<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.2</td>
<td>Examples: Kernel and image</td>
<td>80</td>
</tr>
<tr>
<td>6.3</td>
<td>Example: Subgroup generated by a subset</td>
<td>81</td>
</tr>
<tr>
<td>6.4</td>
<td>Example: Subgroups of cyclic groups</td>
<td>82</td>
</tr>
<tr>
<td>6.5</td>
<td>Monomorphisms</td>
<td>84</td>
</tr>
<tr>
<td></td>
<td>Exercises</td>
<td>85</td>
</tr>
<tr>
<td>7.1</td>
<td>Normal subgroups</td>
<td>88</td>
</tr>
<tr>
<td>7.2</td>
<td>Quotient group</td>
<td>89</td>
</tr>
<tr>
<td>7.3</td>
<td>Cosets</td>
<td>90</td>
</tr>
<tr>
<td>7.4</td>
<td>Quotient by normal subgroups</td>
<td>92</td>
</tr>
<tr>
<td>7.5</td>
<td>Example</td>
<td>94</td>
</tr>
<tr>
<td>7.6</td>
<td>kernel $\iff$ normal</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td>Exercises</td>
<td>95</td>
</tr>
<tr>
<td>8.1</td>
<td>Canonical decomposition</td>
<td>97</td>
</tr>
<tr>
<td>8.2</td>
<td>Presentations</td>
<td>99</td>
</tr>
<tr>
<td>8.3</td>
<td>Subgroups of quotients</td>
<td>100</td>
</tr>
<tr>
<td>8.4</td>
<td>$HK/H$ vs. $K/(H \cap K)$</td>
<td>101</td>
</tr>
<tr>
<td>8.5</td>
<td>The index and Lagrange's theorem</td>
<td>102</td>
</tr>
<tr>
<td>8.6</td>
<td>Epimorphisms and cokernels</td>
<td>104</td>
</tr>
<tr>
<td></td>
<td>Exercises</td>
<td>105</td>
</tr>
<tr>
<td>9.1</td>
<td>Actions</td>
<td>108</td>
</tr>
<tr>
<td>9.2</td>
<td>Actions on sets</td>
<td>109</td>
</tr>
<tr>
<td>9.3</td>
<td>Transitive actions and the category $G$-Set</td>
<td>110</td>
</tr>
<tr>
<td></td>
<td>Exercises</td>
<td>113</td>
</tr>
<tr>
<td>10.1</td>
<td>Categorical viewpoint</td>
<td>115</td>
</tr>
<tr>
<td></td>
<td>Exercises</td>
<td>117</td>
</tr>
<tr>
<td>Chapter III</td>
<td>Rings and modules</td>
<td>119</td>
</tr>
<tr>
<td>1.1</td>
<td>Definition of ring</td>
<td>119</td>
</tr>
<tr>
<td>1.2</td>
<td>First examples and special classes of rings</td>
<td>121</td>
</tr>
<tr>
<td>1.3</td>
<td>Polynomial rings</td>
<td>124</td>
</tr>
<tr>
<td>1.4</td>
<td>Monoid rings</td>
<td>126</td>
</tr>
<tr>
<td></td>
<td>Exercises</td>
<td>127</td>
</tr>
<tr>
<td>2.1</td>
<td>Ring homomorphisms</td>
<td>129</td>
</tr>
<tr>
<td>2.2</td>
<td>Universal property of polynomial rings</td>
<td>130</td>
</tr>
<tr>
<td>2.3</td>
<td>Monomorphisms and epimorphisms</td>
<td>132</td>
</tr>
<tr>
<td>2.4</td>
<td>Products</td>
<td>133</td>
</tr>
<tr>
<td>2.5</td>
<td>$\text{End}_{\text{Ab}}(G)$</td>
<td>134</td>
</tr>
<tr>
<td></td>
<td>Exercises</td>
<td>136</td>
</tr>
</tbody>
</table>
§3. Ideals and quotient rings 138
  3.1. Ideals 138
  3.2. Quotients 139
  3.3. Canonical decomposition and consequences 141

Exercises 143

§4. Ideals and quotients: Remarks and examples. Prime and maximal ideals 144
  4.1. Basic operations 144
  4.2. Quotients of polynomial rings 146
  4.3. Prime and maximal ideals 150

Exercises 153

§5. Modules over a ring 156
  5.1. Definition of (left-)R-module 156
  5.2. The category R-Mod 158
  5.3. Submodules and quotients 160
  5.4. Canonical decomposition and isomorphism theorems 162

Exercises 163

§6. Products, coproducts, etc., in R-Mod 164
  6.1. Products and coproducts 164
  6.2. Kernels and cokernels 166
  6.3. Free modules and free algebras 167
  6.4. Submodule generated by a subset; Noetherian modules 169
  6.5. Finitely generated vs. finite type 171

Exercises 172

§7. Complexes and homology 174
  7.1. Complexes and exact sequences 174
  7.2. Split exact sequences 177
  7.3. Homology and the snake lemma 178

Exercises 183

Chapter IV. Groups, second encounter 187

§1. The conjugation action 187
  1.1. Actions of groups on sets, reminder 187
  1.2. Center, centralizer, conjugacy classes 189
  1.3. The Class Formula 190
  1.4. Conjugation of subsets and subgroups 191

Exercises 193

§2. The Sylow theorems 194
  2.1. Cauchy's theorem 194
  2.2. Sylow I 196
  2.3. Sylow II 197
  2.4. Sylow III 199
  2.5. Applications 200

Exercises 202
§3. Composition series and solvability 205
  3.1. The Jordan-Hölder theorem 205
  3.2. Composition factors; Schreier's theorem 207
  3.3. The commutator subgroup, derived series, and solvability 210
Exercises 213
§4. The symmetric group 214
  4.1. Cycle notation 214
  4.2. Type and conjugacy classes in $S_n$ 216
  4.3. Transpositions, parity, and the alternating group 219
  4.4. Conjugacy in $A_n$; simplicity of $A_n$ and solvability of $S_n$ 220
Exercises 224
§5. Products of groups 226
  5.1. The direct product 226
  5.2. Exact sequences of groups; extension problem 228
  5.3. Internal/semidirect products 230
Exercises 233
§6. Finite abelian groups 234
  6.1. Classification of finite abelian groups 234
  6.2. Invariant factors and elementary divisors 237
  6.3. Application: Finite subgroups of multiplicative groups of fields 239
Exercises 240

Chapter V. Irreducibility and factorization in integral domains 243
§1. Chain conditions and existence of factorizations 244
  1.1. Noetherian rings revisited 244
  1.2. Prime and irreducible elements 246
  1.3. Factorization into irreducibles; domains with factorizations 248
Exercises 249
§2. UFDs, PIDs, Euclidean domains 251
  2.1. Irreducible factors and greatest common divisor 251
  2.2. Characterization of UFDs 253
  2.3. PID $\implies$ UFD 254
  2.4. Euclidean domain $\implies$ PID 255
Exercises 258
§3. Intermezzo: Zorn's lemma 261
  3.1. Set theory, reprise 261
  3.2. Application: Existence of maximal ideals 264
Exercises 265
§4. Unique factorization in polynomial rings 267
  4.1. Primitivity and content; Gauss's lemma 268
  4.2. The field of fractions of an integral domain 270
  4.3. $R$ UFD $\implies$ $R[x]$ UFD 273
Exercises 276
§5. Irreducibility of polynomials
5.1. Roots and reducibility
5.2. Adding roots; algebraically closed fields
5.3. Irreducibility in \( \mathbb{C}[x] \), \( \mathbb{R}[x] \), \( \mathbb{Q}[x] \)
5.4. Eisenstein's criterion
Exercises

§6. Further remarks and examples
6.1. Chinese remainder theorem
6.2. Gaussian integers
6.3. Fermat's theorem on sums of squares
Exercises

Chapter VI. Linear algebra

§1. Free modules revisited
1.1. \( R\text{-Mod} \)
1.2. Linear independence and bases
1.3. Vector spaces
1.4. Recovering \( B \) from \( F^R(B) \)
Exercises

§2. Homomorphisms of free modules, I
2.1. Matrices
2.2. Change of basis
2.3. Elementary operations and Gaussian elimination
2.4. Gaussian elimination over Euclidean domains
Exercises

§3. Homomorphisms of free modules, II
3.1. Solving systems of linear equations
3.2. The determinant
3.3. Rank and nullity
3.4. Euler characteristic and the Grothendieck group
Exercises

§4. Presentations and resolutions
4.1. Torsion
4.2. Finitely presented modules and free resolutions
4.3. Reading a presentation
Exercises

§5. Classification of finitely generated modules over PIDs
5.1. Submodules of free modules
5.2. PIDs and resolutions
5.3. The classification theorem
Exercises

§6. Linear transformations of a free module
6.1. Endomorphisms and similarity
6.2. The characteristic and minimal polynomials of an endomorphism
Contents

6.3. Eigenvalues, eigenvectors, eigenspaces 365
Exercises 368

§7. Canonical forms 371
7.1. Linear transformations of free modules; actions of polynomial rings 371
7.2. \(k[t]\)-modules and the rational canonical form 373
7.3. Jordan canonical form 377
7.4. Diagonalizability 380
Exercises 381

Chapter VII. Fields 385

§1. Field extensions, I 385
1.1. Basic definitions 385
1.2. Simple extensions 387
1.3. Finite and algebraic extensions 391
Exercises 397

§2. Algebraic closure, Nullstellensatz, and a little algebraic geometry 400
2.1. Algebraic closure 400
2.2. The Nullstellensatz 404
2.3. A little affine algebraic geometry 406
Exercises 414

§3. Geometric impossibilities 417
3.1. Constructions by straightedge and compass 417
3.2. Constructible numbers and quadratic extensions 422
3.3. Famous impossibilities 425
Exercises 427

§4. Field extensions, II 428
4.1. Splitting fields and normal extensions 429
4.2. Separable polynomials 433
4.3. Separable extensions and embeddings in algebraic closures 436
Exercises 438

§5. Field extensions, III 440
5.1. Finite fields 441
5.2. Cyclotomic polynomials and fields 445
5.3. Separability and simple extensions 449
Exercises 452

§6. A little Galois theory 454
6.1. The Galois correspondence and Galois extensions 454
6.2. The fundamental theorem of Galois theory, I 459
6.3. The fundamental theorem of Galois theory, II 461
6.4. Further remarks and examples 464
Exercises 466

§7. Short march through applications of Galois theory 468
7.1. Fundamental theorem of algebra 468
7.2. Constructibility of regular $n$-gons ........................................ 469
7.3. Fundamental theorem on symmetric functions .......................... 471
7.4. Solvability of polynomial equations by radicals ...................... 474
7.5. Galois groups of polynomials .................................................. 478
7.6. Abelian groups as Galois groups over $\mathbb{Q}$ ......................... 479

Exercises ......................................................................................... 480

Chapter VIII. Linear algebra, reprise ............................................. 483

§1. Preliminaries, reprise ................................................................. 483
1.1. Functors .................................................................................. 483
1.2. Examples of functors ................................................................. 485
1.3. When are two categories ‘equivalent’? ...................................... 487
1.4. Limits and colimits ................................................................. 489
1.5. Comparing functors ................................................................. 492

Exercises ......................................................................................... 496

§2. Tensor products and the Tor functors ........................................ 500
2.1. Bilinear maps and the definition of tensor product ................... 501
2.2. Adjunction with Hom and explicit computations .................... 504
2.3. Exactness properties of tensor; flatness .................................... 507
2.4. The Tor functors. ................................................................. 509

Exercises ......................................................................................... 511

§3. Base change ............................................................................. 515
3.1. Balanced maps ......................................................................... 515
3.2. Bimodules; adjunction again .................................................. 517
3.3. Restriction and extension of scalars ......................................... 518

Exercises ......................................................................................... 520

§4. Multilinear algebra ................................................................... 522
4.1. Multilinear, symmetric, alternating maps ............................... 522
4.2. Symmetric and exterior powers ............................................... 524
4.3. Very small detour: Graded algebra .......................................... 527
4.4. Tensor algebras ........................................................................ 529

Exercises ......................................................................................... 532

§5. Hom and duals ........................................................................... 535
5.1. Adjunction again ................................................................. 536
5.2. Dual modules ........................................................................... 537
5.3. Duals of free modules ............................................................ 538
5.4. Duality and exactness ............................................................. 539
5.5. Duals and matrices; biduality .................................................. 541
5.6. Duality on vector spaces ......................................................... 542

Exercises ......................................................................................... 543

§6. Projective and injective modules and the Ext functors ............... 545
6.1. Projectives and injectives ....................................................... 546
6.2. Projective modules ............................................................... 547
6.3. Injective modules ..................................................................... 548
6.4. The Ext functors 551
6.5. Ext^*(G, Z) 554
Exercises 555

Chapter IX. Homological algebra 559
§1. (Un)necessary categorical preliminaries 560
1.1. Undesirable features of otherwise reasonable categories 560
1.2. Additive categories 561
1.3. Abelian categories 564
1.4. Products, coproducts, and direct sums 567
1.5. Images; canonical decomposition of morphisms 570
Exercises 574
§2. Working in abelian categories 576
2.1. Exactness in abelian categories 576
2.2. The snake lemma, again 578
2.3. Working with 'elements' in a small abelian category 581
2.4. What is missing? 587
Exercises 589
§3. Complexes and homology, again 591
3.1. Reminder of basic definitions; general strategy 591
3.2. The category of complexes 594
3.3. The long exact cohomology sequence 597
3.4. Triangles 600
Exercises 602
§4. Cones and homotopies 605
4.1. The mapping cone of a morphism 605
4.2. Quasi-isomorphisms and derived categories 607
4.3. Homotopy 611
Exercises 614
§5. The homotopic category. Complexes of projectives and injectives 616
5.1. Homotopic maps are identified in the derived category 616
5.2. Definition of the homotopic category of complexes 618
5.3. Complexes of projective and injective objects 619
5.4. Homotopy equivalences vs. quasi-isomorphisms in K(A) 620
5.5. Proof of Theorem 5.9 624
Exercises 626
§6. Projective and injective resolutions and the derived category 628
6.1. Recovering A 629
6.2. From objects to complexes 631
6.3. Poor man's derived category 635
Exercises 638
§7. Derived functors 641
7.1. Viewpoint shift 641
7.2. Universal property of the derived functor 643
7.3. Taking cohomology 645
7.4. Long exact sequence of derived functors 647
7.5. Relating $\mathcal{F}$, $L_i\mathcal{F}$, $R^i\mathcal{F}$ 653
7.6. Example: A little group cohomology 655

Exercises 658

§8. Double complexes 661
8.1. Resolution by acyclic objects 662
8.2. Complexes of complexes 665
8.3. Exactness of the total complex 670
8.4. Total complexes and resolutions 672
8.5. Acyclic resolutions again and balancing Tor and Ext 675

Exercises 677

§9. Further topics 680
9.1. Derived categories 681
9.2. Triangulated categories 683
9.3. Spectral sequences 686

Exercises 695

Index 699
Introduction

This text presents an introduction to algebra suitable for upper-level undergraduate or beginning graduate courses. While there is a very extensive offering of textbooks at this level, in my experience teaching this material I have invariably felt the need for a self-contained text that would start 'from zero' (in the sense of not assuming that the reader has had substantial previous exposure to the subject) but that would impart from the very beginning a rather modern, categorically minded viewpoint and aim at reaching a good level of depth. Many textbooks in algebra brilliantly satisfy some, but not all, of these requirements. This book is my attempt at providing a working alternative.

There is a widespread perception that categories should be avoided at first blush, that the abstract language of categories should not be introduced until a student has toiled for a few semesters through example-driven illustrations of the nature of a subject like algebra. According to this viewpoint, categories are only tangentially relevant to the main topics covered in a beginning course, so they can simply be mentioned occasionally for the general edification of the reader, who will in time learn about them (by osmosis?). Paraphrasing a reviewer of a draft of the present text, 'Discussions of categories at this level are the reason why God created appendices.'

It will be clear from a cursory glance at the table of contents that I think otherwise. In this text, categories are introduced on page 18, after a scant reminder of the basic language of naive set theory, for the main purpose of providing a context for universal properties. These are in turn evoked constantly as basic definitions are introduced. The word 'universal' appears at least 100 times in the first three chapters.

I believe that awareness of the categorical language, and especially some appreciation of universal properties, is particularly helpful in approaching a subject such as algebra 'from the beginning'. The reader I have in mind is someone who has reached a certain level of mathematical maturity—for example, who needs no
special assistance in grasping an induction argument—but may have only been exposed to algebra in a very cursory manner. My experience is that many upper-level undergraduates or beginning graduate students at Florida State University and at comparable institutions fit this description. For these students, seeing the many introductory concepts in algebra as instances of a few powerful ideas (encapsulated in suitable universal properties) helps to build a comforting unifying context for these notions. The amount of categorical language needed for this catalyzing function is very limited; for example, functors are not really necessary in this acclimatizing stage.

Thus, in my mind the benefit of this approach is precisely that it helps a true beginner, if it is applied with due care. This is my experience in the classroom, and it is the main characteristic feature of this text. The very little categorical language introduced at the outset informs the first part of the book, introducing in general terms groups, rings, and modules. This is followed by a (rather traditional) treatment of standard topics such as Sylow theorems, unique factorization, elementary linear algebra, and field theory. The last third of the book wades into somewhat deeper waters, dealing with tensor products and Hom (including a first introduction to Tor and Ext) and including a final chapter devoted to homological algebra. Some familiarity with categorical language appears indispensable to me in order to appreciate this latter material, and this is hopefully uncontroversial. Having developed a feel for this language in the earlier parts of the book, students find the transition into these more advanced topics particularly smooth.

A first version of this book was essentially a careful transcript of my lectures in a run of the (three-semester) algebra sequence at FSU. The chapter on homological algebra was added at the instigation of Ed Dunne, as were a very substantial number of the exercises. The main body of the text has remained very close to the original ‘transcript’ version: I have resisted the temptation of expanding the material when revising it for publication. I believe that an effective introductory textbook (this is Chapter 0, after all...) should be realistic: it must be possible to cover in class what is covered in the book. Otherwise, the book veers into the ‘reference’ category; I never meant to write a reference book in algebra, and it would be futile (of me) to try to ameliorate excellent available references such as Lang’s ‘Algebra’.

The problem sets will give an opportunity to a teacher, or any motivated reader, to get quite a bit beyond what is covered in the main text. To guide in the choice of exercises, I have marked with a ≥ those problems that are directly referenced from the text, and with a ≠ those problems that are referenced from other problems. A minimalist teacher may simply assign all and only the ≥ problems; these do nothing more than anchor the understanding by practice and may be all that a student can realistically be expected to work out while juggling TA duties and two or three other courses of similar intensity as this one. The main body of the text, together with these exercises, forms a self-contained presentation of essential material. The other exercises, and especially the threads traced by those marked with ≠, will offer the opportunity to cover other topics, which some may well consider just as essential: the modular group, quaternions, nilpotent groups, Artinian rings, the Jacobson radical, localization, Lagrange’s theorem on four squares, projective space and
Grassmannians, Nakayama's lemma, associated primes, the spectral theorem for normal operators, etc., are some examples of topics that make their appearance in the exercises. Often a topic is presented over the course of several exercises, placed in appropriate sections of the book. For example, 'Wedderburn's little theorem' is mentioned in Remark III.1.16 (that is: Remark 1.16 in Chapter III); particular cases are presented in Exercises III.2.11 and IV.2.17, and the reader eventually obtains a proof in Exercise VII.5.14, following preliminaries given in Exercises VII.5.12 and VII.5.13. The \( \Rightarrow \) label and perusal of the index should facilitate the navigation of such topics. To help further in this process, I have decorated every exercise with a list (added in square brackets) of the places in the book that refer to it. For example, an instructor evaluating whether to assign Exercise V.2.25 will be immediately aware that this exercise is quoted in Exercise VII.5.18, proving a particular case of Dirichlet's theorem on primes in arithmetic progressions, and that this will in turn be quoted in §VII.7.6, discussing the realization of abelian groups as Galois groups over \( \mathbb{Q} \).

I have put a high priority on the requirement that this should be a self-contained text which essentially crosses all t's and dots all i's, and does not require that the reader have access to other texts while working through it. I have therefore made a conscious effort to not quote other references: I have avoided as much as possible the exquisitely tempting escape route 'For a proof, see ....' This is the main reason why this book is as thick as it is, even if so many topics are not covered in it. Among these, commutative algebra and representation theory are perhaps the most glaring omissions. The first is represented to the extent of the standard basic definitions, which allow me to sprinkle a little algebraic geometry here and there (for example, see §VII.2), and of a few slightly more advanced topics in the exercises, but I stopped short of covering, e.g., primary decompositions. The second is missing altogether. It is my hope to complement this book with a 'Chapter 1' in an undetermined future, where I will make amends for these and other shortcomings.

By its nature, this book should be quite suitable for self-study: readers working on their own will find here a self-contained starting point which should work well as a prelude to future, more intensive, explorations. Such readers may be helped by the following '9-fold way' diagram of logical interdependence of the chapters:
This may however better reflect my original intention than the final product. For a more objective gauge, this alternative diagram captures the web of references from a chapter to earlier chapters, with the thickness of the lines representing (roughly) the number of references:

With the self-studying reader especially in mind, I have put extra effort into providing an extensive index. It is not realistic to make a fanfare for each and every new term introduced in a text of this size by an official 'definition'; the index should help a lone traveler find the way back to the source of unfamiliar terminology.

Internal references are handled in a hopefully transparent way. For example, Remark III.1.16 refers to Remark 1.16 in Chapter III; if the reference is made from within Chapter III, the same item is called Remark 1.16. The list in brackets following an exercise indicates other exercises or sections in the book referring to that exercise. For example, Exercise 3.1 in Chapter I is followed by [5.1, §VIII.1.1, §IX.1.2, IX.1.10]: this alerts the reader that there are references to this problem in Exercise 5.1 in Chapter I, section 1.1 in Chapter VIII, section 1.2 in Chapter IX, and Exercise 1.10 in Chapter IX (and nowhere else).

Acknowledgments. My debt to Lang's book, to David Dummit and Richard Foote's 'Abstract Algebra,' or to Artin's 'Algebra' will be evident to anyone who is familiar with these sources. The chapter on homological algebra owes much to David Eisenbud's appendix on the topic in his 'Commutative Algebra', to Gelfand-Manin's 'Methods of homological algebra', and to Weibel's 'An introduction to homological algebra'. But in most cases it would simply be impossible for me to retrace the original source of an expository idea, of a proof, of an exercise, or of a specific pedagogical emphasis: these are all likely offsprings of ideas from any one of these and other influential references and often of associations triggered by following the manifold strands of the World Wide Web. This is another reason why, in a spirit of equanimity, I resolved to essentially avoid references altogether. In any case, I believe all the material I have presented here is standard, and I only retain absolute ownership of every error left in the end product.
I am very grateful to my students for the constant feedback that led me to write this book in this particular way and who contributed essentially to its success in my classes. Some of the students provided me with extensive lists of typos and outright mistakes, and I would especially like to thank Kevin Meek, Jay Stryker, and Yong Jae Cha for their particularly helpful comments. I had the opportunity to try out the material on homological algebra in a course given at Caltech in the fall of 2008, while on a sabbatical from FSU, and I would like to thank Caltech and the audience of the course for their hospitality and the friendly atmosphere. Thanks are also due to MSRI for hospitality during the winter of 2009, when the last fine-tuning of the text was performed.

A few people spotted big and small mistakes in preliminary versions of this book, and I will mention Georges Elencwajg, Xia Liao, and Mirroslav Yotov for particularly precious contributions. I also commend Arlene O'Sean and the staff at the AMS for the excellent copyediting and production work.

Special thanks go to Ettore Aldrovandi for expert advice, to Matilde Marcolli for her encouragement and indispensable help, and to Ed Dunne for suggestions that had a great impact in shaping the final version of this book.

Support from the Max-Planck-Institut in Bonn, from the NSA, and from Caltech, at different stages of the preparation of this book, is gratefully acknowledged.
Preliminaries: Set theory and categories

Set theory is a mathematical field in itself, and its proper treatment (say via the famous 'Zermelo-Fränkel' axioms) goes well beyond the scope of this book and the competence of this writer. We will only deal with so-called 'naive' set theory, which is little more than a system of notation and terminology enabling us to express precisely mathematical definitions, statements, and their proofs.

Familiarity with this language is essential in approaching a subject such as algebra, and indeed the reader is assumed to have been previously exposed to it. In this chapter we first review some of the language of naive set theory, mainly in order to establish the notation we will use in the rest of the book. We will then get a small taste of the language of categories, which plays a powerful unifying role in algebra and many other fields. Our main objective is to convey the notion of 'universal property', which will be a constant refrain throughout this book.

1. Naive set theory

1.1. Sets. The notion of set formalizes the intuitive idea of 'collection of objects'. A set is determined by the elements it contains: two sets $A$, $B$ are equal (written $A = B$) if and only if they contain precisely the same elements. 'What is an element?' is a forbidden question in naive set theory: the buck must stop somewhere. We can conveniently pretend that a 'universe' of elements is available to us, and we draw from this universe to construct the elements and sets we need, implicitly assuming that all the operations we will explore can be performed within this universe. (This is the tricky point!) In any case, we specify a set by giving a precise recipe determining which elements are in it. This definition is usually put between braces and may consist of a simple, complete, list of elements:

$$A := \{1, 2, 3\}$$
is\footnote{:= \text{is a notation often used to mean that the symbol on the left-hand side is defined by whatever is on the right-hand side. Logically, this is just expressing the equality of the two sides and could just as well be written `='; the extra : is a psychologically convenient decoration inherited from computer science.}} the set consisting of the integers 1, 2, and 3. By convention, the order\footnote{Ordered lists are denoted with round parentheses: (1, 2, 3) is not the same as (1, 3, 2).} in which the elements are listed, or repetitions in the list, are immaterial to the definition. Thus, the same set may be written out in many ways:
\[
\{1, 2, 3\} = \{1, 3, 2\} = \{1, 2, 1, 3, 3, 2, 3, 1, 1, 2, 1, 3\}.
\]

This way of denoting sets may be quite cumbersome and in any case will only really work for \textit{finite} sets. For infinite sets, a popular way around this problem is to write a list in which some of the elements are understood as being part of a pattern—for example, the set of even integers may be written
\[
E = \{\ldots, -2, 0, 2, 4, 6, \ldots\},
\]
but such a definition is inherently ambiguous, so this leaves room for misinterpretation. Further, some sets are simply ‘too big’ to be listed, even in principle: for example (as one hopefully learns in advanced calculus) there are simply too many \textit{real numbers} to be able to ‘list’ them as one may ‘list’ the integers.

It is often better to adopt definitions that express the elements of a set as elements of some larger (and already known) set $S$, satisfying some property $P$. One may then write
\[
A = \{s \in S \mid s \text{ satisfies } P\}
\]
($\in$ means \textit{element of}...) and this is in general precise and unambiguous$^3$.

We will occasionally encounter a variation on the notion of set, called ‘multiset’. A multiset is a set in which the elements are allowed to appear ‘with multiplicity’: that is, a notion for which $\{2, 2\}$ would be distinct from $\{2\}$. The correct way to define a multiset is by means of \textit{functions}, which we will encounter soon (see Example 2.2).

A few famous sets are
\begin{itemize}
  \item $\emptyset$: the \textit{empty set}, containing no elements;
  \item $\mathbb{N}$: the set of \textit{natural numbers} (that is, nonnegative integers);
  \item $\mathbb{Z}$: the set of \textit{integers};
  \item $\mathbb{Q}$: the set of \textit{rational numbers};
  \item $\mathbb{R}$: the set of \textit{real numbers};
  \item $\mathbb{C}$: the set of \textit{complex numbers}.
\end{itemize}

Also, the term \textit{singleton} is used to refer to any set consisting of precisely one element. Thus $\{1\}$, $\{2\}$, $\{3\}$ are different sets, but they are all \textit{singletons}.

Here are a few useful symbols (called \textit{quantifiers}):
\begin{itemize}
  \item $\exists$ means \textit{there exists}... (the \textit{existential} quantifier);
\end{itemize}
• \(\forall\) means for all... (the universal quantifier).

Also, \(\exists!\) is used to mean there exists a unique...

For example, the set of even integers may be written as

\[E = \{a \in \mathbb{Z} \mid (\exists n \in \mathbb{Z}) \ a = 2n\}\]

in words, "all integers \(a\) such that there exists an integer \(n\) for which \(a = 2n\)." In this case we could replace \(\exists\) by \(\exists!\) without changing the set—but that has to do with properties of \(\mathbb{Z}\), not with mathematical syntax. Also, it is common to adopt the shorthand

\[E = \{2n \mid n \in \mathbb{Z}\},\]

in which the existential quantifier is understood.

Being able to parse such strings of symbols effortlessly, and being able to write them out fluently, is extremely important. The reader of this book is assumed to have already acquired this skill.

Note that the order in which things are written may make a big difference. For example, the statement

\[(\forall a \in \mathbb{Z}) (\exists b \in \mathbb{Z}) \ b = 2a\]

is true: it says that the result of doubling an arbitrary integer yields an integer; but

\[(\exists b \in \mathbb{Z}) (\forall a \in \mathbb{Z}) \ b = 2a\]

is false: it says that there exists a fixed integer \(b\) which is 'simultaneously' twice as much as every integer—there is no such thing.

Note also that writing simply

\[b = 2a\]

by itself does not convey enough information, unless the context makes it completely clear what quantifiers are attached to \(a\) and \(b\): indeed, as we have just seen, different quantifiers may make this into a true or a false statement.

1.2. Inclusion of sets. As mentioned above, two sets are equal if and only if they contain the same elements. We say that a set \(S\) is a subset of a set \(T\) if every element of \(S\) is an element of \(T\), in symbols,

\[S \subseteq T.\]

By convention, \(S \subseteq T\) means the same thing: that is (unlike \(< vs. \leq\)), it does not exclude the possibility that \(S\) and \(T\) may be equal. To avoid any confusion, we will consistently use \(\subseteq\) in this book. One adopts \(S \subset T\) to mean that \(S\) is 'properly' contained in \(T\): that is, \(S \subseteq T\) and \(S \not= T\).

We can think of 'inclusion of sets' in terms of logic: \(S \subseteq T\) means that

\[s \in S \implies s \in T\]

(the quantifier \(\forall\)s is understood); that is, "if \(s\) is an element of \(S\), then \(s\) is an element of \(T\);" that is, all elements of \(S\) are elements of \(T\); that is, \(S \subseteq T\) as promised.

Note that for all sets \(S\), \(\emptyset \subseteq S\) and \(S \subseteq S\).

If \(S \subseteq T\) and \(T \subseteq S\), then \(S = T\).
The symbol $|S|$ denotes the number of elements of $S$, if this number is finite; otherwise, one writes $|S| = \infty$. If $S$ and $T$ are finite, then

$$S \subseteq T \implies |S| \leq |T|.$$ 

The subsets of a set $S$ form a set, called the power set, or the set of parts of $S$. For example, the power set of the empty set $\emptyset$ consists of one element: $\{\emptyset\}$. The power set of $S$ is denoted $\mathcal{P}(S)$; a popular alternative is $2^S$, and indeed $|\mathcal{P}(S)| = 2^{|S|}$ if $S$ is finite (cf. Exercise 2.11).

1.3. Operations between sets. Once we have a few sets to play with, we can obtain more by applying certain standard operations. Here are a few:

- $\cup$: the union;
- $\cap$: the intersection;
- $\setminus$: the difference;
- $\amalg$: the disjoint union;
- $\times$: the (Cartesian) product;
- and the important notion of 'quotient by an equivalence relation'.

Most of these operations should be familiar to the reader: for example,

$$\{1, 2, 4\} \cup \{3, 4, 5\} = \{1, 2, 3, 4, 5\}$$

while

$$\{1, 2, 4\} \setminus \{3, 4, 5\} = \{1, 2\}.$$ 

In terms of Venn diagrams of infamous 'new math' memory:

\[
\begin{align*}
S \cup T & \quad S \cap T & \quad S \setminus T \nend{align*}
\]

(the solid black contour indicates the set included in the operation).

Several of these operations may be written out in a transparent way in terms of logic: thus, for example,

$$s \in S \cap T \iff (s \in S \text{ and } s \in T).$$

Two sets $S$ and $T$ are disjoint if $S \cap T = \emptyset$, that is, if no element is 'simultaneously' in both of them.

The complement of a subset $T$ in a set $S$ is the difference set $S \setminus T$ consisting of all elements of $S$ which are not in $T$. Thus, for example, the complement of the set of even integers in $\mathbb{Z}$ is the set of odd integers.

The operations $\amalg$, $\times$, and quotients by equivalence relations are slightly more mysterious, and it is very instructive to contemplate them carefully. We will look
1. Naive set theory

at them in a particularly naive way first and come back to them in a short while when we have acquired more language and can view them from a more sophisticated viewpoint.

1.4. Disjoint unions, products. One problem with these operations is that their output may not be defined as a set, but rather as a set up to isomorphisms of sets, that is, up to bijections. To make sense out of this, we have to talk about functions, and we will do that in a moment.

Roughly speaking, the disjoint union of two sets $S$ and $T$ is a set $S \sqcup T$ obtained by first producing ‘copies’ $S'$ and $T'$ of the sets $S$ and $T$, with the property that $S' \cap T' = \emptyset$, and then taking the (ordinary) union of $S'$ and $T'$. The careful reader will feel uneasy, since this ‘recipe’ does not define one set: whatever it means to produce a ‘copy’ of a set, surely there are many ways to do so. This ambiguity will be clarified below.

Nevertheless, note that we can say something about $S \sqcup T$ even on these very shaky grounds: for example, if $S$ consists of 3 elements and $T$ consists of 4 elements, the reader should expect (correctly) that $S \cup T$ consists of 7 elements.

Products are marred by the same kind of ambiguity, but fortunately there is a convenient convention that allows us to write down ‘one’ set representing the product of two sets $S$ and $T$: given $S$ and $T$, we let $S \times T$ be the set whose elements are the ordered pairs $\mathbf{(4)}$ $(s, t)$ of elements of $S$ and $T$:

$$S \times T := \{(s, t) \text{ such that } s \in S, t \in T\}.$$

Thus, if $S = \{1, 2, 3\}$ and $T = \{3, 4\}$, then

$$S \times T = \{(1, 3), (1, 4), (2, 3), (2, 4), (3, 3), (3, 4)\}.$$

For a more sophisticated example, $\mathbb{R} \times \mathbb{R}$ is the set of pairs of real numbers, which (as we learn in calculus) is a good way to represent a plane. The set $\mathbb{Z} \times \mathbb{Z}$ could be represented by considering the points in this plane that happen to have integer coordinates. Incidentally, it is common to denote these sets $\mathbb{R}^2$, $\mathbb{Z}^2$; and similarly, the product $A \times A$ of a set by itself is often denoted $A^2$.

If $S$ and $T$ are finite sets, clearly $|S \times T| = |S||T|$.

Also note that we can use products to obtain explicit ‘copies’ of sets as needed for the disjoint union: for example, we could let $S' = \{0\} \times S$, $T' = \{1\} \times T$, guaranteeing that $S'$ and $T'$ are disjoint (why?); and there is an evident way to ‘identify’ $S$ and $S'$, $T$ and $T'$. Again, making this precise requires a little more vocabulary.

The operations $\cup$, $\cap$, $\sqcup$, $\times$ extend to operations on whole ‘families’ of sets: for example, if $S_1, \ldots, S_n$ are sets, we write

$$\bigcap_{i=1}^{n} S_i = S_1 \cap S_2 \cap \cdots \cap S_n$$

---

$\mathbf{(4)}$One can define the ordered pair $(s, t)$ as a set by setting $(s, t) = \{s, \{s, t\}\}$: this carries the information of the elements $s, t$, as well as conveying the fact that $s$ is special (= the first element of the pair).
for the set whose elements are those elements which are simultaneously elements of all sets \( S_1, \ldots, S_n \); and similarly for the other operations. But note that while it is clear from the definitions that, for example,

\[
S_1 \cup S_2 \cup S_3 = (S_1 \cup S_2) \cup S_3 = S_1 \cup (S_2 \cup S_3),
\]

it is not so clear in what sense the sets

\[
S_1 \times S_2 \times S_3, \quad (S_1 \times S_2) \times S_3, \quad S_1 \times (S_2 \times S_3)
\]

should be 'identified' (where we can define the leftmost set as the set of 'ordered triples' of elements of \( S_1, S_2, S_3 \), by analogy with the definition for two sets). In fact, again, we can really make sense of such statements only after we acquire the language of functions. However, all such statements do turn out to be true, as the reader probably expects; by virtue of this fortunate circumstance, we can be somewhat cavalier and gloss over such subtleties.

More generally, if \( \mathcal{S} \) is a set of sets, we may consider sets

\[
\bigcup_{S \in \mathcal{S}} S, \quad \bigcap_{S \in \mathcal{S}} S, \quad \prod_{S \in \mathcal{S}} S,
\]

for the union, intersection, disjoint union, product of all sets in \( \mathcal{S} \). There are important subtleties concerning these definitions: for example, if all \( S \in \mathcal{S} \) are nonempty, does it follow that \( \prod_{S \in \mathcal{S}} S \) is nonempty? The reader probably thinks so, but (if \( \mathcal{S} \) is infinite) this is a rather thorny issue, amounting to the axiom of choice.

By and large, such subtleties do not affect the material in this course; we will partly come to terms with them in due time\(^5\), when they become more relevant to the issues at hand (cf. §V.3).

1.5. Equivalence relations, partitions, quotients. Intuitively, a relation on elements of a set \( S \) is some special affinity among selections of elements of \( S \). For example, the relation \(<\) on the set \( \mathbb{Z} \) is a way to compare the size of two integers: since \( 2 < 5 \), \( 2 \) 'is related to' \( 5 \) in this sense, while \( 5 \) is not related to \( 2 \) in the same sense.

For all practical purposes, what a relation 'means' is completely captured by which elements are related to which elements in the set. We would really know all there is to know about \(<\) on \( \mathbb{Z} \) if we had a complete list of all pairs \((a, b)\) of integers such that \( a < b \). For example, \((2, 5)\) is such a pair, while \((5, 2)\) is not.

This leads to a completely straightforward definition of the notion of relation: a relation on a set \( S \) is simply a subset \( R \) of the product \( S \times S \). If \((a, b) \in R\), we say that \( a \) and \( b \) are 'related by \( R \)' and write

\[ aRb. \]

Often we use fancier symbols for relations, such as \(<\), \(\leq\), \(=\), \(\sim\), \(\ldots\).

\(^5\)The reader will have to employ the axiom of choice in some exercises every now and then, even before we come back to these issues, but this will probably happen below the awareness level, and so it should.
The prototype of a well-behaved relation is '=' which corresponds to the 'diagonal'

\[ \{(a, b) \in S \times S \mid a = b\} = \{(a, a) \mid a \in S\} \subseteq S \times S. \]

Three properties of this very special relation turn out to be particularly important: if \( \sim \) denotes for a moment the relation = of equality, then \( \sim \) satisfies

- **reflexivity**: \((\forall a \in S) \ a \sim a;\)
- **symmetry**: \((\forall a \in S) \ (\forall b \in S) \ a \sim b \implies b \sim a;\)
- **transitivity**: \((\forall a \in S) \ (\forall b \in S) \ (\forall c \in S), \ (a \sim b \text{ and } b \sim c) \implies a \sim c.\)

That is, every \( a \) is equal to itself; if \( a \) is equal to \( b \), then \( b \) is equal to \( a \); etc.

**Definition 1.1.** An equivalence relation on a set \( S \) is any relation \( \sim \) satisfying these three properties.

In terms of the corresponding subset \( R \) of \( S \times S \), 'reflexivity' says that the diagonal is contained in \( R \); 'symmetry' says that \( R \) is unchanged if flipped about the diagonal (that is, if every \((a, b)\) is interchanged with \((b, a)\)); while unfortunately 'transitivity' does not have a similarly nice pictorial translation.

The datum of an equivalence relation on \( S \) turns out to be equivalent to a type of information which looks a little different at first, that is, a partition of \( S \). A partition of \( S \) is a family of disjoint nonempty subsets of \( S \), whose union is \( S \): for example,

\[ \mathcal{P} = \{\{1, 4, 7\}, \{2, 5, 8\}, \{3, 6\}, \{9\}\} \]

is a partition of the set \[ \{1, 2, 3, 4, 5, 6, 7, 8, 9\} \].

Here is how to get a partition of \( S \) from a relation \( \sim \) on \( S \): for every element \( a \in S \), the equivalence class of \( a \) (w.r.t. \( \sim \)) is the subset of \( S \) defined by

\[ [a]_\sim := \{b \in S \mid b \sim a\}; \]

then the equivalence classes form a partition \( \mathcal{P}_\sim \) of \( S \) (Exercise 1.2).

Conversely (Exercise 1.3) every partition \( \mathcal{P} \) is the partition corresponding in this fashion to an equivalence relation. Therefore, the notions of 'equivalence relation on \( S \)' and 'partition of \( S \)' are really equivalent.

Now we can view \( \mathcal{P}_\sim \) as a set (whose elements are the equivalence classes with respect to \( \sim \)). This is the quotient operation mentioned in §1.3.

**Definition 1.2.** The quotient of the set \( S \) with respect to the equivalence relation \( \sim \) is the set

\[ S/\sim := \mathcal{P}_\sim \]

of equivalence classes of elements of \( S \) with respect to \( \sim \).

**Example 1.3.** Take \( S = \mathbb{Z} \), and let \( \sim \) be the relation defined by

\[ a \sim b \iff a - b \text{ is even.} \]

Then \( \mathbb{Z}/\sim \) consists of two equivalence classes:

\[ \mathbb{Z}/\sim = \{[0]_\sim, [1]_\sim\}. \]
Indeed, every integer $b$ is either even (and hence $b - 0$ is even, so $b \sim 0$, and $b \in \{0\}_\sim$) or odd (and hence $b - 1$ is even, so $b \sim 1$, and $b \in \{1\}_\sim$). This is of course the starting point of modular arithmetic, which we will cover in due detail later on (§II.2.3).

One way to think about this operation is that the equivalence relation 'becomes equality in the quotient': that is, two elements of the quotient $S/\sim$ are equal if and only if the corresponding elements in $S$ are related by $\sim$. In other words, taking a quotient is a way to turn any equivalence relation into an equality. This observation will be further formalized in 'categorical terms' in a short while (§5.3).

Exercises

Exercises marked with a ▶ are referred to from the text; exercises marked with a ▼ are referred to from other exercises. These referring exercises and sections are listed in brackets following the current exercise; see the introduction for further clarifications, if necessary.

1.1. Locate a discussion of Russell's paradox, and understand it.

1.2. ▶ Prove that if $\sim$ is a relation on a set $S$, then the corresponding family $\mathcal{P}_\sim$ defined in §1.5 is indeed a partition of $S$: that is, its elements are nonempty, disjoint, and their union is $S$. [§1.5]

1.3. ▼ Given a partition $\mathcal{P}$ on a set $S$, show how to define a relation $\sim$ on $S$ such that $\mathcal{P}$ is the corresponding partition. [§1.5]

1.4. How many different equivalence relations may be defined on the set $\{1, 2, 3\}$?

1.5. Give an example of a relation that is reflexive and symmetric but not transitive. What happens if you attempt to use this relation to define a partition on the set? (Hint: Thinking about the second question will help you answer the first one.)

1.6. ▶ Define a relation $\sim$ on the set $\mathbb{R}$ of real numbers by setting $a \sim b$ if and only if $b - a \in \mathbb{Z}$. Prove that this is an equivalence relation, and find a 'compelling' description for $\mathbb{R}/\sim$. Do the same for the relation $\approx$ on the plane $\mathbb{R} \times \mathbb{R}$ defined by declaring $(a_1, a_2) \approx (b_1, b_2)$ if and only if $b_1 - a_1 \in \mathbb{Z}$ and $b_2 - a_2 \in \mathbb{Z}$. [§II.8.1, II.8.10]

2. Functions between sets

2.1. Definition. A common thread we will follow for just about every structure introduced in this book will be to try to understand both the type of structures and the ways in which different instances of a given structure may interact.

Sets interact with each other through functions. It is tempting to think of a function $f$ from a set $A$ to a set $B$ in 'dynamic' terms, as a way to 'go from $A$ to $B$'. Similarly to the business with relations, it is straightforward to formalize this notion in ways that do not need to invoke any deep 'meaning' of any given $f$: everything that can be known about a function $f$ is captured by the information of
which element $b$ of $B$ is the image of any given element $a$ of $A$. This information is nothing but a subset of $A \times B$:

$$\Gamma_f := \{(a, b) \in A \times B \mid b = f(a)\} \subseteq A \times B.$$ This set $\Gamma_f$ is the graph of $f$; officially, a function really 'is' its graph\(^6\).

Not all subsets $\Gamma \subseteq A \times B$ correspond to ('are') functions: we need to put one requirement on the graphs of functions, which can be expressed as follows:

$$\left( \forall a \in A \right) \left( \exists! b \in B \right) \quad (a, b) \in \Gamma_f,$$

or ('in functional notation')

$$\left( \forall a \in A \right) \left( \exists! b \in B \right) \quad f(a) = b.$$ That is, a function must send each element $a$ of $A$ to exactly one element of $B$, depending on $a$. 'Multivalued functions' such as $\pm \sqrt{x}$ (which are very important in, e.g., the study of Riemann surfaces) are not functions in this sense.

To announce that $f$ is a function from a set $A$ to a set $B$, one writes $f : A \to B$ or draws the following picture ('diagram'):

$$\begin{array}{c}
A \\ \xrightarrow{f} \\ B
\end{array}$$

The action of a function $f : A \to B$ on an element $a \in A$ is sometime indicated by a 'decorated' arrow, as in

$$a \mapsto f(a).$$

The collection of all functions from a set $A$ to a set $B$ is itself a set\(^7\), denoted $B^A$. If we take seriously the notion that a function is really the same thing as its graph, then we can view $B^A$ as a (special) subset of the power set of $A \times B$.

Every set $A$ comes equipped with a very special function, whose graph is the diagonal in $A \times A$: the identity function on $A$

$$\text{id}_A : A \to A$$

defined by $(\forall a \in A)\ id_A(a) = a$. More generally, the inclusion of any subset $S$ of a set $A$ determines a function $S \to A$, simply sending every element $s$ of $S$ to 'itself' in $A$.

If $S$ is a subset of $A$, we denote by $f(S)$ the subset of $B$ defined by

$$f(S) := \{b \in B \mid (\exists a \in A) b = f(a)\}.$$ That is, $f(S)$ is the subset of $B$ consisting of all elements that are images of elements of $S$ by the function $f$. The largest such subset, that is, $f(A)$, is called the image of $f$, denoted 'im $f$'.

Also, $f|_S$ denotes the 'restriction' of $f$ to the subset $S$: this is the function $S \to B$ defined by

$$(\forall s \in S) : \quad f|_S(s) = f(s).$$

---

\(^6\)To be precise, it is the graph $\Gamma_f$ together with the information of the source $A$ and the target $B$ of $f$. These are part of the data of the function.

\(^7\)This is another 'operation among sets', not listed in §1.3. Can you see why we use $B^A$ for this set? (Cf. Exercise 2.10.)