Algebra in Action

Duncan Wilkie

21 November 2023

Preface

If you don't know algebra yet, skip to the roadmap; prefaces are directed at people assigning books.

This book stems from a divine revelation I had while studying the frightening connections between modules over field-valued polynomial rings and linear algebra in the wee hours of a morning my senior year of undergrad. A Discord math friend had demonstrated the clear inadequacy of the geometric intuition for the determinant of a linear transformation, and posed the question of an algebraic motivation for it. Toiling over Dummit and Foote's approach to the structure theory of modules at the time, I quickly found the answer: the product of the roots of the largest invariant factor of the underlying vector space, represented as a module over the polynomial ring with coefficients in the underlying field where the transformation is replaced by an indeterminate.

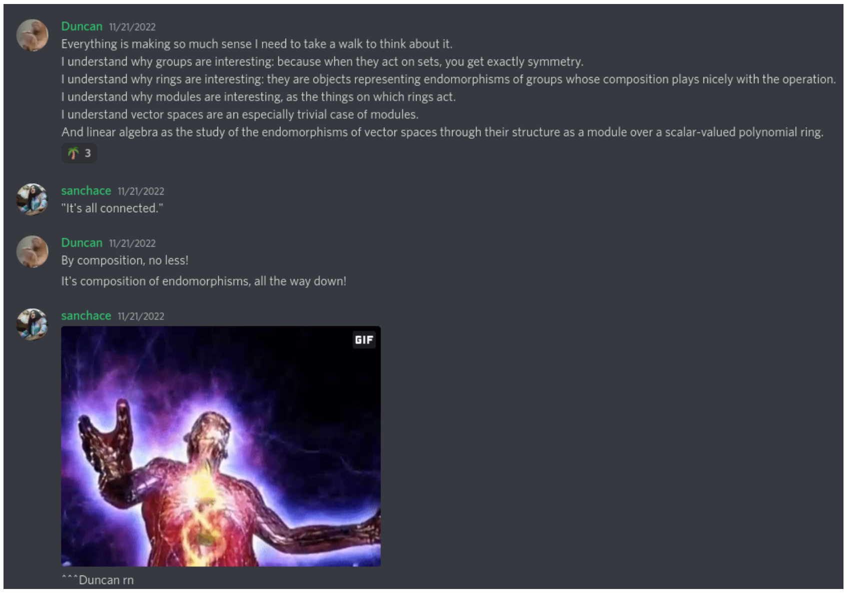
This, alongside the spookily elementary proofs of the Jordan and rational canonical form theorems, threatened my foundational understanding of operator theory, begotten of Axler's text *par excellence*. Surely, despite immediate motivation for the most fundamental concepts emerging from deepest depths, there's something more to the subject? Could it really be, fundamentally, about polynomial rings?

Thinking over how one might deface Axler in this light brought a new perspective on the whole of algebra. You're given a vector space V over a field F, and a linear operator on it, T. What can you do, without giving V the standard agonizing vivisection? Well, you can add any operator expression to any other, via pointwise addition in the vector space, and you can borrow V's scalar multiplication similarly. You can also compose T with itself, forming "powers" of T; this yields polynomials of the form

$$f_0 + f_1T + f_2T^2 + \cdots + f_nT^n, \text{ where} f_i \in F$$

You then investigate the module action of evaluating the polynomial operators in the vector space, to overwhelming effect as described above. The not-so-obvious part of this, at least *ab initio*, is why we should consider opererator composition to be a kind of multiplication. Pondering how to justify considering composition of T as "multiplication," it dawned on me that *every other multiplication is actually composition of endomorphisms*.

The attendant emotional consequences are witnessed in the figure.



Duncan 11/21/2022
Everything is making so much sense I need to take a walk to think about it.
I understand why groups are interesting: because when they act on sets, you get exactly symmetry.
I understand why rings are interesting: they are objects representing endomorphisms of groups whose composition plays nicely with the operation.
I understand why modules are interesting, as the things on which rings act.
I understand vector spaces are an especially trivial case of modules.
And linear algebra as the study of the endomorphisms of vector spaces through their structure as a module over a scalar-valued polynomial ring.
🌴 3

sanchace 11/21/2022
"It's all connected."

Duncan 11/21/2022
By composition, no less!
It's composition of endomorphisms, all the way down!

sanchace 11/21/2022

^^^Duncan rn

This provided an intuitive justification for all my hard-won opinions on proper conventions (e.g. rings not rngs, 0 is a natural number), explained definitions that seemed to fall from the sky (associativity, distributivity, rings are Abelian groups), and provides a more palatable hint towards category theory than "it's what we need to describe functors which are what we need to describe natural transformations."

Practically, this text seeks to be a first course in modern (abstract) algebra from a categorical perspective, correcting the gaps in motivation left by the laudable recent treatments with the same goal. The only prerequisite is an introduction to proofs class covering basic set theory, such as *Book of Proof* or the first chapter of Munkres, save the occasional footnote referencing some other area of mathematics that may be omitted with absolutely no impact on continuity. The background formalism is sumarized in an appendix.

I produce also machine-verified proofs of each result, as an aid in my writing clear and concise arguments and a recourse for the logically-minded reader peeved by my omission of trivial results, or suspicious of my wording. They are in Lean 4, and are in print separately, in addition to freely distributed online. I have made painstaking attempts to establish everything constructively, and any thus-intractable proof is distinguished by ¡a symbol¿ and the nonconstructive step is identified in the text.

How to Read This Book

I write this book to motivate. As such, there's plenty of material in each chapter that's logically unnecessary for its subject, and is there for your philosophical satiety. It's nevertheless my belief that this material is *essential* for an integrated understanding—omit my rambles at your peril.

The text begins with a look at numbers and their history from a mature mathematical perspective. This leads to the core philosophical insight of an **action**: a set whose elements somehow "do something" to another set. In the next part, the idea of an action is made more precise in the context of studying symmetries—following intuitive geometric examples, the abstract definition of a **group** is bootstrapped from laws constraining reasonable interpretations of what "symmetry" means. The third part deals with **rings**, which are groups that act on themselves rather than another set. Equivalently, they describe symmetries that are themselves symmetric; despite seeming like a special requirement, examples in application and other mathematics abound (indeed, arithmetic can be seen as the study of particularly nice rings). In the fourth part, we expound on the observation that in ring theory we often found ourselves almost-but-not-quite repeating our study of groups. The core commonalities are extracted into the notion of a **category**, which provides an even better notion of an action with the idea of a **functor**, and deciding how to tell when two functors are truly different or nothing more than different notation for the same idea (e.g. left/right actions) leads to the **natural transformation**. We also find we can reinterpret the process of going from groups to rings—which seemed to yield richer results—in the new language. This is explored in the final part, where the process is iterated again, forming **modules**: "rings of rings," or "symmetries of symmetries acting on symmetries." We apply our newfound techniques to great effect here. This is the end of the road, though; if you have something that plays nice in the same way rings and modules do (anything with an **Abelian category** structure), then you can always recast it as some kind of module (even though you'd best stay in category-land).

I advise going through this in order. A principal distinction of this book is its development of category-theoretic concepts without requiring of its reader graduate-level familiarity with two or three massive subject areas. The downside of this, however, is comparative over-dependence on the subject areas it does develop. Standalone treatments of category theory (my current favorite is [**?**]) are able to cite dozens of applications of their not-immediately-natural concepts to the most distant areas of mathematics, overwhelming the reader's disquiet with their sheer utility. This is rhetorically effective for the initiate; for the newcomer to formal mathematics, though (welcome!), it's necessary to build those concepts out from scratch, which is what the initial sections do. For some hard evidence, here's a section-level dependency graph of the results save basic set theory and logic:

The situation with examples and problems, by contrast, is more nuanced. I mentally divide these into several categories, based on what I've seen in other pedagogical material; I distinguish them for your convenience. Some examples are throwaway, serving only to illustrate a concept for those finding a mere definition wanting. These I tend to skip, but they're so absolutely helpful when they're necessary that I include them. Do whatever suits you; they're indicated with a ¡symbol here¿. Others are primary mental models for their respective concept: intuitively mapping *any* situation onto the specific case works so well that it's permissible to use the specific as a mental image for the general, and is probably the most practical way to do it. These (indicated by a ¡symbol here¿) I wouldn't omit—I've certainly be guilty of doing it myself, but will appeal to fog of the moment (most texts don't classify like this, and it's hard to tell what's what without foreknowledge of the subject). Still others are critical counterexamples. These demonstrate why and where different mental models for a slippery concept fail, and provide important whetstones against which to throw intuitions, conjectures, and the like. These are absolutely crucial to good understanding: any test worth its ink will ask you to produce one or two, and there are those who have built careers on careful construction of convoluted objects that explode prior understanding[1]. These have a ¡symbol here¿.

Problems in general are the *only* way to truly understand things. Nevertheless, man's finite time makes it likely I'm the only one doing everything, and I classify them for you to prioritize to taste. There are two broad categories: important and trivial problems. These terms are in the sense of likelihood you'll encounter concepts they introduce beyond this book; trivial problems can be quite difficult, but are meant to cultivate skill, and nothing more, whereas important problems ask you to develop the fundamental properties of a construction of genuine mathematical interest. Obviously, the important problems are good candidates for deference—you should at least read them, for nominal familiarity with what they define. In addition, there are different problem difficulties, and different problem targets. The former is fairly obvious: there are easy warm-up problems that merely help embed a definition in your brain (no mark), reasonable problems that involve a nontrivial step or two ($*$), difficult problems with several steps that you need to find yourself ($**$), and open research problems meant as fodder for your daydreams ($***$). Targets indicate what is expected for a solution. These range, in rough order of freedom in your response, from a right answer, a proof, an algorithm, a question, evidence of critical thinking, to multiple of any of the previous. This might correlate with difficulty in practice, independent though it is in theory. For some exercises, my solution requires nonconstructive machinery; this is indicated in the same way as for results in the text. Try to do better than me, but don't worry to much if you can't!

While I've done every problem to ensure they're appropriate, I don't share those solutions, because they will immediately go on libgen, and I feel worse off for having used such pirated solutions on occasion myself. See the appendix for how to determine if a solution is right, even without a key.

[1] For example, it was thought for a good century in topology that continuous structure and smooth structure always agree—that is, homeomorphic $\Leftrightarrow$ diffeomorphic. In fact, not only is this not true, it's not true even for *flat, Euclidean space*! Weirder still, it happens to be true in all dimensions except 4, in which there are uncountably infinitely many incompatible structures! This leads to the mathematical limerick $1, 1, 1, \mathfrak{c}, 1, 1, \ldots$ (which has a Jeapordy answer: "How many distinct manifold structures are there on $\mathbb{R}^n$, for each positive n?")

Contents

Chapter 1

It's All Connected

1.1 Structure

Everything is function composition.

This is mathematics' dirty little secret, usually introduced in a boring corner of a Topology II course for the true believers (graduate students), and even then in a context that seems to reduce it to a computational expedient aiding the dark art of homology theory. This book is an argument for this statement, a tour through the most natural and historically useful guises in which composition presents itself. We'll see this mode of thinking is a useful tool for producing interesting and powerful results about concepts as distant as symmetry, linear algebra, and compass-and-straightedge constructions, and makes otherwise arbitrary definitions seem as natural as breathing. However, you'd be right, at this point, to object to this claim. It's an awfully bold one, and it seems to require total knowledge to substantiate. Let's look a bit deeper at what mathematics is, to assuage these initial concerns, and pave the way for what's to come.

Geometer Michael Attiyah said "mathematics is the science of analogy." It's a science in the sense that it's systematic, and it studies analogies in the sense of precisely-defined abstract concepts. So, less pithily, "mathematics is the systematic study of precise abstraction." Still less pithily, the day-to-day business of mathematics is roughly

1. Have a concept that demands complicated and delicate reasoning,

2. Do some philosophy, and try to locate the core essence of the concept,

3. Encode this core essence symbolically in a formal definition,

4. Perform the complicated and delicate reasoning (almost always) error-free by asking and encoding the natural questions about the definition, proving or disproving theorems (particular answers to those questions) using a system of formal logic[1],

[1] Usually never *quite* formal—most mathematians are content to be convinced beyond reasonable doubt that an argument is formalizable, and don't demand it be actually formalized. This is usually because the burden of formal syntax isn't seen to be worth the additional dB of certainty; machine proof is getting easy enough, and proofs hard enough, that this may change soon, though.

5. Analyze the results philosophically at each step; if they disagree with your original intuitions, weasel out the point of divergence, and decide whether to update the formal definition or your informal thought.

So, "science" to reflect the use of formality and consistent processes to minimize error, and "analogy" to reflect that one works with collections of symbols on paper that shadow real ideas. Let's look at an old friend to see how it all plays out practically.

Natural numbers are the first mathematics, in the history of both the subject and each of its learners. Sheer familiarity tends to obscure their sophistication, hard-won over stagnant millenia[2]. As their arithmetic tends to be taught, the takeaway is a couple algorithms and a collection of manipulation rules: for all numbers a, b, c, $(a+b)+c = a+(b+c)$, $a+b = b+a$, $a\cdot(b+c) = a\cdot b+a\cdot c$, etc. However, this isn't very satisfying. "*Why*," the promising student complains, "are all these properties true? I never would have thought up the associative law in a vacuum. There's got to be some reason why arithmetic has *that* law, some reason I'm not being taught a system of arithmetic with different rules." Indeed, there is a much deeper reason: all of arithmetic is just the properties of a single function "next."

But, this is a spoiler; let's proceed with the process. Our ambient concept, the *raison d'etre* of numbers, is that of (discrete) quantity. We want a way to study quantity in general, insulated from the follow-up question "quantity of what?" Foreknowledge can explain why this is useful: there's no reason to add differently for each *type* of quantity, with a completely different set of numeric symbols to be learned every time. This innocent observation is a rather deep idea in less familiar contexts, so I'll isolate it here for future recall.

$$\boxed{\text{Suppress useless information.}} \tag{P1}$$

Counting is good candidate for what's fundamental about quantity—it's certainly the first thing everyone's taught about it. What's counting? Well, you start somewhere, and repeatedly tick off numbers until whatever you're counting (or you yourself) is exhausted, producing a list $1, 2, 3, 4, \ldots$ So, we should designate a symbol as a starting point, and represent the process of going to the next number by some function[3]. Additionally, if you can't count to it, it's not a number, as defining concepts we can't access with our only tool is a waste of paper. Another fundamental concept is determining when two quantities are the same; this is done by counting them both, and if you end with the same, then the quantities are the same. Remarkably, this turns out to be all the concepts we need[4]. Let's formalize.

Definition 1. The natural numbers are a set $\mathbb{N}$, a successor function succ, and a symbol 0 subject to the following properties:

1. $0 \in \mathbb{N}$, and $\mathrm{succ}(n) = 0$ is false for all $n \in \mathbb{N}$ (a starting point exists).
2. For all $n \in \mathbb{N}$, $\mathrm{succ}(n) \in \mathbb{N}$ (you can count as long as you want).

[2] The numbers we write with can be traced back to a decimal system used in Shang China c. 1300 BC. A decimal system was later adopted by Hindu scholars in what's now India around the turn of the calendar, who invented zero and added it to the system, and introduced it all to Arabic mathematicians who put it to work in their golden era. They made their way to Europe first via merchant contact (traders needed to compare books), and were popularized by Fibbonaci in the 13th century. What makes them so special? See a historical account, or the exercises.

[3] As argued in the next chapter, functions are usually the right way to think about change of any kind.

[4] Realizing these particular few were sufficient took plenty of recalibration; this especially simple formulation is due to Peano in 1889, but Grassman first noted the utility of the successor function in the 1860s.

3. If there's some other set $N \subseteq \mathbb{N}$ containing 0 and having the property $n \in N \Rightarrow \mathrm{succ}(n) \in N$, then $\mathbb{N} \subseteq N$ ($\mathbb{N}$ is the smallest set containing everything you can count to).
4. If $n \neq m$, then $\mathrm{succ}(n) \neq \mathrm{succ}(m)$ (different starting points give different results).

Why pick zero to be the starting point? Usually, most of us learn starting with 1, and historically, the concept of 0 came late (what quantity does 0 count?). The most immediate justification is aesthetic: 0 allows us to write, for example, $3 = \mathrm{succ}(\mathrm{succ}(\mathrm{succ}(0)))$; the quantity counted by the natural number corresponds to the quantity of succ's in this alternative number system[5].

There's an equivalent way of describing the third rule, more familiar to people actually proving things:

Theorem 1. The following is equivalent to (3) above: for any proposition $P(n)$ depending on the natural number n, if $P(n) \Rightarrow P(\mathrm{succ}(n))$ and $P(0)$ hold, then $P(n)$ holds for all n.

Proof. The key observation is that logical formulas correspond to subsets, via the axiom of specification. First, presume the old formulation: $0 \in N \wedge (n \in N \Rightarrow \mathrm{succ}(n) \in N) \Rightarrow \mathbb{N} \subseteq N$. Then, for any logical formula P satisfying the conditions of the new formulation, the set $N = \{n \in \mathbb{N} \mid P(n)\}$ has the property that $0 \in N$ and if $n \in N$ then $\mathrm{succ}(n) \in N$, so by assumption of the old formulation $\mathbb{N} \subseteq N$, i.e. every natural number n satisfies $P(n)$. Conversely, presume the new formulation. Then, for any set N satisfying the conditions of the old formulation, the logical proposition $P(n) = n \in N$ has $P(0)$ and $P(n) \Rightarrow P(\mathrm{succ}(n))$, and so $P(n)$ holds for all $n \in \mathbb{N}$ by assumption of the new formulation, i.e. $n \in \mathbb{N} \Rightarrow n \in N$, or $\mathbb{N} \subseteq N$. □

We can define addition recursively: $a + 0 = a$, and $a + \mathrm{succ}(b) = \mathrm{succ}(a + b)$. This might seem like an unhelpful way to define something (how do I actually *do* addition?), but recursion is actually equivalent to computation[6], and for mathematical purposes, recursive definitions make for very easy inductive arguments. Here's an actual program that implements addition this way, albeit with a rather inconvenient numeral system:

```
(define p0 '())
(define (succ n) (list n))
(define (+ a b)
  (if (equal? b p0)
      a
      (succ (+ a (car b)))))

(+ ;; 3 + 2 = 5
 '((((())))
 '((())))
```

Example 1. Addition, with the definition above, is commutative.

Proof. We certainly have that $a + 0 = 0 + a$ by induction on a, since it's immediate when $a = 0$ and if we assume the result for a, then

$$0 + \mathrm{succ}(a) = \mathrm{succ}(0 + a) = \mathrm{succ}(a + 0) = \mathrm{succ}(a) = \mathrm{succ}(a) + 0.$$

[5] There is actually a *much* deeper reason 0 ought to be a natural number, which we'll explore in a few sections. For the time being, here's an appeal to authority: "The introduction of the digit 0...was general nonsense too, and mathematics was more or less stagnating for thousands of years because nobody was around to take such childish steps" - Grothendieck

[6] Computability theory is often called recursion theory by logicians.

Let $a, b \in \mathbb{N}$, and presume that $a+b = b+a$. Then $a+\mathrm{succ}(b) = \mathrm{succ}(a+b) = \mathrm{succ}(b+a) = b+\mathrm{succ}(a)$. We want to turn this into $\mathrm{succ}(b)+a$ to get our inductive step, so we prove the following intermediate result: for all $n, m \in \mathbb{N}$, $n+\mathrm{succ}(m) = \mathrm{succ}(n)+m$. Clearly, if $m = 0$ then $n+\mathrm{succ}(0) = \mathrm{succ}(n+0) = \mathrm{succ}(n) = \mathrm{succ}(n) + 0$, and if the result holds for m, then

$$n + \mathrm{succ}(\mathrm{succ}(m)) = \mathrm{succ}(n + \mathrm{succ}(m)) = \mathrm{succ}(\mathrm{succ}(n) + m) = \mathrm{succ}(n) + \mathrm{succ}(m),$$

so the intermediate result holds for all m. We can therefore append $b + \mathrm{succ}(a) = \mathrm{succ}(b) + a$ where we left of above, completing the original induction. □

Example 2. Addition, with the definition above, is associative.

Proof. Let $a, b, c \in \mathbb{N}$; we wish to prove $a + (b + c) = (a + b) + c$. We do so by induction on a. If $a = 0$, then $0 + (b + c) = b + c = (0 + b) + c$ (using commutativity to flip $0 + x$ into $x + 0$, so we can apply the definition). If associativity holds for a, then

$$\mathrm{succ}(a)+(b+c) = a+\mathrm{succ}(b+c) = a+(b+\mathrm{succ}(c)) = (a+b)+\mathrm{succ}(c) = \mathrm{succ}(a+b)+c = (\mathrm{succ}(a)+b)+c,$$

where we have made liberal use of the intermediate result used in the previous example. □

To not spoil all the fun, the rest of the laws of arithmetic are left to you in the exercises. This is enough to illustrate the central theme of algebra: recasting the natural numbers in terms of the most obvious *function* associated with them allows us to realize that the laws of arithmetic, which seem strange and arbitrary when they're just asserted, can actually be *derived* from perfectly natural properties of a system for reasoning about quantity.

This definition of the natural numbers is our first example of a mathematical **structure**—roughly, a way of puting math inside logic, in a formal way. Precisely,

Definition 2. A structure is a triple $\mathcal{S} = (U, \sigma, I)$, where U is any set, called the **underlying set**, **universe**, or **domain**, the **signature** $\sigma = (F, \mathrm{ar})$ is a set F of function and relation symbols to be added to the logical language with $\mathrm{ar} : F \to \mathbb{N}$ mapping each symbol to its intended arity (number of arguments, where 0-ary functions are **constants**), and $I : F \to U^n$, the **interpretation**, mapping the symbols in F to an actual function or relation of the appropriate arity on U.

Structures are closely related to **theories**, any collection of logical sentences (usually deduced from some subset of axioms), because those logical sentences may incorporate symbols from the signature, enabling reasoning about non-logical concepts. If all of a theory's sentences are true when the non-logical symbols are interpreted in a particular structure, then the theory is said to be **satisfied** by the structure, and the structure to **model**, **implement**, or be a **model of** the theory[7].

This is all getting very abstract, so let's update our natural numbers example to intuitively ground all these concepts[8].

[7] Given a theory, it's important to show that such a model exists: it's possible for the logical constraints to be **inconsistent** or **unsatisfiable**, in which case the entailing mathematical work is all but worthless.

[8] The astute reader will note that the concept of a structure, as-presented, depends on $\mathbb{N}$; this is merely an example of concepts, so it's but a local concern, though a well-warranted one. A deeper problem with our whole definition of the natural numbers is that the theory is second-order (involves quantification over the uncountable set $\mathcal{P}(\mathbb{N})$); it's easier to just define the numbers via a model, and be done with it.

Definition 3. The natural numbers are any[9] structure with some (concretely defined) underlying set N and signature $(\{\text{succ}, \bar{0}\}, (\text{succ}, 1), (\bar{0}, 0))$ that models the theory generated by deduction from the following axioms:

1. $\bar{0} \in N$
2. $\forall n \in N, \neg(\text{succ}(n) = \bar{0})$
3. $\forall n \in N, \text{succ}(n) \in N$
4. $\forall N \in \mathcal{P}(N), \bar{0} \in N \wedge (n \in N \Rightarrow \text{succ}(n) \in N) \Rightarrow N \subseteq N$
5. $\forall n, m \in N, \text{succ}(n) = \text{succ}(m) \Rightarrow n = m$

An example of such a natural number structure appears in the program for computing addition given above. Defining $\mathbb{N}$ as the set of all possible results of applying `succ` to `p0` repeatedly, we see analogs of everything in the signature, and can verify the axioms; by consistently applying this analogy to any logical formula, we get the interpretation function I.

Theories of mathematical structures are the "encoded core essence" of philosophical concepts mentioned in the outline of mathematics above, and their logical nature enables correct, formal reasoning about those concepts. The rest of this book deals with theories of **algebraic structures**[10], which have no relation symbols in their signature (the most common relation symbols are orders—which induce an analytical flavor). As a shorthand for specifying an algebraic structure and a theory over it, we'll simply mention the set and the signature in an informal way, and carry on doing logic with those symbols injected. The results ought to apply to any model of the theory.

1.2 Functions

The core idea in studying structures is to look at those functions that respect or preserve them. This principle goes back to Emmy Noether, the mother of modern mathematics, and is a watershed between the old and the new. What, precisely, do we mean by "preserving" a structure?

Definition 4. A function $f : A \to B$, where A and B are structures of the same signature, is called **structure-preserving** if

1. for all function symbols g with arity n, $f(g(x_1, x_2, \ldots, x_n)) = g(f(x_1), f(x_2), \ldots, f(x_n))$,
2. for all relation symbols R with arity n, $(x_1, x_2, \ldots, x_n) \in R \Rightarrow (f(x_1), f(x_2), \ldots, f(x_n)) \in R$.

For any function satisfying this property, you can make an argument within theory A, and then apply f to get a proof of a formula in B, immediately apply f and argue in B to prove a formula in A, or anything in between; f is a translation of facts and arguments from one context to another.

Example 3. Structure-preserving maps from the structure $(S, \leq)$ to itself (some set with a non-strict partial order) are increasing functions.

[9] It's a theorem (which we won't prove) that all structures that satisfy this theory are essentially equivalent (the stupid name for this that's become standard is "categorical"—it should be "canonical" to avoid confusion).

[10] It's customary to abuse terminology and refer to a class of algebraic structures modelling the same theory as an algebraic structure.

Proof. A function f is defined to be increasing if $x \leq y \Rightarrow f(x) \leq f(y)$ (think about this graphically!), so the property is immediate.

The immediate consequence of this is that applying increasing functions to both sides of any inequality preserves the truth-value of that inequality—the implicit principle behind the following derivation, for example:

$$\sqrt{2} \leq 1 + \frac{1}{\sqrt{2}} \Leftrightarrow \sqrt{2} \leq \frac{2+\sqrt{2}}{2} \Leftrightarrow 8 \leq 6 + 4\sqrt{2} \Leftrightarrow \frac{1}{2} \leq \sqrt{2} \Leftrightarrow \frac{1}{4} \leq 2 \Leftrightarrow 1 \leq 8$$

At each step, the fact that the operation(s) applied to both sides of the inequality were structure-preserving is what made it work; if, for example, we multiplied by a negative number, it would lead us to absurdity ($1 \leq 8 \not\Leftrightarrow -1 \leq -8$). □

Theorem 2. Structure-preserving maps leave constant symbols unchanged.

Theorem 3. The identity map is always structure-preserving.

$$\boxed{\text{The most obvious thing to do is nothing.}} \tag{P2}$$

1.3 Action

1.4 Exercises

Problem 1 (Those Vedic Guys Were Goated W/Sauce)**.** The Roman numeral system was historically used throughout Europe prior to the introduction of our modern symbols. Why are our current numbers so much better as to drive the change?

1. Develop an addition algorithm for Roman numerals without appealing to the usual one (directly; you can of course use the one you know for motivation).
2. Pick some moderately big number, convert it to Roman numerals, and try to factor it without converting to Arabic.
3. In what sense are Roman numerals base-10? Develop binary and hexidecimal Roman numeral systems, and compare to their place-value equivalents.
4. Imagine a sensible way to extend the system to enable representing rational approximations to real numbers. Is a floating-point standard for the Roman binary developed above more or less obvious than our usual IEEE representation?
5. Are there any intrinsic reasons to prefer IV to IIII? What're a couple disadvantages to having multiple representations for the same quantity?
6. Anticipate some problems with using the same symbols to represent quantity and phonemes in the modern day.

Problem 2 (Induction Can Start Anywhere)**.** Prove by induction that: for all propositions $P(n)$ depending on a natural number n, if $P(n) \Rightarrow P(\text{succ}(n))$ and $P(n_0)$ holds for some $n_0 \in \mathbb{N}$, then $P(m)$ holds for all $m > n_0$.

Problem 3 (Strong Induction)**.** Prove that the following is equivalent to induction: for all propositions $P(n)$ depending on a natural number n, if $P(m)$ holding for all $m < n$ implies $P(\mathrm{succ}(n))$ and $P(0)$ holds, then $P(n)$ holds for all $n \in \mathbb{N}$.

Problem 4 (Peano Multiplication)**.** Just as addition is repeated succession, so multiplication is repeated addition: we define $a \cdot 0 = 0$, and for any number n, $a \cdot \mathrm{succ}(n) = a \cdot n + a$. Prove the following arithmetic laws:

1. $a \cdot b = b \cdot a$
2. $(a \cdot b) \cdot c = a \cdot (b \cdot c)$
3. $a \cdot (b + c) = a \cdot b + a \cdot c$

Problem 5 (Comparing Quantities)**.** We defined abstract quantities, but never talked about how to compare them, despite that being one of the original motivations.

1. Define an ordering on natural numbers in Peano arithmetic that agrees with your understanding of the "usual" $\leq$.
2. Verify that this is indeed a (non-strict, total) ordering; that is, prove for your relation $\preceq$ and all $a, b, c \in \mathbb{N}$:
   (a) $a \preceq a$ (reflexivity),
   (b) if $a \preceq b$ and $b \preceq a$ then $a = b$ (antisymmetry),
   (c) if $a \preceq b$ and $b \preceq c$, then $a \preceq c$ (transitivity),
   (d) $a \preceq b$ or $b \preceq a$ (totality).
3. Think about the way you ordinarily look at two Arabic numbers and decide if one is greater than or less than the other. Formalize this decision proceedure, and try to prove that the resulting algorithm agrees with your order relation.

Problem 6 (A Full Implementation)**.** Let's make a Peano arithmetic calculator.

1. Using the Scheme program given in the footnotes as a guide, implement bare-bones Peano addition in your favorite programming language (if you like Scheme, you can just use my version).
2. WITHOUT using any numeric operations, write a pair of functions that convert between the computational representation of a number to Arabic numerals (preferrably returning a string, so you aren't tempted to cheat!).
3. Do the same for multiplication.
4. If your language supports first-class functions, write a function that performs the common abstract proceedure through which addition and multiplication were defined.
5. Test this abstraction by implementing exponentiation with it. What happens if you apply this more times? To itself?

Problem 7 (Divisibility)**.** For $a, b \in \mathbb{N}$, we say that "a **divides** b," "a is a **divisor** of b," or $a \mid b$ if there exists $k \in \mathbb{N}$ such that $b = ka$. k, if it exists, is called the **quotient** of b by a or b **divided by** a, and is written $\frac{b}{a}$.

1. Give two numbers and prove that neither divides the other.
2. Show by example that the quotient operation is not associative.

Problem 8 (Remainders)**.** An operation related to division is the "modulo" operation: given two natural numbers a and b, $a \pmod b$ is equal to the smallest $r \in \mathbb{N}$ such that $b \mid a - r$.

1. Prove that, unlike division, that r exists for any a and b.
2. Prove that the image of this operation (over all a and fixed b) has cardinality b.

Problem 9 (Prime Numbers)**.** A number is **prime** if it has no nontrivial divisors, i.e. none other than the number itself and 1 (otherwise, it is **composite**). Prove that there are infinitely many primes.

Chapter 2

Symmetries

2.1 What is Symmetry?

2.2 Symmetry-Preserving Maps

2.3 Sub-Symmetries

2.4 Dissecting Symmetries

Chapter 3

Symmetric Symmetries

3.1 Symmetries Acting on Themselves

Chapter 4

Structure, Functions, and Action Done Right

4.1 Spooky Coincidences

4.2 Categories are Functions and Structure

4.3 Functors are Action

Appendix A

Logical Foundations

I consider the standard mathematical foundations (classical logic and Zermelo-Frankel set theory with the axiom of choice) to be true; non-constructive proofs are valid in a philosophical sense. However, *doing* classical logic does absolutely nothing but hide relevant information about computability, and immensely complicates the interpretation of infinity.

The Curry-Howard correspondence lets us say that if we have a constructive proof of a quantified proposition ("for all x, y"), then we have an algorithm for producing a proof of y given any x. Conversely, in classical logic, we often have no way of actually *producing* things we prove exist. This is, in principle, fine (such a proof generally consists of establishing that non-existence is an absurdity), but if we can solve a problem both ways, it's very clear that the former is preferred. Moreover, working in a constructive system, it's possible to gain insight as to what classical logic would say: the Gödel-Gentzen translation is a way of taking a formula, and obtaining an associated formula that has a constructive proof iff the original has a classical proof. This means that constructive logic is *strictly more general* than classical logic, in a practical sense. Here's a quick rundown of those standard mathematical foundations, for reference[1].

Our formal language will be considered to that of Lean 4, with

Definition 5. A **function** or **map** $f : A \to B$ is a subset of the Cartesian product $A \times B$ (with $(a, b) \in f$ written $b = f(a)$) that is **well-defined**: $x = y \Rightarrow f(x) = f(y)$. The set A is called the **domain**, and the set B is the **codomain**. Often, the map f is specified by a **assignment rule** or simply **rule** $a \mapsto g(a)$ or $f(a) = g(a)$, where $g(a)$ is some other concretely-defined function.

Definition 6. Given any subset $X \subseteq A$ and a function $f : A \to B$, the subset $f(X) \subseteq B$ defined by

$$f(X) = \{b \in B \mid \exists x \in X, b = f(x)\}$$

is called the **image** of X under f, or, if $X = A$, the image of f.

Definition 7. A function $f : A \to B$ is **injective** if $f(x) = f(y) \Rightarrow x = y$, for all $x, y \in A$. This can be denoted by $f : A \hookrightarrow B$.

Definition 8. A function $f : A \to B$ is **surjective** if $f(A) = B$, that is, if $b \in B \Rightarrow \exists a \in A, b = f(a)$. This can be denoted by $f : A \twoheadrightarrow B$.

[1] For more detailed exposition, see [**?**] or [**?**].

Definition 9. A function is **bijective** if it is both injective and surjective. This can be denoted by $f : A \hookrightarrow\!\!\!\!\rightarrow B$

Definition 10. When $f : A \to B$ and $g : B \to C$, we may define the **composite map** $g \circ f : A \to C$ (read "g after f") by $(g \circ f)(a) = g(f(a))$.

Definition 11. The **identity function** $id_A : A \to A$ is given by $id_A(a) = a$.

Theorem 4. If a function is injective, then it has a left-inverse: letting $f : A \to B$ be injective, there exists a function $f_L^{-1} : f(B) \to A$ such that $f_L^{-1} \circ f = id_A$.

Theorem 5. If a function is surjective, then it has a right-inverse: letting $f : A \to B$ be surjective, there exists a function $f_R^{-1} : B \to A$ such that $f \circ f_R^{-1} = id_A$.

Appendix B

How Do I Know if My Solution is Right?

Bibliography

[1] E. Reihl, *Category Theory in Context*. Aurora: Dover Modern Math Originalss, Dover Publications, 2016.

[2] P. Halmos, *Naive Set Theory*. Undergraduate Texts in Mathematics, Springer New York, 1998.