

Table of contents	1
Chapter 0. About the course, and these notes.	
0.1. Aims and prerequisites.....	5
0.2. Approach.....	5
0.3. A question a day	6
0.4. Homework.....	6
0.5. The name of the game	8
0.6. Other reading	8
0.7. Numeration; advice; web access; request for corrections.....	8
0.8. Acknowledgements	9
Part I. Motivation and examples.	
Chapter 1. Making some things precise.	
1.1. Generalities	11
1.2. What is a group?.....	11
1.3. Indexed sets	13
1.4. Arity.....	13
1.5. Group-theoretic terms	14
1.6. Evaluation	16
1.7. Terms in other families of operations.....	17
Chapter 2. Free groups.	
2.1. Motivation.....	20
2.2. The logician's approach: construction from group-theoretic terms.....	22
2.3. Free groups as subgroups of big enough direct products	25
2.4. The classical construction: free groups as groups of words.....	29
Chapter 3. A Cook's tour of other universal constructions.	
3.1. The subgroup and normal subgroup of G generated by $S \subseteq G $	35
3.2. Imposing relations on a group; quotient groups	36
3.3. Groups presented by generators and relations	38
3.4. Abelian groups, free abelian groups, and abelianizations.....	43
3.5. The Burnside problem.....	45
3.6. Products and coproducts	47
3.7. Products and coproducts of abelian groups	52
3.8. Right and left universal properties.....	53
3.9. Tensor products.....	56
3.10. Monoids	60
3.11. Groups to monoids and back again.....	65
3.12. Associative and commutative rings	66
3.13. Coproducts and tensor products of rings	73
3.14. Boolean algebras and Boolean rings.....	76
3.15. Sets.....	78
3.16. Some algebraic structures we have not looked at.....	79
3.17. The Stone-Ćech compactification of a topological space	79
3.18. Universal covering spaces.....	85

Part II. Basic tools and concepts

Chapter 4. Ordered sets, induction, and the Axiom of Choice.

4.1. Partially ordered sets.....	88
4.2. Digression: preorders.....	95
4.3. Induction, recursion, and chain conditions	98
4.4. The axioms of set theory	104
4.5. Well-ordered sets and ordinals.....	108
4.6. Zorn's Lemma.....	118
4.7. Some thoughts on set theory.....	123

Chapter 5. Lattices, closure operators, and Galois connections.

5.1. Semilattices and lattices.....	125
5.2. 0, 1, and completeness	132
5.3. Closure operators	140
5.4. Digression: a pattern of threes.....	145
5.5. Galois connections	148

Chapter 6. Categories and functors.

6.1. What is a category?.....	154
6.2. Examples of categories	160
6.3. Other notations and viewpoints.....	166
6.4. Universes.....	169
6.5. Functors.....	173
6.6. Contravariant functors, and functors of several variables.....	180
6.7. Category-theoretic versions of some common mathematical notions: properties of morphisms ...	185
6.8. More categorical versions of common mathematical notions: special objects	191
6.9. Morphisms of functors (or ‘‘natural transformations’’).....	200
6.10. Properties of functor categories	207
6.11. Enriched categories (a sketch)	210

Chapter 7. Universal constructions in category-theoretic terms.

7.1. Universality in terms of initial and terminal objects	213
7.2. Representable functors, and Yoneda's Lemma.....	214
7.3. Adjoint functors	220
7.4. Number-theoretic interlude: the p -adic numbers, and related constructions	229
7.5. Direct and inverse limits.....	233
7.6. Limits and colimits	241
7.7. What respects what	247
7.8. Functors respecting limits and colimits	249
7.9. Interaction between limits and colimits	256
7.10. Some existence theorems	261
7.11. Morphisms involving adjunctions	267
7.12. Contravariant adjunctions	271

Chapter 8. Varieties of algebras.

8.1. The category $\Omega\text{-Alg}$274
 8.2. Generating subalgebras from below.....280
 8.3. Terms and left universal constructions282
 8.4. Identities and varieties289
 8.5. Derived operations297
 8.6. Characterizing varieties and equational theories.....300
 8.7. Lie algebras.....307
 8.8. Some necessary trivialities.....312
 8.9. Clones and clonal theories314
 8.10. Structure and Semantics.....324

Part III. More on adjunctions.

Chapter 9. Algebra and coalgebra objects in categories, and functors having adjoints.

9.1. An example: $SL(n)$330
 9.2. Algebra objects in a category333
 9.3. Coalgebra objects in a category336
 9.4. Freyd’s criterion for the existence of left adjoints339
 9.5. Some corollaries and examples.....341
 9.6. Representable endofunctors of **Monoid**346
 9.7. Functors to and from some related categories355
 9.8. Representable functors among categories of abelian groups and modules.....357
 9.9. More on modules: left adjoints of representable functors361
 9.10. Some general results on representable functors, mostly negative.....366
 9.11. A few ideas and techniques369
 9.12. Contravariant representable functors.....372
 9.13. More on commuting operations378
 9.14. Some further reading on representable functors, and on General Algebra.....383
 References.....385
 Word and Phrase Index.....392
 Symbol Index.....409
 List of Exercises414
 end.....422

(Further topics, some of which I hope to add in the future: Monads and the adjoint tower. Many-sorted algebras. Ultraproducts. Normal forms, and the diamond lemma. Quasivarieties. Subdirect products. Profiniteness. Reflective and coreflective subcategories. Various sorts of duality? ... ?)

Chapter 0. About the course, and these notes.

0.1. Aims and prerequisites. This course will develop some concepts and results which occur repeatedly throughout the various areas of algebra, and sometimes in other fields of mathematics as well, and which can provide valuable tools and perspectives to those working in these fields. There will be a strong emphasis on motivation through examples, and on instructive exercises.

I will assume only an elementary background in algebra, corresponding to an honors undergraduate algebra course or one semester of graduate algebra, plus a moderate level of mathematical sophistication. A student who has seen the concept of free group introduced, but isn't sure he or she thoroughly understood it would be in a fine position to begin. On the other hand, anyone conversant with fewer than three ways of proving the existence of free groups has something to learn from Chapters 1-2.

As a general rule, we will pay attention to petty details when they first come up, but take them for granted later on. So students who find the beginning sections devoted too much to "trivia" should be patient!

In preparing this published version of my course notes, I have not removed remarks about homework, course procedures etc., addressed to students who take the course from me at Berkeley, which make up most of the next two and a half pages, since there are some nonstandard aspects to the way I run the course which I thought would be of interest to others. Anyone else teaching from this text should, of course, let his or her students know which, if any, of these instructions apply to them. In any case, I hope readers elsewhere find this aspect of the book more amusing than annoying.

0.2. Approach. Since I took my first graduate course, it has seemed to me that there is something wrong with our method of teaching. Why, for an hour at a time, should an instructor write notes on a blackboard and students copy them into their notebooks – often too busy with the copying to pay attention to the content – when this work could be done just as well by a photocopying machine? If this is all that happens in the classroom, why not assign a text or distribute duplicated notes, and run most courses as reading courses?

One answer is that this is not all that happens in a classroom. Students ask questions about confusing points and the instructor answers them. Solutions to exercises are discussed. Sometimes a result is developed by the Socratic method through discussion with the class. Often an instructor gives motivation, or explains an idea in an intuitive fashion he or she would not put into a written text.

As for this last point, I think one should not be embarrassed to put motivation and intuitive discussion into a text, and I will include a great deal of both in these notes. In particular, I shall often first approach general results through very particular cases. The other items – answering questions, discussing solutions to exercises, etc. – which seem to me to contain the essential human value of class contact, are what classroom time will be spent on in this course, while these duplicated notes will replace the mechanical copying of notes from the board.

Such a system is not assured of success. Some students may be in the habit of learning material through the process of copying it, and may not get the same benefit by reading it. I advise such students to read these notes with a pad of paper in their hands, and try to anticipate details, work out examples, summarize essential points, etc., as they go.

0.3. A question a day. To help the system described above work effectively, I require every student taking this course to hand in, on each day of class, one question concerning the reading for that day. I strongly encourage you to get your question to me by e-mail by 45 minutes before class. If you do, I will try to work the answer into what I say in class that day. If not, then hand it in at the start of class, and I will generally answer it by e-mail if I feel I did not cover the point in class.

The e-mail or sheet of paper with your question should begin with your name, the point in these notes that the question refers to, and the classifying word “urgent”, “important”, “unimportant” or “pro forma”. (I will illustrate this format on the first-day handout.) The first three choices of classifying word should be used to indicate how important it is to you to have the question answered; use the last one if there was nothing in the reading that you really felt needed clarification. In that case, your “pro forma” question should be one that some reader might be puzzled by; perhaps something that puzzled you at first, but that you then resolved. If you give a “pro forma” question, you must give the answer along with it!

You may ask more than one question; you may ask, in addition to your question on the current reading, questions relating to earlier readings, and you are encouraged to ask questions in class as well. But you must always submit in writing at least one question related to the reading for the day.

0.4. Homework. These notes contain a large number of exercises. I would like you to hand in solutions to an average of one or two problems of medium difficulty per week, or a correspondingly smaller number of harder problems, or a larger number of easier problems. Choose problems that are interesting to you. But please, look at *all* the exercises, and at least think about how you would approach them. A minimum of 5 minutes of thought per exercise, except those to which you can see a solution sooner, is a good rule of thumb. We will discuss many of them in class. The exercises are interspersed through the text; you may sometimes prefer to think about them as you come to them, at other times, to come back to them after you finish the section.

Grades will be based largely on homework. The amount of homework suggested above, reasonably well done, will give an A. I will give partial credit for partial results, as long as you show you realize that they are partial. I would also welcome your bringing to the attention of the class interesting related problems that you think of, or that you find in other sources.

It should hardly need saying that a solution to a homework exercise in general requires a proof. If a problem asks you to find an object with a certain property, it is not sufficient to give a description and say, “This is the desired object”; you must prove that it has the property, unless this is completely obvious. If a problem asks whether a calculation can be done without a certain axiom, it is not enough to say, “No, the axiom is used in the calculation”; you must prove that no calculation not using that axiom can lead to the result in question. If a problem asks whether something is true in all cases, and the answer is no, then to establish this you must, in general, give a counterexample.

I am worried that the amount of “handwaving” (informal discussion) in these notes may lead some students to think handwaving is an acceptable substitute for proof. If you read these notes attentively, you will see that handwaving does not *replace* proofs. I use it to guide us to proofs, to communicate my understanding of what is behind some proofs, and at times to abbreviate a proof which is similar to one we have already seen; but in cases of the last sort there is a tacit challenge to you, to think through whether you can indeed fill in the steps. Homework is meant to develop and demonstrate *your* mastery of the material and methods, so it is not a place for you to follow

this model by challenging the instructor to fill in steps!

Of course, there is a limit to the amount of detail you can and should show. Most nontrivial mathematical proofs would be unreadable if we tried to give every substep of every step. So truly obvious steps can be skipped, and familiar methods can be abbreviated. But more students err in the direction of incomplete proofs than of excessive detail. If you have doubts whether to abbreviate a step, think out (perhaps with the help of a scratch-pad) what would be involved in a more complete argument. If you find that the “step” is more complicated than you had thought, then it should *not* be omitted! But bear in mind that “to show or not to show” a messy step may not be the only alternatives – be on the lookout for a simpler argument, that will avoid the messiness.

I will try to be informative in my comments on your homework. If you are still in doubt as to how much detail to supply, come to my office and discuss it. If possible, come with a specific proof in mind for which you have thought out the details, but want to know how much should be written down.

There are occasional exceptions to the principle that every exercise requires a proof. Sometimes I give problems containing instructions of a different sort, such as “Write down precisely the definition of ...”, or “State the analogous result in the case ...”, or “How would one motivate ...?” Sometimes, once an object with a given property has been found, the verification of that property is truly obvious. However, if direct verification of the property would involve 32 cases each comprising a 12-step calculation, you should, if at all possible, find some argument that simplifies or unifies these calculations.

Exercises frequently consist of several successive parts, and you may hand in some parts without doing others (though when one part is used in another, you should if possible do the former if you are going to do the latter). The parts of an exercise may or may not be of similar difficulty – one part may be an easy verification, leading up to a more difficult part; or an exercise of moderate difficulty may introduce an open question. (Open questions, when included, are always noted as such.)

Homework should be legible and well-organized. If a solution you have figured out is complicated, or your conception of it is still fuzzy, outline it first on scratch paper, and revise the outline until it is clean and elegant before writing up the version to hand in.

If you hand in a proof that is incorrect, I will point this out, and it is up to you whether to try to find and hand in a better proof. If, instead, I find the proof poorly presented, I may request that you redo it.

If you want to see the solution to an exercise that we haven’t gone over, ask in class. But I may postpone answering, or just give a hint, if other people still want to work on it. In the case of an exercise that asks you to supply details for the proof of a result in the text, if you cannot see how to do it you should certainly ask to see it done.

You may also ask for a hint on a problem. If possible, do so in class rather than in my office, so that everyone has the benefit of the hint.

If two or more of you solve a problem together and feel you have contributed approximately equal amounts to the solution, you may hand it in as joint work.

If you turn in a homework solution which is inspired by results you have seen in another text or course, indicate this, so that credit can be adjusted to match your contribution.

0.5. The name of the game. The general theory of algebraic structures has long been called Universal Algebra, but in recent decades, many workers in the field have come to dislike this term, feeling that “it promises too much”, and/or that it suggests an emphasis on universal constructions. Universal constructions are a major theme of this course, but they are not all that the field is about.

The most popular replacement term is General Algebra, and I have used it in the title of these notes; but it has the disadvantage that in some contexts it may not be understood as referring to a specific area. Below, I mostly say “General Algebra”, but occasionally refer to the older term.

0.6. Other reading. Aside from these notes, there is no recommended reading for the course, but I will mention here some items in the list of references at the end that you might like to look at. The books [5], [6], [11], [18] and [20] are other general texts in General (a.k.a. Universal) Algebra. Of these, [11] is the most technical and encyclopedic. [18] and [20] are both, like these notes, aimed at students not necessarily having advanced prior mathematical background; however [20] differs from this course in emphasizing *partial* algebras. [6] has in common with this course the viewpoint that this subject is an important tool for algebraists of all sorts, and it gives some interesting applications to groups, division rings, etc..

[29] and [31] are standard texts for Berkeley’s basic graduate algebra course. (Some subset of Chapters 1-6 of the present notes can, incidentally, be useful supplementary reading for students taking such a course.) Though we will not assume the full material of such a course (let alone the full contents of those books), you may find them useful references. [31] is more complete and rigorous; [29] is sometimes better on motivation. [23]-[25] include similar material. A presentation of the core material of such a course at approximately an honors undergraduate level, with detailed explanations and examples, is [26].

Each of [5], [6], [11], [18], [20] and [24] gives a little of the theory of *lattices*, introduced in Chapter 5 of these notes; a thoroughgoing treatment of this subject may be found in [3].

Chapter 6 of these notes introduces *category theory*. [7] is the paper that created that discipline, and still very stimulating reading; [17] is a general text on the subject. [9] deals with an important area of category theory that our course leaves out. For the thought-provoking paper from which the ideas we develop in Chapter 9 come, see [10].

An amusing parody of some of the material we shall be seeing in Chapters 4-9 is [16].

0.7. Numeration; advice; web access; request for corrections. These notes are divided into chapters, and each chapter into sections. In each section, I use two numbering systems: one that embraces lemmas, theorems, definitions, numbered displays, etc., the other for exercises. The number of each item begins with the chapter-and-section numbers; this is followed by a “.” and the number of the result, or a “:” and the number of the exercise. For instance, in § $m.n$, i.e., section n of Chapter m , we might have display $(m.n.1)$, followed by Definition $m.n.2$, followed by Theorem $m.n.3$, and interspersed among these, Exercises $m.n:1$, $m.n:2$, $m.n:3$, etc.. The reason for using a common numbering system for results, definitions, and displays is that it is easier to find Proposition 3.2.9 if it is between Lemma 3.2.8 and display (3.2.10) than it would be if it were Proposition 3.2.3, located between Lemma 3.2.5 and display (3.2.1). The exercises form a separate system. They are listed in the “List of Exercises” at the end of these notes, along with telegraphic descriptions of their subjects.

What makes this “Edition/Printing 2.4”? I count (retroactively) as Edition/Printings 0.1, 0.2, 0.3 etc., the many partial sets of duplicated notes that I gave my classes between 1971 and 1992, gradually covering more and more of the course. Edition/Printings 1.1 and 1.2, which finally

contained Chapter 9, appeared in the Berkeley Lecture Notes series in Spring and Summer 1995 respectively. Edition/Printing 2.1, further rewritten and with 5 more sections, was the first to be published by Henry Helson. I expect to continue the 2.*n* series until I either have a new chapter to add, or change the form of publication.

To other instructors who may teach from these notes (and myself, in case I forget), I recommend moving quite fast through the easy early material and much more slowly toward the end, where there are more concepts new to the students and more nontrivial proofs. Roughly speaking, the hard material begins with Chapter 7. A finer description of the hard parts would be: §§6.9-6.11, 7.3, 7.9-7.12, 8.9-8.10, and Chapter 9. However, this judgement is based on teaching the course to students most of whom have relatively advanced backgrounds. For students who have not seen ordinals or categories before (the kind I had in mind in writing these notes), the latter halves of Chapters 4 and 6 would also be places to move slowly.

The last two sections of each of Chapters 6, 7 and 8 are sketchy (to varying degrees), so students should be expected either to read them mainly for the ideas, or to put in extra effort to work out details.

After many years of editing, reworking, and extending these notes, I know one reason why the copy-from-the-blackboard system has not been generally replaced by the distribution of material in written form: A good set of notes takes an *enormous* amount of time to develop. But I think that it is worth the effort.

Comments and suggestions on any aspect of these notes – organizational, mathematical or other, including indications of typographical errors – are welcomed! They can be sent to me at the address gbergman@math.berkeley.edu, or Department of Mathematics, University of California, Berkeley, CA 94720-3840.

I presently have PostScript files of these notes accessible through my web page, <http://math.berkeley.edu/~gbergman>. I am not sure whether I will keep them there, but I will certainly continue to maintain there a list of errata, and of notes on important changes between one Edition/Printing and the next.

0.8. Acknowledgements. Though I had picked up some category theory here and there, the first extensive development of it that I read was Mac Lane [17], and much of the material on categories in these notes ultimately derives from that book. I can no longer reconstruct which category-theoretic topics I knew before reading [17], but my debt to that work is considerable. Cohn's book [6] was similarly my first exposure to a systematic development of General Algebra; and Freyd's fascinating paper [10] is the source of the beautiful result of §9.4, which I consider the climax of the course. I am also indebted to more people than I can name for help with specific questions in areas where my background had gaps.

These notes are prepared using a locally enhanced version of the text-formatting program *troff*. I am indebted to Ed Moy, then of U. C. Berkeley's Academic Computing Services, for the many useful features he added to it and the many bug-fixes he cheerfully made, and to Fran Rizzardi and D. Mark Abrahams, then of the U. C. Berkeley Statistics Department's BLSS Project, for their adaptation of the program to a new computing environment.

Finally, I am grateful to the many students who have pointed out corrections to these notes – in particular, in the recent years, Arturo Magidin, David Wasserman, Mark Davis, Joseph Flenner, Boris Bukh, Chris Culter, and Lynn Scow.

Part I. Motivation and examples.

In the next three chapters, we shall look at particular cases of algebraic structures and universal constructions involving them, so as to get some sense of the general results we will want to prove in the chapters that follow.

The construction of free groups will be our first example. We will prepare for it in Chapter 1 by making precise some concepts such as that of a group-theoretic expression in a set of symbols; then, in Chapter 2, we will construct free groups by several complementary approaches. Finally, in Chapter 3 we shall look at a large number of other constructions – from group theory, semigroup theory, ring theory, etc. – which have, to greater or lesser degrees, the same spirit as the free group construction, and also, for a bit of variety, two examples from topology.

Chapter 1. Making some things precise.

1.1. Generalities. Most notation will be explained as it is introduced. We will assume familiarity with basic set-theoretic and logical notation: \forall for “for all” (universal quantification), \exists for “there exists” (existential quantification), \wedge for “and”, and \vee for “or”. Functions will be indicated by arrows \rightarrow , while their behavior on elements will be shown by flat-tailed arrows, \mapsto . That is, if a function $X \rightarrow Y$ carries an element x to an element y , this may be symbolized $x \mapsto y$ (“ x goes to y ”). If S is a set and \sim an equivalence relation on S , the set of equivalence classes under this relation will be denoted S/\sim .

We will (with rare exceptions, which will be noted) write functions on the left of their arguments, i.e., $f(x)$ rather than xf , and understand composite functions fg to be defined so that $(fg)(x) = f(g(x))$.

1.2. What is a group? Loosely speaking, a group is a set G given with a *composition* (or *multiplication*, or *group operation*) $\mu: G \times G \rightarrow G$, an *inverse* operation $\iota: G \rightarrow G$, and a *neutral* element $e \in G$, satisfying certain well-known laws. (We will make a practice of saying “neutral element” rather than “identity element” to avoid confusion with the other important meaning of the word “identity”, namely an equation that holds identically.)

The most convenient way to make precise this idea of a set “given with” three operations is to define the group to be, not the set G , but the 4-tuple (G, μ, ι, e) . In fact, from now on, a letter such as G representing a group will stand for such a 4-tuple, and the first component, called the *underlying set* of the group, will be written $|G|$. Thus

$$G = (|G|, \mu, \iota, e).$$

For simplicity, many mathematicians ignore this formal distinction, and use a letter such as G to represent both a group and its underlying set, writing $x \in G$, for instance, where they mean $x \in |G|$. This is okay, as long as one always understands what “precise” statement such a shorthand statement stands for. Note that to be entirely precise, if G and H are two groups, we should use different symbols, say μ_G and μ_H , ι_G and ι_H , e_G and e_H , for the operations of G and H . How precise and formal one needs to be depends on the situation. Since the aim of this course is to abstract the concept of algebraic structure and study what makes these things tick, we shall be somewhat more precise here than in an ordinary algebra course.

(Many workers in General Algebra use a special type-font, e.g., Fraktur or boldface, to represent algebraic objects, and regular type for their underlying sets. Thus, where we will write $G = (|G|, \mu, \iota, e)$, they might write $\mathbf{G} = (G, \mu, \iota, e)$.)

Perhaps the easiest exercise in the course is:

Exercise 1.2:1. Give a precise definition of a homomorphism from a group G to a group H , distinguishing between the operations of G and the operations of H .

We will often refer to a homomorphism $f: G \rightarrow H$ as a “map” from G to H . That is, unless the contrary is mentioned, “maps” between mathematical objects mean maps between their underlying sets which respect their structure. Note that if we wish to refer to a set map not assumed to respect the group operations, we can call this “a map from $|G|$ to $|H|$ ”.

The use of letters $(\mu$ and $\iota)$ for the operations of a group, and the functional notation

$\mu(x, y)$, $\iota(z)$ which this entails, are desirable for precisely stating results in a form which will *generalize* to a wide class of other sorts of structures. But when actually working with elements of a group, we will generally use conventional notation, writing $x \cdot y$ (or xy , or sometimes in abelian groups $x + y$) for $\mu(x, y)$, and z^{-1} (or $-z$) for $\iota(z)$. When we do this, we may either continue to write $G = (|G|, \mu, \iota, e)$, or we may write $G = (|G|, \cdot, ^{-1}, e)$.

Let us now recall the conditions which must be satisfied by a 4-tuple $G = (|G|, \cdot, ^{-1}, e)$, where $|G|$ is a set, “ \cdot ” is a map $|G| \times |G| \rightarrow |G|$, “ $^{-1}$ ” is a map $|G| \rightarrow |G|$, and e is an element of $|G|$, in order for G to be called a group:

$$(1.2.1) \quad \begin{aligned} (\forall x, y, z \in |G|) \quad (x \cdot y) \cdot z &= x \cdot (y \cdot z), \\ (\forall x \in |G|) \quad e \cdot x &= x = x \cdot e, \\ (\forall x \in |G|) \quad x^{-1} \cdot x &= e = x \cdot x^{-1}. \end{aligned}$$

There is another definition of group that you have probably also seen: In effect, a group is defined to be a pair $(|G|, \cdot)$, such that $|G|$ is a set and \cdot is a map $|G| \times |G| \rightarrow |G|$ satisfying

$$(1.2.2) \quad \begin{aligned} (\forall x, y, z \in |G|) \quad (x \cdot y) \cdot z &= x \cdot (y \cdot z), \\ (\exists e \in |G|) \quad ((\forall x \in |G|) \quad e \cdot x &= x = x \cdot e) \wedge ((\forall x \in |G|) \quad (\exists y \in |G|) \quad y \cdot x = e = x \cdot y). \end{aligned}$$

It is easy to show that given $(|G|, \cdot)$ satisfying (1.2.2), there exist a *unique* operation $^{-1}$ and a unique element e such that $(|G|, \cdot, ^{-1}, e)$ satisfies (1.2.1). (Remember the lemmas saying that neutral elements and 2-sided inverses are unique when they exist.) Thus, these two versions of the concept of group provide equivalent information. Our description in terms of 4-tuples may seem “uneconomical” compared with one using pairs, but we will stick with it. We shall eventually see that, more important than the number of terms in the tuple, is the fact that condition (1.2.1) consists only of identities, i.e., universally quantified equations, while (1.2.2) does not. But we will at times acknowledge the idea of the second definition; for instance, when we ask (imprecisely) whether some semigroup “is a group”.

Exercise 1.2.2. (i) If G is a group, let us define an operation δ_G on $|G|$ by $\delta_G(x, y) = x \cdot y^{-1}$. Does the pair $G' = (|G|, \delta_G)$ determine the group $(|G|, \cdot, ^{-1}, e)$? (I.e., if G_1 and G_2 yield the same pair, $G'_1 = G'_2$, must $G_1 = G_2$?)

(ii) Suppose $|X|$ is any set and $\delta: |X| \times |X| \rightarrow |X|$ any map. Can you write down a set of axioms for the pair $X = (|X|, \delta)$, which will be necessary and sufficient for it to arise from a group G in the manner described above? (That is, assuming $|X|$ and δ given, try to find convenient necessary and sufficient conditions for there to exist a group G such that G' , defined as in (i), is precisely $(|X|, \delta)$.)

If you get such a set of axioms, then try to see how brief and simple you can make it.

Exercise 1.2.3. Again let G be a group, and now define $\sigma_G(x, y) = x \cdot y^{-1} \cdot x$. Consider the same questions for $(|G|, \sigma_G)$ that were raised for $(|G|, \delta_G)$ in the preceding exercise.

My point in discussing the distinction between a group and its underlying set, and between groups described using (1.2.1) and using (1.2.2), was not to be petty, but to make us conscious of the various ways we use mathematical language – so that we can use it without its leading us astray. At times we will bow to convenience rather than trying to be consistent. For instance, since we distinguish between a group and its underlying set, we should logically distinguish between the *set* of integers, the *additive group* of integers, the *multiplicative semigroup* of integers, the *ring* of integers, etc.; but we shall in fact write all of these “ \mathbb{Z} ” unless there is a real danger of

ambiguity, or a need to emphasize a distinction. When there is such a need, we can write $(\mathbb{Z}, +, -, 0) = \mathbb{Z}_{\text{add}}$, $(\mathbb{Z}, \cdot, 1) = \mathbb{Z}_{\text{mult}}$, $(\mathbb{Z}, +, \cdot, -, 0, 1) = \mathbb{Z}_{\text{ring}}$, etc.. Likewise, we may use “ready made” symbols for other objects, such as $\{e\}$ for the trivial subgroup of a group G , rather than interrupting a discussion to set up a notation that distinguishes this subgroup from its underlying set.

The approach of regarding sets with operations as tuples, whose first member is the set and whose other members are the operations, applies, as we have just noted, to other algebraic structures than groups – to semigroups, rings, lattices, and the more exotic beasts we will meet on our travels. To be able to discuss the general case, we must make sure we are clear about what we mean by such concepts as “ n -tuple of elements” and “ n -ary operation”. Hence we shall review these in the next two sections.

1.3. Indexed sets. If I and X are sets, an I -tuple of elements of X , or a family of elements of X indexed by I will be defined formally as a function from I to X , but we shall write it $(x_i)_{i \in I}$ rather than $f: I \rightarrow X$. The difference is one of viewpoint. We think of such families as arrays of elements of X , which we keep track of with the help of an index set I , while when we write $f: A \rightarrow B$, we are most often interested in some properties relating an element of A and its image in B . But the distinction is not sharp. Sometimes there is an interesting functional relation between the indices i and the values x_i ; sometimes typographical or other reasons will dictate the use of $x(i)$ rather than x_i .

There will be a minor formal exception to the above definition when we speak of an n -tuple of elements of X ($n \geq 0$). In these beginning chapters, I will take this to mean a function from $\{1, \dots, n\}$ to X , written (x_1, \dots, x_n) or $(x_i)_{i=1, \dots, n}$, despite the fact that set theorists define the natural number n inductively to be the set $\{0, \dots, n-1\}$. Most set theorists, for consistency with that definition, write their n -tuples (x_0, \dots, x_{n-1}) ; and we shall switch to that notation after reviewing the set theorist’s approach to the natural numbers in Chapter 4.

If I and X are sets, then the set of all functions from I to X , equivalently, of all I -tuples of members of X , is written X^I . Likewise, X^n will denote the set of n -tuples of elements of X , defined as above for the time being.

1.4. Arity. An n -ary operation on a set S means a map $f: S^n \rightarrow S$. For $n = 1, 2, 3$ the words are *unary*, *binary*, and *ternary* respectively. If f is an n -ary operation, we call n the *arity* of f . More generally, given any set I , an I -ary operation on S is defined as a map $S^I \rightarrow S$.

Thus, the definition of a group involves one binary operation, one unary operation, and one distinguished element, or “constant”, e . Likewise, a ring can be described as a 6-tuple $R = (|R|, +, \cdot, -, 0, 1)$, where $+$ and \cdot are binary operations on $|R|$, “ $-$ ” is a unary operation, and $0, 1$ are distinguished elements, all satisfying certain identities.

One may make these descriptions more homogeneous in form by considering “distinguished elements” as 0-ary operations of our algebraic structures. Indeed, since an n -ary operation on S is something that turns out a value in S when we feed in n arguments in S , it makes sense that a 0-ary operation should be something that gives a value in S without our feeding it anything. Or, looking at it formally, S^0 is the set of all maps from the empty set to S , of which there is exactly one; so S^0 is a one-element set, so a map $S^0 \rightarrow S$ determines, and is determined by, a single element of S .

We note also that distinguished elements show the right *numerical* behavior to be called “zeroary operations”. Indeed, if f and g are an m -ary and an n -ary operation on S , and i a

positive integer $\leq m$, then on inserting g in the i th place of f , we get an operation $f(-, \dots, -, g(-, \dots, -), -, \dots, -)$ of arity $m+n-1$. Now if, instead, g is an element of S , then when we put it into the i th place of f we get $f(-, \dots, -, g, -, \dots, -)$, an $(m-1)$ -ary operation, as we should if g is to be thought of as an operation of arity $n=0$.

Strictly speaking, distinguished elements and zeroary operations are in one-to-one correspondence, rather than being the same thing: one may distinguish between a map $S^0 \rightarrow S$, and its (unique) value in S . But since they give equivalent information, we can choose between them in setting up our definitions.

So we shall henceforth treat “distinguished elements” in the definition of groups, rings, etc., as zeroary operations, and we will find that they can be handled essentially like the other operations. I say “essentially” because there are some minor ways in which zeroary operations differ from operations of positive arity. Most notably, on the empty set $X = \emptyset$ there is a *unique* n -ary operation for each positive n , but *no* zeroary operation. Sometimes this trivial fact will make a difference in an argument.

1.5. Group-theoretic terms. One is often interested in talking about what *relations* hold among certain elements of a group or other algebraic structure. For example, *every* pair of elements (ξ, η) of a group satisfies the relation $(\xi \cdot \eta)^{-1} = \eta^{-1} \cdot \xi^{-1}$. Some *particular* pair (ξ, η) of elements of some group may satisfy the relation $\xi \cdot \eta = \eta \cdot \xi^2$.

In general, a “group-theoretic relation” in a family of elements $(\xi_i)_I$ of a group G means an equation $p(\xi_i) = q(\xi_i)$ holding in G , where p and q are *expressions* formed from an I -tuple of symbols using formal group operations \cdot , $^{-1}$ and e . So to study relations in groups, we need to define the set of all “formal expressions” in the elements of a set X under symbolic operations of multiplication, inverse and neutral element.

The technical word for such a formal expression is a “term”. Intuitively, a *group-theoretic term* is a set of instructions on how to apply the group operations to a family of elements. E.g., starting with a set of three symbols, $X = \{x, y, z\}$, an example of a group-theoretic term in X is the symbol $(y \cdot x) \cdot (y^{-1})$; or we might write it $\mu(\mu(y, x), \iota(y))$. Whichever way we write it, the idea is: “apply the operation μ to the pair (y, x) , apply the operation ι to the element y , and then apply the operation μ to the pair of elements so obtained, taken in that order”. The idea can be “realized” when we are given a map f of the set X into the underlying set $|G|$ of a group $G = (|G|, \mu_G, \iota_G, e_G)$, say $x \mapsto \xi, y \mapsto \eta, z \mapsto \zeta$ ($\xi, \eta, \zeta \in |G|$). We can then define the result of “evaluating the term $\mu(\mu(y, x), \iota(y))$ using the map f ” as the *element* $\mu_G(\mu_G(\eta, \xi), \iota_G(\eta)) \in |G|$, that is, $(\eta \cdot \xi) \cdot (\eta^{-1})$.

Let us try to make the concept of group-theoretic term precise. “The set of all terms in the elements of X , under formal operations \cdot , $^{-1}$ and e ” should be a set $T = T_{X, \cdot, ^{-1}, e}$ with the following properties:

- (a_X) For every $x \in X$, T contains a symbol representing x .
- (a) For every $s, t \in T$, T contains a “symbolic combination of s and t under \cdot ”.
- (a₋₁) For every $s \in T$, T contains an element gotten by “symbolic application of $^{-1}$ to s ”.
- (a_e) T contains an element symbolizing e .
- (b) Each element of T can be written in *one and only one way* as one and only one of the following:
 - (b_X) The symbol representing an element of X .

- (b.) The symbolic combination of two members of T under \cdot .
 (b₋₁) The symbol representing the result of applying $^{-1}$ to an element of T .
 (b_e) The symbol representing e .
 (c) Every element of T can be obtained from the elements of X via the given symbolic operations. That is, T contains no proper subset satisfying (a_X)-(a_e).

In functional language, (a_X) says that we are to be given a function $X \rightarrow T$ (the “symbol for x ” function); (a) says we have another function, which we call “formal product”, from $T \times T$ to T ; (a₋₁) gives a function $T \rightarrow T$, the “formal inverse”, and (a_e) a distinguished element of T . Translating our definition into this language, we get

Definition 1.5.1. By “the set of all terms in the elements of X under the formal group operations μ , ι , e ” we shall mean a set T which is:

- (a) given with functions

$$\text{symb}_T: X \rightarrow T, \quad \mu_T: T^2 \rightarrow T, \quad \iota_T: T \rightarrow T, \quad \text{and} \quad e_T: T^0 \rightarrow T,$$

such that

- (b) each of these maps is one-to-one, their images are disjoint, and T is the union of these images, and
 (c) T is generated by $\text{symb}_T(X)$ under the operations μ_T , ι_T , and e_T ; i.e., it has no proper subset which contains $\text{symb}_T(X)$ and is closed under those operations.

The next exercise justifies the use of the word “the” in the above definition.

Exercise 1.5:1. Assuming T and T' are two sets given with functions that satisfy Definition 1.5.1, establish a natural one-to-one correspondence between the elements of T and T' . (You must, of course, show that the correspondence you set up is well-defined, and is a bijection. Suggestion: Let $Y = \{(\text{symb}_T(x), \text{symb}_{T'}(x)) \mid x \in X\} \subseteq T \times T'$, let F be the closure of Y under componentwise application of μ , ι and e , and show that F is the graph of a bijection. What properties will characterize this bijection?)

Exercise 1.5:2. Is condition (c) of Definition 1.5.1 a consequence of (a) and (b)?

How can we obtain a set T with the properties of the above definition? One approach is to construct elements of T as finite *strings of symbols* from some alphabet which contains symbols representing the elements of X , additional symbols μ (or \cdot), ι (or $^{-1}$), and e , and perhaps some symbols of punctuation. But we need to be careful. For instance, if we defined μ_T to take a string of symbols s and a string of symbols t to the string of symbols $s \cdot t$, and ι_T to take a string of symbols s to the string of symbols s^{-1} , then condition (b) would not be satisfied! For a string of symbols of the form $x \cdot y \cdot z$ (where $x, y, z \in X$) could be obtained by formal multiplication either of x and $y \cdot z$, or of $x \cdot y$ and z . In other words, μ_T takes the pairs $(x, y \cdot z)$ and $(x \cdot y, z)$ to the same string of symbols, so it is not one-to-one. Likewise, the expression $x \cdot y^{-1}$ could be obtained either as $\mu_T(x, y^{-1})$ or as $\iota_T(x \cdot y)$, so the images of μ_T and ι_T are not disjoint. (It happens that in the first case, the two interpretations of $x \cdot y \cdot z$ come to the same thing in any group, because of the associative law, while in the second, the two interpretations do not: $\xi \cdot (\eta^{-1})$ and $(\xi \cdot \eta)^{-1}$ are generally distinct for elements ξ, η of a group G . But the point is that in both cases condition (b) fails, making these expressions ambiguous as *instructions* for applying group operations. Note that a notational system in which “ $x \cdot y \cdot z$ ” was

ambiguous in the above way could never be used in *writing down* the associative law; and writing down identities is one of the uses we will want to make of these expressions.)

On the other hand, it is not hard to show that by introducing parentheses among our symbols, and letting $\mu_T(s, t)$ be the string of symbols $(s \cdot t)$, and $\iota_T(s)$ the string of symbols (s^{-1}) , we can get a set of expressions satisfying the conditions of our definition.

Exercise 1.5:3. Verify the above assertion. (How, precisely, will you define T ? What assumptions must you make on the set of symbols representing elements of X ? Can we allow some of these “symbols” to be strings of other symbols?)

Another symbolism that will work is to define the value of μ_T at s and t to be the string of symbols $\mu(s, t)$, and the value of ι_T at s to be the string of symbols $\iota(s)$.

Exercise 1.5:4. Suppose we define the value of μ_T on s and t to be the symbol μst , and the value of ι_T at s to be the symbol ιs . Will the resulting set of strings of symbols satisfy Definition 1.5.1?

A disadvantage of the strings-of-symbols approach is that, though it can be extended to other kinds of algebras with finitary operations (such as rings, lattices, etc.), one cannot use it for algebras with operations of *infinite* arities, because, even if one allows infinite strings of symbols, indexed by the natural numbers or the integers, one cannot string two or more such infinite strings together to get another string of the same sort. One can, however, for an infinite set I , create I -tuples which have I -tuples among their members, and this leads to the more versatile *set-theoretic* approach. Let us show it for the case of group-theoretic terms.

Choose any set of four elements, which will be denoted $*$, \cdot , $^{-1}$ and e . For each $x \in X$, define $\text{symb}_T(x)$ to be the ordered pair $(*, x)$; for $s, t \in T$, define $\mu_T(s, t)$ to be the ordered 3-tuple (\cdot, s, t) ; for $s \in T$ define $\iota_T(s)$ to be the ordered pair $(^{-1}, s)$, and finally, define e_T to be the 1-tuple (e) . Let T be the smallest set closed under the above operations.

Now it is a basic lemma of set theory that no element can be written as an n -tuple in more than one way; i.e., if $(x_1, \dots, x_n) = (x'_1, \dots, x'_{n'})$, then $n' = n$ and $x_i = x'_i$ ($i=1, \dots, n$). It is easy to deduce from this that the above construction will satisfy the conditions of Definition 1.5.1.

Exercise 1.5:5. Would there have been anything wrong with defining $\text{symb}_T(x) = x$ instead of $(*, x)$? If so, can you find a way to modify the definitions of μ_T etc., so that the definition $\text{symb}_T(x) = x$ can always be used?

I leave it to you to decide (or not to decide) which construction for group-theoretic terms you prefer to assume during these introductory chapters. We shall only need the *properties* given in Definition 1.5.1. From now on, we shall often use conventional notation for such terms, e.g., $(x \cdot y) \cdot (x^{-1})$. In particular, we shall often identify X with its image $\text{symb}_T(X) \subseteq T_{X, \cdot, ^{-1}, e}$. We will use the more formal notation of Definition 1.5.1 mainly when we want to emphasize particular distinctions, such as that between the formal operations μ_T etc., and the operations μ_G etc. of a particular group.

1.6. Evaluation. Now suppose G is a group, and $f: X \rightarrow |G|$ a set map, in other words, an X -tuple of elements of G . Given a term in an X -tuple of symbols,

$$s \in T = T_{X, \cdot, ^{-1}, e}$$

we wish to say how to *evaluate* s at this family f of elements, so as to get a value $s_f \in |G|$. We shall do this inductively (or more precisely, “recursively”); we will learn the distinction in §4.3).

If $s = \text{symb}_T(x)$ for some $x \in X$ we define $s_f = f(x)$. If $s = \mu_T(t, u)$, then assuming inductively that we have already defined $t_f, u_f \in |G|$, we define $s_f = \mu_G(t_f, u_f)$. Likewise, if $s = \iota_T(t)$, we assume inductively that t_f is defined, and define $s_f = \iota_G(t_f)$. Finally, for $s = e_T$ we define $s_f = e_G$. Since every element $s \in T$ is obtained from $\text{symb}_T(X)$ by the operations μ_T, ι_T, e_T , and in a unique manner, this construction will give one and only one value s_f for each s .

We have not discussed the general principles that allow one to make recursive definitions like the above. We shall develop these in Chapter 4, in preparation for Chapter 8 where we will do rigorously and in full generality what we are sketching here. Some students might want to look into this question for themselves at this point, so I will make this:

Exercise 1.6:1. Show rigorously that the procedure loosely described above yields a unique well-defined map $T \rightarrow |G|$. (Suggestion: Adapt the method suggested for Exercise 1.5:1.)

In the above discussion of evaluation, we fixed $f \in |G|^X$, and got a function $T \rightarrow |G|$. If we vary f as well as T , we get an ‘‘evaluation map’’,

$$(T_{X, \cdot, -1, e}) \times |G|^X \rightarrow |G|$$

taking each pair (s, f) to s_f . Still another viewpoint is to fix an $s \in T$, and define a map $s_G: |G|^X \rightarrow |G|$ by $s_G(f) = s_f$ ($f \in |G|^X$); this represents ‘‘substitution into s ’’. For example, suppose $X = \{x, y, z\}$, let us identify $|G|^X$ with $|G|^3$, and let s be the term $(y \cdot x) \cdot (y^{-1}) \in T$. Then for each group G , s_G is the operation taking each 3-tuple (ξ, η, ζ) of elements of G to the element $(\eta\xi)\eta^{-1} \in G$. Such operations will be of importance to us, so we give them a name.

Definition 1.6.1. Let G be a group and n a nonnegative integer. Let $T = T_{n, -1, \cdot, e}$ denote the set of group-theoretic terms in n symbols. Then for each $s \in T$, we will let $s_G: |G|^n \rightarrow |G|$ denote the map taking each n -tuple $f \in |G|^n$ to the element $s_f \in |G|$. The n -ary operations s_G obtained in this way from terms $s \in T$ will be called the derived n -ary operations of G . (Some authors call these term operations.)

Note that *distinct terms* can induce the same *derived operation*. E.g., the associative law for groups says that for any group G , the derived ternary operations induced by the terms $(x \cdot y) \cdot z$ and $x \cdot (y \cdot z)$ are the same. As another example, in the particular group S_3 (the symmetric group on three elements), the derived binary operations induced by the terms $(x \cdot x) \cdot (y \cdot y)$ and $(y \cdot y) \cdot (x \cdot x)$ are the same, though this is not true in all groups. (It is true in all dihedral groups.)

Some other examples of derived operations on groups are the binary operation of *conjugation*, commonly written $\xi^\eta = \eta^{-1}\xi\eta$ (induced by the term $y^{-1} \cdot (x \cdot y)$), the binary *commutator* operation, $[\xi, \eta] = \xi^{-1}\eta^{-1}\xi\eta$, the unary operation of *squaring*, $\xi^2 = \xi \cdot \xi$, and the two binary operations δ and σ of Exercises 1.2:2 and 1.2:3. Some trivial examples are also worth noting: the *primitive* group operations – group multiplication, inverse, and neutral element – are by definition also *derived* operations; and finally, one has very trivial derived operations such as the ternary ‘‘second component’’ function, $p_{3,2}(\xi, \eta, \zeta) = \eta$, induced by $y \in T_{\{x, y, z\}, -1, \cdot, e}$.

1.7. Terms in other families of operations. The above approach can be applied to more general sorts of algebraic structures. Let Ω be an ordered pair $(|\Omega|, \text{ari})$, where $|\Omega|$ is a set of symbols (thought of as representing operations), and ari a function associating to each $\alpha \in |\Omega|$ a nonnegative integer $\text{ari}(\alpha)$, the intended *arity* of α (§1.4). (For instance, in the group case which we have been considering, we would take $|\Omega| = \{\mu, \iota, e\}$, $\text{ari}(\mu) = 2$, $\text{ari}(\iota) = 1$, $\text{ari}(e) = 0$.

Incidentally, the commonest symbol, among specialists, for the arity of an operation α is $n(\alpha)$, but I will use $\text{ari}(\alpha)$ to avoid confusion with other uses of the letter n .) Then an Ω -algebra will mean a system $A = (|A|, (\alpha_A)_{\alpha \in |\Omega|})$, where $|A|$ is a set, and for each $\alpha \in |\Omega|$, α_A is some $\text{ari}(\alpha)$ -ary operation on $|A|$:

$$\alpha_A: |A|^{\text{ari}(\alpha)} \rightarrow |A|.$$

For any set X , we can now mimic the preceding development to get a set $T = T_{X, \Omega}$, the set of “terms in elements of X under the operations of Ω ”; and given any Ω -algebra A , we can get substitution and evaluation maps as before, and so define *derived operations* of A .

The long-range goal of this course is to study algebras A in this general sense. In order to discover what kinds of results we want to prove about them, we shall devote Chapters 2 and 3 to looking at specific situations involving familiar sorts of algebras. But let me give here a few exercises concerning these general concepts.

Exercise 1.7:1. On the set $\{0, 1\}$, let M_3 denote the ternary “majority vote” operation; i.e., for $a, b, c \in \{0, 1\}$, let $M_3(a, b, c)$ be 0 if two or more of a, b and c are 0, or 1 if two or more of them are 1. One can form various terms in a symbolic operation M_3 (e.g., $p(w, x, y, z) = M_3(x, M_3(z, w, y), z)$) and then evaluate these in the algebra $(\{0, 1\}, M_3)$ to get operations on $\{0, 1\}$ derived from M_3 .

General problem: Determine which operations (of arbitrary arity) on $\{0, 1\}$ can be expressed as derived operations of this algebra.

As steps toward answering this question, you might try to determine whether each of the following can or cannot be so expressed:

- The 5-ary majority vote function $M_5: \{0, 1\}^5 \rightarrow \{0, 1\}$, defined in the obvious manner.
- The binary operation sup . (I.e., $\text{sup}(a, b) = 0$ if $a = b = 0$; otherwise $\text{sup}(a, b) = 1$.)
- The unary “reversal” operation r , defined by $r(0) = 1$, $r(1) = 0$.
- The 4-ary operation N_4 , described as “the majority vote function, where the first voter has extra tie-breaking power”; i.e., $N_4(a, b, c, d) =$ the majority value among a, b, c, d if there is one, while if $a + b + c + d = 2$ we set $N_4(a, b, c, d) = a$.

Advice: (i) If you succeed in proving that some operation s is *not* derivable from M_3 , try to abstract your argument by establishing a general property that all operations derived from M_3 must have, but which s clearly does not have. (ii) A mistake some students make is to think that a formula such as $s(\xi, \eta) = M_3(0, \xi, \eta)$ defines a derived operation. However, since our system $(\{0, 1\}, M_3)$ does not include the *zeroary operation* 0 (nor 1), “ $M_3(0, x, y)$ ” is not a term.

Exercise 1.7:2. (Question raised by Jan Mycielski, letter dated 1/17/83.) Let \mathbb{C} denote the set of complex numbers, and exp the exponential function $\text{exp}(x) = e^x$, a unary operation on \mathbb{C} .

(i) Does the algebra $(\mathbb{C}, +, \cdot, \text{exp})$ have any automorphisms other than the identity and complex conjugation? (An *automorphism* means a bijection of the underlying set with itself, which respects the operations.) I don’t know the answer to this question.

It is not hard to prove using the theory of transcendence bases of fields ([29, §VI.1], [31, §VIII.1]) that the automorphism group of $(\mathbb{C}, +, \cdot)$ is infinite (cf. [29, Exercise VI.6(b)], [31, Exercise VIII.1]). A couple of easy results in the opposite direction, which you may prove and hand in, are

- The algebra $(\mathbb{C}, +, \cdot)$ has no *continuous* automorphisms other than those two mentioned.
- If we write “cj” for the unary operation of complex conjugation, then the algebra $(\mathbb{C}, +, \cdot, \text{cj})$ has no automorphisms other than id and cj .
- A map $\mathbb{C} \rightarrow \mathbb{C}$ is an automorphism of $(\mathbb{C}, +, \cdot, \text{exp})$ if and only if it is an automorphism of $(\mathbb{C}, +, \text{exp})$.

Exercise 1.7:3. Given operations $\alpha_1, \dots, \alpha_r$ (of various arities) on a *finite* set S , and another operation β on S , describe a test that will determine in a finite number of steps whether β is a derived operation of $\alpha_1, \dots, \alpha_r$.

The arities considered so far have been finite; the next exercise will deal with terms in operations of possibly *infinite* arities. To make this reasonable, let me give some naturally arising examples of operations of countably infinite arity on familiar sets:

On the real unit interval $[0, 1]$:

- (a) the operation \limsup (“limit superior”, defined by $\limsup x_i = \lim_{i \rightarrow \infty} \sup_{j \geq i} x_j$),
- (b) the operation defined by $s(a_1, a_2, \dots) = \sum 2^{-i} a_i$.

On the set of real numbers ≥ 1 :

- (c) the continued fraction operation $c(a_1, a_2, \dots) = a_1 + 1/(a_2 + 1/(...))$.

On the class of subsets of the set of integers:

- (d) the operation $\cup a_i$,
- (e) the operation $\cap a_i$.

Exercise 1.7:4. Suppose Ω is a pair $(|\Omega|, \text{ari})$, where $|\Omega|$ is again a set of operation symbols, but where the arities $\text{ari}(\alpha)$ may now be finite or infinite cardinals; and let X be a set of variable-symbols. Suppose we can form a set T of expressions satisfying the analogs of conditions (a)-(c) of §1.5. For $s, t \in T$, let us write $s \succ t$ if t is “immediately involved” in s , that is, if s has the form $\alpha(u_1, u_2, \dots)$ where $\alpha \in |\Omega|$, and $u_i = t$ for *some* i .

- (i) Show that if all the arities $\text{ari}(\alpha)$ are *finite*, then for each s we can find a finite bound $B(s)$ on the lengths n of sequences $s_1, \dots, s_n \in T$ such that $s = s_1 \succ \dots \succ s_n$.
- (ii) If not all $\text{ari}(\alpha)$ are finite, and X is nonempty, show that there exists s for which no such finite bound exists.
- (iii) In the situation of (ii), is it possible to have an infinite chain $s = s_1 \succ \dots \succ s_n \succ \dots$ in T ?
- (iv) Show that one cannot have a “cycle” $s_1 \succ \dots \succ s_n \succ s_1$ in T .

Until we come to Chapter 8, we shall rarely use the word “algebra” in the general sense of this section. But the reader consulting the index should keep this sense in mind, since it used there with reference to general concepts of which we will be considering specific cases in these intervening chapters.