f is *differentiable at* $\vec{x} \in A$ if there is a linear transformation $Df|_{\vec{x}} : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\lim_{\vec{h}\to 0} \frac{\left\| f(\vec{x}+\vec{h}) - f(\vec{x}) - Df|_{\vec{x}}(\vec{h}) \right\|}{\|\vec{h}\|} = 0.$$

$Df|_{\vec{x}}$ is the *derivative of* f *at* $\vec{x}$. As in the one-variable case, if f is continuous then it suffices to check this only for those $\vec{h}$ all of whose coordinates are rational. If f is differentiable at all $\vec{x} \in A$ then we just say it is *differentiable*.

This second definition of differentiability is equivalent to the earlier one in the $n = m = 1$ case, but $f'(x)$ and $Df|_x$ are are different things — one is a real number, the other is the linear transformation that multiplies by that number. However, we can be a little sloppy and identify $Df|_{\vec{x}}$, when it exists, with the corresponding $m \times n$ matrix (relative to the standard basis), or in the $n = 1$ case, with an element of $\mathbb{R}^m$. In the latter case we also write $f'(\vec{x}) \in \mathbb{R}^m$ for this element.

If f is a function from an open subclass of $\mathbb{R}^n$ to $\mathbb{R}$ and $\vec{a}$ belongs to its domain, then for each $1 \le j \le n$ we can form a function of one variable $\tilde{f}$ by letting $\tilde{f}(t)$ be the value $f(a_1, \ldots, a_n)$ with the jth entry, a_j, replaced by t. The derivative of $\tilde{f}$ at $t = a_j$, if it exists, is called a *partial derivative* of f and denoted $\frac{\partial f}{\partial x_j}(\vec{a})$. For functions from $\mathbb{R}^n$ to $\mathbb{R}^m$, we can write $f = (f_1, \ldots, f_m)$ and consider the partial derivatives $\frac{\partial f_i}{\partial x_j}$ for various i and j.

Proposition 13.5. *Let $f : A \to \mathbb{R}^m$ be a function from an open class $A \subseteq \mathbb{R}^n$ into $\mathbb{R}^m$. If each of the partial derivatives of f exists and is continuous on A, then f is differentiable on A and the matrix of $Df|_{\vec{x}}$ has $\frac{\partial f_i}{\partial x_j}$ in the (i,j) entry.*

Proof. Suppose the partial derivatives all exist and are continuous, and fix $\vec{x} \in A$. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be the linear transformation whose matrix has $\frac{\partial f_i}{\partial x_j}$ in the (i,j) entry. We will show that for every $\epsilon > 0$ there exists $\delta > 0$ such that every $\vec{y} \in A$ with $\|\vec{x} - \vec{y}\| < \delta$ satisfies

$$\|f(\vec{y}) - f(\vec{x}) - T(\vec{y} - \vec{x})\| < \epsilon \|\vec{y} - \vec{x}\|.$$

We can treat each component of f separately and thereby reduce to the case where $m = 1$.

Now $f(\vec{y}) - f(\vec{x})$ can be written as a telescoping sum

$$\begin{aligned} f(y_1, \ldots, y_n) - f(x_1, y_2, \ldots, y_n) + f(x_1, y_2, \ldots, y_n) - f(x_1, x_2, y_3, \ldots, y_n) \\ + \cdots + f(x_1, \ldots, x_{n-1}, y_n) - f(x_1, \ldots, x_n), \end{aligned}$$

and we can apply the mean value theorem (Corollary 13.3) to rewrite this sum as

$$\begin{aligned} \left(\frac{\partial f}{\partial x_1}(u_1, y_2, \ldots, y_n)\right)(y_1 - x_1) + \left(\frac{\partial f}{\partial x_2}(x_1, u_2, y_3, \ldots, y_n)\right)(y_2 - x_2) \\ + \cdots + \left(\frac{\partial f}{\partial x_n}(x_1, \ldots, x_{n-1}, u_n)\right)(y_n - x_n) \end{aligned}$$

for some u_1 between x_1 and y_1, u_2 between x_2 and y_2, etc. Comparing this expression for $f(\vec{y}) - f(\vec{x})$ to $T(\vec{y} - \vec{x})$, which is

$$\left(\frac{\partial f}{\partial x_1}(x_1, \ldots, x_n)\right)(y_1 - x_1) + \cdots + \left(\frac{\partial f}{\partial x_n}(x_1, \ldots, x_n)\right)(y_n - x_n),$$

yields that $f(\vec{y}) - f(\vec{x}) - T(\vec{y} - \vec{x})$ equals

$$\begin{aligned} \Big(\frac{\partial f}{\partial x_1}(u_1, y_2, \ldots, y_n) - \frac{\partial f}{\partial x_1}(x_1, \ldots, x_n)\Big)(y_1 - x_1) \\ + \cdots + \Big(\frac{\partial f}{\partial x_n}(x_1, \ldots, x_{n-1}, u_n) - \frac{\partial f}{\partial x_n}(x_1, \ldots, x_n)\Big)(y_n - x_n). \end{aligned}$$

Since the partial derivatives are continuous and u_i lies between x_i and y_i, we can find $\delta > 0$ such that each of the n differences is at most $\frac{\epsilon}{n}$ in absolute value. Together with the inequality $|y_i - x_i| \leq \|\vec{y} - \vec{x}\|$, this yields the desired estimate on $\|f(\vec{y}) - f(\vec{x}) - T(\vec{y} - \vec{x})\|$. □

Thus, if each of the partial derivatives of f exists and is continuous on A, then $Df|_{\vec{x}}$ exists on A and its matrix entries vary continuously as a function of $\vec{x}$. We say that f is *continuously differentiable*.

The mean value theorem (Corollary 13.3) becomes an inequality for vector-valued functions. For $\vec{x}, \vec{y} \in \mathbb{R}^n$, define the *inner product* $\langle \vec{x}, \vec{y} \rangle$ to be $x_1y_1 + \cdots + x_ny_n$. Also, in the following result, regard $f'(c) = Df|_c$ as a vector in $\mathbb{R}^m$.

Corollary 13.6. *(Mean value inequality) Let $f : [a, b] \to \mathbb{R}^m$ be continuous and suppose f is differentiable on (a, b). Then there exists $c \in (a, b)$ with $\|f(b) - f(a)\| \leq \|f'(c)\| \cdot (b - a)$.*

Proof. Consider the function $g(t) = \langle f(b) - f(a), f(t) \rangle$. It is continuous on $[a, b]$ and differentiable on (a, b) with derivative $g'(t) = \langle f(b) - f(a), f'(t) \rangle$. So Corollary 13.3 says that there exists $c \in (a, b)$ such that $g(b) - g(a) = g'(c)(b - a)$, and thus

$$\begin{aligned} \|f(b) - f(a)\|^2 &= |g(b) - g(a)| \\ &= |g'(c)|(b - a) \\ &= |\langle f(b) - f(a), f'(c) \rangle| \, (b - a) \\ &\leq \|f(b) - f(a)\| \|f'(c)\| (b - a) \end{aligned}$$

(using the Cauchy-Schwarz inequality, Lemma 10.3). Thus $\|f(b) - f(a)\| \leq \|f'(c)\| \cdot (b - a)$. □

Proposition 13.7. *(Multivariable chain rule) Let $A \subseteq \mathbb{R}^n$ and $B \subseteq \mathbb{R}^m$ be open, let $f : A \to \mathbb{R}^m$ and $g : B \to \mathbb{R}^p$ be differentiable, and suppose $f(A) \subseteq B$. Then $g \circ f : A \to \mathbb{R}^p$ is differentiable and we have $D(g \circ f)|_{\vec{x}} = Dg|_{f(\vec{x})} \cdot Df|_{\vec{x}}$ for all $\vec{x} \in A$.*

Proof. Fix $\vec{x} \in A$. We must show that

$$\lim_{\vec{u} \to \vec{x}} \frac{\left\| (g \circ f)(\vec{u}) - (g \circ f)(\vec{x}) - Dg|_{f(\vec{x})} \cdot Df|_{\vec{x}}(\vec{u} - \vec{x}) \right\|}{\|\vec{u} - \vec{x}\|} = 0.$$

Let $\epsilon > 0$ and write the numerator as the sum of

$$g(f(\vec{u})) - g(f(\vec{x})) - Dg|_{f(\vec{x})}(f(\vec{u}) - f(\vec{x}))$$

and

$$Dg|_{f(\vec{x})}(f(\vec{u}) - f(\vec{x}) - Df|_{\vec{x}}(\vec{u} - \vec{x})).$$

We will show that for $\vec{u}$ sufficiently close to $\vec{x}$, both parts have norm at most $\epsilon \cdot \|\vec{u} - \vec{x}\|$. For the first part, since f is differentiable, there exist

$\delta_1, M_1 > 0$ such that $\|\vec{u} - \vec{x}\| < \delta_1$ implies $\|f(\vec{u}) - f(\vec{x})\| \leq M_1\|\vec{u} - \vec{x}\|$. Since g is differentiable, there exists $\delta_2 > 0$ such that $\|f(\vec{u}) - f(\vec{x})\| < M_1\delta_2$ implies

$$\left\|g(f(\vec{u})) - g(f(\vec{x})) - Dg|_{f(\vec{x})}(f(\vec{u}) - f(\vec{x}))\right\| < \frac{\epsilon}{M_1}\|f(\vec{u}) - f(\vec{x})\|.$$

Combining these yields that $\|\vec{u} - \vec{x}\| < \min(\delta_1, \delta_2)$ implies $\left\|g(f(\vec{u})) - g(f(\vec{x})) - Dg|_{f(\vec{x})}(f(\vec{u}) - f(\vec{x}))\right\| \leq \epsilon \cdot \|\vec{u} - \vec{x}\|$.

For the second term, since $Dg|_{f(\vec{x})}$ is linear, it is M_2-Lipschitz for some $M_2 > 0$, and differentiability of f yields that there exists $\delta_3 > 0$ such that $\|\vec{u} - \vec{x}\| < \delta_3$ implies

$$\|f(\vec{u}) - f(\vec{x}) - Df|_{\vec{x}}(\vec{u} - \vec{x})\| < \frac{\epsilon}{M_2}\|\vec{u} - \vec{x}\|.$$

Combining these two facts yields that $\|\vec{u} - \vec{x}\| < \delta_3$ implies

$$\left\|Dg|_{f(\vec{x})}(f(\vec{u}) - f(\vec{x}) - Df|_{\vec{x}}(\vec{u} - \vec{x}))\right\| \leq \epsilon \cdot \|\vec{u} - \vec{x}\|.$$

So we have the desired estimates on both parts when $\|\vec{u} - \vec{x}\| < \min(\delta_1, \delta_2, \delta_3)$. □

Lemma 13.8. *Let g be a c-Lipschitz function from $B_r(0) \subset \mathbb{R}^n$ to $\mathbb{R}^n$ with $g(0) = 0$, for some $c < 1$. Then $f(\vec{x}) = \vec{x} + g(\vec{x})$ is injective on $B_r(0)$ and satisfies $B_{(1-c)r}(0) \subseteq f(B_r(0)) \subseteq B_{(1+c)r}(0)$.*

Proof. First, f is injective because $f(\vec{x}_1) = f(\vec{x}_2)$ implies $\vec{x}_1 - \vec{x}_2 = g(\vec{x}_2) - g(\vec{x}_1)$, which contradicts the assumption that g is c-Lipschitz unless $\vec{x}_1 = \vec{x}_2$. We have $f(B_r(0)) \subseteq B_{(1+c)r}(0)$ because $\|g(\vec{x})\| = \|g(\vec{x}) - g(0)\| \leq c\|\vec{x}\|$ and so $\|f(\vec{x})\| = \|\vec{x} + g(\vec{x})\| \leq (1 + c)\|\vec{x}\|$ for any $\vec{x} \in B_r(0)$.

To prove $B_{(1-c)r}(0) \subseteq f(B_r(0))$, let $r' < r$ and let $\vec{y} \in \overline{B}_{(1-c)r'}(0)$; we must find $\vec{x} \in \overline{B}_{r'}(0)$ such that $\vec{x} + g(\vec{x}) = f(\vec{x}) = \vec{y}$, i.e., $\vec{x} = \vec{y} - g(\vec{x})$. For all $\vec{x} \in \overline{B}_{r'}(0)$ define $F(\vec{x}) = \vec{y} - g(\vec{x})$. Then $\|F(\vec{x})\| \leq (1-c)r' + c\|\vec{x}\| \leq r'$, so F maps $\overline{B}_{r'}(0)$ into itself. Also, F is c-Lipschitz because $F(\vec{x}_1) - F(\vec{x}_2) = g(\vec{x}_2) - g(\vec{x}_1)$. So according to the Banach fixed-point theorem (Theorem 10.7), F has a fixed point, i.e., there exists $\vec{x} \in B_{r'}(0)$ with $\vec{x} = \vec{y} - g(\vec{x})$, as desired. □

For later reference, let us note that the lemma is still true if we replace balls by open cubes, i.e., replace every B_r with $(-\frac{r}{2}, \frac{r}{2})^n$. Another way to say this is that we work with open balls for the l^∞ metric mentioned in Section 10.1 rather than the Euclidean metric. The proof is literally identical if we simply reinterpret $\|\vec{a}\|$ to be $\|\vec{a}\|_\infty = \max_{1 \leq i \leq n}|a_i|$. Taking

closures, we get $C_{(1-c)r} \subseteq f(C_r) \subseteq C_{(1+c)r}$ where $C_r = \left[-\frac{r}{2}, \frac{r}{2}\right]^n$ (provided g is defined on C_r).

In the next proof we use the elementary fact that $\|Tv\| \leq \sqrt{n}\|T\|_\infty \cdot \|v\|$, for any linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$, where $\|T\|_\infty = \max|t_{ij}|$ is the maximum of the absolute values of the entries of the matrix $[T]$.

Theorem 13.9. *(Inverse function theorem) Let $A \subseteq \mathbb{R}^n$ be open, let $\vec{a} \in A$, and let $f : A \to \mathbb{R}^n$ be continuously differentiable. Suppose $Df|_{\vec{a}}$ is invertible. Then there exist open classes U and V with $\vec{a} \in U \subseteq A$ and $\vec{b} = f(\vec{a}) \in V \subseteq \mathbb{R}^n$ such that $f|_U$ is a bijection between U and V whose inverse $g : V \to U$ is also continuously differentiable and satisfies $Dg|_{\vec{b}} = Df|_{\vec{a}}^{-1}$.*

Proof. Normalize so that $\vec{a} = \vec{b} = 0$ (by shifting) and $Df|_0 = I$ (by composing f with $Df|_0^{-1}$). Let $\tilde{f} = f - I$. Then for any distinct $\vec{x}$ and $\vec{y}$ in any ball about 0 contained in A, the function $\tilde{f}(\vec{x} + t(\vec{y} - \vec{x}))$, as a function of $t \in [0, 1]$, has derivative

$$\|\vec{y} - \vec{x}\| \cdot D\tilde{f}|_{\vec{x}+t(\vec{y}-\vec{x})}(\vec{z})$$

where $\vec{z} = \frac{\vec{y}-\vec{x}}{\|\vec{y}-\vec{x}\|}$ and so

$$\|\tilde{f}(\vec{y}) - \tilde{f}(\vec{x})\| \leq \|\vec{y} - \vec{x}\| \cdot \sup_{0<t<1} \left\| D\tilde{f}|_{\vec{x}+t(\vec{y}-\vec{x})}\vec{z} \right\|$$

by the mean value inequality (Corollary 13.6). Using the remark made just before the statement of the theorem, plus continuous differentiability and the fact that $D\tilde{f}|_0$ is the zero matrix, we see that there exists $r > 0$ such that

$$\|\tilde{f}(\vec{y}) - \tilde{f}(\vec{x})\| \leq \frac{1}{2}\|\vec{y} - \vec{x}\|$$

for all $\vec{x}, \vec{y} \in B_r(0)$. By the lemma, $f = \tilde{f} + I$ is injective on $B_r(0)$ and $B_{r/2}(0) \subseteq f(B_r(0))$, so f takes $U = B_r(0) \cap f^{-1}(B_{r/2}(0))$ bijectively onto $V = B_{r/2}(0)$. Possibly making r smaller, we can ensure that $Df|_{\vec{x}}$ is invertible throughout U. (It is invertible if and only if its determinant is nonzero, and $\det([T])$ is a continuous function of the entries of $[T]$.)

We must also show that the inverse function $g : V \to U$ is continuously differentiable. Fix $\vec{x} \in U$ and let $\vec{y} = f(\vec{x})$. For $\vec{h}$ small enough that $\vec{x} + \vec{h} \in U$, define $\vec{k} = f(\vec{x} + \vec{h}) - f(\vec{x})$, so that $f(\vec{x} + \vec{h}) = \vec{y} + \vec{k}$. By the choice of r we have

$$\|\vec{h} - \vec{k}\| = \|f(\vec{x} + \vec{h}) - f(\vec{x}) - \vec{h}\| = \|\tilde{f}(\vec{x} + \vec{h}) - \tilde{f}(\vec{x})\| \leq \frac{\|\vec{h}\|}{2},$$

and so $\frac{\|\vec{h}\|}{2} \leq \|\vec{k}\|$. Letting $T = Df|_{\vec{x}}$, we then have

$$\begin{aligned}\|g(\vec{y}+\vec{k}) - g(\vec{y}) - T^{-1}\vec{k}\| &= \|\vec{h} - T^{-1}(f(\vec{x}+\vec{h}) - f(\vec{x}))\| \\ &\leq L \cdot \|T\vec{h} - f(\vec{x}+\vec{h}) + f(\vec{x})\|\end{aligned}$$

where L satisfies $\|T^{-1}\vec{z}\| \leq L\|\vec{z}\|$ for all $\vec{z}$.

We have $\|\vec{h}\| \leq 2L \cdot \|\vec{k}\|$ for sufficiently small $\vec{h}$, $\vec{k}$. Since

$$\lim_{\vec{h}\to 0} \frac{\|f(\vec{x}+\vec{h}) - f(\vec{x}) - T\vec{h}\|}{\|\vec{h}\|} = 0,$$

the above yields

$$\lim_{\vec{k}\to 0} \frac{\|g(\vec{y}+\vec{k}) - g(\vec{y}) - T^{-1}\vec{k}\|}{\|\vec{k}\|} = 0.$$

Thus g is differentiable at $\vec{y}$ with derivative $T^{-1} = Df|_{\vec{x}}^{-1}$. One can explicitly describe the inverse of an invertible matrix in terms of determinants of its minors, in a way that shows the inverse mapping is continuous on the class of invertible matrices, so g is continuously differentiable. □

Corollary 13.10. *(Implicit function theorem) Suppose $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ is continuously differentiable in an open ball about $(\vec{a}, \vec{b}) \in \mathbb{R}^n \times \mathbb{R}^m$, and the derivative of the map $\vec{y} \mapsto f(\vec{a}, \vec{y})$ at $\vec{y} = \vec{b}$ is invertible. Then there exist open classes U and V about $\vec{a}$ and $\vec{b}$ and a differentiable function $g : U \to V$ such that $f(\vec{x}, g(\vec{x})) = 0$ on U.*

Proof. Find open balls B_1 and B_2 about $\vec{a}$ and $\vec{b}$ such that f is continuously differentiable on $B_1 \times B_2$. Then define $F(\vec{x}, \vec{y}) = (\vec{x}, f(\vec{x}, \vec{y})) \in \mathbb{R}^n \times \mathbb{R}^m$ for $\vec{x} \in B_1$ and $\vec{y}$ in B_2. Then $DF|_{(\vec{a},\vec{b})}$ is invertible since its determinant is nonzero, and therefore by the inverse function theorem (Theorem 13.9) there are open classes U, V, and W about $\vec{a}$, $\vec{b}$, and $F(\vec{a}, \vec{b})$ such that F takes $U \times V$ bijectively onto W. Letting $G : W \to U \times V$ be the inverse and writing $G(\vec{x}) = (G_1(\vec{x}), G_2(\vec{x}))$, we have

$$(\vec{x}, \vec{y}) = F(G_1(\vec{x}, \vec{y}), G_2(\vec{x}, \vec{y})) = (G_1(\vec{x}, \vec{y}), f(G_1(\vec{x}, \vec{y}), G_2(\vec{x}, \vec{y}))).$$

So $\vec{y} = f(\vec{x}, G_2(\vec{x}, \vec{y}))$, for all $\vec{x} \in U$ and $\vec{y} \in V$, and the desired function is $g(\vec{x}) = G_2(\vec{x}, 0)$. □

13.2 Measure theory

Definition 13.11. Let X be a class. Subclasses R_i ($1 \leq i \leq n$ or $i \in \mathbb{N}$) constitute a *measurable basis* if $\bigcup R_i = X$, $R_i \neq R_j$ for $i \neq j$, and for all

i and j there exist k and l such that $R_k = R_i \cup R_j$ and $R_l = R_i \setminus R_j$. A *measure* on X is given by a choice of measurable basis together with a sequence $\mu = (a_i)$ of elements of $[0, \infty)$, over the same range of indices, such that, writing $\mu[R_i] = a_i$, we have (1) $\mu[\emptyset] = 0$ and (2) if the classes R_{i_n} ($1 \leq n \leq N$ or $n \in \mathbb{N}$) are disjoint and $R_i = \bigcup R_{i_n}$ for some i, then $\mu[R_i] = \sum \mu[R_{i_n}]$. A *measure space* is a class equipped with a measure.

We can call the R_i *basic measurable classes*. The entire theory is more or less trivial when there are only finitely many basic measurable classes, so I will assume throughout that the index i ranges over $\mathbb{N}$. The expression "$\mu[R_i]$" is just notation: it is not the value of a function on R_i.

Note that $R_i \cap R_j = R_i \setminus (R_i \setminus R_j)$, so any finite intersection of basic measurable classes is a basic measurable class. Note also that if $R_i \subseteq R_j$ then $R_j = (R_j \setminus R_i) \cup R_i$ is a disjoint union, and therefore $\mu[R_j] = \mu[R_j \setminus R_i] + \mu[R_i]$; in particular, $\mu[R_i] \leq \mu[R_j]$.

If A and B are subclasses of the same class X, we write $A \triangle B$ for the *symmetric difference* $(A \setminus B) \cup (B \setminus A)$.

Definition 13.12. Let μ be a measure on a class X. A subclass $A \subseteq X$ is *ϵ-small* ($\epsilon > 0$) if there exist indices i_n ($1 \leq n \leq N$ or $n \in \mathbb{N}$) such that $A \subseteq \bigcup R_{i_n}$ and $\sum \mu[R_{i_n}] < \epsilon$. The class A is *null* if it is ϵ-small for every $\epsilon > 0$ and *finitely measurable* if for every $\epsilon > 0$ there exists an index i such that $A \triangle R_i$ is ϵ-small. It is *measurable* if its intersection with every finitely measurable class is finitely measurable.

In the next several results we fix a measure μ on a class X.

Lemma 13.13. *Suppose $R_i \subseteq \bigcup R_{i_n}$, for some i and i_n; then $\mu[R_i] \leq \sum \mu[R_{i_n}]$. We have $\mu[R_i] < \epsilon$ if and only if R_i is ϵ-small ($\epsilon > 0$). If $R_i \triangle R_j$ is ϵ-small then $|\mu[R_i] - \mu[R_j]| < \epsilon$.*

Proof. Let $A_1 = R_i \cap R_{i_1}$, $A_2 = (R_i \cap R_{i_2}) \setminus A_1$, $A_3 = (R_i \cap R_{i_3}) \setminus (A_1 \cup A_2)$, etc. Then the A_n are disjoint and their union equals R_i, so $\mu[R_i] = \sum \mu[A_n] \leq \sum \mu[R_{i_n}]$ (since $A_n \subseteq R_{i_n}$ for all n).

For the second part, if R_i is ϵ-small then the R_{i_n} can be chosen so that $\sum \mu[R_{i_n}] < \epsilon$. Then the first part implies that $\mu[R_i] < \epsilon$. The converse implication is trivial because one can just take $R_{i_1} = R_i$ and $R_{i_2} = R_{i_3} = \cdots = \emptyset$.

Finally, if $R_i \triangle R_j$ is ϵ-small then

$$|\mu[R_i] - \mu[R_j]| = |\mu[(R_i \setminus R_j) \cup (R_i \cap R_j)] - \mu[(R_j \setminus R_i) \cup (R_i \cap R_j)]|$$

$$\begin{aligned}&= |\mu[R_i \setminus R_j] - \mu[R_j \setminus R_i]| \\ &\leq \mu[R_i \setminus R_j] + \mu[R_j \setminus R_i] \\ &= \mu[R_i \triangle R_j] < \epsilon,\end{aligned}$$

using the second part of the lemma in the final inequality. □

If $A \triangle R_i$ is ϵ_1-small and $A \triangle R_j$ is ϵ_2-small, then $R_i \triangle R_j \subseteq (A \triangle R_i) \cup (A \triangle R_j)$ is $(\epsilon_1 + \epsilon_2)$-small and therefore $|\mu[R_i] - \mu[R_j]| < \epsilon_1 + \epsilon_2$ by Lemma 13.13. Thus if $A \triangle R_{i_n}$ is $\frac{1}{n}$-small for all n, then the values $\mu[R_{i_n}]$ converge and the limit is independent of the choice of R_{i_n}'s. For any finitely measurable class A we define the measure of A to be this limit and denote it $\mu[A]$. This is clearly consistent with the previous notation $\mu[R_i]$ because we can just take $R_{i_n} = R_i$ for all n. In order to define the measure of any measurable class, we need the following proposition.

Proposition 13.14. *Suppose $A \subseteq X$ is finitely measurable. Then $A \cap R_i$ is finitely measurable for any basic measurable class R_i, and $\mu[A] = \sup \mu[A \cap R_i]$.*

Proof. $A \cap R_i$ is finitely measurable because if $A \triangle R_j$ is ϵ-small then $(A \cap R_i) \triangle (R_j \cap R_i) = (A \triangle R_j) \cap R_i$ is also ϵ-small. This also shows that we can approximate $A \cap R_i$ with basic measurable classes that are contained in basic measurable classes approximating A, so $\mu[A \cap R_i] \leq \mu[A]$, for all i. Conversely, for any $\epsilon > 0$ we can find i such that $A \triangle R_i$ is ϵ-small, and this implies $|\mu[A] - \mu[R_i]| < \epsilon$ directly from the definition of $\mu[A]$. But $(A \cap R_i) \triangle R_i \subseteq A \triangle R_i$ is also ϵ-small, so that we also get $|\mu[A \cap R_i] - \mu[R_i]| < \epsilon$ and therefore $|\mu[A] - \mu[A \cap R_i]| \leq 2\epsilon$. Taking $\epsilon \to 0$, this shows that $\mu[A] = \sup \mu[A \cap R_i]$. □

Therefore we can define the measure of any measurable class A to be $\mu[A] = \sup \mu[A \cap R_i] \in [0, \infty]$. Proposition 13.14 guarantees that this is consistent with the original definition of the measure of A when A is finitely measurable. We will see below (Lemma 13.16) that the terminology "finitely measurable" is well-chosen — if A is measurable but not finitely measurable, then its measure is infinite.

We can generalize the second and third parts of Lemma 13.13 to all finitely measurable classes.

Lemma 13.15. *Suppose $A, B \subseteq X$ are finitely measurable. Then $\mu[A] < \epsilon$ if and only if A is ϵ-small ($\epsilon > 0$). If $A \triangle B$ is ϵ-small then $|\mu[A] - \mu[B]| < \epsilon$.*

Proof. Suppose $\mu[A] < \epsilon$. Let $\delta = \frac{\epsilon - \mu[A]}{2}$ and find R_i such that $A \triangle R_i$ is δ-small. Then $A \subseteq (A \triangle R_i) \cup R_i$ is $\delta + (\mu[A] + \delta) = \epsilon$-small.

Conversely, suppose A is ϵ-small; it will suffice to show that $\mu[A] \leq \epsilon$. First find $\epsilon' > 0$ such that A is $(\epsilon - \epsilon')$-small and R_i such that $A \triangle R_i$ is ϵ'-small; then $R_i \subseteq A \cup (A \triangle R_i)$ is ϵ-small and therefore $\mu[R_i] < \epsilon$ by Lemma 13.13. Taking ϵ' to zero yields $\mu[A] \leq \epsilon$.

For the other statement, suppose $A \triangle B$ is ϵ-small. Then it is ϵ_0-small for some $\epsilon_0 < \epsilon$. Find R_i and R_j with $A \triangle R_i$ and $B \triangle R_j$ both $\frac{\epsilon - \epsilon_0}{4}$-small. Then $R_i \triangle R_j \subseteq (R_i \triangle A) \cup (A \triangle B) \cup (B \triangle R_j)$, which implies that $R_i \triangle R_j$ is $\frac{\epsilon + \epsilon_0}{2}$-small and therefore

$$\begin{aligned} |\mu[A] - \mu[B]| &\leq |\mu[A] - \mu[R_i]| + |\mu[R_i] - \mu[R_j]| + |\mu[R_j] - \mu[B]| \\ &< \frac{\epsilon - \epsilon_0}{4} + \frac{\epsilon + \epsilon_0}{2} + \frac{\epsilon - \epsilon_0}{4} = \epsilon, \end{aligned}$$

as desired. □

Lemma 13.16. *Let A and B be disjoint finitely measurable classes. Then $A \cup B$ is finitely measurable and $\mu[A \cup B] = \mu[A] + \mu[B]$. If C is measurable and $\mu[C] < \infty$ then C is finitely measurable.*

Proof. Let $\epsilon > 0$ and find i and j such that $A \triangle R_i$ and $B \triangle R_j$ are both ϵ-small. Then $(A \cup B) \triangle (R_i \cup R_j) \subset (A \triangle R_i) \cup (B \triangle R_j)$ is 2ϵ-small, which shows that $A \cup B$ is finitely measurable and also yields $|\mu[A] - \mu[R_i]| < \epsilon$, $|\mu[B] - \mu[R_j]| < \epsilon$, and $|\mu[A \cup B] - \mu[R_i \cup R_j]| < 2\epsilon$.

Moreover, $R_i \cap R_j \subseteq (A \triangle R_i) \cup (B \triangle R_j)$ is 2ϵ-small, so that $\mu[R_i \cap R_j] < 2\epsilon$. As we also have $\mu[R_i \cup R_j] = \mu[R_i] + \mu[R_j \setminus R_i] = \mu[R_i] + \mu[R_j] - \mu[R_i \cap R_j]$, this shows that $|\mu[R_i \cup R_j] - (\mu[R_i] + \mu[R_j])| < 2\epsilon$. Combining this with the result of the last paragraph and letting ϵ decrease to 0 yields $\mu[A \cup B] = \mu[A] + \mu[B]$.

Now suppose $\mu[C] < \infty$. Let $R'_1 = R_1$ and $R'_n = R_n \setminus (R_1 \cup \cdots \cup R_{n-1})$ for $n \geq 2$. By finite additivity, we have

$$\sum_{n=1}^{N} \mu[C \cap R'_n] = \mu \left[C \cap \bigcup_{n=1}^{N} R_n \right] \leq \mu[C] < \infty$$

for all N, and therefore the infinite sum $\sum \mu[C \cap R'_n]$ is also finite. So given $\epsilon > 0$, we can find N such that $\sum_{n=N+1}^{\infty} \mu[C \cap R'_n] < \epsilon$, and by Lemma 13.15 this implies that $\bigcup_{n=N+1}^{\infty} (C \cap R'_n) = C \cap \bigcup_{n=N+1}^{\infty} R'_n$ is ϵ-small. But since C is measurable, $C \cap \bigcup_{n=1}^{N} R'_n$ is finitely measurable, so there exists j such that $\left(C \cap \bigcup_{n=1}^{N} R'_n \right) \triangle R_j$ is also ϵ-small. These two facts yield that $C \triangle R_j$ is 2ϵ-small. We have shown that C is finitely measurable. □

In part (e) of the next proposition, we write $\lim a_i = \infty$ if for every $N \in \mathbb{N}$ we have $a_i > N$ for all but finitely many i.

Proposition 13.17. *Let X be a measure space with measure μ and let A, B, A_1, A_2, ... be measurable classes. Then*

(a) $A \setminus B$ is measurable
(b) if $A \subseteq B$ then $\mu[A] \leq \mu[B]$
(c) (countable subadditivity) if $\sum \mu[A_n]$ converges then $\bigcup A_n$ is finitely measurable with $\mu\left[\bigcup A_n\right] \leq \sum \mu[A_n]$
(d) (countable additivity) if $\sum \mu[A_n]$ converges and the A_n are disjoint then $\mu\left[\bigcup A_n\right] = \sum \mu[A_n]$
(e) if $A_1 \subseteq A_2 \subseteq \cdots$ then $\bigcup A_n$ is measurable and $\mu\left[\bigcup A_n\right] = \lim \mu[A_n]$
(f) if $A_1 \supseteq A_2 \supseteq \cdots$ and $\mu[A_1] < \infty$ then $\bigcap A_n$ is finitely measurable and $\mu\left[\bigcap A_n\right] = \lim \mu[A_n]$.

Proof. (a) Let C be finitely measurable and write $A' = A \cap C$ and $B' = B \cap C$. We need $(A \setminus B) \cap C = A' \setminus B'$ to be finitely measurable. But if $A' \triangle R_i$ and $B' \triangle R_j$ are both ϵ-small, then $(A' \setminus B') \triangle (R_i \setminus R_j) \subseteq (A' \triangle R_i) \cup (B' \triangle R_j)$ is 2ϵ-small, so $A' \setminus B'$ is indeed finitely measurable.

(b) We can assume that $\mu[B] < \infty$. Let $A \subseteq B$ and suppose first that $\mu[A] < \infty$. Then B is ϵ-small for any $\epsilon > \mu[B]$ by Lemma 13.15, so A is also ϵ-small for any such ϵ, and again by Lemma 13.15 this implies that $\mu[A] \leq \mu[B]$. This also shows that $\mu[A] = \infty$ is impossible, because measurability of A now implies that $\mu[A \cap C] \leq \mu[B]$ for any finitely measurable C, in particular for any R_i.

(c) Suppose $\sum \mu[A_n]$ converges, let $\epsilon > 0$, and find $N \in \mathbb{N}$ such that $\sum_{n=N+1}^{\infty} \mu[A_n] < \epsilon$. By Lemma 13.15, this means that $\bigcup_{n=N+1}^{\infty} A_n$ is ϵ-small. Then for $n = 1, \ldots, N$ find indices i_n such that $A_n \triangle R_{i_n}$ is $\frac{\epsilon}{N}$-small. Say $R_i = \bigcup_{n=1}^{N} R_{i_n}$; then

$$\left(\bigcup_{n=1}^{\infty} A_n\right) \triangle R_i \subseteq \bigcup_{n=1}^{N} (A_n \triangle R_{i_n}) \cup \bigcup_{n=N+1}^{\infty} A_n,$$

showing that $(\bigcup A_n) \triangle R_i$ is 2ϵ-small. As $\mu[R_i] \leq \sum_{n=1}^{N} \mu[A_n] + \epsilon$, and ϵ was arbitrary, this shows that $\bigcup A_n$ is finitely measurable with $\mu\left[\bigcup A_n\right] \leq \sum \mu[A_n]$.

(d) Suppose $\sum \mu[A_n] < \infty$ and the A_n's are disjoint. Now for any N Lemma 13.16 yields $\mu\left[\bigcup_{n=1}^{N} A_n\right] = \sum_{n=1}^{N} \mu[A_n]$. Together with part (b), this shows that $\sum_{n=1}^{N} \mu[A_n] \leq \mu\left[\bigcup_{n=1}^{\infty} A_n\right]$, and since N was arbitrary, simply that $\sum \mu[A_n] \leq \mu\left[\bigcup A_n\right]$. For the reverse inequality, let $\epsilon > 0$ and find N such that $\sum_{n=N+1}^{\infty} \mu[A_n] < \epsilon$. Then $\bigcup_{n=N+1}^{\infty} A_n$ is ϵ-small by

Lemma 13.15. Find R_i such that $\left(\bigcup_{n=1}^N A_n\right) \triangle R_i$ is ϵ-small; then

$$\left(\bigcup A_n\right) \triangle R_i \subseteq \left(\left(\bigcup_{n=1}^N A_n\right) \triangle R_i\right) \cup \bigcup_{n=N+1}^\infty A_n$$

is 2ϵ-small, so that (using finite additivity)

$$\mu\left[\bigcup A_n\right] \leq \mu[R_i] + 2\epsilon \leq \mu\left[\bigcup_{n=1}^N A_n\right] + 3\epsilon = \sum_{n=1}^N \mu[A_n] + 3\epsilon.$$

Since ϵ was arbitrary, we get $\mu\left[\bigcup A_n\right] \leq \sum \mu[A_n]$.

(e) Suppose $A_1 \subseteq A_2 \subseteq \cdots$. Assume first that $\sup \mu[A_n] < \infty$. Let $B_1 = A_1$, $B_2 = A_2 \setminus A_1$, $B_3 = A_3 \setminus A_2$, ... Then the B_n are disjoint and for each n we have $\mu[A_n] = \mu[B_1] + \cdots + \mu[B_n]$ by finite additivity. Thus the Nth partial sum of $\sum \mu[B_n]$ equals $\mu[A_N]$, and so the assumption that $\sup \mu[A_n] < \infty$ means that $\sum \mu[B_n]$ converges to $\sup \mu[A_n]$. And by parts (c) and (d), we have that $\bigcup B_n = \bigcup A_n$ is measurable and $\sum \mu[B_n] = \mu\left[\bigcup B_n\right] = \mu\left[\bigcup A_n\right]$.

If $\sup \mu[A_n] = \infty$, we can apply the preceding to the sequence $A_1 \cap C \subseteq A_2 \cap C \subseteq \cdots$ for an arbitrary finitely measurable class C. This yields that $(\bigcup A_n) \cap C = \bigcup (A_n \cap C)$ is measurable, and as C was arbitrary that means $\bigcup A_n$ is measurable. Part (b) then shows that $\mu\left[\bigcup A_n\right] = \infty$.

(f) Suppose $A_1 \supseteq A_2 \supseteq \cdots$ and $\mu[A_1] < \infty$. For each n let $A'_n = A_1 \setminus A_n$; then $A'_1 \subseteq A'_2 \subseteq \cdots$, so part (e) says that $\mu[A'_n] \to \mu\left[\bigcup A'_n\right] = \mu\left[A_1 \setminus \bigcap A_n\right]$. By finite additivity $\mu[A'_n] = \mu[A_1] - \mu[A_n]$ and $\mu\left[A_1 \setminus \bigcap A_n\right] = \mu[A_1] - \mu\left[\bigcap A_n\right]$; we conclude that $\mu[A_n] \to \mu\left[\bigcap A_n\right]$. □

Corollary 13.18. *Complements, countable unions, and countable intersections of measurable classes are measurable.*

Proof. Complements are measurable by Proposition 13.17 (a). If A and B are finitely measurable then $A \setminus B$ is also finitely measurable by Proposition 13.17 (a) and (b) plus the last part of Lemma 13.16; thus the first part of Lemma 13.16 tells us that $A \cup B$ is finitely measurable. From this it easily follows that any finite union of measurable classes is measurable, and then from Proposition 13.17 (e) that countable unions of measurable classes are measurable. The same is true of countable intersections by taking complements. □

Although I write the measure of A as "$\mu[A]$", it is not the value of a function μ on a measurable class A. There is no class, nor even a pseudoclass, of all measurable classes. However, for many purposes it suffices to

work with measurable classes of the following type, which can be indexed as a pseudoclass.

Definition 13.19. A subclass of a measure space is an M_δ *class* if it can be expressed as a countable intersection of countable unions of basic measurable classes.

Proposition 13.20. *Let A be a measurable subclass of a measure space. Then there is an M_δ class B and a null class N such that $A = B \setminus N$.*

Proof. By Lemma 13.15, for each i and n we can find a countable union B^i_n of basic measurable classes such that $A \cap R_i \subseteq B^i_n$ and $\mu[B^i_n] < \mu[A \cap R_i] + \frac{1}{n \cdot 2^i}$. Then for each n the class $\bigcup_i B^i_n$ is a countable union of basic measurable classes that contains A and whose difference with A has measure at most $\frac{1}{n}$. So $\bigcap_n \bigcup_i B^i_n$ is an M_δ class that contains A and whose difference with A is null. □

We can index the M_δ classes with the class $\mathcal{P}(\mathbb{N}\times\mathbb{N})$. The set $Z \subseteq \mathbb{N}\times\mathbb{N}$ can index the M_δ class which is the intersection over $m \in \mathbb{N}$ of the union of the basic measurable classes R_n such that $(m, n) \in Z$. Thus we can effectively quantify over the M_δ classes in any measure space, and we can interpret μ as a legitimate function from the pseudoclass of all M_δ classes to $[0, \infty]$.

The single most important example of a measure is *Lebesgue measure* on $\mathbb{R}$. Here we take the basic measurable classes to be the empty class plus all finite disjoint unions of the form $R_i = [a_1, b_1) \cup \cdots \cup [a_n, b_n)$ with $a_1, \ldots, a_n, b_1, \ldots, b_n \in \mathbb{Q}$ and $a_1 < b_1 < \cdots < a_n < b_n$. We define Lebesgue measure by $m[R_i] = (b_1 - a_1) + \cdots + (b_n - a_n)$. It is easy to see that this is finitely additive. To verify countable additivity, suppose some R_i is partitioned into R_{i_1}, R_{i_2}, ... First use finite additivity to reduce to the case where $R_i = [a, b)$ and $R_{i_n} = [a_n, b_n)$ for all n. Finite additivity then yields the inequality $\sum m[R_{i_n}] \leq m[R_i]$ in the usual way. For the reverse inequality, let $\epsilon > 0$ and for each n let $R'_{i_n} = (a_n - \frac{\epsilon}{2^n}, b_n)$. Then $\bigcup R'_{i_n}$ contains $R_i = [a, b)$, so by compactness finitely many of the R'_{i_n}'s cover $[a, b - \epsilon]$. Finite additivity now yields

$$m[R_i] - \epsilon = (b - a) - \epsilon \leq \sum m[R'_{i_n}] = \sum m[R_{i_n}] + \epsilon.$$

So $m[R_i] \leq \sum m[R_{i_n}] + 2\epsilon$, and taking ϵ to zero yields the desired inequality. Thus Lebesgue measure is a measure.

Proposition 13.21. *Every open subclass of $\mathbb{R}$ is Lebesgue measurable.*

This just follows from Corollary 13.18, as every open class is a countable union of open intervals and every open interval is a countable union of half-open intervals with rational endpoints.

A subclass A of a topological space is said to be "G_δ" if it can be expressed as an intersection of countably many open classes.

Theorem 13.22. *(Regularity of Lebesgue measure on $\mathbb{R}$) Let $A \subseteq \mathbb{R}$ be Lebesgue measurable with $m[A] < \infty$. Then*

(a) for any $\epsilon > 0$ there exists an open class U with $A \subseteq U$ and $m[U] < m[A] + \epsilon$
(b) for any $\epsilon > 0$ there exists a compact class K with $K \subseteq A$ and $m[K] > m[A] - \epsilon$
(c) we can write $A = A' \setminus N$ where A' is G_δ and N is null.

Proof. (a) Let $\epsilon > 0$ and find a finite disjoint union of half-open intervals R such that $A \triangle R$ is ϵ-small. Then $A \setminus R$ is contained in a countable union of half-open intervals whose total measure is less than ϵ, and by enlarging each of these half-open intervals to a slightly larger open interval, we get that $A \setminus R$ is contained in an open class U_1 whose total measure is less than ϵ. We can also find an open class U_2 which contains R such that $m[U_2] < m[R] + \epsilon$. Then A is contained in $U = U_1 \cup U_2$ and

$$m[U] \leq m[U_1] + m[U_2] < m[R] + 2\epsilon < m[A] + 3\epsilon,$$

which is enough.

(b) Since $A = \bigcup_n (A \cap [-n, n])$, by Proposition 13.17 (e) we can reduce to the case where $A \subseteq [-n, n]$ for some n. Then by part (a) we can find an open class U containing $[-n, n] \setminus A$ with $m[U] < m\left[[-n, n] \setminus A\right] + \epsilon$; letting $K = [-n, n] \setminus U$, we then have $K \subseteq A$ and

$$m[K] \geq 2n - m[U] > 2n - (2n - m[A] + \epsilon) = m[A] - \epsilon.$$

(c) By part (a), for each n we an find an open class U_n with $A \subseteq U_n$ and $m[U_n] < m[A] + \frac{1}{n}$. Then $A' = \bigcap U_n$ is G_δ, contains A, and satisfies $m[A'] \leq \inf m[U_n] \leq m[A]$. So $m[A'] = m[A]$ and $A' \setminus A$ is null. □

Now let us consider product measures. Let X and Y be measure spaces, with measures μ and ν defined on basic measurable classes R_i and S_j. We wish to define a measure $\mu \times \nu$ on $X \times Y$. The natural choice for basic measurable classes T_k in $X \times Y$ is finite unions of "rectangles" of the form $R_i \times S_j$. Since the T_k are required to be distinct, we must check that any two such finite unions definitely are or are not equal, in order that duplicates

may be discarded. This is an exercise in Boolean algebra: given two finite unions, all the R_i's that appear generate a finite Boolean algebra under unions and complements, and similarly for the S_j's. Each of the two finite unions can then be expressed as a finite union of rectangles $R_i \times S_j$ with R_i and S_j atoms in their respective Boolean algebras. Once this is done, the two unions are equal if and only if they are composed of the same rectangles. The *product measure* $\mu \times \nu$ is given by $(\mu \times \nu)[R_i \times S_j] = \mu[R_i]\nu[S_j]$, extending to all the T_k's by finite additivity. One must check that this extension is consistent, i.e., if T_k is partitioned into rectangles in more than one way, they yield the same value for $(\mu \times \nu)[T_k]$. This is also an easy exercise (hint: common refinement).

The harder part is showing that $\mu \times \nu$ is countably additive on the T_k's in the sense of Definition 13.11. I will just sketch the proof. First, reduce to the case of a single rectangle $R_i \times S_j$ that is partitioned into countably many smaller rectangles $R_{i_n} \times S_{j_n}$. Finite additivity gives the inequality

$$\sum_{n=1}^{N} (\mu \times \nu)[R_{i_n} \times S_{j_n}] \leq (\mu \times \nu)[R_i \times S_j],$$

and taking N to infinity yields the same inequality with a sum from 1 to ∞ on the left side. For the reverse inequality, fix $\epsilon > 0$ and for each $N \in \mathbb{N}$ define $A_N \subseteq X$ by letting $x \in A_N$ if the sum of $\nu[S_{j_n}]$, taken over all $n \leq N$ such that $x \in R_{i_n}$, is at least $\nu[S_j] - \epsilon$. Then each A_N will be a basic measurable class and $\bigcup_N A_N = R_i$. So by Proposition 13.17 (e) we get $\sup \mu[A_N] = \mu[R_i]$; in particular, we must have $\mu[A_N] \geq \mu[R_i] - \epsilon$ for some value of N, and then finite additivity yields

$$\sum_{n=1}^{N} (\mu \times \nu)[R_{i_n} \times S_{j_n}] \geq (\mu[R_i] - \epsilon)(\nu[S_j] - \epsilon).$$

Taking ϵ to zero completes the proof of countable additivity. Thus $\mu \times \nu$ is a legitimate measure on $X \times Y$.

We consider two measures to be eqivalent if they have the same measurable classes and take the same values on them, even if their basic measurable classes differ. The next result shows that the choice of basic measurable classes does not essentially affect the product measure construction.

Proposition 13.23. *Let μ be a measure defined on basic measurable classes R_i. Suppose some μ-measurable classes R'_i also constitute a measurable basis and the values $\mu[R'_i]$ are all finite. Then $\mu'[R'_i] = \mu[R'_i]$ defines a measure μ' on the measurable basis (R'_i). If every R_i is μ'-measurable then*

μ' and μ have the same measurable classes and take the same values on them.

Proof. Proposition 13.17 (d) shows that μ' is a measure. Now suppose every R_i is μ'-measurable. Then for any i and j the class $R_i \cap R'_j$ is finitely μ'-measurable, so for any $\epsilon > 0$ there exists k for which $\mu'[(R_i \cap R'_j) \triangle R'_k] < \epsilon$. But ϵ-small relative to μ' implies ϵ-small relative to μ (Lemma 13.15), so $\mu[(R_i \cap R'_j) \triangle R'_k] < \epsilon$ and hence $\mu[R_i \cap R'_j]$ and $\mu'[R_i \cap R'_j]$ are both within ϵ of $\mu[R'_k] = \mu'[R'_k]$. Taking ϵ to zero yields that $\mu[R_i \cap R'_j] = \mu'[R_i \cap R'_j]$, and since j was arbitrary this shows that $\mu[R_i] = \mu'[R_i]$ for all i. Now, by virtue of Lemma 13.15, μ and μ' have the same measurable classes and take the same values on them. □

The point about product measures is that if μ' and ν' are equivalent to μ and ν, then the basic measurable rectangles for $\mu \times \nu$ will be measurable relative to $\mu' \times \nu'$, and vice versa, with the same measures. So Proposition 13.23 implies that $\mu' \times \nu'$ is equivalent to $\mu \times \nu$.

As an application of the product measure construction, we can construct Lebesgue measure in $\mathbb{R}^n$, still denoted m, as a product of copies of the one-dimensional Lebesgue measure constructed earlier. The next result is a straightforward generalization of Theorem 13.22.

Theorem 13.24. *(Regularity of Lebesgue measure on $\mathbb{R}^n$) Let $A \subseteq \mathbb{R}^n$ be Lebesgue measurable with $m[A] < \infty$. Then*

(a) for any $\epsilon > 0$ there exists an open class U with $A \subseteq U$ and $m[U] < m[A] + \epsilon$
(b) for any $\epsilon > 0$ there exists a compact class K with $K \subseteq A$ and $m[K] > m[A] - \epsilon$
(c) we can write $A = A' \setminus N$ where A' is G_δ and N is null.

We conclude this section with a discussion of signed and complex measures. First, we need the following basic notions.

Definition 13.25. Let μ and ν be measures defined on the same basic measurable classes. Then μ and ν are *mutually singular*, written $\mu \perp \nu$, if there is a M_δ class A such that $\mu[A^c] = \nu[A] = 0$. The measure ν is *absolutely continuous* with respect to μ, written $\nu \ll \mu$, if every μ-null class is also ν-null.

Again, quantification over all M_δ classes is bounded. There is no loss in defining mutual singularity in terms of M_δ classes, because if A is any class

which is both μ- and ν-measurable and satisfies $\mu[A^c] = \nu[A] = 0$, then we can find an M_δ class B with

$$(\mu + \nu)[A \triangle B] = (\mu + \nu)[A^c \triangle B^c] = 0,$$

and then $\mu[B^c] = \mu[A^c] = 0$ and $\nu[B] = \nu[A] = 0$.

Absolute continuity can be diagnosed by only looking at the behavior of μ and ν on the basic measurable classes.

Proposition 13.26. *Let μ and ν be finite measures defined on the same measurable basis. The following are equivalent:*

(a) $\nu \ll \mu$
(b) every μ-measurable class is also ν-measurable and for every $\epsilon > 0$ there exists $\delta > 0$ such that every δ-small class for μ is ϵ-small for ν
(c) for every $\epsilon > 0$ there exists $\delta > 0$ such that $\mu[R_i] < \delta$ implies $\nu[R_i] < \epsilon$, for every basic measurable class R_i.

Proof. Suppose $\nu \ll \mu$. Let us show that for every rational $\epsilon > 0$ there exists a rational $\delta > 0$ such that $\mu[R_i] < \delta$ implies $\nu[R_i] < \epsilon$, for any basic measurable class R_i. This assertion is clearly definite. Suppose it is false; then there exists $\epsilon > 0$ and basic measurable classes R_{i_n} satisfying $\mu[R_{i_n}] < \frac{1}{2^n}$ and $\nu[R_{i_n}] \geq \epsilon$. The class $A = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} R_{i_k}$ is then μ-null but (using the fact that ν is finite) $\nu[A] \geq \epsilon$, a contradiction. This proves that (a) implies (c).

Given (c), it follows from Proposition 13.17 (e) that for every $\epsilon > 0$ there exists $\delta > 0$ such that every δ-small class for μ is ϵ-small for ν. Then every finitely μ-measurable class B is finitely ν-measurable because for any $\epsilon > 0$ we can find a basic measurable class R_i with $\mu[B \triangle R_i] < \delta$, hence $\nu[B \triangle R_i] < \epsilon$. This implies that every μ-measurable class is also ν-measurable. So (c) implies (b).

Finally, (b) trivially implies that every μ-null class is also ν-null. □

Definition 13.27. A *complex measure* on a class X equipped with a measurable basis $R_1, R_2, \ldots$ is a sequence $\mu = (a_i)$ of elements of $\mathbb{C}$ such that, writing $\mu[R_i] = a_i$, we have (1) $\mu[\emptyset] = 0$, (2) $\sup |a_i| < \infty$, and (3) if the classes R_{i_n} are disjoint and $R_i = \bigcup R_{i_n}$ for some i, then $\mu[R_i] = \sum \mu[R_{i_n}]$ with absolute convergence. If the a_i are all real then μ is a *signed measure*.

For clarity, we can call the measures discussed earlier, which take values in $[0, \infty]$, *positive*.

Theorem 13.28. *(Hahn-Jordan decomposition) Let μ be a signed measure. Then there uniquely exist mutually singular, finite, positive measures μ^+ and μ^- defined on the same measurable basis as μ, such that $\mu[R_i] = \mu^+[R_i] - \mu^-[R_i]$ for all i.*

Proof. For each i define $\mu^+[R_i]$ to be the supremum of $\{\mu[R_j] : R_j \subseteq R_i\}$. Note that R_j definitely is or is not contained in R_i, for all i and j, because there exists a unique k with $R_k = R_j \setminus R_i$, and there exists a unique l with $R_l = \emptyset$, and k definitely does or does not equal l.

To show that μ^+ is a positive measure we must demonstrate countable additivity in the sense that whenever R_i is the disjoint union of R_{i_n}, we have $\mu^+[R_i] = \sum \mu^+[R_{i_n}]$. So suppose R_i is a disjoint union of basic measurable classes R_{i_n}. Let $\epsilon > 0$ and find j such that $R_j \subseteq R_i$ and $\mu[R_j] > \mu^+[R_i] - \epsilon$. Then $\mu[R_j] = \sum \mu[R_{i_n} \cap R_j] \leq \sum \mu^+[R_{i_n}]$ by condition (3) of Definition 13.27, which yields that $\mu^+[R_i] - \epsilon < \sum \mu^+[R_{i_n}]$. Since ϵ was arbitrary we get $\mu^+[R_i] \leq \sum \mu^+[R_{i_n}]$. For the reverse inequality, observe that boundedness (Definition 13.27 (2)) forces $\sum \mu^+[R_{i_n}]$ to be finite. So given $\epsilon > 0$, we can find N such that $\sum_{n=N+1}^{\infty} \mu^+[R_{i_n}] < \epsilon$. Then for each $1 \leq n \leq N$, find an index j_n such that $R_{j_n} \subseteq R_{i_n}$ and $\mu[R_{j_n}] \geq \mu^+[R_{i_n}] - \frac{\epsilon}{N}$. The union $\bigcup_{n=1}^{N} R_{j_n}$ is then a basic measurable class R_j contained in R_i with the property that

$$\mu^+[R_i] \geq \mu[R_j] \geq \sum_{n=1}^{N} \mu^+[R_{i_n}] - \epsilon > \sum_{n=1}^{\infty} \mu^+[R_{i_n}] - 2\epsilon.$$

Since ϵ was arbitrary, we get that $\mu^+[R_i] \geq \sum \mu^+[R_{i_n}]$. This completes the proof that μ^+ is a finite positive measure.

Define $\mu^-[R_i] = \mu^+[R_i] - \mu[R_i]$ for all i. This is obviously positive, and it is countably additive in the desired sense because μ^+ and μ are both countably additive. So μ^- is a finite positive measure too. We must now show that μ^+ and μ^- are mutually singular.

Let $a = \sup a_i = \sup \mu[R_i]$. For each $m \in \mathbb{N}$ find R_{j_m} with $a \geq \mu[R_{j_m}] \geq a - \frac{1}{2^m}$, which implies that $\mu^-[R_{j_m}] = \mu^+[R_{j_m}] - \mu[R_{j_m}] \leq a - (a - \frac{\epsilon}{2^m}) = \frac{1}{2^m}$. Then let $A_m = \bigcup_{k \geq m+1} R_{j_k}$ for all m and $A = \bigcap A_m$. For each m we have $\mu^+[A_m] \geq \sup_{k \geq m+1} \mu^+[R_{j_k}] \geq a$, and as $\mu^+[A]$ cannot be strictly greater than a, this means that $\mu^+[A] = a$. Also, $\mu^-[A_m] \leq \sum_{k=m+1}^{\infty} \frac{1}{2^k} = \frac{1}{2^m}$, so $\mu^-[A] = 0$. And finally, $a \geq \mu^+[A \cup A^c] = a + \mu^+[A^c]$, showing that $\mu^+[A^c] = 0$. So μ^+ and μ^- are mutually singular.

Finally, for the uniqueness statement, suppose $\tilde{\mu}^+$ and $\tilde{\mu}^-$ are two other mutually singular, finite, positive measures with $\mu[R_i] = \tilde{\mu}^+[R_i] - \tilde{\mu}^-[R_i]$

for all i. Then, with A as above, $\nu^+[R_i] = \tilde{\mu}^+[R_i \cap A]$ and $\nu^-[R_i] = \tilde{\mu}^-[R_i \cap A]$ define mutually singular positive measures whose difference satisfies $\nu^+[R_i] - \nu^-[R_i] = \mu^+[R_i] \geq 0$ for any basic measurable class R_i, and hence $\nu^+[B] \geq \nu^-[B]$ for any class B which is both ν^+- and ν^--measurable, which is only possible if ν^- is constantly zero. Thus $\tilde{\mu}^-[A] = 0$, and $\tilde{\mu}^+[A^c] = 0$ similarly, so we conclude that $\tilde{\mu}^+ = \mu^+$ and $\tilde{\mu}^- = \mu^-$. □

Definition 13.29. Let μ be a signed measure on a class X and let μ^+ and μ^- be as in Theorem 13.28. We say that $A \subseteq X$ is *measurable* with respect to μ if it is measurable with respect to both μ^+ and μ^-; if so, we define $\mu[A] = \mu^+[A] - \mu^-[A]$. Likewise, if μ is a complex measure then its real and imaginary parts are signed measures μ^r and μ^i, and we define $\mu[A] = \mu^r[A] + i\mu^i[A]$ for all A which are μ-measurable, meaning both μ^r- and μ^i-measurable.

Given a complex measure μ, for each i define $|\mu|[R_i]$ to be the supremum of $|\mu[R_{i_1}]| + \cdots + |\mu[R_{i_m}]|$, taken over all $m \in \mathbb{N}$ and all partitions of R_i into basic measurable classes $R_{i_1}, \ldots, R_{i_m}$.

Proposition 13.30. *Let μ be a complex measure on a class X. Then $|\mu|$ is a finite positive measure, and a subclass of X is $|\mu|$-measurable if and only if it is μ-measurable. For any such class A the supremum $\sup\{|\mu[A_1]| + \cdots + |\mu[A_m]|\}$, taken over all partitions of A into μ-measurable classes $A_1, \ldots, A_m$, exists and equals $|\mu|[A]$. If μ is a signed measure and μ^+ and μ^- are as in the Hahn-Jordan decomposition (Theorem 13.28) then $|\mu| = \mu^+ + \mu^-$.*

Proof. To prove that $|\mu|$ is a positive measure, we must verify countable additivity when R_i is the disjoint union of R_{i_n}. This is done using an argument similar to the first part of the proof of Theorem 13.28.

The positive and negative parts of the real and imaginary parts of μ are all dominated by $|\mu|$, so any $|\mu|$-measurable class is also measurable for their sum, and hence for μ. Conversely, $|\mu|$ is dominated by this sum, so any μ-measurable class is also $|\mu|$-measurable.

Since $|\mu[R_i]| \leq |\mu|[R_i]$ for all i, it follows that $|\mu[A]| \leq |\mu|[A]$ for all μ-measurable A. Thus if $A_1, \ldots, A_m$ are μ-measurable classes that partition a μ-measurable class A, then

$$|\mu[A_1]| + \cdots + |\mu[A_m]| \leq |\mu|[A_1] + \cdots + |\mu|[A_m] = |\mu|[A].$$

At the same time, given any μ-measurable subclass A, for any $\epsilon > 0$ we can find a basic measurable class R_i with $|\mu|[A \triangle R_i] < \epsilon$ and a measurable

partition $R_{i_1}, \ldots, R_{i_n}$ of R_i with $|\mu[R_{i_1}]| + \cdots + |\mu[R_{i_n}]| \geq |\mu|[R_i] - \epsilon$, and then $B_1 = A \cap R_{i_1}, \ldots, B_n = A \cap R_{i_n}$, $B_{n+1} = A \setminus R_i$ will be a measurable partition of A. Setting $B'_1 = R_{i_1} \setminus A, \ldots, B'_n = R_{i_n} \setminus A$, we have (as above)

$$|\mu[B'_1]| + \cdots + |\mu[B'_n]| \leq |\mu|[R_i \setminus A] < \epsilon,$$

so

$$\begin{aligned} |\mu[B_1]| + \cdots + |\mu[B_{n+1}]| &= |\mu[R_{i_1}] - \mu[R_{i_1} \setminus A]| + \cdots + |\mu[R_{i_n}] - \mu[R_{i_n} \setminus A]| \\ &\geq |\mu[R_{i_1}]| + \cdots + |\mu[R_{i_n}]| - \epsilon \\ &\geq |\mu|[R_i] - 2\epsilon \geq |\mu|[A] - 3\epsilon. \end{aligned}$$

This shows that $|\mu|[A]$ can be approximated arbitrarily well from below by measurable partitions of A. We conclude that $|\mu|[A] = \sup\{|\mu[A_1]| + \cdots + |\mu[A_m]|\}$, taken over all measurable partitions of A.

Now suppose μ is signed and let R_i be a basic measurable class. If $\mu[R_i] < \mu^+[R_i]$, then for any $0 < \epsilon < \mu^+[R_i] - \mu[R_i]$ we can find $R_j \subseteq R_i$ with $\mu[R_j] \geq \mu^+[R_i] - \epsilon$; and then $\mu[R_i \setminus R_j] \leq -\mu^-[R_i] + \epsilon < 0$ and

$$|\mu|[R_i] \geq |\mu[R_j]| + |\mu[R_i \setminus R_j]| \geq \mu^+[R_i] + \mu^-[R_i] - 2\epsilon.$$

Since ϵ was arbitrary, we get $|\mu|[R_i] \geq \mu^+[R_i] + \mu^-[R_i]$ for all i. The same conclusion is trivial if $\mu[R_i] = \mu^+[R_i]$. Conversely, for any partition of R_i into $R_{i_1}, \ldots, R_{i_n}$ we have

$$\sum |\mu[R_{i_k}]| \leq \sum \mu^{\pm}[R_{i_k}] \leq \mu^+[R_i] + \mu^-[R_i]$$

where the sign of $\mu^{\pm}$ is determined depending on whether $\mu[R_{i_k}]$ is positive or negative. Taking the supremum over all such partitions completes the proof that $|\mu|[R_i] = \mu^+[R_i] + \mu^-[R_i]$.

Finiteness of $|\mu|$ in the signed case follows immediately, and in the complex case by taking real and imaginary parts. □

The measure $|\mu|$ is the *absolute value* of μ. Proposition 13.30, together with Proposition 13.23, shows that this construction is essentially independent of the choice of basic measurable classes. Note that, writing $\mu = \mu_1 - \mu_2 + i\mu_3 - i\mu_4$, we have

$$\mu_1, \mu_2, \mu_3, \mu_4 \leq |\mu| \leq \mu_1 + \mu_2 + \mu_3 + \mu_4;$$

since a class is μ-measurable if and only if it is μ_i-measurable for $i = 1, 2, 3, 4$, this implies that μ-measurability is equivalent to $|\mu|$-measurability.

If μ is a positive measure and ν is a complex measure, we say that ν is absolutely continuous with respect to μ and write $\nu \ll \mu$ if $|\nu|$ is absolutely

continuous with respect to μ, and we say that μ and ν are mutually singular and write $\mu \perp \nu$ if μ and $|\nu|$ are mutually singular.

Theorem 13.31. *(Lebesgue decomposition) Let μ be a positive measure and let ν be a complex measure defined on the same basic measurable classes. Then there are complex measures ν^1 and ν^2 defined on the same basic measurable classes with $\nu = \nu^1 + \nu^2$ that satisfy $\nu^1 \ll \mu$ and $\nu^2 \perp \mu$.*

Proof. By separating ν into real and imaginary parts, then separating each of those into positive and negative parts, we can reduce to the case where ν is a finite positive measure.

For each $n \in \mathbb{N}$ define $b_n = \sup\left\{\nu[R_i] : \mu[R_i] \leq \frac{1}{2^n}\right\}$. Then let $b = \lim b_n$ and, for each n, find R_{i_n} such that $\mu[R_{i_n}] \leq \frac{1}{2^n}$ and $\nu[R_{i_n}] \geq b_n - \frac{1}{2^n}$. Then $A = \bigcap_n A_n$ with $A_n = \bigcup_{m \geq n+1} R_{i_m}$ satisfies $\mu[A] = 0$ and $\nu[A] = b$ (since $\mu[A_n] \leq \frac{1}{2^n}$ and therefore $b_n - \frac{1}{2^{n+1}} \leq \nu(A_n) \leq b_n$). Let ν^2 be the restriction of ν to A, i.e., the measure satisfying $\nu^2[R_i] = \nu[R_i \cap A]$ for all i, and let $\nu^1 = \nu - \nu^2$.

It is clear that ν^2 and μ are mutually singular. We must show that $\nu^1 \ll \mu$. To do this, let B be a μ-null class which is disjoint from A. Given $n \in \mathbb{N}$, find basic measurable classes R_{j_k} such that $A \cup B \subseteq \bigcup R_{j_k}$ and $\sum \mu[R_{j_k}] < \frac{1}{2^n}$. Let $B_n = \bigcup R_{j_k}$. Then $\nu\left[\bigcup_{k=1}^m R_{j_k}\right] \leq b_n$ for all $m \in \mathbb{N}$, and therefore $\nu[B_n] \leq b_n$. Since $\nu[A] = b$, that means that $\nu[B_n \setminus A] \leq b_n - b$. Thus for any n the class B is contained in a ν-measurable class $B_n \setminus A$ with $\nu[B_n \setminus A] < b_n - b$. As $b_n - b \to 0$, this shows that B is ν-null, and hence $A \cup B$ is ν_1-null. Since every μ-null class is contained in $A \cup B$ for some such B, this shows that ν_1 is absolutely continuous with respect to μ. □

13.3 Integration

Definition 13.32. A function $f : X \to Y$ from a measure space to a topological space is *measurable* if $f^{-1}(U)$ is a measurable subclass of X for every open subclass U of Y.

One easily checks that when $Y = \mathbb{R}$ it suffices for $f^{-1}((a, \infty))$ to be measurable for all $a \in \mathbb{R}$, or even all $a \in \mathbb{Q}$. Also, trivially, if Z is another topological space and $g : Y \to Z$ is continuous then measurability of f implies measurability of $g \circ f$.

A function into $\mathbb{C}^2$ is measurable if and only if each component is. The forward direction of this statement follows from the previous observation,

since the coordinate projections are continuous. The reverse direction follows from the fact that every open subclass of $\mathbb{C}^2$ is a countable union of classes of the form $U \times V$ with $U, V \subseteq \mathbb{C}$ open.

Proposition 13.33. *Let X be a measure space.*

(a) If $f, g : X \to \mathbb{C}$ are measurable then so are the pointwise sum and product $f + g$ and fg.
(b) The pointwise complex conjugate of a measurable function from X into $\mathbb{C}$ is measurable. A function $f : X \to \mathbb{C}$ is measurable if and only if its real and imaginary parts are measurable.
(c) Let f_1, f_2, ... be measurable functions from X into $\mathbb{R}$ and suppose $\sup_n f_n(x)$ and $\inf_n f_n(x)$ are finite for every $x \in X$. Then the pointwise supremum and infimum $\sup f_n$ and $\inf f_n$ are measurable.
(d) Let f_1, f_2, ... be measurable functions from X into $\mathbb{C}$ and suppose $f_n(x) \to f(x)$ for all $x \in X$ (pointwise convergence). Then f is measurable.

Proof. (a) Addition and multiplication are continuous functions from $\mathbb{C}^2$ to $\mathbb{C}$, and, as we noted just above, if f and g are measurable functions into $\mathbb{C}$ then $x \mapsto (f(x), g(x))$ is a measurable function into $\mathbb{C}^2$. So $f + g$ and fg are both compositions of a measurable function with a continuous function and hence are measurable.

(b) The first part holds because the map $z \mapsto \overline{z}$ is continuous from $\mathbb{C}$ to itself (and again using the fact that the composition of a continuous function with a measurable function is measurable). The second part now follows from (a) because $\mathrm{Re}(f) = \frac{1}{2}(f + \overline{f})$ and $\mathrm{Im}(f) = \frac{1}{2i}(f - \overline{f})$.

(c) Let $g = \sup f_n$ be the pointwise supremum. Then for each $a \in \mathbb{R}$

$$g^{-1}((a, \infty)) = \bigcup f_n^{-1}((a, \infty))$$

is a countable union of measurable classes and is therefore measurable. Measurability of the infimum can be deduced from this or proven in a similar way.

(d) If f is real-valued, we have $f = \limsup f_n$, so this follows by a double application of part (c). For complex-valued f the result is inferred by considering the real and imaginary parts of f separately. □

Let X be a measure space with (positive) measure μ. A *simple function* on X is a function of the form $f = \sum_{i=1}^{n} a_i \cdot 1_{A_i}$ where the a_i's are non-negative real numbers, the A_i's are measurable subclasses of X, and 1_{A_i} is the indicator function which takes the value 1 on A_i and 0 elsewhere. If

the A_i's partition X and the a_i's are distinct then f is *in standard form*. If f is in standard form, we define $\int f\,d\mu$ to be $\sum_{i=1}^{n} a_i \cdot \mu[A_i]$, using the convention $0 \cdot \infty = 0$.

Now let $f : X \to [0,\infty]$ be a measurable function. For any n-tuple $\omega = (r_1, \ldots, r_n)$ of rational numbers satisfying $0 = r_1 < \cdots < r_n$, let $A_i = f^{-1}([r_i, r_{i+1}))$ $(1 \le i < n)$ and define f_ω to be the function $\sum_{i=1}^{n} r_i \cdot 1_{A_i}$. If ω' is a refinement of ω, i.e., a tuple that includes all the r_i's, then it is easy to see that $\int f_{\omega'}\,d\mu \ge \int f_\omega\,d\mu$. We define $\int f\,d\mu$ (or $\int f(x)\,d\mu(x)$, or $\int_X f\,d\mu$, or $\int f$, if more or less detail is needed — or sometimes $\int f(x)\,dx$ when f is a function of x and μ is understood) to be $\sup \int f_\omega\,d\mu$, taking the supremum over all ω. This supremum could be infinite. If it is finite then f is *integrable*. One easily checks that if f is simple then this definition of $\int f\,d\mu$ agrees with the earlier one.

If $f \le g$ then $f_\omega \le g_\omega$ for all ω, and therefore $\int f \le \int g$.

Theorem 13.34. *(Monotone convergence theorem) Let X be a measure space and let f_1, f_2, ... be measurable functions into $[0,\infty]$ satisfying $f_1 \le f_2 \le \cdots$. Then $\int \lim f_n = \lim \int f_n$.*

Proof. Let $f = \lim f_n$ be the pointwise limit. Then $f_n \le f$ for all n and therefore $\int f_n \le \int f$ for all n, which entails that $\lim \int f_n \le \int f$. For the reverse inequality, let $0 < \epsilon < 1$ and fix ω; it will suffice to show that $\lim \int f_n \ge (1-\epsilon) \int f_\omega$.

For each $n \in \mathbb{N}$ define $A_n = \{x : f_n(x) \ge (1-\epsilon) f_\omega(x)\}$. Then $A_1 \subseteq A_2 \subseteq \cdots$ and $\bigcup A_n = X$. We have

$$\int f_n \ge \int_{A_n} f_n \ge \int_{A_n} (1-\epsilon) f_\omega = (1-\epsilon) \int_{A_n} f_\omega,$$

where the last equality is trivial since f_ω is simple.

For any measurable class $A \subseteq X$, we have $\int_{A_n} 1_A = \mu[A_n \cap A] \to \mu[A] = \int 1_A$, and taking linear combinations yields $\int_{A_n} f_\omega \to \int f_\omega$. We conclude that $\lim \int f_n \ge (1-\epsilon) \int f_\omega$, as desired. □

Corollary 13.35. *Let X be a measure space with measure μ and let $f, g, f_n : X \to [0,\infty]$ be measurable functions. Then*

(a) if $f \le g$ then $\int f \le \int g$
(b) $\int af = a \int f$ for all $a \ge 0$
(c) $\int (f+g) = \int f + \int g$
(d) $\sum \int f_n = \int \sum f_n$.

Proof. Part (a) was already noted above. For part (b), suppose $a \geq 0$ is rational and observe that for any $\omega = (a_1, \ldots, a_n)$, if we set $a\omega = (a \cdot a_1, \ldots, a \cdot a_n)$ then $a \int f_\omega = \int (af)_{a\omega}$. Taking limits yields $a \int f = \int af$. For arbitrary $a > 0$, find a sequence of positive rationals which increase to a and apply Theorem 13.34 to get $a \int f = \int af$.

For part (c), let $h_1 \leq h_2 \leq \cdots$ and $k_1 \leq k_2 \leq \cdots$ be simple functions which converge pointwise to f and g, respectively. Then $h_n + k_n$ converges pointwise to $f+g$, and hence $\int (h_n+k_n) \to \int (f+g)$ by Theorem 13.34. Since h_n and k_n are simple, one can show by hand that $\int (h_n+k_n) = \int h_n + \int k_n$, and Theorem 13.34 again yields $\lim \int h_n \to \int f$ and $\lim \int k_n \to \int g$. That suffices to show that $\int (f+g) = \int f + \int g$.

Part (d) follows from Theorem 13.34 by the calculation

$$\sum_{n=1}^{\infty} \int f_n = \lim_{n\to\infty} \sum_{k=1}^{n} \int f_k = \lim_{n\to\infty} \int \sum_{k=1}^{n} f_k = \int \sum_{n=1}^{\infty} f_n.$$

□

Corollary 13.36. *(Fatou's lemma) Let X be a measure space and let f_1, f_2, ... be measurable functions into $[0, \infty]$. Then $\int (\liminf f_n) \leq \liminf \int f_n$.*

Proof. For each n let $g_n = \inf_{k \geq n} f_k$. Then $g_1 \leq g_2 \leq \cdots$ and $\liminf f_n = \lim g_n$. Theorem 13.34 then yields

$$\int (\liminf f_n) = \int \lim g_n = \lim \int g_n;$$

but since $\int g_n \leq \int f_k$ for all $k \geq n$, this shows that $\int (\liminf f_n) \leq \liminf \int f_n$. □

Now we can proceed to integrate complex-valued functions.

Definition 13.37. Let X be a measure space. A measurable function $f : X \to \mathbb{C}$ is *integrable* if $\int |f| < \infty$, i.e., $|f|$ is integrable. If f is integrable and real-valued then $f_+ = f \cdot 1|_A$ with $A = \{x : f(x) \geq 0\}$ and $f_- = f_+ - f$ are both integrable and we define $\int f = \int f_+ - \int f_-$. For any integrable f, both $\mathrm{Re}(f)$ and $\mathrm{Im}(f)$ are integrable and we define $\int f = \int \mathrm{Re}(f) + i \int \mathrm{Im}(f)$.

Theorem 13.38. *Let X be a measure space, let $f, g : X \to \mathbb{C}$ be integrable, and let $a \in \mathbb{C}$. Then*

(a) $\int (f+g) = \int f + \int g$
(b) $\int af = a \int f$
(c) $\left| \int f \right| \leq \int |f|$.

Proof. (a) Integrability of $f+g$ follows from the inequality $|f+g| \leq |f|+|g|$. Treating real and imaginary parts separately, we can assume f and g are real. Then letting $h = f+g$ and decomposing $f = f_+ - f_-$, $g = g_+ - g_-$, $h = h_+ - h_-$, we have

$$h_+ - h_- = f + g = f_+ - f_- + g_+ - g_-,$$

so $h_+ + f_- + g_- = h_- + f_+ + g_+$. Corollary 13.35 (c) then yields

$$\int h_+ + \int f_- + \int g_- = \int h_- + \int f_+ + \int g_+,$$

so

$$\int h = \int h_+ - \int h_- = \int f_+ - \int f_- + \int g_+ - \int g_- = \int f + \int g.$$

(b) For $f : X \to [0,\infty]$ and $a \geq 0$ this was Corollary 13.35 (b). For general f and a the result follows by writing $a = a_1 - a_2 + ia_3 - ia_4$ and $f = f_1 - f_2 + if_3 - if_4$ with $a_1, a_2, a_3, a_4 \geq 0$ and $f_1, f_2, f_3, f_4 \geq 0$, and invoking part (a).

(c) Suppose $\int f \neq 0$ and define $a = \frac{|\int f|}{\int f}$. Then $\int af = \left|\int f\right|$ is real and $|a| = 1$, so

$$\left|\int f\right| = \int af = \int \mathrm{Re}(af) \leq \int |f|$$

since $\mathrm{Re}(af)(x) \leq |f(x)|$ for all x. □

Theorem 13.39. *(Dominated convergence theorem) Let X be a measure space, let f_1, f_2, ... be integrable functions on X, and let g be a positive integrable function on X. Suppose $f_n \to f$ pointwise and $|f_n| \leq g$ for all n. Then f is integrable and $\int f_n \to \int f$.*

Proof. By considering real and imaginary parts separately, we can reduce to the real case. Observe first that $|f(x)| = \lim |f_n(x)| \leq g(x)$ for all x, so integrability of g implies integrability of f. Now $g \pm f_n \geq 0$ for all n, so by Fatou's lemma (Corollary 13.36)

$$\int g + \int f \leq \liminf \int (g + f_n) = \int g + \liminf \int f_n$$

and

$$\int g - \int f \leq \liminf \int (g - f_n) = \int g - \limsup \int f_n.$$

Thus $\liminf \int f_n \geq \int f \geq \limsup \int f_n$, which shows that $\int f = \lim \int f_n$. □

For $a, b \in \mathbb{R}$, $a < b$, write $\int_a^b f(t)\,dt$ for the integral of a function f defined on $[a, b]$ with respect to Lebesgue measure. We define differentiability for complex-valued functions on $[a, b]$ exactly the same way we defined it for real-valued functions at the beginning of Section 14.1; equivalently, $f : [a, b] \to \mathbb{C}$ is differentiable if and only if its real and imaginary parts f_r and f_i are, with $f' = f_r' + if_i'$.

Theorem 13.40. *(Fundamental theorem of calculus) Let $a, b \in \mathbb{R}$, $a < b$, and let $f : [a, b] \to \mathbb{C}$ be continuous. Define $F : [a, b] \to \mathbb{C}$ by $F(x) = \int_a^x f(t)\,dt$. Then F is continuous on $[a, b]$ and differentiable on (a, b) with derivative f.*

Proof. If $x_n \to x$ in $[a, b]$ then by Theorem 13.39

$$F(x_n) = \int_a^{x_n} f(t)\,dt = \int 1_{[a,x_n]} \cdot f\,dt \to \int 1_{[a,x]} \cdot f\,dt = \int_a^x f(t)\,dt = F(x),$$

so F is continuous. Now fix $x \in (a, b)$ and for any $\epsilon > 0$ find $k > 0$ such that $a < x - k < x + k < b$ and $|f(y) - f(x)| \le \epsilon$ for all $y \in [x - k, x + k]$. Then for any $h \in (-k, k)$

$$\begin{aligned} f(x) - \epsilon &= \frac{1}{h} \int_x^{x+h} (f(x) - \epsilon)\,dt \\ &\le \frac{1}{h} \int_x^{x+h} f(t)\,dt \\ &\le \frac{1}{h} \int_x^{x+h} (f(x) + \epsilon)\,dt \\ &= f(x) + \epsilon \end{aligned}$$

and thus

$$\frac{F(x+h) - F(x)}{h} = \frac{1}{h} \int_x^{x+h} f(t)\,dt$$

lies between $f(x) + \epsilon$ and $f(x) - \epsilon$. Thus F is differentiable at x with $F'(x) = f(x)$. □

Corollary 13.41. *Let $a, b \in \mathbb{R}$, $a < b$, let $f : [a, b] \to \mathbb{C}$ be continuous, and suppose $F : [a, b] \to \mathbb{C}$ is continuous on $[a, b]$ and differentiable on (a, b) with derivative f. Then $\int_a^b f(x)\,dx = F(b) - F(a)$.*

Proof. Define $G(x) = F(a) + \int_a^x f(t)\,dt$. According to Theorem 13.40, G is continuous on $[a, b]$ and differentiable on (a, b) with derivative f. Thus $G - F$ is continuous on $[a, b]$ and differentiable on (a, b) with derivative 0. It follows from the mean value theorem (Corollary 13.3) that $G - F$ is constant on $[a, b]$. Since $G(a) = F(a)$, we must therefore have $G(x) = F(x)$ for all $x \in [a, b]$; thus $\int_a^b f(t)\,dt = G(b) - F(a) = F(b) - F(a)$. □

If something is true for all x outside of a null class, we say it is true *almost everywhere*.

Theorem 13.42. *(Cavalieri's principle) Let X and Y be measure spaces with measures μ and ν and let C be a measurable subclass of $X \times Y$. Then $C_x = \{y : (x, y) \in C\} \subseteq Y$ and $C^y = \{x : (x, y) \in C\} \subseteq X$ are measurable for almost every x and y, the functions $x \mapsto \nu[C_x]$ and $y \mapsto \mu[C^y]$ are measurable, and*

$$(\mu \times \nu)(C) = \int_X \nu[C_x]\, dx = \int_Y \mu[C^y]\, dy.$$

Proof. Let the basic measurable classes in X and Y be R_1, R_2, ... and S_1, S_2, ..., respectively. First, assume μ and ν are finite. The theorem is trivial when C is a basic measurable rectangle, $C = R_i \times S_j$, and then Proposition 13.33 (a) and (d) (for measurability of the functions $x \mapsto \nu[C_x]$ and $y \mapsto \mu[C^y]$) and Corollary 13.35 (d) (for the equality of integrals) yield the theorem when C is a disjoint union of basic measurable rectangles. In this special case C_x and C^y are measurable for all x and y.

Now suppose $C = \bigcap C_n$ where each C_n is a disjoint union of basic measurable rectangles. Since the intersection of two disjoint unions of basic measurable rectangles is itself a disjoint union of basic measurable rectangles, we can assume that $C_1 \supseteq C_2 \supseteq \cdots$. The theorem then holds for C by Corollary 13.18 (for measurability of C_x and C^y, still true for all x and y), Proposition 13.33 (d) (for measurability of the functions $x \mapsto \nu[C_x]$ and $y \mapsto \mu[C^y]$), and the dominated convergence theorem (Theorem 13.39) (for the equality of integrals).

If C is an arbitrary measurable subclass of $X \times Y$ then we can find a basic measurable class — a finite disjoint union of basic measurable rectangles — whose symmetric difference with C is ϵ-small for any desired ϵ; this shows that C can be covered by a disjoint union of basic measurable rectangles whose total measure is at most $\frac{1}{n}$ more than $(\mu \times \nu)[C]$. Doing this for each n and intersecting the sequence then yields a class C' which contains C and such that $C' \setminus C$ is null, for which the theorem holds. Making the same argument for the complement of C yields another class C'' which is contained in C and such that $C \setminus C''$ is null, for which the theorem holds. Then letting $f(x) = \nu[C'_x]$ and $g(x) = \nu[C''_x]$, we get $g \leq f$, both functions are measurable, and

$$(\mu \times \nu)[C] = \int f\, d\mu = \int g\, d\mu.$$

This implies that $f = g$ almost everywhere, and since $C_x'' \subseteq C_x \subseteq C_x'$ for all x we conclude that C_x is measurable for almost every x and $(\mu \times \nu)[C] = \int \nu[C_x]\, d\mu$. A similar argument leads to the same conclusion for the y sections. This completes the proof in the case where μ and ν are finite.

In general, let $R_1' = R_1$, $R_2' = R_2 \setminus R_1$, $R_3' = R_3 \setminus (R_1 \cup R_2)$, etc., and define S_1', S_2', ... similarly. Then, after $\emptyset$ has been removed anywhere it appears in either list, we are left with partitions of X and Y into basic measurable classes. For each i and j we can then intersect C with $R_i' \times S_j'$, and we know that the theorem holds for each of these intersections, so by summing everything up it holds for the original class C. □

In the next theorem $\int f\, dxdy$ means $\int f\, d(\mu \times \nu)$.

Theorem 13.43. *(Tonelli's theorem) Let X and Y be measure spaces with measures μ and ν and let $f : X \times Y \to [0, \infty]$ be measurable. Then $f_x(y) = f(x, y)$ and $f^y(x) = f(x, y)$ are measurable functions for almost every x and y, the functions $x \mapsto \int f_x\, d\nu$ and $y \mapsto \int f^y\, d\mu$ are measurable, and*

$$\int f\, dxdy = \int \left[\int f(x, y)\, dy \right] dx = \int \left[\int f(x, y)\, dx \right] dy.$$

Proof. The result holds when $f = 1_C$ for some measurable $C \subseteq X \times Y$ by Theorem 13.42, and it follows for simple functions by linearity. For general f, find simple functions $h_1 \leq h_2 \leq \cdots$ from X into $[0, \infty]$ which increase pointwise to f. Then the monotone convergence theorem (Theorem 13.34) implies that $\int (h_n)_x\, dy \to \int f_x\, dy$ and $\int (h_n)^y\, dx \to \int f^y\, dx$ for all x and y off of null classes on which $(h_n)_x$ and $(h_n)^y$ are not measurable for some n. Measurability of the functions $x \mapsto \int (h_n)_x\, dy$ and $y \mapsto \int (h_n)^y\, dx$ then entails measurability of the functions $x \mapsto \int f_x\, dy$ and $y \mapsto \int f^y\, dx$. Monotone convergence implies the equality of the integrals. □

Theorem 13.44. *(Fubini's theorem) Let X and Y be measure spaces with measures μ and ν and let $f : X \times Y \to \mathbb{C}$ be integrable. Then f_x and f^y are integrable for almost every x and y, the functions $x \mapsto \int f_x\, dy$ and $y \mapsto \int f^y\, dx$ are integrable, and*

$$\int f\, dxdy = \int \left[\int f(x, y)\, dy \right] dx = \int \left[\int f(x, y)\, dx \right] dy.$$

Proof. Since f is integrable, Theorem 13.43 implies that $\int |f|_x\, d\nu$ and $\int |f|^y\, d\mu$ are finite almost everywhere. This shows that f_x and f^y are integrable for almost every x and y. The remainder of the theorem follows from

Theorem 13.43 by writing f as a linear combination of positive integrable functions. □

Next I will prove a theorem which describes how integration in $\mathbb{R}^n$ is affected by a change of variables. We need the following basic fact about determinants.

Lemma 13.45. *Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be an invertible linear transformation. Then the image under T of any Lebesgue measurable subclass A of $\mathbb{R}^n$ is Lebesgue measurable, and $\mu[T(A)] = |\det([T])| \cdot \mu[A]$, where μ is Lebesgue measure in $\mathbb{R}^n$.*

Proof. The image $T(A)$ exists as a subclass of $\mathbb{R}^n$ because T is invertible and inverse images of subclasses are subclasses.

We know from Section 12.1 that T can be decomposed as $T = DPV_1 \cdots V_s$ where $[D]$ is a diagonal matrix whose diagonal entries are all nonzero, $[P]$ is a permutation matrix, i.e., P permutes the elements of the standard basis, and each $[V_i]$ has all 1's on the diagonal and a single nonzero off-diagonal entry. V_i is a *skew transformation*; multiplying on the right by $[V_i]$ alters a matrix by adding a nonzero multiple of one column to another. The $[U_i]$'s from Section 12.1 can be written as products of such matrices.

P clearly takes basic measurable classes to basic measurable classes, preserving measure; it follows that P preserves measurability. If all the diagonal entries of $[D]$ are positive then D too takes basic measurable classes to basic measurable classes, and we have $\mu[D(R_j)] = \det([D]) \cdot \mu[R_j]$ for any basic measurable class R_j since D scales the coordinate axes by its diagonal entries, and its determinant is the product of its diagonal entries. This implies that D also preserves measurability. If D has some negative entries it can still be written as $D = D'J$ where D' is diagonal with positive entries down the diagonal and J is diagonal with ± 1 entries down the diagonal; by the symmetry of Lebesgue measure on $\mathbb{R}$ we then get that D preserves measurability regardless of the signs of its diagonal entries.

To complete the proof that $T(A)$ is measurable we must check that each V_i preserves measurability. It is a simple exercise to show that the image under V_i of any class of the form $[a_1, b_1) \times \cdots \times [a_n, b_n)$ is measurable, with the same measure — all that really matters is the $n = 2$ case — and from this it follows that the image under V_i of any basic measurable class is measurable, with the same measure. Thus, if A is finitely measurable then $V_i(A) \triangle V_i(R_j) = V_i(A \triangle R_j)$ can be made ϵ-small by a suitable choice of basic measurable class R_j, and since $V_i(R_j)$ is finitely measurable, $V_i(R_j) \triangle R_k$

can be made ϵ-small by a suitable choice of R_k. This shows that if A is finitely measurable then so is $V_i(A)$, i.e., V_i preserves finite measurability, and this implies that it preserves measurability.

Since the determinant is multiplicative (Theorem 12.21), we now have

$$\mu[T(A)] = \mu[D(A)] = |\det([D])| \cdot \mu[A] = |\det([T])| \cdot \mu[A],$$

as desired. □

The *Jacobian* of a function f from an open subclass of $\mathbb{R}^n$ into $\mathbb{R}^n$, denoted J_f, is the determinant of the matrix of Df.

Theorem 13.46. *(Change of variables) Let $A \subseteq \mathbb{R}^n$ be open and let $f : A \to \mathbb{R}^n$ be continuously differentiable. Suppose $Df|_{\vec{x}}$ is invertible for all $\vec{x} \in A$, i.e., J_f is nowhere zero. If $g : \mathbb{R}^n \to \mathbb{R}$ is a continuous function which is constantly zero outside of a compact class contained in $f(A)$, then*

$$\int_{f(A)} g(\vec{y})\, d\vec{y} = \int_A g(f(\vec{x})) \cdot |J_f(\vec{x})|\, d\vec{x}$$

(integration against Lebesgue measure on $\mathbb{R}^n$).

Proof. We will prove the desired equality for integration over a cube $C = [a_1, a_1 + b] \times \cdots \times [a_n, a_n + b]$. This will suffice because any open class A can be written as a union of countably many cubes with disjoint interiors, and their boundaries have measure zero. (Suppose A is a union of open balls B_i. First find all cubes of side length 1 with vertices in $\mathbb{Z}^n$ that are contained in some B_i. Then find all cubes of side length $\frac{1}{2}$ with vertices in $\frac{1}{2}\mathbb{Z}^n$ that are contained in some B_i and whose interiors are disjoint from all cubes chosen so far. Continue through all powers of 2; every element of every B_i must be contained in a cube that is selected at some point.)

We know from the inverse function theorem (Theorem 13.9) that the matrix entries of $(Df|_{\vec{x}})^{-1}$ are continuous functions of $\vec{x}$, and therefore the function $(\vec{x}, v) \mapsto (Df|_{\vec{x}})^{-1}v$ is continuous on $C \times \mathbb{R}^n$. By compactness, the supremum $\lambda = \sup \left\|(Df|_{\vec{x}})^{-1}v\right\|$, taken over all $\vec{x} \in C$ and all *unit vectors* $v \in \mathbb{R}^n$ with $\|v\| = 1$, exists and is finite. Also let κ be the supremum of $|g|$ on C, and let ρ be the supremum of $|J_f|$ over C.

Fix $\epsilon = \frac{1}{r}$ with $r \in \mathbb{N}$, $r \geq 2$. By Theorem 10.48 there exists $\delta > 0$ such that $\|\vec{x} - \vec{y}\| \leq \delta$ implies that $\|Df|_{\vec{x}}v - Df|_{\vec{y}}v\| \leq \frac{\epsilon}{\lambda}$ and that $\left\|\frac{\partial f_i}{\partial x_j}(\vec{x}) - \frac{\partial f_i}{\partial x_j}(\vec{y})\right\|$, $|J_f(\vec{x}) - J_f(\vec{y})|$, and $|g(f(\vec{x})) - g(f(\vec{y}))|$ are all less than ϵ, for all $\vec{x}, \vec{y} \in C$, all unit vectors $v \in \mathbb{R}^n$, and all i and j. Then we can subdivide C into a union of finitely many little cubes with disjoint interiors,

each of which has side length $l < \frac{1}{\sqrt{n}}\delta$, so the distance between any two points in each little cube is less than δ. I claim that the difference between the two integrals, of $g(\vec{y})$ and $g(f(\vec{x}))\cdot|J_f(\vec{x})|$, relative to any of these little cubes, is at most $M\epsilon\cdot l^n$ for some M which depends only on κ and ρ; since the volume of each little cube is l^n, summing over all the cubes yields that the difference between the two integrals relative to C is at most $M\epsilon\cdot\mu[C]$. Since ϵ could be arbitrarily small, this will establish the result.

Fix one of the little cubes and without loss of generality suppose it is the cube $C_l = \left[-\frac{l}{2},\frac{l}{2}\right]^n$. Normalize f so that it takes $0\in\mathbb{R}^n$ to itself and let $T = Df|_0$. Then let $\tilde{f} = T^{-1}\circ f - I$. For all $\vec{x},\vec{y}\in C_l$, by the mean value inequality (Corollary 13.6) there exists $\vec{z}$ on the line segment joining them such that

$$\begin{aligned}\|\tilde{f}(\vec{y})-\tilde{f}(\vec{x})\| &\le \left\|D\tilde{f}|_{\vec{z}}\frac{\vec{y}-\vec{x}}{\|\vec{y}-\vec{x}\|}\right\|\cdot\|\vec{y}-\vec{x}\|\\ &\le \lambda\cdot\left\|(Df|_{\vec{z}}-Df|_0)\left(\frac{\vec{y}-\vec{x}}{\|\vec{y}-\vec{x}\|}\right)\right\|\cdot\|\vec{y}-\vec{x}\|\\ &\le \epsilon\cdot\|\vec{y}-\vec{x}\|.\end{aligned}$$

Thus $\tilde{f}$ is ϵ-Lipschitz on C_l, so by the comment just after Lemma 13.8 the function $T^{-1}\circ f$ is injective on C_l and satisfies $C_{(1-\epsilon)l}\subseteq(T^{-1}\circ f)(C_l)\subseteq C_{(1+\epsilon)l}$.

We can now estimate the difference between $\int_{f(C_l)} g(\vec{y})\,d\vec{y}$ and $\int_{C_l} g(f(\vec{x}))\cdot|J_f(\vec{x})|\,d\vec{x}$ by comparing both to the quantity $l^n\cdot|g(0)|\cdot|J_f(0)| = l^n\mu[C_l]$. Rather than write out the gory details I will just explain the general idea. First, we know the volume of $f(C_l)$ is close to $l^n\cdot|J_f|_0$ by the fact that $f(C_l)$ is nested between $T(C_{(1-\epsilon)l})$ and $T(C_{(1+\epsilon)l})$. We also know that g varies by less than ϵ over $f(C_l)$. Together, these facts imply that $\int_{f(C_l)} g(\vec{y})\,d\vec{y}$ is close to $l^n\cdot|g(0)|\cdot|J_f(0)|$. And we can estimate $\int_{C_l} g(f(\vec{x}))\cdot|J_f(\vec{x})|\,d\vec{x}$ by noting that the volume of C_l is l^n, that again g varies by less than ϵ over $f(C_l)$, and that $J_f(\vec{x})$ is within ϵ of $J_f(0)$ throughout C_l. In each case "close" can be described explicitly by ϵl^n times a formula involving only κ and ρ. We conclude that

$$\left|\int_{f(A)} g(\vec{y})\,d\vec{y} - \int_A g(f(\vec{x}))\cdot|J_f(\vec{x})|\,d\vec{x}\right| < M\epsilon\cdot l^n$$

for some constant M that depends only on λ, κ, and ρ. □

And here is one final, slightly more advanced theorem about integration.

Theorem 13.47. *(Radon-Nikodym theorem) Let X be a measure space with measure μ and let ν be a complex measure that is absolutely continuous*

with respect to μ. Then there exists a μ-integrable function $f : X \to \mathbb{C}$ such that $\nu[A] = \int_A f\,d\mu$ for every μ-measurable class A.

Proof. By writing ν as a linear combination of the positive and negative parts of its real and imaginary parts — all of which are absolutely continuous with respect to μ — we can reduce to the case where ν is positive.

We can also use the technique employed in the last part of the proof of Cavalieri's principle (Theorem 13.42) to reduce to the case where μ is finite. Now fix $n \in \mathbb{N}$ and for each $k \in \mathbb{N}$ let A_k be an M_δ class with $(\nu - \frac{k}{2^n}\mu)_-[A_k^c] = 0$ and $(\nu - \frac{k}{2^n}\mu)_+[A_k] = 0$. By replacing A_k with $A_1 \cup \cdots \cup A_k$, we can assume $A_1 \subseteq A_2 \subseteq \cdots$; then define $f_n = \sum_{k=1}^\infty \frac{k}{2^n} 1_{A_{k+1} \setminus A_k}$. For any μ-measurable subclass A of $A_{k+1} \setminus A_k$ we have

$$\int_A f_n\,d\mu = \frac{k}{2^n}\mu[A] \le \nu[A] \le \frac{k+1}{2^n}\mu[A] = \int_A \left(f_n + \frac{1}{2^n}\right) d\mu.$$

Also, for $B = (\bigcup A_k)^c$ we have $\nu[B] \ge \frac{k}{2^n}\mu[B]$ for all k, which implies that B is μ-null and hence ν-null. For any μ-measurable class A, summing over $A \cap (A_k \setminus A_{k-1})$, setting $A_0 = \emptyset$, yields

$$\int_A f_n\,d\mu \le \nu[A] \le \int_A f_n\,d\mu + \frac{\mu[X]}{2^n}.$$

Letting n vary, we have $f_1 \le f_2 \le \cdots$ almost everywhere, so the monotone convergence theorem (Theorem 13.34) now establishes that $\nu[A] = \int_A f\,d\mu$ where $f = \lim f_n$. □

In the preceding proof we affirmed that there exists a function with certain properties. This is an unbounded existential quantifier, so we need to be sure that an explicit description of the function can be extracted from the proof. In fact it can, because the classes A_k provided by Theorem 13.28 (Hahn-Jordan decomposition) were explicitly constructed in the proof of that result. Another minor point: in the proof the function $f = \lim f_n$ is only defined almost everywhere, but we can simply set $f(x) = 0$ for any $x \in X$ for which $\lim f_n(x)$ does not exist.

The function f in Theorem 13.47 is the *Radon-Nikodym derivative* of ν with respect to μ. It may be denoted $\frac{d\nu}{d\mu}$, but this is just a formal expression. It is convenient notation because of the following result.

Corollary 13.48. *Let X be a measure space with measure μ and let ν be a complex measure that is absolutely continuous with respect to μ. Then for any ν-integrable function g the function $g \cdot \frac{d\nu}{d\mu}$ is μ-integrable and satisfies*

$$\int g\,d\nu = \int g \cdot \frac{d\nu}{d\mu} d\mu.$$

Proof. We can reduce to the case that ν is positive in the usual way. The conclusion follows immediately from Theorem 13.47 when $g = 1_A$. It is then true for simple functions by linearity, for positive integrable functions by the monotone convergence theorem (Theorem 13.34), and then for all integrable functions by linearity again. □

Corollary 13.49. *Let ν be a complex measure on a measurable space X. Then there is a ν-measurable function $f : X \to \mathbb{C}$ with $|f(x)| = 1$ $|\nu|$-almost everywhere, such that*

$$\nu[A] = \int_A f \, d|\nu|$$

for any measurable class A.

Proof. Put $\mu = |\nu|$ in Theorem 13.47. This yields a measurable function f with $\nu[A] = \int_A f \, d|\nu|$ for every measurable class A.

For any measurable class A we have

$$|\nu[A]| = \left| \int_A f \, d|\nu| \right| \leq \int_A |f| \, d|\nu|,$$

and conversely, if $h = \sum a_i 1_{A_i}$ is any simple function with disjoint classes A_i whose union is A and with $|a_i| = 1$ for all i, then

$$|\nu|[A] \geq \sum |\nu[A_i]| \geq \left| \int_A h \cdot f \, d|\nu| \right|;$$

by letting h converge to $\frac{|f|}{f}$ pointwise and invoking the dominated convergence theorem (Theorem 13.39) we then get $\int_A |f| \, d|\nu| = \int_A \frac{|f|}{f} \cdot f \, d|\nu| \leq |\nu|[A]$. We conclude that $|\nu|[A] = \int_A |f| \, d|\nu|$, but we also have $|\nu|[A] = \int_A 1 \, d|\nu|$, and since A was arbitrary we must have $|f| = 1$ almost everywhere. □

13.4 Complex analysis

In Section 13.1 we talked about derivatives of functions defined on subclasses of $\mathbb{R}^n$. Now we will switch to the setting of complex numbers.

Definition 13.50. Let A be an open subclass of $\mathbb{C}$ and let f be a function from A to $\mathbb{C}$. Then f is *differentiable at $z \in A$* if the limit

$$\lim_{h \to 0} \frac{f(z+h) - f(z)}{h}$$

exists, and we define the *derivative* of f, denoted f', to be this limit when it exists. We say that f is *analytic* or *holomorphic* on A if it is differentiable at every point of A. If f is analytic on C then we say that it is *entire*.

In this definition it is important to note that h is a complex number, not a real number, tending to 0. Since we can approach 0 with purely real numbers and with purely imaginary numbers, the requirement that the limit exist is a strong restriction. Write $z = x + iy$ and $f(z) = u(z) + iv(z)$; if f is differentiable on some open region A then taking only real values of h yields that $\frac{\partial u}{\partial x}$ and $\frac{\partial v}{\partial x}$ exist on A and

$$f'(z) = \frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x},$$

while taking purely imaginary values of h yields that $\frac{\partial u}{\partial y}$ and $\frac{\partial v}{\partial y}$ exist on A and

$$f'(z) = -i\frac{\partial u}{\partial y} + \frac{\partial v}{\partial y}.$$

Equating the real and imaginary parts of the two expressions reveals that

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \qquad \text{and} \qquad \frac{\partial v}{\partial x} = -\frac{\partial u}{\partial y}.$$

These are the *Cauchy-Riemann equations*.

Conversely, if the partial derivatives of u and v exist and are continuous on some open region A, then f is differentiable as a map from $\mathbb{R}^2$ to $\mathbb{R}^2$ by Proposition 13.5. If the Cauchy-Riemann equations are satisfied, it follows that f is differentiable in the sense given here.

Any polynomial in z is analytic on $\mathbb{C}$. Here is another example of an entire function.

Definition 13.51. For $z \in \mathbb{C}$ define $\exp z = \sum_{n=0}^{\infty} \frac{z^n}{n!}$. This is the *exponential function*.

This series converges for all z by the ratio test (Proposition 10.26), and $\exp z$ is continuous by Corollary 10.31. Using Proposition 10.29, a short calculation shows that $\exp z \exp w = \exp(z+w)$. Defining e to be $\exp(1)$, it follows that $\exp(n) = e^n$ for any $n \in \mathbb{N}$. If $p = \frac{n}{m}$ is positive and rational we then have $(\exp(p))^m = \exp(mp) = \exp(n) = e^n$, so that $\exp(p) = e^p$. And for any positive rational number q we have $\exp(q)\exp(-q) = \exp(0) = 1$, so $\exp(-q) = \frac{1}{\exp(q)} = e^{-q}$. This suggests the notation $e^z = \exp z$, for any $z \in \mathbb{C}$.

For any $z \in \mathbb{R}$ we have

$$\frac{e^{z+h} - e^z}{h} = e^z \cdot \frac{e^h - 1}{h} \to e^z$$

as $h \to 0$ ($h \in \mathbb{C}$). Thus e^z is analytic on $\mathbb{C}$ and $(e^z)' = e^z$.

Next, for any $x \in \mathbb{R}$ define $\cos x$ and $\sin x$ to be the real and imaginary parts, respectively, of e^{ix}. Note that $|e^{ix}|^2 = e^{ix}e^{-ix} = 1$, so the point $(\cos x, \sin x)$ lies on the unit circle for all $x \in \mathbb{R}$. We can generalize to complex numbers by setting $\cos z = \frac{e^{iz}+e^{-iz}}{2}$ and $\sin(z) = \frac{e^{iz}-e^{-iz}}{2i}$. We can then use the fact that e^z is analytic on $\mathbb{C}$ to infer that $\cos z$ and $\sin z$ are also analytic on $\mathbb{C}$, with $(\cos z)' = -\sin z$ and $(\sin z)' = \cos z$. In terms of infinite series, we have $\cos z = 1 - \frac{z^2}{2!} + \frac{z^4}{4!} - \cdots$ and $\sin z = z - \frac{z^3}{3!} + \frac{z^5}{5!} - \cdots$, both of which converge absolutely for all z by the ratio test (Proposition 10.26).

Define π by letting $\frac{\pi}{2}$ be the smallest positive real number satisfying $\cos(\frac{\pi}{2}) = 0$. There must be such a value because cos is continuous and we have $\cos(0) = 1 > 0$ and, for instance, $\cos(2) < 0$, because $\cos(2)$ lies between $1 - \frac{2^2}{2!} = -1$ and $1 - \frac{2^2}{2!} + \frac{2^4}{4!} = -\frac{1}{3}$. We know that $(\cos(\frac{\pi}{2}), \sin(\frac{\pi}{2}))$ lies on the unit circle, so $\sin(\frac{\pi}{2}) = \pm 1$; since $(\sin x)' = \cos x$ and $\cos x > 0$ on $[0, \frac{\pi}{2})$, we must have $\sin(\frac{\pi}{2}) = 1$. Thus $e^{i\pi/2} = 0 + 1 \cdot i = i$ and $e^{\pi i} = (e^{i\pi/2})^2 = -1$.

On the region $0 \leq y < 2\pi$ the exponential function is injective and therefore has an inverse function ln that takes $\mathbb{C}^*$ onto this region. Differentiating the equation $e^{\ln z} = z$ using the chain rule yields $(\ln z)' = \frac{1}{z}$. One can also infer the other basic properties of the logarithm from corresponding properties of e^z.

In Section 9.9 I promised to prove that $(\cos\theta + i\sin\theta)(\cos\phi + i\sin\phi) = \cos(\theta + \phi) + i\sin(\theta + \phi)$. This is trivial now that we can rewrite it as $e^{i\theta}e^{i\phi} = e^{i(\theta+\phi)}$.

If $\gamma : [a, b] \to \mathbb{C}$ is continuous on $[a, b]$ and differentiable on (a, b) and $f : A \to \mathbb{C}$ is continuous, where $\gamma(t) \in A \subseteq \mathbb{C}$ for all $t \in [a, b]$, then we define the *line integral* $\int_\gamma f(z)\,dz$ to be $\int_a^b f(\gamma(t))\gamma'(t)\,dt$, provided the latter exists. Here γ is a *path*; if $\gamma(a) = \gamma(b)$ then it is *closed*. A closed path is also called a *loop*. This same formula can be used if γ fails to be differentiable at finitely many points.

Lemma 13.52. *Let f be analytic on an open class containing the rectangle*

$$R = \{z : a \leq \mathrm{Re}(z) \leq b,\, c \leq \mathrm{Im}(z) \leq d\}.$$

Let $\gamma : [0, 2(b-a) + 2(d-c)] \to \mathbb{C}$ be the loop which starts at $a + ic$ and

moves counterclockwise around the perimeter of the rectangle at unit speed. Then $\int_\gamma f(z)\,dz = 0$.

Proof. Divide R into four congruent sub-rectangles R^1, R^2, R^3, and R^4. Let γ (respectively, γ^i) be the loop which moves counterclockwise around the perimeter of R (respectively, R^i) at unit speed; then

$$\int_\gamma f(z)\,dz = \int_{\gamma^1} f(z)\,dz + \int_{\gamma^2} f(z)\,dz + \int_{\gamma^3} f(z)\,dz + \int_{\gamma^4} f(z)\,dz$$

because the integrals on the right cancel each other except for the loop segments on the boundary of R, where they agree with the integral on the left. Writing η for the integral on the left and η^i $(i = 1, 2, 3, 4)$ for the integrals on the right, one of these rectangles R^i must therefore satisfy

$$|\eta^i = \left|\int_{\gamma^i} f(z)\,dz\right| \geq \frac{1}{4}\left|\int_\gamma f(z)\,dz\right| = \frac{1}{4}\eta.$$

Set $R_1 = R^i$ and $\eta_1 = \eta^i$ and continue in this way to create rectangles R_2, R_3, …, each one half the dimensions of, and contained in, the previous one, such that the corresponding integrals satisfy $|\eta_n| \geq \frac{1}{4}|\eta_{n-1}|$ for all $n > 1$. Inductively we have $|\eta| \leq 4^n|\eta_n|$ for all n.

Let z^* be the single point contained in $\bigcap R_n$. A direct calculation shows that $\int_{\gamma_n} dz = \int_{\gamma_n} z\,dz = 0$ for each n, where γ_n is the loop which moves around the perimeter of R_n at unit speed. Thus $\int_{\gamma_n} f(z^*)\,dz = \int_{\gamma_n} z^* f'(z^*)\,dz = \int_{\gamma_n} z f'(z^*)\,dz = 0$ and therefore

$$\eta_n = \int_{\gamma_n} (f(z) - f(z^*) - (z - z^*)f'(z^*))\,dz.$$

Let $\epsilon > 0$ and choose n large enough that $\left|\frac{f(z)-f(z^*)}{z-z^*} - f'(z^*)\right| < \epsilon$ within R_n; this entails that

$$\begin{aligned}
|\eta_n| &\leq \epsilon \cdot \int_{\gamma_n} |z - z^*|\,dz \\
&\leq \epsilon \cdot \int_0^{L_n} |\gamma_n(t) - z^*||\gamma_n'(t)|\,dt \\
&= \epsilon \cdot \int_0^{L_n} |\gamma_n(t) - z^*|\,dt
\end{aligned}$$

where L_n is the length of the perimeter of R_n. If L is the length of the perimeter of R then we have $L_n = \frac{1}{2^n}L$. Also, since $\gamma_n(t)$ ranges over the boundary of R_n and $z^* \in R_n$, we have $|\gamma_n(t) - z^*| \leq D_n = \frac{1}{2^n}D$ where D_n is the length of a diagonal of R_n and D is the length of a diagonal of R. We conclude that $|\eta_n| \leq \epsilon \cdot D_n L_n = \epsilon \cdot \frac{1}{4^n}DL$. Recalling that $|\eta| \leq 4^n|\eta_n|$, this yields $|\eta| \leq \epsilon \cdot DL$. As ϵ was arbitrary, we get $\eta = 0$. □

The hypothesis of Lemma 13.52 can be weakened slightly to say that f is continuous on an open class U containing R and analytic on $U \setminus \{p\}$ for some p in R. If p lies on the boundary of R, then we can simply bring the sides of R in by a small amount ϵ and apply Lemma 13.52 to get that the integral over the shrunken rectangle is zero; taking ϵ to zero then shows that the integral around R is also zero. If p lies in the interior of R then subdivide R into two rectangles, each of which has p on its boundary, and reduce to the first case.

Lemma 13.53. *Suppose $\gamma : [a, b] \to \mathbb{C}$ is continuous on $[a, b]$ and differentiable on (a, b), $U \subseteq \mathbb{C}$ is an open region which contains the image of γ, and $f : U \to \mathbb{C}$ is continuous. If γ is closed and there exists an analytic function $F : U \to \mathbb{C}$ such that $F' = f$ on U, then $\int_\gamma f(z)\, dz = 0$.*

Proof. It follows from the multivariable chain rule (Proposition 13.7) that the function $F \circ \gamma$ is differentiable on (a, b) with $(F \circ \gamma)'(t) = F'(\gamma(t))\gamma'(t)$. Thus

$$\begin{aligned}\int_\gamma f(z)\, dz &= \int_a^b f(\gamma(t))\gamma'(t)\, dt \\ &= \int_a^b (F \circ \gamma)'(t)\, dt \\ &= (F \circ \gamma)(b) - (F \circ \gamma)(a) \\ &= 0\end{aligned}$$

by the fundamental theorem of calculus (Theorem 13.40). □

An *open disk* is an open ball in $\mathbb{C}$, in the metric space sense.

Theorem 13.54. *(Cauchy's theorem) Let $\Delta \subseteq \mathbb{C}$ be an open disk and let $f : \Delta \to \mathbb{C}$ be analytic. Then $\int_\gamma f(z)\, dz = 0$ for every loop γ whose image is contained in Δ.*

Proof. Without loss of generality suppose Δ is centered at the origin. For any $z \in \Delta$, $z = x + iy$, let $a = 0$ and $b = |x| + |y|$, and define $\gamma_1 : [a, b] \to \mathbb{C}$ to be the path which starts at the origin and moves horizontally at unit speed until it reaches the point x, and then moves vertically at unit speed until it reaches $x + iy$. Define γ_2 in the same way, but this time with the vertical segment first. Now define $F(z) = \int_{\gamma_1} f(z)\, dz = \int_{\gamma_2} f(z)\, dz$, where the two integrals are equal by Lemma 13.52.

Using the γ_1 definition we then see that $\frac{\partial F}{\partial y}(z) = if(z)$ and using the γ_2 definition we see that $\frac{\partial F}{\partial x}(z) = f(z)$. Thus the partial derivatives of F

exist and are continuous, and F satisfies the Cauchy-Riemann equations, so F is analytic and $F' = f$. The result now follow from Lemma 13.53. □

By the comment following Lemma 13.52, the hypothesis of Cauchy's theorem can be weakened to say that f is continuous on Δ and analytic on $\Delta \setminus \{p\}$ for some $p \in \Delta$. That lets us deduce the next result.

Corollary 13.55. *(Cauchy's formula) Let $\Delta \subseteq \mathbb{C}$ be an open disk with radius $R > 0$ and center $p \in \mathbb{C}$, and let $f : \Delta \to \mathbb{C}$ be analytic. Let γ be a loop that moves counterclockwise at unit speed about the circle of radius $r < R$ centered at p and let a belong to the interior of this circle. Then $f(a) = \frac{1}{2\pi i} \int_\gamma \frac{f(z)}{z-a}\, dz$.*

Proof. By translating we can assume $p = 0$. Define $F(z) = \frac{f(z)-f(a)}{z-a}$ for $z \neq a$ and $F(a) = f'(a)$. Then F is continuous on Δ and analytic on $\Delta \setminus \{a\}$, so the comment just before this corollary yields that

$$\int_\gamma \frac{f(z) - f(a)}{z - a}\, dz = 0,$$

that is,

$$\int_\gamma \frac{f(z)}{z - a}\, dz = f(a) \int_\gamma \frac{1}{z - a}\, dz.$$

The integral on the right equals $\int_0^{2\pi} \frac{1}{re^{it}-a}\, ire^{it}\, dt$. Recalling that we have defined $\ln(z)$ so that it is analytic on $\mathbb{C} \setminus [0, \infty)$, find a point $t_0 \in [0, 2\pi)$ with $\mathrm{Im}(re^{it} - a) = 0$ and then consider $\gamma(t) = re^{it}$ on $t_0 < t < t_0 + 2\pi$; then $\ln(re^{it} - a)$ is analytic on this region and its derivative is $\frac{ire^{it}}{re^{it}-a}$, so the fundamental theorem of calculus (Theorem 13.40) yields that $\int_{t_0}^{t_0+2\pi} \frac{1}{re^{it}-a}\, ire^{it}\, dt = \ln(re^{it_0+2\pi i} - a) - \ln(re^{it_0} - a) = 2\pi i$. Thus $\int_\gamma \frac{f(z)}{z-a}\, dz = 2\pi i f(a)$, as desired. □

Theorem 13.56. *(Power series representation) Let f be analytic in an open disk $\Delta \subset \mathbb{C}$ with radius R and center p. Then f has a power series expansion, i.e., for some choice of coefficients c_n we have*

$$f(z) = \sum_{n=0}^{\infty} c_n (z - p)^n$$

for all $z \in \Delta$, with absolute convergence on the right side.

Proof. Without loss of generality assume $p = 0$. Let γ be as in Cauchy's formula (Corollary 13.55). According to that result, if $|z| < r < R$ then we have $f(z) = \frac{1}{2\pi i} \int_0^{2\pi} \frac{f(re^{it})}{re^{it}-z} \cdot ire^{it}\, dt$ for all $z \in \Delta$.

For fixed z the geometric series $1 + \frac{z}{re^{it}} + \left(\frac{z}{re^{it}}\right)^2 + \cdots = \frac{1}{1-z/re^{it}}$ converges uniformly for all t, so the dominated convergence theorem (Theorem 13.39) yields that

$$\begin{aligned} f(z) &= \frac{1}{2\pi}\int_0^{2\pi} f(re^{it}) \cdot \left(1 + \frac{z}{re^{it}} + \left(\frac{z}{re^{it}}\right)^2 + \cdots\right) dt \\ &= \frac{1}{2\pi}\sum_{n=0}^{\infty}\left(\int_0^{2\pi} \frac{f(re^{it})}{r^n e^{int}}\, dt\right) z^n, \end{aligned}$$

with uniform convergence on $|z| < r$. □

Corollary 13.57. *(Identity theorem) Suppose f is analytic on a connected open region $U \subseteq \mathbb{C}$ and not constantly zero. Then every zero of f in U is isolated.*

Proof. Let a be a zero of f and suppose there is a disk centered at a and contained in U on which f is not constantly zero. Then we can write $f(z) = \sum c_n(z-a)^n$ for some coefficients c_n within this disk. If $m \geq 1$ is the smallest index with $c_m \neq 0$, then we have $f(z) = (z-a)^m(c_m + c_{m+1}(z-a) + \cdots)$, and neither factor of this product is zero near a.

We now know that every zero is either isolated or the center of a disk on which f is constantly zero. It follows that the union of all the disks with rational centers and radii on which f is constantly zero is clopen in U. Since f was assumed not to be constantly zero, connectedness implies that this union must be empty. □

Corollary 13.58. *(Liouville's theorem) Any bounded entire function is constant.*

Proof. Let f be an entire function. By the power series representation (Theorem 13.56) we can write $f(z) = \frac{1}{2\pi}\sum_{n=0}^{\infty} c_n z^n$ where $c_n = \int_0^{2\pi} \frac{f(re^{it})}{r^n e^{int}}\, dt$. If $|f(z)| \leq M$ for all z then for any $n \geq 1$

$$\left|\int_0^{2\pi} \frac{f(re^{it})}{r^n e^{int}}\, dt\right| \leq \int_0^{2\pi} \frac{M}{r^n}\, dt = \frac{2\pi M}{r^n}.$$

But $r > 0$ is arbitrary, so this shows that every coefficient c_n with $n \geq 1$ vanishes. □

Corollary 13.59. *(Maximum modulus principle) Let f be a nonconstant analytic function on a connected open region $U \subseteq \mathbb{C}$. Then $|f|$ has no local maximum within U.*

Proof. Let $a \in U$ and find $r > 0$ such that $\overline{B}_r(a) \subseteq U$. By translating U, we can assume $a = 0$. Then the power series representation (Theorem 13.56) lets us write $f(z) = \sum c_n z^n$. Since f is nonconstant on U, it must be nonconstant on $B_r(a)$, as otherwise the difference between it and a constant function will vanish on $B_r(a)$ and hence be constantly zero by the identity theorem (Corollary 13.57). Thus there is a minimal $m \geq 1$ with $c_m \neq 0$. Then $f(z) = c_0 + c_m z^m + \cdots$ and for small $|z|$ the remaining terms are negligible. But $c_0 + c_m z^m$ takes values throughout a circle of radius $|z|^m$ about c_0, so somewhere on that circle we must have $|f(z)| > |c_0|$. Thus f does not have a local maximum at a, and since the original choice of a was arbitrary we conclude that $|f|$ has no local maximum within U. □

Corollary 13.60. *(Infinite differentiability of analytic functions) Let $f(z) = \sum_{n=0}^{\infty} c_n(z-a)^n$, with absolute convergence in $B_R(a)$. Then f is analytic in $B_R(a)$ and its derivative is $f'(z) = \sum_{n=1}^{\infty} nc_n(z-a)^{n-1}$, also with absolute convergence within $B_R(a)$. Every analytic function is infinitely differentiable.*

Proof. Define $g(z) = \sum_{n=1}^{\infty} nc_n(z-a)^{n-1}$. By Proposition 10.28, the series for g has the same radius of convergence as the one for f. For each N, the derivative of $\sum_{n=0}^{N} c_n(z-a)^n$ is $\sum_{n=1}^{N} nc_n(z-a)^{n-1}$. But for any $r < R$ the former converge uniformly on $B_r(a)$ to f and the latter converge uniformly on $B_r(a)$ to g, and that suffices to establish that f is differentiable with derivative g. Thus f is analytic in $B_R(a)$.

Since g is itself represented by a power series, the same argument now implies that it is analytic in $B_R(a)$. Repeating this argument indefinitely, we see that f must be infinitely differentiable. By the power series representation (Theorem 13.56) this conclusion applies to any analytic function□

Corollary 13.61. *(Morera's theorem) Let $U \subseteq \mathbb{C}$ be open and let $f : U \to \mathbb{C}$ be continuous. If $\int_\gamma f(z)\,dz = 0$ for all loops in U then f is analytic.*

Proof. Define a function F on any disk Δ contained in U as in the proof of Cauchy's theorem (Theorem 13.54); now the two integrals defining F agree because the line integral of f over any rectangle contained in Δ is assumed to be zero. As in the proof of Theorem 13.54 we get that $F' = f$. Thus F is analytic, and then f is analytic by the infinite differentiability of analytic functions (Corollary 13.60). □

13.5 Ordinary differential equations

A function $F : U \to \mathbb{R}^n$, with $U \subseteq \mathbb{R}^m$ open, is *smooth* if its partial derivatives of all orders exist and are continuous. This is equivalent to each of its components $f_i : U \to \mathbb{R}$ being smooth, where $F(\vec{x}) = (f_1(\vec{x}), \ldots, f_n(\vec{x}))$.

A *system of ordinary differential equations* is a finite family of equations involving one or more functions of an independent variable t, together with their derivatives. If $U \subseteq \mathbb{R}^n$ is open and we are given functions $f_i : (-\epsilon, \epsilon) \times U \to \mathbb{R}$, $1 \leq i \leq n$, then we can write the system of n ordinary differential equations

$$g_i'(t) = f_i(t, g_1(t), \ldots, g_n(t)),$$

$1 \leq i \leq n$, as a single equation

$$G'(t) = F(t, G(t))$$

with $F(t, \vec{x}) = (f_1(t, \vec{x}), \ldots, f_n(t, \vec{x}))$ and $G(t) = (g_1(t), \ldots, g_n(t))$.

Theorem 13.62. *(Picard's existence theorem) Let $U \subseteq \mathbb{R}^n$ be open and let $\epsilon > 0$. Also let $F : (-\epsilon, \epsilon) \times U \to \mathbb{R}^n$ be smooth. Then for each $\vec{x}_0 \in U$ there exists $\delta > 0$ and an open neighborhood $V \subseteq U$ of $\vec{x}_0$ such that for all $\vec{a} \in V$ there is a unique smooth function $G : (-\delta, \delta) \to U$ which satisfies $G(0) = \vec{a}$ and $G'(t) = F(t, G(t))$ for $t \in (-\delta, \delta)$.*

Proof. Fix $\vec{x}_0 \in U$ and find $r \in (0, 1)$ such that $\overline{B}_{3r}(\vec{x}_0) \subseteq U$. Also let $\epsilon' < \epsilon$. Then find $M > 1$ such that the supremum of $\|F(t, \vec{x})\|$ over $[-\epsilon', \epsilon'] \times \overline{B}_{3r}(\vec{x}_0)$ is at most M, as is the supremum of $\frac{\|F(t,\vec{x}) - F(t,\vec{y})\|}{\|\vec{x} - \vec{y}\|}$ over all $t \in [-\epsilon', \epsilon']$ and distinct $\vec{x}, \vec{y} \in \overline{B}_{3r}(\vec{x}_0)$. This is possible because F is continuously differentiable on a neighborhood of $[-\epsilon', \epsilon'] \times \overline{B}_{3r}(\vec{x}_0)$. Let $\vec{a} \in B_r(\vec{x}_0)$.

Set $\delta = \frac{r}{M}$. Consider the pseudoclass $C([-\delta, \delta], \overline{B}_{2r}(\vec{x}_0))$. By Proposition 10.30 this is a complete metric space relative to the metric $d(\phi, \psi) = \sup_{t \in [-\delta, \delta]} \|\phi(t) - \psi(t)\|$. Let $\mathcal{F} \subset C([-\delta, \delta], \overline{B}_{2r}(\vec{x}_0))$ be the subclass consisting of those (codes of) ϕ with $\phi(0) = \vec{a}$. Clearly this is a closed subspace and therefore it is complete.

Define $L : \mathcal{F} \to \mathcal{F}$ by $L(\phi)(t) = \vec{a} + \int_0^t F(\tau, \phi(\tau))\, d\tau$, where the integral is performed on each component separately. Here $L(\phi)(0) = \vec{a}$ is obvious, and continuity of $L(\phi)$ follows from the fundamental theorem of calculus (Theorem 13.40). In addition, $L(\phi)$ maps $[-\delta, \delta]$ into $\overline{B}_{2r}(\vec{x}_0)$ because $\|\vec{a} - \vec{x}_0\| < r$ and $\|L(\phi)(t) - \vec{a}\| < r$ for $|t| \leq \delta = \frac{r}{M}$ by the mean value inequality (Corollary 13.6). Thus L does map $\mathcal{F}$ into itself.

For any $\phi, \psi \,\varepsilon\, \mathcal{F}$ we have

$$\begin{aligned}\|L(\phi)(t) - L(\psi)(t)\| &\leq \int_0^t \|F(\tau, \phi(\tau)) - F(\tau, \psi(\tau))\|\, d\tau \\ &\leq \delta M \cdot \sup_{t\in[-\delta,\delta]} \|\phi(t) - \psi(t)\| \\ &\leq \delta M \cdot d(\phi, \psi) = rd(\phi, \psi).\end{aligned}$$

So L is r-Lipschitz, and hence by the Banach fixed-point theorem (Theorem 10.7) it has a fixed point. That is, there exists $G \,\varepsilon\, \mathcal{F}$ with $L(G) = G$. Then $G(0) = \vec{a}$ and $G'(t) = F(t, G(t))$ for $t \in (-\delta, \delta)$ by Theorem 13.40. Since F is smooth, this shows that G is smooth too.

Finally, G is unique because any other solution would also be a fixed point for L, and a c-Lipschitz function with $c < 1$ cannot have more than one fixed point. □

Chapter 14

Algebraic Topology

14.1 Simplicial complexes

A nonempty finite subset $S = \{v_0, \ldots, v_p\}$ of a real vector space is in *general position* if the p vectors $v_1 - v_0$, ..., $v_p - v_0$ are linearly independent. A *convex combination* of the vectors v_i is a vector of the form $a_0v_0 + \cdots + a_pv_p$ where $a_i \geq 0$ for all i and $\sum a_i = 1$. It is *strict* if $a_i > 0$ for all i. The *convex hull* of S is the class $[S]$ of all convex combinations of the v_i's, and its *interior* (S) is the class of all strict convex combinations of the v_i's. General position ensures that every element of $[S]$ has a unique representation as a convex combination of the v_i's.

(The convex hull exists as a class because it is the closure of the set of all rational combinations of the stated form; the interior exists because it is the complement of the boundary, which is the union of the convex hulls of all p-element subsets of S.)

Our discussion of algebraic topology will center around the following kind of object. Recall that $\mathbb{R}^{<\omega}$ is the class of those sequences in $\mathbb{R}^\omega$ all but finitely many of whose entries are zero.

Definition 14.1. A *simplicial complex* is a set $K \subseteq \mathcal{P}(V_K)$ of nonempty finite subsets of some *vertex set* $V_K \subset \mathbb{R}^n$ (for some n) or $\mathbb{R}^{<\omega}$, such that

(1) any subset of a set in K is also a set in K
(2) every set in K is in general position
(3) if $S, S' \,\varepsilon\, K$ are distinct then $(S) \cap (S') = \emptyset$
(4) each $v \in V_K$ belongs to at least one, but only finitely many, sets in K.

An element of a simplicial complex is a *simplex*. A simplex with exactly $p + 1$ elements is a *p-simplex*.

Every simplicial complex K has a *carrier* $[K] = \bigcup_{S \varepsilon K}[S]$. The carrier of any simplicial complex is a *polytope*. I will use this term even if K is infinite. The classes $[S]$, for S a simplex of K, are the *faces* of $[K]$; if S is a p-simplex then $[S]$ is a *p-dimensional* face. An *edge* is a one-dimensional face; thus it is the line segment joining two vertices v and w. Any 0-simplex $S = \{v\}$ is also a face. In my terminology the point v itself is a vertex, but not a face.

If K is finite then $[K]$ inherits a compact topology from $\mathbb{R}^n$ or $\mathbb{R}^{<\omega}$. Every polytope can be written as an increasing union of compact polytopes, $[K] = \bigcup[K_n]$ with $K_1 \subseteq K_2 \subseteq \cdots$, so we can give $[K]$ the associated compactly coherent topology. This topology is independent of the specific choice of K_n's (Proposition 10.86 (d)). If K is infinite then according to Proposition 10.86 (d) $[K]$ cannot be compact. It is not hard to show that $[K]$ will always be locally compact (in our sense, which includes metrizability).

A *triangulation* of a locally compact space X is given by specifying a simplicial complex K and a homeomorphism between X and $[K]$. A locally compact space that has a triangulation is *triangulable*. (I.e., X is triangulable if there exists a simplicial complex K and a homeomorphism in $C(X, \mathbb{R}^n)$ or $C(X, \mathbb{R}^{<\omega})$ between X and $[K]$.)

Here is a very basic, almost trivial example.

Example 14.2. Let K be a simplicial complex with three vertices $1, e^{2\pi i/3}, e^{4\pi i/3} \in \mathbb{C} \cong \mathbb{R}^2$, and which contains all three possible 1-simplices but not the one possible 2-simplex. Then $[K]$ is a triangle and is homeomorphic to the unit circle $\mathbb{S}^1$ by a map which moves each point of $[K]$ radially into $\mathbb{S}^1$. This shows one way of triangulating $\mathbb{S}^1$.

The natural "morphisms" between simplicial complexes are described as follows.

Definition 14.3. Let K and L be simplicial complexes. A *simplicial transformation* from K to L is a function from V_K to V_L with the property that the image of any simplex of K is a simplex of L.

Any simplicial transformation ϕ from K to L extends to an *affine* function, which I will also denote ϕ, from $[K]$ to $[L]$. This function takes vertices to vertices according to ϕ and extends linearly to all faces. Thus if $S = \{v_0, \ldots, v_p\}$ is a simplex of K then ϕ maps $[S]$ into $[\phi(S)]$ via the formula

$$\phi(a_0 v_0 + \cdots + a_p v_p) = a_0 \phi(v_0) + \cdots + a_p \phi(v_p).$$

Since multiple simplicial complexes can have the same carrier, whether $\phi : [K] \to [L]$ is affine depends not just on the carriers $[K]$ and $[L]$, but also on the simplicial complexes K and L. However, it is convenient to just say "let $\phi : [K] \to [L]$ be affine" and take it as understood that this is relative to the K and L listed there.

Sometimes we need to refine a polytope by chopping up some of its faces. A *subdivision* of a simplicial complex K is a simplicial complex K' with the same carrier such that every face of $[K]$ is the union of finitely many faces of $[K']$. The simplest way to do this is *stellar subdivision*. Here one selects a single simplex $S = \{v_0, \ldots, v_p\}$ and adds a single vertex, the *barycenter* $v_S = \frac{1}{p+1}v_0 + \cdots + \frac{1}{p+1}v_p$ of $[S]$. Any simplex that contains S is replaced by $p+1$ simplices, each of which replaces one of the vertices of S by v_S, and then we must also add in all of their sub-simplices.

This process can be iterated. For instance, if K is finite we can subdivide each of its simplices once, starting with the p-simplices with maximal p, and after all of these are taken care of, continuing on to progressively smaller values of p. The end result of this process is called the *barycentric subdivision of* K. For any sequence $S_0 \subset \cdots \subset S_p$ of simplices of K we have a corresponding p-simplex $\{v_{S_0}, \ldots, v_{S_p}\}$ of the barycentric subdivision of K, and these are all the simplices of the barycentric subdivision. Geometrically, we subdivide $[K]$ by placing a new vertex at the center of each face and then chopping $[K]$ up in a natural way.

Barycentric subdivision is a special case of *iterated stellar subdivision*. By that I mean any subdivision obtained by applying a finite or infinite sequence of stellar subdivisions, making sure that any face of the original polytope gets subdivided only finitely many times.

We will use subdivision to relate arbitrary continuous functions between polytopes to affine functions.

The *star* of a vertex v of a polytope $[K]$ is the union of (S) over all simplices S that contain v. We denote it $\mathrm{St}(v)$. Note that this includes the vertex v itself, as every 0-dimensional simplex equals its own interior.

Definition 14.4. Let $f : [K] \to [L]$ be a continuous function between two polytopes. A *simplicial approximation* of f is an affine map $\phi : [K] \to [L]$ satisfying $f(\mathrm{St}(v)) \subseteq \mathrm{St}(\phi(v))$ for every vertex v of $[K]$.

If we are allowed to subdivide the domain, we can find a simplicial approximation of any continuous function between two polytopes. If K is finite, define its *mesh* to be the maximum length of an edge of $[K]$. Taking repeated barycentric subdivisions of a finite simplicial complex yields

subdivisions with arbitrarily small mesh.

Theorem 14.5. *Let $[K]$ and $[L]$ be polytopes and let $f : [K] \to [L]$ be continuous. If the image under f of the star of any vertex of K is contained in the star of some vertex of L, then there is a simplicial approximation $\phi : [K] \to [L]$ of f. If K is finite then there exists $\delta > 0$ such that every subdivision of K with mesh less than δ satisfies this condition.*

Proof. Suppose that for every vertex v of K the image under f of $\mathrm{St}(v)$ is contained in $\mathrm{St}(w)$ for some vertex w of L. Choose one such w for each v and define $\phi(v) = w$. We must show that ϕ is a simplicial transformation; then one easily sees that its affine extension is a simplicial approximation of f. To do this, let $S = \{v_0, \ldots, v_p\}$ be a simplex in K. Any point in (S) belongs to the star of every v_i, and therefore the image under f of such a point must belong to the star of every $\phi(v_i)$. But this image lies in the interior of a unique face of $[L]$, so every $\phi(v_i)$ must be a vertex of that face, which means that $\phi(S) = \{\phi(v_0), \ldots, \phi(v_p)\}$ must be a simplex of L.

Now suppose K is finite, so that $[K]$ is compact. The stars of the vertices of L constitute an open cover of $[L]$ and therefore their inverse images in $[K]$, the classes $f^{-1}(\mathrm{St}(w))$ for w a vertex of $[L]$, constitute an open cover of $[K]$. Thus, giving $[K]$ the metric it inherits from $\mathbb{R}^n$ or $\mathbb{R}^{<\omega}$, there exists $\delta > 0$ such that for every $x \in [K]$ the ball $B_\delta(x)$ is wholly contained in some $f^{-1}(\mathrm{St}(w))$ (existence of Lebesgue number, Theorem 10.37). Then any subdivision of K with mesh less than this δ satisfies the desired condition. □

Corollary 14.6. *(Simplicial approximation) Let $f : [K] \to [L]$ be a continuous function between two polytopes. Then we can find an iterated stellar subdivision K' of K and a simplicial approximation $\phi : [K'] \to [L]$ of f.*

Proof. If K is finite, this corollary is an immediate consequence of Theorem 14.5 together with the comment just preceding it which pointed out that repeated barycentric subdivisions will have arbitrarily small mesh. If K is infinite, write it as an increasing union $K = \bigcup K_n$ of finite subcomplexes K_n. This can be done in such a way that every simplex containing a vertex in K_n belongs to K_{n+1}. Thus, stellar subdivision applied to any simplex in $K \setminus K_{n+1}$ will not affect anything in K_n. That means that we can successively apply the second part of Theorem 14.5 to K_{n+2}, only subdividing (perhaps multiple times) simplices in $K_{n+2} \setminus K_{n+1}$, to ensure that the image under f of the star of any vertex of the domain is contained in the star of some vertex of the range. Since this will not affect anything in K_n,

the process can be repeated indefinitely, ultimately yielding a subdivision of K which satisfies the condition of the first part of Theorem 14.5. That theorem then yields the desired conclusion. □

With this tool, we will be able to study homotopy using simplicial complexes.

Definition 14.7. Let $f, g : X \to Y$ be continuous functions between locally compact topological spaces. Let A be a subclass of X on which f and g agree. Then a *homotopy relative to* A between f and g is a continuous function $F : X \times [0,1] \to Y$ satisfying

(a) $F(x,0) = f(x)$ for all $x \in X$
(b) $F(x,1) = g(x)$ for all $x \in X$
(c) $F(x,t) = f(x) = g(x)$ for all $x \in A$.

When such a homotopy exists, we say that f and g are *homotopic relative to* A and write $f \sim g \ (\text{rel}\, A)$.

To say that a homotopy exists is to say that there is an element of the pseudoclass $C(X \times [0,1], Y)$ with certain properties. Thus it is a bounded quantifier.

Often we just say that f and g are homotopic and write $f \sim g$, when the "rel A" is understood. I omit the easy proof of the next result.

Proposition 14.8. *Let $f, g, h : X \to Y$ be continuous functions between locally compact spaces, let $A \subseteq X$, and suppose $f|_A = g|_A = h|_A$. Then*

(a) $f \sim f \ (\text{rel}\, A)$
(b) if $f \sim g \ (\text{rel}\, A)$ then $g \sim f \ (\text{rel}\, A)$
(c) if $f \sim g \ (\text{rel}A)$ and $g \sim h \ (\text{rel}\, A)$ then $f \sim h \ (\text{rel}\, A)$.

Now here is why we can study homotopy simplicially. If K is a simplicial complex then a *subcomplex* of K is a subset $K_0 \subseteq K$ which is itself a simplicial complex.

Theorem 14.9. *Let K and L be simplicial complexes, let $f : [K] \to [L]$ be continuous, and let $K_0 \subseteq K$ be a subcomplex. Assume that $f|_{[K_0]} : [K_0] \to [L]$ is affine and let $\phi : [K] \to [L]$ be a simplicial approximation of f. Then f is homotopic to ϕ relative to $[K_0]$.*

Proof. Note first that ϕ has to agree with f on $[K_0]$, since every star in $[L]$ contains only one vertex, so if $f(v) = w$ is a vertex of $[L]$ (for $v \in V_{K_0}$)

then simplicial approximation implies that w must belong to $\mathrm{St}(\phi(v))$ and hence must equal $\phi(v)$.

Any $x \in [K]$ belongs to the interior of exactly one face $[S]$ of $[K]$, so that x belongs to $\mathrm{St}(v)$ for every vertex v of $[S]$. By simplicial approximation we then have $f(x) \in \mathrm{St}(\phi(v))$ for all such v. But also $f(x)$ lies in the interior of a face $[S']$ of $[L]$, and thus $\phi(v)$ must be a vertex of $[S']$ for every vertex v of $[S]$. Thus $f(x)$ and $\phi(x)$ lie in the same face and we can define $F(x,t) = (1-t)f(x) + t\phi(x)$, giving us the desired homotopy. □

The proof of Theorem 14.9 is trivial, but it is still a very important result. In combination with simplicial approximation (Corollary 14.6) it shows that every continuous function between polytopes is, after sufficient subdivision, homotopic to an affine function.

The homotopy used in the proof of Theorem 14.9 is called a *straight line homotopy*.

14.2 The fundamental group

An *n-route* in a simplicial complex K is a finite sequence $\omega = (\omega_0, \ldots, \omega_n)$, $n \in \mathbb{N}$, of vertices of K with the property that $\{\omega_i, \omega_{i+1}\}$ is a 0- or 1-simplex of K for $0 \leq i < n$.

If $\omega_i = \omega_{i+1}$ then we say this is a *repeated vertex*.

To each route ω is associated a path $\tilde{\omega} : [0,1] \to [K]$ which takes the value ω_i at the point $\frac{i}{n}$ $(0 \leq i \leq n)$ and which is linear on each interval $[\frac{i}{n}, \frac{i+1}{n}]$. I will call $\tilde{\omega}$ the path *associated to* ω. In this chapter we will adopt the convention that every path has domain $[0,1]$.

Fix a "base point" $e \in V_K$. An n-route ω with $\omega_0 = \omega_n = e$ is an *n-course based at e*. The path associated to any course is closed, i.e., a loop. I will also say that this loop is based at e and that e is the base point of $[K]$. The *trivial course* (based at e) is the course ω with $n = 1$ and $\omega_0 = \omega_1 = e$.

We say that two n-routes ω and ω' are *contiguous* if for all $i < n$ the set $\{\omega_i, \omega_{i+1}, \omega'_i, \omega'_{i+1}\}$ is a simplex of K. We do not assume that these four vertices are distinct.

Now let us define an equivalence relation on the courses of K. Let two courses ω and ω' based at e be equivalent if there exist courses ω^1, …, ω^k, all based at e, with $\omega = \omega^1$ and $\omega' = \omega^k$, and such that for every $i < k$ either ω^i and ω^{i+1} are contiguous or ω^{i+1} is obtained from ω^i by either inserting or deleting a repeated vertex. This is clearly an equivalence relation.

Concatenation of courses is compatible with this equivalence relation:

if ω is equivalent to ω' and χ is equivalent to χ' then $\omega \frown \chi$ is equivalent to $\omega' \frown \chi'$. Thus concatenation defines an operation on the quotient space of courses modulo equivalence. It is straightforward to check that this is a group operation, with the class of the trivial course being its identity element and the inverse of the class of ω being the class of the reverse route ω^r which satisfies $\omega_i^r = \omega_{n-i}$.

Definition 14.10. The *fundamental group* of $[K]$ is the group of courses based at e modulo equivalence. It is denoted $\pi_1([K])$ or $\pi_1([K], e)$.

Say that K is *connected* if there is a route from any vertex to any other vertex. This is equivalent to $[K]$ being connected in the topological sense. If K is connected, then $\pi_1([K])$ as a group does not depend on the choice of base point, because any course based at e can be converted into a course based at e' by pre- and post-concatenating with a route from e' to e and its reverse, which yields a homomorphism from $\pi_1([K], e)$ to $\pi_1([K], e')$; and courses based at e' can be converted to courses based at e in the same way, using the same route, yielding a homomorphism in the opposite direction which is easily seen to be the inverse of the first one. Thus any route from e to e' induces an isomorphism between $\pi_1([K], e)$ and $\pi_1([K], e')$. This justifies the notation "$\pi_1([K])$" as opposed to "$\pi_1([K], e)$".

We say that $[K]$ is *simply connected* if it is connected and $\pi_1([K])$ is trivial, consisting only of its identity element. That is, every course in K is equivalent to the trivial course.

The elements of the fundamental group of $[K]$ exactly represent the loops in $[K]$ based at e up to homotopy (rel $\{0, 1\}$). This will be shown in the next sequence of results. From here on I will omit the "rel $\{0, 1\}$".

For each $n \in \mathbb{N}$ let I_n be the simplicial complex in $\mathbb{R}$ with vertices 0, $\frac{1}{n}$, …, 1 and 1-simplices $\{0, \frac{1}{n}\}$, $\{\frac{1}{n}, \frac{2}{n}\}$, …, $\{\frac{n-1}{n}, 1\}$. Its carrier is the unit interval $[0, 1]$.

Lemma 14.11. *Let K be a simplicial complex with base point e. Then every loop in $[K]$ based at e is homotopic to the loop $\tilde{\omega}$ for some course ω based at e.*

Proof. Let $f : [0, 1] \to [K]$ be a loop based at e. We can apply Theorem 14.5 to find a value of n and a simplicial approximation $\phi : [I_n] \to [K]$ of f. Then let ω be the n-course with entries $\phi(0)$, $\phi(\frac{1}{n})$, …, $\phi(1)$, so that $\tilde{\omega} = \phi$. The proof is completed by an appeal to Theorem 14.9. □

Lemma 14.12. *Let ω and ω' be equivalent courses in a simplicial complex*

K. Then $\tilde{\omega}$ is homotopic to $\tilde{\omega}'$.

Proof. By Proposition 14.8 (c) it will suffice to consider the cases that ω and ω' are contiguous, or that ω' is obtained from ω by either inserting or deleting a repeated vertex. The latter case is left as an easy exercise. The former case is also easy, because if ω and ω' are contiguous, then for every $i < n$ the points ω_i, ω_{i+1}, ω'_i, and ω'_{i+1} all lie in a single face of $[K]$, so there is a straight line homotopy from $\tilde{\omega}$ to $\tilde{\omega}'$. □

Lemma 14.13. *Let ω and ω' be courses in a simplicial complex K and suppose $\tilde{\omega}$ and $\tilde{\omega}'$ are homotopic. Then ω and ω' are equivalent courses.*

Proof. By adding repeated vertices to either ω or ω', we can ensure that they have the same length n. This affects neither their equivalence nor the existence of a homotopy from $\tilde{\omega}$ to $\tilde{\omega}'$.

We have a continuous function $F : [0,1]^2 \to [K]$ with the property that $F(s,0) = \tilde{\omega}(s)$ and $F(s,1) = \tilde{\omega}'(s)$ for all s, and additionally $F(0,t) = F(1,t) = e$ for all t. Now $[0,1]^2$ is a polytope with four vertices, two horizontal edges, two vertical edges, and one diagonal edge, say from $(0,0)$ to $(1,1)$ (SW to NE). According to Theorem 14.5, there is a simplicial approximation of F on any subdivision with mesh less than some δ. Let us choose a subdivision with vertices appearing at the grid points $(\frac{i}{nk}, \frac{j}{nk})$ with $0 \le i, j \le nk$ and k chosen sufficiently large that a square with side length $\frac{1}{nk}$ has diagonal shorter than δ. The subdivision will include all the obvious horizontal and vertical line segments (from $(\frac{i}{nk}, \frac{j}{nk})$ to $(\frac{i+1}{nk}, \frac{j}{nk})$ and to $(\frac{i}{nk}, \frac{j+1}{nk})$) as well as, on each $\frac{1}{nk} \times \frac{1}{nk}$ square, a diagonal edge connecting either the NW and SE corners or the NE and SW corners. Let us alternate between the two diagonal choices in a checkerboard pattern; thus we have an edge from $(\frac{i}{nk}, \frac{j}{nk})$ to $(\frac{i+1}{nk}, \frac{j+1}{nk})$ if $i + j$ is even, and from $(\frac{i+1}{nk}, \frac{j}{nk})$ to $(\frac{i}{nk}, \frac{j+1}{nk})$ if $i + j$ is odd. (Draw a picture!)

Let Φ be a simplicial approximation of F for this subdivision, and for $0 \le j \le nk$ let ω^j be the nk-course whose vertices are the points $\Phi(0, \frac{j}{nk})$, $\Phi(\frac{1}{nk}, \frac{j}{nk})$, …, $\Phi(1, \frac{j}{nk})$. Modify ω by inserting $k-1$ copies of ω_i between ω_i and ω_{i+1}, for each $i < n$, increasing the length of ω to $nk+1$, creating a new course which is equivalent to ω; but it is also contiguous to ω^0, because simplicial approximation means that $\Phi(\frac{i'}{nk}, 0) = \omega_i$ or ω_{i+1} when $ik \le i' \le (i+1)k$. (The point $\tilde{\omega}(\frac{i'}{nk})$ lies on the edge joining ω_i and ω_{i+1}, since $\tilde{\omega}$ is affine.) Thus ω is equivalent to ω^0, and a similar argument shows that ω' is equivalent to ω^{nk}. It will now suffice to show that for all $j < nk$ the courses ω^j and ω^{j+1} are equivalent. We do this by interpolating a

zigzag course whose vertices are alternately ω_i^j and ω_i^{j+1} for $i = 0, \ldots, nk$. This is possible because we have placed the diagonals in a checkerboard pattern. But both ω^j and ω^{j+1} are contiguous with the zigzag course (Φ is a simplicial transformation), and therefore they are equivalent to each other. We conclude that ω and ω' are equivalent. □

We have proven:

Theorem 14.14. *Let K be a simplicial complex with base point e. Then every loop in $[K]$ based at e is homotopic to the loop $\tilde{\omega}$ for some course ω based at e, and two courses ω and ω' are equivalent if and only if $\tilde{\omega}$ is homotopic to $\tilde{\omega}'$.*

Thus $\pi_1([K])$ exactly captures the notion "loops based at e modulo homotopy". Every loop based at e is represented by exactly one element of $\pi_1([K])$, and two loops are homotopic if and only if they are represented by the same element of $\pi_1([K])$. Since the notions of "loop" and "homotopy" are purely topological, it follows that $\pi_1([K])$ is invariant under homeomorphisms, i.e., if $[K]$ and $[L]$ are homeomorphic then $\pi_1([K]) \cong \pi_1([L])$ (assuming base points are chosen in a compatible manner, although if K and L are connected then it doesn't matter). Therefore, it is sensible to use the notation $\pi_1(X)$ for any connected triangulable locally compact space X.

To illustrate this, let us identify the fundamental group of a circle. In the next section I will present a tool which easily solves this problem for us, but here is an argument that goes directly from the definition. In Example 14.2 I described a simplicial complex K with three vertices whose geometric realization is homeomorphic to $\mathbb{S}^1$. Fix one of the three elements as the base point and let ω be any course based there. We can remove all repeated vertices to get an equivalent course ω' that is either trivial or has no repeated vertices. At that point, if ω' goes only clockwise or only counterclockwise, it must be a course with $3n + 1$ elements for some $n \in \mathbb{N}$, where n is the number of times it goes all the way around, either clockwise or counterclockwise. But if for some i the ω'_i to ω'_{i+1} step goes clockwise and for some j the ω'_j to ω'_{j+1} step goes counterclockwise, then since there are no repeated vertices there must be someplace where the direction switches, i.e., $\omega'_i = \omega'_{i+2}$. But then ω' is contiguous to the course ω'' which agrees with ω' at every point except $i+1$, where it takes the value ω'_i. We can then remove repeated vertices to get a shorter course which is equivalent to ω. Keep doing this until we reach a course which goes only

purely clockwise or counterclockwise (or is trivial). This shows that every element of the fundamental group is represented either by a trivial course or by a course with $3n + 1$ elements that wraps around n times, either clockwise or counterclockwise.

We can now define a homomorphism from $\mathbb{Z}$ to $\pi_1([K])$ which takes 0 to the identity of $\pi_1([K])$, $n > 0$ to the course that wraps around counterclockwise n times, and $n < 0$ to the course that wraps around clockwise $-n$ times. The argument made just above shows that this homomorphism is surjective. I claim it is an isomorphism; to see this we need only show that it has trivial kernel, i.e., a course that wraps around n times is not equivalent to the trivial course. That is because every step of any course either stays at the same point or moves clockwise or counterclockwise, and we can let m be the total number of counterclockwise steps minus the total number of clockwise steps. It is clear that inserting or deleting repeated vertices does not affect the value of m, and the same is true of passing to a contiguous course. For instance, if ω' is another course which completely agrees with ω except at i, and ω and ω' are contiguous, then the points $\omega_{i-1} = \omega'_{i-1}$, ω_i, and ω'_i must all belong to the same simplex, as must the points ω_i, ω'_i, and $\omega_{i+1} = \omega'_{i+1}$. An easy examination of cases then shows that ω and ω' have the same value of m. Since any two contiguous courses can be linked by a sequence of single-step contiguities, this completes the argument that any two equivalent courses have the same value of m. Thus a course that wraps around $n = \frac{|m|}{3}$ times, $n \in \mathbb{N}$, cannot be equivalent to a trivial course with $m = 0$. We conclude that $\pi_1(\mathbb{S}^1)$ is isomorphic to $\mathbb{Z}$.

Suppose K and L are simplicial complexes and ϕ is a simplicial transformation from K to L. Fix a base point e of K and designate $e' = \phi(e)$ as the base point of L. Then any course $\omega = (\omega_0, \ldots, \omega_n)$ in K based at e pushes forward to a course $\phi_*(\omega) = (\phi(\omega_0), \ldots, \phi(\omega_n))$ in L based at e'. Moreover, equivalent courses push forward to equivalent courses, so we get a function from $\pi_1([K], e)$ to $\pi_1([L], e')$. This function is a homomorphism. Moving to the topological setting, if $f : [K] \to [L]$ is any continuous map with $f(e) = e'$, we can subdivide K and find a simplicial approximation ϕ of f (Corollary 14.6), and then use this ϕ to push forward from $\pi_1([K])$ to $\pi_1([L])$. We denote this homomorphism $f_* : \pi_1([K]) \to \pi_1([L])$ and call it the *pushforward map*. If g is a loop in $[K]$ based at e then $f \circ g$ is a loop in $[L]$ based at e', and f_* takes the element of $\pi_1([K])$ represented by g to the element of $\pi_1([L])$ represented by $f \circ g$. This shows that the definition of f_* is independent of the choice of subdivision and the choice of ϕ.

Here is a nice application of pushforward maps. It will be generalized

to higher dimensions in Section 14.5. Let $\mathbb{D}^2$ be the closed unit disk in $\mathbb{R}^2$.

Theorem 14.15. *(Brouwer's fixed point theorem for $\mathbb{D}^2$) Every continuous function from the closed unit disk to itself has a fixed point.*

Proof. Let $f : \mathbb{D}^2 \to \mathbb{D}^2$ be continuous. We can diagnose whether f has a fixed point by evaluating $\inf\{\|x - f(x)\| : x \in \mathbb{D}^2 \text{ has rational coordinates}\}$: if this infimum is zero then there is a fixed point, otherwise there is not. So we can use a proof by contradiction to establish the existence of a fixed point.

Suppose there is no fixed point. Then for every $x \in \mathbb{D}^2$ there is a unique ray which begins at $f(x)$, extends through $x \neq f(x)$, and goes on to pass through a point y in the unit circle $\mathbb{S}^1$ (i.e., the boundary of $\mathbb{D}^2$). Define $g : \mathbb{D}^2 \to \mathbb{S}^1$ by $g(x) = y$. Continuity of f implies continuity of g, and it is clear that g fixes every element of $\mathbb{S}^1$.

After triangulating $\mathbb{D}^2$ and $\mathbb{S}^1$ we can find a simplicial approximation $\phi : \mathbb{D}^2 \to \mathbb{S}^1$ of g. The fundamental group of $\mathbb{S}^1$ is generated by (the class of) a loop that goes around the circle once, and this is also a loop in $\mathbb{D}^2$, so the fact that g fixes every element of the boundary (and hence is already affine on $\mathbb{S}^1$ with respect to any triangulation) means that the pushforward map $g_* : \pi_1(\mathbb{D}^2) \to \pi_1(\mathbb{S}^1)$ is surjective. But every loop in $\mathbb{D}^2$ is homotopic to a constant loop (straight line homotopy), so $\pi_1(\mathbb{D}^2)$ contains only an identity element. Therefore its image cannot be all of $\pi_1(\mathbb{S}^1) \cong \mathbb{Z}$, a contradiction. We conclude that f must have had a fixed point. □

14.3 Calculating $\pi_1(X)$

Actually calculating fundamental groups straight from the definition is usually more or less impossible, but there are good tools that can help. We will discuss two such tools in this section; here is the first.

Definition 14.16. Let $[K]$ be a connected polytope. A *cover* of $[K]$ consists of a connected polytope $[L]$ together with an affine map $\pi : [L] \to [K]$ such that for each vertex v of $[K]$ the class $\pi^{-1}(\mathrm{St}(v))$ is a disjoint union of open classes, each of which is taken homeomorphically onto $\mathrm{St}(v)$ by π. Here π is the *covering map*. A *universal* cover is a cover that is simply connected. An affine homeomorphism $f : [L] \cong [L]$ is a *deck transformation* if $\pi \circ f = \pi$.

Universal covers always exist.

Lemma 14.17. *Let ω be a route in a connected simplicial complex K and let $[L]$ be a cover of $[K]$, with covering map π. For any $v \in \pi^{-1}(\omega_0)$ there is a unique route ω' in L with $\omega'_0 = v$ and $\pi(\omega'_i) = \omega_i$ for all i. If ω is a course that is equivalent to the trivial course then ω' is too.*

Proof. Since the covering map is a local homeomorphism there is exactly one choice for each ω'_{i+1} given the value of ω'_i. If ω is a course based at e then neither inserting or removing vertices nor passing to a contiguous course based at e affects the starting or ending points of the lifts, and the lifts are related in the same way (by inserting or removing vertices, or passing to a contiguous route). So an equivalence of ω to the trivial course in K lifts to an equivalence of ω' to the trivial course in L. □

At this point it will be helpful to introduce the notion of an *abstract simplicial complex*. This is simply a set $\mathcal{K} \subseteq \mathcal{P}(V_\mathcal{K})$ of nonempty finite subsets of some abstract vertex set $V_\mathcal{K}$, such that (1) any subset of a set in $\mathcal{K}$ is also a set in $\mathcal{K}$, and (2) every vertex belongs to at least one, but only finitely many, sets in $\mathcal{K}$. The vertices of any abstract simplicial complex can be situated in $\mathbb{R}^n$ or $\mathbb{R}^{<\omega}$ so as to produce a simplicial complex in our original sense. The simplest way to do that is to replace its abstract vertices with the standard basis vectors in $\mathbb{R}^n$ or $\mathbb{R}^{<\omega}$.

Theorem 14.18. *Every connected simplicial complex has a universal cover.*

Proof. Let K be a connected simplicial complex with base point e. We define an abstract simplicial complex $\mathcal{K}$ as follows. The abstract vertices of $\mathcal{K}$ are the routes in K starting at e, modulo the equivalence relation which makes two routes equivalent if they terminate at the same vertex of K and one followed by the other's reverse is equivalent to the trivial course.

For any p-simplex $S = \{v_0, \ldots, v_p\}$ in K and any route ω from e to v_0, we introduce a p-simplex in $\mathcal{K}$ whose elements are (the classes of) ω, $\omega \frown v_1$, …, $\omega \frown v_p$. This makes $\mathcal{K}$ into an abstract simplicial complex. We take (the class of) the trivial course in K to be the base point of $\mathcal{K}$, and denote it $\tilde{e}$. Fix a geometric realization $\widetilde{K}$ of $\mathcal{K}$.

$[\widetilde{K}]$ is a cover of $[K]$ with covering map π that takes (the geometric realization of the class of) a route ω based at e to its terminal point, and extends linearly to p-simplices with $p \geq 1$. It is clear that this does describe a cover of $[K]$.

Every route ω' in $\widetilde{K}$ based at $\tilde{e}$ projects onto a route ω in K, and the terminal point of ω' is the class of ω. So if ω' is a course in $\widetilde{K}$ based at

$\tilde{e}$ then it projects onto a course ω that is equivalent to the trivial course based at e, which implies that ω' is equivalent to a trivial course by Lemma 14.17. This shows that $\widetilde{K}$ is simply connected. □

Moreover, they are essentially unique.

Proposition 14.19. *Let K be a connected simplicial complex and let $\pi_1 : [K_1] \to [K]$ and $\pi_2 : [K_2] \to [K]$ be universal covers of $[K]$. Then there is an affine homeomorphism $f : [K_1] \cong [K_2]$ satisfying $\pi_2 \circ f = \pi_1$.*

Proof. Let e be the base point of K and find $e_1 \in V_{K_1}$ and $e_2 \in V_{K_2}$ satisfying $\pi_1(e_1) = e = \pi_2(e_2)$. Take these to be the base points of K_1 and K_2. Then define a simplicial transformation from K_1 to K_2 as follows. For any vertex v in $[K_1]$ there is a route in K_1 from e_1 to v that pushes down to a route in K, which can then, by Lemma 14.17, be lifted to a route in K_2 starting at e_2 and terminating at some point w. Since $[K_1]$ is simply connected, for any two routes $\tilde{\omega}_1$ and $\tilde{\omega}_2$ from e_1 to v the concatenation $\tilde{\omega}_1 \frown \tilde{\omega}_2^r$ is equivalent to the trivial course based at e_1; it follows that they push down to routes ω_1 and ω_2 in K for which $\omega_1 \frown \omega_2^r$ is equivalent to the trivial course based at e, and then Lemma 14.17 says that the lifts of ω_1 and ω_2 to K_2 satisfy the condition that one followed by the reverse of the other is a course. This shows that the point w at which the lifts of ω_1 and ω_2 terminate does not depend on the choice of route from e_1 to v. We can now define an affine function that takes each vertex v of $[K_1]$ to the corresponding vertex w of $[K_2]$. The same construction with K_1 and K_2 interchanged describes an affine map from $[K_2]$ to $[K_1]$ which is the inverse of the first map. Therefore $[K_1]$ and $[K_2]$ are affinely homeomorphic, in a way which is compatible with the covering maps. □

If you can guess what the universal cover looks like, you don't need Theorem 14.18 to prove it exists. You just have to check that the space you have in mind is simply connected and really is a cover. For instance, let K be the simplicial complex of Example 14.2 for which $[K]$ is homeomorphic to a circle. Its universal cover can be identified as the simplicial complex $\widetilde{K}$ with vertex set $\mathbb{Z} \subset \mathbb{R}$ and which has the 0-simplices $\{n\}$ and the 1-simplices $\{n, n+1\}$ for all $n \in \mathbb{Z}$, but nothing else. Then $[\widetilde{K}] = \mathbb{R}$, which is simply connected (straight line homotopy), and it covers $[K]$ by the affine map which takes n to $e^{2n\pi i/3}$. So it is a universal cover.

The fundamental group of a polytope can be identified using universal covers in the following way. The deck transformations of a cover $[L]$ constitute a subgroup of Homeo$([L])$.

Theorem 14.20. *Let $[K]$ be a connected polytope with universal cover $[\widetilde{K}]$. Then the group of deck transformations of $[\widetilde{K}]$ is isomorphic to $\pi_1([K])$.*

Proof. It is clear that the deck transformations form a group. Let e be the base point of K and fix a base point $e' \in [\widetilde{K}]$ with $\pi(e') = e$. For any deck transformation $f : [\widetilde{K}] \to [\widetilde{K}]$ let $v = f(e')$, find a route in $\widetilde{K}$ from e' to v, and project it to a course in K based at e. I claim that any two routes in $\widetilde{K}$ from e' to v project to equivalent courses in K, so we get a map from the group of deck transformations into $\pi_1([K])$. I also claim that this map is an isomomorphism.

First, if ω_1 and ω_2 are two routes in $\widetilde{K}$ from e' to v then $\omega_1 \frown \omega_2^r$ is a course based at e', and since $\widetilde{K}$ is simply connected it is equivalent to the trivial course based at e'. Projecting down to K therefore yields a course in K that is equivalent to the trivial course based at e. This means that the projection of ω_1 followed by the projection of ω_2^r is a course that represents the identity element of $\pi_1([K])$, and therefore the projections of ω_1 and ω_2 represent the same element of $\pi_1([K])$. We have established that the construction given above induces a function ψ from the group of deck transformations into $\pi_1([K])$. It is clearly a homomorphism.

If two courses in K are equivalent then one followed by the reverse of the other is equivalent to the trivial course based at e, and therefore lifts to a course in $\widetilde{K}$ that is equivalent to the trivial course there (Lemma 14.17), in particular, that both starts and ends at e'. This shows that the lifts of the two courses terminate at the same point. Moreover, if two deck transformations take e' to the same point v, then they take any route in $\widetilde{K}$ starting at e' to the same route starting at v. Since every vertex in $\widetilde{K}$ is the endpoint of some route starting at e', this shows that the two deck transformations must be equal. Thus if two deck transformations are distinct then they take e' to two different points, and therefore routes from e' to the two points project down to inequivalent courses in K. So our map from the deck transformations to $\pi_1([K])$ is injective.

Finally, let ω be a course in K based at e. According to Lemma 14.17, it lifts to a route in $\widetilde{K}$ that starts at e'. To establish surjectivity we must show that if v is the endpoint of this route then there is a deck transformation that takes e' to v. We construct the desired deck transformation as follows. For any vertex w of $\widetilde{K}$ find a route from e' to w, project this down to a route in K, and then use Lemma 14.17 to lift it to a route starting at v. The endpoint of that lift indicates the image of w under the desired deck transformation that takes e' to v. □

Going back to Example 14.2, the deck transformations of the universal cover described above are exactly the maps $n \mapsto n + 3k$ for some $k \in \mathbb{Z}$, and the group of deck transformations is isomorphic to $\mathbb{Z}$. We infer that $\pi_1(\mathbb{S}^1) \cong \mathbb{Z}$.

I promised two tools for computing fundamental groups. The first one involved universal covers; the second relates the fundamental group of a simplicial complex to the fundamental groups of two subcomplexes. We need the following notions from group theory.

Definition 14.21. Let G_1 and G_2 be countable groups. Define an equivalence relation $\sim$ on the set S of all finite sequences of elements of $G_1 \sqcup G_2$ — *words* in $G_1 \sqcup G_2$ — by making two words equivalent if one can be transformed into the other by a finite sequence of moves of the form (1) removing or inserting an instance of the identity element of G_1 or the identity element of G_2, or (2) replacing two consecutive entries both lying in either G_1 or G_2 with the single entry consisting of their product in G_1 or G_2, or replacing a single entry with a pair that has that entry as it product. Then $S/\sim$ is a group, with group operation given by concatention of words. This is the *free product* of G_1 and G_2, denoted $G_1 * G_2$.

Now let F be another countable group and let $\psi_1 : F \to G_1$ and $\psi_2 : F \to G_2$ be homomorphisms, let $\widetilde{G}_1$ and $\widetilde{G}_2$ be the obvious copies of G_1 and G_2 sitting inside $G_1 * G_2$, and let $\tilde{\psi}_1$ and $\tilde{\psi}_2$ be the corresponding homomorphisms from F into $\widetilde{G}_1$ and $\widetilde{G}_2$. Let N be the set of all elements of $G_1 * G_2$ which can be expressed as a finite product of elements of the form $h\tilde{\psi}_1(g)\tilde{\psi}_2(g)^{-1}h^{-1}$ or $h\tilde{\psi}_2(g)\tilde{\psi}_1(g)^{-1}h^{-1}$ with $h \in G_1 * G_2$ and $g \in F$. This is a normal subgroup of $G_1 * G_2$, the smallest normal subgroup containing $\tilde{\psi}_1(g)\tilde{\psi}_2(g)^{-1}$ for all $g \in F$. The quotient of $G_1 * G_2$ by N is the *amalgamated free product* of G_1 and G_2 relative to ψ_1 and ψ_2, denoted $G_1 *_F G_2$.

Definitions of amalgamated free products usually specify that ψ_1 and ψ_2 are injective, but we will require the greater generality of the definition used here.

Free products are characterized by a universal property.

Proposition 14.22. *Let G_1, G_2, and H be countable groups and let $\phi_1 : G_1 \to H$ and $\phi_2 : G_2 \to H$ be homomorphisms. Also let $\theta_1 : G_1 \to G_1 * G_2$ and $\theta_2 : G_2 \to G_1 * G_2$ be the natural embeddings. Then there is a unique homomorphism $\phi : G_1 * G_2 \to H$ which satisfies $\phi_1 = \phi \circ \theta_1$ and $\phi_2 = \phi \circ \theta_2$. This property characterizes $G_1 * G_2$ up to an isomorphism respecting the*

embeddings θ_1 and θ_2.

If F is another countable group and $\psi_1 : F \to G_1$ and $\psi_2 : F \to G_2$ are homomorphisms, and we assume that $\phi_1 \circ \psi_1 = \phi_2 \circ \psi_2$, then there is a unique homomorphism $\tilde{\phi} : G_1 *_F G_2 \to H$ which satisfies $\phi_1 = \tilde{\phi} \circ \tilde{\theta}_1$ and $\phi_2 = \tilde{\phi} \circ \tilde{\theta}_2$ where $\tilde{\theta}_1 : G_1 \to G_1 *_F G_2$ and $\tilde{\theta}_2 : G_2 \to G_1 *_F G_2$ are the natural homomorphisms. We also have $\tilde{\theta}_1 \circ \psi_1 = \tilde{\theta}_2 \circ \psi_2$. These properties characterize $G_1 *_F G_2$ up to an isomorphism respecting the homomorphisms $\tilde{\theta}_1$ and $\tilde{\theta}_2$.

Proof. For the first part, we define $\phi : G_1 * G_2 \to H$ as follows. Given a word in $G_1 \cup G_2$, apply ϕ_1 to each entry from G_1 and ϕ_2 to each entry from G_2, then evaluate the product in H. For this to define a map ϕ from $G_1 * G_2$ to H, we must show that equivalent words map to the same element of H, but that is practically trivial. It is also trivial to check that ϕ is a homomorphism. Uniqueness follows from the simple fact that every element of $G_1 * G_2$ is a product of elements of $\widetilde{G}_1$ and $\widetilde{G}_2$, whose images in H are determined by ϕ_1 and ϕ_2.

The analogous conclusions can be drawn for amalgamated free products by essentially the same arguments. Here one must check that the homomorphism from $G_1 * G_2$ to H vanishes on all elements of the form $h\tilde{\psi}_1(g)\tilde{\psi}_2(g)^{-1}h^{-1}$ or $h\tilde{\psi}_2(g)\tilde{\psi}_1(g)^{-1}h^{-1}$ with $h \in G_1 * G_2$ and $g \in F$. This is a trivial consequence of the condition that $\phi_1 \circ \psi_1 = \phi_2 \circ \psi_2$.

To show that the universal properties characterize $G_1 * G_2$ and $G_1 *_F G_2$, let H' be any group with the same properties relative to some homomorphisms from G_1 and G_2 into H'. The universal properties then provide homomorphisms in both directions between H' and $G_1 * G_2$ or $G_1 *_F G_2$, and uniqueness guarantees that the composition of the two homomorphisms, in either order, is the identity map. □

This construction will now be used to evaluate the fundamental group of a simplicial complex in terms of the fundamental groups of two subcomplexes.

Theorem 14.23. *(Seifert-van Kampen theorem) Let K be a connected simplicial complex and let K_1 and K_2 be connected subcomplexes such that $K_0 = K_1 \cap K_2$ is connected and $K = K_1 \cup K_2$. Fix $e \in V_{K_0}$ and use this as a base point for K_0, K_1, K_2, and K. Letting $\psi_1 : \pi_1([K_0]) \to \pi_1([K_1])$ and $\psi_2 : \pi_1([K_0]) \to \pi_1([K_2])$ be the natural maps, we get that $\pi_1([K])$ is isomorphic to the amalgamated free product $\pi_1([K_1]) *_{\pi_1([K_0])} \pi_1([K_2])$.*

Proof. We prove this by showing that $\pi_1([K])$ has the universal prop-

erty described in Proposition 14.22. Thus let H be a group and let $\phi_1 : \pi_1([K_1]) \to H$ and $\phi_2 : \pi_1([K_2]) \to H$ be homomorphisms satisfying $\phi_1 \circ \psi_1 = \phi_2 \circ \psi_2$. Also let $\tilde{\theta}_1 : \pi_1([K_1]) \to \pi_1([K])$ and $\tilde{\theta}_2 : \pi_1([K_2]) \to \pi_1([K])$ be the natural maps. It is clear that $\tilde{\theta}_1 \circ \psi_1 = \tilde{\theta}_2 \circ \psi_2$.

We define a map from courses in K to H as follows. First, for each vertex v of K_0, fix a route ω^v from e to v in K_0. Then, given any course ω in K, for each $1 \leq i \leq n-1$ such that ω_i belongs to V_{K_0}, replace the single vertex $v = \omega_i$ with the route $(\omega^v)^r \frown \omega^v$. Also add or remove repeated e's to ensure that every maximal string of e's within $(\omega_1, \ldots, \omega_{n-1})$ has even length. We can then write ω (so modified) in a unique way as a concatenation of courses each of which lies entirely in K_0, K_1, or K_2, and in each of which e does not appear anywhere in the course besides the first and last entries. Each course can then be mapped into H using ψ_1 on courses in K_1 and ψ_2 on courses in K_2 (noting that these agree on courses in K_0). We must show that this map lifts to $\pi_1([K])$, i.e., any two equivalent courses map to the same element of H. Given this, it is easy to see that the result is a homomorphism $\tilde{\phi}$ from $\pi_1([K])$ to H that has the desired universal property that $\phi_1 = \tilde{\phi} \circ \tilde{\theta}_1$ and $\phi_2 = \tilde{\phi} \circ \tilde{\theta}_2$.

To verify well-definedness on $\pi_1([K])$, it will suffice to check that either adding or deleting a repeated vertex in a course does not affect its image in H, and that two contiguous courses map to the same element of H. The first statement is trivial. For the second, observe that any two contiguous courses are joined by a sequence of courses each of which is contiguous to its successor and differs from its successor in only a single entry. (Instead of moving all the points in the first course to all the points in the second course at once, move the points one at a time.) So we reduce to the problem of checking that two contiguous courses that differ in only a single entry map to the same element of H. If this single entry, its immediate predecessor, and its immediate successor all belong to K_1 or all belong to K_2, then the fact that contiguous courses represent the same element of the fundamental group entails that they map to the same element of H. Otherwise, the key observation is that if $v \in V_{K_1} \setminus V_{K_0}$ and $w \in V_{K_2} \setminus V_{K_0}$ then $\{v, w\}$ cannot be a 1-simplex. Because if it were, then it would have to lie in either K_1 or K_2, forcing v and w to both belong to either V_{K_1} or V_{K_2}. Thus the only way for a single entry, its immediate predecessor, and its immediate successor not to all lie in K_1 or all lie in K_2 is if the single entry belongs to K_0 and either its predecessor or its successor belongs to K_1 while the other one belongs to K_2. Contiguity then implies that the new version of the single entry must also belong to K_0. It is easy to see that in this case

the image in H has not changed. □

Corollary 14.24. *Let K be a connected simplicial complex and let K_1 and K_2 be connected subcomplexes such that $K = K_1 \cup K_2$ and $[K_1 \cap K_2]$ is simply connected. Then $\pi_1([K]) \cong \pi_1([K_1]) * \pi_1([K_2])$.*

For example, a figure eight can be triangulated into two triangles with one vertex in common. Letting K_1 and K_2 be the two triangles, we get that K_0 consists only of the single point in common and therefore $[K_0]$ is certainly simply connected. Since the fundamental group of a circle is $\mathbb{Z}$, it follows that the fundamental group of a figure eight is $\mathbb{Z} * \mathbb{Z}$.

More generally, if n circles are joined at a point then the same argument can be repeated to show that the fundamental group is an n-fold free product $\mathbb{Z} * \cdots * \mathbb{Z}$.

14.4 Surfaces

An *n-manifold* is a locally compact topological space which is covered by open classes U_i ($1 \leq i \leq k$ or $i \in \mathbb{N}$), each of which is homeomorphic to an open subclass of $\mathbb{R}^n$. Thus $\mathbb{R}^n$ is itself an example of an n-manifold. The unit sphere (or *n-sphere*) $\mathbb{S}^n$ in $\mathbb{R}^{n+1}$ is also an n-manifold, for any n. The n-fold product $\mathbb{S}^1 \times \cdots \times \mathbb{S}^1$ is an n-dimensional manifold denoted $\mathbb{T}^n$ and called an *n-torus*.

Connected 2-dimensional manifolds are also called *surfaces*. Thus $\mathbb{S}^2$ and $\mathbb{T}^2$ are surfaces. Both can be constructed by the following general method for producing surfaces.

Let P be a polygon in $\mathbb{R}^2$ with an even number of edges, together with its interior. Informally, we can create a surface by pairing off the edges and "identifying" the two edges of each pair. This identification must also state which end of the first edge is identified with which end of the second edge.

Now, if the polygon P is triangulated by a simplicial complex K, then we can create an abstract simplicial complex $\mathcal{K}$ by letting its abstract vertex set be $V_K/\!\sim$, where $\sim$ identifies vertices of P in the manner prescribed by the edge pairing. We then simply replace every element of every simplex in K with the class of that element in $V_K/\!\sim$. In order to produce a genuine abstract simplicial complex, we need only require that the triangulation we start with has to be sufficiently fine that no two verties or two edges from a single 2-simplex get identified.

If we assign a letter to each edge pair, we can express the identification

by starting at some vertex, going counterclockwise around P, and writing either α or α^{-1} for each edge of P. Here α is an arbitrary letter and we choose α or α^{-1} depending on how the edges of P are to be identified. If two edges are identified and the first vertex of one edge is identified with the first vertex of the other as we go around P counterclockwise (the two edges are *coherent*), then we write α both times; if the first vertex of one edge is identified with the second vertex of the other as we go around P counterclockwise (the two edges are *incoherent*), then we write α for one and α^{-1} for the other. The result is a *boundary word*. ("A", not "the", because we could have interchanged α and α^{-1}, or used different letters, or started somewhere else on the boundary of P.)

For example, $\mathbb{T}^2$ can be described as a square with top and bottom edges identified, left to left and right to right, and with right and left edges identified, top to top and bottom to bottom. The associated boundary word could be $aba^{-1}b^{-1}$. One way to describe $\mathbb{S}^2$ is also as a square, with edges identified according to the boundary word $aa^{-1}bb^{-1}$ (so "zipping up" from the vertex between the "a" edges to the vertex between the "b" edges). A third basic surface modelled on a square is the *real projective plane* $\mathbb{P}^2$; here the boundary word is $abab$.

A surface formed in this manner is a *polygonal surface*. The significance of this construction is shown by the next result.

Proposition 14.25. *Every compact triangulable surface is homeomorphic to a polygonal surface.*

In fact, every compact surface is triangulable, so that qualification is unnecessary, but the proof of this is complicated.

Proposition 14.25 is actually quite easy to visualize: given a compact two-dimensional polytope, start with one triangular face chosen at random, lay it on the table, then choose a triangular face which shares an edge with the first one and attach it there, and so on. At each step we must find a thus far unused triangle which shares an edge with some already placed triangle, and attach it there. Every time we add a triangle we get a new polygon (which may be nonconvex, but that is okay). Since the original polytope was compact, the associated simplicial complex is finite, so the construction terminates after finitely many steps. At the end, the edges which remain exposed are supposed to be identified in pairs, so we have shown how to realize the original triangulated surface as a polygonal surface.

Now let us identify the fundamental groups of our three basic surfaces, $\mathbb{S}^2$, $\mathbb{T}^2$, and $\mathbb{P}^2$. The easiest is $\mathbb{S}^2$. We can treat it as two closed disks

attached along their boundary $\mathbb{S}^1$'s; this leads to a Seifert-van Kampen setup with $[K_1] \cong [K_2] \cong \mathbb{D}^2$ and $[K_0] \cong \mathbb{S}^1$. Since the fundamental group of $\mathbb{D}^2$ is trivial (straight line homotopy) it follows immediately from the Seifert-van Kampen theorem (Theorem 14.23) that $\pi_1(\mathbb{S}^2)$ is trivial. (One could also model $\mathbb{S}^2$ by a tetrahedron and easily see how to make any course equivalent to the trivial one.)

We can identify the fundamental group of $\mathbb{T}^2$ using the following result.

Proposition 14.26. *Let X and Y be connected triangulable spaces. Then $\pi_1(X \times Y) = \pi_1(X) \times \pi_1(Y)$.*

Proof. Fix base points e_1 and e_2 in X and Y, respectively, and designate (e_1, e_2) as the base point of $X \times Y$. Then every loop f in $X \times Y$ has the form $t \mapsto (f_1(t), f_2(t))$ $(t \in [0,1])$, where f_1 and f_2 are loops in X and Y, respectively; conversely, if f_1 and f_2 are loops in X and Y then $t \mapsto (f_1(t), f_2(t))$ is a loop in $X \times Y$. One straightforwardly checks that f is homotopic to g in $X \times Y$ if and only if f_1 is homotopic to g_1 in X and f_2 is homotopic to g_2 in Y. Thus we have established an isomorphism between $\pi_1(X \times Y)$ and $\pi_1(X) \times \pi_1(Y)$. □

So $\pi_1(\mathbb{T}^2) \cong \mathbb{Z}^2$. Another way to see this is to identify $\mathbb{R}^2$ as the universal cover of the unit square model of $\mathbb{T}^2$ via the covering map which takes the point $\vec{x} \in \mathbb{R}^2$ to the unique point $\lfloor \vec{x} \rfloor$ in $[0,1)^2$ satisfying $\vec{x} - \lfloor \vec{x} \rfloor \in \mathbb{Z}^2$. The deck transformations of $\mathbb{R}^2$ are then the maps $(s,t) \mapsto (s+m, t+n)$ for $m, n \in \mathbb{Z}$. The group of all deck transformations is therefore isomorphic to $\mathbb{Z}^2$, and so $\pi_1(\mathbb{T}^2)$ is too by Theorem 14.20.

The easiest way to identify $\pi_1(\mathbb{P}^2)$ is by recognizing that its universal cover is $\mathbb{S}^2$. Taking the square model of $\mathbb{P}^2$ described above, we can model $\mathbb{S}^2$ as an octahedron, project the bottom half vertically onto a square, and rotate the top half 180° and then project it vertically onto the same square. Since opposite sides of the square are identified in $\mathbb{P}^2$, this describes a covering map. The only nontrivial deck transformation is given, in the octahedron model, by taking each vertex to the polar opposite vertex. Thus the fundamental group of $\mathbb{P}^2$ has two elements, that is, $\pi_1(\mathbb{P}^2) \cong \mathbb{Z}/2\mathbb{Z}$.

In fact, there is a uniform method for evaluating the fundamental group of any polygonal surface other than $\mathbb{S}^2$. First, instead of the polygon construction, we can start with the unit disk $\mathbb{D}^2$, then split up its boundary $\mathbb{S}^1$ into an even number of disjoint arcs separated by points (which I will still call "vertices"). Finally, pair off the arcs, coherently or incoherently. This pairing can still be described by a boundary word, but the disk models are

slightly more general than the polygon models because they include the case of only two boundary arcs, whereas polygon models must have at least four sides. In this picture we can recover $\mathbb{S}^2$ with the boundary word aa^{-1} and $\mathbb{P}^2$ with the boundary word aa. I will still use the term "polygonal surface" for this construction, because if there are more than two boundary arcs then they can be straightened out, converting $\mathbb{D}^2$ into a polygon. $\mathbb{S}^2$ and $\mathbb{P}^2$ are the only two spaces which can be realized using only two boundary arcs, and we already know that they are both polygonal surfaces, so in fact we are still talking about the same polygonal surfaces discussed earlier.

Suppose the pairing has been done in such a way that all the vertices are identified to a single point. This includes the case of the projective plane (boundary word aa) and the 2-torus (boundary word $aba^{-1}b^{-1}$), but not the 2-sphere (boundary word aa^{-1}). Then if Y is the space formed from just the boundary $\mathbb{S}^1$, identifying the boundary arcs in the prescribed manner but discarding the interior of the disk, what we will get is n circles joined at a point, where $2n$ is the total number of boundary arcs. And we know from the comment at the end of the last section that its fundamental group is the n-fold free product $\mathbb{Z} * \cdots * \mathbb{Z}$.

Let Z be the "punctured surface" obtained by removing not all of the interior of $\mathbb{D}^2$, but only a smaller open disk, with the same center as $\mathbb{D}^2$ but smaller radius, and then identifying the boundary arcs. We have a straight line homotopy between the identity map from Z to itself and the map from Z to $Y \subset Z$ which transports every element of Z radially onto an element of the bounding $\mathbb{S}^1$. Any loop in Z can therefore be deformed to a loop in Y, and two loops are homotopic to each other before the deformation if and only if they are homotopic to each other after the deformation (the key point being that any loop in Z is homotopic to its image when projected into Y). Thus $\pi_1(Z) \cong \pi_1(Y) \cong \mathbb{Z} * \cdots * \mathbb{Z}$.

We now have a Seifert-van Kampen setup involving Z, the closure of the smaller removed disk, and the circle in which these two pieces intersect. The smaller removed disk has trivial fundamental group, and we know that the fundamental group of the circle is $\mathbb{Z}$. So the Seifert-van Kampen theorem (Theorem 14.23) tells us that the fundamental group of the polygonal surface obtained from $\mathbb{D}^2$ by identifying the boundary arcs is an amalgamated free product of $\pi_1(Y) \cong \mathbb{Z} * \cdots * \mathbb{Z}$ with a trivial group. But the unamalgamated free product is just $\mathbb{Z} * \cdots * \mathbb{Z}$, and the amalgamation simply has the effect of factoring out by the normal subgroup generated by the boundary word. In more detail, we can treat $\mathbb{Z} * \cdots * \mathbb{Z}$ as the group of all words in n

letters and their inverses (α and α^{-1}) modulo inserting or removing strings of the form $\alpha\alpha^{-1}$ or $\alpha^{-1}\alpha$. The boundary word w belongs to this model of $\mathbb{Z} * \cdots * \mathbb{Z}$, and it generates a normal subgroup which consists of all finite products of words of the form $hw^k h^{-1}$ where $h \in \mathbb{Z} * \cdots * \mathbb{Z}$ and w^k is a string of k copies of w when $k \geq 0$, or $-k$ copies of the inverse word w^{-1} when $k < 0$. By the "inverse word" I mean w written in reverse order and with every α replaced by α^{-1} and every α^{-1} replaced by α. If we let $\langle w \rangle$ denote this normal subgroup, we get the following result.

Proposition 14.27. *Let X be a polygonal surface formed by identifying boundary arcs in $\mathbb{S}^1 \subset \mathbb{D}^2$ according to the boundary word w. Suppose that all the vertices are identified to a single point. Then*

$$\pi_1(X) \cong (\mathbb{Z} * \cdots * \mathbb{Z})/\langle w \rangle,$$

where the free product has n factors if there were $2n$ boundary arcs.

For example, in the projective plane model described above with boundary word aa, we get the 1-fold free product $\mathbb{Z}$ modulo the (automatically normal) subgroup generated by the element aa, taking a to be the generator of $\mathbb{Z}$. This subgroup is just $2\mathbb{Z}$, so $\pi_1(\mathbb{P}^2) \cong \mathbb{Z}/2\mathbb{Z}$. In the torus model with boundary word $aba^{-1}b^{-1}$ we get $(\mathbb{Z} * \mathbb{Z})/\langle aba^{-1}b^{-1} \rangle$, and here factoring out the normal subgroup has the effect of making the generators of the two copies of $\mathbb{Z}$ in $\mathbb{Z} * \mathbb{Z}$ commute ($aba^{-1}b^{-1} = e$ if and only if $ab = ba$). Thus the quotient is just $\mathbb{Z}^2$.

These two examples can be generalized as follows. First, instead of $aba^{-1}b^{-1}$ use the word $a_1 b_1 a_1^{-1} b_1^{-1} \cdots a_g b_g a_g^{-1} b_g^{-1}$ for any $g \in \mathbb{N}$. The polygonal surface this describes is the *genus g handlebody* $\mathcal{H}_g$, which is like a torus but with g holes. Second, instead of aa use the word $a_1 a_1 \cdots a_g a_g$. This describes the *genus g cross-cap surface* $\mathcal{C}_g$.

The fundamental groups of these surfaces are a little complicated, but they have simple abelianizations. The *abelianization* of a countable group G is the quotient G/N where N is the normal subgroup generated by elements of the form $ghg^{-1}h^{-1}$. Factoring out N has the effect of making everything in G commute.

Theorem 14.28. *The fundamental group of $\mathcal{H}_g$ is*

$$(\mathbb{Z} * \cdots * \mathbb{Z})/\langle a_1 b_1 a_1^{-1} b_1^{-1} \cdots a_g b_g a_g^{-1} b_g^{-1} \rangle$$

(with a $2g$-fold free product) and the fundamental group of $\mathcal{C}_g$ is

$$(\mathbb{Z} * \cdots * \mathbb{Z})/\langle a_1 a_1 \cdots a_g a_g \rangle$$

(with a g-fold free product). The abelianization of $\pi_1(\mathcal{H}_g)$ is $\mathbb{Z}^{2g}$ and the abelianization of $\pi_1(\mathcal{C}_g)$ is $\mathbb{Z}^{g-1} \times (\mathbb{Z}/2\mathbb{Z})$.

I will sketch how to identify the abelianizations in this theorem. First, abelianizing $\mathbb{Z} * \cdots * \mathbb{Z}$ means factoring out the normal subgroup generated by all elements of the form $ghg^{-1}h^{-1}$, and this subgroup includes the word $a_1b_1a_1^{-1}b_1^{-1}\cdots a_gb_ga_g^{-1}b_g^{-1}$, so the abelanization of $\pi_1(\mathcal{H}_g)$ is just the abelianization of a $2g$-fold free power of $\mathbb{Z}$, which is a $2g$-fold direct sum of copies of $\mathbb{Z}$. For $\pi_1(\mathcal{C}_g)$ we can first abelianize the free product to get a g-fold direct sum of copies of $\mathbb{Z}$, then factor out the subgroup generated by $a_1a_1\cdots a_ga_g$. That means we get the group $\mathbb{Z}^g$ modulo the subgroup $\{(2n,\ldots,2n) : n \in \mathbb{Z}\}$. But one easily checks that this is isomorphic to $(\mathbb{Z}\oplus\cdots\oplus\mathbb{Z})/(2\mathbb{Z}\oplus\{0\}\oplus\cdots\oplus\{0\}) \cong \mathbb{Z}^{g-1}\oplus(\mathbb{Z}/2\mathbb{Z})$ by considering the map from $\mathbb{Z}\oplus\cdots\oplus\mathbb{Z}$ to itself which takes $(n_1,\ldots,n_g)$ to $(n_1, n_2-n_1,\ldots,n_g-n_1)$. This map is an isomorphism which takes the subgroup generated by $(2,2,\ldots,2)$ to the subgroup generated by $(2,0,\ldots,0)$. Thus factoring out these subgroups yields isomorphic groups.

Corollary 14.29. *$\mathbb{S}^2$, $\mathcal{H}_g$, and $\mathcal{C}_g$ (both for $g \in \mathbb{N}$) are all topologically distinct.*

Because their fundamental groups have distinct abelianizations.

The final result of this section states that this list exhausts all compact triangulable surfaces, up to homeomorphism.

Lemma 14.30. *Every polygonal surface besides $\mathbb{S}^2$ and $\mathbb{P}^2$ is homeomorphic to one coming from a polygon model in which all vertices are identified to a single point.*

Proof. Every polygonal surface comes from a polygon model with at least four vertices. First, every time aa^{-1} or $a^{-1}a$ appears in the boundary word, "zip up" and delete that pair of letters. It is clear that the resulting surface is homeomorphic to the original one.

At this point, if there are still vertices that aren't identified, then there must be some adjacent pair of vertices v and v' which aren't identified. Let v'' be the next vertex after v', create an edge between v and v'', and sever and reattach the triangle with vertices v, v', v'' to the edge that was paired with the v to v' edge. The number of vertices identified with v has now gone down by one. Then remove any aa^{-1} or $a^{-1}a$ pairs and continue. Eventually we will eliminate all vertices equivalent to v; then we can move on to the next equivalence class of vertices. As long as the surface is not

$\mathbb{S}^2$ or $\mathbb{P}^2$, we will never have fewer than four vertices, and the only way the process can stop is when all vertices are identified. □

In this proof we introduced the technique of slicing off part of the polygon model and reattaching it elsewhere on the boundary. The same technique can be used to modify the boundary word in several ways. In the following result we use the term "interlocked pair" for edges labelled a, b, a^{-1}, b^{-1} which appear in that order as we move counterclockwise around the boundary.

Lemma 14.31. *Given any polygon model, we can find a new polygon model that produces the same surface (up to homeomorphism) whose boundary word has been modified in any of the following ways:*

(1) make a coherent pair adjacent (convert $AaBaC$ to $AB^{-1}bbC$)
(2) separate an adjacent coherent pair (convert $AaaBC$ to $AbB^{-1}bC$)
(3) make an interlocked pair adjacent (convert $AaBbCa^{-1}Db^{-1}E$ to $ADcdc^{-1}d^{-1}CBE$)
(4) convert an adjacent interlocked pair that follows an adjacent coherent pair into three adjacent coherent pairs (convert $Aaabcb^{-1}c^{-1}B$ to $Addeeff B$).

Proof. Here the capital letters A, B, C, D, E represent substrings of the boundary word. We accomplish (1) by introducing an edge between the aB vertex (the vertex between a and B) and the aC vertex, then chopping off and reattaching along a. We accomplish (2) by introducing an edge between the aa and BC vertices, chopping off and reattaching along a. We accomplish (3) by introducing an edge d between the bC and Db^{-1} vertices, chopping off and reattaching along a, and then introducing an edge between the Dd^{-1} and db^{-1} vertices, chopping off and reattaching along b. And we accomplish (4) by applying (2) once and then (1) twice, going from $Aaabcb^{-1}c^{-1}B$ to $Adbc^{-1}b^{-1}dc^{-1}B$ to $Adbd^{-1}bffB$ to $Addeeff B$. □

Theorem 14.32. *(Classification of compact surfaces) Every compact triangulable surface is homeomorphic to exactly one of $\mathbb{S}^2$, $\mathcal{H}_g$ for some g, or $\mathcal{C}_g$ for some g.*

Proof. The distinctness of $\mathbb{S}^2$, $\mathcal{H}_g$, and $\mathcal{C}_g$ was shown in Corollary 14.29. For the rest, use Proposition 14.25 to convert any triangulable surface into a polygonal surface, then if the result is not $\mathbb{S}^2$ use Lemma 14.30 to convert it into a polygonal surface in which all vertices are identified, and then use

Lemma 14.31 (1) to make all coherent pairs adjacent. If there is now at least one incoherent pair, there must be an interlocked pair because $AaBa^{-1}$ must contain an edge in A that is paired with an edge in B: otherwise the initial and final vertices of the a edge would not be identified. Since all the coherent pairs have already been made adjacent, that means we have an interlocked pair, and we can now use (3) repeatedly to make all interlocked pairs adjacent. At this point, if all pairs are incoherent then we have a handlebody $\mathcal{H}_g$ for some g, and if not then we can use (4) to make the boundary word consist entirely of adjacent coherent pairs, in which case we have a cross-cap surface $\mathcal{C}_g$ for some g. □

14.5 Simplicial homology

Let $S = \{v_0, \ldots, v_p\}$ be a simplex. Any two orderings of the elements of S are related by a permutation of $\{0, \ldots, p\}$ which has a sign, ± 1 (Section 12.2). So the orderings of S are partitioned into two sets, those which differ from some fixed ordering by a permutation with sign $+1$ and those which differ from it by a permutation with sign -1. An *orientation* of S is a choice of one of these two sets. A simplex for which an orientation has been chosen is an *oriented simplex* and may be denoted $\hat{S}$ here. In the case of 0-simplices there is only one orientation, but in every other case there are exactly two and I will denote the other one $\check{S}$.

I will also write $[v_0, \ldots, v_p]$ for the simplex $\{v_0, \ldots, v_p\}$ with orientation represented by the ordering $v_0, \ldots, v_p$. We can model the oriented p-simplices in a simplicial complex K as follows. First consider the set of $(p+1)$-tuples $\vec{v} \in V_K^{p+1}$ for which $\{v_0, \ldots, v_p\}$ is a p-simplex in K; then make $\vec{v}$ and $\vec{w}$ equivalent if $[v_0, \ldots, v_p]$ and $[w_0, \ldots, w_p]$ represent the same simplex with the same orientation. The quotient is the desired model. We denote it $K^{(p)}$.

Definition 14.33. Let K be a simplicial complex and let $p \geq 1$. A *p-chain* in K is a function c from the set of oriented p-simplices to $\mathbb{Z}$ which is zero on all but finitely many oriented p-simplices, and such that $c(\hat{S}) = -c(\check{S})$ for every oriented p-simplex $\hat{S}$. A *0-chain* is a function from the set of 0-simplices to $\mathbb{Z}$ which is zero on all but finitely many 0-simplices.

Any two p-chains can be added pointwise, yielding an abelian group denoted $C_p(K) \subseteq \mathrm{fn}(K^{(p)}, \mathbb{Z})$.

Formally, $C_p(K)$ is a subset of the pseudoclass of functions from the set

of oriented p-simplices to $\mathbb{Z}$.

For any oriented p-simplex $\hat{S}$, $p \geq 1$, there is an *elementary* p-chain which takes the value 1 on $\hat{S}$, the value -1 on $\check{S}$, and 0 on all other oriented p-simplices. I will write $c_{\hat{S}}$ for this p-chain. If S is a 0-simplex then c_S is the 0-chain which takes the value 1 on S and 0 on all other 0-simplices. $C_p(K)$ is isomorphic to $\mathbb{Z}^n$ for some $n \geq 0$ or to $\mathbb{Z}^{<\omega}$: fix an orientation of each p-simplex and map the corresponding elementary p-chains to the standard generators of $\mathbb{Z}^n$ or $\mathbb{Z}^{<\omega}$.

Definition 14.34. Let K be a simplicial complex. For $p > 0$ define the *boundary operator* $\partial_p : C_p(K) \to C_{p-1}(K)$ by putting

$$\partial_p(c_{[v_0,\ldots,v_p]}) = \sum_{i=0}^{p} (-1)^i c_{[v_0,\ldots,\dot{v}_i,\ldots,v_p]}$$

for each oriented p-simplex $[v_0, \ldots, v_p]$ and extending linearly, where "$\dot{v}_i$" means that the v_i entry is omitted. Thus ∂_p is a homomorphism of abelian groups. For $p = 0$ we let the boundary operator $\partial_0 : C_0(K) \to \{0\}$ be the trivial homomorphism into the trivial group.

Proposition 14.35. *Let K be a simplicial complex. Then $\partial_{p-1} \circ \partial_p = 0$ for all p.*

Proof. It suffices to check this on elementary p-chains. We have

$$\begin{aligned}\partial_{p-1}\partial_p(c_{[v_0,\ldots,v_p]}) &= \sum_{i=0}^{p} (-1)^i \partial_{p-1}(c_{[v_0,\ldots,\dot{v}_i,\ldots,v_p]}) \\ &= \sum_{j<i} (-1)^i (-1)^j c_{[v_0,\ldots,\dot{v}_j,\ldots,\dot{v}_i,\ldots,v_p]} \\ &\quad + \sum_{j>i} (-1)^i (-1)^{j-1} c_{[v_0,\ldots,\dot{v}_i,\ldots,\dot{v}_j,\ldots,v_p]},\end{aligned}$$

and the terms of the last two sums cancel in pairs. □

Definition 14.36. The kernel of ∂_p is the *group of p-cycles* $Z_p(K)$, and its image is the *group of $(p-1)$-boundaries* $B_{p-1}(K)$. The *pth homology group of K* is the quotient $H_p(K) = Z_p(K)/B_p(K)$.

(We have $B_p(K) \subseteq Z_p(K)$ by Proposition 14.35.)

A valuable tool in what follows is the notion of a cone. If K is a simplicial complex contained in $\mathbb{R}^n$ or $\mathbb{R}^{<\omega}$, it is homeomorphic to a simplicial complex in $\mathbb{R} \oplus \mathbb{R}^n \cong \mathbb{R}^{n+1}$ or $\mathbb{R} \oplus \mathbb{R}^{<\omega} \cong \mathbb{R}^{<\omega}$ every vertex of which is zero on the first coordinate. Assume K has this form and let $v^* = e_1$ be the standard

basis vector of the initial $\mathbb{R}$. The *cone over* K, $\mathcal{C}K$, is then the simplicial complex which contains all the simplices of K plus the simplex $\{v^*\}$ and, for every p-simplex S of K, the $(p+1)$-simplex $S \cup \{v^*\}$. The homology groups of cones are very simple.

Proposition 14.37. *(Homology of a cone) Let K be a simplicial complex. Then $H_0(\mathcal{C}K) \cong \mathbb{Z}$ and $H_p(\mathcal{C}K) = \{0\}$ for all $p \geq 1$.*

Proof. For $p = 0$, observe that every 0-chain is a 0-cycle. If $v \in V_K$ then $\partial c_{[v^*,v]} = c_{\{v\}} - c_{\{v^*\}}$, so every 0-chain in $\mathcal{C}K$ differs from an integer multiple of $c_{\{v^*\}}$ by a 0-boundary. However, no integer multiple of $c_{\{v^*\}}$ is itself a boundary. So $Z_0(\mathcal{C}K)/B_0(\mathcal{C}K) \cong \mathbb{Z}$.

For $p \geq 1$, define a homomorphism $d_p : C_p(\mathcal{C}K) \to C_{p+1}(\mathcal{C}K)$ by putting $d_p(c_{[v_0,\ldots,v_p]}) = c_{[v^*,v_0,\ldots,v_p]}$ if v^* is not among $v_0, \ldots, v_p$, and letting it be zero on all other elementary p-chains. A short calculation shows that $\partial_{p+1} d_p(c_{\hat{S}}) + d_{p-1}\partial_p(c_{\hat{S}}) = c_{\hat{S}}$ for every oriented simplex $\hat{S}$ of $\mathcal{C}K$, and hence $\partial_{p+1} d_p(c) + d_{p-1}\partial_p(c) = c$ for every p-chain c. Thus if c is a p-cycle then $\partial_{p+1} d_p(c) = c$, showing that c is also a p-boundary. We conclude that $H_p(\mathcal{C}K)$ is trivial. □

Now we must show that the homology groups are topologically invariant, i.e., if $[K]$ and $[L]$ are homeomorphic then K and L have the same homology groups. The proof will proceed through a series of lemmas. Some of the proofs are simple computations that I will omit.

First, if ϕ is a simplicial transformation from K to L then we can define an associated homomorphism of abelian groups $\phi_p : C_p(K) \to C_p(L)$. If ϕ takes a p-simplex S of K to a q-simplex of L, where $q < p$, then we define $\phi_p(c_{\hat{S}}) = 0$. If $q = p$ and $\hat{S} = [v_0, \ldots, v_p]$ then we define $\phi_p(c_{\hat{S}}) = c_{\hat{S}'}$ where $\hat{S}' = [\phi(v_0), \ldots, \phi(v_p)]$. Then $\phi_p(c_{\check{S}}) = -\phi_p(c_{\hat{S}})$, so ϕ_p is well-defined on elementary p-chains, and it extends linearly to all p-chains.

Lemma 14.38. *Let K and L be simplicial complexes and let ϕ be a simplicial transformation from K to L. Then $\partial_p \circ \phi_p = \phi_{p-1} \circ \partial_p$ for all p.*

It suffices to check this on elementary p-chains. We call the sequence (ϕ_p) a *chain map*, meaning that it commutes with the boundary operator in the sense of Lemma 14.38.

As a consequence, if c is a p-cycle, so that $\partial_p(c) = 0$, then

$$\partial_p(\phi_p(c)) = \phi_{p-1}(\partial_p(c)) = 0,$$

showing that $\phi_p(c)$ is a p-cycle, and if c is a p-boundary, so that $c = \partial_{p+1}(c')$ for some $(p+1)$-chain c', then

$$\partial_{p+1}(\phi_{p+1}(c')) = \phi_p(\partial_{p+1}(c')) = \phi_p(c),$$

showing that $\phi_p(c)$ is a p-boundary. Thus ϕ_p takes p-cycles to p-cycles and p-boundaries to p-boundaries, and it therefore induces a homomorphism $\overline{\phi}_p : H_p(K) \to H_p(L)$.

Lemma 14.39. *Let K, L, and M be simplicial complexes and let ϕ be a simplicial transformation from K to L and ψ a simplicial transformation from L to M. Then $\overline{(\psi \circ \phi)}_p = \overline{\psi}_p \circ \overline{\phi}_p$ for all p.*

Our next step is to check that iterated stellar subdivision does not alter the homology groups. Recall that in the basic step, a single stellar subdivision, we introduce a new vertex v_S that is to be situated at the barycenter of some face $[S]$, and then replace every simplex containing S with the simplices obtained by replacing some vertex of S with v_S, also adding in all of their sub-simplices. If we are given an orientation of any simplex that contains S, this induces an orientation on all of the simplices that replace it in an obvious way (replace the chosen vertex with v_S and leave everything else alone).

If K is a simplicial complex and K' is the stellar subdivision of K for some simplex S, then we can define a function $\chi_p : C_p(K) \to C_p(K')$ by leaving alone all p-chains that are common to K and K', and if S' is a p-simplex that contains S, taking $c_{\hat{S}'}$ to the sum of $c_{\hat{S}''}$ over all the simplices S'' into which S' is subdivided, with the induced orientation. This is a chain map:

Lemma 14.40. *Let K be a simplicial complex and let K' be a stellar subdivision of K. Then $\partial_p \circ \chi_p = \chi_{p-1} \circ \partial_p$ for all p.*

Again, it suffices to check this on elementary p-chains. Just as for the maps ϕ_p, this result entails that χ_p induces a homomorphism $\overline{\chi}_p : H_p(K) \to H_p(K')$. By repeating this construction and composing, we get a homomorphism from $H_p(K)$ to $H_p(K')$ for any finitely iterated stellar subdivision K' of K; I will also call this map $\overline{\chi}_p$. Since we require general iterated stellar subdivisions to subdivide any original face only finitely many times, it is not hard to see that the construction still works for infinitely iterated stellar subdivisions.

We can also go from a single stellar subdivision K' back to K, in the following way. Let θ be a simplicial transformation from K' to K which

fixes all the vertices of K and takes the one new vertex v_S to one of the vertices — it doesn't matter which one — of S. Then θ is a simplicial approximation of the identity map from $[K']$ to $[K]$ (and every simplicial approximation of the identity map from $[K']$ to $[K]$ has this form). As we saw above, θ induces a homomorphism $\overline{\theta}_p : H_p(K') \to H_p(K)$ for all p.

Lemma 14.41. *Let K be a simplicial complex and let K' be a stellar subdivision of K. Then $\overline{\chi}_p$ and $\overline{\theta}_p$ are inverse to each other, for all p.*

Proof. It is easy to check that $\theta_p(\chi_p(c)) = c$ for every $c \,\varepsilon\, C_p(K)$, so $\overline{\theta}_p \circ \overline{\chi}_p$ is the identity map from $H_p(K)$ to itself. Now let $c \,\varepsilon\, Z_p(K')$ and consider $c - \chi_p(\theta_p(c))$. The union L of all the simplices in K' which contain the new vertex v_S, together with all of their sub-simplices, is a sub-complex of K', and its carrier $[L]$ is a cone. Since

$$\partial_p(c - \chi_p(\theta_p(c))) = \partial_p(c) - \chi_{p-1}(\theta_{p-1}(\partial_p(c))) = 0,$$

Proposition 14.37 yields that if $p \geq 1$ then $c - \chi_p(\theta_p(c))$ is the boundary of a $(p+1)$-chain of $L \subseteq K'$. Thus c and $\chi_p(\theta_p(c))$ represent the same element of $H_p(K')$, and this shows that $\overline{\chi}_p \circ \overline{\theta}_p$ is the identity map from $H_p(K')$ to itself. When $p = 0$ we have $\chi_0(\theta_0(c)) = c + (\partial_1 c_{[v_S, \theta(v_S)]}) \cdot c(\{v_S\})$, so again c and $\chi_0(\theta_0(c))$ represent the same element of $H_0(K')$. □

It follows that $H_p(K')$ and $H_p(K)$ are isomorphic groups; as above, we can iterate to get an isomorphism from $H_p(K')$ to $H_p(K)$ for any finitely iterated stellar subdivision K' of K, and I will also call this map $\overline{\theta}_p$. Inductively we have that $\overline{\chi}_p$ and $\overline{\theta}_p$ are inverse to each other for any finitely iterated stellar subdivision K' of K. Again, the local finiteness of iterated stellar subdivisions allows us to generalize these statements to the infinitely iterated setting.

We need three more lemmas. First, let us generalize the notion of contiguity used in Section 14.2.

Definition 14.42. Let K and L be simplicial complexes and let ϕ and ψ be simplicial transformations from K to L. We say that ϕ and ψ are *contiguous* if for every simplex S of K, the set $\phi(S) \cup \psi(S)$ is a simplex of L.

Lemma 14.43. *Let K and L be simplicial complexes and let ϕ and ψ be contiguous simplicial transformations from K to L. Then we have $\overline{\phi}_p = \overline{\psi}_p$ for all p.*

Proof. We show that there is a *chain homotopy* between ϕ and ψ, a sequence of homomorphisms $d_p : C_p(K) \to C_{p+1}(L)$ for all p such that

$$d_{p-1}\partial_p + \partial_{p+1}d_p = \phi_p - \psi_p.$$

Then $\phi_p - \psi_p$ takes every cycle to a boundary, so $\overline{\phi}_p = \overline{\psi}_p$. The existence of d_p is shown by first putting $d_0(c_{\{v\}}) = 0$ if $\phi(v) = \psi(v)$, and $d_0(c_{\{v\}}) = c_{[\phi(v),\psi(v)]}$ if $\phi(v) \neq \psi(v)$. Then, for an oriented p-simplex $\hat{S}$ with $p \geq 1$, let $c = \phi_p(c_{\hat{S}}) - \psi_p(c_{\hat{S}}) - d_{p-1}\partial_p(c_{\hat{S}})$ and use the chain homotopy condition for $p-1$ plus Proposition 14.35 and Lemma 14.38 to calculate

$$\begin{aligned}\partial_p(c) &= \partial_p(\phi_p(c_{\hat{S}}) - \psi_p(c_{\hat{S}}) - d_{p-1}\partial_p(c_{\hat{S}})) \\ &= \partial_p(\phi_p(c_{\hat{S}})) - \partial_p(\psi_p(c_{\hat{S}})) \\ &\qquad - [\phi_{p-1}(\partial_p(c_{\hat{S}})) - \psi_{p-1}(\partial_p(c_{\hat{S}})) - d_{p-2}(\partial_{p-1}(\partial_p(c_{\hat{S}})))] = 0.\end{aligned}$$

For any p-simplex S of K the sets $\phi(S)$ and $\psi(S)$ belong to a simplex of L, and every simplex is a cone, so by Proposition 14.37 this implies that c is a p-boundary, say $c = \partial_{p+1}(c')$. We can then define $d_p(c_{\hat{S}}) = c'$ and extend linearly. That completes the construction of d_p. □

Now suppose $f : [K] \to [L]$ is continuous. By simplicial approximation (Corollary 14.6) we can find an iterated stellar subdivision K' of K and a simplicial approximation ϕ of f that is affine on $[K']$. Then, as we saw after Lemma 14.38, this simplicial approximation induces a homomorphism $\overline{\phi}_p$ from $H_p(K')$ to $H_p(L)$ for all p, and thus a homomorphism $\overline{\phi}_p \circ \overline{\chi}_p = \overline{\phi}_p \circ (\overline{\theta}_p)^{-1}$ from $H_p(K)$ to $H_p(L)$.

Lemma 14.44. *Let K and L be simplicial complexes and let $f : [K] \to [L]$ be continuous. Then $\overline{\phi}_p \circ \overline{\chi}_p$ does not depend on the choice of subdivision K' or the choice of ϕ.*

Proof. Given the subdivision of K, all choices of ϕ yield the same homomorphism from $H_p(K)$ to $H_p(L)$ by Lemma 14.43 because they are all simplicial approximations of f and are therefore contiguous.

Let K' and K'' be iterated stellar subdivisions of K and suppose $\phi' : [K'] \to [L]$ and $\phi'' : [K''] \to [L]$ are simplicial approximations of f. Let $\theta' : K' \to K$ and $\theta'' : K'' \to K$ be simplicial approximations of the identity map, as in Lemma 14.41. We must show that $\overline{\phi}'_p \circ (\overline{\theta}'_p)^{-1} = \overline{\phi}''_p \circ (\overline{\theta}''_p)^{-1}$.

As in the proof of simplicial approximation (Corollary 14.6) we can find an iterated stellar subdivision K''' of K and simplicial approximations $\psi' : [K'''] \to [K']$ and $\psi'' : [K'''] \to [K'']$ of the identity map. Then $\theta' \circ \psi'$ and $\theta'' \circ \psi''$ are simplicial approximations of the identity map from $[K''']$

to $[K]$, and $\phi' \circ \psi'$ and $\phi'' \circ \psi''$ are simplicial approximations of the map f from $[K''']$ to $[L]$. So

$$(\overline{\phi' \circ \psi'})_p \circ (\overline{\theta' \circ \psi'})_p^{-1} = (\overline{\phi}'_p \circ \overline{\psi}'_p) \circ (\overline{\theta}'_p \circ \overline{\psi}'_p)^{-1} = \overline{\phi}'_p \circ (\overline{\theta}'_p)^{-1},$$

i.e., the homomorphism from $H_p(K)$ to $H_p(L)$ induced by the choice of subdivision K''' equals the one induced by choosing K'. Similarly, the homomorphism from $H_p(K)$ to $H_p(L)$ induced by the choice of subdivision K''' also equals the one induced by choosing K''. Thus K' and K'' induce the same homomorphism from $H_p(K)$ to $H_p(L)$. □

Consequently, any continuous function $f : [K] \to [L]$ induces a well-defined homomorphism of homology groups $\overline{f}_p : H_p(K) \to H_p(L)$. We can now strengthen Lemma 14.39 to apply to all continuous functions, not just affine ones.

Lemma 14.45. *Let K, L, and M be simplicial complexes and let $f : [K] \to [L]$ and $g : [L] \to [M]$ be continuous. Then $(\overline{g \circ f})_p = \overline{g}_p \circ \overline{f}_p$.*

Proof. Find a subdivision L' of L and a simplicial approximation $\psi : [L'] \to [M]$ of g, then find a subdivision K' of K and a simplicial approximation $\phi : [K'] \to [L']$ of f. Also let $\theta : [L'] \to [L]$ simplicially approximate the identity map and let $\chi_p : C_p(K) \to C_p(K')$ and $\chi'_p : C_p(L) \to C_p(L)$ be as before. Then $\theta \circ \phi$ simplicially approximates f and $\psi \circ \phi$ simplicially approximates $g \circ f$, so

$$\begin{aligned} \overline{g}_p \circ \overline{f}_p &= (\overline{\psi}_p \circ \overline{\chi}'_p) \circ (\overline{\theta}_p \circ \overline{\phi}_p \circ \overline{\chi}_p) \\ &= \overline{\psi}_p \circ \overline{\phi}_p \circ \overline{\chi}_p \\ &= (\overline{\psi \circ \phi})_p \circ \overline{\chi}_p \\ &= (\overline{g \circ f})_p, \end{aligned}$$

as desired. □

We can now finally show that homology groups are topologically invariant.

Theorem 14.46. *(Topological invariance of simplicial homology) Let K and L be simplicial complexes and suppose $[K] \cong [L]$. Then $H_p(K) \cong H_p(L)$ for all p.*

Proof. Let $f : [K] \to [L]$ be a homeomorphism with inverse $g : [L] \to [K]$. Then $\overline{g}_p \circ \overline{f}_p = (\overline{g \circ f})_p$ is the identity map from $H_p(K)$ to itself, for any p, and similarly $\overline{f}_p \circ \overline{g}_p$ is the identity map from $H_p(L)$ to itself. So $H_p(K) \cong H_p(L)$. □

Consequently, the notation $H_p(X)$ now makes sense for any triangulable space X. Let us evaluate the homology groups of $\mathbb{S}^n$ and $\mathbb{R}^n$. Here is the result:

Proposition 14.47. *$H_0(\mathbb{S}^n) \cong H_n(\mathbb{S}^n) \cong \mathbb{Z}$ and $H_p(\mathbb{S}^n) = \{0\}$ for $1 \leq p \leq n-1$. $H_0(\mathbb{R}^n) \cong \mathbb{Z}$ and $H_p(\mathbb{R}^n) = \{0\}$ for $p \geq 1$.*

To see this, first observe that the closed unit ball $\mathbb{D}^n$ of $\mathbb{R}^n$ can be triangulated as a simplicial complex consisting of a single n-simplex and all of its sub-simplices. As this is a cone, we already know its homology groups from Proposition 14.37: $H_0(\mathbb{D}^n) \cong \mathbb{Z}$ and $H_p(\mathbb{D}^n) = \{0\}$ for $p \geq 1$. This yields that $H_0(\mathbb{S}^n) \cong \mathbb{Z}$ and $H_p(\mathbb{S}^n) = \{0\}$ for $1 \leq p \leq n-1$, because if we triangulate $\mathbb{D}^{n+1}$ into a single $(n+1)$-simplex and triangulate $\mathbb{S}^n$ into all of its n-subsimplices, then we see that every p-chain in $\mathbb{S}^n$ is also a p-chain in $\mathbb{D}^{n+1}$, and conversely, if $p \leq n$ then every p-chain in $\mathbb{D}^{n+1}$ is also a p-chain in $\mathbb{S}^n$. (The one $(n+1)$-simplex in $\mathbb{D}^{n+1}$ cannot contribute to any p-chain if $p \leq n$.) So the homology groups of $\mathbb{S}^n$ and $\mathbb{D}^{n+1}$ agree for $p < n$. For $p = n$, observe that there are no n-boundaries in $\mathbb{S}^n$, so $H_n(\mathbb{S}^n) = Z_n(\mathbb{S}^n) = Z_n(\mathbb{D}^{n+1})$. But since $H_n(\mathbb{D}^{n+1}) = \{0\}$, we have $Z_n(\mathbb{D}^{n+1}) = B_n(\mathbb{D}^{n+1})$. And since every $(n+1)$-chain in $\mathbb{D}^{n+1}$ is a multiple of $c_{\hat{S}}$ where S is the one $(n+1)$-simplex in $\mathbb{D}^{n+1}$, and $\partial_{n+1}(a \cdot c_{\hat{S}}) = a \cdot \partial_{n+1}(c_{\hat{S}})$ for all $a \in \mathbb{Z}$, we see that $B_n(\mathbb{D}^{n+1}) \cong \mathbb{Z}$.

Also, we can triangulate $\mathbb{R}^n$ as an increasing union of polytopes each homeomorphic to $\mathbb{D}^n$: first triangulate the closed unit ball $\overline{B}_1(0)$ as a single n-simplex, then extend that to a triangulation of $\overline{B}_2(0)$, and so on. Thus any p-chain in $\mathbb{R}^n$, for any p, is also a p-chain in $\overline{B}_N(0)$ for some $N > 0$, which is homeomorphic to $\mathbb{D}^n$, and therefore if $p \geq 1$ it is also a p-boundary in $\overline{B}_N(0)$. This shows that $H_p(\mathbb{R}^n) = \{0\}$ for all $p \geq 1$, and we also have $H_0(\mathbb{R}^n) \cong \mathbb{Z}$ by an argument similar to the one used in the proof of Proposition 14.37.

Corollary 14.48. *(Invariance of dimension) If $n \neq m$ then $\mathbb{S}^n$ is not homeomorphic to $\mathbb{S}^m$ and $\mathbb{R}^n$ is not homeomorphic to $\mathbb{R}^m$.*

It is clear that $\mathbb{S}^n \ncong \mathbb{S}^m$, because these two spaces have different homology groups. And then since $\mathbb{R}^n$ and $\mathbb{R}^m$ are homeomorphic to $\mathbb{S}^n$ and $\mathbb{S}^m$ minus a single point, it follows that $\mathbb{R}^n \ncong \mathbb{R}^m$. Here one has to check that any putative homeomorphism between $\mathbb{R}^n$ and $\mathbb{R}^m$ could be extended to the one extra point, but this is easy. If $\mathbb{R}^n$ and $\mathbb{R}^m$ were homeomorphic, then every sequence in $\mathbb{R}^n$ that converges to the extra point in $\mathbb{S}^n$ must

eventually escape every compact subclass of $\mathbb{R}^n$; therefore, its image in $\mathbb{R}^m$ must also eventually escape every compact subclass there, and hence must converge to the extra point in $\mathbb{S}^m$. Thus we get continuity from $\mathbb{S}^n$ to $\mathbb{S}^m$, and continuity in the reverse direction is proven in the same way.

Corollary 14.49. *(Brouwer's fixed point theorem for $\mathbb{D}^n$) Every continuous function from $\mathbb{D}^n$ to itself has a fixed point.*

The proof is almost identical to the proof of Theorem 14.15. We use a hypothetical fixed-point free map to construct a retraction of $\mathbb{D}^n$ onto its boundary, then let this induce a surjective homomorphism from $H_{n-1}(\mathbb{D}^n) = \{0\}$ onto $H_{n-1}(\mathbb{S}^{n-1}) \cong \mathbb{Z}$, yielding a contradiction.

Chapter 15

Functional Analysis

15.1 Banach spaces

In this chapter I will use the symbols $\mathbb{F}$ and $\mathbb{F}_0$ to stand for the scalar field and its standard dense countable subfield, which could be either $\mathbb{R}$ and $\mathbb{Q}$, or else $\mathbb{C}$ and $\mathbb{Q} + i\mathbb{Q}$.

Definition 15.1. A *norm* on an $\mathbb{F}$-vector space E is a function $\|\cdot\| : E \to [0, \infty)$ that satisfies

(a) $\|v\| = 0$ if and only if $v = 0$
(b) $\|av\| = |a|\|v\|$
(c) $\|v + w\| \leq \|v\| + \|w\|$

for all $v, w \in E$ and $a \in \mathbb{F}$. A vector space equipped with a norm is a *normed vector space*. It is a *Banach space* if the associated metric $d(v, w) = \|v - w\|$ is complete.

Any linear subspace of a normed vector space that is complete for the inherited norm is a Banach space. I will call such a subspace a *Banach subspace*. Thus a linear subspace of a Banach space is Banach if and only if it is sequentially closed.

Every normed vector space can be completed (Theorem 10.8), and it is routine to check that the vector space operations and the norm extend to the completion, so the completion is a Banach space. Now here is a simple condition for completeness of the metric in a normed vector space. Say that an infinite series $\sum v_n$ is *summable* if the partial sums $\sum_{n=1}^{N} v_n$ converge, and *absolutely summable* if the sum $\sum \|v_n\|$ is finite.

Proposition 15.2. *Let E be a normed vector space in which every absolutely summable series in E is summable. Then E is a Banach space.*

Proof. Let (v_n) be a Cauchy sequence in E. It will suffice to show that some subsequence converges. Thus, by passing to a subsequence we may assume $\|v_{n+1} - v_n\| \leq 2^{-n}$ for $n \geq 1$. Now define $w_1 = v_1$ and $w_{n+1} = v_{n+1} - v_n$ for $n \geq 1$; then $\sum w_n$ is absolutely summable because

$$\sum_{n=1}^{\infty} \|w_n\| \leq \|w_1\| + \sum_{n=1}^{\infty} 2^{-n} < \infty.$$

So by hypothesis the partial sums of $\sum w_n$ converge. But the partial sums are just the vectors v_n, so (v_n) converges. □

Example 15.3. Let K be a compact space and write $C(K) = C(K, \mathbb{F})$, the pseudoclass of continuous functions from K into $\mathbb{F}$. $C(K)$ is a vector space with norm $\|f\|_\infty = \sup_{x \in \mathring{K}} |f(x)|$ where $\mathring{K}$ is some (any) dense subset of K. Indeed, it is a Banach space by Proposition 15.2 together with Corollary 10.31. If K is locally compact, we define $C_0(K)$ to be the class of functions $f \,\varepsilon\, C(K, \mathbb{F})$ such that, fixing a boundedly compact metric on K (Proposition 10.47), for all n we have $|f(x)| \leq \frac{1}{n}$ outside some ball. One easily checks that this condition is definite (and does not depend on the choice of boundedly compact metric), and that $C_0(K)$ is a Banach space under the norm $\|\cdot\|_\infty$ defined in the same way as above.

We call $\|\cdot\|_\infty$ the *sup norm* or *uniform norm*.

Separable Banach spaces have a lot of nice properties, so it is good to know that every $C(K)$ space is separable.

Proposition 15.4. *Let K be a compact space. Then $C(K)$ is separable.*

Proof. Every compact space is separable by definition. Fix a metric on K and let $\mathring{K}$ be a dense subset of K. Consider the open balls $B_r(x)$ with $r > 0$ rational and $x \in \mathring{K}$. These constitute a basis for the topology. Then consider all open classes U and V which can be written as finite unions of these basic open classes and such that $d(U, V) = \inf\{d(x, y) : x \in U \cap \mathring{K}$ and $y \in V \cap \mathring{K}\} > 0$. By Urysohn's lemma for metric spaces (Proposition 10.22) there is, for any such U and V, a continuous function $f = f_{U,V} : K \to [0, 1]$ with $U \subseteq f^{-1}(0)$ and $V \subseteq f^{-1}(1)$. Enumerate the $\mathbb{F}_0$-linear combinations of the functions $f_{U,V}$ together with 1_K as f_1, f_2, ... We must show that every $f \,\varepsilon\, C(K)$ can be uniformly approximated by the f_n's.

Fix a continuous function $g : K \to [0, 1]$. Since every function in $C(K)$ can be written as a linear combination of such functions, it will suffice to show that this g can be uniformly approximated by the f_n's. Let $C = g^{-1}([0, \frac{1}{3}])$, $C' = g^{-1}([0, \frac{2}{5}))$, $D = g^{-1}([\frac{2}{3}, 1])$, and $D' = g^{-1}((\frac{3}{5}, 1])$. Both

C' and D' are open, and hence each is a union of basic open classes. These basic open classes cover C and D, respectively, which are both compact, so we can find open classes U and V, each a finite union of basic open classes, such that $C \subseteq U \subseteq C'$ and $D \subseteq V \subseteq D'$. Then $g_1 = \frac{1}{3}(1 + f_{U,V})$, for this U and V, satisfies $\|g - g_1\|_\infty \leq \frac{1}{3}$. (If U is empty, interpret $f_{U,V}$ as 1_K, and if V is empty interpret it as the zero function. If both U and V are empty, either choice will work.) We can then find g_2 with $\|g - g_1 - g_2\|_\infty \leq \frac{1}{9}$, and so on. So g can be uniformly approximated by the f_n's. □

Before proceeding to other examples of Banach spaces, let us prove one important theorem about $C(K)$ spaces.

Lemma 15.5. *Let $a > 0$. Then there is a sequence of polynomial functions which converges uniformly to $|t|$ on the interval $[-a, a]$.*

Proof. If $\big||t| - p(t)\big| \leq \epsilon$ on $[-1, 1]$ then $\big||\frac{t}{a}| - p(\frac{t}{a})\big| \leq \epsilon$ on $[-a, a]$, and hence $\big||t| - ap(\frac{t}{a})\big| \leq a\epsilon$ on $[-a, a]$. So by scaling we can assume $a = 1$.

Recursively define a sequence of polynomials p_n by putting $p_0 = 0$ and $p_{n+1}(t) = p_n(t) + \frac{1}{2}(t - p_n(t)^2)$. We must show that $p_n(t)$ increases unformly to $\sqrt{t}$ on $[0, 1]$. This is enough because it implies that $p_n(t^2)$ increases uniformly to $|t|$ on $[-1, 1]$.

Short inductive calculations show that $0 \leq p_n(t) \leq \sqrt{t}$ and $\sqrt{t} - p_n(t) \leq \left(1 - \frac{\sqrt{t}}{2}\right)^n$ on $[0, 1]$, for all n. Now let $\epsilon \in (0, 1)$. If $t < \epsilon^2$ then trivially $\sqrt{t} - p_n(t) \leq \epsilon$ for any n, while if $t \geq \epsilon^2$ then $\sqrt{t} - p_n(t) \leq (1 - \frac{\epsilon}{2})^n$, which is also at most ϵ for sufficiently large n. Thus $p_n(t) \to \sqrt{t}$ uniformly on $[0, 1]$. □

We say that a linear subspace $\mathcal{A} \subseteq C(K)$ is *self-adjoint* if $f \,\varepsilon\, \mathcal{A}$ implies $\overline{f} \,\varepsilon\, \mathcal{A}$ (taking the adjoint pointwise; this condition is vacuous if $\mathbb{F} = \mathbb{R}$), and a *subalgebra* if $f, g \,\varepsilon\, \mathcal{A}$ implies $fg \,\varepsilon\, \mathcal{A}$, where this product is taken pointwise. It is *unital* if it contains the function 1_K. It *separates points* if for all distinct $x, y \in K$ there exists $f \,\varepsilon\, \mathcal{A}$ with $f(x) \neq f(y)$.

Theorem 15.6. *(Stone-Weierstrass theorem) Let K be a compact space and let $\mathcal{A}$ be a closed, self-adjoint, separable, unital subalgebra of $C(K)$ which separates points. Then $\mathcal{A} = C(K)$.*

Proof. Consider the case $\mathbb{F} = \mathbb{R}$ first. Fix $f \,\varepsilon\, C(K)$ and $\epsilon > 0$. Also fix a metric on K.

Let $\mathring{\mathcal{A}}$ be a dense subset of $\mathcal{A}$. For any distinct $x, y \in K$ and any $a, b \in \mathbb{R}$ there exists $g \,\varepsilon\, \mathcal{A}$ such that $g(x) = a$ and $g(y) = b$; this follows

easily from separation of points plus the fact that $1_K \,\varepsilon\, \mathcal{A}$. Thus, for any distinct $x, y \in K$, there exists $g \,\varepsilon\, \mathring{\mathcal{A}}$ with $|g(x) - f(x)|, |g(y) - f(y)| < \epsilon$.

Let $\mathring{K}$ be a dense subset of K and fix $x \in K$. I claim that there exists $\delta > 0$ with the property that for every $y \in \mathring{K}$ there exists $g \,\varepsilon\, \mathring{\mathcal{A}}$ with $|g(x) - f(x)| < \epsilon$ and $g(z) < f(z) + \epsilon$ for all $z \in \mathring{K} \cap B_\delta(y)$. To see this, first observe that by quantifying over $\mathring{\mathcal{A}}$ we can find, for each $y \in \mathring{K}$, the largest value δ_y for which these inequalities hold. But if the δ_y are not bounded away from zero then we can take a cluster point and get a contradiction; this proves the claim.

Since K is compact, we can now find finitely many $y_1, \ldots, y_n \in \mathring{K}$ and $g_1, \ldots, g_n \,\varepsilon\, \mathring{\mathcal{A}}$ such that $K = B_\delta(y_1) \cup \cdots \cup B_\delta(y_n)$, and $|g_i(x) - f(x)| < \epsilon$ for all i, and $g_i(z) < f(z) + \epsilon$ for all i and all $z \in \mathring{K} \cap B_\delta(y_i)$. The formulas

$$\max(h, k) = \frac{h + k + |h - k|}{2} \qquad \text{and} \qquad \min(h, k) = \frac{h + k - |h - k|}{2}$$

plus the lemma show that $g^x = \min(g_1, \ldots, g_n)$ belongs to $\mathcal{A}$. This function satisfies $|g^x(x) - f(x)| < \epsilon$ and $g^x(z) < f(z) + \epsilon$ for all $z \in \mathring{K}$.

By another compactness argument there exists $\delta' > 0$ such that for every $x \in \mathring{K}$ there exists $g_x \,\varepsilon\, \mathring{\mathcal{A}}$ satisfying $|g_x(x) - f(x)| < \epsilon$, $g_x(z) < f(z) + \epsilon$ for all $z \in \mathring{K}$, and $g_x(z) > f(z) - \epsilon$ for all $z \in \mathring{K} \cap B_{\delta'}(x)$. Find $x_1, \ldots, x_m \in \mathring{K}$ such that $K = B_{\delta'}(x_1) \cup \cdots \cup B_{\delta'}(x_m)$; then $g = \max(g_{x_1}, \ldots, g_{x_m}) \,\varepsilon\, \mathcal{A}$ satisfies $f(z) - \epsilon < g(z) < f(z) + \epsilon$ for all $z \in \mathring{K}$, and therefore $\|f - g\|_\infty \leq \epsilon$. Since $\mathcal{A}$ is closed and ϵ was arbitrary, we conclude that $f \,\varepsilon\, \mathcal{A}$. We have shown that $\mathcal{A} = C(K)$.

In the complex case let $\mathcal{A}_{sa}$ be the pseudoclass of real-valued functions in $\mathcal{A}$ and let $C_{sa}(K)$ be the pseudoclass of real-valued functions in $C(K)$. Since $\mathcal{A}$ is self-adjoint, the real and imaginary parts of any function in $\mathcal{A}$ belong to $\mathcal{A}$, and it follows that $\mathcal{A}_{sa}$ is a closed, separable, unital subalgebra of $C_{sa}(K)$ that separates points. By the real version of the theorem, this implies that $\mathcal{A}_{sa} = C_{sa}(K)$, and hence $\mathcal{A} = C(K)$. □

Now here is another, rather different class of examples.

Example 15.7. Let l^∞ be the linear subspace of $\mathbb{F}^\omega$ consisting of all bounded sequences in $\mathbb{F}$, with norm $\|\vec{a}\,\|_\infty = \sup |a_n|$. Let c_0 be the linear subspace of l^∞ consisting of all sequences whose entries converge to zero, with the inherited norm. For $1 \leq p < \infty$ let l^p be the linear subspace of $\mathbb{F}^\omega$ consisting of all sequences for which $\sum |a_n|^p$ is finite, with norm $\|\vec{a}\,\|_p = \left(\sum |a_n|^p\right)^{1/p}$.

These are all Banach spaces. In order to check this we must verify that $\|\cdot\|_p$ is a norm; after that it is a simple matter to show completeness using

Proposition 15.2. It is easy to see that $\|\cdot\|_1$ and $\|\cdot\|_\infty$ are norms on l^1 and l^∞, but for $p \in (1,\infty)$ we need some inequalities. Let us prove them in a more general form that will be useful later.

Definition 15.8. Let X be a measure space and let $1 \le p < \infty$. Say that a measurable function $f : X \to \mathbb{F}$ is *p-integrable* if $\int |f|^p < \infty$, and let the *p-norm* of such a function be $\|f\|_p = \left(\int |f|^p\right)^{1/p}$.

(Thus f is 1-integrable if and only if it is integrable in the sense of Section 13.3.)

Every $p \in (1,\infty)$ has a *conjugate exponent* $q = \frac{p}{p-1}$ which satisfies $\frac{1}{p} + \frac{1}{q} = 1$.

Proposition 15.9. *(Hölder and Minkowski inequalities) Let $1 < p < \infty$, let X be a measure space, and let $f, g : X \to \mathbb{F}$ be measurable. Then (a) $\|fg\|_1 \le \|f\|_p \|g\|_q$, where q is the conjugate exponent to p, and (b) $\|f+g\|_p \le \|f\|_p + \|g\|_p$.*

Proof. I claim that $st \le \frac{s^p}{p} + \frac{t^q}{q}$ for all $s, t \in [0,\infty)$. This can be proven using elementary calculus by fixing s and minimizing $\frac{t^q}{q} - st$. Now $\|f\|_p = 0$ implies $f = 0$ almost everywhere, which makes the first inequality trivial, and the same is true if $\|g\|_q = 0$, so we may assume $\|f\|_p$ and $\|g\|_q$ are both nonzero but finite. Then let $\tilde{f} = \frac{1}{\|f\|_p} f$ and $\tilde{g} = \frac{1}{\|g\|_q} g$ and estimate

$$\|\tilde{f}\tilde{g}\|_1 = \int |\tilde{f}\tilde{g}| \le \int \left(\frac{1}{p}|\tilde{f}|^p + \frac{1}{q}|\tilde{g}|^q\right) = \frac{1}{p} + \frac{1}{q} = 1.$$

Multiplying through by $\|f\|_p \|g\|_q$ yields the first (Hölder's) inequality.

For the second (Minkowski's) inequality, use Hölder's inequality to get

$$\begin{aligned}\int |f+g|^p &\le \int |f| \cdot |f+g|^{p-1} + \int |g| \cdot |f+g|^{p-1} \\ &\le \|f\|_p \||f+g|^{p-1}\|_q + \|g\|_p \||f+g|^{p-1}\|_q \\ &= (\|f\|_p + \|g\|_p) \left(\int |f+g|^p\right)^{1/q}.\end{aligned}$$

Thus $\left(\int |f+g|^p\right)^{1/p} = \left(\int |f+g|^p\right)^{1-1/q} \le \|f\|_p + \|g\|_p$, as desired. □

Taking $X = \mathbb{N}$ with counting measure (i.e., $\mu[A]$ is just the number of elements of A), we see that $\|\cdot\|_p$ is a norm on l^p. Thus the l^p spaces are Banach spaces. The spaces c_0 and l^p for $1 \le p < \infty$ are all separable, with $\mathbb{F}_0^{<\omega}$ as a countable dense subspace of each of them. However, l^∞ is not separable, because there are uncountably many sequences $\vec{a} \in \mathbb{F}^\omega$ all

of whose entries are either 0 or 1 (Theorem 9.6), and the distance in l^∞ between any two of them is 1.

We will give a measurable version of Example 15.7 using the following tool. Let X be a measure space with basic measurable classes R_i ($i \in \mathbb{N}$). As usual, measure-theoretic complications vanish when there are only finitely many basic measurable classes, so I will restrict to the case where the index set is infinite. I will also assume that the measure of every R_i besides $\emptyset$ is strictly positive; this can always be arranged by removing the union of the null R_i's from X. Let $S \subset \mathbb{F}^{<\omega}$ be the class of sequences $\vec{a}$ with only finitely many nonzero entries, such that (1) if $R_m = \emptyset$ then $a_m = 0$ and (2) if a_m and a_n are both nonzero then R_m and R_n are disjoint. Let $f_{\vec{a}}$ be the function $f_{\vec{a}} = \sum a_n 1_{R_n}$ and put $\vec{a} \sim \vec{b}$ if $f_{\vec{a}} = f_{\vec{b}}$. This condition is definite because we just have to check, for each nonzero value c in either sequence, whether the union of the R_n's with $a_n = c$ equals the union of the R_n's with $b_n = c$; the symmetric difference of the two unions is a basic measurable class, and every basic measurable class definitely is or is not the empty class. Thus $\sim$ is an equivalence relation on S, and we define the *basic function space* over X to be the quotient $\mathcal{M}_0(X) = S/\!\sim$.

For every element $[\vec{a}\,]$ of $\mathcal{M}_0(X)$ there is a corresponding function $f_{\vec{a}}$ on X. Thus we have an indexed family of functions, and it is exact in the sense of Section 9.3. Every function that can be written as $\sum a_n 1_{R_n}$ with only finitely many a_n's nonzero equals $f_{\vec{a}}$ for some $\vec{a} \in S$ and hence belongs to $\mathcal{M}_0(X)$. Since sums and scalar multiples of functions of this form have the same form, $\mathcal{M}_0(X)$ is an $\mathbb{F}$-vector space. For each $1 \leq p < \infty$, Proposition 15.9 (plus the assumption that $\emptyset$ is the only basic measurable class with measure zero) ensures that the p-norm $\|f\|_p$, for $f \,\varepsilon\, \mathcal{M}_0(X)$, is indeed a norm on $\mathcal{M}_0(X)$.

Definition 15.10. Let X be a measure space and assume $\emptyset$ is the only null basic measurable class. For any $1 \leq p < \infty$ we define $\tilde{L}^p(X) \subseteq \mathcal{M}_0(X)^\omega$ to be the class of sequences (f_n) in $\mathcal{M}_0(X)$ which satisfy $\|f_{n+1} - f_n\|_p \leq 2^{-n}$ for all n. Then we let $L^p(X)$ be the quotient of $\tilde{L}^p(X)$ by the equivalence relation which sets $(f_n) \sim (g_n)$ when $\|f_n - g_n\|_p \to 0$.

Since every sequence in $\mathcal{M}_0(X)$ that is Cauchy for the norm $\|\cdot\|_p$ has a subsequence in $\tilde{L}^p(X)$, it follows that $L^p(X)$ is "the" completion of $\mathcal{M}_0(X)$ for $\|\cdot\|_p$ (cf. the comment following Theorem 10.8). Thus $L^p(X)$ is a separable Banach space. The reason for detouring through $\tilde{L}^p(X)$ rather than just defining $L^p(X)$ to be this completion of $\mathcal{M}_0(X)$ is explained by the following result.

Proposition 15.11. *Let X be a measure space. Then every sequence (f_n) in $\tilde{L}^p(X)$ converges almost everywhere and in p-norm to a p-integrable function f. Every p-integrable function agrees almost everywhere with the limit of a sequence in $\tilde{L}^p(X)$, and two sequences in $\tilde{L}^p(X)$ are equivalent (i.e., they instantiate the same element of $L^p(X)$) if and only if their limits agree almost everwhere.*

Proof. Given a sequence (f_n) in $\tilde{L}^p(X)$, let $g_1 = f_1$ and $g_{n+1} = f_{n+1} - f_n$ for all $n \geq 1$, and then define $h_n = \sum_{i=1}^{n} |g_i|$ for all n and let $h = \sum_{i=1}^{\infty} |g_i|$. By the monotone convergence theorem (Theorem 13.34) plus Minkowski's inequality (Proposition 15.9 (b))

$$\|h\|_p = \lim_{n\to\infty} \|h_n\|_p \leq \lim_{n\to\infty} \sum_{i=1}^{n} \|g_i\|_p \leq \|f_1\|_p + 1 < \infty,$$

so that h must be finite almost everywhere. Thus $\sum g_i$ converges absolutely almost everywhere, to some limit function f. That is, $f_n \to f$ almost everywhere. Moreover, f is p-integrable because $|f| \leq h$ almost everywhere, and the dominated convergence theorem (Theorem 13.39) implies that $\|f - f_n\|_p \to 0$.

Now let g be any p-integrable function on X. We can find a sequence (g_n) of simple functions with $|g_n| \leq |g|$ which converges pointwise to g, and then by dominated convergence $\|g - g_n\|_p \to 0$. That is, g is approximated in p-norm by simple functions. As any simple function is approximated in p-norm by functions in $\mathcal{M}_0(X)$, it follows that we can find a sequence of functions $f_n \,\varepsilon\, \mathcal{M}_0(X)$ with $\|g - f_n\|_p \to 0$. By passing to a subsequence, we can assume that $(f_n) \,\varepsilon\, \tilde{L}^p(X)$, and by the first part of the proof its pointwise limit f satisfies $\|f - f_n\|_p \to 0$, so that $\|f - g\|_p = 0$, which implies that $f = g$ almost everywhere. Similar reasoning yields the last part of the theorem, that two sequences in $\tilde{L}^p(X)$ are equivalent if and only if their limits agree almost everywhere. □

This result shows that the elements of $L^p(X)$ exactly capture the notion "p-integrable function on X, up to modification on a null class". However, $L^p(X)$ does not literally index a family of functions on X in the sense of Section 9.3. That is done by $\tilde{L}^p(X)$. For every sequence (f_n) in $\tilde{L}^p(X)$, we have a literal function $f : X \to \mathbb{F}$ which is defined to be the pointwise limit of (f_n) wherever that limit exists, and zero where it does not. Not every p-integrable function is indexed by an element of $\tilde{L}^p(X)$, but every p-integrable function agrees almost everywhere with one that is.

Thus, when we quantify over $L^p(X)$ we are effectively quantifying over all p-integrable functions modulo equality almost everywhere, but $\tilde{L}^p(X)$ is the class that actually indexes a family of functions. I will follow the usual practice of writing "$f \,\varepsilon\, L^p(X)$", even though the function associated to an element of $L^p(X)$ is only defined almost everywhere.

For $p = \infty$ we need the following notion. Define the *essential sup* norm of a measurable function $f : X \to \mathbb{F}$ to be

$$\begin{aligned}\|f\|_\infty &= \inf\{r \geq 0 \text{ rational} : \mu[f^{-1}(B_r(0)^c)] = 0\} \\ &= \sup\{r \geq 0 \text{ rational} : \mu[f^{-1}(B_r(0)^c)] > 0\},\end{aligned}$$

taking $\inf \emptyset$ to be ∞. Thus the class of points on which $|f|$ is greater than $\|f\|_\infty$ has measure zero, and $\|f\|_\infty$ is the smallest value for which this is true. Say that f is *essentially bounded* if $\|f\|_\infty < \infty$.

Definition 15.12. Let X be a measure space with measure μ defined on basic measurable classes R_i, and assume $\emptyset$ is the only null basic measurable class. We define $L^\infty(X)$ to be the class of signed or complex measures ν for which there exists $C > 0$ such that $|\nu[R_i]| \leq C \cdot \mu[R_i]$ for all i. The norm of ν is the smallest possible value of C, i.e., the value $\sup \frac{|\nu[R_i]|}{\mu[R_i]}$ taken over all basic measurable classes besides $\emptyset$.

The definition is equivalent to saying that $|\nu[A]| \leq C \cdot \mu[A]$ for all μ-measurable classes A. Note that a signed or complex measure (relative to the given basic measurable classes) is, formally, just a sequence of real or complex numbers, so $L^\infty(X)$ is a legitimate class. It is an $\mathbb{F}$-vector space relative to the usual sum and scalar multiplication of measures, and it is a Banach space by Proposition 15.2. Its elements exactly capture the notion "essentially bounded function on X, up to modification on a null class":

Proposition 15.13. *Let X be a measure space. For any $\nu \in L^\infty(X)$ the Radon-Nikodym derivative (Theorem 13.47) $\frac{d\nu}{d\mu}$ is an essentially bounded measurable function on X (defined almost everywhere). The essential sup norm of $\frac{d\nu}{d\mu}$ equals the norm of ν in $L^\infty(X)$. Every essentially bounded measurable function on X arises in this way, up to modification on a null class, from an element of $L^\infty(X)$.*

Proof. Given $\nu \in L^\infty(X)$, let $f = \frac{d\nu}{d\mu}$ and let C be the norm of ν. Then $\|f\|_\infty \leq C$ because for any measurable class A we have $\left|\int_A f \, d\mu\right| = |\nu[A]| \leq C \cdot \mu[A]$. Also, for any A we have $|\nu[A]| = \left|\int_A f \, d\mu\right| \leq \|f\|_\infty \cdot \mu[A]$, showing that $C \leq \|f\|_\infty$.

Conversely, for any essentially bounded measurable function f the formula $\nu[A] = \int_A f\,d\mu$ defines a signed or complex measure in $L^\infty(X)$ for which $f = \frac{d\nu}{d\mu}$ almost everywhere. □

In contrast to the L^p spaces, we actually do have a family of bounded measurable functions indexed by $L^\infty(X)$. That is because the proof of Theorem 13.47 provided an explicit, albeit noncanonical, construction of the Radon-Nikodym derivative.

We conclude this section with some basic facts about finite dimensional subspaces of normed vector spaces.

Proposition 15.14. *Let E be a normed vector space. Then any finite subset of E is definitely either dependent or independent, it has a span, and the span is separable and complete.*

Proof. We prove all three statements by induction on the number of elements of the subset. We can take the base case to be the empty set, which is trivially independent and whose span is the zero subspace, which is trivially separable and complete. Now suppose the statement is true for all sets with n elements, and let A be an n-element subset of E and let $v \in E$. Then the span of A exists as a subclass of E, so v definitely does or does not belong to it, which means that $A \cup \{v\}$ is definitely either dependent or independent. To show that the span of $A \cup \{v\}$ exists, assume v does not belong to the span of A and let $w \in E$. Consider the function $f(t) = d(w - tv, \operatorname{span}(A))$ from $\mathbb{F}$ to $[0, \infty)$. Here the distance is defined as the infimum of $d(w - tv, u)$ for u ranging over some (any) dense subset of $\operatorname{span}(A)$. This function is continuous (even Lipschitz), and since $d(tv, \operatorname{span}(A)) = |t| \cdot d(v, \operatorname{span}(A)) \to \infty$ as $t \to \pm\infty$, it follows from the triangle inequality that $f(t) \to \infty$ as $t \to \pm\infty$. Therefore it attains its minimum value (the extreme value theorem, Theorem 10.36), and since $\operatorname{span}(A)$ is sequentially closed, w can be written as $w = u + tv$ for some $u \in \operatorname{span}(A)$ and $t \in \mathbb{F}$ if and only if the minimum value of f is zero. We have proven that $\operatorname{span}(A \cup \{v\})$ exists as a subclass of E.

Separability of this span is easy: if B is a dense subset of $\operatorname{span}(A)$ then $\{tv + w : t \in \mathbb{F}_0, w \in B\}$ is dense in $\operatorname{span}(A \cup \{v\})$. To show completeness, let (t_n) be a sequence in $\mathbb{F}$ and (w_n) a sequence in $\operatorname{span}(A)$ such that $(t_n v + w_n)$ is Cauchy. Then $\|t_n v + w_n\|$ must be bounded, so the sequence (t_n) must be bounded; otherwise we can pass to a subsequence with $v + \frac{1}{t_n} w_n \to 0$, which implies $v \in \operatorname{span}(A)$, contrary to our assumption. So by passing to a subsequence we can assume (t_n) converges, say $t_n \to t$. Then $t_n v \to tv$

and hence (w_n) is Cauchy, so (w_n) converges to an element of $\text{span}(A)$ by the induction hypothesis. This shows that $\text{span}(A \cup \{v\})$ is complete. □

Corollary 15.15. *Any subset of any normed vector space is definitely either dependent or independent and has a span.*

15.2 Bounded linear transformations

The natural morphisms between normed vector spaces are these.

Definition 15.16. Let E and F be normed vector spaces. A linear transformation $T : E \to F$ is said to be *bounded* if there exists $C > 0$ such that $\|Tv\| \leq C \cdot \|v\|$ for all $v \in E$. Equivalently, T is Lipschitz. It is an *isomorphism* if it is bijective and both it and its inverse are bounded. If E is separable we define the *operator norm* of T to be $\|T\| = \sup\{\|Tv\| : v \in \mathring{E}, \|v\| \leq 1\}$ where $\mathring{E}$ is some (any) dense subset of E.

A bijective linear transformation that is isometric is an *isometric isomorphism*, and two normed vector spaces E and F are *isometrically isomorphic*, written $E \cong F$, if there exists an isometric isomorphism between them. As usual, this is an unbounded existential quantifier, but we will see in the next section (Definition 15.31) that if the spaces are separable then we can take it to be bounded.

The importance of boundedness is seen in the next result. In its proof we use the notation $(E)_r$ for the open ball of E of radius r, i.e., $(E)_r = \{v \in E : \|v\| < r\}$. The closed ball $\{v \in E : \|v\| \leq r\}$ will be denoted $[E]_r$. A vector whose norm is 1 is a *unit vector*.

Proposition 15.17. *Let $T : E \to F$ be a linear transformation between normed vector spaces. If T is bounded then it is sequentially continuous. If E and F are both separable then T is bounded if and only if it is continuous.*

Proof. The first implication (bounded implies sequentially continuous) is easy. Now suppose E and F are separable and T is continuous. Then $T^{-1}((F)_1)$ is open in E and therefore contains a ball about the origin, say $(E)_r$ with $r > 0$. For any $r' < r$ and any unit vector $v \in E$ we have $r'v \in (E)_r$, hence $\|T(r'v)\| < 1$, hence $\|Tv\| < \frac{1}{r'}$. Thus $\|Tv\| \leq \frac{1}{r}$ for all unit vectors $v \in E$. This proves that T is bounded. Boundedness implies continuity by the first part of the proposition plus Theorem 10.20. □

A linear transformation $T : E \to F$ with $F = \mathbb{F}$ is a *linear functional*. Now let us prove the single most important theorem about bounded linear functionals.

Lemma 15.18. *Let E be a separable complex normed vector space. Then the real part of any bounded complex linear functional on E is a bounded real linear functional with the same norm, and every bounded real linear functional is the real part of a unique bounded complex linear functional.*

Proof. If $\phi : E \to \mathbb{C}$ is a bounded complex linear functional then the real part of ϕ, the map $v \to \mathrm{Re}(\phi(v))$, is clearly a bounded real linear functional, with norm at most $\|\phi\|$. Conversely, if $v \in E$ and $\phi(v) \neq 0$ then for $a = \frac{|\phi(v)|}{\phi(v)}$ we have $\|av\| = \|v\|$ and $\mathrm{Re}(\phi(av)) = |\phi(v)|$; letting v range over a dense subset of E yields that the norm of $v \mapsto \mathrm{Re}(\phi(v))$ is at least $\|\phi\|$.

Now let $\psi : E \to \mathbb{R}$ be a bounded real linear functional. Define $\phi : E \to \mathbb{C}$ by putting $\phi(v) = \psi(v) - i\psi(iv)$. This is a bounded complex linear functional whose real part is ψ. Uniqueness follows from the fact that the difference of two complex functionals with the same real part has null real part and hence is the zero functional. □

Theorem 15.19. *(Hahn-Banach theorem) Let E be a separable normed vector space, let E_0 be a separable linear subspace of E, and let $\phi : E_0 \to \mathbb{F}$ be a bounded linear functional. Then ϕ extends to a bounded linear functional ϕ on E with the same norm.*

Proof. Assume real scalars. It is easy to extend ϕ to the closure of E_0 without increasing its norm, so assume E_0 is closed. Let $\mathring{E}$ be a dense subset of E and $\mathring{E}_0$ a dense subset of E_0.

The key step is to show that we can extend ϕ to one additional dimension without increasing its norm. Thus let $v \in E_0^c$. We can extend linearly to $E_0 + v \cdot \mathbb{R} = \{av + w : a \in \mathbb{Q},\, w \in \mathring{E}_0\}$ by specifying a value b for $\phi(v)$. The norm will not increase provided $|ab + \phi(w)| \leq \|\phi\|\|av + w\|$ for all $a \in \mathbb{Q}$ and $w \in \mathring{E}_0$, and by dividing through by $|a|$ we can assume $a = 1$. But a short calculation with the triangle inequality shows that

$$\sup_{w \in \mathring{E}_0} (-\|\phi\|\|v + w\| - \phi(w)) \leq \inf_{w \in \mathring{E}_0} (\|\phi\|\|v + w\| - \phi(w)),$$

and any b lying between these two values works.

Say $\mathring{E} = \{v_1, v_2, \ldots\}$ and recursively define $E_n = E_{n-1} + v_n \cdot \mathbb{R}$ for $n \geq 1$. By the above we get a norm-preserving extension of ϕ to $\bigcup E_n$, which is dense in E. This is enough.

The theorem follows for complex scalars by the lemma. □

Our next topic is the open mapping theorem and its consequences. First we need a lemma (which already appeared, in disguised form, in the proof of the Tietze extension theorem, Theorem 10.71).

Lemma 15.20. *(Approximation lemma) Let E and F be Banach spaces and let $T : E \to F$ be a bounded linear transformation. Let $m > 0$ and $0 < r < 1$, and suppose that for each $w \in [F]_1$ there exists $v' \in E$ with $\|v'\| \leq m$ and $\|w - Tv'\| \leq r$. Then for each $w \in F$ there exists $v \in E$ with $\|v\| \leq \frac{m\|w\|}{1-r}$ and $Tv = w$.*

Proof. The hypothesis implies that for every $w \in F$ there exists $v' \in E$ with $\|v'\| \leq m\|w\|$ and $\|w - Tv'\| \leq r\|w\|$. Now fix $w \in F$. By scaling we can assume $\|w\| = 1$. Find v_1 with $\|v_1\| \leq m$ and $\|w - Tv_1\| \leq r$. Next, apply the hypothesis to $w - Tv_1$ in place of w to find v_2 with $\|v_2\| \leq rm$ and $\|w - T(v_1 + v_2)\| \leq r^2$. Continuing in this way, we obtain a sequence (v_n) in E such that $\|v_n\| \leq r^{n-1}m$ and $\|w - T(v_1 + \cdots + v_n)\| \leq r^n$ for all n. Thus $\sum v_n$ is absolutely summable to a vector v of norm at most $\frac{m}{1-r}$, and $\|w - Tv\| = 0$, i.e., $w = Tv$. □

Theorem 15.21. *(Open mapping theorem) Let E and F be separable Banach spaces and let $T : E \to F$ be a bounded linear transformation which is surjective. Then there exists $\epsilon > 0$ such that every element of $(F)_\epsilon$ is the image of an element of $(E)_1$. The image of any open subclass of E under T exists and is open.*

Proof. Let $\mathring{E}$ and $\mathring{F}$ be dense subsets of E and F, respectively, and for each $n \in \mathbb{N}$ let F_n be the closure of $T(\mathring{E} \cap (E)_n)$. Since T is surjective, we must have $F = \bigcup F_n$. The complements F_n^c are open, therefore separable, and therefore have closures which definitely do or do not contain $\mathring{F}$ and hence definitely are or are not all of F. Thus each F_n definitely does or does not contain a nonempty open class, and by the Baire category theorem (Theorem 10.18) plus the LPO it follows that one of them definitely does. By scaling this must be true of F_1. Let $B_\epsilon(w)$ be an open ball contained in F_1; then since any element of $(F)_{2\epsilon}$ is the difference of two elements of $B_\epsilon(w)$ (write $u \in (F)_{2\epsilon}$ as $u = (w + \frac{1}{2}u) - (w - \frac{1}{2}u)$), we have $(F)_{2\epsilon} \subseteq F_2$. Thus $(F)_\epsilon \subseteq F_1$, which shows that the hypothesis of the approximation lemma (Lemma 15.20) is satisfied with $m = \frac{1}{\epsilon}$ and arbitrary r. It follows that every element of $(F)_\epsilon$ is the image of an element of $(E)_1$.

Let $B_r(v)$ be an open ball in E. For each $w \in \mathring{E} \cap B_r(v)$ let $r_w = r - \|v - w\|$, so that $B_{r_w}(w) \subseteq B_r(v)$. Then $U = \bigcup_w B_{\epsilon \cdot r_w}(Tw)$ is an open subclass of F. I claim that every element of U is the image under T of an element of $B_r(v)$, and no element outside of U is; thus $T(B_r(v))$ exists and equals U. The first part of the claim follows from the fact that every element of $B_{\epsilon \cdot r_w}(Tw)$ is the image of an element of $B_{r_w}(w) \subseteq B_r(v)$. For the second part, let $u \in F$ and suppose $u = Tw'$ for some $w' \in B_r(v)$. Then we can find a sequence (w_n) in $\mathring{E} \cap B_r(v)$ that converges to w', so that $Tw_n \to u$, and $r_{w_n} \to r - \|v - w'\| > 0$, so u must belong to $B_{\epsilon \cdot r_{w_n}}(Tw_n)$ for sufficiently large n. But this shows that $u \in U$. We have shown that the image of $B_r(v)$ under T exists and is open; since every open subclass of E is the union of countably many open balls, it follows that the image of any open subclass of E exists and is open. □

The first corollary of the open mapping theorem is immediate:

Corollary 15.22. *(Banach isomorphism theorem) Let E and F be separable Banach spaces and let $T : E \to F$ be a bounded linear transformation which is bijective. Then T is an isomorphism.*

If E and F are normed vector spaces, their cartesian product is usually denoted $E \oplus F$, with a typical element written as $v \oplus w$ for $v \in E$ and $w \in F$. It can be given a variety of norms; maybe the simplest is $\|v \oplus w\|_\infty = \max(\|v\|, \|w\|)$.

Corollary 15.23. *(Closed graph theorem) Let E and F be Banach spaces with E separable and let $T : E \to F$ be a linear map. Then the following are equivalent:*

(a) T is bounded;
(b) if $v_n \to 0$ in E and $Tv_n \to w$ in F then $w = 0$;
(c) $\Gamma(T) = \{v \oplus Tv : v \in E\}$ is a Banach subspace of $E \oplus F$.

Proof. The implication (a) → (b) is trivial. Now assume (b) and let $(v_n \oplus Tv_n)$ be a sequence in $\Gamma(T)$ which converges to a limit point $v \oplus w$. Then $v_n - v \to 0$ in E and $T(v_n - v) \to w - Tv$ in F, so (b) implies that $w - Tv = 0$, i.e., $Tv = w$. Thus $v \oplus w$ also belongs to $\Gamma(T)$, proving (c). Finally, assume (c). Note that $\Gamma(T)$ is separable since E is separable. Let $P_1 : \Gamma(T) \to E$ and $P_2 : \Gamma(T) \to F$ be the two coordinate projections; then P_1 is a bounded linear bijection, so by Corollary 15.22 its inverse P_1^{-1} is bounded. Hence $T = P_2 \circ P_1^{-1}$ is bounded, proving (a). □

If E_1, E_2, ... are normed vector spaces, let $\bigoplus^\infty E_n$ be the class of all sequences $\vec{v} = (v_n)$ with $v_n \in E_n$ for all n and $\|\vec{v}\|_\infty = \sup \|v_n\| < \infty$.

Corollary 15.24. *(Principle of uniform boundedness; Banach-Steinhaus theorem) Let E and F be Banach spaces with E separable and let T_1, T_2, ... be bounded linear transformations from E to F. Suppose that $\sup \|T_n v\| < \infty$ for every $v \in E$. Then $\sup \|T_n\| < \infty$. If the sequence $(T_n v)$ converges in F for every $v \in E$, then there exists a bounded linear transformation $T : E \to F$ such that $T_n v \to Tv$ for all $v \in E$.*

Proof. Let $F_n = F$ for all $n \in \mathbb{N}$ and define $T_\infty : E \to \bigoplus^\infty F_n$ by $T_\infty v = (T_n v)$. I claim that T_∞ is bounded. To see this, suppose $v_k \to 0$ in E and $T_\infty v_k \to \vec{w}$ in $\bigoplus^\infty F_n$. Then $T_\infty v_k$ must converge to $\vec{w}$ on each coordinate, i.e., we must have $T_n v_k \to w_n$ for all n. But each T_n is bounded, so w_n must be zero for all n. That is, $\vec{w} = 0$. Corollary 15.23 now tells us that T_∞ is bounded. Now $\|T_n\| \leq \|T_\infty\|$ for all n, so $\sup \|T_n\| < \infty$. If $(T_n v)$ converges for all v then we can define $T : E \to F$ by putting $Tv = \lim T_n v$ for all $v \in E$; then T must be bounded because $\|Tv\| = \lim \|T_n v\| \leq \|T_\infty\| \cdot \|v\|$ for all $v \in E$. □

We conclude this section with a basic construction.

Definition 15.25. Let E be a Banach space and let E_0 be a separable Banach subspace of E. We define the *quotient space* E/E_0 to be the quotient of E by the equivalence relation which puts $v \sim w$ if $v - w \in E_0$. We define a *quotient norm* $\|\cdot\|_\sim$ on E/E_0 by putting $\|[v]\|_\sim = \inf\{\|v + w\| : w \in \mathring{E}_0\}$ for some (any) dense subset $\mathring{E}_0$ of E_0.

One easily verifies that this is a norm on E/E_0. The next result follows from Proposition 15.2.

Proposition 15.26. *Let E be a Banach space and let E_0 be a separable Banach subspace of E. Then E/E_0 with the quotient norm is a Banach space.*

It is clear that the natural projection from E onto E/E_0 is nonexpansive, and any bounded linear transformation with domain E whose kernel contains E_0 lifts to a bounded linear transformation with domain E/E_0. Conversely, if $T : E \to F$ is any surjective bounded linear transformation between separable Banach spaces, then $\ker(T)$ is separable and T lifts to an isomorphism between $E/\ker(T)$ and F. The last part follows from the

Banach isomorphism theorem (Theorem 15.22), but separability requires a separate proof.

Proposition 15.27. *Let E and F be separable Banach spaces and let $T : E \to F$ be a surjective bounded linear transformation. Then* $\ker(T)$ *is separable.*

Proof. Let $\mathring{E}$ be a dense subset of E and let ϵ be as in the open mapping theorem (Theorem 15.21). For all $v \in \mathring{E}$ and all $n \in \mathbb{N}$ such that $\|Tv\| < \frac{1}{n}$ there exists $w \in (E)_{\frac{1}{\epsilon n}}$ such that $Tv = Tw$. Thus $v - w \in \ker(T)$, and its distance to v is less than $\frac{1}{\epsilon n}$. Select one such w for all choices of v and n, and let $\mathring{E}_0$ be the resulting set of differences $v - w$. Then $\mathring{E}_0 \subseteq \ker(T)$, and for any $u \in \ker(T)$ and $n \in \mathbb{N}$ we can find $v \in \mathring{E}$ with $\|v - u\| < \frac{1}{\epsilon n}$ and $\|Tv\| < \frac{1}{n}$. The corresponding $v - w \in \mathring{E}_0$ then satisfies $\|(v - w) - u\| \leq \|v - u\| + \|w\| \leq \frac{2}{\epsilon n}$, so we see that $\mathring{E}_0$ is dense in $\ker(T)$. □

In particular, the kernel of any bounded linear functional on a separable Banach space is separable.

15.3 Dual Banach spaces

The dual Banach spaces that will be the main focus of this section are usually best thought of in topological, not metric, terms. Here is the general definition:

Definition 15.28. Let E be an $\mathbb{F}$-vector space. Then a compactly coherent topology on E is *compatible* with the vector space structure if the function $(v, w) \mapsto v + w$ is continuous from $E \times E$ to E and the function $(a, v) \mapsto av$ is continuous from $\mathbb{F} \times E$ to E. A *compactly coherent vector space* is an $\mathbb{F}$-vector space equipped with a compactly coherent topology that is compatible with the vector space structure.

The three isomorphism theorems for compactly coherent groups and rings discussed in Chapter 11 carry over with no essential alteration to the vector space setting. Indeed, every compactly coherent vector space is a compactly coherent abelian group, so the isomorphism theorems for compactly coherent vector spaces can be easily deduced from the isomorphism theorems for compactly coherent groups. We also have basic facts like the following.

Proposition 15.29. *Every finite subset of a compactly coherent vector space is definitely either dependent or independent and has a span which is closed and homeomorphically isomorphic to $\mathbb{F}^k$ for some k.*

Proof. The proof resembles the proof of Proposition 15.14. Let E be a compactly coherent vector space, let $v_1, \ldots, v_n \in E$, and let $v \in E$. We will prove the proposition for $\{v, v_1, \ldots, v_n\}$, assuming it holds for $\{v_1, \ldots, v_n\}$. Thus we assume that the latter is definitely either dependent or independent and has a span F which is closed and homeomorphically isomorphic to $\mathbb{F}^k$ for some k. This span definitely does or does not contain v. If it does contain v, or if $\{v_1, \ldots, v_n\}$ is dependent, then $\{v, v_1, \ldots, v_n\}$ is dependent, and otherwise it is independent. To show that the span of $\{v, v_1, \ldots, v_n\}$ exists, we can assume v lies outside of F (as otherwise F is itself the span). The key claim is that any limit of a sequence of elements of the form $av+w$ with $w \in F$ has the same form. To see this, let $(a_i v + w_i)$ be a sequence of elements of this form and suppose $a_i v + w_i \to u$. Then the sequence (a_i) definitely is or is not bounded in $\mathbb{F}$; it if is not, then we can pass to a subsequence and assume $|a_i| \to \infty$, and then $\frac{1}{a_i}(a_i v + w_i) \to 0$, so that $-\frac{1}{a_i} w_i \to v$, contradicting the facts that F is closed and v does not belong to F. Thus the sequence (a_i) must be bounded in $\mathbb{F}$, and by passing to a subsequence we can assume it converges, say to $a \in \mathbb{F}$, so that $a_i v \to av$ and hence $w_i \to u - av$. Then $u - av \in F$ and hence u is a linear combination of v and the v_i's. This proves the claim.

Let $\mathring{F}$ be a dense subset of F (since it is homeomorphic to $\mathbb{F}^k$) and let F' be the sequential pseudoclosure (Definition 10.92) of the set of points of the form $av + w$ with $a \in \mathbb{F}_0$ and $w \in \mathring{F}$. The claim proven above implies that F' is precisely the span of $\{v, v_1, \ldots, v_n\}$, so this span exists, and the claim also shows that F' is itself sequentially closed and therefore closed (Proposition 10.91).

We still have to show that F' is homeomorphically isomorphic to $\mathbb{F}^{k+1}$. Let $\phi : \mathbb{F}^k \to F$ be a homeomorphic isomorphism. It extends to an isomorphism $\phi : \mathbb{F}^{k+1} \to F'$ with $\phi(e_{k+1}) = v$. Sequential continuity, and therefore continuity (Proposition 10.90), of this map follows directly from the axioms of a compactly coherent vector space. Conversely, suppose $u_i \to 0$ in F'; we must show that $\phi^{-1}(u_i) \to 0$ in $\mathbb{F}^{k+1}$. This condition is definite, so suppose it fails. Then by passing to a subsequence we can assume there exists $r > 0$ such that $\|\phi^{-1}(u_i)\| \geq r$ for all i. The values $a_i = \frac{1}{\|\phi^{-1}(u_i)\|}$ are then bounded and so, by passing to another subsequence, we can assume (a_i) converges, say to a. But then the sequence $(a_i \phi^{-1}(u_i))$

lies on the unit sphere in $\mathbb{F}^{k+1}$, so by passing to one final subsequence we can assume it converges to a unit vector. Now continuity of ϕ implies that $(a_i u_i)$ must converge to something nonzero in F', but at the same time $a_i \to a$ and $u_i \to 0$ imply $a_i u_i \to 0$, a contradiction. We conclude that ϕ^{-1} is continuous. □

Corollary 15.30. *Every subset of a compactly coherent vector space has a span and is definitely either dependent or independent.*

The most important examples of compactly coherent vector spaces arise in the following way.

Definition 15.31. Let E and F be Banach spaces with E separable. Let $Z \subseteq E^\omega \times F^\omega$ be the class of pairs of sequence $((v_n), (w_n))$ such that the v_n are pairwise distinct and constitute a dense $\mathbb{F}_0$-linear subspace of E, and the function that takes v_n to w_n for all n is bounded and $\mathbb{F}_0$-linear. Any such function extends to a bounded linear transformation from E to F; let two elements of Z be equivalent if they extend to same bounded linear transformation, and define $B(E, F)$ to be the quotient of Z by this equivalence relation. For $F = \mathbb{F}$ this is the *dual space* of E, denoted $E^* = B(E, \mathbb{F})$. When $F = E$ we write $B(E) = B(E, E)$.

$B(E, F)$ is a pseudoclass that exactly indexes the bounded linear transformations from E to F. The operator norm makes every $B(E, F)$ into a Banach space (Proposition 15.2).

Here are two simple examples of dual Banach spaces.

Proposition 15.32. *We have $c_0^* \cong l^1$ and $(l^1)^* \cong l^\infty$.*

Proof. Every $\vec{a} \in l^1$ determines a linear functional $\phi_{\vec{a}}$ on c_0 by $\phi_{\vec{a}}(\vec{b}) = \sum a_n b_n$. Since

$$\left|\sum a_n b_n\right| \leq \sum |a_n| \cdot |b_n| \leq \|\vec{b}\|_\infty \cdot \sum |a_n| = \|\vec{a}\|_1 \cdot \|\vec{b}\|_\infty,$$

the operator norm of $\phi_{\vec{a}}$ is at most $\|\vec{a}\|_1$. Similarly, every $\vec{a} \in l^\infty$ determines a linear functional $\psi_{\vec{a}}$ on l^1 with operator norm at most $\|\vec{a}\|_\infty$.

Now let e_n be the sequence with a 1 in the nth place and zeros everywhere else. Given any bounded linear functional ϕ on c_0, define $a_n = \phi(e_n)$ for all n. Let $N \in \mathbb{N}$ and put $b_n = \frac{|a_n|}{a_n}$ (interpreting this quotient as zero if $a_n = 0$) for $1 \leq i \leq N$, and $b_n = 0$ for $n > N$. This yields a sequence $\vec{b}$ whose sup norm is at most 1 and which satisfies

$$|\phi(\vec{b})| = \sum_{n=1}^{N} |a_n|;$$

taking $N \to \infty$ yields $\|\vec{a}\,\|_1 \leq \|\phi\|$. So $\vec{a} \in l^1$, and $\phi_{\vec{a}}$ and ϕ agree on the linear span of the e_n's, which is dense in c_0, so $\phi = \phi_{\vec{a}}$. We conclude that every bounded linear functional on c_0 arises from pairing with an element of l^1, and its norm as a linear functional equals the 1-norm of the corresponding sequence. We have shown that $c_0^* \cong l^1$.

Given any bounded linear functional ψ on l^1, a similar argument shows that the values $b_n = \psi(e_n)$ determine a bounded sequence $\vec{b}$ with $\psi(\vec{a}\,) = \sum a_n b_n$ and $\|\vec{b}\,\|_\infty \leq \|\psi\|$, and such that $\psi = \psi_{\vec{b}}$. Together with the first paragraph, this completes the proof that $(l^1)^* \cong l^\infty$. □

The vectors e_n constitute a *Schauder basis* of either c_0 or l^1. This means that for every vector v in the space there is a unique sequence of scalars (a_n) with the property that $v = \sum a_n e_n$. This is the most important notion of "basis" in Banach spaces, a point I alluded to in Section 12.1.

If F is itself a dual Banach space then $B(E, F)$ can be given the following topology, which makes it a compactly coherent vector space.

Definition 15.33. Let E and F be separable Banach spaces with dense subsets $\mathring{E}$ and $\mathring{F}$. For any $r > 0$ the *weak** topology on the closed ball $[B(E, F^*)]_r$ is the weak topology (see the end of Section 10.4) for the functions $T \mapsto (Tv)(w)$ with $v \in \mathring{E}$ and $w \in \mathring{F}$.

The weak* topology on $[B(E, F^*)]_r$ does not depend on the choices of $\mathring{E}$ and $\mathring{F}$. A sequence (T_n) in $[B(E, F^*)]_r$ converges to $T \,\varepsilon\, [B(E, F^*)]_r$ if and only if $(T_n v)(w) \to (Tv)(w)$ for all $v \in E$ and $w \in F$. Also, for each $v \in \mathring{E}$ and $w \in \mathring{F}$ with $\|v\|, \|w\| \leq 1$ let $F_{v,w} = [\mathbb{F}]_r$; then $[B(E, F^*)]_r$ homeomorphically embeds in $\prod F_{v,w}$ as a closed subspace via the map that takes $T \,\varepsilon\, [B(E, F^*)]_r$ to the element of the product whose (v, w)-coordinate is $(Tv)(w)$. Since $\prod F_{v,w}$ is compact (the Tychonoff product theorem, Theorem 10.77), it follows that $[B(E, F^*)]_r$ is also compact for the weak* topology (Proposition 10.45). That establishes the next result.

Theorem 15.34. *(Banach-Alaoglu theorem) Let E and F be separable Banach spaces and let $r > 0$. Then $[B(E, F^*)]_r$ is weak* compact.*

The weak* topologies on $[B(E, F^*)]_r$ are clearly compatible for different values of r, so we can define the *coherent weak* topology* on $B(E, F^*)$ to be the compactly coherent topology arising from the filtration $[B(E, F^*)]_1 \subseteq [B(E, F^*)]_2 \subseteq \cdots$, equipping each $[B(E, F^*)]_n$ with the weak* topology.

Proposition 15.35. *Let E and F be separable Banach spaces. Then the coherent weak* topology makes $B(E, F^*)$ a compactly coherent vector space.*

Proof. We check that addition is continuous. By Propositions 10.90 and 10.95 it will suffice to show that if $T_n \to T$ and $T'_n \to T'$ in $B(E, F^*)$ then $T_n + T'_n \to T + T'$. By Proposition 10.86 (c) there exists N such that $T_n, T'_n, T, T' \varepsilon [B(E, F^*)]_N$ for all n. So $(T_n v + T'_n v)(w) = (T_n v)(w) + (T'_n v)(w) \to (Tv)(w) + (T'v)(w) = (Tv + T'v)(w)$ for all $v \in E$ and $w \in F$, and hence $T_n + T'_n \to T + T'$ weak* in $[B(E, F^*)]_{2N}$. The proof that scalar multiplication is continuous is similar. □

Although every compact space is, by definition, separably metrizable, the weak* topology generally does not arise from the Banach space metric on $[B(E, F^*)]_r$. Indeed, the latter is often nonseparable.

The scalar field $\mathbb{F}$ can be identified with its own dual and hence is itself a dual Banach space. In this case the coherent weak* topology really is just the usual (norm) topology.

Now we can pick up some basic properties of dual spaces. The next result is a disguised version of the Krein-Smulian theorem. For $v \in E$ define the *evaluation functional* $\dot{v} : E^* \to \mathbb{F}$ by $\dot{v}(\phi) = \phi(v)$.

Theorem 15.36. *Let E be a separable Banach space and let $\Phi : E^* \to \mathbb{F}$ be a linear functional which is continuous for the coherent weak* topology on E^*. Then there exists $v \in E$ such that $\Phi = \dot{v}$.*

Proof. Let $\mathring{E}$ be a dense subset of E. Note that $\ker(\Phi)$ definitely does or does not equal E (Proposition 10.56), so Φ definitely does or does not equal 0. If $\Phi = 0$ then we can take $v = 0$. Otherwise, the class $C = \Phi^{-1}(1)$ is coherent weak* closed in E^* and does not contain the origin. Now for all n the ball $[E^*]_{1/n}$ definitely does or does not meet C (Proposition 10.56), and if they all meet C then $0 \varepsilon C$, which is impossible. (Norm convergence implies weak* convergence.) Thus some $[E^*]_{1/n}$ does not meet C, and by scaling Φ we can assume $[E^*]_1$ does not meet C.

Let $C_1 = C \cap [E^*]_2$. Every element ϕ of C_1 has norm strictly greater than 1 so satisfies $\mathrm{Re}(\phi(v)) > 1$ for some $v \in \mathring{E} \cap [E]_1$. Thus the coherent weak* open classes $\{\phi \varepsilon E^* : \mathrm{Re}(\phi(v)) > 1\}$ for $v \in \mathring{E} \cap [E]_1$ cover C_1, and since C_1 is coherent weak* compact we can extract finitely many elements $v_1, \ldots, v_{i_1} \in \mathring{E} \cap [E]_1$ for which these classes cover C_1. In other words, C_1 is contained in the coherent weak* open class $V_1 = \{\phi \varepsilon E^* : \mathrm{Re}(\phi(v_i)) > 1$ for some $1 \le i \le i_1\}$.

Recursively, for $n > 1$ define

$$C_n = (C \cap [E^*]_{n+1}) \setminus (V_1 \cup \cdots \cup V_{n-1});$$

each of these will be coherent weak* compact, and inductively every element of C_n has norm strictly greater than n. So we can find $v_{i_{n-1}+1}, \ldots, v_{i_n} \in [E]_{1/n}$ such that $C_n \subseteq V_n = \{\phi \,\varepsilon\, E^* : \mathrm{Re}(\phi(v_i)) > 1 \text{ for some } i_{n-1} < i \leq i_n\}$.

Now since $\|v_i\| \to 0$ as $i \to \infty$, we have a linear map $\Psi : E^* \to c_0$ defined by letting the ith entry of $\Psi(\phi)$ be $\phi(v_i)$, for all $\phi \,\varepsilon\, E^*$. Let $\mathring{C}$ be a dense subset of C, let $X = \mathrm{span}(\Psi(\mathring{C}))$, and define $\psi : X \to \mathbb{F}$ by $\psi(\sum a_i \Psi(\phi_i)) = \sum a_i$ for $a_i \in \mathbb{F}$ and $\phi_i \,\varepsilon\, \mathring{C}$. To see that ψ is well-defined, it suffices to show that $\sum a_i \Psi(\phi_i) = 0$ implies $\sum a_i = 0$. But if $a = \sum a_i \neq 0$ then $\frac{1}{a} \sum a_i \phi_i \,\varepsilon\, C$, hence $\Psi(\frac{1}{a} \sum a_i \phi_i) \neq 0$, hence $\sum a_i \Psi(\phi_i) \neq 0$. This shows that ψ is well-defined. Moreover, if $|a| \geq 1$ then $\sum a_i \Psi(\phi_i) = a \cdot \Psi(\frac{1}{a} \sum a_i \phi_i) \notin [c_0]_1$ because $\Psi(\frac{1}{a} \sum a_i \phi_i)$ is the image under Ψ of an element of C and hence does not belong to $[c_0]_1$. This shows that $\|\psi\| \leq 1$, so by the Hahn-Banach theorem (Theorem 15.19) and Proposition 15.32 there exists $\vec{b} \in l^1$ such that $\sum b_i \cdot \phi(v_i) = \psi(\Psi(\phi)) = 1$ for all $\phi \,\varepsilon\, \mathring{C}$.

Define $v = \sum b_i v_i$. Then for all $\phi \,\varepsilon\, \mathring{C}$ we have $\phi(v) = \sum b_i \cdot \phi(v_i) = 1$, and this implies that $\Phi = \dot{v}$. □

Corollary 15.37. *Let E and F be separable Banach spaces and let $\Psi^* : F^* \to E^*$ be a coherent weak* continuous linear transformation. Then there is a bounded linear transformation $\Psi : E \to F$ such that $\Psi^*(\phi) = \phi \circ \Psi$ for all $\phi \,\varepsilon\, F^*$.*

Proof. For each $v \in E$ the map $\phi \mapsto \Psi^*(\phi)(v)$ is a coherent weak* continuous linear functional on F^*, so by Theorem 15.36 it has the form $\dot{w}$ for a unique $w \in F$. Define $\Psi(v) = w$. It is straightforward to check that Ψ is linear. To see that it is bounded, note first that by Proposition 10.86 (e) we have $\Psi^*([F^*]_1) \subseteq [E^*]_N$ for some $N \in \mathbb{N}$, i.e., Ψ^* is bounded. Now for all $v \in E$ and $\phi \,\varepsilon\, F^*$ we have

$$|\phi(\Psi(v))| = |(\Psi^*(\phi))(v)| \leq \|\Psi^*(\phi)\| \cdot \|v\| \leq N \cdot \|\phi\| \cdot \|v\|.$$

By the Hahn-Banach theorem (Theorem 15.19) there exists $\phi \,\varepsilon\, [F^*]_1$ with $\phi(\Psi(v)) = \|\Psi(v)\|$, so this yields $\|\Psi(v)\| \leq N \cdot \|v\|$. This shows that Ψ is bounded with $\|\Psi\| \leq N$. □

Theorem 15.36 can also be leveraged in the context of separation theorems.

Lemma 15.38. *Let U be an open convex subclass of $\mathbb{R}^2$ which does not contain the origin. Then there is a line through the origin that does not intersect U.*

Proof. Consider $V = \bigcup_{a\in\mathbb{Q}\cap(0,\infty)} aU$. This is an open wedge in $\mathbb{R}^2$, and since it is convex and does not contain the origin, it cannot subtend an angle greater than π. So there is a line through the origin that does not intersect it. □

Theorem 15.39. *Let E be a separable Banach space and let U be a coherent weak* open, convex subclass of E^* that does not contain the origin. Then there exists $v \in E$ such that* $\mathrm{Re}(\phi(v)) > 0$ *for all $\phi \in U$.*

Proof. By Lemma 15.18 it suffices to consider the case of real scalars. Let (ϕ_n) be a coherent weak* dense sequence in E^*. We may assume that $\phi_1 \in U$ and no other ϕ_n is a scalar multiple of ϕ_1.

Let $U_1 = U$, and recursively construct open, convex classes U_n that do not contain the origin together with real scalars a_n for $n \geq 2$ as follows. Given U_{n-1}, identify the span of ϕ_1 and ϕ_n with $\mathbb{R}^2$. The intersection of U_{n-1} with this plane is then an open convex subclass of $\mathbb{R}^2$ that does not contain the origin. By the lemma we can find $a_n \in \mathbb{R}$ such that U_{n-1} does not intersect the line $L_n = \mathbb{R}\cdot\tilde{\phi}_n$ spanned by the vector $\tilde{\phi}_n = a_n\phi_1 + \phi_n$. Let $U_n = U_{n-1} + L_n$ (cf. Proposition 11.10). This is a coherent weak* open, convex subclass of E^* that does not intersect the origin.

After this construction is completed, let F be the set of all finite linear combinations of the elements $\tilde{\phi}_2, \tilde{\phi}_3, \ldots$ with coefficients in $\mathbb{Q}$, with coherent weak* sequential pseudoclosure $\breve{F}$ (Definition 10.92). Observe that $U_n = U + \mathrm{span}\,\{\tilde{\phi}_2, \ldots, \tilde{\phi}_n\}$; since U_n does not contain the origin for any n, it follows that F does not intersect U, and hence $\breve{F}$ does not intersect U either. I claim that $\breve{F}$ has codimension 1; specifically, every element of E^* has the form $a\phi_1 + \psi$ for some $a \in \mathbb{R}$ and $\psi \in \breve{F}$. That is because the ϕ_n are dense in E^*, and hence the span of the set $\{\phi_1, \tilde{\phi}_2, \tilde{\phi}_3, \ldots\}$ is also dense. Thus for any $\phi\,\varepsilon\,E^*$ there exists a sequence $(b_i\phi_1 + \psi_i)$, with $b_i \in \mathbb{R}$ and $\psi_i\,\varepsilon\,F$, that converges to it in the coherent weak* topology. By replacing the sequence (b_i) in $\mathbb{R}$ with a sequence (b_i') in $\mathbb{Q}$ such that $b_i - b_i' \to 0$, we can assume the b_i are rational. Note that ϕ_1 does not belong to $\breve{F}$ since $\breve{F}$ is disjoint from U. Thus, as in the proof of Proposition 15.29, we can infer that the sequence (b_i) is bounded, then pass to a subsequence so that $b_i\phi_1 \to b\phi_1$ for some $b \in \mathbb{R}$; this implies that $\psi_i \to \phi - b\phi_1$, so that $\phi - b\phi_1\,\varepsilon\,\breve{F}$, and this proves the claim.

We can now show that $\breve{F}$ is in fact coherent weak* closed. By Proposition 10.91 this will follow from sequential closure. But if any sequence (ψ_n') in $\breve{F}$ converged to a point outside of $\breve{F}$, that limit point would have the form $a\phi_1 + \psi$ for some nonzero $a \in \mathbb{R}$ and some $\psi\,\varepsilon\,\breve{F}$. Then $\frac{1}{a}(\psi_n' - \psi)$ would be

a sequence in $\breve{F}$ that converges to $\phi_1 \varepsilon U$, a contradiction. So $\breve{F}$ is coherent weak* closed and $E^*/\breve{F} \cong \mathbb{R}$, so by Theorem 15.36 there exists $v \in E$ such that $\ker \dot{v} = \breve{F}$. Now, if some $\phi, \phi' \varepsilon U$ satisfy $\dot{v}(\phi) < 0 < \dot{v}(\phi')$, then since the line segment joining ϕ and ϕ' is contained in U and $\dot{v}$ is continuous, there would be an element of U that $\dot{v}$ maps to 0, a contradiction. Thus, possibly replacing v with $-v$, we conclude that $\phi(v) > 0$ for all $\phi \varepsilon U$. □

Corollary 15.40. *Let E be a separable Banach space and let K_1 and K_2 be disjoint coherent weak* compact, convex subclasses of E^*. Then there exist $v \in E$ and $a, b \in \mathbb{R}$ with*

$$\mathrm{Re}(\phi_1(v)) \leq a < b \leq \mathrm{Re}(\phi_2(v))$$

for all $\phi_1 \varepsilon K_1$ and $\phi_2 \varepsilon K_2$.

Proof. Let $C = K_2 - K_1$. This exists by Proposition 11.10, it is coherent weak* compact and convex, and it does not contain the origin. By Proposition 10.86 (d) there exists N such that $C \subset [E^*]_N$. Then we can find v_1, …, $v_n \in E$ and $\epsilon > 0$ such that C is disjoint from $U = \{\phi \varepsilon E^* : |\phi(v_i)| < \epsilon$ for all $i\}$. But U is convex, and therefore $C - U$ is a coherent weak* open, convex subclass of E^* that contains C but not the origin. Applying Theorem 15.39 to this class, we get $v \in E$ such that $\mathrm{Re}(\phi_2(v) - \phi_1(v)) > 0$ for all $\phi_1 \varepsilon K_1$ and $\phi_2 \varepsilon K_2$, and strict separation follows by compactness. □

Corollary 15.41. *Let E be a separable Banach space and let K be a coherent weak* compact, convex subclass of E^*. Suppose that $\|v\| \leq \sup_{\phi \varepsilon K} \mathrm{Re}(\phi(v))$ for all $v \in E$. Then $[E^*]_1 \subseteq K$.*

Proof. Let $\psi \varepsilon K^c$. According to Corollary 15.40 there exist $v \in E$ and real scalars $a < b$ such that $\mathrm{Re}(\phi(v)) \leq a < b \leq \mathrm{Re}(\psi(v))$ for all $\phi \varepsilon K$. Taking the supremum over ϕ (possible since K is compact) then yields $\|v\| \leq a$, so we must have $\|\psi\| > 1$. This shows that $[E^*]_1 \subseteq K$. □

15.4 The duals of $L^p(X)$ and $C(K)$

In this section we will work out a few more examples of dual spaces. First we consider L^p spaces.

Theorem 15.42. *Let X be a measure space, let $p \in (1, \infty)$, and let q be the conjugate exponent. Then $L^p(X)^* \cong L^q(X)$.*

Proof. By Hölder's inequality (Proposition 15.9 (a)), if f and g are, respectively, p- and q-integrable functions on X, then fg is integrable and $\left|\int fg\right| \leq \|fg\|_1 \leq \|f\|_p \|g\|_q$. Thus every element g of $L^q(X)$ induces a bounded linear functional ϕ_g on $L^p(X)$ defined by $\phi_g(f) = \int fg$. We have $\|\phi_g\| \leq \|g\|_q$.

Now let $\phi \,\varepsilon\, L^p(X)^*$ and let μ be the given measure on X. Suppose first that $\mu[X] < \infty$ and define $\nu[R_i] = \phi(1_{R_i})$ for every basic measurable class R_i. This is a signed or complex measure that absolutely continuous with respect to μ (Proposition 13.26) and satisfies $\nu[A] = \phi(1_A)$ for every μ-measurable class A. (As $\mu[A \triangle R_i]$ goes to zero, so do both $|\nu[A] - \nu[R_i]|$ and $|\phi(1_A) - \phi(1_{R_i})|$.) So according to the Radon-Nikodym theorem (Theorem 13.47) there exists an integrable function g satisfying $\phi(1_A) = \nu[A] = \int_A g$ for every μ-measurable class A. Taking linear combinations yields $\phi(f) = \int fg$ for all simple functions f.

I claim that g is q-integrable and $\|g\|_q \leq \|\phi\|$. This will complete the proof in the $\mu[X] < \infty$ case. First, for $n \in \mathbb{N}$ let $A_n = \{x : |g(x)| \leq n\}$ and define $g_n = g \cdot 1_{A_n}$. Then any p-integrable function f which is zero off of A_n can be approximated by simple functions f_k, and by dominated convergence (Theorem 13.39) we have both $\int f_k g \to \int fg$ and $\|f_k - f\|_p \to 0$, which implies $\phi(f_k) \to \phi(f)$. Thus $\phi(f) = \int fg$ for all such functions, and therefore

$$\left|\int f g_n\right| = \left|\int (f \cdot 1_{A_n}) g\right| = |\phi(f \cdot 1_{A_n})| \leq \|\phi\| \|f\|_p$$

for every p-integrable function f. Since $\int |g_n|^q \to \int |g|^q$ by the monotone convergence theorem (Theorem 13.34), it will suffice to show that $\|g_n\|_q \leq \|\phi\|$ for all n. But, interpreting $\frac{0}{0}$ as 0, we have

$$\|g_n\|_q^q = \left|\int g_n \cdot \frac{|g_n|^q}{g_n}\right| \leq \|\phi\| \cdot \| \, |g_n|^{q-1} \|_p = \|\phi\| \cdot \|g_n\|_q^{q/p}$$

since $\frac{|g_n|^q}{g_n}$ is bounded and hence p-integrable, so we get $\|g_n\|_q \leq \|\phi\|$ as desired.

If $\mu[X] = \infty$ then we can partition X into basic measurable classes R_{i_k}, then define $w = \sum 2^k \mu[R_{i_k}] \cdot 1_{R_{i_k}}$, and let μ' be the measure $\mu'[A] = \int_A \frac{1}{w} \, d\mu$. For any function f on X that is p-integrable with respect to μ, the function $w^{1/p} f$ is p-integrable with respect to μ', such that the map $f \mapsto w^{1/p} f$ is an isometric isomorphism between $L^p(X, \mu)$ and $L^p(X, \mu')$. We can now invoke the finite measure case to get a natural isometric isomorphism between $L^p(X, \mu')^*$ and $L^q(X, \mu')$ and then multiply by $w^{-1/q}$ to get back to $L^q(X, \mu)$. The pairing agrees because $\int w^{1/p} f \cdot w^{1/q} g \, d\mu' = \int w \cdot fg \, d\mu' = \int fg \, d\mu$. □

Theorem 15.43. *Let X be a measure space. Then $L^1(X)^* \cong L^\infty(X)$.*

Proof. Given $g \,\varepsilon\, L^\infty(X)$ and $f \,\varepsilon\, L^1(X)$, define $\phi_g(f) = \int fg$. The inequality $\left|\int fg\right| \leq \|g\|_\infty \int |f|$ is trivial, and this shows that ϕ_g is well-defined on $L^1(X)$ and satisfies $\|\phi_g\| \leq \|g\|_\infty$. For the reverse inequality, let $0 < \epsilon < \|g\|_\infty$ and let A be a positive measure subclass of X such that $|g(x)| \geq \|g\|_\infty - \epsilon$ on A. We can assume $\mu[A] < \infty$. Then $f = \frac{|g|}{g} \cdot 1_A$ is integrable and satisfies $\|f\|_1 = \mu[A]$ and

$$\left|\int fg\right| = \int_A |g| \geq (\|g\|_\infty - \epsilon)\mu[A].$$

This shows that $\|\phi_g\| \geq \|g\|_\infty - \epsilon$, and since ϵ was arbitrary we conclude that $\|\phi_g\| = \|g\|_\infty$.

Now let $\phi \,\varepsilon\, L^1(X)^*$. Then $\nu[A] = \phi(1_A)$ defines an element of the literal class $L^\infty(X)$ (Definition 15.12). The corresponding essentially bounded function $g = \frac{d\nu}{d\mu}$ satisfies $\int fg\, d\mu = \phi(f)$ for all $f \,\varepsilon\, L^1(X)$ by Corollary 13.48, so $\phi = \phi_g$. □

Now let us analyze the dual of $C(K)$.

A *Borel measure* on a compact space is a measure whose basic measurable classes are generated from a topological basis using finite unions and complements. Since every open class is a countable union of basic open classes, it follows from Corollary 13.18 that every open class is measurable, and then Proposition 13.23 shows that we get the same measurable classes with the same measures independently of the choice of topological basis used to generate the basic measurable classes.

A minor subtlety here is that we require basic measurable classes to be distinct, so we need it to be the case that any two classes generated from a topological basis using finite unions and complements definitely are or are not equal, in order that duplicates can be discarded. It is a little exercise in using de Morgan's laws to infer this definiteness from Proposition 10.56. (Use the fact that $A = B$ if and only if $A \cap B^c = A^c \cap B = \emptyset$, so it is enough for each basic measurable class to definitely be either empty or inhabited. Then write any basic measurable class as a finite union of classes of the form $U \cap C$ with U open and C closed.)

Borel measures can be positive, or signed, or complex. However, even in the positive case they are necessarily finite because K is a union of basic open classes, and hence by compactness a union of finitely many basic open classes. (Recall that positive measures are, by definition, finite on every basic measurable class.)

In fact, Borel measures are already determined by their values on open classes. The following definition will be used to show this. Since every signed or complex measure is a linear combination of positive measures, it suffices to consider the positive case.

Definition 15.44. Let K be a compact space. Open classes U_i ($1 \leq i \leq n$ or $i \in \mathbb{N}$) constitute a *full basis* if they are a basis for the topology on K, for every i and j the classes $U_i \cap U_j$ and $U_i \cup U_j$ appear somewhere in the list, and there is no repetition. Given a full basis U_1, U_2, ..., a *(positive) Borel premeasure* is a sequence $\mu_0 = (a_i)$ of elements of $[0, \infty)$ such that, writing $\mu_0[U_i] = a_i$, we have (1) $\mu_0[\emptyset] = 0$, (2) $U_i \subset U_j$ implies $\mu_0[U_i] \leq \mu_0[U_j]$, (3) $\mu_0[U_i] + \mu_0[U_j] = \mu_0[U_i \cup U_j] + \mu_0[U_i \cap U_j]$ for all i and j, and (4) for any i and any $\epsilon > 0$ there exists j such that $K = U_i \cup U_j$ and $\mu_0[U_i \cap U_j] \leq \epsilon$.

Fix a full basis. Note that if $C \subseteq K$ is compact, U is open, and $C \subseteq U$, then there must be a basic open class V satisfying $C \subseteq V \subseteq U$. That is because U is a union of basic open classes, and these classes cover C, therefore finitely many of them cover C, therefore a single one covers C. Indeed, using normality we can first shrink U, then make the above argument to get that $\overline{V} \subseteq U$. We will use this observation several times.

Lemma 15.45. *Let K be a compact space, let U_1, U_2, ... constitute a full basis, and let μ_0 be a Borel premeasure defined on this basis. Let $A \subseteq K$ be a class with the property that for every $\epsilon > 0$ there exist j and k such that $A \subseteq U_j$, $A^c \subseteq U_k$, and $\mu_0[U_j \cap U_k] \leq \epsilon$. Then A^c has this property as well. If B has the same property then so do $A \cup B$ and $A \cap B$.*

Proof. The claim for A^c is trivial because the condition is symmetric under complementation. If U_j and U_k verify the condition for A and $U_{j'}$ and $U_{k'}$ verify it for B then $A \cup B \subseteq U_j \cup U_{j'}$, $(A \cup B)^c \subseteq U_k \cap U_{k'}$, and

$$\begin{aligned}\mu_0\left[(U_j \cup U_{j'}) \cap (U_k \cap U_{k'})\right] &\leq \mu_0\left[(U_j \cap U_k) \cup (U_{j'} \cap U_{k'})\right] \\ &\leq \mu_0[U_j \cap U_k] + \mu_0[U_{j'} \cap U_{k'}] \leq 2\epsilon.\end{aligned}$$

The claim for $A \cap B$ follows by taking complements. □

Theorem 15.46. *Every positive Borel measure on a compact space induces a Borel premeasure on any full basis. Conversely, every Borel premeasure extends to exactly one Borel measure on the measurable basis generated by the full basis on which it was defined.*

Proof. Let μ be a positive Borel measure on a compact space and let V_1, V_2, ... constitute a full basis. Then the values $\mu[V_i]$ clearly satisfy the

first three conditions to be a Borel premeasure given in Definition 15.44. For the fourth condition, fix i and $\epsilon > 0$. Then V_i equals the union of the V_k's whose closure is contained in it, so there exists k with $\overline{V}_k \subseteq V_i$ and $\mu[\overline{V}_k] \geq \mu[V_i] - \epsilon$. Then by normality there exists an open class U that contains V_i^c and whose closure is disjoint from $\overline{V}_k$, and by compactness a basic open class V_j that contains $\overline{U}$ and is contained in $\overline{V}_k^c$. We now have $V_i \cup V_j = K$ and

$$\mu[V_i \cap V_j] = \mu[V_i] - \mu[V_j^c] \leq \mu[V_i] - \mu[\overline{V}_k] \leq \epsilon.$$

This verifies the fourth condition from Definition 15.44.

Now let $U_1, U_2, \ldots$ constitute a full basis and let μ_0 be a Borel premeasure defined on this basis. Let $R_1, R_2, \ldots$ be the classes generated from the U_i via finite unions and complements, without repetition. Then define $\nu[R_i] = \inf\{\mu_0[U_j] : R_i \subseteq U_j\}$; monotonicity (condition (2) in Definition 15.44) entails that $\nu[U_i]$ equals $\mu_0[U_i]$ for every U_i. We will show that ν is a measure.

We must show that if R_{i_k} are disjoint basic measurable classes whose union is also a basic measurable class R_i, then $\sum \nu[R_{i_k}] = \nu[R_i]$. First, note that it follows from the lemma plus condition (4) in Definition 15.44 that for every i and every $\epsilon > 0$ there exist j and k such that $R_i \subseteq U_j$, $R_i^c \subseteq U_k$, and $\mu_0[U_j \cap U_k] \leq \epsilon$. This already shows that if ν is a measure, it is the only measure that extends μ_0. Now let us check finite additivity. Suppose R_i and $R_{i'}$ are disjoint. Then $\nu[R_i \cup R_{i'}] \leq \nu[R_i] + \nu[R_{i'}]$ because $R_i \subseteq U_j$ and $R_{i'} \subseteq U_{j'}$ imply $R_i \cup R_{i'} \subseteq U_j \cup U_{j'}$, and therefore

$$\nu[R_i \cup R_{i'}] \leq \mu_0[U_j \cup U_{j'}] \leq \mu_0[U_j] + \mu_0[U_{j'}];$$

taking an infimum yields the stated inequality. For the reverse inequality, let $\epsilon > 0$ and find U_k and U_l such that $R_i \subseteq U_k$, $R_i^c \subseteq U_l$, and $\mu_0[U_k \cap U_l] \leq \epsilon$; then for any $U_{k'}$ that contains $R_i \cup R_{i'}$ we have $R_i \subseteq U_k \cap U_{k'}$ and $R_{i'} \subseteq U_l \cap U_{k'}$, so that

$$\begin{aligned}
\nu[R_i] + \nu[R_{i'}] &\leq \mu_0[U_k \cap U_{k'}] + \mu_0[U_l \cap U_{k'}] \\
&= \mu_0\left[(U_k \cap U_{k'}) \cup (U_l \cap U_{k'})\right] + \mu_0\left[(U_k \cap U_{k'}) \cap (U_l \cap U_{k'})\right] \\
&= \mu_0[U_{k'}] + \mu_0[U_k \cap U_l \cap U_{k'}] \\
&\leq \mu_0[U_{k'}] + \epsilon.
\end{aligned}$$

Taking an infimum completes the proof that $\nu[R_i \cup R_{i'}] = \nu[R_i] + \nu[R_{i'}]$.

We can now verify countable additivity. Let R_{i_k} be disjoint basic measurable classes whose union is a basic measurable class R_i. Finite additivity immediately yields that

$$\sum_{k=1}^{N} \nu[R_{i_k}] = \nu\left[\bigcup_{k=1}^{N} R_{i_k}\right] \leq \nu[R_i]$$

for all N, so that $\sum \nu[R_{i_k}] \leq \nu[R_i]$. For the reverse inequality, fix ϵ and for each k find U_{j_k} with $R_{i_k} \subseteq U_{j_k}$ and $\mu_0[U_{j_k}] \leq \nu[R_{i_k}] + \frac{\epsilon}{2^k}$. Then find j and j' with $R_i \subseteq U_j$, $R_i^c \subseteq U_{j'}$, and $\mu_0[U_j \cap U_{j'}] \leq \epsilon$. Since $U_{j'}^c$ is compact and contained in $\bigcup U_{j_k}$, it must be contained in $\bigcup_{k=1}^N U_{j_k}$ for some N. But $U_j = U_{j'}^c \cup (U_j \cap U_{j'})$, so (using finite additivity)

$$\begin{aligned}\mu_0[U_j] &= \nu[U_{j'}^c] + \nu[U_j \cap U_{j'}] \\ &\leq \sum_{k=1}^N \nu[U_{j_k}] + \epsilon \\ &\leq \sum_{k=1}^N \nu[R_{i_k}] + 2\epsilon.\end{aligned}$$

Since $R_i \subseteq U_j$, this yields $\nu[R_i] \leq \sum \nu[R_{i_k}] + 2\epsilon$, and taking $\epsilon \to 0$ gives us countable additivity. This completes the proof that ν is a measure that agrees with μ_0 on the U_i's. □

Theorem 15.47. *Let μ be a positive Borel measure on a compact space K and suppose $A \subseteq K$ is μ-measurable. Then*

$$\mu[A] = \inf\{\mu[U] : A \subseteq U \text{ open}\} = \sup\{\mu[K] : A \supseteq K \text{ compact}\}.$$

Proof. We can assume that the basic measurable classes for μ are generated (via finite unions and complements) from a full basis U_1, U_2, ... We know that $\mu[A] \leq \mu[U]$ for any open class U containing A. Conversely, for any $\epsilon > 0$ we can find basic measurable classes R_{i_k} which cover A and satisfy $\sum \mu[R_{i_k}] \leq \mu[A] + \epsilon$ (Lemma 13.15); then, by Lemma 15.45, for each k we can find a basic open class U_k which contains R_{i_k} and satisfies $\mu[U_k] \leq \mu[R_{i_k}] + \frac{\epsilon}{2^k}$. Then $U = \bigcup U_k$ is an open class that contains A and satisfies $\mu[A] \geq \mu[U] - \epsilon$. This proves the first equality; the second follows by taking complements. □

Definition 15.48. Let K be a compact space. Depending on whether $\mathbb{F}$ is $\mathbb{R}$ or $\mathbb{C}$, we define $M(K)$ to be the class of all signed or complex Borel measures on K (relative to some choice of basic measurable classes). This is a normed vector space with the obvious addition and scalar multiplication and the norm $\|\mu\| = |\mu|(K)$.

It straightforwardly follows from Proposition 15.2 that $M(K)$ is in fact a Banach space.

Lemma 15.49. *Let K be a compact space and let μ be a signed or complex Borel measure on K. Then $\phi_\mu : f \mapsto \int f \, d\mu$ is a bounded linear functional on $C(K)$ and its norm equals $\|\mu\|$.*

Proof. Every continuous function on K is bounded and therefore integrable relative to μ. The map ϕ_μ is clearly linear, and it is bounded with norm $\|\phi_\mu\| \leq \|\mu\| = |\mu|[K]$ because

$$\left|\int f\, d\mu\right| = \left|\int f \cdot \frac{d\mu}{d|\mu|}\, d|\mu|\right| \leq \int |f|\, d|\mu| \leq \|f\|_\infty \cdot |\mu|[K].$$

For the reverse inequality, let $K = A_1 \cup \cdots \cup A_n$ be a partition of K into measurable classes and let $\epsilon > 0$. By Theorem 15.47 we can find compact subclasses $K_k \subseteq A_k$ with $|\mu|[A_k \setminus K_k] \leq \frac{\epsilon}{n}$ $(1 \leq k \leq n)$. Then apply the Tietze extension theorem (Theorem 10.71) to find a continuous function f on K satisfying $\|f\|_\infty \leq 1$ and $f|_{K_k} = \frac{|\mu[K_k]|}{\mu[K_k]}$ for all k, so that

$$\int_{\bigcup K_k} f\, d\mu = \sum |\mu[K_k]| \geq \sum |\mu[A_k]| - \epsilon$$

and

$$\left|\int_{(\bigcup K_k)^c} f\, d\mu\right| \leq \int_{(\bigcup K_k)^c} |f|\, d|\mu| \leq \epsilon.$$

Since ϵ was arbitrary, it follows that $\|\phi_\mu\| \geq \sum |\mu[A_k]|$, and taking the supremum over all partitions of K (which exists by Proposition 13.30) yields $\|\phi_\mu\| \geq \|\mu\|$. □

The *support* of a continuous function $f : K \to \mathbb{F}$ is the closure of the complement of $f^{-1}(0)$.

Theorem 15.50. *(Riesz-Markov) We have $C(K)^* \cong M(K)$ for every compact space K.*

Proof. Let K be a compact space. Lemma 15.49 isometrically embeds $M(K)$ in $C(K)^*$; to complete the proof we must show that every bounded linear functional on $C(K)$ arises from integration against a Borel measure. We can reduce to the case of real scalars by restricting the real and imaginary parts of a bounded linear functional to the real part of $C(K)$, finding signed measures μ_1 and μ_2 which implement them, and then setting $\mu = \mu_1 + i\mu_2$.

Thus assume real scalars. Let C be the class of positive $\phi \,\varepsilon\, [C(K)^*]_1$ which satisfy $\phi(1_K) \leq 1$, where ϕ is *positive* if $\phi(f) \geq 0$ for all positive $f \,\varepsilon\, C(K)$. This condition is definite because it can be checked on the positive functions in a dense subset of $C(K)$ (and whether a function is positive can be checked on a dense subset of K). It is easy to see that C is convex and coherent weak* compact. Thus the class $C - C$, which is the image of

$C \times C$ under the map $(\phi, \psi) \mapsto \phi - \psi$ from $C(K)^* \times C(K)^*$ to $C(K)^*$, is also convex and weak* compact. Now for any continuous function f there exists $x \in K$ such that $f(x) = \pm\|f\|_\infty$; then the linear functional $\phi : g \mapsto \pm g(x)$ belongs to $C - C$ and satisfies $\phi(f) = \|f\|_\infty$, so by Corollary 15.41 we get $[C(K)^*]_1 \subseteq C - C$. By scaling, we conclude that every bounded linear functional is the difference of two positive linear functionals. So we need only show that any positive linear functional on $C(K)$ arises from integration against a Borel measure.

Let ϕ be a positive linear functional on $C(K)$. We define a Borel premeasure μ_0 as follows. First, for any open class $U \subseteq K$ and any continuous function $f : K \to \mathbb{R}$ write $f \prec U$ to mean that $0 \leq f(x) \leq 1$ everywhere and f is supported on U. Fix a full basis. For any basic open class U, let V range over all basic open classes whose closure is contained in U, and for each such V find a continuous function $f_V \prec U$ with $f_V = 1$ on V. Then define $\mu_0[U]$ to be $\sup_V \phi(f_V)$.

We must show that μ_0 is a Borel premeasure. First, note that any $f \prec U$ satisfies $\phi(f) \leq \mu_0[U]$, so that $\mu_0[U]$ is the supremum of $\phi(f)$ over all $f \prec U$, i.e., the specific choice of the f_V's doesn't matter. This is because any such function f is supported on a compact subclass of U, and is therefore supported on a single basic open class V whose closure is contained in U; we then have $\phi(f) \leq \phi(f_V) \leq \mu_0[U]$.

Conditions (1) and (2) of Definition 15.44 are trivial. For condition (4), given a basic open class U and $\epsilon > 0$, find $f \prec U$ with $\phi(f) \geq \mu_0[U] - \epsilon$. Then find a basic open class V that contains U^c and is contained in $f^{-1}(0)$. For any $g \prec U \cap V$ we then have $f + g \prec U$ and therefore $\phi(f + g) \leq \mu_0[U]$. Thus $\phi(g) \leq \mu_0[U] - \phi(f) \leq \epsilon$, showing that $\mu_0[U \cap V] \leq \epsilon$. This verifies (4).

To verify condition (3), let U and V be basic open classes. Given $f \prec U$ and $g \prec V$, let $C \subseteq U$ and $D \subseteq V$ be compact classes on which f and g are respectively supported and use Urysohn's lemma (Theorem 10.69) to find h and k satisfying $1_{C \cup D} \leq h \prec U \cup V$ and $1_{C \cap D} \leq k \prec U \cap V$. Then $f + g \leq h + k$, and we can now infer that $\mu_0[U] + \mu_0[V] \leq \mu_0[U \cup V] + \mu_0[U \cap V]$. For the reverse inequality, choose any h and k with $h \prec U \cup V$ and $k \prec U \cap V$. Then $U \cup V$ is the union of $U' \cup V'$ over all basic open classes U' and V' whose closures are contained in U and V, respectively, so there exist such U' and V' with $h \prec U' \cup V'$. Similarly, $k \prec U' \cap V'$ for some such U' and V', and by taking unions we can assume it is the same U' and V'. Then for any $1_{U'} \leq f \prec U$ and $1_{V'} \leq g \prec V$ we have $h + k \leq f + g$, and we can now infer that $\mu_0[U \cup V] + \mu_0[U \cap V] \leq \mu_0[U] + \mu_0[V]$. This completes the

proof that μ_0 is a Borel premeasure.

Let μ be the Borel measure generated from μ_0. We must show that $\int f\,d\mu = \phi(f)$ for all $f \,\varepsilon\, C(K)$. Observe that $\mu[U]$ is the supremum of $\phi(f)$ over all $f \prec U$, for any open class U; this can be seen by approximating U from the inside by basic open classes.

By linearity it will suffice to consider the case where $0 \leq f \leq 1$ everywhere. Fix $n \in \mathbb{N}$ and for $1 \leq i \leq n$ define

$$f_i = \min\left(\max\left(f - \frac{i-1}{n}, 0\right), \frac{1}{n}\right).$$

Thus, f_i is obtained by truncating f above at $\frac{i}{n}$ and below at $\frac{i-1}{n}$, and shifting the result down by $\frac{i-1}{n}$. Also define $U_i = f^{-1}((\frac{i}{n}, 1])$ for $0 \leq i \leq n$. Then $f = f_1 + \cdots + f_n$ and $\frac{1}{n}1_{U_{i-1}} \geq f_i \geq \frac{1}{n}1_{U_i}$ for each i. Thus $\mu[U_i] \leq n \cdot \phi(f_i)$ for each i, by the observation made in the last paragraph, and also $n \cdot \phi(f_i) \leq \mu[U_{i-1}]$ since $n \cdot f_i$ can be uniformly approximated by functions $f \prec U_{i-1}$. Summing over i yields

$$\frac{1}{n}\sum_{i=1}^{n} \mu[U_i] \leq \phi(f) \leq \frac{1}{n}\sum_{i=0}^{n-1} \mu[U_i].$$

But we also have $\frac{1}{n}\mu[U_i] \leq \int f_i\,d\mu \leq \frac{1}{n}\mu[U_{i-1}]$ for all i, and hence

$$\frac{1}{n}\sum_{i=1}^{n} \mu[U_i] \leq \int f\,d\mu \leq \frac{1}{n}\sum_{i=0}^{n-1} \mu[U_i].$$

We conclude that $|\phi(f) - \int f\,d\mu| \leq \frac{1}{n}\mu[U_0] - \frac{1}{n}\mu[U_n] \leq \frac{1}{n}\mu[X]$; since n was arbitrary, this yields that $\int f\,d\mu = \phi(f)$. □

15.5 Hilbert spaces

Definition 15.51. An *inner product* on an $\mathbb{F}$-vector space E is a map $\langle\cdot,\cdot\rangle : E \times E \to \mathbb{F}$ satisfying the conditions

(a) $\langle v, v\rangle \geq 0$, with equality if and only if $v = 0$
(b) $\langle w, v\rangle = \overline{\langle v, w\rangle}$
(c) $\langle av + bw, u\rangle = a\langle v, u\rangle + b\langle w, u\rangle$

for all $v, w, u \in E$ and $a, b \in \mathbb{F}$. We define $\|v\| = \langle v, v\rangle^{1/2}$.

We must show that this $\|\cdot\|$ is a norm. First we generalize the Cauchy-Schwarz inequality in $\mathbb{R}^n$ (Lemma 10.3).

Proposition 15.52. *(Cauchy-Schwarz inequality) Let v and w be elements of an $\mathbb{F}$-vector space equipped with an inner product. Then $|\langle v, w\rangle| \leq \|v\|\|w\|$.*

Proof. If $w = 0$ then $\langle v, w\rangle = 0$ and we are done. Otherwise, evaluating the inner product of the vector $\|w\|v - \frac{\langle v,w\rangle}{\|w\|}w$ with itself yields the result $\|w\|^2\|v\|^2 - |\langle v, w\rangle|^2$. Since this quantity must be greater than or equal to zero, we infer the stated inequality. □

Corollary 15.53. *Let v and w be elements of an $\mathbb{F}$-vector space equipped with an inner product. Then $\|v + w\| \leq \|v\| + \|w\|$.*

Proof. Using the proposition, we calculate:

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

Taking square roots yields the triangle inequality. □

Together with the easy observation that $\|av\| = |a| \cdot \|v\|$, this shows that $\|\cdot\|$ is a legitimate norm.

Definition 15.54. A *Hilbert space* is an $\mathbb{F}$-vector space equipped with an inner product for which the associated norm is complete.

Example 15.55. The Euclidean norm in $\mathbb{F}^n$ comes from the inner product $\langle \vec{v}, \vec{w}\rangle = \sum_{i=1}^n v_i\overline{w}_i$; the norm in l^2 comes from the inner product $\langle \vec{v}, \vec{w}\rangle = \sum_{i=1}^\infty v_i\overline{w}_i$; and the norm in $L^2(X)$ comes from the inner product $\langle f, g\rangle = \int f\overline{g}$. These are all Hilbert spaces.

Hilbert spaces are a special kind of Banach space. The key geometric concept that we have in Hilbert spaces is orthogonality.

Definition 15.56. Let H be a Hilbert space. We say that $v, w \in H$ are *orthogonal*, and write $v \perp w$, if $\langle v, w\rangle = 0$. For subspaces E and F of H we write $E \perp F$, and say that E and F are orthogonal, if $v \perp w$ for all $v \in E$ and $w \in F$. The *orthocomplement* of a separable subspace E is the class $E^\perp$ of all $w \in H$ such that $v \perp w$ for all $v \in E$.

Here $E^\perp$ exists because $v \perp w$ for all $v \in E$ if and only if $v \perp w$ for all v in a dense subset of E. It is a sequentially closed subspace. We will see below (Corollary 15.60) that if H is separable, then $E^\perp$ is also separable.

We say that vectors e_i ($1 \leq i \leq n$ or $i \in \mathbb{N}$) are *orthonormal* if $\|e_i\| = 1$ for all i and $e_i \perp e_j$ for all $i \neq j$. The e_i constitute an *orthonormal*

basis of H if they are orthonormal and their span is dense in H. (Despite the terminology, in general they will not constitute a basis in the sense of Definition 12.3. They will constitute a Schauder basis, however.)

Lemma 15.57. *Let e_1, ..., e_n be orthonormal vectors in a Hilbert space H and let $v \in H$. Then we can write $v = v_1 + v_2$ with*

$$v_1 = \langle v, e_1\rangle e_1 + \cdots + \langle v, e_n\rangle e_n$$

and $v_2 \in \operatorname{span}(e_1, \ldots, e_n)^\perp$.

Proof. Let v_1 be as stated and define $v_2 = v - v_1$. By orthonormality, for each i we have

$$\langle v_1, e_i\rangle = \langle\langle v, e_i\rangle e_i, e_i\rangle = \langle v, e_i\rangle.$$

Thus $\langle v_2, e_i\rangle = \langle v - v_1, e_i\rangle = 0$ for all i, so $v_2 \in \operatorname{span}(e_1, \ldots, e_n)^\perp$. □

Theorem 15.58. *Every separable Hilbert space has an orthonormal basis.*

Proof. Let H be a separable Hilbert space and let (v_n) be a finite or infinite sequence of linearly independent vectors whose span is dense in H. Such a sequence can be extracted from a dense subset by enumerating the subset and removing every vector which belongs to the span of the vectors that precede it in the enumeration.

Let $e_1 = \frac{1}{\|v_1\|}v_1$. Given e_1, ..., e_{n-1}, use the lemma to write $v_n = w_n + v'_n$ with $w_n \in \operatorname{span}(e_1, \ldots, e_{n-1})$ and $v'_n \in \operatorname{span}(e_1, \ldots, e_{n-1})^\perp$. Then v'_n must be nonzero because v_n is not a linear combination of v_1, ..., v_{n-1}, and we inductively have $\operatorname{span}(e_1, \ldots, e_{n-1}) = \operatorname{span}(v_1, \ldots, v_{n-1})$. So we can define $e_n = \frac{1}{\|v'_n\|}v'_n$ and proceed with the construction. This yields an orthonormal basis. □

Corollary 15.59. *Let H be a separable Hilbert space. If H is finite dimensional then it is isometrically isomorphic to $\mathbb{F}^n$ for $n = \dim(H)$; otherwise it is isometrically isomorphic to l^2.*

Proof. Suppose $\{f_1, \ldots, f_n\}$ is an orthonormal basis of H. Then the function $f_i \mapsto e_i$ extends linearly to an isometric isomorphism between H and $\mathbb{F}^n$, where $\{e_1, \ldots, e_n\}$ is the standard basis of $\mathbb{F}^n$. If H does not have a finite orthonormal basis then it has an infinite orthonormal basis $\{f_1, f_2, \ldots\}$ and the function $f_i \mapsto e_i$ extends linearly to an isometric isomorphism between a dense subspace of H and a dense subspace of l^2, where $\{e_1, e_2, \ldots\}$ is the standard orthonormal basis of l^2. This implies that $H \cong l^2$. □

Corollary 15.60. *Let E be a separable closed subspace of a separable Hilbert space H. Then $E^\perp$ is a separable closed subspace of H and $H = E + E^\perp$. Every vector in H can be uniquely decomposed as the sum of an element of E and an element of $E^\perp$, yielding maps $P_E : H \to E$ and $P_{E^\perp} : H \to E^\perp$ which are linear and nonexpansive.*

Proof. First, note that decompositions must be unique because if $v, v' \in E$, $w, w' \in E^\perp$, and $v + w = v' + w'$ then $v - v' = w' - w \in E \cap E^\perp = \{0\}$.

By Theorem 15.58 E has an orthonormal basis $\{e_1, \ldots, e_n\}$ or $\{e_1, e_2, \ldots\}$. In the former case, $H = E + E^\perp$ follows immediately from Lemma 15.57. In the latter case, let $v \in H$ and for each n define $a_n = \langle v, e_n\rangle$ and $v_n = a_1 e_1 + \cdots + a_n e_n$. Observe that $u \perp w$ implies $\|u + w\|^2 = \langle u + w, u + w\rangle = \|u\|^2 + \|w\|^2$; in particular, $\|u\| \le \|u + w\|$ whenever $u \perp w$. Then Lemma 15.57 implies that $\|v_n\| \le \|v\|$ for all n, so

$$|a_1|^2 + \cdots + |a_n|^2 = \|v_n\|^2 \le \|v\|^2,$$

and therefore the series $\sum a_n e_n$ converges in $E \cong l^2$.

Let $w = \sum a_n e_n$. Then $w \in E$, and $v - w = v - \lim v_n \in E^\perp$ since $e_n \perp v - v_m$ for all $m \ge n$. Thus we have expressed v as the sum of a vector in E and a vector in $E^\perp$. Define $P_E v = w$ and $P_{E^\perp} v = v - w$.

If $v, w \in H$ and $a, b \in \mathbb{F}$, write $v = v_1 + v_2$ and $w = w_1 + w_2$ with $v_1, w_1 \in E$ and $v_2, w_2 \in E^\perp$. Then $av + bw = (av_1 + bw_1) + (av_2 + bw_2)$ decomposes $av + bw$ as the sum of a vector in E and a vector in $E^\perp$, and this shows that P_E and $P_{E^\perp}$ are both linear. They are nonexpansive by the observation about $\|u\| \le \|u + w\|$ made above.

We still have to show that $E^\perp$ is separable. To do this, let $\mathring{H}$ be a dense subset of H; I claim that $P_{E^\perp}(\mathring{H})$ is dense in $E^\perp$. For if $w \in E^\perp$ and (w_n) is a sequence in $\mathring{H}$ that converges to w, then $\|w - P_{E^\perp} w_n\| \le \|w - P_{E^\perp} w_n - P_E w_n\| = \|w - w_n\|$ for all n, so $P_{E^\perp} w_n \to w$. □

We have already essentially proven the first part of the next corollary.

Corollary 15.61. *(Bessel's inequality) Let e_i ($1 \le i \le n$ or $i \in \mathbb{N}$) be orthonormal vectors in a Hilbert space H and let $v \in H$. Then*

$$\sum |\langle v, e_i\rangle|^2 \le \|v\|^2,$$

with equality if the e_i constitute an orthonormal basis. In that case we also have

$$\langle v, w\rangle = \sum \langle v, e_i\rangle\langle e_i, w\rangle$$

for all $w \in H$.

Proof. By working in the sequential closure of the span of v and the e_i's, we can assume H is separable. Let E be the closure of the span of the e_i's. By Corollary 15.60 plus Theorem 15.58, $E^\perp$ has an orthonormal basis, and combining this with the e_i's yields an orthonormal basis of H. As the corollary is clear in $\mathbb{F}^n$ and l^2, we conclude the proof by appealing to Corollary 15.59. □

Corollary 15.62. *(Riesz representation) Let H be a separable Hilbert space. Then every $w \in H$ gives rise to a bounded linear functional $\phi_w : v \mapsto \langle v, w\rangle$ with norm equal to $\|w\|$, and every bounded linear functional on H has this form. This describes a bijective isometry between H and H^*.*

Proof. It follows immediately from the definition of an inner product that the map $v \mapsto \langle v, w\rangle$ is a linear functional on H. It is bounded, with norm at most $\|w\|$, by the Cauchy-Schwarz inequality (Proposition 15.52), and taking $v = w$ shows that its norm is exactly $\|w\|$.

Now let $\phi : H \to \mathbb{F}$ be any bounded linear functional on H. If $\phi = 0$ then $\phi(v) = \langle v, 0\rangle$ for all v; otherwise ϕ is surjective, and its kernel is a separable closed subspace of H by Proposition 15.27. The orthocomplement of $\ker(\phi)$ is a one-dimensional subspace, say $\ker(\phi)^\perp = \mathbb{F} \cdot w_0$. We can assume $\|w_0\| = 1$.

Let $w = \overline{\phi(w_0)}w_0$. Then $\phi(v) = 0 = \langle v, w\rangle$ for all $v \in \ker(\phi)$, and $\phi(w) = |\phi(w_0)|^2 = \langle w, w\rangle$; since every element of H is the sum of something in $\ker(\phi)$ and a scalar multiple of w, this proves that $\phi(v) = \langle v, w\rangle$ for all $v \in H$. □

We have $\phi_{av+bw} = \overline{a}\phi_v + \overline{b}\phi_w$, so Corollary 15.62 describes an *antilinear* isometric isomorphism between H and H^*.

Definition 15.63. Let H be a separable Hilbert space and let $T \,\varepsilon\, B(H)$. Retaining the notation of the last corollary, for every $v \in H$ we have a bounded linear functional ϕ_v on H, so that $\phi_v \circ T$ is also a bounded linear functional, and by that corollary it has the form $\phi_v \circ T = \phi_{v'}$ for some $v' \in H$. We have $\|v'\| = \|\phi_{v'}\| \leq \|\phi_v\|\|T\| = \|v\|\|T\|$, so that the map $v \mapsto v'$ from H to itself is bounded and, using the comment just after the corollary, linear. Thus there exists $T^* \,\varepsilon\, B(H)$ with $T^*v = v'$ for all $v \in H$. This is the *adjoint* of T.

We call the elements of $B(H)$ *operators*. Here is a simpler characterization of the adjoint.

Proposition 15.64. *Let H be a separable Hilbert space and let $T \varepsilon B(H)$. Then $\langle Tv, w\rangle = \langle v, T^*w\rangle$ for all $v, w \in H$.*

Proof. Let $w \in H$ and let $w' = T^*w$, so that $\phi_w \circ T = \phi_{w'}$. Then for any $v \in H$ we have

$$\langle v, T^*w\rangle = \langle v, w'\rangle = \phi_{w'}(v) = \phi_w(Tv) = \langle Tv, w\rangle.$$

Thus $\langle Tv, w\rangle = \langle v, T^*w\rangle$. □

Define a product on $B(H)$ by setting $ST = S \circ T$. Also let $I \varepsilon B(H)$ be the *identity* operator with $Iv = v$ for all $v \in H$.

Definition 15.65. Let H be a separable Hilbert space and let $T \varepsilon B(H)$. Then T is *self-adjoint* if $T = T^*$, *unitary* if $T^*T = TT^* = I$, and a *projection* if $T = T^* = T^2$.

Proposition 15.66. *P_E is a projection, for any separable closed subspace E of a separable Hilbert space H. Every projection in $B(H)$ has this form.*

Proof. Let E be a separable closed subspace of a separable Hilbert space H. We have $P_E^2 = P_E$ because $P_E(v) \in E$ for all $v \in H$, and P_E fixes every element of E. For self-adjointness, calculate

$$\langle P_E v, w\rangle = \langle P_E v, P_E w + P_{E^\perp} w\rangle = \langle P_E v, P_E w\rangle,$$

and observe that similarly $\langle v, P_E w\rangle = \langle P_E v, P_E w\rangle$. This completes the proof that P_E is a projection.

Now let $P \varepsilon B(H)$ be any projection, let $\mathring{H}$ be a dense subset of H, and let E be the closure of $P(\mathring{H})$. Since P fixes every element of $P(\mathring{H})$, it follows that P fixes every element of E, and by continuity we have $Pv \in E$ for every $v \in H$. Thus the image of P exists and equals E. We have $Pv = v = P_E v$ for all $v \in E$, and for $v \in E^\perp$ we have $P_E v = 0$ and $\langle Pv, w\rangle = \langle v, Pw\rangle = 0$ for all $w \in H$, showing that $Pv = 0$. We conclude that $P = P_E$. □

Proposition 15.67. *The operator $U \varepsilon B(H)$ is unitary if and only if it is bijective and preserves inner products.*

Proof. Suppose U is unitary. Then $\langle Uv, Uw\rangle = \langle U^*Uv, w\rangle = \langle v, w\rangle$ for all $v, w \in H$, so U preserves inner products. This implies that it is injective. The condition $UU^* = I$ forces U to be surjective.

Conversely, suppose U is a bijection that preserves inner products. Then we have $\langle v, w\rangle = \langle Uv, Uw\rangle = \langle U^*Uv, w\rangle$ for all $v, w \in H$. This shows that $U^*Uv = v$ for all v, i.e., $U^*U = I$. Since U is invertible, U^{-1} must equal U^*, and therefore $UU^* = I$ holds as well. □

The preceding results give a pretty complete description of the structure of projections and unitaries. We can also analyze in detail the structure of self-adjoint operators; I will do this here only for the finite dimensional case with complex scalars, when $H \cong \mathbb{C}^n$.

Recall from the proof of Theorem 12.51 that an eigenvector of a linear transformation $T : \mathbb{C}^n \to \mathbb{C}^n$ is a nonzero vector $v \in \mathbb{C}^n$ for which $Tv = \lambda v$, for some eigenvalue $\lambda \in \mathbb{C}$, and that every operator on $\mathbb{C}^n$ has an eigenvector.

Theorem 15.68. *(Finite dimensional spectral theorem) Let $A \varepsilon B(\mathbb{C}^n)$ be self-adjoint. Then every eigenvalue of A is real, and $\mathbb{C}^n$ has an orthonormal basis consisting of eigenvectors for A.*

Proof. Let $v \in \mathbb{C}^n$ be a nonzero eigenvector with eigenvalue λ. Then

$$\lambda\langle v, v\rangle = \langle Av, v\rangle = \langle v, Av\rangle = \overline{\lambda}\langle v, v\rangle,$$

so $\lambda = \overline{\lambda}$ and hence λ must be real.

We prove the existence of an orthonormal basis of eigenvectors by induction on n. Suppose it is true for all smaller values of n. Let v be an eigenvector for A; we can assume $\|v\| = 1$. Let E be the orthocomplement of the one-dimensional subspace $\mathbb{C} \cdot v \subseteq \mathbb{C}^n$. Then for any $w \in E$ we have

$$\langle v, Aw\rangle = \langle Av, w\rangle = \lambda\langle v, w\rangle = 0,$$

so that A restricts to a linear transformation from $E \cong \mathbb{C}^{n-1}$ to itself. Also, we have $\langle Aw, w'\rangle = \langle w, Aw'\rangle$ for all $w, w' \in E$, showing that $A|_E$ is self-adjoint. By the induction hypothesis, E has an orthonormal basis consisting of eigenvectors for $A|_E$, and together with v this yields an orthonormal basis of $\mathbb{C}^n$ consisting of eigenvectors for A. □

15.6 Fourier analysis

We work in $\mathbb{R}^n$ but take the scalar field to be $\mathbb{C}$. *Normalized Lebesgue measure* in $\mathbb{R}^n$ is the measure $\tilde{m} = \frac{1}{(2\pi)^{n/2}} m$ where m is ordinary Lebesgue measure. Throughout this section all L^p spaces are complex and are defined relative to normalized Lebesgue measure, i.e., $L^1(\mathbb{R}^n) = L^1(\mathbb{R}^n, \tilde{m})$, etc.

We define the *Fourier transform* $\mathcal{F}(f) = \hat{f}$ of $f \varepsilon L^1(\mathbb{R}^n)$ by

$$\hat{f}(\vec{t}) = \int f(\vec{x}) e^{-i\langle \vec{t}, \vec{x}\rangle}\, d\tilde{m}(\vec{x})$$

for all $\vec{t} \in \mathbb{R}^n$. The integral exists because f is integrable and $e^{-i\langle \vec{t}, \vec{x}\rangle}$ is bounded.

The *convolution* $f * g$ of $f, g \,\varepsilon\, L^1(\mathbb{R}^n)$ is defined by

$$(f * g)(\vec{x}\,) = \int_{\mathbb{R}^n} f(\vec{x} - \vec{y})g(\vec{y}\,)\, d\tilde{m}(\vec{y}).$$

The integral is well-defined for almost every $\vec{x}$ because

$$\begin{aligned}\int\int |f(\vec{x} - \vec{y})||g(\vec{y})|\, d\tilde{m}(\vec{y})d\tilde{m}(\vec{x}\,) &= \int\int |f(\vec{x} - \vec{y})||g(\vec{y})|\, d\tilde{m}(\vec{x}\,)d\tilde{m}(\vec{y}\,)\\ &= \|f\|_1 \int |g(\vec{y})|\, d\tilde{m}(\vec{y}\,)\\ &= \|f\|_1\|g\|_1 < \infty\end{aligned}$$

by Tonelli's theorem (Theorem 13.43). Indeed, this calculation shows not only that $f * g$ is defined almost everywhere, but also that it belongs to $L^1(\mathbb{R}^n)$ with $\|f * g\|_1 \le \|f\|_1\|g\|_1$.

For $\vec{y} \in \mathbb{R}^n$ we define the *translation* of $f \,\varepsilon\, L^1(\mathbb{R}^n)$ by $\tau_{\vec{y}} f(\vec{x}\,) = f(\vec{x} - \vec{y}\,)$.

Proposition 15.69. *Let* $f, g \,\varepsilon\, L^1(\mathbb{R}^n)$ *and* $\vec{y} \in \mathbb{R}^n$. *Then* $(\tau_{\vec{y}} f)\hat{}\,(\vec{t}\,) = e^{-i\langle \vec{t}, \vec{y}\,\rangle}\hat{f}(\vec{t}\,)$ *and* $(f * g)\hat{} = \hat{f}\hat{g}$.

The proof is a straightforward calculation (using Fubini's theorem, Theorem 13.44, in the second part).

A function $f : \mathbb{R}^n \to \mathbb{C}$ is *rapidly decreasing* if $x_i^k f(\vec{x}\,) \to 0$ as $\vec{x} \to \infty$, for every $1 \le i \le n$ and $k \in \mathbb{N}$. That is, for every $\epsilon > 0$ there is a ball about the origin outside of which $|x_i^k f(\vec{x}\,)| < \epsilon$. The *Schwartz space* $\mathcal{S}_n$ consists of all smooth complex-valued functions on $\mathbb{R}^n$ all of whose partial derivatives, of all orders, are rapidly decreasing. It is a linear subspace of $C(\mathbb{R}^n) = C(\mathbb{R}^n, \mathbb{C})$.

We need to check that the condition "$f \,\varepsilon\, C(\mathbb{R}^n)$ is smooth" is definite, in order for $\mathcal{S}_n$ to exist as a subclass of $C(\mathbb{R}^n)$. It will suffice to check the definiteness of the condition "$\frac{\partial f}{\partial x_i}$ exists and is continuous" for a function f on the cube $[0, 1]^n$. That is because if each partial derivative exists and is continuous, then we can check whether each of their partial derivatives exist and are continuous, and so on. Since $\mathbb{R}^n$ is a countable union of cubes, with a countable quantification we can determine whether f is smooth.

To see this, without loss of generality let $i = 1$ and for each rational $\vec{a} \in [0, 1]^{n-1}$ determine whether the function $f(t, \vec{a}\,)$ is differentiable at all rational values of t in $(0, 1)$. The definiteness of this condition was noted at the beginning of Section 13.1. Then, if the function is differentiable at all rational t and $\vec{a}$, check whether these derivatives extend to a continuous function g on $[0, 1]^n$ (Theorem 10.48). If they do, then for each rational t

and $\vec{a}$ check whether $f(t,\vec{a}\,) = f(0,\vec{a}\,)+\int_0^t g(s,\vec{a}\,)\,ds$. This condition implies that $f(t,\vec{a}\,)$ is continuously differentiable for all rational $\vec{a}$ (the fundamental theorem of calculus, Theorem 13.40), and then $\frac{\partial f}{\partial x_1}$ exists and is continuous on $[0,1]^n$ by the dominated convergence theorem (Theorem 13.39). Thus, for $\frac{\partial f}{\partial x_1}$ to exist and be continuous on $[0,1]^n$ it is necessary and sufficent that f be differentiable in the x_1 direction at all rational t and $\vec{a}$, that these derivatives extend continuously to $[0,1]^n$, and that f and g satisfy the equality stated above. And these conditions are definite.

Lemma 15.70. *Let $f\,\varepsilon\,\mathcal{S}_n$ and let $1 \le j \le n$. Then $(x_j f)\hat{} = i\frac{\partial}{\partial t_j}\hat{f}$ and $\left(\frac{\partial f}{\partial x_j}\right)\hat{} = it_j\hat{f}$. The Fourier transform takes $\mathcal{S}_n$ into itself.*

Proof. It is clear that $x_j f$ and $\frac{\partial}{\partial x_j}f$ both belong to $\mathcal{S}_n$, and hence to $L^1(\mathbb{R}^n)$. By Fubini's theorem (Theorem 13.44) and Theorem 13.41 plus Proposition 13.1, we have

$$\begin{aligned} 0 &= \int \frac{\partial}{\partial x_j}\left(f(\vec{x})e^{-i\langle\vec{t},\vec{x}\,\rangle}\right)\,d\tilde{m}(\vec{x}\,) \\ &= \int \frac{\partial f}{\partial x_j}(\vec{x})e^{-i\langle\vec{t},\vec{x}\,\rangle}\,d\tilde{m}(\vec{x}\,) - \int it_j f(\vec{x})e^{-i\langle\vec{t},\vec{x}\,\rangle}\,d\tilde{m}(\vec{x}\,) \\ &= \left(\frac{\partial f}{\partial x_j}\right)\hat{}\,(\vec{t}) - it_j\hat{f}(\vec{t}). \end{aligned}$$

This shows that $\left(\frac{\partial f}{\partial x_j}\right)\hat{} = it_j\hat{f}$.

Next, by the dominated convergence theorem (Theorem 13.39)

$$\begin{aligned} \frac{\hat{f}(\vec{t}+he_j)-\hat{f}(\vec{t}\,)}{h} &= \frac{1}{h}\int f(\vec{x})(e^{-i\langle\vec{t},\vec{x}\,\rangle-ihx_j} - e^{-i\langle\vec{t},\vec{x}\,\rangle})\,d\tilde{m}(\vec{x}\,) \\ &= \int \frac{e^{-ihx_j}-1}{h}f(\vec{x})e^{-i\langle\vec{t},\vec{x}\,\rangle}\,d\tilde{m}(\vec{x}\,) \\ &\to -i\int x_j f(\vec{x})e^{-i\langle\vec{t},\vec{x}\,\rangle}\,d\tilde{m}(\vec{x}\,) \\ &= -i(x_j f)\hat{}\,(\vec{t}\,) \end{aligned}$$

as $h \to 0$. This shows that $\frac{\partial}{\partial t_j}\hat{f}$ exists everywhere and equals $-i(x_j f)\hat{}$.

Repeated application of these two facts yields that $\hat{f}$ is smooth and all of its partial derivatives, of all orders, are rapidly decreasing. □

Lemma 15.71. *$\mathcal{S}_n$ is dense in $L^1(\mathbb{R}^n)$.*

Proof. The function

$$\xi(t) = \begin{cases} 1 & \text{if } t \leq 1 \\ \frac{e^{-1/(2-t)}}{e^{-1/(2-t)}+e^{-1/(t-1)}} & \text{if } 1 < t < 2 \\ 0 & \text{if } t \geq 2 \end{cases}$$

is smooth, constantly one on $[0,1]$ and constantly zero on $[2,\infty)$. Thus the functions $\xi(|x_1|)^k \cdots \xi(|x_n|)^k$ belong to $\mathcal{S}_n$ and (by the dominated convergence theorem, Theorem 13.39) converge in 1-norm to $1_{[-1,1]^n}$. Since $\mathcal{S}_n$ is invariant under scaling and translation in each coordinate, it follows that the indicator function of any rectangle $[a_1,b_1]\times\cdots\times[a_n,b_n]$ lies in its closure in $L^1(\mathbb{R}^n)$. By Theorem 13.39 again, the indicator function of any finitely measurable class in $\mathbb{R}^n$ belongs to that closure, which then implies that that closure is all of $L^1(\mathbb{R}^n)$. □

Theorem 15.72. *(Riemann-Lebesgue lemma) If $\check{f}\,\varepsilon\, L^1(\mathbb{R}^n)$ then $\hat{f}\,\varepsilon\, C_0(\mathbb{R}^n)$ and $\|\hat{f}\|_\infty \leq \|f\|_1$.*

Proof. The inequality $\|\hat{f}\|_\infty \leq \|f\|_1$ is trivial since $|e^{-i\langle \vec{t},\vec{x}\,\rangle}| = 1$. Now let f be integrable and find a sequence (f_k) in $\mathcal{S}_n$ that converges to f in 1-norm, as per Lemma 15.71. Then $\hat{f}_k\,\varepsilon\,\mathcal{S}_n \subset C_0(\mathbb{R}^n)$ by Lemma 15.70 and $\hat{f}_k \to \hat{f}$ in sup norm by the norm inequality. Thus $\hat{f}\,\varepsilon\, C_0(\mathbb{R}^n)$. □

Lemma 15.73. *The function $\phi(\vec{x}\,) = e^{-\frac{1}{2}|\vec{x}\,|^2}$ is a fixed point of $\mathcal{F}$.*

Proof. We know from Lemma 15.70 that $\hat{\phi}\,\varepsilon\,\mathcal{S}_n$. In the $n=1$ case, ϕ is a solution of the differential equation $\phi' + x\phi = 0$; applying $\mathcal{F}$ to this equation and using Lemma 15.70 yields that $\hat{\phi}$ is a solution of the same equation. Also

$$\hat{\phi}(0) = \int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2}\, d\tilde{m}(x) = 1 = \phi(0),$$

so the same initial condition is satisfied and the uniqueness part of Picard's existence theorem (Theorem 13.62 yields) $\hat{\phi} = \phi$.

For general n, we can write $\phi(\vec{x}\,) = e^{-\frac{1}{2}|x_1|^2}\cdots e^{-\frac{1}{2}|x_n|^2}$ and reduce to the one-dimensional case. □

Theorem 15.74. *(Fourier inversion) We have*

$$g(\vec{x}\,) = \int \hat{g}(\vec{t}\,) e^{i\langle \vec{x},\vec{t}\,\rangle}\, d\tilde{m}(\vec{t}\,)$$

for all $g\,\varepsilon\,\mathcal{S}_n$. The Fourier transform takes $\mathcal{S}_n$ onto itself.

Proof. Let $g \varepsilon \mathcal{S}_n$. For any $f \varepsilon \mathcal{S}_n$ we can use Fubini's theorem (Theorem 13.44) to perform the integral

$$\int \int f(\vec{x})g(\vec{y})e^{-i\langle \vec{x},\vec{y}\rangle}\, d\tilde{m}(\vec{x})d\tilde{m}(\vec{y})$$

in either order; this shows that

$$\int \hat{f}g\, d\tilde{m} = \int f\hat{g}\, d\tilde{m}.$$

Applying this to the function $f_k(\vec{x}) = \phi(\vec{x}/k)$, where ϕ is as in Lemma 15.73, we have $\hat{f}_k(\vec{t}) = k^n\hat{\phi}(k\vec{t})$ and

$$\begin{aligned}\int f_k(\vec{y})\hat{g}(\vec{y})\, d\tilde{m}(\vec{y}) &= \int g(\vec{t})k^n\hat{\phi}(k\vec{t})\, d\tilde{m}(\vec{t}) \\ &= \int g(\vec{t}/k)\hat{\phi}(\vec{t})\, d\tilde{m}(\vec{t}) \\ &= \int g(\vec{t}/k)\phi(\vec{t})\, d\tilde{m}(\vec{t}).\end{aligned}$$

Taking $k \to \infty$ and applying the dominated convergence theorem (Theorem 13.39), plus the fact that $\hat{\phi}(0) = 1$, yields

$$\int \hat{g}(\vec{y})\, d\tilde{m}(\vec{y}) = g(0),$$

the $\vec{x} = 0$ case of the inversion formula. Replacing g with $\tau_{\vec{x}}g$ yields the formula for arbitrary $\vec{x}$.

Replacing $\vec{x}$ with $-\vec{x}$ in the inversion formula shows that $g(-\vec{x})$ is the Fourier transform of $\hat{g}$, which belongs to $\mathcal{S}_n$ by Lemma 15.70. Since every function in $\mathcal{S}_n$ can be written as $g(-\vec{x})$ with $g \varepsilon \mathcal{S}_n$, this shows that $\mathcal{F}$ takes $\mathcal{S}_n$ onto itself. □

Theorem 15.75. *(Plancharel's theorem) We have*

$$\int f\overline{g}\, d\tilde{m} = \int \hat{f}\overline{\hat{g}}\, d\tilde{m}$$

for all $f, g \varepsilon \mathcal{S}_n$. The Fourier transform extends from $\mathcal{S}_n \subset L^2(\mathbb{R}^n)$ to a unitary operator on $B(L^2(\mathbb{R}^n))$.

Proof. Let $f, g \varepsilon \mathcal{S}_n$. By the Fourier inversion theorem (Theorem 15.74)

$$\begin{aligned}\int f(\vec{x})\overline{g}(\vec{x})\, d\tilde{m}(\vec{x}) &= \int \int \overline{g}(\vec{x}) \cdot \hat{f}(\vec{t})e^{i\langle \vec{x},\vec{t}\rangle}\, d\tilde{m}(\vec{t})d\tilde{m}(\vec{x}) \\ &= \int \int \hat{f}(\vec{t}) \cdot \overline{g}(\vec{x})e^{i\langle \vec{x},\vec{t}\rangle}\, d\tilde{m}(\vec{x})d\tilde{m}(\vec{t})\end{aligned}$$

$$= \int \hat{f}(\vec{t})\overline{\hat{g}}(\vec{t})\, d\tilde{m}(\vec{t}).$$

This proves the first statement. Taking $g = f$ yields $\|f\|_2 = \|\hat{f}\|_2$, so $\mathcal{F}|_{\mathcal{S}_n}$ implements an isometric map from a dense subclass of $L^2(\mathbb{R}^n)$ into itself, indeed, onto itself by Theorem 15.74. Thus it extends to an isometric linear transformation U from $L^2(\mathbb{R}^n)$ onto itself. Since $\langle \check{f}, \check{g}\rangle = \langle U\check{f}, U\check{g}\rangle$ for all $\check{f}, \check{g} \varepsilon\, \check{\mathcal{S}}_n$, the same formula holds for all $\check{f}, \check{g} \varepsilon\, L^2(\mathbb{R}^n)$, and this shows that U is unitary. □

Chapter 16

Geometry

16.1 Smooth manifolds

In Chapter 14 I introduced manifolds, locally compact topological spaces with the property that every point has an open neighborhood that is homeomorphic to an open subclass of $\mathbb{R}^n$, for some fixed n. Let $\mathbb{H}^n \subset \mathbb{R}^n$ be the *half-space*

$$\mathbb{H}^n = \{\vec{x} \in \mathbb{R}^n : x_1 \geq 0\};$$

a *manifold with boundary*, or *n-manifold with boundary* if we want to make the dimension explicit, is a locally compact topological space M for which every point has an open neighborhood that is homeomorphic to a (relatively) open subclass of $\mathbb{H}^n$. The *boundary* of M, denoted ∂M, consists of the points in M which are taken by the homeomorphism into the subspace $\{\vec{x} \in \mathbb{R}^n : x_n = 0\}$. The *interior* of M is the class $M \setminus \partial M$.

I defined smoothness (= infinite differentiability) of functions from an open subclass of $\mathbb{R}^n$ into $\mathbb{R}^m$ in Section 13.5. We can slightly generalize that definition to say that a function ϕ from an arbitrary subclass $A \subseteq \mathbb{R}^n$ into $\mathbb{R}^m$ is *smooth* if there is an open class U containing A and a function from U to $\mathbb{R}^m$ which is smooth in the previous sense and whose restriction to A equals ϕ.

Definition 16.1. Let M be an n-manifold with boundary. A *chart* on M is a homeomorphism ϕ from an open subclass U of M to an open subclass of $\mathbb{H}^n$. Two charts $\phi : U \to \mathbb{H}^n$ and $\psi : V \to \mathbb{H}^n$ are *compatible* if $\psi \circ \phi^{-1}$ (defined on $\phi(U \cap V)$) and $\phi \circ \psi^{-1}$ (defined on $\psi(U \cap V)$) are both smooth. Open subclasses $U_\lambda \subseteq M$ and charts $\phi_\lambda : U_\lambda \to \mathbb{H}^n$ ($1 \leq \lambda \leq k$ or $\lambda \in \mathbb{N}$) constitute an *atlas* on M if ϕ_λ and ϕ_κ are compatible for all λ and κ and every element of M lies in at least one, but only finitely many, of the U_λ's.

It follows from the inverse function theorem (Theorem 13.9) that if $x \in U_\lambda \cap U_\kappa$, then either both $\phi_\lambda(x)$ and $\phi_\kappa(x)$ lie in the boundary of $\mathbb{H}^n$ or neither does, so our definition of ∂M is self-consistent. (If the transition functions are merely continuous, not smooth, this statement is still true, and can be proven using the techniques developed in Section 14.5.)

Given an atlas on M, I will say that a chart $\phi : U \to \mathbb{H}^n$ is *smooth* if it is compatible with every ϕ_λ. We consider two atlases to be equivalent if every chart from one of them is smooth relative to the other. A *smooth manifold (with boundary)* is a manifold (with boundary) equipped with an atlas, but any equivalent atlas is considered to represent the same smooth structure.

Definition 16.2. Let M be a smooth n-manifold with boundary, with an atlas consisting of charts $\phi_\lambda : U_\lambda \to \mathbb{H}^n$. Let $x \in M$. A *tangent vector* at x is a sequence $v = (\vec{v}_\lambda)$ of vectors in $\mathbb{R}^n$, over the same range of values of λ, with the properties that (1) $\vec{v}_\lambda = 0$ if $x \notin U_\lambda$ and (2) if $x \in U_\lambda \cap U_\kappa$ then $\vec{v}_\kappa = D(\phi_\kappa \circ \phi_\lambda^{-1})|_{\phi_\lambda(x)} \vec{v}_\lambda$. The *tangent space* at x is the n-dimensional vector space $T_x M$ consisting of all tangent vectors at x.

Informally, we use a chart to model an open subclass U of M by an open subclass of $\mathbb{H}^n$, and then model tangent vectors at elements of U by vectors in $\mathbb{R}^n$. The only constraint is a simple consistency condition which relates any two models that both apply to the same point.

The definition of tangent vectors still makes sense if x lies in the boundary of M, because we require that $\phi_\kappa \circ \phi_\lambda^{-1}$ extend to a smooth function on a neighborhood of $\phi_\lambda(x)$ in $\mathbb{R}^n$, so this extension has a derivative at $\phi_\lambda(x)$, but the derivative $D(\phi_\kappa \circ \phi_\lambda^{-1})|_{\phi_\lambda(x)} : \mathbb{R}^n \to \mathbb{R}^n$ is a linear transformation and hence is already determined by the behavior of $\phi_\kappa \circ \phi_\lambda^{-1}$ in a neighborhood of $\phi_\lambda(x)$ in $\mathbb{H}^n$.

A defect of Definition 16.2 is that it apparently depends on a choice of atlas. Let us identify an equivalent definition that does not exhibit this dependence.

A function $f : M \to \mathbb{R}$ is *smooth* if for each λ the function $f \circ \phi_\lambda^{-1} : \phi_\lambda(U_\lambda) \to \mathbb{R}$ is smooth. It is clear from the definition of compatibility of charts that smoothness of f does not depend on the choice of atlas. If v is a tangent vector at $x \in M$, then the *directional derivative* $df|_x v$ of a smooth function f at x in the direction v is the value

$$df|_x v = D(f \circ \phi_\lambda^{-1})|_{\phi_\lambda(x)} \vec{v}_\lambda,$$

for any λ with $x \in U_\lambda$. This definition is consistent because if $x \in U_\lambda \cap U_\kappa$

then $f \circ \phi_\lambda^{-1} = f \circ \phi_\kappa^{-1} \circ \phi_\kappa \circ \phi_\lambda^{-1}$ in a neighborhood of $\phi_\lambda(x)$, and so

$$\begin{aligned} D(f \circ \phi_\lambda^{-1})|_{\phi_\lambda(x)} \vec{v}_\lambda &= D(f \circ \phi_\kappa^{-1})|_{\phi_\kappa(x)} \cdot D(\phi_\kappa \circ \phi_\lambda^{-1})|_{\phi_\lambda(x)} \vec{v}_\lambda \\ &= D(f \circ \phi_\kappa^{-1})|_{\phi_\kappa(x)} \vec{v}_\kappa \end{aligned}$$

(the multivariable chain rule, Proposition 13.7).

Directional derivatives can be used to give a "coordinate free" definition of tangent vectors.

Definition 16.3. Let M be a smooth n-manifold with boundary. We define $C^\infty(M)$ to be the class of smooth functions from M into $\mathbb{R}$. It is a linear subspace of the real vector space $C(M) = C(M, \mathbb{R})$ of continuous real-valued functions (with operations defined pointwise).

A linear transformation $\mathcal{D} : C^\infty(M) \to \mathbb{R}$ is a *point derivation* at $x \in M$ if it satisfies the Leibniz rule

$$\mathcal{D}(fg) = f(x) \cdot \mathcal{D}g + \mathcal{D}f \cdot g(x)$$

for all $f, g \,\varepsilon\, C^\infty(M)$.

$C^\infty(M)$ is a legitimate subclass of $C(M)$ because every open class $\phi_\lambda(U_\lambda)$ in $\mathbb{H}^n$ is a countable union of cubes in $\mathbb{R}^n$, and as we saw in Section 15.6, the condition that a function be smooth on a cube is definite.

Every directional derivative at $x \in M$ is a point derivation at x, and conversely, we will show that for every point derivation $\mathcal{D}$ at x there is a tangent vector v at x such that $\mathcal{D}f = df|_x v$ for all $f \,\varepsilon\, C^\infty(M)$. As it is easy (using Lemma 16.4 below) to find, for any distinct tangent vectors v and v' at x, a smooth function whose directional derivatives in the v and v' directions are different, it follows that point derivations at x exactly correspond to tangent vectors at x. Thus the notion "point derivation at x" is an atlas-independent interpretation of "tangent vector at x".

First, let us note that smooth "bump functions" exist. Recall the function ξ from Lemma 15.71.

Lemma 16.4. *For any $n \in \mathbb{N}$ the function $\xi(\|\vec{x}\|)$ on $\mathbb{R}^n$ is smooth, constantly one on $B_1(0)$, and constantly zero outside $B_2(0)$.*

We also need the following variation on Taylor's theorem.

Lemma 16.5. *Let $f : C \to \mathbb{R}$ be a smooth function defined on a closed, convex subclass of $\mathbb{R}^n$ which contains the origin. Then there exist smooth functions $g_1, \ldots, g_n : C \to \mathbb{R}$ with $g_i(0) = \frac{\partial f}{\partial x_i}(0)$ and $g_i(\vec{x}) = 0$ for $\vec{x} \notin B_2(0) \cap C$ $(1 \leq i \leq n)$ and such that*

$$f(\vec{x}) - f(0) = x_1 g_1(\vec{x}) + \cdots + x_n g_n(\vec{x})$$

on $B_1(0) \cap C$.

Proof. We have (treating $f(t\vec{x})$ as a function of one variable, t)

$$f(\vec{x}) - f(0) = \int_0^1 f'(t\vec{x})\,dt = \int_0^1 \left(\sum x_i \cdot \frac{\partial f}{\partial x_i}(t\vec{x})\right) dt$$

for all $\vec{x} \in B_2(0) \cap C$ (the multivariable chain rule, Proposition 13.7). So we can take

$$g_i(\vec{x}) = \xi(\|\vec{x}\|) \cdot \int_0^1 \frac{\partial f}{\partial x_i}(t\vec{x})\,dt$$

for $1 \leq i \leq n$. These are evidently smooth, and we have $f(\vec{x}) - f(0) = \sum x_i g_i(\vec{x})$ on $B_1(0) \cap C$, where $\xi(\|\vec{x}\|)$ is constantly one. □

Lemma 16.6. *Let M be a smooth manifold with boundary and let $\mathcal{D} : C^\infty(M) \to \mathbb{R}$ be a point derivation at $x \in M$. Then (1) $\mathcal{D}(1_M) = 0$ and (2) if $f_1, f_2 \,\varepsilon\, C^\infty(M)$ agree in a neighborhood of x then $\mathcal{D}(f_1) = \mathcal{D}(f_2)$.*

Proof. We have

$$\mathcal{D}(1_M) = \mathcal{D}(1_M^2) = 1_M(x)\mathcal{D}(1_M) + \mathcal{D}(1_M) \cdot 1_M(x) = 2\mathcal{D}(1_M),$$

so $\mathcal{D}(1_M) = 0$.

For the second part, let $h = f_1 - f_2$ and suppose h is zero in a neighborhood U of x; we must show that $\mathcal{D}h = 0$. To do this, use Lemma 16.4 to find $g \,\varepsilon\, C^\infty(M)$ which is 0 in a neighborhood $V \subset U$ of x and 1 outside U. Then $h = gh$, so

$$\mathcal{D}h = \mathcal{D}gh = g(x) \cdot \mathcal{D}h + \mathcal{D}g \cdot h(x) = 0$$

since $g(x) = h(x) = 0$. □

Theorem 16.7. *Let M be a smooth manifold with boundary and let $\mathcal{D} : C^\infty(M) \to \mathbb{R}$ be a point derivation at $y \in M$. Then there exists $v \in T_yM$ such that $\mathcal{D}f = df|_y v$ for all $f \,\varepsilon\, C^\infty(M)$.*

Proof. Let $\phi : U \to \mathbb{H}^n$ be a smooth chart with $y \in U$. Without loss of generality we can assume $\phi(U)$ contains $\overline{B}_3(\phi(y)) \cap \mathbb{H}^n$. Define $h(x) = \phi(x) - \phi(y)$ and write $h(x) = (h_1(x), \ldots, h_n(x))$. Then for any $f \,\varepsilon\, C^\infty(M)$ apply Lemma 16.5 to $(\overline{B}_3(\phi(y)) \cap \mathbb{H}^n) - \phi(y)$ to find $g_i \,\varepsilon\, C^\infty(M)$, $1 \leq i \leq n$, satisfying $f(x) - f(y) = h_1(x)g_1(x) + \cdots + h_n(x)g_n(x)$ in a neighborhood of y and $g_i(y) = d(f \circ \phi^{-1})|_{\phi(y)} e_i$ for all i. Note that the g_i from Lemma 16.5 smoothly extend to all of M by letting them vanish on U^c. According to Lemma 16.6, $\mathcal{D}$ vanishes on every constant function, in particular on the function which takes the constant value $f(y)$, and also $\mathcal{D}$ takes the same

value on $f - f(y)$ and $g_1h_1 + \cdots + g_nh_n$ since these two functions agree in a neighborhood of y. So

$$\begin{aligned}\mathcal{D}f &= \mathcal{D}(g_1h_1 + \cdots + g_nh_n)\\ &= \sum_{i=1}^{n}(g_i(y)\mathcal{D}h_i + \mathcal{D}g_i \cdot h_i(y))\\ &= \sum_{i=1}^{n} g_i(y)\mathcal{D}h_i = df|_y v\end{aligned}$$

where v is the tangent vector whose representation in the chart $\phi : U \to \mathbb{H}^n$ is the vector $(\mathcal{D}h_1, \ldots, \mathcal{D}h_n)$. We have shown that $\mathcal{D}f = df|_y v$ for all $f \,\varepsilon\, C^\infty(M)$. □

The notion of the derivative at $x \in \mathbb{R}^m$ of a smooth function from $\mathbb{R}^m$ to $\mathbb{R}^n$ can be generalized to maps between arbitrary smooth manifolds.

Definition 16.8. Let M and N be smooth m- and n-manifolds with boundary and let $\phi_\lambda : U_\lambda \to \mathbb{H}^m$ and $\psi_\kappa : V_\kappa \to \mathbb{H}^n$ be charts that constitute their atlases. A continuous function $f : M \to N$ is *smooth* if for all λ and κ the function $\psi_\kappa \circ f \circ \phi_\lambda^{-1}$ from $\phi_\lambda(U_\lambda \cap f^{-1}(V_\kappa)) \subseteq \mathbb{R}^m$ to $\mathbb{R}^n$ is smooth. The *derivative of f at x* is the linear transformation $Df|_x : T_xM \to T_{f(x)}N$ whose value at $v \in T_xM$ is the vector $w \in T_{f(x)}N$ with

$$\vec{w}_\kappa = D(\psi_\kappa \circ f \circ \phi_\lambda^{-1})|_{\phi_\lambda(x)}\vec{v}_\lambda$$

for any λ and κ such that $x \in U_\lambda$ and $f(x) \in V_\kappa$.

A smooth function $f : M \to N$ is a *smooth immersion* if $Df|_x$ has rank m for all $x \in M$, and a *smooth embedding* if it is also a homeomorphism between M and $f(M)$. A homeomorphism f between M and N for which both f and f^{-1} are smooth is a *diffeomorphism*.

It is clear that this definition of smoothness does not depend on the choice of atlas. The definition of $Df|_x$ is consistent by essentially the same reasoning that showed the consistency of our definition of the directional derivative. An equivalent atlas-independent interpretation of $Df|_x v$ is: the point derivation $h \mapsto d(h \circ f)|_x v$ (for $h \,\varepsilon\, C^\infty(N)$) at $f(x)$.

The following variation on the inverse function theorem (Theorem 13.9) is basic.

Theorem 16.9. *Let M and N be smooth n-manifolds, let $x \in M$, and let $f : M \to N$ be smooth. Suppose $Df|_x$ is invertible. Then there exist open classes $U \subseteq M$ and $V \subseteq N$ with $x \in U$ such that $f|_U$ is a bijection between U and V whose inverse $g : V \to U$ is also smooth.*

Since this is a local result, we can work in a single chart and reduce to maps from open subsets of $\mathbb{R}^n$ to $\mathbb{R}^n$, and we can assume that Df is invertible throughout U. (Note that M and N are assumed to have no boundaries.) The new fact here is that g is infinitely differentiable. We know from the inverse function theorem that g is differentiable at $f(\vec{y})$ with derivative $Df|_{\vec{y}}^{-1}$ for all $\vec{y} \in U$, but the map that takes a matrix to its inverse is smooth on the class of invertible matrices (this follows from the explicit expression of its entries in terms of determinants, as a ratio of two polynomials), so infinite differentiability of f implies infinite differentiability of the matrix entries of $Df|_{\vec{y}}^{-1}$ and hence of g.

Let us close this section with a basic tool for studying smooth manifolds, plus one easy application of it.

Lemma 16.10. *Let M be a smooth n-manifold with boundary, with an atlas consisting of charts $\phi_\lambda : U_\lambda \to \mathbb{H}^n$. Then there exist smooth charts $\psi_\kappa : V_\kappa \to \mathbb{H}^n$ and elements $\vec{a}_\kappa \in \psi_\kappa(V_\kappa)$ with the properties (1) each V_κ is contained in some U_λ; (2) every element of M belongs to at least one, but only finitely many, V_κ's; (3) $\psi_\kappa(V_\kappa) = B_3(\vec{a}_\kappa) \cap \mathbb{H}^n$; and (4) the classes $\psi_\kappa^{-1}(B_1(\vec{a}_\kappa) \cap \mathbb{H}^n)$ cover M.*

Proof. Let $P_1, P_2, \ldots$ constitute a basis for the topology on M for which $\overline{P}_i$ is compact for all i. Define $K_1 = \overline{P}_1$, and inductively define $K_{i+1} = \overline{P}_1 \cup \cdots \cup \overline{P}_k$ where k is the smallest number greater than i such that $K_i \subseteq P_1 \cup \cdots \cup P_k$. This is possible because K_i is compact.

Let $K_i^{int} = K_i \setminus (\overline{K_i^c})$. ($M$ is hereditarily separable by Corollary 10.42, so the closure of K_i^c exists.) Then put $B_{1,\lambda} = K_3^{int} \cap U_\lambda$ for all λ and $B_{k,\lambda} = (K_{k+2}^{int} \setminus K_{k-1}) \cap U_\lambda$ for $k \geq 2$ and all λ. For every $\vec{b} \in \phi_\lambda(B_{k,\lambda})$ with all rational coordinates, let $V_{\vec{b},\lambda} = \phi_\lambda^{-1}(B_r(\vec{b}))$ where r is the distance from $\mathbb{H}^n \setminus \phi_\lambda(B_{k,\lambda})$ to $\vec{b}$. Also let $\vec{a}_{\vec{b},\lambda} = \frac{3}{r}\vec{b}$ and define $\psi_{\vec{b},\lambda} : V_{\vec{b},\lambda} \to B_3(\vec{a}_{\vec{b},\lambda})$ by

$$\psi_{\vec{b},\lambda}(y) = \frac{3}{r} \cdot \phi_\lambda(y).$$

The classes $\psi_{\vec{b},\lambda}^{-1}(B_1(\vec{a}_{\vec{b},\lambda}))$ cover each $K_{k+1} \setminus K_k^{int}$, which is compact, so we can find finitely many of them that cover $K_{k+1} \setminus K_k^{int}$. Making such a choice of finitely many $V_{\vec{b},\lambda}$'s that cover $K_{k+1} \setminus K_k^{int}$ for each k yields countably many $V_{\vec{b},\lambda}$'s that satisfy the desired conditions. □

As in the last chapter, the support of a continuous function $f : X \to \mathbb{R}$ on a second countable space X, denoted $\mathrm{supp}(f)$, is the closure of $f^{-1}(\mathbb{R} \setminus \{0\})$.

Theorem 16.11. *(Partition of unity) Let M be a smooth n-manifold with boundary, with an atlas consisting of charts $\phi_\lambda : U_\lambda \to \mathbb{R}^n$. Then there are smooth functions $f_\kappa : M \to [0,1]$ such that for every $x \in M$ we have $\sum f_\kappa(x) = 1$, with only finitely many nonzero terms in each sum, and such that for every κ the support of f_κ is contained in some U_λ.*

Proof. Let $\psi_\kappa : V_\kappa \to \mathbb{H}^n$ and $\vec{a}_\kappa$ be as in the lemma, and for each κ let $g_\kappa(y) = \xi(\|\psi_\kappa(y) - \vec{a}_\kappa\|)$ on V_κ, where ξ is the bump function from Lemma 15.71. Put $g_\kappa = 0$ outside of V_κ. Then the functions

$$f_\kappa = \frac{g_\kappa}{\sum_{\mu=1}^{\infty} g_\mu}$$

have the desired properties. □

If the support of every f_κ is contained in some U_λ, as in the theorem, then we say that the f_κ's are *subordinate* to the U_λ's.

In the next result we will work with maps into $\mathbb{R}^{<\omega}$, which is not a manifold in our sense. However, if $f : M \to \mathbb{R}^{<\omega}$ is continuous then every $x \in M$ has a neighborhood U with compact closure such that $f(U) \subseteq \mathbb{R}^k \subset \mathbb{R}^{<\omega}$ for some $k \in \mathbb{N}$ (Proposition 10.86 (d)). So there is a clear meaning to the statement that f is a smooth embedding.

Corollary 16.12. *(Baby Whitney embedding) Every smooth manifold with boundary smoothly embeds in $\mathbb{R}^{<\omega}$. Every compact smooth manifold with boundary smoothly embeds in some $\mathbb{R}^k$.*

Proof. Fix an atlas consisting of charts $\phi_\lambda : U_\lambda \to \mathbb{H}^n$ on a smooth n-manifold M with boundary and let f_κ be as in partition of unity (Theorem 16.11). For each κ let λ_κ be an index with $\operatorname{supp}(f_\kappa) \subseteq U_{\lambda_\kappa}$. Then define $F : M \to \mathbb{R}^\omega$ by

$$\begin{aligned} F(x) &= (f_1(x), f_1(x) \cdot \phi_{\lambda_1}(x), f_2(x), f_2(x) \cdot \phi_{\lambda_2}(x), \ldots) \\ &\in (\mathbb{R} \times \mathbb{R}^n) \times (\mathbb{R} \times \mathbb{R}^n) \times \cdots \cong \mathbb{R}^\omega \end{aligned}$$

(taking $f_\kappa \cdot \phi_{\lambda_\kappa}$ to be zero outside of U_{λ_κ}). The image of F is actually contained in $\mathbb{R}^{<\omega}$ since for every $x \in M$ we have $f_\kappa(x) \neq 0$ for only finitely many values of κ. Smoothness of F is clear, and F is injective because for every $x \in M$ we have $f_\kappa(x) \neq 0$ for some value of κ, and then $f_\kappa(y) \neq f_\kappa(x)$ immediately implies $F(x) \neq F(y)$, while $f_\kappa(y) = f_\kappa(x) \neq 0$ implies that $f_\kappa(x) \cdot \phi_{\lambda_\kappa}(x) \neq f_\kappa(y) \cdot \phi_{\lambda_\kappa}(y)$ if $x \neq y$. Finally, F is a homeomorphism by Propositions 10.90 and 10.86 (c) plus Theorem 10.44.

If M is compact, then by Proposition 10.86 (d) we get a smooth embedding in $\mathbb{R}^k \subset \mathbb{R}^{<\omega}$ for some $k \in \mathbb{N}$. □

In fact every smooth manifold smoothly embeds in $\mathbb{R}^k$ for some $k \in \mathbb{N}$; this is the full Whitney embedding theorem. It is substantially harder to prove.

16.2 Differential forms

We need some preliminaries about finite dimensional real vector spaces.

Let $\vec{y} \in \mathbb{R}^n$. Then the map $\vec{x} \mapsto \langle \vec{x}, \vec{y} \rangle$ is a linear functional on $\mathbb{R}^n$, and every linear functional has this form. Thus any $\vec{y} \in \mathbb{R}^n$ can be thought of as an n-dimensional real vector, or alternatively, as representing a linear functional on $\mathbb{R}^n$. When we are thinking of it in the latter way I will use the notation $\breve{y}$ and refer to it as a *covector*, and I will write the space of covectors as $(\mathbb{R}^n)^*$. This is the *dual space* of $\mathbb{R}^n$. I will write $\breve{y}|\vec{x}$ or $\breve{y}(\vec{x})$ for $\breve{y}$, as a linear functional, applied to $\vec{x}$.

To clarify the meaning of "thinking of it as a linear functional", let A be an invertible linear map from $\mathbb{R}^n$ to itself. Then for all $\breve{y} \in (\mathbb{R}^n)^*$ the map $\vec{x} \mapsto \breve{y}(A^{-1}\vec{x})$ is a linear functional on $\mathbb{R}^n$ and so determines a unique covector $\breve{A}\breve{y} = (A^{-1})^*\breve{y}$ (see Definition 15.63) with the property that $\breve{A}\breve{y}|A\vec{x} = \breve{y}|\vec{x}$ for all $\vec{x} \in \mathbb{R}^n$. Thus, when vectors transform according to A, covectors transform by their action on vectors via $\breve{A}$, so that their pairing is unchanged.

If M is a smooth manifold with boundary, we define the *covector space* at $x \in M$ to be the dual space T_x^*M of T_xM. If M has an atlas consisting of charts $\phi_\lambda : U_\lambda \to \mathbb{H}^n$, then an element of T_x^*M can be represented as a sequence $(\breve{w}_\lambda)$, over the same values of λ, with each $\breve{w}_\lambda \in (\mathbb{R}^n)^*$, $\breve{w}_\lambda = 0$ if $x \notin U_\lambda$, and $\breve{w}_\kappa = \breve{A}\breve{w}_\lambda$, where $A = D(\phi_\kappa \circ \phi_\lambda^{-1})|_{\phi_\lambda(x)}$, whenever $x \in U_\lambda \cap U_\kappa$. This has the effect of ensuring that $\breve{w}_\lambda|\vec{v}_\lambda = \breve{w}_\kappa|\vec{v}_\kappa$ for any $v \in T_xM$ and $w \in T_x^*M$, and all λ and κ with $x \in U_\lambda \cap U_\kappa$.

The dual space construction can be applied to $(\mathbb{R}^n)^*$, and we find that the linear functionals on $(\mathbb{R}^n)^*$ are precisely the maps of the form $\breve{y} \mapsto \breve{y}|\vec{x}$ for $\vec{x} \in \mathbb{R}^n$. Thus the dual space of $(\mathbb{R}^n)^*$ is naturally identified with $\mathbb{R}^n$.

The representation of linear functionals can be generalized to *multilinear functionals* on products of the form

$$(\mathbb{R}^n)^* \times \cdots \times (\mathbb{R}^n)^* \times \mathbb{R}^n \times \cdots \times \mathbb{R}^n,$$

where there are, say, p copies of $(\mathbb{R}^n)^*$ and q copies of $\mathbb{R}^n$. These are maps into $\mathbb{R}$ which are linear on each factor, when all the other coordinates are held constant. Now consider elements of the product of the form

$(\tilde{e}_{i_1}, \ldots, \tilde{e}_{i_p}, e_{j_1}, \ldots, e_{j_q})$ where $\{e_1, \ldots, e_n\}$ is the standard basis for $\mathbb{R}^n$ and $\{\tilde{e}_1, \ldots, \tilde{e}_n\}$ is the standard basis for $(\mathbb{R}^n)^*$. If T is a multilinear functional, then by linearity in the first entry its values on these elements determine its value on any element of the form $(\breve{y}^1, \tilde{e}_{i_2}, \ldots, \tilde{e}_{i_p}, e_{j_1}, \ldots, e_{j_q})$. But then linearity in the second entry shows that this determines the value on any element of the form $(\breve{y}^1, \breve{y}^2, \tilde{e}_{i_3}, \ldots, \tilde{e}_{i_p}, e_{j_1}, \ldots, e_{j_q})$, and so on. Carrying this out to the end shows that T is completely determined by its values on elements of the form $(\tilde{e}_{i_1}, \ldots, \tilde{e}_{i_p}, e_{j_1}, \ldots, e_{j_q})$; conversely, any choice of values for these elements determines a multilinear map. This means that the multilinear maps from $(\mathbb{R}^n)^* \times \cdots \times (\mathbb{R}^n)^* \times \mathbb{R}^n \times \cdots \times \mathbb{R}^n$, into $\mathbb{R}$ can be represented by the elements of an n^{p+q}-dimensional real vector space. This is a *tensor product*, specifically the *space of* (p, q)*-tensors* over $\mathbb{R}^n$, and is denoted $\mathbb{R}^n \otimes \cdots \otimes \mathbb{R}^n \otimes (\mathbb{R}^n)^* \otimes \cdots \otimes (\mathbb{R}^n)^*$. A natural basis for the space of (p, q)-tensors consists of elements of the form $e_{i_1} \otimes \cdots \otimes e_{i_p} \otimes \tilde{e}_{j_1} \otimes \cdots \otimes \tilde{e}_{j_q}$, which take $(\breve{y}^1, \ldots, \breve{y}^p, \vec{x}^1, \ldots, \vec{x}^q)$ to

$$(\breve{y}^1 | e_{i_1}) \cdots (\breve{y}^p | e_{i_p}) \cdot (\tilde{e}_{j_1} | \vec{x}^1) \cdots (\tilde{e}_{j_q} | \vec{x}^q) = y^1_{i_1} \cdots y^p_{i_p} \cdot x^1_{j_1} \cdots x^q_{j_q}.$$

A (p, q)-tensor of the form $\vec{x}^1 \otimes \cdots \otimes \vec{x}^p \otimes \breve{y}^1 \otimes \cdots \otimes \breve{y}^q$ is *elementary*. Every (p, q)-tensor is a linear combination of elementary tensors with each $\vec{x}^i$ belonging to the standard basis of $\mathbb{R}^n$ and each $\vec{y}^j$ belonging to the standard basis of $(\mathbb{R}^n)^*$.

Any invertible linear map A from $\mathbb{R}^n$ to itself induces a linear map $A^{(p,q)}$ from $\mathbb{R}^n \otimes \cdots \otimes \mathbb{R}^n \otimes (\mathbb{R}^n)^* \otimes \cdots \otimes (\mathbb{R}^n)^*$ to itself which is defined on elementary tensors by

$$A^{(p,q)}(\vec{x}^1 \otimes \cdots \otimes \vec{x}^p \otimes \breve{y}^1 \otimes \cdots \otimes \breve{y}^q) = (A\vec{x}^1 \otimes \cdots \otimes A\vec{x}^p \otimes \breve{A}\breve{y}^1 \otimes \cdots \otimes \breve{A}\breve{y}^q).$$

If M is a smooth manifold with boundary, with an atlas consisting of charts $\phi_\lambda : U_\lambda \to \mathbb{H}^n$, then a (p, q)*-tensor at* $x \in M$ is an element of $T_xM \otimes \cdots \otimes T_xM \otimes T^*_xM \otimes \cdots \otimes T^*_xM$, with p factors of T_xM and q factors of T^*_xM. Relative to an atlas we can take it it be a sequence $\alpha = (\alpha_\lambda)$ with $\alpha_\lambda \in \mathbb{R}^n \otimes \cdots \otimes \mathbb{R}^n \otimes (\mathbb{R}^n)^* \otimes \cdots \otimes (\mathbb{R}^n)^*$ such that $\alpha_\lambda = 0$ if $x \notin U_\lambda$ and $\alpha_\kappa = A^{(p,q)}\alpha_\lambda$ where $A = D(\phi_\kappa \circ \phi_\lambda^{-1})|_{\phi_\lambda(x)}$ whenever $x \in U_\lambda \cap U_\kappa$. The class of all (p, q)-tensors at x constitute the (p, q)*-tensor space at* x, which we denote $T^{(p,q)}_x M$.

This definition reduces to our previous definitions of tangent and cotangent vectors at x when $p = 1$ and $q = 0$, or $p = 0$ and $q = 1$, respectively. Informally, we again use a chart to model an open subclass U of M by an open subclass of $\mathbb{H}^n$, and then model (p, q)-tensors at elements of U by (p, q)-tensors over $\mathbb{R}^n$. As before, the only constraint is a simple consistency condition which relates any two models that both apply to the same point.

The *tangent bundle* over M, relative to a choice of atlas, is the class $TM = \{(x, v) : x \in M \text{ and } v \in T_xM\}$, regarded as a smooth $2n$-manifold with charts given by the open classes $TU_\lambda = \{(x, v) \in TM : x \in U_\lambda \text{ and } v \in T_xU_\lambda\}$ mapping into $\mathbb{H}^n \times \mathbb{R}^n = \mathbb{H}^{2n}$ via $(x, v) \mapsto (\phi_\lambda(x), \vec{v}_\lambda)$. Technically, TM is a subclass of $M \times (\mathbb{R}^n)^{<\omega}$. Likewise, the *cotangent bundle* over M is the smooth $2n$-manifold $T^*M = \{(x, w) : x \in M \text{ and } w \in T^*_xM\}$, with charts that map T^*U_λ into $\mathbb{H}^n \times (\mathbb{R}^n)^* = \mathbb{H}^{2n}$ via $(x, w) \mapsto (\phi_\lambda(x), \vec{w}_\lambda)$. In general, the (p, q)-*tensor bundle* over M is the smooth manifold

$$T^{(p,q)}M = \{(x, \alpha) : x \in M \text{ and } \alpha \in T^{(p,q)}_x M\}$$

with charts given by the open classes $T^{(p,q)}U_\lambda$ mapping into $\mathbb{H}^n \times (\mathbb{R}^n \otimes \cdots \otimes \mathbb{R}^n \otimes (\mathbb{R}^n)^* \otimes \cdots \otimes (\mathbb{R}^n)^*)$ via $(x, \alpha) \mapsto (\phi_\lambda(x), \alpha_\lambda)$.

We have a smooth map $\pi : T^{(p,q)}M \to M$ given by $\pi(x, \alpha) = x$. A *smooth* (p, q)-*tensor field* on M is a smooth function $\mathcal{T} : M \to T^{(p,q)}M$ satisfying $\pi(\mathcal{T}|_x) = x$ for all $x \in M$. Here I write $\mathcal{T}|_x$ for the value of $\mathcal{T}$ at x. Thus, $\mathcal{T}$ smoothly selects a (p, q)-tensor at each point of M. When $p = 1$ and $q = 0$, this is a *smooth vector field*, a smooth function $X : M \to TM$ satisfying $\pi(X|_x) = x$ for all $x \in M$; when $p = 0$ and $q = 1$, it is a *smooth covector field* $\omega : M \to T^*M$. Again, $X|_x$ and $\omega|_x$ are the values of X and ω at x. The space of smooth vector fields on M is denoted $\mathfrak{X}M$, and the space of smooth covector fields is denoted ΛM.

In any chart $\phi_\lambda : U_\lambda \to \mathbb{H}^n$ we can model a (p, q)-tensor field $\mathcal{T}$ as a smooth function from U_λ into $\mathbb{R}^n \otimes \cdots \otimes \mathbb{R}^n \otimes (\mathbb{R}^n)^* \otimes \cdots \otimes (\mathbb{R}^n)^*$, and then we can project onto each of the n^{p+q} coordinates to decompose $\mathcal{T}$ on U_λ as

$$\sum f_{i_1,\ldots,i_p,j_1,\ldots,j_q} \cdot e_{i_1} \otimes \cdots \otimes e_{i_p} \otimes \tilde{e}_{j_1} \otimes \cdots \otimes \tilde{e}_{j_q},$$

taking the sum over all $(p + q)$-tuples of indices i_1, …, i_p, j_1, …, j_q, with each $f_{i_1,\ldots,i_p,j_1,\ldots,j_q} \varepsilon\, C^\infty(U_\lambda)$.

Tensor fields with $p = 0$ are *covariant*, and tensor fields with $q = 0$ are *contravariant*. If $\mathcal{T}$ is a covariant $(0, q)$-tensor field and X_1, …, X_q are vector fields then at each $x \in M$ the tensor $\mathcal{T}|_x$ is a multilinear functional on $T_xM \times \cdots \times T_xM$, and therefore can be evaluated on $(X_1|_x, \ldots, X_q|_x)$. In local coordinates, if $x \in U_\lambda$, for each $1 \leq k \leq q$ we can write $X_k = \sum g^k_i \cdot e_i$ on U_λ with each $g^k_i \varepsilon\, C^\infty(U_\lambda)$, and then decompose $\mathcal{T}$ as above and evaluate $\mathcal{T}$ on $(X_1, \ldots, X_q)$ by

$$\mathcal{T}(X_1, \ldots, X_q) = \sum f_{j_1,\ldots,j_q} \cdot g^1_{j_1} \cdots g^q_{j_q} \varepsilon\, C^\infty(U_\lambda).$$

It would be simpler to just say $\mathcal{T}(X_1, \ldots, X_q)(x) = \mathcal{T}|_x(X_1|_x, \ldots, X_q|_x)$ for all $x \in M$, but writing the result in local coordinates is helpful because

it shows that the result is smooth. That is, we can regard $\mathcal{T}$ as a function from $\mathfrak{X}M \times \cdots \times \mathfrak{X}M$ to $C^\infty(M)$.

An important difference between covariant and contravariant tensor fields is that the former can be "pulled back" along a smooth function $f : M \to N$, while the latter cannot. Note that a tangent vector v at a single point $x \in M$ can be "pushed forward" to the vector $Df|_x v$, but if $f(x) = f(y)$ for some distinct $x, y \in M$ this can produce a conflict when we try to push forward tangent vectors at both points. In contrast, if $\mathcal{T}$ is a $(0,q)$-tensor field on N we can define a $(0,q)$-tensor field $f^*\mathcal{T}$ on M by putting, at each $x \in M$,

$$(f^*\mathcal{T})|_x(v^1, \ldots, v^q) = \mathcal{T}|_{f(x)}(Df|_x v^1, \ldots, Df|_x v^q)$$

where v^1, ..., v^q are tangent vectors at x.

Proposition 16.13. *Let $f : M \to N$ be a smooth map between smooth manifolds and let $\mathcal{T}$ be a smooth $(0,q)$-tensor field on N. Then $f^*\mathcal{T}$ is a smooth $(0,q)$-tensor field on M.*

Multilinearity is easy, and smoothness of $f^*\mathcal{T}$ can be seen by working in local coordinates, where it is clear that $f^*\mathcal{T}$ is smooth if and only if $f^*\mathcal{T}(X_1, \ldots, X_q)$ is smooth for all smooth vector fields X_1, ..., X_q. That is, we need $f^*\mathcal{T}|_x(X_1|_x, \ldots, X_q|_x) = \mathcal{T}|_{f(x)}(Df|_x X_1|_x, \ldots, Df|_x X_q|_x)$ to be a smooth function of x. By decomposing $\mathcal{T}$ as $\sum f_{j_1,\ldots,j_p} \cdot \tilde{e}_{j_1} \otimes \cdots \otimes \tilde{e}_{j_q}$ and each X_i as something of the form $\sum g_k e_k$ with each $f_{j_1,\ldots,j_p}, g_k \, \varepsilon \, C^\infty(M)$, this reduces to the observation that the matrix entries of $Df|_x$ are smooth functions of x.

A little more about $(0,q)$-tensors, where we are working with $(\mathbb{R}^n)^* \otimes \cdots \otimes (\mathbb{R}^n)^*$. We can apply any permutation σ of $\{1, \ldots, q\}$ to the $(0,q)$-tensors by applying the inverse permutation to $\mathbb{R}^n \times \cdots \times \mathbb{R}^n$ before evaluating the $(0,q)$-tensor; this is an isomorphism of the space of $(0,q)$-tensors with itself. Another way to say this is that σ applied to an elementary tensor $\tilde{y}^1 \otimes \cdots \otimes \tilde{y}^q$ is the elementary tensor $\tilde{y}^{\sigma(1)} \otimes \cdots \otimes \tilde{y}^{\sigma(q)}$. A $(0,q)$-tensor is *symmetric* if it is invariant under all permutations of $\{1, \ldots, q\}$. Given any elementary tensor $\tilde{y}^1 \otimes \cdots \otimes \tilde{y}^q$, if we apply all possible permutations, sum up the result, and normalize, we get a symmetric tensor. This extends to a linear map, the *symmetrizing map* $\mathcal{S}$, from the $(0,q)$-tensors to the symmetric $(0,q)$-tensors. Thus

$$\mathcal{S}(\tilde{y}^1 \otimes \cdots \otimes \tilde{y}^q) = \frac{1}{q!} \sum \tilde{y}^{\sigma(1)} \otimes \cdots \otimes \tilde{y}^{\sigma(q)},$$

taking the sum over all permutations σ of $\{1,\dots,q\}$. When $\mathcal{S}$ is applied to the standard basis consisting of elementary tensors of the form $\tilde{e}_{j_1}\otimes\cdots\otimes\tilde{e}_{j_q}$ we get a basis for the space of symmetric tensors (with repetition, because any permutation of an elementary tensor of this form gives rise to the same basis element).

A $(0,q)$-tensor is *antisymmetric* if permutations with sign 1 leave it fixed, while permutations with sign -1 multiply it by -1. Here we have an *antisymmetrizing map* $\mathcal{A}$ given by

$$\mathcal{A}(\overline{y}^1\otimes\cdots\otimes\overline{y}^q)=\frac{1}{q!}\sum\operatorname{sign}(\sigma)\cdot\overline{y}^{\sigma(1)}\otimes\cdots\otimes\overline{y}^{\sigma(q)},$$

again taking the sum over all permutations σ of $\{1,\dots,q\}$. In this case we define the *wedge product* of an antisymmetric $(0,q_1)$-tensor α_1 with an antisymmetric $(0,q_2)$-tensor α_2 to be antisymmetric $(0,q_1+q_2)$-tensor $\alpha_1\wedge\alpha_2=\mathcal{A}(\alpha_1\otimes\alpha_2)$. A basis of the space of antisymmetric $(0,q)$-tensors may be obtained by applying $\mathcal{A}$ to all elementary tensors of the form $\tilde{e}_{j_1}\otimes\cdots\otimes\tilde{e}_{j_q}$ with j_1, …, j_q distinct. (When they are not distinct, the antisymmetrizing map sends the tensor to zero.) Thus we have one element of the desired basis for the space of antisymmetric tensors corresponding to each choice of q distinct indices between 1 and n.

For $1\le q\le n$ a *q-form* on M is a smooth $(0,q)$-tensor field which is antisymmetric at each point. The space of all q-forms is denoted $\Lambda^qM\subseteq C(M,T^{(0,q)}M)$. If ω_1 is a q_1-form and ω_2 is a q_2-form, then we define their wedge product to be the (q_1+q_2)-form $\omega_1\wedge\omega_2$ whose value at $x\in M$ is $\omega_1|_x\wedge\omega_2|_x$.[1] The wedge product takes $\Lambda^{q_1}M\times\Lambda^{q_2}M$ to $\Lambda^{q_1+q_2}M$. Also, the pullback operation described in Proposition 16.13 takes q-forms on N to q-forms on M.

Λ^1M is just the space ΛM of smooth covector fields. It is also convenient to define $\Lambda^0M=C^\infty(M)$ and $g\wedge\omega=g\omega$ (the pointwise product) when g is a 0-form and ω is a q-form. If $f:M\to N$ is smooth then we define the pullback f^*g to be $g\circ f$, for $g\,\varepsilon\,C^\infty(N)$.

If $f\,\varepsilon\,C^\infty(M)=\Lambda^0M$, we have a directional derivative $df|_x$ at each point of M. This is a linear functional on T_xM, so it describes a covector at each point, and by working in local coordinates one easily sees that these covectors vary smoothly, i.e., they constitute a 1-form df. This gives a linear map $d:\Lambda^0M\to\Lambda^1M$. In local coordinates, we can write $df=\frac{\partial f}{\partial x_1}\tilde{e}_1+\cdots+\frac{\partial f}{\partial x_n}\tilde{e}_n$, or, letting x_j be the jth coordinate function on $\mathbb{H}^n$, so that dx_j is the covector field which takes the constant value $\tilde{e}_j$, as $df=$

[1] Old MacDonald had a form, $e_i\wedge e_i=0$.

$\frac{\partial f}{\partial x_1}dx_1 + \cdots + \frac{\partial f}{\partial x_n}dx_n$. The map can be extended to higher order forms in the following way.

Proposition 16.14. *Let M be a smooth n-manifold with boundary. Then for each $1 \leq q \leq n$ there is a linear map $d = d_M : \Lambda^{q-1}M \to \Lambda^q M$ with the properties*

(1) if $f \,\varepsilon\, \Lambda^0 M$ then df is the 1-form whose value at x is the directional derivative $df|_x$
(2) if ω_1 and ω_2 are q_1- and q_2-forms ($0 \leq q_1, q_2 \leq n$) then $d(\omega_1 \wedge \omega_2) = (d\omega_1) \wedge \omega_2 + (-1)^{q_1}\omega_1 \wedge (d\omega_2)$
(3) $d^2 = 0$.

These conditions uniquely determine d. If $U \subseteq M$ is open then $(d_M\omega)|_U = d_U(\omega|_U)$ for any q-form ω on M.

Proof. First, let V be an open subclass of $\mathbb{H}^n$. Then every q-form on V can be uniquely written as a sum of q-forms $a dx_{i_1} \wedge \cdots \wedge dx_{i_q}$ with $i_1 < \cdots < i_q$ and $a \,\varepsilon\, C^\infty(V)$. Conditions (2) and (3) entail that $d(dx_{i_1} \wedge \cdots \wedge dx_{i_q}) = 0$, and then by (1) and (2) we have no choice but to define $d(a dx_{i_1} \wedge \cdots \wedge dx_{i_q}) = da \wedge dx_{i_1} \wedge \cdots \wedge dx_{i_q}$ and extend linearly. It is straightforward to verify that the three conditions are satisfied by this definition.

Now, passing to M, for any $\omega \,\varepsilon\, \Lambda^q M$ and any coordinate chart $\phi_\lambda : U_\lambda \to \mathbb{H}^n$, we define $d_\lambda(\omega|_{U_\lambda})$ as above. We must show that this definition is consistent on $U_\lambda \cap U_\kappa$. To see this, let $x \in U_\lambda \cap U_\kappa$ and find $g \,\varepsilon\, C^\infty(M)$ which is constantly 1 in a neighborhood of x and constantly zero outside of $U_\lambda \cap U_\kappa$. Then $d_\lambda(g\omega) = d_\lambda(\omega)$ and $d_\kappa(g\omega) = d_\kappa(\omega)$ in a neighborhood of x, and, as above, the given conditions uniquely determine d_λ and d_κ on all q-forms $g\omega'$ with $\omega' \,\varepsilon\, \Lambda^q(U_\lambda \cap U_\kappa)$, so we must have $d_\lambda(g\omega) = d_\kappa(g\omega)$. Thus $d_\lambda(\omega) = d_\kappa(\omega)$ in a neighborhood of any point in $U_\lambda \cap U_\kappa$, and this shows that we can consistently define $d\omega|_{U_\lambda} = d_\lambda\omega$ on every chart in the atlas.

Uniqueness of d almost follows from the uniqueness of the d_λ, but we need to show that any d satisfying the three conditions is locally determined. To see this, suppose ω is constantly zero in a neighborhood of some point x and find $g \,\varepsilon\, C^\infty(M)$ supported on that neighborhood with $g(x) = 1$. Then $g\omega = 0$ so $0 = d(g\omega) = dg \wedge \omega + g \cdot d\omega = g \cdot d\omega$. This shows that $d\omega$ must be zero at x, which is enough. The final claim that $(d_M\omega)|_U = d_U(\omega|_U)$ is clear. □

The map $d : \Lambda^{q-1}M \to \Lambda^q M$ is the *exterior derivative* on M. It is compatible with pullbacks:

Proposition 16.15. *Let M and N be smooth manifolds. Then $f^* \circ d_N = d_M \circ f^*$, as maps from $\Lambda^q N$ to $\Lambda^{q+1}M$, for any smooth $f : M \to N$.*

Proof. Let $\phi : U \to \mathbb{H}^m$ and $\psi : V \to \mathbb{H}^n$ be charts on M and N, respectively, and let $U' = U \cap f^{-1}(V)$. By the last part of Proposition 16.14 it will suffice to show that $f^* \circ d_V = d_{U'} \circ f^*$ for any smooth $f : U' \to V$. For the sake of notational simplicity, assume $U' \subseteq \mathbb{H}^m$ and $V \subseteq \mathbb{H}^n$, with ϕ and ψ the identity maps.

For any 0-form $\omega \,\varepsilon\, C^\infty(V)$ we have $d_V\omega = \frac{\partial \omega}{\partial y_1} dy_1 + \cdots + \frac{\partial \omega}{\partial y_n} dy_n$ where $y_1, \ldots, y_n$ are the coordinate functions on V. So for any tangent vector v at $x \in U'$ we have

$$(f^*(d_V\omega))|_x v = (d_V\omega)|_{f(x)}(Df|_x v) = D(\omega \circ f)|_x v = d_{U'}(f^*\omega)|_x v,$$

which shows that $f^*(d_V\omega) = d_{U'}(f^*\omega)$. This proves the proposition when $q = 0$.

For 1-forms, by linearity it will suffice to consider the case $\omega = g dy_i$ for some $g \,\varepsilon\, C^\infty(V)$ and $1 \leq i \leq n$. Here

$$f^*(d\omega) = f^*(dg \wedge dy_i) = f^*(dg) \wedge f^*(dy_i) = d(f^*g) \wedge d(f^*y_i)$$

and

$$\begin{aligned} d(f^*\omega) &= d((f^*g) \cdot (f^*dy_i)) = f^*(dg) \wedge f^*(dy_i) + f^*g \cdot f^*(d(dy_i)) \\ &= f^*(dg) \wedge f^*(dy_i) = d(f^*g) \wedge d(f^*y_i), \end{aligned}$$

so the proposition is true for $q = 1$.

Now let ω be a q-form on V, for some $q \geq 2$, and suppose the result is true for all forms of lower degree. By linearity, it will suffice to consider the case where $\omega = g \cdot dy_{i_1} \wedge \cdots \wedge dy_{i_q}$. Let $\omega_1 = g \cdot dy_{i_1}$ and $\omega_2 = dy_{i_2} \wedge \cdots \wedge dy_{i_q}$; then

$$\begin{aligned} d(f^*(\omega)) &= d(f^*(\omega_1 \wedge \omega_2)) \\ &= d(f^*\omega_1 \wedge f^*\omega_2) \\ &= (df^*\omega_1) \wedge (f^*\omega_2) - (f^*\omega_1) \wedge (df^*\omega_2) \\ &= f^*(d\omega_1) \wedge (f^*\omega_2) - (f^*\omega_1) \wedge f^*(d\omega_2) \\ &= f^*(d\omega_1 \wedge \omega_2 - \omega_1 \wedge d\omega_2) = f^*(d\omega), \end{aligned}$$

as desired. □

Let X be a smooth vector field on a smooth manifold with boundary M and let $f \varepsilon C^\infty(M)$. We have a 1-form df which can be applied to X. Then $df(X) \varepsilon C^\infty(M)$, so this yields a linear map $f \mapsto df(X)$ from $C^\infty(M)$ into itself, which we also denote X. If Y is another smooth vector field and $x \in M$ then we define $[X,Y]_x(f) = (X(Y(f)) - Y(X(f)))(x)$. This is a map from $C^\infty(M)$ to $\mathbb{R}$, and it is a point derivation at x:

Lemma 16.16. *Let M be a smooth manifold with boundary, let $X, Y \varepsilon \mathfrak{X}M$, let $f, g \varepsilon C^\infty(M)$, and let $a, b \in \mathbb{R}$. Then*

(a) $[X,Y]_x(af + bg) = a[X,Y]_x(f) + b[X,Y](g)$
(b) $[X,Y]_x(fg) = [X,Y]_x(f) \cdot g(x) + f(x) \cdot [X,Y]_x(g)$.

The proof is a straightforward computation. Thus $[X,Y]_x$ can be identified with a tangent vector at x, and we can define $[X,Y]$ as a function from M into TM. We call $[X,Y]$ a *Lie bracket*.

Proposition 16.17. *Let M be a smooth manifold with boundary and let $X, Y \varepsilon \mathfrak{X}M$. Then $[X,Y]$ is a smooth vector field.*

The proof is easy. Working in a chart $\phi : U \to \mathbb{H}^n$, it suffices to check that $X(Y(f)) - Y(X(f))$ is smooth for every $f \varepsilon C^\infty(U)$. But $X(f) = df(X)$ is the pairing of a 1-form with a smooth vector field, and we already know the result is smooth. Repeated application of this fact yields that $X(Y(f)) - Y(X(f))$ is smooth.

Here are some more basic properties of the Lie bracket.

Proposition 16.18. *Let M be a smooth manifold with boundary, let $X, Y, Z \varepsilon \mathfrak{X}M$, and let $a, b \in \mathbb{R}$. Then*

(a) $[aX + bY, Z] = a[X,Z] + b[Y,Z]$
(b) $[Y,X] = -[X,Y]$
(c) $[X,[Y,Z]] + [Z,[X,Y]] + [Y,[Z,X]] = 0$
(d) $[fX,Y] = f \cdot [X,Y] - Y(f) \cdot X$.

Again, these are simple computations.

16.3 Integration on Manifolds

Every antisymmetric $(0,n)$-tensor over $\mathbb{R}^n$ has the form

$$a\tilde{e}_1 \wedge \cdots \wedge \tilde{e}_n$$

with $a \in \mathbb{R}$. In particular, for any $\vec{x}^1, \ldots, \vec{x}^n \in (\mathbb{R}^n)^*$ we have

$$\vec{x}^1 \wedge \cdots \wedge \vec{x}^n = \det([T]) \cdot \tilde{e}_1 \wedge \cdots \wedge \tilde{e}_n$$

where $T : (\mathbb{R}^n)^* \to (\mathbb{R}^n)^*$ is the linear transformation that takes $\tilde{e}_j$ to $\vec{x}^j$, $1 \leq j \leq n$. Writing $\vec{x}^j = t_{1j}\tilde{e}_1 + \cdots + t_{nj}\tilde{e}_n$, so that the matrix elements of $[T]$ are the t_{ij}, this becomes a routine calculation.

Indeed, there is nothing special about the standard basis in this calculation. If $(\vec{x}^1, \ldots, \vec{x}^n)$ and $(\vec{y}^1, \ldots, \vec{y}^n)$ are two ordered bases of $\mathbb{R}^n$ (i.e., n-tuples whose elements constitute a basis), then we have

$$\vec{y}^1 \wedge \cdots \wedge \vec{y}^n = \det([T]) \cdot \vec{x}^1 \wedge \cdots \wedge \vec{x}^n$$

where T is the linear transformation that takes $\vec{x}^j$ to $\vec{y}^j$.

If $\det([T]) > 0$ then T is *orientation preserving* and the two ordered bases *have the same orientation*; if $\det([T]) < 0$ it is *orientation reversing* and they have *opposite orientations*. We can specify an orientation of $\mathbb{R}^n$ by choosing an ordered basis, and two ordered bases specify the same orientation if and only if the transition matrix $[T]$ has positive determinant. Multiplicativity of the determinant shows that there are two disjoint classes of ordered bases, such that any two bases in the same class have the same orientation and every basis in one class has the opposite orientation to every basis in the other class.

Equivalently, an orientation can be specified by choosing a nonzero antisymmetric $(0, n)$-tensor. Such a tensor can be applied to any ordered basis, and the sign of the result tells which ordered bases have positive orientation. Two such tensors specify the same orientation if and only if they are positive multiples of each other.

An orientation on a smooth n-manifold M is then given by specifying a nowhere vanishing n-form, and M is *orientable* if there exists such an n-form. The n-form induces an orientation on every tangent space T_xM.

Any chart $\phi : U \to \mathbb{H}^n$ is a diffeomorphism between $U \subseteq M$ and an open subclass of $\mathbb{H}^n$, and we can pull back the n-form $dx_1 \wedge \cdots \wedge dx_n$ on $\phi(U)$ to U along ϕ. This yields an orientation on U. We say that two compatible charts $\phi : U \to \mathbb{H}^n$ and $\psi : V \to \mathbb{H}^n$ are *coherently oriented* if the n-form $dx_1 \wedge \cdots \wedge dx_n$ on $\phi(U \cap V)$ pulls back via $\phi \circ \psi^{-1}$ to an n-form $f \cdot dx_1 \wedge \cdots \wedge dx_n$ on $\psi(U \cap V)$ with $f > 0$ everywhere.

Lemma 16.19. *Every smooth manifold with boundary has an atlas consisting of charts $\phi_\lambda : U_\lambda \to \mathbb{H}^n$ for which $\phi_\lambda(U_\lambda)$ is an open ball in $\mathbb{H}^n$ for all λ.*

This lemma follows from the fact that every open subclass of $\mathbb{H}^n$ can be written as a union of countably many open balls, such that each point belongs to only finitely many of them.

Proposition 16.20. *Let M be a smooth n-manifold with boundary. Then M is orientable if and only if it has an atlas whose charts are all coherently oriented.*

Proof. Suppose M is orientable. Then there is a nowhere-vanishing n-form ω on M, and by the lemma M is equipped with an atlas for which every U_λ is connected (Corollary 10.63). For every λ we then have $(\phi_\lambda^{-1})^*\omega = f_\lambda \cdot dx_1 \wedge \cdots \wedge dx_n$ for some $f_\lambda \,\varepsilon\, C^\infty(\phi_\lambda(U_\lambda))$ which vanishes nowhere. Since U_λ is connected, it follows that f is either positive everywhere or negative everywhere, and by possibly multiplying one of the coordinates by -1 we can arrange that it is positive everywhere. The charts $\phi_\lambda : U_\lambda \to \mathbb{H}^n$ are then coherently oriented. (The $n = 1$ case must be handled separately, but it is easy.)

Conversely, suppose we are given coherently oriented smooth charts $\phi_\lambda : U_\lambda \to \mathbb{H}^n$ with $M = \bigcup U_\lambda$. Let $f_\kappa \,\varepsilon\, C^\infty(M)$ constitute a partition of unity subordinate to the open classes U_λ (Theorem 16.11). Then for each κ choose an index λ_κ with $\operatorname{supp}(f_\kappa) \subseteq U_{\lambda_\kappa}$ and define

$$\omega = \sum_\kappa f_\kappa \cdot \phi_{\lambda_\kappa}^*(dx_1 \wedge \cdots \wedge dx_n).$$

We need $\omega|_x$ to be nonzero for all $x \in M$. But coherent orientation of the charts entails that in local coordinates around x the tensor $\omega|_x$ is a sum of finitely many tensors of the form $a \cdot dx_1 \wedge \cdots \wedge dx_n$ with $a > 0$, and therefore $\omega|_x$ also has this form. Thus $\omega|_x \neq 0$. □

Thus an orientation on M can be specified by identifying an atlas all of whose charts are coherently oriented. Any smooth chart that is coherently oriented with all the charts in the given atlas is said to be *positively oriented*.

Proposition 16.21. *Let M be an oriented smooth n-manifold with boundary. Then ∂M is a smooth $(n-1)$-manifold (without boundary), the orientation on M induces an orientation on ∂M, and the inclusion of ∂M into M is a smooth embedding.*

This proposition is mostly trivial. We induce an orientation on ∂M by restricting an atlas on M all of whose charts are coherently oriented to ∂M. These charts map into $\partial\mathbb{H}^n \cong \mathbb{R}^n$.

Now we can define the integral of an n-form over an oriented n-manifold with boundary M. First, let $Q \subseteq M$ be contained in the domain U of a positively oriented smooth chart $\phi : U \to \mathbb{H}^n$ and suppose $\phi(Q)$ is a cube $C = [a_1, a_1 + b] \times \cdots \times [a_n, a_n + b]$. Let ω be an n-form that is supported in Q. Then $(\phi^{-1})^*(\omega)$ has the form $f(\vec{x})\, dx_1 \wedge \cdots \wedge dx_n$ for some smooth function f on C. We define the integral of ω over M, $\int_M \omega$, to equal $\int_C f\, d\vec{x}$.

I claim that this construction does not depend on the choice of Q. To see this, suppose ω is supported in another class $Q' \subseteq M$ for which $\psi(Q')$ is a cube C', for some positively oriented smooth chart ψ. Write $(\psi^{-1})^*(\omega) = g(\vec{y})\, dy^1 \wedge \cdots \wedge dy^n$. Then $f(\vec{x}) = J_{\psi\circ\phi^{-1}}(\vec{x}) \cdot g(\psi(\phi^{-1}(\vec{x})))$ for $\vec{x} \in \phi(Q \cap Q')$ by a comment made at the beginning of this section, and $J_{\psi\circ\phi^{-1}}$ is positive everywhere since ϕ and ψ are coherently oriented, so

$$\int_C f\, d\vec{x} = \int_C (g \circ (\psi \circ \phi^{-1}))\, |J_{\psi\circ\phi^{-1}}|\, d\vec{x} = \int_{C'} g\, d\vec{y}$$

by the change of variables theorem (Theorem 13.46). This proves the claim.

Say that a cube $[a_1, b_1] \times \cdots \times [a_n, b_n]$ in $\mathbb{H}^n$ is *rational* if the a_i's and b_i's are all rational. Thus there are countably many rational cubes in $\mathbb{H}^n$. Now if M is a smooth n-manifold with boundary, with an atlas consisting of charts $\phi_\lambda : U_\lambda \to \mathbb{H}^n$, then for every rational cube C contained in any $\phi_\lambda(U_\lambda)$, let $Q = \phi_\lambda^{-1}(C)$; if ω is any compactly supported n-form then we can find finitely many such $Q_1, \ldots, Q_k$ such that $Q_1^{int} \cup \cdots \cup Q_k^{int}$ contains the support of ω, where $Q_i^{int} = Q_i \setminus (\overline{M \setminus Q_i})$. Let $f_1, \ldots, f_k$ be a partition of unity on $Q_1^{int} \cup \cdots \cup Q_k^{int}$ subordinate to its covering by $Q_1^{int}, \ldots, Q_k^{int}$, with f_i supported on Q_i for $1 \leq i \leq k$. Then define

$$\int_M \omega = \int_M f_1\omega + \cdots + \int_M f_k\omega.$$

This definition is independent of all the choices we made because given any other $Q'_1, \ldots, Q'_l$ we can find a corresponding partition of unity $f'_0, \ldots, f'_l$ and then note that for each $1 \leq i \leq l$ we have $\int f'_i\omega = \int f_1 f'_i\omega + \cdots + \int f_k f'_i\omega$ by the linearity of integration in $\mathbb{R}^n$. Summing everything up shows that the value of $\int_M \omega$ is the same regardless of which covering was used.

Reversing the orientation of M multiplies $\int_M \omega$ by -1.

Theorem 16.22. *(Stokes' theorem) Let M be an oriented n-manifold and let ω be a compactly supported $(n-1)$-form on M. Then $\int_M d\omega = \int_{\partial M} i^*\omega$, where $i : \partial M \to M$ is the inclusion map.*

Proof. It suffices to check this for an $(n-1)$-form ω supported in $C = [0,1]^n$ which is zero on all the sides except possibly the $x_1 = 0$ side C'. For

simplicity suppose $\omega = f\,dx_1 \wedge \cdots \wedge dx_n$ with one dx_j removed; this suffices by linearity. Then $d\omega = (-1)^{j+1}\frac{\partial f}{\partial x_j}\,dx_1 \wedge \cdots \wedge dx_n$ and Fubini's theorem (Theorem 13.44) yields

$$\int_C d\omega = \int_0^1 \cdots \int_0^1 (-1)^{j+1}\frac{\partial f}{\partial x_j}\,dx_1 \cdots dx_n.$$

Performing the x_j integral first yields zero if $j \neq 1$ (the fundamental theorem of calculus, Theorem 13.40) because in that case f vanishes on both the $x_j = 0$ and $x_j = 1$ sides. If $j = 1$ then we get $\int_C d\omega = -\int_0^1 \cdots \int_0^1 f(0, x_2, \ldots, x_n)\,dx_2 \cdots dx_n$ since f vanishes on the $x_1 = 0$ side. But that is precisely $\int_{C'} i^*\omega$, as desired. □

If $\gamma : [a, b] \to \mathbb{R}^2$ is a smooth loop and $f, g : A \to \mathbb{R}$ are continuous, where $\gamma(t) = (\gamma_1(t), \gamma_2(t)) \in A$ for all $t \in [a, b]$, we define

$$\int_\gamma f\,dx = \int_a^b f(\gamma(t))\gamma_1'(t)\,dt \qquad \int_\gamma g\,dy = \int_a^b g(\gamma(t))\gamma_2'(t)\,dt.$$

We also compress notation and write

$$\int_\gamma (f\,dx + g\,dy) = \int_a^b (f(\gamma(t))\gamma_1'(t) + g(\gamma(t))\gamma_2'(t))\,dt.$$

This is a variation on the notion of line integral used in Section 13.4 and is also called a "line integral".

Corollary 16.23. *(Green's theorem) Let $\gamma : [a, b] \to \mathbb{R}^2$ be a smooth loop which encloses a region M and let $f, g : M \to \mathbb{R}$ be smooth. Then*

$$\int_M \left(\frac{\partial g}{\partial x} - \frac{\partial f}{\partial y}\right) d\mu = \int_\gamma (f\,dx + g\,dy)$$

where μ is Lebesgue measure on $\mathbb{R}^2$.

Proof. Let ω be the 1-form $\omega = f\,dx + g\,dy$. Then $d\omega = \left(\frac{\partial g}{\partial x} - \frac{\partial f}{\partial y}\right) dx \wedge dy$ and Theorem 16.22 says

$$\int_M \left(\frac{\partial g}{\partial x} - \frac{\partial f}{\partial y}\right) dx \wedge dy = \int_{\partial M} (f\,dx + g\,dy).$$

The proof is completed by observing that the right side is just the line integral $\int_\gamma (f\,dx + g\,dy)$. □

Corollary 16.24. *(Divergence theorem) Let M be the region in $\mathbb{R}^3$ enclosed by a smooth surface ∂M and let $P, Q, R : M \to \mathbb{R}$ be smooth. Then*

$$\int_M \left(\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z}\right) dx \wedge dy \wedge dz = -\int_{\partial M} (P\,dy \wedge dz + Q\,dz \wedge dx + R\,dx \wedge dy).$$

Proof. This is an immediate consequence of Theorem 16.22 with $\omega = P\,dy \wedge dz + Q\,dz \wedge dx + R\,dx \wedge dy$. □

16.4 Riemannian manifolds

A $(0,2)$-tensor α over $\mathbb{R}^n$ is *positive definite* if $\alpha(\vec{v},\vec{v}) \geq 0$ for all $\vec{v} \in \mathbb{R}^n$, with equality only when $\vec{v} = 0$. Thus a symmetric positive definite $(0,2)$-tensor is simply an inner product on $\mathbb{R}^n$. A *Riemannian metric* on a smooth manifold with boundary is a smooth $(0,2)$-tensor field g which is symmetric and positive definite at every point. A *Riemannian manifold* is a smooth manifold with boundary equipped with a Riemannian metric.

For any $x \in M$, we denote the inner product g induces on T_xM by $\langle v,w\rangle_g = g|_x(v,w)$, with corresponding norm $\|v\|_g = \sqrt{g|_x(v,v)}$. We can also write $\langle X,Y\rangle_g = g(X,Y) \,\varepsilon\, C^\infty(M)$ for the inner product of two vector fields X and Y, taken pointwise, and $\|X\|_g = \sqrt{\langle X,X\rangle_g} \,\varepsilon\, C(M)$ (in $C^\infty(M)$ if it never vanishes).

The first easy thing to note is that every smooth manifold can be equipped with a Riemannian metric.

Proposition 16.25. *Let M be a smooth manifold with boundary. Then there is a Riemannian metric on M.*

Proof. Given an atlas consisting of charts $\phi_\lambda : U_\lambda \to \mathbb{H}^n$, find a partition of unity subordinate to the U_λ, consisting of functions f_κ (Theorem 16.11). Every relatively open subclass of $\mathbb{H}^n$ carries a "standard" Riemannian metric which is just the standard inner product $\langle\cdot,\cdot\rangle$ on $\mathbb{R}^n$, identified with the tangent space at each point in the usual way; for each λ let g_λ be this Riemannian metric on $\phi_\lambda(U_\lambda)$. For each κ choose an index λ_κ such that the support of f_κ is contained in U_{λ_κ}, and define $g = \sum f_\kappa \cdot \phi^*_{\lambda_\kappa}(g_{\lambda_\kappa})$. For any $x \in M$ and $v \in T_xM$, we have $g|_x(v,v) = \sum f_\kappa(x) \cdot \langle D\phi_{\lambda_\kappa} v, D\phi_{\lambda_\kappa} v\rangle$; all the terms in the sum are nonnegative so $g|_x(v,v) \geq 0$, and if $v \neq 0$ then $f_\kappa(x) \cdot \langle D\phi_{\lambda_\kappa} v, D\phi_{\lambda_\kappa} v\rangle > 0$ for any κ with $f_\kappa(x) > 0$. Thus $g|_x(v,v) = 0$ only when $v = 0$. □

If M is a Riemannian manifold, the *length* of a smooth path $\gamma : [a,b] \to M$ is

$$L(\gamma) = \int_a^b \|\gamma'(t)\|_g \, dt,$$

where $\gamma'(t) \in T_{\gamma(t)}M$ is the *velocity vector* $\gamma'(t) = D\gamma|_t e_1$. Note that if γ is reparametrized by a smooth homeomorphism $f : [c,d] \to [a,b]$ then the new path $\gamma \circ f$ has the same length. This is a consequence of the change of variables theorem (Theorem 13.46).

We want to define the distance between two points of a connected Riemannian manifold to be the infimum of the lengths of the paths that join them. However, we can only take infima of sets, not classes in general. We must also verify that any two points can be joined by a path. Both problems can be solved using the notion of a "rational point" of a smooth manifold with boundary M with an atlas consisting of charts $\phi_\lambda : U_\lambda \to \mathbb{H}^n$ for which every $\phi_\lambda(U_\lambda)$ is an open ball in $\mathbb{H}^n$. (Such an atlas always exists by Lemma 16.19.) Say that $x \in M$ is λ-*rational* if $x \in U_\lambda$ and $\phi_\lambda(x) \in \mathbb{H}^n$ has all rational coordinates, and just *rational* if it is λ-rational for some λ. Also, say that two points $x, y \in M$ are *rationally linked* if there is a finite sequence of points $x_1, \ldots, x_k$ and indices $\lambda_1, \ldots, \lambda_{k+1}$ such that x_i is λ_i-rational for $1 \leq i \leq k$ and $x_i \in U_{\lambda_{i+1}}$ for $1 \leq i \leq k$, and with $x \in U_{\lambda_1}$ and $y \in U_{\lambda_{k+1}}$.

Proposition 16.26. *Let M be a smooth manifold with boundary. If any two points in M can be joined by a path then M is connected, and if M is connected then any two points can be joined by a smooth path.*

Proof. The first statement is easy: suppose any two points in M can be joined by a path and let $C \subseteq M$ be clopen. Then C definitely is or is not inhabited, and so is its complement (Proposition 10.55); if both C and C^c are inhabited, then there is a path γ that joins an element of one to an element of the other, and then $\gamma^{-1}(C)$ and $\gamma^{-1}(C^c)$ contradict Proposition 10.62. Thus any clopen subclass of M is either $\emptyset$ or M, showing that M is connected.

Conversely, suppose M is connected. Fix an atlas consisting of charts $\phi_\lambda : U_\lambda \to \mathbb{H}^n$ with the property that every $\phi_\lambda(U_\lambda)$ is an open ball in $\mathbb{H}^n$. By the LPO, any two rational points in M definitely are or are not rationally linked, and by the LPO again it definitely is or is not the case that any two rational points are rationally linked. Suppose there are rational points x and y that are not rationally linked. Then let A be the set of all rational points to which x is rationally linked, and let U be the union of the U_λ with the property that x is rationally linked to some λ-rational point. Then U is clopen, it contains x, and it does not contain y. This contradicts connectedness, so we conclude that any two rational points are rationally linked, and it easily follows that any two points in M are rationally linked. Now given any finite sequence that rationally links x and y we can reparametrize the straight line paths joining successive points so that all their derivatives are zero at the linking points, and then concatenating these straight line paths yields a smooth path from x to y. □

In order to define a distance function on a connected Riemannian manifold, we first need the following simple fact about $\mathbb{R}^n$.

Lemma 16.27. *Let $\vec{x}, \vec{y} \in \mathbb{R}^n$. Then every smooth path $\gamma : [a,b] \to \mathbb{R}^n$ with $\gamma(a) = \vec{x}$, $\gamma(b) = \vec{y}$ satisfies $L(\gamma) \geq \|\vec{x} - \vec{y}\|$, and this infimum is only attained by a straight line path.*

Proof. Suppose $\gamma(a) = \vec{x}$ and $\gamma(b) = \vec{y}$, and without loss of generality assume $\vec{x} = 0$. Let $a' = \sup\{t : \gamma(t) = 0\} = \inf_n \sup\{t \in [a,b] \cap \mathbb{Q} : \|\gamma(t)\| < \frac{1}{n}\}$; then by restricting γ to $[a', b]$ we get a shorter path joining $\vec{x}$ and $\vec{y}$. So we can assume $a' = a$.

For $t \in (a,b]$ write $\gamma(t) = f(t)\bar{\gamma}(t)$ where $\bar{\gamma}(t) = \frac{\gamma(t)}{\|\gamma(t)\|}$ and $f(t) \in (0,\infty)$. Then $\gamma'(t) = f'(t)\bar{\gamma}(t) + f(t)\bar{\gamma}'(t)$, and as $\bar{\gamma}'(t)$ is orthogonal to $\bar{\gamma}(t)$ we get that

$$\|\gamma'(t)\|^2 = (f'(t))^2 + (f(t))^2\|\bar{\gamma}'(t)\|^2 \geq (f'(t))^2.$$

Thus $L(\gamma) \geq \int_a^b |f'(t)|\,dt \geq f(b) - f(a) = \|y\|$. Moreover, equality obtains only if $\bar{\gamma}'(t) = 0$ for all t, i.e., $\bar{\gamma}$ is constant, in which case γ is a straight line. □

If $x, y \in U_\lambda$ then we define $d_\lambda(x,y)$ to be the length of the path $\phi_\lambda^{-1} \circ \gamma$ where $\gamma(t) = (1-t)\phi_\lambda(x) + t\phi_\lambda(y)$, $0 \leq t \leq 1$ (the straight line from $\phi_\lambda(x)$ to $\phi_\lambda(y)$). This is possible if we still assume that each $\phi_\lambda(U_\lambda)$ is a ball in $\mathbb{H}^n$ and hence convex. Now if M is connected then for any x and y we can define $d_g(x,y)$ to be the infimum of the values

$$d_{\lambda_1}(x,x_1) + d_{\lambda_2}(x_1,x_2) + \cdots + d_{\lambda_k}(x_{k-1},x_k) + d_{\lambda_{k+1}}(x_k,y)$$

over all finite sequences $x_1, \ldots, x_k$ and $\lambda_1, \ldots, \lambda_{k+1}$ which rationally link x and y.

Theorem 16.28. *Let M be a connected Riemannian manifold with boundary with Riemannian metric g. Then d_g is a metric (i.e., distance function) on M, and the resulting metric topology equals the original topology on M. The value $d_g(x,y)$ is the infimum of the lengths of the smooth paths that join x to y.*

Proof. Given any smooth path $\gamma : [a,b] \to M$ we can find a finite sequence which rationally links $x = \gamma(a)$ and $y = \gamma(b)$ and such that the sum $d_{\lambda_1}(x,x_1) + \cdots + d_{\lambda_{k+1}}(x_k,y)$ approximates $L(\gamma)$. I omit the details of this estimate. (Without loss of generality $a = 0$ and $b = 1$. First find $k \in \mathbb{N}$ such that for all $1 \leq i \leq k$ the points $\gamma(\frac{i-1}{k})$ and $\gamma(\frac{i}{k})$ lie in the same chart and

the straight line path between them has approximately the same length as $\gamma|_{[\frac{i-1}{k},\frac{i}{k}]}$. Then perturb them slightly to make them rational.) This implies that $d_g(x,y)$ is less than or equal to the length of any smooth path that joins x to y. Conversely, the smoothing argument used at the end of the proof of Proposition 16.26 shows that for any finite sequence that rationally links x and y the value $d_{\lambda_1}(x,x_1)+d_{\lambda_2}(x_1,x_2)+\cdots+d_{\lambda_k}(x_{k-1},x_k)+d_{\lambda_{k+1}}(x_k,y)$ equals the length of a smooth path that joins x to y. This proves the last statement of the theorem.

It is easy to see that d_g is symmetric and satisfies the triangle inequality, and that the distance from any point to itself is zero. Now suppose $x \neq y$; we must show that $d_g(x,y) \neq 0$. Fix λ with $x \in U_\lambda$ and $r > 0$ such that $\overline{B}_r(\phi_\lambda(x)) \cap \mathbb{H}^n \subseteq \phi_\lambda(U_\lambda)$ and $y \notin \phi_\lambda^{-1}(B_r(\phi_\lambda(x)) \cap \mathbb{H}^n)$. Any smooth path γ from x to y must include an earliest point $\gamma(b') = z \in U_\lambda$ with $\|\phi_\lambda(x) - \phi_\lambda(z)\| = r$. So it will suffice to show that $d_g(x,z) > 0$. And for this, by the last paragraph it will suffice to show that there is a positive lower bound on the length of the smooth path γ_z from x to z obtained by restricting γ to $[a,b']$ (where its original domain was $[a,b]$).

Identify U_λ with $\phi_\lambda(U_\lambda)$ and identify the tangent space to any point in $\overline{B}_r(\phi_\lambda(x))$ with $\mathbb{R}^n$. Then let $s = \inf\{\|v\|_g : v \in T_{x'}M$ for some $x' \in \overline{B}_r(\phi_\lambda(x))$, and $\|v\| = 1\}$, where $\|v\|$ is the usual Euclidean norm of v as an element of $\mathbb{R}^n$. The infimum exists and is attained by compactness. So the length of γ_z is bounded below by

$$\int_a^{b'} \|\gamma'(t)\|_g \, dt \geq \int_a^{b'} s \cdot \|\gamma'(t)\| \, dt,$$

where the right side is s times the length of γ' relative to the usual Euclidean metric. By the lemma, we get that $L(\gamma_z) \geq sr > 0$.

We still have to show that the metric topology equals the original topology on M. We can do this by showing that the identity map is continuous from either topology to the other. Let (x_n) be a sequence in M and suppose $d_g(x_n,x) \to 0$, i.e., (x_n) converges to x in the metric topology. If $x \in U_\lambda$ then the previous paragraph shows that (x_n) eventually lies in U_λ and $(\phi_\lambda(x_n))$ eventually lies in a ball about x of any positive radius, so (x_n) converges to x in the original topology. Conversely, suppose (x_n) converges to x in the original topology and $x \in U_\lambda$, and fix $r > 0$ such that $\overline{B}_r(\phi_\lambda(x)) \cap \mathbb{H}^n \subseteq \phi_\lambda(U_\lambda)$. Then $\phi_\lambda(x_n)$ lies in this ball for all sufficiently large n, and then we can bound $d(x_n,x)$ above by the straight line distance between $\phi_\lambda(x_n)$ and $\phi_\lambda(x)$ relative to g. But, letting $S = \sup\{\|v\|_g : v \in T_{x'}M$ for some $x' \in \overline{B}_r(\phi_\lambda(x))$, and $\|v\| = 1\}$, we get

that this straight line distance relative to g is at most S times the Euclidean distance from $\phi_\lambda(x_n)$ to $\phi_\lambda(x)$, which goes to zero. So (x_n) converges to x in the metric topology. □

16.5 Affine connections

If $\phi_\lambda : U_\lambda \to \mathbb{H}^n$ is a coordinate chart on a smooth manifold, then the vector field on $\phi_\lambda(U_\lambda)$ which takes the constant value e_i corresponds to a vector field on U_λ, which we denote E_i.

Definition 16.29. Let M be a smooth manifold with boundary. An *affine connection* on M is a bilinear map $\nabla : \mathfrak{X}M \times \mathfrak{X}M \to \mathfrak{X}M$ satisfying (writing $\nabla_X Y = \nabla(X, Y)$)

(i) $\nabla_{fX} Y = f \cdot \nabla_X Y$
(ii) $\nabla_X (fY) = X(f)Y + f \cdot \nabla_X Y$

for all $X, Y \,\varepsilon\, \mathfrak{X}M$ and $f \,\varepsilon\, C^\infty(M)$. The *torsion* of ∇ is defined by

$$T(X, Y) = \nabla_X Y - \nabla_Y X - [X, Y]$$

for $X, Y \,\varepsilon\, \mathfrak{X}M$. If the torsion is constantly zero then ∇ is *torsion-free*. ∇ is a *Riemannian connection* relative to a Riemannian metric g if it is torsion-free and satisfies

$$X(\langle Y, Z\rangle_g) = \langle \nabla_X Y, Z\rangle_g + \langle Y, \nabla_X Z\rangle_g$$

for all $X, Y, Z \,\varepsilon\, \mathfrak{X}M$.

According to (i), connections are not just linear in the first variable, they are $C^\infty(M)$*-linear*. This has a strong consequence.

Proposition 16.30. *Let M be a smooth manifold with boundary, let $x \in M$, and let $\phi : \mathfrak{X}M \to \mathbb{R}$ be a linear functional which satisfies $\phi(fX) = f(x)\phi(X)$ for all $f \,\varepsilon\, C^\infty(M)$ and $X \,\varepsilon\, \mathfrak{X}M$. Then $\phi(X)$ depends only on $X|_x$, the value of X at x.*

Proof. First observe that $\phi(X)$ only depends on the values of X in a neighborhood of x. That is because we can use Lemma 16.4 to find a function $f \,\varepsilon\, C^\infty(M)$ which is 1 at x and 0 outside a given neighborhood of x; then $\phi(fX) = f(x)\phi(X) = \phi(X)$. So if $X = Y$ in the given neighborhood then $fX = fY$ and therefore $\phi(X) = \phi(Y)$.

In particular, this is true for any coordinate neighborhood that contains x, which means that we can reduce to the case where $M = U$ is an open subclass of $\mathbb{H}^n$. In that case any smooth vector field can be written as $X = \sum f_i E_i$ $(1 \leq i \leq n)$ where the f_i belong to $C^\infty(U)$, and $C^\infty(M)$-linearity yields

$$\phi(X) = \sum f_i(x)\phi(E_i),$$

which only involves the values of the f_i at x and therefore depends only on the value of X at x. □

Thus, if ∇ is a connection, $x \in M$, and $X, Y\, \varepsilon\, \mathfrak{X}M$, then $\nabla_X Y|_x$ depends only on $X|_x$. We can think of $\nabla_X Y|_x$ as telling us the rate of change of Y at x in the direction $X|_x$, and we can also use the notation $\nabla_v Y|_x$ for a single tangent vector v at x.

Theorem 16.31. *Let M be a Riemannian manifold with boundary. Then*

$$\begin{aligned}\langle \nabla_X Y, Z\rangle_g = \frac{1}{2}\big(&X(\langle Y, Z\rangle_g) + Y(\langle Z, X\rangle_g) - Z(\langle X, Y\rangle_g)\\ &+ \langle Z, [X, Y]\rangle_g + \langle Y, [Z, X]\rangle_g - \langle X, [Y, Z]\rangle_g\big)\end{aligned}$$

defines a Riemannian connection ∇, and it is the unique Riemannian connection on M.

Proof. We start by checking that at each point $x \in M$, the expression for $\langle \nabla_X Y, Z\rangle_g$ depends only on the value of Z at x. We calculate

$$\begin{aligned}\langle \nabla_X Y, fZ\rangle_g &= \frac{1}{2}\big(X(f\cdot\langle Y, Z\rangle_g) + Y(f\cdot\langle Z, X\rangle_g) - f\cdot Z(\langle X, Y\rangle_g)\\ &\qquad + f\cdot\langle Z, [X, Y]\rangle_g + \langle Y, [fZ, X]\rangle_g - \langle X, [Y, fZ]\rangle_g\big)\\ &= \frac{1}{2}\big(X(f)\cdot\langle Y, Z\rangle_g + fX(\langle Y, Z\rangle_g) + Y(f)\langle Z, X\rangle_g\\ &\qquad + fY(\langle Z, X\rangle_g) - f\cdot Z(\langle X, Y\rangle_g) + f\cdot\langle Z, [X, Y]\rangle_g\\ &\qquad + f\cdot\langle Y, [Z, X]\rangle_g - X(f)\cdot\langle Y, Z\rangle_g - f\cdot\langle X, [Y, Z]\rangle_g\\ &\qquad - Y(f)\langle X, Z\rangle_g\big)\\ &= f\cdot\langle \nabla_X Y, Z\rangle_g.\end{aligned}$$

This shows that for fixed X and Y the map $Z \mapsto \langle \nabla_X Y, Z\rangle_g$ is $C^\infty(M)$-linear, and so it only depends on $Z|_x$ by Proposition 16.30. That means that the formula for $\langle \nabla_X Y, Z\rangle_g$ defines $\nabla_X Y$ as an element of $T_x M$ at each $x \in M$. Evidently the pairing of $\nabla_X Y$ with any smooth Z is smooth, so $\nabla_X Y$ is a smooth vector field. Thus we have defined a function $\nabla : \mathfrak{X}M \times \mathfrak{X}M \to \mathfrak{X}M$. Bilinearity is clear.

Condition (i) in Definition 16.29, $C^\infty(M)$-linearity in X, is checked by a computation similar to the one given just above that showed $C^\infty(M)$-linearity in Z. For condition (ii) we have

$$\begin{aligned}\langle\nabla_X(fY),Z\rangle_g &= \frac{1}{2}\big(X(f\cdot\langle Y,Z\rangle_g)+fY(\langle Z,X\rangle_g)-Z(f\cdot\langle X,Y\rangle_g)\\ &\qquad +\langle Z,[X,fY]\rangle_g+f\cdot\langle Y,[Z,X]\rangle_g-\langle X,[fY,Z]\rangle_g\big)\\ &= \frac{1}{2}\big(X(f)\cdot\langle Y,Z\rangle_g+fX(\langle Y,Z\rangle_g)+fY(\langle Z,X\rangle_g)\\ &\qquad -Z(f)\cdot\langle X,Y\rangle_g-fZ(\langle X,Y\rangle_g)\\ &\qquad +\langle Z,f[X,Y]+X(f)\cdot Y\rangle_g+f\cdot\langle Y,[Z,X]\rangle_g\\ &\qquad -\langle X,f\cdot[Y,Z]-Z(f)\cdot Y\rangle_g\big)\\ &= X(f)\cdot\langle Y,Z\rangle_g+f\cdot\langle\nabla_XY,Z\rangle_g\\ &= \langle X(f)\cdot Y+f\cdot\nabla_XY,Z\rangle_g.\end{aligned}$$

Next, ∇ is torsion-free because

$$\langle\nabla_XY-\nabla_YX,Z\rangle_g=\frac{1}{2}\big(\langle[X,Y],Z\rangle_g-\langle[Y,X],Z\rangle_g\big)=\langle[X,Y],Z\rangle_g$$

and it is Riemannian because

$$\langle\nabla_XY,Z\rangle_g+\langle Y,\nabla_XZ\rangle_g=\frac{1}{2}\big(X(\langle Y,Z\rangle_g)+X(\langle Z,Y\rangle_g)\big)=X(\langle Y,Z\rangle_g).$$

For uniqueness, suppose $\widetilde{\nabla}$ is any Riemannian connection on M. Then we have

$$\langle\widetilde{\nabla}_XY,Z\rangle_g=X(\langle Y,Z\rangle_g)-\langle Y,\widetilde{\nabla}_XZ\rangle_g$$

by the Riemannian condition and

$$\begin{aligned}\langle\widetilde{\nabla}_XY,Z\rangle_g &= \langle\widetilde{\nabla}_YX,Z\rangle_g+\langle[X,Y],Z\rangle_g\\ &= Y(\langle X,Z\rangle_g)-\langle X,\widetilde{\nabla}_YZ\rangle_g+\langle[X,Y],Z\rangle_g\end{aligned}$$

by the torsion-free condition followed by the Riemannian condition. Adding these two equations yields

$$\begin{aligned}2\langle\widetilde{\nabla}_XY,Z\rangle_g = X(\langle Y,Z\rangle_g)&-\langle Y,\widetilde{\nabla}_XZ\rangle_g\\ &+Y(\langle X,Z\rangle_g)-\langle X,\widetilde{\nabla}_YZ\rangle_g+\langle[X,Y],Z\rangle_g,\end{aligned}$$

and finally we have

$$\begin{aligned}0 &= -Z(\langle X,Y\rangle_g)+\langle\widetilde{\nabla}_ZX,Y\rangle_g+\langle X,\widetilde{\nabla}_ZY\rangle_g\\ &= -Z(\langle X,Y\rangle_g)+\langle\widetilde{\nabla}_XZ+[Z,X],Y\rangle_g+\langle X,\widetilde{\nabla}_YZ-[Y,Z]\rangle_g\end{aligned}$$

by the Riemannian condition followed by the torsion-free condition used twice. Combining yields the formula for $\langle\nabla_XY,Z\rangle_g$ given in the statement of the theorem. □

The unique Riemannian connection on M is also called the *Levi-Civita connection*.

Within a coordinate chart $\phi_\lambda : U_\lambda \to \mathbb{H}^n$ we can always decompose a connection ∇ by writing

$$\nabla_{E_i} E_j = \sum_k \Gamma_{ij}^k E_k$$

where the Γ_{ij}^k are smooth real-valued functions and, again, the E_i are the coordinate vector fields. The Γ_{ij}^k are *Christoffel symbols*. Then, for arbitrary $X = \sum a_i E_i$ and $Y = \sum b_j E_j$, with the a_i and b_j smooth real-valued functions, we get

$$\begin{aligned}
\nabla_X Y &= \sum_{i,j} a_i \nabla_{E_i}(b_j E_j) \\
&= \sum_{i,j} a_i (E_i(b_j) E_j + b_j \nabla_{E_i} E_j) \\
&= \sum_{i,j} \left(a_i E_i(b_j) E_j + \sum_k a_i b_j \Gamma_{ij}^k E_k \right) \\
&= \sum_k X(b_k) E_k + \sum_{i,j,k} a_i b_j \Gamma_{ij}^k E_k.
\end{aligned}$$

As we noted above, when evaluating $\nabla_X Y$ at a point x, all we need to know about X is the values $a_i(x)$. That is, it only depends on $X|_x$. Thus the derivative of Y along a smooth path $\gamma : [a, b] \to U_\lambda$ can be defined as

$$\nabla_{\gamma'} Y = \sum_k (b_k \circ \gamma)' E_k + \sum_{i,j,k} \gamma_i' b_j \Gamma_{ij}^k E_k$$

where $\gamma(t) = (\gamma_1(t), \ldots, \gamma_n(t))$ in local coordinates. This expression depends in its second argument only on the values of the b_j on the path. In other words, if $Y_1 = Y_2$ on the range of γ then $\nabla_{\gamma'} Y_1 = \nabla_{\gamma'} Y_2$. If $\gamma : [a, b] \to U_\lambda$ is a smooth path and $v : [a, b] \to TU_\lambda$ is a smooth path with $\pi(v(t)) = \gamma(t)$ for all t — that is, at each t the function v selects a tangent vector at $\gamma(t)$ — then we can therefore define the derivative of v along γ to be

$$\nabla_{\gamma'} v = \sum_k v_k' E_k + \sum_{i,j,k} \gamma_i' v_j \Gamma_{ij}^k E_k,$$

where $v(t) = (v_1(t), \ldots, v_n(t))$ in local coordinates. If on sufficiently short segments of t the vector functions γ' and v extend to smooth vector fields X and Y on M, then $\nabla_{\gamma'} v$ is just the restriction of $\nabla_X Y$ to the range of γ on those segments. This shows that the formula we have given for $\nabla_{\gamma'} v$

is, in this case, independent of the choice of coordinate chart. But the γ_i' and v_i can always be approximated by polynomials (the Stone-Weierstrass theorem, Theorem 15.6), which always extend locally, so $\nabla_{\gamma'} v$ is always independent of the choice of coordinate chart, without qualification.

Applying the above to $v(t) = \gamma'(t)$, we can consider the derivative of the tangent vector to γ along γ, $\nabla_{\gamma'}\gamma'$. If this value is constantly zero then γ is a *geodesic*. According to the above, γ is a geodesic if and only if

$$\gamma_k''(t) + \sum_{i,j} \gamma_i'(t)\gamma_j'(t)\Gamma_{ij}^k(\gamma(t)) = 0$$

for $1 \leq k \leq n$. We can easily deduce from Picard's existence theorem (Theorem 13.62) that a geodesic emanating from any point in the direction of any tangent vector at that point exists and is unique on $t \in (-\epsilon, \epsilon)$ for sufficiently small ϵ. Actually we get a little more:

Proposition 16.32. *Let M be a Riemannian manifold and let $x \in M$. Then there is a neighborhood V of x and $r, \delta > 0$ such that for each $y \in V$ and $v \in T_yM$ with $\|v\|_g < r$ there is a unique geodesic $\gamma_{y,v} : (-\delta, \delta) \to M$ with $\gamma_{y,v}(0) = y$ and $\gamma_{y,v}'(0) = v$. The map $(t, y, v) \mapsto \gamma_{y,v}(t)$ from $(-\delta, \delta) \times TV$ into M is smooth.*

This is proven by working in local coordinates and using Picard's existence theorem (Theorem 13.62) to solve the system of n ordinary differential equations

$$\begin{aligned} f_k' &= g_k \\ g_k' &= -\sum_{i,j} \Gamma_{ij}^k(f(t))g_i(t)g_j(t) \end{aligned}$$

for $1 \leq k \leq n$. These equations are defined on $U \times \mathbb{R}^n$ where $U \subseteq \mathbb{R}^n$ is open. The desired geodesic is $\gamma(t) = (f_1(t), \ldots, f_n(t))$ with $\gamma(0) \in V \subseteq U$ and $\gamma'(0) \in \mathbb{R}^n$.

We can now define the exponential map.

Definition 16.33. Let M be a Riemannian manifold and let $x \in M$. For any $v \in T_xM$ we have a value of $\epsilon > 0$ and a unique geodesic $\gamma : [-\epsilon, \epsilon] \to M$ satisfying $\gamma(0) = x$ and $\gamma'(0) = v$. The *exponential map* is defined as $\mathrm{Exp}_x(rv) = \gamma(r)$ for $r \in [-\epsilon, \epsilon]$.

The consistency of this definition when v is multiplied by a scalar is a trivial consequence of the local uniqueness of geodesics, up to the exact choice of ϵ, which is left undetermined here. That defect will be remedied in the Hopf-Rinow theorem (Theorem 16.36).

Next, we will derive the basic property of geodesics as locally length-minimizing. The proof is a straightforward generalization of the proof of Lemma 16.27.

Theorem 16.34. *Let M be a Riemannian manifold and let $x \in M$. Then there is a value of $\epsilon > 0$ such that the exponential map is defined on the ball about the origin of T_xM of radius ϵ and takes it diffeomorphically onto an open neighborhood U of x. For any $y = \mathrm{Exp}_x(v) \in U$ we have $d(x, y) = \|v\|_g$, and the geodesic emanating from x in the direction v is, up to reparametrization, the unique length-minimizing path from x to y.*

Proof. The derivative of the map Exp_x at $0 \in T_xM$ is the identity map, so the first part follows from Theorem 16.9.

We must show that any smooth path from x to y has length at least $\|v\|_g < \epsilon$. Let γ be such a path, say $\gamma : [0, 1] \to U$. We can assume its range is contained in U because the argument will show that any path that escapes U must have length at least ϵ. Write $\gamma(t) = \mathrm{Exp}_x(r(t)v(t))$ where $r : [0, 1] \to [0, \|v\|_g]$ and $v(t) : [0, 1] \to T_xM$ with $\|v(t)\|_g = 1$ for all t. As in the proof of Lemma 16.27, we can assume that $r(t) > 0$ for $t > 0$, and similarly we can assume that $r(t) < \|v\|_g$ for $t < 1$. We must show that $v(t) = v$ for all t.

Define $\tilde{\gamma}(s, t) = \mathrm{Exp}_x(s \cdot v(t))$ for $s \in [0, \epsilon)$ and $t \in [0, 1]$, so that $\gamma(t) = \tilde{\gamma}(r(t), v(t))$. Then

$$\gamma'(t) = \frac{\partial \tilde{\gamma}}{\partial s} \cdot r'(t) + \frac{\partial \tilde{\gamma}}{\partial t}.$$

I claim that the the two terms on the right are orthogonal to each other. First,

$$\frac{\partial}{\partial s}\left(\left\langle \frac{\partial \tilde{\gamma}}{\partial s}, \frac{\partial \tilde{\gamma}}{\partial t} \right\rangle_g\right) = \left\langle \frac{\partial}{\partial s}\left(\frac{\partial \tilde{\gamma}}{\partial s}\right), \frac{\partial \tilde{\gamma}}{\partial t} \right\rangle_g + \left\langle \frac{\partial \tilde{\gamma}}{\partial s}, \frac{\partial}{\partial s}\left(\frac{\partial \tilde{\gamma}}{\partial t}\right) \right\rangle_g$$

where the first term is zero because for fixed t the function $\tilde{\gamma}$ is a geodesic, so $\frac{\partial}{\partial s}(\frac{\partial \tilde{\gamma}}{\partial s}) = 0$, and the second term is zero because

$$\begin{aligned}\left\langle \frac{\partial \tilde{\gamma}}{\partial s}, \frac{\partial}{\partial s}\left(\frac{\partial \tilde{\gamma}}{\partial t}\right) \right\rangle_g &= \left\langle \frac{\partial \tilde{\gamma}}{\partial s}, \frac{\partial}{\partial t}\left(\frac{\partial \tilde{\gamma}}{\partial s}\right) \right\rangle_g \\ &= \frac{1}{2} \cdot \frac{\partial}{\partial t}\left(\left\langle \frac{\partial \tilde{\gamma}}{\partial s}, \frac{\partial \tilde{\gamma}}{\partial s} \right\rangle_g\right) \\ &= \frac{1}{2} \cdot \frac{\partial}{\partial t}(1) = 0.\end{aligned}$$

This shows that $\frac{\partial}{\partial s}(\langle\frac{\partial\tilde{\gamma}}{\partial s},\frac{\partial\tilde{\gamma}}{\partial t}\rangle_g = 0$, but the inner product is zero at $(0,0)$ because $\tilde{\gamma}(0,t) = 0$ for all t and therefore $\frac{\partial\tilde{\gamma}}{\partial t}$ vanishes at $(0,0)$. So the inner product must be zero, i.e., the two terms in the expression for γ' are orthogonal, and therefore

$$\|\gamma'(t)\|_g^2 = \|v(t)\|_g^2 \cdot |r'(t)|^2 + \left\|\frac{\partial\tilde{\gamma}}{\partial t}\right\|^2 \geq |r'(t)|^2$$

(since $\|v(t)\|_g = 1$ for all t), so that

$$\begin{aligned} L(\gamma) &= \int_0^1 \|\gamma'(t)\|_g \, dt \geq \int_0^1 |r'(t)| \, dt \\ &\geq \left|\int_0^1 r'(t)\, dt\right| = r(1) - r(0) = \|v\|_g, \end{aligned}$$

with equality only when $\frac{\partial\tilde{\gamma}}{\partial t}$ is constantly 0, i.e., $v(t)$ is constant. But then γ can only go from x to y if $v(t) = v$ for all t. □

Finally, here is the promised condition under which the exponential map is defined on all of TM.

Lemma 16.35. *Every complete Riemannian manifold is boundedly compact, in the metric sense.*

Proof. Let M be a complete Riemannian manifold and let $x \in M$. For each $r > 0$, the closed ball $\overline{B}_r(x)$ is complete and therefore definitely is or is not compact (Lemma 10.34 and Theorem 10.35). Thus the supremum R of the set of rational values of r for which $\overline{B}_r(x)$ is compact exists. Suppose it is finite; then $\overline{B}_R(x)$ is easily seen to be totally bounded and therefore compact.

Let $A = \{x_1, x_2, \ldots\}$ be a dense subset of $\overline{B}_R(x)$ and for each n let R_n be the largest radius such that $\overline{B}_{R_n}(x_n)$ is compact. If the R_n were not bounded away from zero then, by compactness, there would be a point with no compact neighborhoods, but in any manifold every point has compact neighborhoods. Thus the R_n are bounded away from zero, and then taking a finite subcover (Corollary 10.38) yields a finite union of compact classes which contains $\overline{B}_{R+\epsilon}(x)$ for some $\epsilon > 0$, contradicting the choice of R. We conclude that $R = \infty$. □

Theorem 16.36. *(Hopf-Rinow theorem) Let M be a connected Riemannian manifold. If M is complete as a metric space, then the geodesic ray emanating from any point in M in any direction can be extended indefinitely.*

Proof. For all $x \in M$ we know by Proposition 16.32 that there exists $\epsilon > 0$ such that for all $v \in T_xM$ with $\|v\|_g = 1$ there exists a unique geodesic $\gamma : [0, \epsilon) \to M$ with $\gamma(0) = x$ and $\gamma'(0) = v$. I claim that for all $x \in M$, all $v \in T_xM$ with $\|v\|_g = 1$, and all $r > 0$ there definitely does or does not exist a geodesic ray of length r emanating from x in the v direction. This can be diagnosed by asking whether there exist points $x_1, \ldots, x_n$ and indices $\lambda_1, \ldots, \lambda_n$ for the given atlas on M, with $x \in U_{\lambda_1}$, x_i is λ_i-rational for all i, and $x_i \in U_{\lambda_{i+1}}$ for $1 \leq i < n$, and such that (1) $d(x, x_n) \geq r$, (2) the length of the piecewise straight line path from x to x_n (as in the proof of Proposition 16.26) is arbitrarily close to r, and (3) the starting direction is arbitrarily close to v. Using compactness (Lemma 16.35), we can take a pointwise limit of such paths parametrized by arc length to get a geodesic of length r that starts in the v direction. Conversely, any such geodesic can be approximated in such a way. This proves the claim.

Fix $x \in M$ and $N \in \mathbb{N}$. I now claim that there exists $r > 0$ such that every $y \in \overline{B}_N(x)$ has geodesic rays of length r emanating in all directions. It will suffice to prove this for a dense set of y's and a dense set of unit vectors, as we can again take a pointwise limit of geodesics to get a geodesic. By the first claim, for every rational $r > 0$, every rational point $y \in \overline{B}_N(x)$, and every unit vector w in T_yM there definitely does or does not exist a geodesic of length r emanating from y in the w direction, so we can use the LPO to either find a single $r > 0$ that works or a sequence of y's and w's that don't work. In the latter case, taking a cluster point yields a contradiction to Proposition 16.32. This proves the second claim.

We now know that every element of $B_N(x)$ has geodesic rays of length r emanating in all directions. So given any $v \in T_xM$ with $\|v\|_g = 1$ we can find a length r geodesic γ_1 with $\gamma_1(0) = x$ and $\gamma_1'(0) = v$, and then let $x_1 = \gamma_1(r)$ and find a length r geodesic γ_2 with $\gamma_2(0) = x_1$ and $\gamma_2'(0) = \gamma_1'(r)$, and so on. Concatenating, we get a geodesic of length at least N emanating from the given point in the given direction. Since N was arbitrary this completes the proof. □

Bibliography

[1] C. Akemann and N. Weaver, Consistency of a counterexample to Naimark's problem, *Proc. Natl. Acad. Sci. USA* **101** (2004), 7522-7525.

[2] J. L. Austin, *Sense and Sensibilia*, 1962.

[3] P. Benacerraf, What numbers could not be, *Philos. Rev.* **74** (1965), 47-73.

[4] E. Bishop and D. Bridges, *Constructive Analysis*, 1985.

[5] M. Black, The elusiveness of sets, *Review of Metaphysics* **24** (1971), 614-636.

[6] G. Boolos, The iterative conception of set, *J. Philos.* **68** (1971), 215-231.

[7] E. B. Davies, Building infinite machines, *British J. Philos. Sci.* **52** (2001), 671-682.

[8] M. Dummett, Wittgenstein's philosophy of mathematics, *The Philosophical Review* **68** (1959), 324-348.

[9] ———, The philosophical significance of Gödel's theorem, *Ratio* **5** (1963), 140-155.

[10] ———, *The Seas of Language*, 1993.

[11] G. Frege, *Die Grundlagen der Arithmetic: eine logisch mathematische Untersuchung über der Begriff der Zahl*, 1884.

[12] K. Gödel, Russell's mathematical logic, in *The Philosophy of Bertrand Russell*, P. A. Schilpp, ed., 1944, 125-153.

[13] ———, What is Cantor's continuum problem?, *Amer. Math. Monthly* **54** (1947), 515-525.

[14] P. Halmos, *Naive Set Theory*, 1960.

[15] J. D. Hamkins, The set-theoretic multiverse, *Rev. Symb. Logic* **5** (2012), 416-449.

[16] G. Hellman, *Mathematics Without Numbers: Towards a Modal-Structural Interpretation*, 1989.

[17] A. Heyting, Sur la logique intuitionniste, *Acad. Roy. Belg. Bull. Cl. Sci.* **16** (1930), 957-963.

[18] I. Jané, Reflections on Skolem's relativity of set-theoretical concepts, *Philosophia Mathematica* **9** (2001), 129-153.

[19] ———, The iterative conception of sets from a cantorian perspective, in *Logic, Methodology, and Philosophy of Science: Proceedings of the Twelfth International Congress*, P. Hájek, L. Valdés-Villanueva, and D. Westerstål,

eds., 2005, 373-393.
[20] S. Kripke, Outline of a theory of truth, *The Journal of Philosophy* **72** (1975), 690-716.
[21] D. Lewis, *On the Plurality of Worlds*, 1986.
[22] P. Maddy, A theory of sets and classes, in *Between Logic and Intuition*, G. Sher and R. Tiedzen, eds., 2000, 299-316.
[23] J. B. Manchak, On the possibility of supertasks in general relativity, *Found. Phys.* **40** (2010), 276-288.
[24] C. Menzel, On the iterative explanation of the paradoxes, *Philos. Stud.* **49** (1986), 37-61.
[25] C. Parsons, What is the iterative conception of set?, in *Logic, Foundations of Mathematics and Computability Theory*, Part I, R. E. Butts and J. Hintikka, eds., 1977, 335-367.
[26] H. Poincaré, Les mathématiques et la logique, *Revue de métaphysique et de morale* **14** (1906), 294-317.
[27] D. Prawitz, Truth and objectivity from a verificationist point of view, in *Truth in Mathematics (Mussomeli, 1995)*, 1998, 41-51.
[28] H. Putnam, What is mathematical truth, *Historia Math.* **2** (1975), 529-533.
[29] W. V. O. Quine, On what there is, *Rev. Metaphysics* **2** (1948), 21-38.
[30] ———, *The Ways of Paradox and Other Essays*, 1976.
[31] ———, *Mathematical Logic* (revised edition), 1981.
[32] ———, *Philosophy of Logic* (second edition), 1986.
[33] F. P. Ramsey, *The Foundations of Mathematics*, 1931.
[34] B. Rang and W. Thomas, Zermelo's discovery of the "Russell Paradox", *Historia Math.* **8** (1981), 15-22.
[35] B. Russell, Letter to Frege, translated in [49], pp. 124-125.
[36] ———, On some difficulties in the theory of transfinite numbers and order types, *Proc. London Math. Soc.* **4** (1907), 29-53.
[37] ———, Mathematical logic as based on the theory of types, *Amer. J. Math.* **30** (1908), 222-262.
[38] ———, *Introduction to Mathematical Philosophy*, 1919.
[39] ———, The limits of empiricism, *Proceedings of the Aristotelian Society* **34** (1936), 131-150.
[40] B. Russell and A. N. Whitehead, *Principia Mathematica*, 1910-1913.
[41] M. F. Sharlow, Proper classes via the iterative conception of set, *J. Symb. Log.* **52** (1987), 636-650.
[42] T. Skolem, Some remarks on axiomatized set theory, in [49], pp. 291-301.
[43] ———, Sur la portée du théorème de Löwenheim-Skolem, in [44], pp. 455-482.
[44] ———, *Selected Works in Logic*, J. E. Fenstad, ed., 1970.
[45] B. H. Slater, Grammar and sets, *Australas. J. Philos.* **84** (2006), 59-73.
[46] T. Tao, comment in https://mathoverflow.net/questions/336191/ cauchy-reals-and-dedekind-reals-satisfy-the-same-mathematical-theorems (2019).
[47] A. Tarski, *Logic, Semantics, Metamathematics* (second edition), 1983.
[48] J. F. Thomson, Tasks and super-tasks, *Analysis* **15** (1954), 1-13.
[49] J. van Heijenoort (ed.), *From Frege to Gödel: A Source Book in Mathemat-*

ical Logic, 1879-1931, 1967.
[50] J. von Neumann, An axiomatization of set theory, in [49], pp. 394-413.
[51] H. Wang, *A Survey of Mathematical Logic*, 1962.
[52] N. Weaver, Is set theory indispensible? arXiv:math/0905.1680 (2009).
[53] ———, *Forcing for Mathematicians*, 2014
[54] ———, *Truth & Assertibility*, 2015.
[55] ———, Hierarchies of Tarskian truth predicates, arXiv:math/2202.00851 (2022).
[56] H. Weyl, *Das Kontinuum. Kritische Untersuchungen über die Grundlagen der Analysis*, 1918.
[57] ———, *Philosophy of Mathematics and Natural Science*, 1949.
[58] L. Wittgenstein, Tractatus Logico-Philosophicus, 1922.
[59] E. Zermelo, Investigations in the foundations of set theory I, in [49], pp. 199-215.

Notation

V_α, 7
$\mathbb{N}$, 8
$\mathcal{P}(\mathbb{N})$, 8
Λ_q, 45
Λ'_q, 46
$\mathbb{T}[\phi]$, 55
$\hat{x}$, 55
$\mathbb{A}[\phi]$, 66
$\bot$, 67
Λ_a, 68
Λ, 80
$\mathbb{H}[\phi, x]$, 83
$\mathcal{R}$, 83
$\mathcal{R}_a$, 83
$\mathbb{M}[\phi]$, 85
L_1, L_2, 102
θ_1, θ_2, 102
$\mathrm{Ind}(\phi)$, 141
$\mathrm{Prog}\,\phi$, $\mathrm{Prog}_{\preceq}\,\phi$, 143
T_v, 147
$\mathcal{L}$, $\mathcal{L}_a$, 150
$\mathbb{T}_a$, 150
$M(a)$, $W(a)$, 151
v_a, 151
γ_n, Γ_0, 151
$\mathcal{L}^p_a$, 152
$B^c_a(b)$, $C(a)$, 152
$\bot$, $\wedge$, $\vee$, $\to$, 166
$\neg$, $\leftrightarrow$, 166
$\forall x$, $\exists x$, 166
$\theta(\alpha, \beta)$, 171
$x \in X$, 172
$X \subseteq Y$, $X \subset Y$, 172
$x \notin X$, 172
$X_1 \setminus X_2$, 173
Y^c, 173
$X_1 \cup X_2$, $X_1 \cap X_2$, 173
$\emptyset$, 173
$2^\alpha 3^\beta$, 173
$X \times Y$, 173
(x, y), 173
X^ω, 173
(x_n), $\vec{x}$, 173
α_p, 173
$\mathbb{N}$ (v 2.0), 174
Ω, 174
X_λ, 176
$X \,\varepsilon\, I$, 176
$\bigcup X_\lambda$, $\bigcap X_\lambda$, 176
$\mathbb{M}$, $\mathbb{A}$, 179
$\exists!$, 185
$g \circ f$, 185
$f|_{X_0}$, 186
$f : X \cong Y$, 186
f^{-1}, 186
$f(X_0)$, $f^{-1}(Y_0)$, 187
$\alpha_\to$, 187
$\langle X_n \rangle$, 188, 189
$\langle X_1, \ldots, X_n \rangle$, 188
$X_1 \sqcup \cdots \sqcup X_n$, 188
$\prod X_n$, 188
X^n, 189

π_{i_0}, 190
$\varprojlim X_i$, 190
$\widetilde{X}/\!\sim$, 191
$[x]$, 191
$\mathrm{fn}(A, X)$, 193
$\mathcal{P}(X)$, 193
$\varinjlim X_i$, 194
$\overrightarrow{\mathbb{Z}}$, $\mathbb{Q}$, 194
$\mathbb{Z}_+$, 195
$\mathbb{R}$, 195
(α, β), $[\alpha, \infty)$, 196
$\overline{\mathbb{R}}$, 196
$\sup A$, $\inf A$, 197
$\min(a, b)$, $\max(a, b)$, 197
$\mathbb{C}$, 198
$\mathrm{Re}(z)$, $\mathrm{Im}(z)$, 198
$|z|$, 198
$\overline{z}$, 198
$\cos\theta$, $\sin\theta$, 198
$d(x, y)$, 203
$\|\vec{a}\|$, 204
$x_n \to x$, 205, 229
$\lim x_n = x$, 205
$B_r(x), \overline{B}_r(x)$, 207
$\mathring{X}$, 207
$\overline{A}$, 210, 230
$d_Y(x)$, 214
$\limsup a_n$, $\liminf a_n$, 214
$\sum a_i$, 214
$\|f\|_\infty$, 217
$C(X, Y)$, 227, 250
$\mathrm{Homeo}(X)$, 228, 250
$\breve{A}$, 249
$G \cong H$, 255
ℓ_g, ℓ, 255
G/H, 259
$[g]$, $[g]_H$, 259
$d|n$, 266
$\phi(n)$, 267
$R[x_1, \ldots, x_n]$, 268
R/I, 270
$[a]$, $[a]_I$, 270
R^*, 271
B^*, 289
$\hom(B, \{0, 1\})$, 289
X^*, 289
ϕ^*, f^*, 290
σ, τ, 290
1_C, 292
$F^{<\omega}$, 293
$[T]$, 296
$M_n(F)$, $M_{m,n}(F)$, 296
$[I]$, 296
$\mathrm{span}(S)$, 298
$\dim(V)$, 302
$\mathrm{Im}(T)$, 302
$\mathcal{L}(V)$, $\mathcal{L}(V, W)$, 304
$\mathrm{Sub}(V)$, 304
$\mathrm{sign}(\sigma)$, 304
$\det([T])$, 305
$[K : F]$, 306
$F[a]$, 309
$\overline{x}$, 309
$\phi(p)$, 311
$\mathrm{Gal}(K/F)$, 314
K^H, 314
Df, 331
$\frac{\partial f}{\partial x_j}$, 331
$\langle \vec{x}, \vec{y} \rangle$, 333
$\|\vec{a}\|_\infty$, 334
R_i, 337
$\mu[R_i]$, 337
$A \triangle B$, 337
$\mu[A]$, 338
$\mu \times \nu$, 343, 344
$\mu \perp \nu$, 345
$\nu \ll \mu$, 345
μ^+, μ^-, 347
μ^r, μ^i, 348
$|\mu|$, 348
$\int f$, $\int f\, d\mu$, 352
f_+, f_-, 353
$\int_a^b f(t)\, dt$, 355
$\int f\, dxdy$, 357
J_f, 359
$\frac{d\nu}{d\mu}$, 361
$\exp z$, 363
e^z, 363
$\cos z$, $\sin z$, 364
π, 364
$\ln z$, 364

$\int_\gamma f(z)\,dz$, 364
$[S]$, (S), 373
V_K, 373
$[K]$, 374
$\mathbb{S}^1$, 374
v_S, 375
$\mathrm{St}(v)$, 375
$f \sim g \ (\mathrm{rel}\, A)$, 377
$\tilde{\omega}$, 378
$\omega \frown \chi$, 379
$\pi_1([K])$, $\pi_1([K], e)$, 379
$\pi_1(X)$, 381
$\mathbb{D}^2$, 383
$G_1 * G_2$, $G_1 *_F G_2$, 387
$\mathbb{S}^n$, $\mathbb{T}^n$, 390
$\mathbb{P}^2$, 391
$\mathcal{H}_g$, $\mathcal{C}_g$, 394
$\hat{S}$, $\check{S}$, 397
$[v_0, \ldots, v_p]$, 397
$K^{(p)}$, 397
$C_p(K)$, 397
$c_{\hat{S}}$, 398
∂_p, 398
$Z_p(K)$, $B_p(K)$, $H_p(K)$, 398
$\mathcal{C}K$, 399
$\underline{\phi}_p$, 399
$\overline{\phi}_p$, 400
$\underline{\chi}_p$, $\overline{\chi}_p$, 400
$\overline{\theta}_p$, 401
$\underline{d}_p$, 402
$\overline{f}_p$, 403
$H_p(X)$, 404
$\mathbb{F}$, $\mathbb{F}_0$, 407
$\|v\|$, 407
$C(K)$, $C_0(K)$, 408
$\|f\|_\infty$, 408, 414
l^∞, c_0, l^p, 410
$\|\vec{a}\|_\infty$, $\|\vec{a}\|_p$, 410
$\mathcal{M}_0(X)$, 412
$L^p(X)$, 412
$L^\infty(X)$, 414
$\|T\|$, 416
$E \cong F$, 416
$(E)_r$, $[E]_r$, 416
$E \oplus F$, $v \oplus w$, 419
E/E_0, 420
$\|[v]\|_\sim$, 420
$B(E, F)$, $B(E)$, E^*, 423
$\dot{v}$, 425
$M(K)$, 433
$\|\mu\|$, 433
$\langle v, w \rangle$, 436
$v \perp w$, 437
$E^\perp$, 437
P_E, 439
ϕ_w, 440
T^*, 440
I, 441
$\tilde{m}$, 442
$\mathcal{F}$, 442
$\hat{f}$, 442
$f * g$, 443
$\tau_{\vec{y}} f$, 443
$\mathcal{S}_n$, 443
$\mathbb{H}^n$, 449
∂M, 449
U_λ, ϕ_λ, 449
$T_x M$, 450
$df|_x$, 450
$C^\infty(M)$, 451
$Df|_x$, 453
K^{int}, 454
$\mathrm{supp}(f)$, 454
$(\mathbb{R}^n)^*$, 456
$\vec{y}|\vec{x}$, 456
$\check{A}$, 456
$T_x^* M$, 456
$\tilde{e}_i$, 457
$\mathbb{R}^n \otimes \cdots \otimes \mathbb{R}^n$, 457
$\vec{x}^1 \otimes \cdots \otimes \vec{x}^p$, 457
$A^{(p,q)}$, 457
$T_x^{(p,q)} M$, 457
TM, T^*M, $T^{(p,q)}M$, 458
$\mathcal{T}|_x$, $X|_x$, $\omega|_x$, 458
$\mathfrak{X}M$, ΛM, 458
$f^*\mathcal{T}$, 459
$\mathcal{S}$, 459
$\mathcal{A}$, 460
$\alpha_1 \wedge \alpha_2$, 460
$\Lambda^q M$, 460
df, 460

$d : \Lambda^{q-1}M \to \Lambda^q M$, 461
$[X, Y]_x$, $[X, Y]$, 463
$\int_M \omega$, 466
$\int_\gamma (f\,dx + g\,dy)$, 467
$\langle v, w\rangle_g$, $\|v\|_g$, 468
$\langle X, Y\rangle_g$, $\|X\|_g$, 468
$L(\gamma)$, 468
$d_\lambda(x, y)$, $d_g(x, y)$, 470
E_i, 472
∇, $\nabla_X Y$, 472
$T(X, Y)$, 472
$\nabla_v Y|_x$, 473
Γ^k_{ij}, 475
$\nabla_{\gamma'} v$, 475
$\mathrm{Exp}_x(rv)$, 476

Index

0-chain, 397
$C^\infty(M)$-linearity, 472
Δ_0-formula, 149
Γ_0, 158
$\perp$-release, 67–69, 78
ϵ-net, 217
ϵ-small class, 337
$\forall$-capture, 77, 141, 149, 151
γ_n-predicate, 151
λ-rational point, 469
λ-slice, 176
σ-compactness, 217, 241
1-1 correspondence, 186
1-1 function, 186

abelian group, 253
abelianization, 394
absolute continuity, 345, 349–350
absolute convergence, 214
absolute value of a measure, 349
absolutely summable, 407
abstract object, xii, 10, 12, 16, 19–40, 93
abstract simplicial complex, 384
adjoint operator, 440
affine connection, 472–479
affine function, 374
algebraic closure, 316
algebraic element, 309
algebraic over F, 309
algebraically closed field, 316
almost everywhere, 356
amalgamated free product, 387
analytic function, 363
anomalous sentence, 78, 82, 83, 88, 90
antilinear map, 440
antisymmetric relation, 185
antisymmetric tensor, 460
antisymmetrizing map, 460
Apple II computer, 22–24
approximation lemma, 418
Arc de Triomphe, 34, 38
arithmetical soundness, viii, 121–126
assertibility, 61–78, 95, 111, 177–181
assertible liar paradox, 68, 70, 77–78, 95–110
associate element, 273
atemporal process, 9
atlas, 449
ATM, 101–110, 157
    atomic sentence of, 101
    term of, 101
atom of a Boolean algebra, 282
atomic formula, 166
Austin, J. L., 19
axiom
    choice, 6
    extensionality, 6
    foundation, 6
    pairing, 5

power set, 6, 8–9
union, 6
axiom of choice, 6
axioms for assertibility, 74–78

Bézout's identity, 265
baby Whitney embedding theorem, 455
Baire category theorem, 211
Banach fixed-point theorem, 205
Banach isomorphism theorem, 419
Banach space, 407–416
dual, 421–428
Banach subspace, 407
Banach-Alaoglu theorem, 424
Banach-Steinhaus theorem, 420
barycenter, 375
barycentric subdivision, 375
base point, 378
based vector space, 294
Basic Law V, 23
basic measurable classes, 337
basic open classes, 229
basis
full, 431
measurable, 336–337
topological, 229
vector space, 294
Bell's inequality, 57
belonging
to a class, 172
to a pseudoclass, 176
Benacerraf, Paul, 30
Bessel's inequality, 439
bijection, 186
Bishop, Errett, xii
bit, 171
bitstring, 171
bitstring modulo equivalence, 171
Black, Max, 20
block of an equivalence relation, 191
Bolzano-Weierstrass theorem, 220
Boolean algebra, 281–292
Boolean ring, 283
Boolean subalgebra, 283
Boolos, George, 7, 9
Borel measure, 430
Borel premeasure, 431
borogrove, 179
bound variable, 166
boundary of a manifold, 449
boundary operator, 398
boundary word, 391
bounded compactness, 217
bounded linear transformation, 416–421
bounded quantifier, 183
Bridges, Douglas, xii
Brouwer's fixed point theorem
for $\mathbb{D}^2$, 383
for $\mathbb{D}^n$, 405
Brouwer, L. E. J., xi
bump function, 451

c-Lipschitz function, 203
Calkin algebra, 129
Cantor space, 243
middle thirds, 162
Cantor, Georg, 6, 133
Cantorianism, 29, 36, 93, 111, 114, 118–126, 148
capture law, 66–69, 71
as a scheme, 66, 97, 180
predicate form, 76–77
carrier of a simplicial complex, 374
Cauchy completion, 206
Cauchy sequence, 205
Cauchy's formula, 367
Cauchy's theorem, 366
Cauchy-Riemann equations, 363
Cauchy-Schwarz inequality, 204, 437
Cavalieri's principle, 356
Cayley's theorem, 256
center of a ball, 207
chain, 397
chain homotopy, 402
chain map, 399
chain rule, 333
change of variables theorem, 359
character string, 26–29, 32, 34–38, 45–46, 74, 93, 136
chart, 449

coherently oriented, 464
positively oriented, 465
smooth, 450
child of a vertex, 39
Christoffel symbols, 475
class, viii–ix, xiii, 3, 16, 20, 26, 41–43, 79, 90–93, 111, 121, 137–139, 147, 171, 174, 181–184
empty, 173
proper, viii, x, 11, 13, 15, 17, 58–59, 70, 91, 94, 119, 126–127, 137–139, 145, 174, 178
class comprehension, 172
classification of compact surfaces, 396
classification of finitely generated abelian groups, 325
clopen subspace, 233
closed ball, 207
closed class, 207, 229
closed graph theorem, 419
closed path, 364
closure, 211, 230
code, 176
exact, 176
coextensive (set and class), 92, 93
Cohen, Paul, 127
coherent density, 249
coherent edges, 391
coherent product topology, 250
coherent topology, 244–252
coherent weak* topology, 424
coherently oriented charts, 464
collection, 4, 12, 39, 41–42, 119, 121
commutative ring, 268
compact coherence, 245
compact filtration, 245
compactly coherent
group, 254
module, 319
ring, 268
vector space, 421
compactness, 217, 241
compatibility (of topology with group structure), 254
compatibility (of topology with vector space structure), 421
compatible charts, 449
complement, class-theoretic, 173
complete metric space, 205
complex analysis, 362–369
complex conjugate, 198
complex measure, 346
complex plane, 198
compositionality, 79, 85
comprehension, 4–6, 15, 124, 144
full, 4, 5
computable well-ordering, 148
concatenation, 378–379
concept, viii, 3, 5, 26, 79, 87–90, 111, 112, 127
constructive, 89
conceptual afterimage, 10
concrete object, 19–20, 22–30, 35, 36, 40
cone, 398–399
conjugate exponent, 411
conjunction law, 55, 56, 86–87
conjunction rule, 103
connected
simplicial complex, 379
connected space, 233
connection, 472–479
Levi-Civita, 475
Riemannian, 472
torsion-free, 472
conservative extension, xiii, 31–33, 35
constructive countablism, vii–xv, 65, 111, 136, 142, 145, 148–158, 161–166
constructive truth, 50, 59, 61, 95
constructivism, vii, ix–x, xii, xiii, 8, 57, 61–78, 91, 145–147, 178
objects, 9, 14–15
truth, vii, xii, 14–15, 65, 66, 111, 114, 137, 142, 164–166
containment, 172
proper, 172
contiguity, 378, 401
continuity, 212, 232

at x, 329
continuous differentiability, 333
continuum hypothesis, 6, 127, 130
contradiction, proof by, 167
contravariant tensor field, 458
convergence
metric, 205
topological, 229
uniform, 217
convergent sequence, 205, 229
convergent series, 214
convex combination, 373
strict, 373
convex hull, 373
convexity, 234
convolution, 443
coordinate of a sequence, 188
coordinate projection, 190
cotangent bundle, 458
countability, 163, 174
countable additivity, 340
countable choice, 177, 181
countable subadditivity, 340
countablism, vii–xi, xiii, 29, 36, 93, 111–139, 141–142, 147, 148, 161–164
course, 378
trivial, 378
covariant tensor field, 458
covector, 456
covector field, 458
covector space, 456
cover, 383
universal, 383
covering (by open classes), 217
covering map, 383
cross-cap surface, 394
crossing pair, 304
cumulative hierarchy, 8–12
cut, 195
cyclic group, 265
cyclic module, 320

Das Kontinuum, xi, 112–113, 126
Davies machine, 117–120
Davies, E. B., 117
de Morgan laws, 282
deck transformation, 383
Dedekind cut, 195
defining characteristic, 3
definite, 164, 171
identity criterion, 26
object, 20, 26, 34–35, 92, 136, 137, 147
predicate, 84, 90, 92, 137–139, 142, 144, 145
definitely applicable predicate, 90, 92
deflationism, 54
Deimos, 30
dense
in itself, 199
in order, 199
metrically, 207
topologically, 229
dense embedding, 206
dependent choice, 177, 181
dependent vectors, 294
derivative, 329, 331, 363, 453
partial, 331
descendant of a vertex, 39
determinant, 305
dialetheism, 68, 97
diffeomorphism, 453
difference, class-theoretic, 173
differentiable at x, 331
differentiable function, 329, 331, 362
differential form, 456–463
dimension of a vector space, 302
direct limit, 194
direction of a line, 30–32
directional derivative, 450
disconnected space, 233
discrete topology, 254
disjoint union, 188
disjointness, 173
distance function, 203
distance to a subspace, 214
divergence theorem, 467
divergent series, 214
domain of a function, 185

dominated convergence theorem, 354
double negation, 72, 73, 77–78, 82, 83, 88, 98, 100, 104–105, 167–169
double negation elimination, 73, 82, 98, 100, 167–169
dual Banach space, 421–428
dual of $\mathbb{R}^n$, 456
dual of $L^1(X)$, 430
dual of $L^p(X)$, 428
Dummett, Michael, 16, 64, 65, 69–70

edge
  coherent, 391
  incoherent, 391
  of a polytope, 374
eigenvalue, 327
eigenvector, 327
element
  of a class, 91, 172
  of a set, 34
elementary p-chain, 398
elementary tensor, 457
elimination rule, 166–167
embedding in cubes, 243
empty class, 173
entire function, 363
entry of a matrix, 296
enumerability, 163, 174
epistemological problem, 32
equivalence relation, 191
equivalent
  atlases, 450
  measures, 344
  topologies, 229
essential sup norm, 414
essentially bounded function, 414
Euclid's lemma, 266
Euclidean domain, 273–274
Euclidean metric, 204
Euclidean norm, 204
Euler's theorem, 267
Euler's totient function, 267
evaluation functional, 425
ex falso, 73–74, 167
exact code, 176
exact pseudoclass, 176
excluded middle, ix, x, 73, 96, 136, 143, 165–170
exhaustive sequence, 174
existence and uniqueness of algebraic closures, 318
existence of maximal ideals, 272
explicit knowledge, 63, 66
explosion, 73–74, 167
exponential function, 363
exponential map, 476
extension
  field, 306
  finite, 306
  Galois, 312
  normal, 312
  of a concept, 3
  of a partial order, 201
extensionality axiom, 6
exterior derivative, 462
extreme value theorem, 219

F-automorphism, 313
F-homomorphism, 310
F-vector space, 293
face of a polytope, 374
falling under, 89
Fatou's lemma, 353
Fermat's last theorem, 113, 162
Fermat's little theorem, 267
Fermat's sum of two squares theorem, 280
field, 271
field extension, 306
  finite, 306
filtration, 244
finite dimensional spectral theorem, 442
finite dimensional vector space, 294
finite extension, 306
finite sequence of classes, 188
finitely generated free module, 321
finitely generated module, 320
finitely measurable class, 337

finitism, 16, 29, 36, 93, 111, 112, 114–118
first isomorphism theorem
  for based vector spaces, 300
  for groups, 263
  for rings, 271
fixed field, 314
fixed point, 205
formalism, vii, x, xi
  radical, 61–62
forming a set, 8–12
formula, 166
foundation axiom, 6
Fourier analysis, 442–447
Fourier inversion theorem, 445
Fourier transform, 442
free product, 387
  amalgamated, 387
free variable, 166
Frege, Gottlob, 3, 23, 30, 123
Fubini's theorem, 357
Fuji apple, 23
full basis, 431
full well-foundedness, 146–147
full well-ordering, 146–158
function, 185
  1-1, 186
  analytic, 363
  differentiable, 329, 363
  entire, 363
  essentially bounded, 414
  exponential, 363
  holomorphic, 363
  indicator, 292
  injective, 186
  integrable, 353
  onto, 186
  simple, 351
  smooth, 370, 449, 450, 453
  subjective, 186
  surjective, 186
fundamental group, 378–390
fundamental theorem of algebra, 317
fundamental theorem of calculus, 355
fundamental theorem of Galois theory, 315

G_δ class, 343
Gödel's incompleteness theorems, 46, 62
Gödel, Kurt, 5, 7, 63
Galois extension, 312
Galois group, 314
"gathering together", 7–12, 119
Gauss, C. F., vii
general position, 373
generating set, 320
genus g cross-cap surface, 394
genus g handlebody, 394
geodesic, 476
gluing lemma, 232
Goldbach's conjecture, 117, 161, 164
Gravity's Rainbow, 21
Green's theorem, 467
grounded sentence, 102, 110
group, 253–265
  p-th homology, 398
  compactly coherent, 254
  cyclic, 265
  fundamental, 378–390
  of p-boundaries, 398
  of p-cycles, 398
  quotient, 262
group operation, 253

H-invariance, 259
H-scheme, 50–51, 53, 55
Hölder's inequality, 411
Hahn-Banach theorem, 417
Hahn-Jordan decomposition, 347
half-space, 449
Halmos, Paul, 20
Hamkins, Joel, 129–130, 132, 134–136
handlebody, 394
Hausdorff topological space, 235
Heine-Borel theorem, 224
hereditary separability, 222, 229
Heyting, Arend, 70
hidden variables, 57

Hilbert axioms, 75, 102
Hilbert space, 436–442
Hilbert space operator, 440
Hilbert, David, xi, 6
holding, 43, 47–48, 79
holomorphic function, 363
homeomorphism, 212, 232
homology, 397–405
  of a cone, 399
homology group, 398
homomorphism
  of Boolean algebras, 283
  of groups, 255
  of rings, 269
  unital, 269
homotopic rel A, 377
homotopy, 377
  straight line, 378
Honda Odyssey
  transmission problems, 31
Hopf-Rinow theorem, 478
Husserl, Edmund, 3

ideal, 270, 283
  maximal, 272, 288
  maximal closed, 273
  prime, 288
  principal, 270
  proper, 270
identity criterion, 21–26, 35–36, 171
  definite, 26
identity element
  in a group, 253
  in a ring, 268
identity operator, 441
identity theorem, 368
image of a set, 186–187
imaginary part of a complex number, 198
imcomparable elements, 201
implicit function theorem, 336
implicit knowledge, 63, 66
impredicative definition, xi, 12, 15
incoherent edges, 391
indefinite extensibility, 16–17, 28, 93
indefinite object, 88, 91–92, 94
independent vectors, 294
index class, 175
indexed family of subclasses, 175
indicator function, 292
indispensibility argument, 124–125
induction, 38, 143–146
  on $\mathbb{N}$, 135, 136, 141–143, 163, 180
  predicate, 144–146
  set, 144–146
  strong, 142, 143
infimum, 197
infinitary release rule, 156–158
infinite bitstring, 171
infinite differentiability of analytic functions, 369
infinite series, 214–217
inhabited class, 175
injective function, 186
inner product, 333, 436
instantiation, 171
integers, 194
integrable function, 352, 353
integral domain, 273–281
integration, 350–362
  on manifolds, 463–467
interior
  of a class, 454
  of a manifold, 449
  of a simplex, 373
internal direct sum, 320
intersection, 173, 176
interval in $\mathbb{R}$, 233
introduction rule, 166–167
intuitionism, vii, x, xi, 14–15, 91, 114, 137, 166
intuitionistic logic, ix, x, 72–74, 78, 83, 96, 102, 165–170
invariance of dimension, 404
inverse element, 253
inverse function theorem, 335
inverse limit, 190
invertible element, 271
irreducible element, 273
isometric isomorphism, 416
isometry, 203

isomorphic
  Banach spaces, 416
  Boolean algebras, 284
  groups, 255
  rings, 269
  vector spaces, 294
isomorphism
  isometric, 416
  of Banach spaces, 416
  of Boolean algebras, 284
  of groups, 255
  of rings, 269
  of vector spaces, 294
isomorphism theorems
  for based vector spaces, 300
  for compactly coherent vector spaces, 421
  for groups, 263–265
  for rings, 271
iterated stellar subdivision, 375
iterative conception, 7–12, 37, 119, 122

Jacobian of a function, 359
Jané, Ignacio, 132–133
Jeffrey the string, 35
join (in a lattice), 281
join (of cuts), 196
Jordan chain, 326
Jordan form, 326
Jordan matrix, 326

kernel
  of a Boolean algebra homomorphism, 283
  of a group homomorphism, 261
  of a ring homomorphism, 270
knowability of facts, 66–67
knowledge
  explicit, 63, 66
  implicit, 63, 66
König's infinity lemma, 200
KP^i, 148
Krein-Smulian theorem, 425
Kripke, Saul, 102
Kripke-Platek set theory, 148
Kronecker, Leopold, 171

l^∞-metric, 204
Löwenheim-Skolem theorem, 112, 120, 121
Lagrange's theorem, 261
lattice operations, 281
law of excluded middle, ix, x, 73, 96, 136, 143, 165–170
least upper bound principle, 197
Lebesgue decomposition theorem, 350
Lebesgue measure, 342, 345
  normalized, 442
Lebesgue number, 219
left R-module, 319
left translation of a subgroup, 259
Leibniz rule, 451
length of a smooth path, 468
level (in a tree), 200
Levi-Civita connection, 475
liar paradox, ix, 43–49, 54, 57, 61, 68, 78–83, 88, 95–111
Lie bracket, 463
limit of a sequence, 205
limited principle of omniscience
  see LPO, 170
line integral, 364, 467
linear combination, 294
linear functional, 417
linear subspace, 294
linear transformation, 294
linguistic object, x, 20, 35–37, 113, 178, 181
Liouville's theorem, 368
Lipschitz function, 203
list of names, 34–38, 40, 42, 93, 119–122, 134, 136, 137
local compactness, 217, 241
logicism, xi
loop, 364
LPO, 170, 181

M_δ class, 342
Malament-Hogarth spacetime, 116–118

manifold, 390
orientable, 464
Riemannian, 468–472
smooth, 449–456
with boundary, 449–456
Mars, 30
matrix, 296
matrix of a permutation, 306
matrix product, 296
maximal
closed ideal, 273
ideal, 272, 288
maximum modulus principle, 368
mean value inequality, 333
mean value theorem, 330
meaning, 50, 69–72
meaningfulness, 44, 54, 57, 75, 79–87, 98–110, 177–181
measurable
basis, 336–337
class, 337, 348
function, 350
measure, 337
Borel, 430
complex, 346
normalized Lebesgue, 442
positive, 346
signed, 346
measure space, 337
measure theory, 336–350
meet (in a lattice), 281
meld, 48
membership criterion, x, 42
Menzel, Christopher, 7, 11
mereological sum, 20, 23, 24
mesh, 375
metaphor, 10
metric, 203
Euclidean, 204
Riemannian, 468
metric convergence, 205
metric space, 203–214
metric subspace, 203
metric topology, 232
mimsiness, 179
minimal polynomial, 309
Minkowski's inequality, 411
modal nominalism, 27–33
modal realism, 27, 28
model existence theorem, 65
module, 319–327
compactly coherent, 319
cyclic, 320
finitely generated, 320
torsion, 320
torsion-free, 321
modulus of a complex number, 198
modus ponens, 103, 167
monic polynomial, 309
Monod, Jacques, 44
monotone convergence theorem, 352
Morera's theorem, 369
multilinear functional, 456
multiplicity theory, 131
multivariable chain rule, 333
multiverse view, 132, 134–136
mutually singular measures, 345, 350

n times differentiable, 330
n-course, 378
n-manifold, *see* manifold
n-manifold with boundary, *see* manifold with boundary
n-route, 378
n-sphere, 390
n-torus, 390
n-tuple, 189
Naimark's problem, 129
natural deduction, 73–74, 80, 166–167, 181
natural number, 30–32, 62–64, 171, 174
negation, 72, 73, 166
nonempty set, 175
nonexpansive function, 203
norm, 407
L^p, 411
l^p, 410
essential sup, 414
Euclidean, 204
sup, 217, 408, 410

uniform, 408
normal extension, 312
normal subgroup, 261
normal topological space, 235
normalized Lebesgue measure, 442
normed vector space, 407
nth order derivative, 330
null class, 337
number, 30–32
number theory, 265–267, 279–281

object
abstract, xii, 10, 12, 16, 19–40, 93
concrete, 19–20, 22–30, 35, 36, 40
definite, 20, 26, 34, 92, 136, 137, 147
indefinite, 88, 91–92, 94
linguistic, x, 20, 35–37, 113, 178, 181
objects constructivism, 9, 14–15
objects platonism, 14
onto function, 186
ontological priority, 10–11
open ball, 207
open class, 207, 229
open disk, 366
open mapping theorem, 418
operator
Hilbert space, 440
projection, 441
self-adjoint, 441
unitary, 441
operator norm, 416
order
of a group element, 261
of a module, 321
of a module element, 321
of a subgroup, 261
order-isomorphism, 199
ordered class, 199
ordered pair, 173, 189
ordered triple, 189
ordinary differential equation, 370–371
orientable manifold, 464
orientation, 397
orientation preserving, 464
orientation reversing, 464
oriented simplex, 397
orthocomplement, 437
orthogonality, 437
orthonormal
basis, 437–438
vectors, 437

p, i, n numbers, 106
(p, q)-tensor, 457
(p, q)-tensor bundle, 458
(p, q)-tensor field, 458
(p, q)-tensor space at x, 457
p-chain, 397
elementary, 398
p-dimensional face, 374
p-integrable function, 411
p-norm, 410, 411
p-simplex, 373
PA, 55, 62, 63, 121, 127, 135–136, 142, 150
pairing axiom, 5
paradox
assertible liar, 68, 70, 77–78, 95–110
liar, ix, 43–49, 54, 57, 61, 68, 78–83, 88, 95–111
release rule, 157
Richard's, 113
Russell's, vii, xii, 3–17, 23, 83
Russell's (class-theoretic), x, 42–43, 92, 111
Russell's (constructive), 83
Russell's (semantic), ix, 3, 15, 44, 47–49, 54, 61, 78, 83, 90, 111
Russell's (set-theoretic), 3, 4, 8, 11, 12, 14–16, 40
Skolem's, 112, 132–134
parity, 304
Parsons, Charles, 7, 8
partial derivative, 331
partial order, 198

partial sum, 214
partition of unity, 455
path, 364
    associated to a route, 378
    closed, 364
Peano arithmetic, *see* PA
Pedersen, Gert, 128
permutation matrix, 306
Phobos, 30
Picard's existence theorem, 370
PID, 274
Plancharel's theorem, 446
platonism, vii, x, xi, 137
    objects, 14
    set-theoretic, vii
    truth, 14
Poincaré, Henri, xi, 12, 112–113
point derivation, 451
pointwise convergence, 289
pointwise sum, 216
polar representation, 198
poltergeist explanation, 10
polygonal surface, 391
polynomial
    minimal, 309
    monic, 309
polynomial ring, 268
polytope, 374
positive Borel premeasure, 431
positive definite tensor, 468
positive measure, 346
positively oriented chart, 465
possible world, 26–29, 36–42, 58–59, 93, 115–122
potentialism, 9, 14
power class, 193
power of a class, 189
power series representation of analytic functions, 367
power set axiom, 6, 8–9
Prawitz, Dag, 70
predicate, 47–49, 166
    definite, 84, 90, 92, 137–139, 142, 144, 145
    definitely applicable, 90, 92
    universally meaningful, 84
predicate induction, 144–146
predicativism, xi–xii, 12–15, 111–149
preimage of a class, 186–187
prime element, 273
prime ideal, 288
prime ideal theorem, 288
principal ideal, 270
principal ideal domain, *see* PID
Principia Mathematica, xi
principle of uniform boundedness, 420
product
    measure, 344
    of classes, 173, 187–191
    of groups, 256
    of rings, 269
    topology, 238
progressive
    predicate, 143
    set, 143
progressive up to γ_n, 151
projection operator, 441
proof by contradiction, 167
proper class, viii, x, 11, 13, 15, 17, 70, 91, 94, 119, 126–127, 137–139, 145, 174, 178
proper containment, 172
proper ideal, 270
proposition, 44–46, 79, 87–91, 111
    constructive, 88
proxy, 32
pseudoclass, 176
    exact, 176
pth homology group, 398
pure set, 37–42, 147
pushforward map, 382

q-form, 460
quantified variable, 166
quantifier
    bounded, 183
    unbounded, 183
quasi-quotation, 52
Quine, W. V. O., 45, 51

Quine-Putnam indispensibility argument, 124–125
quining, 45
quotient, 191–194
  group, 262
  norm, 420
  ring, 270
  space (Banach), 420
  topology, 259

R-module, 319
radical formalism, 61–62
radius of a ball, 207
radius of convergence, 215, 216
Radon-Nikodym derivative, 361
Radon-Nikodym theorem, 360–361
Ramsey, Frank, 13
range of a function, 185
rank-nullity theorem, 302
rapidly decreasing function, 443
ratio test, 215
rational cube, 466
rational numbers, 194
rational point, 469
rationally linked points, 469
real analysis, 329–336
real line, 195
real part of a complex number, 198
real projective plane, 391
reasoning subjunctively, 179
reducible element, 273
reflexive relation, 185
regular topological space, 235
regularity of Lebesgue measure, 343, 345
Reinhardt cardinal, 123
relation, 87–90, 185
  antisymmetric, 185
  reflexive, 185
  symmetric, 185
  transitive, 185
relational phrase, 47–49
relative closedness, 229
relative complement, 173
relative openness, 229
release law, 66–69, 83, 97
  as a rule, 69, 97, 100–101, 103, 141–142, 149, 180
  as a scheme, 67
  infinitary rule, 156–158
  predicate form, 76–77
release rule paradox, 157
repeated vertex, 378
residue of x, 309
restriction
  of a function, 186
  of a relation, 185
revenge problem, ix, 45
reverse mathematics, 125, 126
Richard's paradox, 113
Riemann-Lebesgue lemma, 445
Riemannian connection, 472
Riemannian manifold, 468–472
Riemannian metric, 468
Riesz representation theorem, 440
Riesz-Markov theorem, 434
rigid tree, 39
ring, 267–281
  of polynomials, 268
  compactly coherent, 268
  unital, 268
ring operations, 267–268
Rolle's theorem, 330
root of a tree (graph-theoretic), 39
root of a tree (order-theoretic), 200
root test, 215
rooted tree (graph-theoretic), 39, 146
rooted tree (order-theoretic), 200
Rosser, J. B., 63
route, 378
Russell's paradox, vii, xii, 3–17, 23, 83
  class-theoretic, x, 42–43, 92, 111
  constructive, 83
  semantic, ix, 3, 15, 44, 47–49, 54, 61, 78, 83, 90, 111
  set-theoretic, 3, 4, 8, 11, 12, 14–16, 40
Russell, Bertrand, viii, xi, 3, 12, 15, 41, 112, 114

scalar, 293
scalar field, 293
Schauder basis, 424
schematic proof, 56, 86
Schwartz space, 443
second countability, 229
second isomorphism theorem
for based vector spaces, 300
for groups, 263–264
for rings, 271
Seifert-van Kampen theorem, 388
self-adjoint operator, 441
self-adjoint subspace, 409
self-reproductive property, 15
sentence, 47–49, 166
separability, 207, 229
separable metrizability, 232
separating points, 409
sequence, 188
sequence in X, 173
sequence of classes, 187, 188
finite, 188
sequence space, 173
sequential closure, 210, 229
sequential continuity, 212, 232
sequential pseudoclosure, 249
sequentially closed subclass, 207
set, viii–x, xii–xiii, 3, 5–14, 16, 20, 25–26, 29–30, 33–37, 41, 42, 91, 93, 121, 133–139, 145, 147, 171, 173–174
as a nonexistent bag, 10
pure, 37–42, 147
set induction, 144–146
set-theoretic platonism, vii
set-tree, 39–42, 93, 147
setwise well-foundedness, 146–147
setwise well-ordering, 143–146
Sharlow, Mark, 7, 10
ship of Theseus, 19, 22, 24
sign of a permutation, 304
signed measure, 346
simple function, 351
simplex, 373
simplicial approximation, 375
simplicial approximation theorem, 376
simplicial complex, 373
abstract, 384
simplicial homology, 397–405
simplicial transformation, 374
simply connected, 379
Skolem's paradox, 112, 132–134
Skolem, Thoralf, xi, 112, 114, 132–135
slice, 176
smooth
chart, 450
covector field, 458
embedding, 453
function, 370, 449, 450, 453
immersion, 453
manifold, 449–456
manifold with boundary, 449–456
tensor field, 458
vector field, 458
soundness (arithmetical), viii, 122
soundness(arithmetical), 121–126
span, 298
spanning set, 294
spectral theorem (in finite dimensions), 442
standard basis, 295
standard basis vector, 295
standard form of a simple function, 352
standard topology on a finite field extension, 308
star of a vertex, 375
stellar subdivision, 375
Stokes' theorem, 466
Stone duality, 290
Stone space, 289
Stone-Weierstrass theorem, 409
straight line homotopy, 378
strict convex combination, 373
string of characters, 26–29, 32, 34–38, 45–46, 74, 93, 136
strong induction, 142, 143

structure theorem for finitely generated modules over a PID, 325
subalgebra, 409
subbasis, 232
subclass, 172
subcomplex, 377
subdivision, 375
  barycentric, 375
  iterated stellar, 375
  stellar, 375
subgroup, 255
  normal, 261
subjective function, 186
subjunctive reasoning, 179
subordinate function, 455
subring, 270
  unital, 270
subset, 175
subspace
  metric, 203
  topological, 229
  vector space, 294
successor (in a tree), 200
sum of a series, 214
summable, 407
  absolutely, 407
sup norm, 217, 408, 410
supertask, 116
support of a continuous function, 434
supremum, 197
surface, 390–397
surjective function, 186
surveyability, viii, x, 36
symmetric difference, 337
symmetric relation, 185
symmetric tensor, 459
symmetrizing map, 459
syntactic list, 34–38, 40
system of ordinary differential equations, 370

T-scheme, 43, 50–51, 53–57, 85, 86, 99–100, 179–180
T_1 topological space, 235
Taj Mahal, 34, 38
tangent bundle, 458
tangent space, 450
tangent vector, 450
Tao, Terry, xii
Tarski's undefinability theorem, 43–44
Tarski, Alfred, 43, 55, 56, 179–180
Tarskian catastrophe, 54, 86
Taylor's theorem, 330
tensor, 457
  antisymmetric, 460
  elementary, 457
  positive definite, 468
  symmetric, 459
tensor bundle, 458
tensor field, 458
  contravariant, 458
  covariant, 458
tensor product, 457
tensor space, 457
"there all at once", 9, 13, 16, 28, 40
third isomorphism theorem
  for based vector spaces, 300
  for groups, 264
  for rings, 271
third order arithmetic, xiii, 174
Thomson's lamp, 115–116
Thomson, J. F., 115
Tietze extension theorem, 238
token, xii, 21–24, 30
Tonelli's theorem, 357
topological invariance of simplicial homology, 403
topological space, 228–244
topology of pointwise convergence, 289
topology on a class, 228
torsion
  element, 320
  module, 320
  of a connection, 472
torsion-free connection, 472
torsion-free module, 321
total boundedness, 217
total order, 198
totality, 3, 12, 13, 16, 17, 93

totally disconnected space, 289
totient function, 267
transitive relation, 185
translation of a function, 443
tree (graph-theoretic), 39
    rigid, 39
    rooted, 39
    well-founded, 39
tree (order-theoretic), 200
triangle inequality, 203
triangulable space, 374
triangulation, 374
trichotomy, 196
trivial course, 378
truth, 5, 41–59, 61, 79, 111, 150–158
    constructive, 50, 59, 61, 95
truth constructivism, vii, xii, 14–15, 65–67, 111, 114, 137, 142, 164–166
truth platonism, 14
truth value, vii, ix, x, 14–15, 23, 34, 36, 50, 55, 57–59, 64–67, 71–73
Tychonoff product theorem, 241
type, xii, 21–24, 27, 30

UFD, 274
ultrafinitism, vii, 115
unbounded quantifier, 183
unbounded total order, 199
uncountability, 174
undefinability theorem, 43–44
uniform continuity, 225
uniform convergence, 217
uniform norm, 408
union, 173, 176
union axiom, 6
unique factorization domain, *see* UFD
uniqueness of the countable atomless Boolean algebra, 286
unit vector, 416
unital
    homomorphism, 269
    ring, 268
    subalgebra, 409
    subring, 270
unitary operator, 441
universal class, 174
universal cover, 383
universal property
    of product groups, 257
    of product rings, 269–270
    of product topologies, 239
    of products, 190
    of quotient groups, 262
    of quotient rings, 271
    of quotients, 191–192
    of the Cantor space, 244
universally meaningful predicate, 84, 90
urelement, 37, 39–40
Urysohn metrization theorem, 242
Urysohn's lemma, 236
    for metric spaces, 214
Urysohn, Pavel, 242

$V = L$, 129
Veblen function, 151–158
Veblen hierarchy, 151–158
vector, 293
vector field, 458
vector space, 293–306
    based, 294
    compactly coherent, 421
    normed, 407
velocity vector, 468
vertex set, 373
vicious circle principle, xi–xii, 12–15, 111
von Neumann ordinal, 30–32
von Neumann, John, 6

Wang, Hao, 148
weak topology, 244
weak* topology, 424
    coherent, 424
Weaver, Nadia, 21, 23
Weaver, Pepsi, 29
wedge product, 460
well-founded tree, 39
well-foundedness, 93
    full, 146–147
    setwise, 146–147

well-ordering, xiii, 15, 36, 38, 41, 93, 120, 141–158
  computable, 148
  full, 146–158
  setwise, 143–146
Weyl, Hermann, xi, 112–126
Whitehead's problem, 129
Whitney embedding theorem, 456
Wilson's theorem, 279
Wittgenstein, Ludwig, 19, 21, 23, 164
word in G, 387

Zermelo ordinal, 30–32
Zermelo, Ernst, xi, 3, 5, 120, 132, 135
Zermelo-Frankel set theory, *see* ZFC
zero class of a function, 221
zero divisor, 273
ZFC, viii, 5–12, 33, 55–56, 112, 122–127, 132–136