Math250

Lecture Notes

on

Discrete Mathematics

M. Andrew Moshier

January 2015

Version: 0.0.0 (February 2, 2015) Format: print

Abstract things very often can stand in for, or *represent*, concrete, practical things while stripping away unwelcome clutter. For example, it is certainly useful to know that addition is a commutative operation ($x + y = y + x$ is always true no matter what numbers x and y happen to be). It does not matter what x and y represent (the number of math and the number of computer science students, or the length of two sticks, or the sizes of two bank accounts). The fact is we can think about addition *abstractly* without needing to know what the numbers stand for. This is incredibly useful, and is at the heart of mathematics. Multiplication is also commutative. This, surely, is also a useful thing to know.

We now have two operations (addition, multiplication) that are quite different from one another, but they share some properties: both are commutative. You can easily think of other propertie You can probably also come up with some properties that they do not share (hint: Does every number have a negative? Does every number have a reciprocal?). Thinking about arithmetic in abstract terms helps us understand where the abstractions are appropriate.

Actually, the answers to most questions depend on further levels of abstraction. For example, whether or not every number has a negative depends on what sort of numbers we care about. If we only want to think about counting numbers – 0, 1, 2, and so on – then the only number that has a negative is 0 because $0 + 0 = 0$, but for example "-1" does not even exist as far as we are concerned.

For another example, the claim that every number has a square root is obviously false if our numbers are the real numbers, since negative numbers do not have square roots. But it happens that every complex number does have exactly two square roots (except 0 which has only itself as a root). And though a formula for the square root is not easy to find, the property that all complex numbers have two square roots (except for 0) is part of what makes complex numbers useful. So we can not just think about things like addition, multiplication and so on, without also thinking about the *type* of data we are concerned with.

A big advantage of dealing in abstraction is that, by separating mathematical *ideas* from any concrete interpretation, we sometimes discover other completely novel uses. Try the following procedure:

1. Write down your credit card number (I will not ask you to give anything away). For example, suppose the number is 845937493485 (not a real card).

2. Starting from the 2nd digit on the right, double every other digit and leave the other digits as is. For our example, the result would be 164109678964165. I have put spaces in here to make it more readable.

3. Now add all the individual digits of the result. That is, count 16 as $1+6$. For the example, the result is $1+6+4+1+0+9+6+7+8+9+6+4+1+6+5=73$.

Your result will be a number that ends with 0, assuming you have a valid credit card. Now try the same procedure using your credit card number with a mistake. Try swapping two adjacent digits, or change any single digit. The result will not end in 0 (with only one exception – swapping a 9 with an adjacent 0 does not do anything to the result).

This procedure (called *Luhn's checksum*) uses simple arithmetic to test that a number is, at least possibly, a valid card number. It catches the easiest errors (adjacent digit swaps and single digit errors). But this has nothing to do with the usual concrete interpretation of addition, as in 4 apples plus 3 apples give us 7 apples. Arithmetic is used here as purely abstract operations on these very large numbers, but the numbers themselves do not represent counting anything, or measuring anything.

In this course, we deal head on with mathematics as the study of abstract stucture. This can be frustrating at first because the concrete applications are not always obvious. The pay off comes when we see that abstraction from particulars leads to much wider applications than we could have anticipated.

We have claimed that addition and multiplication are commutative. But is that really true? Is it the case that $1239283 + 11^{11^{1}1}$ (a number with more that 100 trillion digits) is the same as $11^{11^{1}1} + 1239283$? These numbers are far too big to check explicitly. But we are completely confident that the two numbers are equal. Why?

We should be able to convince ourselves that $m+n=n+m$ is true for any two counting numbers, not just take it as being "obvious". Indeed, figuring out how to convince ourselves (and other mathematicians) about what is true is one of the main activities of mathematics. In a sense, this makes mathematics a human activity. It is mainly concerned with persuasion of the sort: I know that addition is commutative, here is why you should know it too ...

The standards for persuasion in mathematics are high. We do not settle for "preponderance of evidence" or "beyond reasonable doubt". We aim for "beyond any doubt". In most human endeavours (including in the sciences), that standard would be paralyzing, but in mathematics, where abstract structure is the object of investigation, it is not only within reach, but it a standard that makes sense.

In mathematics, a **proof** is a convincing argument. We will spend a great deal of our time discussing and constructing proofs. Proofs that meet the standards of mathematics have a vocabulary, grammar and prosody of their own. That is, there are specific word usages that have precise technical meanings (vocabulary). There are basic rules of proof construction (grammar). And there are general ways to make a proof more understandable, and therefore more convincing (prosody).

Part I

Natural Numbers and Their Relatives

The *natural numbers* are the numbers we use to count things. Our earliest mathematical experience is learning to count. So, in a sense, a study of the natural numbers is a return to our childhood. Arithmetic is connected to counting. But also, our ability to prove things about natural numbers is deeply bound up in counting.

Another early mathematical experience is learning to arrange things in some order. The informal concept of a *list* captures this. We discuss lists formally, showing how they relate to natural numbers.

Of course, natural numbers are not enough. We need to be able account for deficits as well a surpluses. Integers allow this. We also need to be able to reason about proportions (4 out of 5 mathematics student choose Discrete). Rational numbers allow this. Real and complex numbers also have their uses, but we will not descuss them there. That's the job of courses in analysis.

LECTURE **1**

The Natural Numbers

Goals

Lecture

- Present the natural numbers as comprising a structure suited to counting.
- Identify similar structures that can not properly represent counting.
- Rule out "bad" structures via postulates.

Study

- Gain facility in the course's *successor* notation, including translating between successor notation and base 10 notation.
- Commit to memory the postulates of natural numbers.
- Demonstrate ability to recognize failures of the postulates.

The *natural numbers* have to do with counting: $0, 1, 2, 3, \ldots$. They do not include negatives or fractions or irrationals. In this lecture, the structure of natural numbers is the topic. To hone in on that structure, we look at structures similar to the natural numbers, but that fail to capture some basic aspects of counting. Bogus structures are ruled out by *postulates* (also known as *axioms*) that distinguish the structure of natural numbers from others.

1.1 The Basic Picture

Our first task is to look back to one of the very first mathematical concepts we all learned, namely, *counting*. Some numbers obviously are meant for conuting. 1, 2, 3, and so on are counting numbers, whereas -1, $\frac{1}{2}$ and $\sqrt{2}$ are not. So let's start with a basic intuitive definition: a *counting* or *natural* number is a number that can be used to answer a qeustion of the form "How many x are there?" This is distinct, somehow, from "How *much* x?" The latter sort of question could be answered with $\frac{1}{2}$ (as long as we know how we are measuring). On the other hand, the answer to a "how many" question could be 0 ("how many professors own unicorns?"). So 0 shoud be included in our thinking.

Natural numbers are pictured like stepping stones in Figure 1.1.

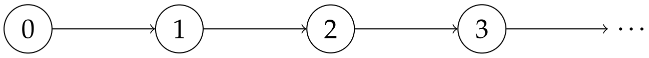


Figure 1.1: A picture of the natural numbers

Not all "stepping stone" pictures are acceptable. Figures 1.2, 1.3 and 1.4 illustrate three ways *not* to picture the natural numbers.

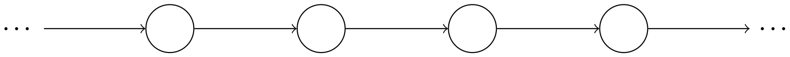


Figure 1.2: Nowhere to start

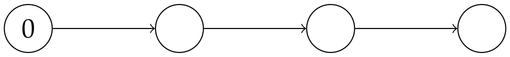


Figure 1.3: Nowhere to go

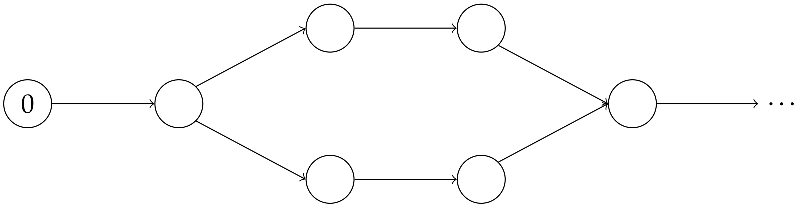


Figure 1.4: Forks in the path

These incorrect pictures can be ruled out by explaining the basic structure of counting. We will "explain the obvious" by stating things like this as *postulates*.

Postulate 1: Basic Structure of Natural Numbers

The **natural numbers** have the following basic structure.

- There is a special natural number. We denote this by 0.
- For any natural number n, there is a unique *next* natural number. We call this the **successor of** n. In these lectures, we denote the successor of n by $n^\frown$.

According to Postulate 1, 0, $0^\frown$, $0^{\frown\frown}$, $0^{\frown\frown\frown}$ each denote a natural number. Of course, we usually abbreviate them by writing $0, 1, 2, 3$. But the *characters* 1, 2, 3, etc., are not related to each other in any way. The notation we are using here makes it completely clear that $0^\frown$ is the number after 0, and so on. We will want to be able to switch between the familiar "decimal' notation and "successor" notation whenever it is convenient.

Exercises for Lecture 1

Convert the following from decimal notation to successor notation.

1. 9
2. 10
3. $4 + 3$
4. $n + 4$

Convert the following to from successor notation to decimal notation.

1. $0^{\curvearrowright\curvearrowright\curvearrowright\curvearrowright}$
2. $n^{\curvearrowright\curvearrowright\curvearrowright\curvearrowright\curvearrowright}$
3. $5^{\curvearrowright\curvearrowright}$
4. $0^{\curvearrowright} + 0^{\curvearrowright\curvearrowright}$

1.2 Narrowing the possibilities

Figures 1.5 and 1.6 illustrate problems that Postulate 1 does not avoid.

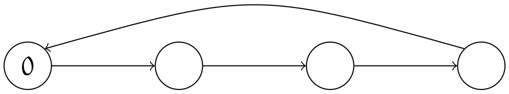


Figure 1.5: A strange way to count

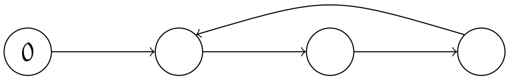


Figure 1.6: Another strange way to count

Exercises for Lecture 1

1. Explain, in one or two sentences each, why Figures 1.5 and 1.6 depict systems that agree with Postulate 1.

Figure 1.5 is flawed because 0 has a *predecessor*: a value n satisfying $0^{\curvearrowright\curvearrowright\curvearrowright\curvearrowright} = 0$. Figure 1.6 is flawed because an element has two distinct predecessors: $0^{\curvearrowright} = 0^{\curvearrowright\curvearrowright\curvearrowright\curvearrowright}$. We can insist that these flaws do not happen in the natural numbers. That is, we rule them out with axioms.

Postulate 2

Nothing Precedes 0 For every natural number n, $n^{\curvearrowright} \neq 0$.

Postulate 3

Predecessors are Unique For any natural numbers m and n, if $m^{\curvearrowright} = n^{\curvearrowright}$ then $m = n$.

These postulates eliminate Figures 1.5, 1.6 and similar pictures. But there is still a subtle problem. Consider Figure 1.7.

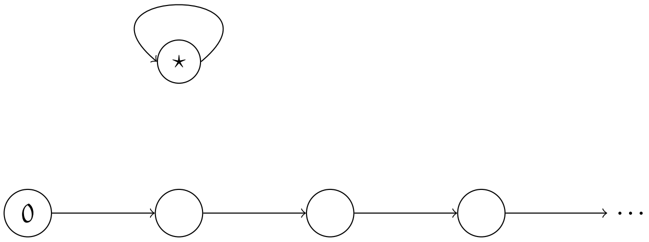


Figure 1.7: A model of the natural numbers?

This picture satisfies the first three postulates. Yet, it is not a picture of natural numbers because it has "extra stuff" in it ($\star$).

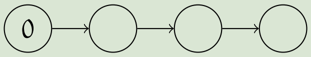
To rule out "extra stuff", we formulate our final postulate for natural numbers. We diagnose the problem as follows. Were we to erase the circle labelled $\star$ and any the arrows leading to and from it, the remaining part of Figure 1.7 would still live up to Postulate 1. This is exactly what we mean by "extra stuff": elements that can be removed without violating the Postulate 1 (the essential structure). This leads to our last axiom.

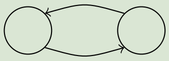
Postulate 4

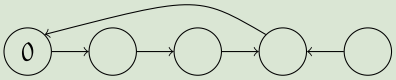
The Axiom of Induction No natural numbers can be removed without violating 1.

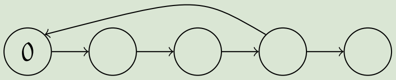
Exercises for Lecture 1

1. Each of the following pictures fails to satisfy either the one or more of our axioms. For each, explain which axioms are violated.

   1. 

   2. 

   3. 

   4. 

2. I have in mind a picture for the Basic Vocabulary 1 and that satisfies Axioms 2 and 3. Furthermore, in that picture, I have in mind and element n for which (a) $n \neq 0$ and (b) n has no predecessor (that is, $n \neq m^{\curvearrowright}$ for every m). Convince me that the picture fails to satisfy Axiom 4.

3. Draw three different pictures of situations that satisfies all the postulates except that they fail Postulate 2. So there will be an arrow from a bubble into the bubble 0. The result must satisfy all other postulates including the Axiom of Induction.

The latest exercise shows that in the natural numbers, if $n \neq 0$, then $n = m^{\curvearrowright}$ for some m. In other words, every non-zero natural number has a predecessor.

LECTURE **2**

Arithmetic

Goals

Lecture

- Present addition and multiplication via defining equations.
- Practice using the defining equations to calculate sums and products.

Study

- Understand addition and multiplication as characterized by defining equations.
- Be able to explain how addition and multiplication relate to counting.
- Exhibit competence in calculating sums and products from the defining equations.

Adding and multiplying arise from counting. In this section, we explore how to define them purely in terms of counting.

2.1 Basic Arithmetic Operations

We know that addition "works" by counting ahead. For example, to *add* $4 + 5$, we can start with 4 and then count up five more. Likewise, multiplication "works" by counting a number of additions. For example, to multiply $2 \cdot 3$ we can add 2 three times: $2 + 2 + 2$. The following definitions capture the idea.

Definition 1: Arithmetic Operations

The **sum** of two natural numbers, m and n, is a natural number (denoted by $m + n$). For every natural number m, the following are true:

$$\begin{aligned} m + 0 &= m \\ m + k^{\curvearrowright} &= (m + k)^{\curvearrowright} && \text{for any natural number } k \end{aligned}$$

The **product** of two natural numbers, m and n, is a natural number (denoted by $m \cdot n$). For every natural number m, the following are true:

$$\begin{aligned} m \cdot 0 &= 0 \\ m \cdot k^{\curvearrowright} &= m + (m \cdot k) && \text{for any natural number } k \end{aligned}$$

A moment's thought about arithmetic should convince you that these equations are reasonable. Certainly $m + 0 = m$ and $m \cdot 0 = 0$ should be true for any m. The second equation for $+$ can be read as saying "to add m to the successor of k, simply add m to k, then take the successor." The second equation for $\cdot$ can be read as saying "to multiply m by the successor of k, simply multiply m by k, and add m to the result."

The Axiom of Induction ensures that there are indeed unique operations $+$ and $\cdot$ that satisfy the equations. A proof of this fact is not particularly illuminating right now, so let us agree to take it for granted.

Example 2

Do the defining equations for addition really explain how to add? Let's use them to calculate $4+3$:

$$\begin{aligned}
4+3 &= 4+0^{\curvearrowright\curvearrowright\curvearrowright} && [\text{3 abbreviates } 0^{\curvearrowright\curvearrowright\curvearrowright}]\\
&= 4^{\curvearrowright}+0^{\curvearrowright\curvearrowright} && [m+k^{\curvearrowright} = m^{\curvearrowright}+k]\\
&= 4^{\curvearrowright\curvearrowright}+0^{\curvearrowright} && [\text{Same reason}]\\
&= 4^{\curvearrowright\curvearrowright\curvearrowright}+0 && [\text{Same reason}]\\
&= 4^{\curvearrowright\curvearrowright\curvearrowright} && [m+0=m]\\
&= (0^{\curvearrowright\curvearrowright\curvearrowright\curvearrowright})^{\curvearrowright\curvearrowright\curvearrowright} && [\text{4 abbreviates } 0^{\curvearrowright\curvearrowright\curvearrowright\curvearrowright}]\\
&= 0^{\curvearrowright\curvearrowright\curvearrowright\curvearrowright\curvearrowright\curvearrowright\curvearrowright} && [\text{Remove unneeded parentheses}]\\
&= 7 && [\text{7 abbreviates } 0^{\curvearrowright\curvearrowright\curvearrowright\curvearrowright\curvearrowright\curvearrowright\curvearrowright}]
\end{aligned}$$

Example 3

A product can be calculated similarly. Consider $2 \cdot 2$.

$$\begin{aligned}
2 \cdot 2 &= 2 \cdot 0^{\curvearrowright\curvearrowright} && [\text{2 abbreviates } 0^{\curvearrowright\curvearrowright}]\\
&= 2+(2 \cdot 0^{\curvearrowright}) && [m \cdot k^{\curvearrowright} = m+(m \cdot k)]\\
&= 2+(2+(2 \cdot 0)) && [\text{Same reason}]\\
&= 2+(2+0) && [m \cdot 0 = 0]\\
&= 2+2 && [m+0=m]\\
&= 2+0^{\curvearrowright\curvearrowright} && [\text{2 abbreviates } 0^{\curvearrowright\curvearrowright}]\\
&= 2^{\curvearrowright}+0^{\curvearrowright} && [m+k^{\curvearrowright} = m^{\curvearrowright}+k]\\
&= 2^{\curvearrowright\curvearrowright}+0 && [\text{Same reason}]\\
&= 2^{\curvearrowright\curvearrowright} && [m+0=m]\\
&= (0^{\curvearrowright\curvearrowright})^{\curvearrowright\curvearrowright} && [\text{2 abbreviates } 0^{\curvearrowright\curvearrowright}]\\
&= 0^{\curvearrowright\curvearrowright\curvearrowright\curvearrowright} && [\text{Remove unnecessary parentheses}]\\
&= 4 && [\text{4 abbreviates } 0^{\curvearrowright\curvearrowright\curvearrowright\curvearrowright}]
\end{aligned}$$

We certainly will not want to calculate this way in real life. After all, it took twelve steps just to figure $2 \cdot 2 = 4$. But these examples and the following exercises show how addition and multiplication are closely tied to simple counting.

Exercises for Lecture 2

1. Calculate these sums, following the previous example to write each step of your calculation explicitly. Include the reason for each step (as in the previous example).Take care to lay out the chain of equalities correctly, and do not skip any steps.

    1. $2 + 4$
    2. $4 + 2$
    3. $3 + (3 + 1)$
    4. $(3 + 3) + 1$
    5. $0 + 3$

2. Notice that it takes more steps to calculate $2 + 4$ than $4 + 2$, even though we know they will produce the same answer. Explain why.

3. Calculate the following values, writing each step explicity.

    1. $2 \cdot 3$
    2. $0 \cdot 2$
    3. $2 \cdot (2 \cdot 2)$
    4. $3 \cdot (2 + 1)$
    5. $(3 \cdot 2) + (3 \cdot 1)$

4. Write a definition of exponentiation via defining equations. Follow the pattern of definition I have written for addition and multiplication.

LECTURE 3

Laws of Arithmetic

Goals

Lecture

- Present the most common Laws of Arithmetic for natural numbers.
- Explain the method of *proof by simple induction*
- Prove a representative sample of the laws by simple induction.

Study

- Become familiar with the common names for the Laws of Arithmetic.
- Pay particular attention to the Laws of Positivity and Cancellativity (they may be the least familiar to you).
- Demonstrate the ability to identify the main parts of a proof by simple induction.
- Demonstrate the ability to construct the parts of a proof by simple induction.
- Prove the remaining laws for yourself.

Before working the last exercises, you knew that $3 \cdot (2+1)$ and $3 \cdot 2 + 3 \cdot 1$ would come out the same because of a law of arithmetic known as *distributivity*. Addition and multiplication satisfy several other laws.

The following list summarizes several useful laws of arithmetic on the natural numbers. They are organized to emphasize similarities between addition and multiplication.

Laws 1

For any natural numbers, m, n and p:

Associativity $m + (n + p) = (m + n) + p$
$m \cdot (n \cdot p) = (m \cdot n) \cdot p$

Commutativity $m + n = n + m$
$m \cdot n = n \cdot m$

Identity $m + 0 = m$
$m \cdot 1 = m$

Positivity if $m + n = 0$ then $m = 0$
if $m \cdot n = 1$ then $m = 1$

Cancellativity if $m + p = n + p$ then $m = n$
if $m \cdot p^{\frown} = n \cdot p^{\frown}$ then $m = n$

Distributivity $m \cdot (n + p) = (m \cdot n) + (m \cdot p)$

Case Distinction if $m \neq 0$ then $m = k^{\frown}$ for some k

Most of these laws are familiar and are listed with their common names. The Law of Case Distinction was the subject of Lecture **??** Exercise 2.. *Go back and look at that exercise again*. The Law of Positivity for multiplication is not a common name, but I have used it to emphasize the analogies between addition and multiplication. Also Case Distinction does not really have a common name. I made that up.

3.1 Inductive Proofs

Suppose we wish to prove that every natural number has some property. For example, let us suppose we wish to prove that every natural number is *mimsy*. I have no idea what a mimsy number is, but let us try to prove this anyway. We could try proving that 0 is mimsy, 1 is mimsy, 2 is mimsy, and so on. But this won't work because our proof will never end. In fact, it is not so obvious that we, humans with finite minds, can ever prove that some property is true for *all* natural numbers, since it seems to involve checking infinitely many individual cases.

The Axiom of Induction provides a way forward in spite of our limitations. Suppose we were to show that the mimsy natural numbers all by themselves constitute a picture of Signature 1. Then there could not be any natural numbers left out, for otherwise, we could erase all the non-mimsy natural numbers and still have a picture of 1. This is exactly what the Axiom of Induction forbids: we can not erase *anything* without breaking the signature.

So to prove that all natural numbers are mimsy, we simply need to prove that

- 0 is mimsy, and
- for all natural numbers k, if k is mimsy so is $k^{\frown}$.

From these, we conclude that the mimsy natural numbers by themselves form a picture of 1. So the Axiom of Induction ensures that all natural numbers are mimsy.

To make inductive proofs easier to understand, we often write them using a three step outline, as illustrated here.

- [Basis] Prove that 0 is mimsy.
- [Inductive Hypothesis] Assume that k is mimsy.
- [Inductive Step] Prove that $k^{\curvearrowright}$ is mimsy. [You may use the assumption that k is mimsy in this part of the proof.]

More practical examples are next.

Proposition 2

Addition is associative.

Proof: We need to show that $m + (n + p) = (m + n) + p$ for all m, n and p. Let us suppose that m and n are fixed values (not known to us). We now prove that the values p for which $m + (n + p) = (m + n) + p$ holds form a picture of 1.

- [Basis] $m + (n + 0) = m + n = (m + n) + 0$. Both steps are due to the defining equations of $+$.
- [Inductive Hypothesis] Assume $m + (n + k) = (m + n) + k$.
- [Inductive Step] We must show that $m + (n + k^{\curvearrowright}) = (m + n) + k^{\curvearrowright}$.

$$\begin{aligned}
m + (n + k^{\curvearrowright}) &= m + (n + k)^{\curvearrowright} && \text{[Def. of +]} \\
&= (m + (n + k))^{\curvearrowright} && \text{[Same]} \\
&= ((m + n) + k)^{\curvearrowright} && \text{[Inductive Hypothesis]} \\
&= (m + n) + k^{\curvearrowright} && \text{[Def. of +]}
\end{aligned}$$

Therefore (by the Axiom of Induction), $m + (n + p) = (m + n) + p$ holds for all p. Since the argument does not depend on any extra assumptions about m and n, it holds for all m and n. □

In the remainder of this section, we further illustrate the technique of simple arithmetic induction via proofs of other laws of arithmetic.

Proposition 3

0 is the identity for addition.

Proof: We must prove that $m + 0 = m = 0 + m$ for all m. The first equality is true by the definition of $+$. But the second equality, $m = 0 + m$, is not explicitly one of the defining facts about $+$. So we proceed by induction on m.

- [Basis] $0 + 0 = 0$ is true by definition of $+$.
- [Inductive Hypothesis] Assume $0 + k = k$.
- [Inductive Step] We must show that $0 + k^{\curvearrowright} = k^{\curvearrowright}$.

$$\begin{aligned} 0 + k^{\curvearrowright} &= (0 + k)^{\curvearrowright} && \text{[Def. of $+$]} \\ &= k^{\curvearrowright} && \text{[Inductive hypothesis]} \end{aligned}$$

Therefore, $0 + m = m$ holds for all m. □

To prove that addition is commutative, we need an additional fact about how successor and addition interact. Mathematicians use the word *lemma* to indicate that a certain fact is only needed to make others proofs easier and is not necessarily valuable in its own right.

Lemma 4

For any m and n, $(m+n)^{\curvearrowright} = m^{\curvearrowright} + n$.

Proof: By induction on n:

- [Basis]

$$\begin{aligned} (m+0)^{\curvearrowright} &= m^{\curvearrowright} && \text{[Def. of +]} \\ &= m^{\curvearrowright} + 0 && \text{[Def. of +]} \end{aligned}$$

- [Inductive Hypothesis] Assume $(m+k)^{\curvearrowright} = m^{\curvearrowright} + k$ for some k.
- [Inductive Step] We must show that $(m+k^{\curvearrowright})^{\curvearrowright} = m^{\curvearrowright} + k^{\curvearrowright}$.

$$\begin{aligned} (m+k^{\curvearrowright})^{\curvearrowright} &= ((m+k)^{\curvearrowright})^{\curvearrowright} && \text{[Def. of +]} \\ &= (m^{\curvearrowright}+k)^{\curvearrowright} && \text{[Inductive Hypothesis]} \\ &= m^{\curvearrowright} + k^{\curvearrowright} && \text{[Def. of +]} \end{aligned}$$

So $(m+n)^{\curvearrowright} = m^{\curvearrowright} + n$. Because the proof does not depend on any assumption about m, it is valid for all m. □

Roughly speaking this lemma permits us to move $\curvearrowright$ anywhere within an addition: $m^{\curvearrowright} + n = (m+n)^{\curvearrowright} = m + n^{\curvearrowright}$. So we are free to move a successor "out of the way" whenever we need to. The next proof illustrates the point.

Proposition 5

Addition is commutative.

Proof: We need to show that $m + n = n + m$ for all m and n. This time, the proof is by induction on m. Fix a value for n.

- [Basis] $0 + n = n = n + 0$ holds because of Proposition 3 and the definition of $+$.
- [Inductive Hypothesis] Assume that $k + n = n + k$ for some k.
- [Inductive Step] We must show that $k^{\curvearrowright} + n = n + k^{\curvearrowright}$.

$$\begin{aligned} k^{\curvearrowright} + n &= (k + n)^{\curvearrowright} && \text{[Lemma 4]} \\ &= (n + k)^{\curvearrowright} && \text{[Inductive Hypothesis]} \\ &= n + k^{\curvearrowright} && \text{[Def. of } +\text{]} \end{aligned}$$

Therefore, $m+n = n+m$ for all m. Since this argument does not depend on any assumptions about n, it is valid for all n. □

The next law may be less familiar to you. Roughly, it says that we can "subtract" equals and get equals. But note that actual subtraction does not always make sense for natural numbers. We can not, for example, say what $5 - 7$ means without introducing negative numbers.

Proposition 6

Addition is cancellative.

Proof: We need to prove that if $m + p = n + p$, then $m = n$. This proof is a little subtler than the previous ones. But notice that is still follows the same form.

The proof is by induction on p. Assume that m and n are some fixed natural numbers.

- [Basis] Suppose $m + 0 = n + 0$. Then immediately by definition of $+$, $m = n$.
- [Inductive Hypothesis] Assume that the following statement is true for some k: if $m + k = n + k$ then $m = n$.
- [Inductive Step] We must show that if $m + k^{\curvearrowright} = n + k^{\curvearrowright}$ then $m = n$. Suppose $m + k^{\curvearrowright} = n + k^{\curvearrowright}$ [call this (*) for reference]. Then

$$\begin{aligned}(m+k)^{\curvearrowright} &= m + k^{\curvearrowright} && \text{[Def. of +]}\\ &= n + k^{\curvearrowright} && \text{[By the supposition (*)]}\\ &= (n+k)^{\curvearrowright} && \text{[Definition of +]}\end{aligned}$$

  Hence, by Axiom **??** $m + k = n + k$. So by the Inductive Hypothesis, $m = n$.

Therefore, $m + p = n + p$ implies $m = n$ for all p. Since this argument does not depend on any assumptions regarding m and n, it is valid for all m and n. □

To prove that multiplication is commutative and cancellative, we will need the following technical facts (analogous to Proposition 3 and Lemma 4).

Lemma 7

For any n, $0 \cdot n = 0$

Proof: The proof is by induction on n.

- [Basis] $0 \cdot 0 = 0$ by definition of $\cdot$.
- [Inductive Hypothesis] Assume that $0 \cdot k = 0$ for some k.

Lemma 7 (cont.)

- [Inductive Step] We must show that $0 \cdot k^{\curvearrowright} = 0$.

$$\begin{aligned} 0 \cdot k^{\curvearrowright} &= 0 + 0 \cdot k && \text{[Definition of $\cdot$]} \\ &= 0 + 0 && \text{[Inductive Hypothesis]} \\ &= 0 && \text{[Definition of $+$]} \end{aligned}$$

□

Lemma 8

For any m and n, $m^{\curvearrowright} \cdot n = m \cdot n + n$

Proof: The proof is by induction on n.

- [Basis] $m^{\curvearrowright} \cdot 0 = 0 = 0 + 0 = m \cdot 0 + 0$ all follow from the definitions of $+$ and $\cdot$.
- [Inductive Hypothesis] Assume that $m^{\curvearrowright} \cdot k = m \cdot k + k$ for some k.
- [Inductive Step] We must show that $m^{\curvearrowright} \cdot k^{\curvearrowright} = m \cdot k^{\curvearrowright} + k^{\curvearrowright}$.

$$\begin{aligned} m^{\curvearrowright} \cdot k^{\curvearrowright} &= m^{\curvearrowright} + m^{\curvearrowright} \cdot k && \text{[Exercise]} \\ &= (m + m^{\curvearrowright} \cdot k)^{\curvearrowright} && \text{[Exercise]} \\ &= (m + (m \cdot k + k))^{\curvearrowright} && \text{[Exercise]} \\ &= ((m + m \cdot k) + k)^{\curvearrowright} && \text{[Exercise]} \\ &= (m \cdot k^{\curvearrowright} + k)^{\curvearrowright} && \text{[Exercise]} \\ &= m \cdot k^{\curvearrowright} + k^{\curvearrowright} && \text{[Exercise]} \end{aligned}$$

□

Some of the other laws are left as exercises.

Exercises for Lecture 3

1. Prove that 1 is the identity for multiplication. That is $1 \cdot m = m = m \cdot 1$.

2. Write out the entire proof of Lemma 8 providing the justifications for each line of the equational calculation in the Inductive Step.

3. Prove that multiplication distributes over addition $[m \cdot (n + p) = m \cdot n + m \cdot p]$ by induction on p. You can use the any of the lemmas and propositions we have already proved.

    1. Prove the basis: $m \cdot (n + 0) = m \cdot n + m \cdot 0$.
    2. Write the inductive hypothesis.
    3. Prove the inductive step: $m \cdot (n + k^{\curvearrowright}) = m \cdot n + m \cdot k^{\curvearrowright}$

4. Prove that multiplication is associative $[m \cdot (n \cdot p) = (m \cdot n) \cdot p]$ by induction on p.

    1. Prove the basis: $m \cdot (n \cdot 0) = (m \cdot n) \cdot 0$.
    2. Write the inductive hypothesis.
    3. Prove the Inductive Step: $m \cdot (n \cdot k^{\curvearrowright}) = (m \cdot n) \cdot k^{\curvearrowright}$. Hint: Use the Law of Distribution, which you just proved.

5. Prove that multiplication is commutative. Hint: Use the two Lemmas we proved right before these exercises.

Lecture 4

Lists

Natural numbers constitute an important example of something more general, where objects are built up from simpler ones. The Axiom of Induction captures the idea of building "up" and provides an important method for proving facts about natural numbers.

In this lecture, we develop an analogous way to think about *lists*.

Goals

Lecture Goals

- Introduce a formal counterpart to the informal concept of a list
- Emphasize the close analogy between lists and natural numbers
- Introduce basic operations on lists.

Study Goals

- Demonstrate facility with basic list manipulation including calculating length and concatenation of lists.

4.1 List Basics

In this section, we concentrate on the fundamental concept of *lists*. The idea is really meant to be the familiar one, so a list of "to do" items is a list. The alphabetized names on a class roster is a list. We will write lists using square brackets. So for example, $[2, 3, 5, 7]$ is the list of the prime numbers less than 10 in ascending order. For lists, we expect the order to matter. So $[7, 5, 3, 2]$ is a different list.

Something that occurs on a list is called an *item* of the list. We can even specify where it is. So we can talk about the "first", "second" item, and so on, assuming the list has enough items.

Because we have already agreed that natural numbers begin with 0, it turns out to make many things easier if we change the way we talk about items on a list to gibe with the natural numbers. So instead of refering to the "first" item, we might call it the "initial" item. Furthermore, we will number them to start with 0. What I mean is that if $L = [2, 3, 5, 7]$, we will write L_0, L_1, L_2, L_3 for the elements 2, 3, 5, 7, respectively. In short, the "initial" item is indexed by the "initial" natural number 0. The next item after that is indexed by next natural number, $0^\frown$, and so on.

Like natural numbers, lists can be built up by starting with an empty list and incrementally adding items. We have choices for how we might formalize the idea. We will follow a standard that has developed in computer science. Clearly, since we use square brackets to punctuate lists, the empty list should be written as $[\,]$. To add an item to a list, we will conventionally put it on the front.

Given the list $[x, y, z]$, we may build a new list with initial item w and the given list as the rest, resulting in $[w, x, y, z]$. The operation of *prepending* an item to a list is denoted by a colon (:). So $w : [x, y, z]$ *is* the list $[w, x, y, z]$.

The empty list, together with prepending items, gives us a way to construct any list we want.

Example 1

Here are some examples.

- $5 : 6 : [4, 5]$ is the same as $5 : [6, 4, 5]$, which is the same as $[5, 6, 4, 5]$.
- $[\,]$ is the empty list
- $1 : [\,]$ is the same as $[1]$
- $1 : 2 : 3 : 4 : [\,]$ is the same as $[1, 2, 3, 4]$.

Notice that every list is either empty ($[\,]$) or not. If not, it has the form $x : L$ where x is the initial item and L is the rest of the list. This suggests a signature for lists, not so different from the signature for natural numbers.

Postulate 2: Basic Structure of Lists

Lists have the following basic structure.

- There is a special list, which we call *the empty list* and denote by $[\,]$.
- For any thing x and any list L, there is another list, obtained by *prepending* x to L. We denote the result by $x : L$.

As with the natural numbers, we need to think about axioms that prevent strange behavior. These are exactly analogous to the axioms of natural numbers. First, $[\,]$ can not be obtained by adding a new initial item to another list. So

Postulate 3

For any list L and any thing x, $[] \neq x : L$.

Likewise, a list that is not empty can only be built one way.

Postulate 4

For any things x and y and lists L and M, if $x : L = y : M$, then $x = y$ and $L = M$.

For example, if I tell you that $[2, 3, 4, 5] = x : L$, then you know immediately that $x = 2$ and $L = [3, 4, 5]$.

Finally, lists need an induction axiom that ensures that all lists are built up from $[]$.

Postulate 5: The Axiom of List Induction

No lists can be removed without violating Postulate **??**.

This axiom justifies conducting proofs about all lists by a scheme almost identical to simple arithmetic induction. That is, to prove some property is true about all lists, it is enough to show

- [Basis] The property is true about $[]$.
- [Inductive Hypothesis] Assume that the property is true for from list K.
- [Inductive Step] Prove that for any thing x, the property is true about $x : K$. [You may use the assumption about K in this part of the proof.]

Operations on lists can now also be defined by schemes similar to how we defined addition and multiplication on natural numbers. For example, every list has a length. Writing $\mathsf{len}(L)$ for the length of a list, $\mathsf{len}([2, 3, 4]) = 3$. A precise definition is now easy to formulate.

Definition 6

For a list L. the *length* of L, denoted by $\mathsf{len}(L)$, is the natural number. This satisfies the following equalities.

$$\mathsf{len}([]) = 0$$
$$\mathsf{len}(x : L) = len(L)^{\curvearrowright}$$

Example 7

$$\begin{aligned}
\mathsf{len}([2,3,4]) &= \mathsf{len}(2:[3,4]) \\
&= \mathsf{len}([3,4])^{\curvearrowright} \\
&= \mathsf{len}(3:[4])^{\curvearrowright} \\
&= \mathsf{len}([4])^{\curvearrowright\curvearrowright} \\
&= \mathsf{len}(4:[\,])^{\curvearrowright\curvearrowright} \\
&= \mathsf{len}(,])^{\curvearrowright\curvearrowright\curvearrowright} \\
&= 0^{\curvearrowright\curvearrowright\curvearrowright} \\
&= 3
\end{aligned}$$

Another common operation on lists is *concatenation*: $[2,3,4] \otimes [4,1,3] = [2,3,4,4,1,3]$, whereby the two lists are simply glued together in their original orders. This is defined precisely by the following.

Definition 8

For lists L and M, their *concatenation*, denoted by $L \otimes M$, is a list. For all lists M, the following are true.

$$\begin{aligned}
[\,] \otimes M &= M \\
(x:K) \otimes M &= x:(K \otimes M) && \text{for any thing } x \text{ and any list } K
\end{aligned}$$

Example 9

To calculate $[4,5,2,1] \otimes [3,4,1]$, we can follow a method similar to arithmetic:

$$\begin{aligned}
[4,5,2,1] \otimes [3,4,1] &= (4:5:2:1:[\,]) \otimes [3,4,1] && [[4,5,2,1] \text{ abbreviates } 4:5:2:1:[\,]] \\
&= 4:((5:2:1:[\,]) \otimes [3,4,1]) && [\text{Def. of } \otimes] \\
&= 4:5:((2:1:[\,]) \otimes [3,4,1]) && [\text{Same}] \\
&= 4:5:2:((1:[\,]) \otimes [3,4,1]) && [\text{Same}] \\
&= 4:5:2:1:([\,] \otimes [3,4,1]) && [\text{Same}] \\
&= 4:5:2:1:[3,4,1] && [\text{Same}] \\
&= [4,5,2,1,3,4,1] && [\text{Abbreviation}]
\end{aligned}$$

Now we can prove some useful facts about lists.

Lemma 10

On lists, $[]$ is the identity for $\otimes$,

Proof: By definition $[] \otimes L = L$ always true. But $[]$ must also satisfy $L \otimes [] = L$ always. We can proceed by induction on L. The proof should look familiar (see the proof of Lemma **??**).

- [Basis] $[] \otimes [] = []$ is true by definition of $\otimes$.
- [Inductive Hypothesis] Assume $K \otimes [] = K$ for some list K.
- [Inductive Step] Suppose x is some thing. We need to show that $(x : K) \otimes [] = x : K$.

$$\begin{aligned} (x : K) \otimes [] &= x : (K \otimes []) && \text{[by definition of } \otimes] \\ &= x : K && \text{[by the Inductive Hypothesis]} \end{aligned}$$

Thus (by the Axiom of List Induction), the lists for which $L \otimes [] = L$ constitute all lists. □

Lemma 11

On lists, $\otimes$ is associative.

Proof: We prove $L \otimes (M \otimes N) = (L \otimes M) \otimes N$ using induction on L. This should look familiar. It is almost identicial to the proofs that addition and multiplication are associative.

- [Basis] $[] \otimes (M \otimes N) = M \otimes N = ([] \otimes M) \otimes N$. Both steps are by the definition of $\otimes$.
- [Inductive hypothesis] Suppose $K \otimes (M \otimes N) = (K \otimes M) \otimes N$ for some particular list K.
- [Inductive step]

$$\begin{aligned} (x : K) \otimes (M \otimes N) &= x : (K \otimes (M \otimes N)) && \text{Def. of } \otimes \\ &= x : ((K \otimes M) \otimes N) && \text{Inductive Hypothesis} \\ &= (x : (K \otimes M)) \otimes N && \text{Def. of } \otimes \\ &= ((x : K) \otimes M) \otimes N && \text{Def. of } \otimes \end{aligned}$$

Lemma 11 (cont.)

So $L \otimes (M \otimes N) = (L \otimes M) \otimes N$ is true for all L. Since the proof does not depend on any special propertis of M and N (except that they are both lists), the result is true for all lists M and N. □

Here is another nice fact that we can prove by induction relating length to concatenation.

Lemma 12

For any lists L and M, $\mathsf{len}(L \otimes M) = \mathsf{len}(L) + \mathsf{len}(M)$.

Proof: [This claim is probably fairly obvious to you. Nevertheless, to illustrate the technique of list induction again, we prove it explicitly.]

- [Basis] $\mathsf{len}([\,]) + \mathsf{len}(M) = 0 + \mathsf{len}(M) = \mathsf{len}(M) = \mathsf{len}([\,] \otimes M)$. These are by definition of $\otimes$ and $+$.
- [Inductive Hypothesis] Suppose $\mathsf{len}(K \otimes M) = \mathsf{len}(K) + \mathsf{len}(M)$ holds for some particular list K.
- [Inductive Step]

$$\begin{aligned}
\mathsf{len}((x : K) \otimes M) &= \mathsf{len}(x : (K \otimes M)) && \text{Def. of } \otimes \\
&= \mathsf{len}(K \otimes M)^{\curvearrowright} && \text{Def. of } \mathsf{len} \\
&= (\mathsf{len}(K) + \mathsf{len}(M))^{\curvearrowright} && \text{Inductive Hypothesis} \\
&= \mathsf{len}(K)^{\curvearrowright} + \mathsf{len}(M) && \text{Lemma 4} \\
&= \mathsf{len}(x : K) + \mathsf{len}(M) && \text{Def. of } \mathsf{len}
\end{aligned}$$

□

Often we will use a list somewhat informally without all the punctuation. For example, we might say "Consider a list $a_0, a_1, \ldots, a_{n-1}$ of real numbers." If we do not intend to use the list itself for anything special, but only want to think about the numbers a_0 through a_n, then there is no need to be formal about it. Also, there is no harm in writing something like this: a_5, a_6, a_7, a_8, where the indices start at 5. The default is to start at 0, but that is merely a convention.

Lemma 13

$\otimes$ is cancellative on the left and on the right. That is,

Lemma 13 (cont.)

- $L \otimes M = L \otimes N$ implies $M = N$; and
- $L \otimes N = M \otimes N$ implies $L = M$.

Proof: Exercise. □

4.2 List Itemization

In a list L, the items are in order. So we can refer to items by their position in the list. There are two standards in mathematics for doing this. Either we start counting from 1 or from 0. Although it may seem unintuitive at first to start from 0 (meaning that the "initial" item of a list is item number 0), this actually makes many calculations simpler. For that reason, most programming languages use this convention for a lists and arrays. So I will consistently start with 0.

The idea can be made precise as follows.

Definition 14

Suppose L is a list and $i < \mathsf{len}(L)$. Then L_i is an item on the list defined as follows.

$$\begin{aligned} [\,]_i &\text{is never defined because } 0 \not< \mathsf{len}([\,]) \\ (x : L)_0 &= x \\ (x : L)_{k^\frown} &= L_k && \text{provided that } L_k \text{ is defined} \end{aligned}$$

This is a precise way of explaining that in a list, for example $L = [a, b, c, d, e]$, we can refer to an item by its *index*, so that $L_0 = a$, $L_1 = b$ and so on, up to $L_4 = e$. Notice that L_k is undefined if $k \geq \mathsf{len}(L)$.

Example 15

Suppose $L = [a, b, c, d, e]$. We can calculate L_3 explicitly step by step.

$$\begin{aligned} L_3 &= [a, b, c, d, e]_3 \\ &= (a : b : c : d : e : [\,])_{0\curvearrowright\curvearrowright\curvearrowright} \\ &= (b : c : d : e : [\,])_{0\curvearrowright\curvearrowright} \\ &= (c : d : e : [\,])_{0\curvearrowright} \\ &= (d : e : [\,])_0 \\ &= d \end{aligned}$$

Of course, this is just a very careful (you might even say fussy) way to find item number 3 in the list. In every day use, we humans would not do this. We would simply count forward from the beginning of the list.

Exercises for Lecture 4

1. Suppose $L = [3, 2, 3, 3, 5]$ and $M = [0, 1, 2, 3, 4, 5]$. Calculate the following explicitly step by step.

   1. $\mathsf{len}(L)$
   2. L_4
   3. $(L \otimes M)_9$

4.3 Lists of a Particular Type

We will commonly need to consider lists in which all elements are similar, such as a list consisting of natural numbers. For example, because we know how arithmetic operations work on natural numbers, we can also define operations on lists of natural numbers using arithmetic. Similar extensions are possible for other operations defined on other types of elements.

To illustrate, suppose L is a list of natural numbers. We can define the *sum* of items on the list in the obvious way, so that the sum of the list $[2, 3, 4]$ is $2 + 3 + 4 = 9$. We make this precise with the following.

Definition 16

For a list L of natural numbers, the *sum of* L, denoted by $\sum L$, is a natural number, satisfying

$$\begin{aligned} \sum [\,] &= 0 \\ \sum m : L &= m + \sum L \qquad \text{for any natural number } m \text{ and any list of natural numbers } L \end{aligned}$$

We will introduce variations and extensions of this notation this later. For now, we look only at lists.

Exercises for Lecture 4

1. Prove using list induction that for any lists of natural numbers,
$$\sum L + \sum M = \sum (L \otimes M)$$
2. Define the product of lists of natural numbers, following the pattern of our definition for $\sum L$. The standard notation for a product is $\prod L$. The result should be that $\prod[2,3,4]$ equals 24. Pay close attention the base case $\prod[\,]$.
3. Using your definition of products, prove by list induction that for any lists of natural numbers,
$$\prod L \cdot \prod M = \prod (L \otimes M)$$
.

We can also consider lists of integers, lists of real numbers, and so on. We can even think about lists of lists. For example, $[[2,3,4],[4,3,2],[5]]$ is a list consisting of two items: $[2,3,4]$, $[4,3,2]$ and $[5]$. Written this using :, this is list is $[2,3,4] : [4,3,2] : [5] : [\,]$. Suppose we have a list of lists like this we can define the concatenation of all the items. For this example, the result should be $[2,3,4,4,3,2,5]$. The definition of this is exactly analogous to sums and products.

Definition 17

For a list $\mathcal{L}$ of lists, the *fold of* $\mathcal{L}$, denoted by $\bigotimes \mathcal{L}$, is a list, satisfying

$$\begin{aligned} \bigotimes [\,] &= [\,] \\ \sum M : \mathcal{L} &= M \otimes \otimes \mathcal{L} \qquad \text{for any list } M \text{ and any list of lists } \mathcal{L} \end{aligned}$$

Compare the definitions of $\sum$, $\prod$ and $\otimes$. They differ only in terms of (i) what is the result for an empty list and (ii) what binary operation is used in the second equation.

Suppose we are given a list $\mathcal{L}$ of lists of natural numbers (like the example just above the latest definition). Then its fold is a list of natural numbers. So this can be summed. That is, $\bigotimes \mathcal{L}$ is a list of natural numbers, and $\sum(\otimes\mathcal{L})$ is a natural number. But we might also apply the summation operation to each list on $\mathcal{L}$ separately, resulting in another list of natural numbers. The idea of applying an operation to each element of a list is called "mapping". In this case, we intend to "map" the operation $\sum$ across lists of lists of natural numbers. Here is a suitable definition.

Definition 18

For a list $\mathcal{L}$ of lists of natural numbers, the *mapping of* $\sum$ *on* $\mathcal{L}$, denoted by $\mathsf{map}_{\sum}(\mathcal{L})$ is a list of natural numbers, satisfying

$$\bigotimes[\,] = [\,]$$

$$\sum M : \mathcal{L} = (\,) \sum M) : \mathcal{L} \quad \text{for any list of natural numbers } M \text{ and any list of lists of natural numbers } \mathcal{L}$$

Exercises for Lecture 4

1. Calculate $\sum(\otimes[[3,4,5],[6,3]])$.
2. Calculate $\mathsf{map}_{\sum}([[3,4,5],[6,3]])$.
3. Calculate $\sum(\mathsf{map}_{\sum}([[3,4,5],[6,3]]))$.
4. Prove that $\sum(\otimes\mathcal{L}) = \sum(\mathsf{map}_{\sum}(\mathcal{L}))$ for any list of lists of natural numbers $\mathcal{L}$.

4.4 Other Inductively Defined Collections

The structure of natural numbers and the structure of lists are very similar. This similarity can be exploited to develop a simple way of summarizing their properties.

For natural numbers, 0 and $n^{\curvearrowright}$ are the only ways to construct them. Operations like addition and multiplication are defined in terms of 0 and ${}^{\curvearrowright}$, so they do not contribute directly to the *construction* of natural numbers. So we refer to 0 and $\curvearrowright$ as *constructors*.

Axioms **??** and **??** spell out how these constructors behave. Namely, Axiom **??** captures the idea that the two constructors are entirely different from the other: $0 \neq n^{\curvearrowright}$. Axiom **??** captures the idea that ${}^{\curvearrowright}$ constructs distinct natural numbers from distinct natural numbers: $m^{\curvearrowright} = n^{\curvearrowright}$ implies $m = n$ (or equivalently, $m \neq n$ implies $m^{\curvearrowright} \neq n^{\curvearrowright}$).

So the basic ingredients of natural numbers are the constructors 0 and $^\curvearrowright$ with the understanding that (a) each produces different results and (b) from different ingredients, $^\curvearrowright$ produces different results. In fact, point (b) also applies to 0 trivially, because 0 does not use any ingredients.

We can summarize everything we want to say about natural numbers concisely in the following way.

Definition 19

The *natural numbers* are defined *inductively* by

$$n := 0 \mid n^\curvearrowright$$

In this notation, the vertical bar separates the different constructors for natural numbers. The first constructor (0) does not depend on anything else. The second constructor depends on a natural number n and produces a new one $n^\curvearrowright$. So this gives a very concise description of the signature of natural numbers. Implicitly, this notation is meant to indicate that the two alternatives are completely distinct. This is Axiom **??**. Also implicitly, the notation is meant indicate that $n^\curvearrowright$ produces distinct results from distinct n's. This is Axiom **??**. By declaring saying that this defines natural numbers *inductively*, we also mean that no natural numbers can be removed without violating the signature.

Now let's consider lists. Again, there are two ways to construct lists. $[]$ and $x : L$ for any thing x and any list L. Likewise, the constructrs are distinct, and $x : L = y : M$ is true if and only if both $x = y$ and $L = M$. So we can encapsulate the definition of lists similarly.

Definition 20

The *lists* are defined *inductively* by

$$L := [] \mid x : L \qquad\qquad \text{for any thing } x$$

Notice that x can also be a list.

Later in the course, we will make these definitions, and many others like them, rigorous. For now, we just draw attention to the similarity between natural numbers and lists, and point out that proofs by induction work thanks to the structure of these definitions.

4.5 Binary Trees

Simple binary trees are structures that play a role in many parts of computer science and mathematics. Figure 4.1 illustrates an example. There are many variations on the basic idea, but we concentrate on the simplest version (where there is no extra structure).

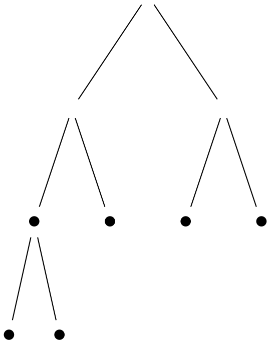


Figure 4.1: A simple binary tree

Such structures are built from *leaves* (denoted here by $\bullet$) by “grafting” two smaller trees to form a larger one as pictured in Figure 4.2. Trees are usually depicted “upside down”, so the *root* is at the top and the leaves are at the bottom. What can you do? It’s tradition.

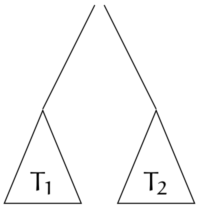


Figure 4.2: Constructing a tree from subtrees

In addition to drawing pictures of binary trees, we can use a linear notation (something that can we written in the midst of prose). If T_1 and T_2 are simple binary trees, we may denote the tree constructed by grafting T_1 on the left and T_2 on the right as $(T_1 \curlywedge T_2)$ (as in Figure 4.2. Thus we define simple binary trees, and two operations on them, as follows.

Definition 21

Simple binary trees are defined inductively by

$$T \coloneqq \bullet \mid (T_1 \curlywedge T_2)$$

The *size* of a simple binary tree is a natural number defined by the equations

$$\begin{aligned} \mathsf{sz}(\bullet) &\coloneqq 0 \\ \mathsf{sz}(T_1 \curlywedge T_2) &\coloneqq 1 + \mathsf{sz}(T_1) + \mathsf{sz}(T_2) \end{aligned}$$

The *height* of a simply binary tree a natural number is defined by the equations

$$\begin{aligned} \mathsf{ht}(\bullet) &:= 0 \\ \mathsf{ht}(T_1 \curlywedge T_2) &:= 1 + \max(\mathsf{ht}(T_1), \mathsf{ht}(T_2)) \end{aligned}$$

where $\max(m, n)$ is the larger of the two numbers.

These definitions suggest a relation between the size and height of a tree.

Lemma 22

For any simple binary tree T,

$$\mathsf{ht}(T) \leq \mathsf{sz}(T) < 2^{\mathsf{ht}(T)}.$$

Proof: To prove this by *structural induction*, we must prove it for the basis (•) and that if it is true for some T_1 and T_2, then it is true for $(T_1 \wedge T_2)$.

- [Basis] $\mathsf{ht}(\bullet) = 0$, $\mathsf{sz}(\bullet) = 0$ and $2^{\mathsf{ht}(\bullet)} = 1$. So the claim is true for •.
- [Inductive Hypothesis] Suppose the inequalities holds for some T_1 and T_2.
- [Inductive Step] We must prove the two inequalities for $T = (T_1 \curlywedge T_2)$. Without loss of generality, assume that $\mathsf{ht}(T_1) \leq \mathsf{ht}(T_2)$. That is, if this is not so, then we may swap T_1 for T_2 in the following.

$$\begin{aligned} \mathsf{ht}(T) &= 1 + \max(\mathsf{ht}(T_1), \mathsf{ht}(T_2)) && \text{[Definition of ht]} \\ &\leq 1 + \mathsf{ht}(T_1) + \mathsf{ht}(T_2) && \text{[Arithmetic]} \\ &\leq 1 + \mathsf{sz}(T_1) + \mathsf{sz}(T_2) && \text{[Inductive Hypothesis]} \\ &= \mathsf{sz}(T) && \text{[Definition of sz]} \end{aligned}$$

And

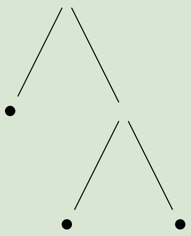
$$\begin{aligned} \mathsf{sz}(T) &= 1 + \mathsf{sz}(T_1) + \mathsf{sz}(T_2) && \text{[Definition of sz]} \\ &\leq 1 + 2^{\mathsf{ht}(T_1)} - 1) + (2^{\mathsf{ht}(T_2)} - 1) && \text{[Inductive Hypothesis]} \\ &\leq 2 \cdot 2^{\mathsf{ht}(T_2)} - 1 && \text{[Assumption that } \mathsf{ht}(T_1) \leq \mathsf{ht}(T_2)\text{]} \\ &= 2^{\mathsf{ht}(T_2)+1} - 1 && \text{[Arithmetic]} \\ &= 2^{\mathsf{ht}(T)} - 1 && \text{[Definition of ht]} \end{aligned}$$

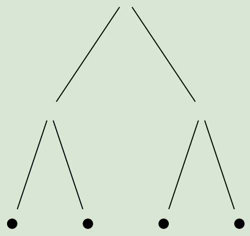
So $\mathsf{sz}(T) < 2^{\mathsf{ht}(T)}$

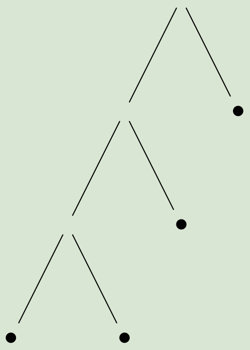
□

Exercises for Lecture 4

1. Calculate the height and size of the following simple binary trees.
   1. $(\bullet \curlywedge \bullet)$
   2. $(\bullet \curlywedge (\bullet \curlywedge \bullet))$
   3. $((\bullet \curlywedge \bullet) \curlywedge (\bullet \curlywedge (\bullet \curlywedge \bullet)))$
   4. $((\bullet \curlywedge (\bullet \curlywedge \bullet)) \wedge ((\bullet \curlywedge \bullet) \curlywedge (\bullet \curlywedge (\bullet \curlywedge \bullet))))$
2. Draw diagrams (similar to those in Figure 4.1) for the following simple binary trees.
   1. $((\bullet \curlywedge \bullet) \curlywedge (\bullet \curlywedge \bullet))$
   2. $(((\bullet \curlywedge \bullet) \curlywedge \bullet) \curlywedge ((\bullet \curlywedge \bullet) \curlywedge (\bullet \curlywedge (\bullet \curlywedge \bullet))))$
3. For each of the following diagrams, write the expression using $\bullet$ and $\curlywedge$ defining the same tree.

1. 

2. 

3. 

We will study binary trees in more depth later in the course.

LECTURE **5**

Ordering the Natural Numbers by Addition

We have used the ordering of numbers informally without much comment because $\leq$ has its obvious meaning. In this lecture we exploit the fact that $\leq$ on the natural numbers is defined by addition, setting the stage for an analogous concept defined by multiplication. The multiplicative analogue of "m is less than or equal n" is "m divides n".

Goals

Lecture

- Introduce a formal definition of $\leq$ on natural numbers, and discuss simple facts about order
- Review basic laws involving min and max.
- Introduce Strong Induction and the Principle of Well-foundedness as alternatives to Simple Induction.

Study

- Prove basic facts about $\leq$, min and max.
- Commit to memory the concepts of reflexivity, transitivity, anti-symmetry and linearity.
- Practice using Strong Induction.

5.1 Less or Equal

For the natural numbers, $\leq$ has a particularly simple definition.

Definition 1

For natural numbers n and m, say that m *is less than or equal to* n (written $m \leq n$) if and only if there is a natural number d so that $m + d = n$. [I've used the letter d because d is the

Definition 1 (cont.)

difference of m from n.]

Also, say that m **is (strictly) less than** n if and only if there is a natural number d so that $m + d^{\frown} = n$.

All of the usual properties of $\leq$ for natural numbers follow easily from this definition using only the laws of addition. The important properties have names.

Laws 2

Reflexivity $\leq$ is *reflexive*: $m \leq m$ is true for any m. This is simply because because $m+0 = m$.

Transitivity $\leq$ is *transitive*: if $m \leq n$ and $n \leq p$, then $m \leq p$. Transitivity follows from the law of associativity for addition (you will prove this as an exercise).

Anti-symmetry $\leq$ is *anti-symmetric*: if $m \leq n$ and $n \leq m$, then $m = n$. This follows from cancellativity and positivity (another exercise).

Linearity $\leq$ is *linear*: for any m and n, either $m \leq n$ or $n \leq m$. This requires a separate proof using induction, dealt with below.

Clearly, $\leq$ is also meaningful for integers, rational numbers and real numbers, and is still reflexive, transitive and anti-symmetric and linear for them. For our purposes, though, the interesting features are already present in the natural numbers. For one thing, if $m + d = n$ and $m + e = n$ then $d = e$ (why?). This tells us that $m \leq n$ can only be true for one reason.

Exercises for Lecture 5

In the following you will use the Laws of Arithmetic from Lecture 3 to prove the basic facts about $\leq$.

1. Show that $\leq$ is transitive by the following steps.
    1. Assume that $m \leq n$. By definition, there is some d_0 so that $m + d_0 = n$
    2. Assume that $n \leq p$. Write out what this means. “By definition, there is some …”
    3. Now find an e so that $m + e = p$.
    4. Write out the conclusion.
2. Show that $\leq$ is anti-symmetric by the following steps.
    1. Assume $m \leq n$. Write out what this means: “There is some d_0 so that ….”

Exercises for Lecture 5 (cont.)

2. Assume $n \leq m$. Write out what this means. “There is some d_1 so that . . . .”
3. Show that $d_0 + d_1 = 0$. [Think about using cancellativity.]
4. Conclude that $d_0 = 0$. [Which law justifies this?]
5. Conclude that $m = n$.

As we mentioned, $\leq$ is *linear*, meaning that for any m and n, either $m \leq n$ or $n \leq m$. A proof of this is not terribly difficult, but is more subtle than transitivity and anti-symmetry.

Lemma 3

$\leq$ is linear.

Proof: Note that we defined linearity in terms of $\leq$, but it is equivalent to saying that for any m and n, either $m < n$)or $n \leq m$. After all, if $m < n$ then $m \leq n$. And if $m \leq n$, then either $m = n$ and hence $n \leq n$, or $m < n$. A proof is by induction on m.

- [Basis] $0 + n = n$ is always true. So $0 \leq n$.
- [Inductive Hypothesis] Suppose that for some k, it is the case that either $k \leq n$ or $n \leq k$ (but we do not know which comparison is true).
- [Inductive Step] We must show that either $k^{\curvearrowright} \leq n$ or $n \leq k^{\curvearrowright}$. According to the Inductive Hypothesis, there are two cases to consider. Either $k \leq n$ or $n \leq k$. We take them in reverse.

  1. Suppose $n \leq k$. Then obviously $n \leq k^{\curvearrowright}$ (meaning, it is so obvious that we do not bother to fill in the details).
  2. Suppose $k < n$. So for some d, $k + d^{\curvearrowright} = n$. Hence $k^{\curvearrowright} + d = n$. if $d = 0$, then $k^{n}xt = n$, so $n \leq k^{\curvearrowright}$. Otherwise, $k^{\curvearrowright} < n$.

□

Exercises for Lecture 5

The following proofs should require you only to refer to the definition of $\leq$ and to the basic Laws of Arithmetic.

1. Prove that addition is **monotonic** with respect to $\leq$. This means that $m \leq n$ implies $m + p \leq n + p$ for all natural numbers m, n and p.

Exercises for Lecture 5 (cont.)

2. Prove that addition is **order reflectiing** with respect to $\leq$. This means that if $m + p \leq n + p$, then $m \leq n$.

5.2 Other Forms of Induction

Using $\leq$, we can formulate a more flexible method of proof by induction.

Suppose we wish to prove that all natural numbers are *outgrabe*. Again this is nonsense, but let's try anyway. We prove the basis with no trouble; formulate the Inductive Hypothesis; get stuck on the Inductive Step. Here is a simple idea that can help. Define a *hypergrabe* number to be a natural number k so that for every $j < k$, j is outgrabe. In other words, k does not necessarily have the property we really care about, but all the numbers below it do. Now suppose every natural number is outgrabe. Then clearly, every natural number is also hypergrabe. Likewise, if every natural number is hypergrabe, then every natural number is outgrabe. So proving either property by induction will suffice to prove the other. A proof by simple induction that all natural numbers are hypergrabe amounts to this:

- [Basis] Show that every natural number $j < 0$ is outgrabe. But since there are no such numbers, this is trivially true no matter what outgrabe happens to mean.

- [Inductive Hypothesis] Suppose that k is hypergrabe for some k. That is, suppose that for all $j < k$, j is outgrabe.

- [Inductive Step] Prove that $k^\frown$ is hypergrabe. By the Inductive Hypothesis, all $j < k$ are outgrabe. So to prove that every $j < k^\frown$ is outgrabe just amounts to proving it for k. So the Inductive Step (for proving that $k^\frown$ is hypergrabe) amounts to proving that k is outgrabe.

This can now be reformulated without mentioning *hypergrabe* at all. The Basis is not needed, because it is true automatically. The Inductive Hypothesis can be reformulated as supposing that for all $j < k$, j is outgrabe. The proof of the Inductive Step amounts to proving that k is outgrabe.

This leads to a formulation of induction that is typically called *Strong Induction*, even though it is not really any stronger (i.e., it is not able to prove more facts). A proof of property P by Strong Induction has the outline:

- [Strong Inductive Hypothesis] Suppose that $P(j)$ is true for all $j < k$.

- [Strong Inductive Step] Prove that $P(k)$ is true.

Exercises involving strong inductive proofs are not easy to come by right now. We will see a very important example when we prove that every positive natural number can be factored into primes.

Example 4

Here is an example that uses strong induction to prove something. Define Fibonnacci number f_n by the equations

$$\begin{aligned} f_0 &= 0 \\ f_1 &= 1 \\ f_{n+2} &= f_{n+1} + f_n && \text{for any } n \end{aligned}$$

We claim that for all n, $2f_n \leq f_{n+2}$.

- [Strong Inductive Hypothesis] Suppose that for all $j < k$, $2f_j \leq f_{j+2}$.
- [Strong Inductive Step]. We need to show that the claim also holds for k. In case $k = 0$, this is obviously true because $2f_0 = 0$. In case $k = 1$, it is also obviously true because $2f_1 = 2 = f_3$. In all other cases $k = i + 2$ for some i. So

$$\begin{aligned} 2f_k &= 2f_i + 2f_{i+1} && \text{[By definition of } f \text{ and distributivity]} \\ &\leq f_{i+2} + f_{i+3} && \text{[By Inductive Hypothesis]} \\ &= f_k + f_{k+1} && [i + 2 = k] \\ &= f_{k+2} && \text{[By definition of } f\text{]} \end{aligned}$$

We close this section with another method for using induction that is sometimes useful.

Lemma 5

Suppose $P(-)$ is a property that makes sense for natural numbers. Suppose, furthermore, that $P(m)$ is true for some m. Then there is a smallest m for which $P(m)$ is true. That is, there is a natural number m_0 so that $P(m_0)$ and so that $P(n)$ implies $m_0 \leq n$.

Proof: Suppose $P(-)$ is a property that makes sense for natural numbers and that there is no minimal m for which $P(m)$ is true. We will show that $P(m)$ is false for all m by strong induction.

- [Strong Inductive Hypothesis] Assume that $P(j)$ is false for all $j < k$.
- [Inductive Step] We must show that $P(k)$ is also false. Suppose $P(k)$ were true. By the strong inductive hypothesis, $P(j)$ is false for all $j < k$. So k would be the smallest natural number for which $P(k)$ is true. This contradicts the assumption that there is no minimal value for which $P(m)$ is true. So $P(k)$ must not be true.

□

Exercises for Lecture 5

1. Define the "tribonacci" numbers as follows:

$$\begin{aligned} t_0 &= 0 \\ t_1 &= 1 \\ t_2 &= 2 \\ t_{n+3} &= t_{n+2} + t_{n+1} + t_n && \text{for each } n \end{aligned}$$

   Prove by strong induction that $t_n < 2^{n-1}$ is true for all natural numbers n. [Recall that $2^{-1} = \frac{1}{2}$.]

2. Prove that $n^3 < 3^n$ for all natural numbers n.

5.3 Minimum and Maximum

The minimum of two numbers m and n is, of course, the smaller of the two. We write $\min(m, n)$ for this. The maximum is written as $\max(m, n)$. To make this precise, we can write a formal definition.

Definition 6

For natural numbers m and n, $\min(m,n)$ is a natural number satisfying:

- $\min(m,n) \leq m$ and $\min(m,n) \leq n$;
- for any natural number p, if $p \leq m$ and $p \leq n$, then $p \leq \min(m,n)$.

Also, $\max(m,n)$ is defined *dually*. For natural numbers m and n, $\max(m,n)$ is a natural number satisfying:

- $m \leq \max(m,n)$ and $n \leq \max(m,n)$;
- for any natural number p, if $m \leq p$ and $n \leq p$, then $\max(m,n) \leq p$.

Both min and max are characterized by what we may call an "adjoint situation".

$$\begin{aligned} p \leq \min(m,n) &\iff p \leq m \text{ and } p \leq n \\ \max(m,n) \leq p &\iff m \leq p \text{ and } n \leq p \end{aligned}$$

This shows that comparing p to two numbers on the right is the same as comparing p to their minimum on the right; and similarly for comparison on the left and maximum. We say that $\min(m,n)$ is the *greatest lower bound* of m and n and that $\max(m,n)$ is the *least upper bound* of m and n.

Some simple facts about min and max derive directly from this characterization by adjointness. The two operations, min and max, have many useful properties, all of which can be proved using the above characterizations plus some facts about addition.

In particular, together with addition, min and max make the natural numbers into something called a *distributive lattice ordered monoid*. Basically, this means that addition together with min and max cooperate in specific ways that are akin to the basic laws of arithmetic. The most useful laws having to do with min and max are:

Laws 7: min **and** max

For all natural numbers m, n and p:

Commutativity

$$\min(m,n) = \min(n,m) \qquad \max(m,n) = \max(n,m)$$

Associativity

$$\min(\min(m,n),p) = \min(m,\min(n,p)) \qquad \max(\max(m,n),p) = \max(m,\max(n,p))$$

Laws 7 (cont.)

Idempotency

$$\min(m, m) = m \qquad \max(m, m) = m$$

Absorption

$$\min(m, \max(n, m)) = m \qquad \max(m, \min(n, m)) = m$$

Distributivity

$$m + \min(n, p) = \min(m + n, m + p) \qquad m + \max(n, p) = \max(m + n, m + p)$$
$$\max(m, \min(n, p)) = \min(\max(m, n), \max(m, p)) \qquad \min(m, \max(n, p) = \max(\min(m, n), \min(m, p))$$

Modularity

$$m + n = \min(m, n) + \max(m, n)$$

We will not prove most of these as they follow easily from arithmetic laws. But the most subtle is worth looking at.

Lemma 8

$m + \min(n, p) = \min(m + n, m + p)$ for all $m, n, p \in \mathbb{N}$.

Proof: By definition, $\min(n, p) \leq n$. Because addition is monotonic, $m + \min(n, p) \leq m + n$. Likewise $m + \min(n, p) \leq m + p$. So $m + \min(n, p) \leq \min(m + n, m + p)$.

So to complete the proof, we must show that $\min(m+n, m+p) \leq m+\min(n, p)$. Suppose $k \leq \min(m + n, m + p)$. Then if $k \leq m$, then $k \leq m + \min(n, p)$ obviously. Otherwise, $m \leq k$ by Linearity. So $m + d = k$ for some d. Hence $m + d \leq m + n$ and $m + d \leq m + p$. Since $\leq$ is order reflecting, $d \leq \min(n, p)$. Consequently, $k = m + d \leq m + \min(n, p)$. We have thus shown that $k \leq \min(m + n, m + p)$ implies $k \leq m + \min(n, p)$. In particular, this applies to $\min(m + n, m + p)$. □

Exercises for Lecture 5

1. Calculate the following values. Show work.

Exercises for Lecture 5 (cont.)

1. $\min(5, \min(4, 6))$
2. $\min(5, \max(4, 6))$
3. $\min(340, \max(234, 340))$
4. $\min(5, \max(3, \min(\max(1, 2), 7)))$
5. $\min(5 + \max(4 + \min(3 + \max(7, 8), 3 + \min(7, 8)), 6), 7)$

2. Prove that min distributes over max.

Lecture 6

Divisibility

Goals

Lecture

- Develop the analogue of $\leq$ defined by multiplication instead of by addition.
- Illustrate the value of the analogy to prove useful properties of divisibility.
- Introduce general division for natural numbers.

Study

- Demonstrate competence in determining divisibility and division facts.

Study

For natural numbers, $m \leq n$ *means* that $m + d = n$ for some d. Since multiplication satisfies many of the same laws (it is commutative, associative, etc.), a similar definition is possible in terms of multiplication.

Definition 1

For natural numbers m and n, say that m **divides** n, if and only if $m \cdot q = n$ for some natural number q. We write $m \mid n$ when m divides n.

We have an analogy between "m is less than or equal to n" and "m divides n." The difference is precisely that the former is defined by addition and the latter by multiplication. This is useful because we can sometimes transfer a fact about $\leq$ to a fact about $\mid$ simply by noticing that they both depend on analogous laws of arithmetic.

For example, the relation $\leq$ is reflexive *because* 0 is the identity for addition. The relation $\mid$ is reflexive because 1 is the identity for multiplication. Likewise, $\leq$ is transitive *because* addition is associative; so $\mid$ is transitive because multiplication is associative.

Anti-symmtry is also true, but a proof hints at why our analogy is not perfect. Recall that we proved that $m \leq n$ and $n \leq m$ implies $m = n$. using cancellativity of addition. But multiplication is only cancellative for non-zeros. That is, $m \cdot p = n \cdot p$ implies $m = n$ only when $p \neq 0$. This means that we need to treat 0 as a special case. Suppose $m \cdot q = n$ and $n \cdot r = m$. If $m = 0$, then obviously $n = 0$. So $m = n$. If $m \neq 0$, then $m \cdot q \cdot r = m$ and $m \neq 0$. So by cancellativity $q \cdot r = 1$, and so $q = 1$. Hence $m = m \cdot 1 = n$.

The divisibility relation begins to be more interesting when we realize that it is *not* linear. For example, 4 does not divide 13 and 13 does not divide 4. Apparently, the structure of the natural numbers with respect to $|$ is much more complicated that with respect to $\leq$.

Note that $1 \mid m$ is true for any m, simply because $1 \cdot m = m$. And $m \mid 0$ is true for any m because $m \cdot 0 = 0$. So 1 is "at the bottom" of the divisibility relation and 0 is "at the top". This may seem strange. Some people find it so irritating that they simply rule 0 out of consideration, and declare that $0 \mid 0$ is undefined. This is fine, but I prefer to understand $0 \mid 0$ to mean $0 \cdot q = 0$ for *some* q.

Figure 6 shows a fragment of the natural numbers with respect to divisibility.

Exercises for Lecture 6

For each of the following pairs (m, n) of natural numbers, determine whether or not $m \mid n$.

1. $(4, 202)$
2. $(7, 49)$
3. $(11, 1232)$
4. $(9, 19384394)$
5. $(n, 6n)$
6. $(26, 65)$

Divisibility also makes sense for integers, but it is no longer anti-symmetric for a somewhat trivial reason. For example, $5 \mid -5$ and $-5 \mid 5$, but obviously $5 \neq -5$. In short, divisibility ignores the sign of an integer. This is a good reason for us to concentrate our attention on natural numbers, bearing in mind that much of what we can say about divisibility is true for integers as well.

Before closing this section, we note additional facts about how $|$ interacts with arithmetic.

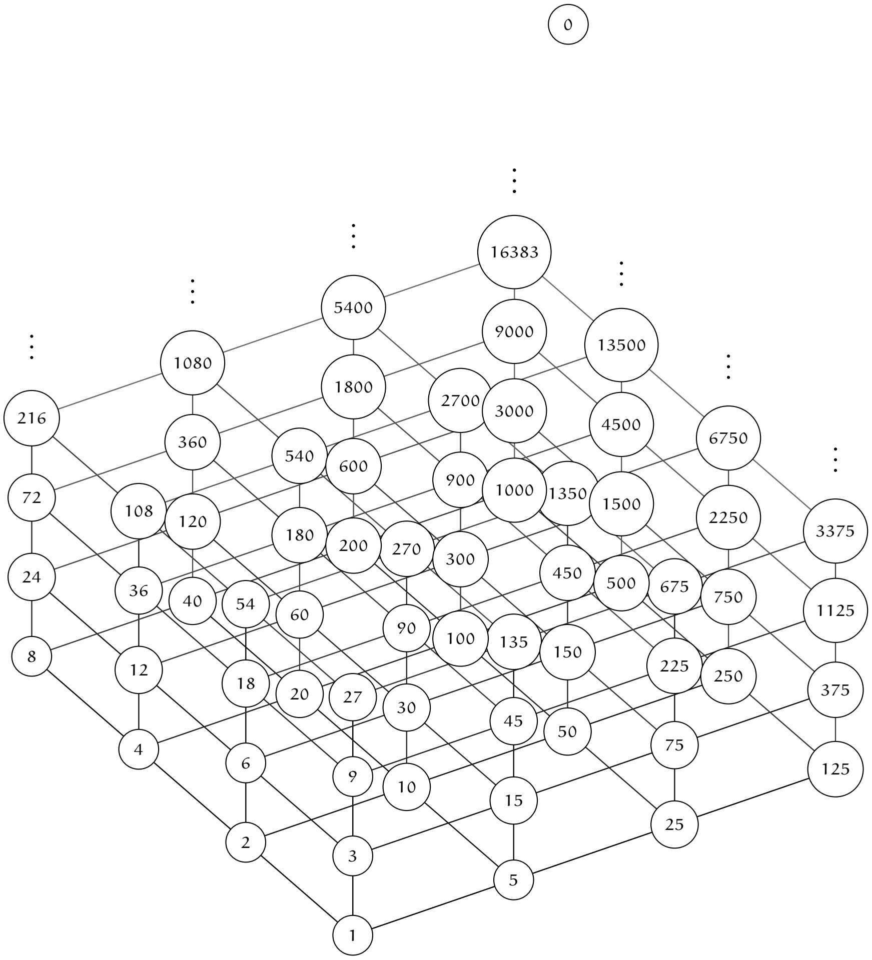


Figure 6.1: Part of the divisibility relation

Proposition 2

1. $m \mid n$ implies $m \mid np$
2. $m \mid n$ and $m \mid p$ implies $m \mid (n+p)$
3. $0 < n < m$ implies $m \nmid pm + n$

Proposition 2 (cont.)

4. $m \mid n$ and $n > 0$ implies $0 < m \leq n$.

Proof:

1. is true by associativity.
2. is due to distributivity.
3. can be proved by showing that $mq \neq pm + n$ for all q. We leave this as a voluntary exercise.
4. uses (3). That is, suppose $m \mid n$ and $n > 0$. Then we can write this as $m \mid 0 \cdot m + n$. Thus $0 < m$. And according to (3), $0 < n < m$ has to fail. Since $0 < n$ holds by assumption, $m \leq n$.

□

6.1 Quotients and Remainders

Without rational numbers, division still makes sense. We only need to account for the fact that sometimes numbers don't divide evenly. To make this work properly, we can speak of a *quotient* and *remainder*. For example, dividing 23 by 7 results in a quotient of 3 and a remainder of 2. That is, $3 \cdot 7 + 2 = 23$. Let us make this precise.

Theorem 3: Natural Number Division

For any natural number m and any positive natural number n, there is a unique pair of natural numbers q and r satisfying

$$m = qn + r$$

and

$$r < n$$

Proof: First, we prove that if q and r satisfy the stated conditions, and so do q' and r', then $q = q'$ and $r = r'$. This will prove that there can be at most one pair of natural numbers satisfying the stated conditions. Suppose $qn + r = q'n + r'$ and $r < n$ and $r' < n$. If $r < r$, then there is some positive e so that $r + e = r'$. Hence $qn = q'n + e$. But this means that $n \mid q'n + e$. Clearly, $e < n$, so is impossible. Thus it can not be the case that $r < r'$. For the same reason, it can not be the case that $r' < r$. So $r = r'$.

Theorem 3 (cont.)

Now we prove that there actually is a pair of numbers q and r satisfying the condtions. For this, we proceed by induction on m.

- [Basis] $0 = 0 \cdot n = 0$. So $q = 0$ and $r = 0$ do the job.
- [Inductive hypothesis] Assume that for some k, natural numbers p and s exist for which $k = p \cdot n + s$ and $s < n$.
- [Inductive Step] We must find q and r so that $m^{\curvearrowright} = q \cdot n + r$ and $r < n$. There are two cases: either $s^{\curvearrowright} = n$ or $s^{\curvearrowright} < n$. Suppose $s^{\curvearrowright} < n$. Then $m^{\curvearrowright} = p \cdot n + s^{\curvearrowright}$. So let $q = p$ and let $r = s^{\curvearrowright}$. Suppose $s^{\curvearrowright} = n$. Then $p \cdot n + s^{\curvearrowright} = p^{n}x \cdot n + 0$. So let $q = p^{\curvearrowright}$ and $r = 0$.

□

This theorem indicates that division of natural numbers works, as long as we account for both the quotient and the remainder. It is useful to have notation for both of these. So we will sometimes use the same notation for an integer divided by a positive natural number.

Definition 4

For any integer a and positive natural number n, let $a \mathbin{/\!/} n$ denote the *quotient* and $a \bmod n$ the *remainder* of dividing a by n. That is, $a \mathbin{/\!/} n$ and $a \bmod n$ are the unique two integers so that

$$a = (a \mathbin{/\!/} n) \cdot n + (a \bmod n)$$

and

$$0 \leq a \bmod n < n$$

The notation // is borrowed from the programming langauge Python. It is intended to avoid confusion with real number division x/y. Python, Java, C and many other languages use a percent sign for remainder, but unfortunately, its precise meaning differs from langauge to language. The notation mod avoids this ambiguity.

Exercises for Lecture 6

1. Calculate the following:

    1. 24 // 7
    2. 10000000 mod 10000001
    3. 13 mod 8

Exercises for Lecture 6 (cont.)

4. 8 mod 5
5. 5 mod 3
6. 3 mod 2
7. 2 mod 1

2. Show that for any natural number m, any positive natural number n and any positive natural number p, it is the case that $pm \mathbin{/\!/} pn = m \mathbin{/\!/} n$ and that $pm \bmod pn = p(m \bmod n)$.

Lecture 7

Prime Numbers

Goals

Lecture

- Prove the Fundamental Theorem of Arithmetic
- Prove that the prime numbers are unbounded in the natural numbers.

Study

- Demonstrate competence in prime factorization.

We all know what a prime number is. It is a number that is not 1 and has no non-trivial factors. One might ask why 1 does not count as a prime number since it seems like an arbitrary thing to exclude. We settle that question here. The key insight (from an earlier lecture) is that an *empty product* is 1.

Recall that the product of a list of natural numbers is defined inductively by

- $\prod[] = 1$
- $\prod n : L = n \cdot \prod L$

Definition 1

A **prime number** is a positive natural number p so that for any list L of natural numbers, if $p \mid \prod L$ holds, then $p \mid L_i$ holds for some $i < \mathsf{len}(L)$. In other words, if p divides a product of natural numbers, it divides one of them.

With this definition, 1 is not prime for exactly the same reason that 6 is not prime. That is, 6 is not prime because, for example, $6 \mid \prod[2, 3]$ but 6 does not divide any item on the list $[2, 3]$. Likewise, $1 \mid \prod[]$ but 1 does not divide any item on the list $[]$ because there are no items on the empty list. In contrast, 2 is prime because if $2 \mid \prod L$ with L being a list of natural numbers, at least one of the items on the list must be even. It is not hard to check that this definition of primality agrees with the more familiar one.

Our first task involving primes is to remind ourselves of the *Fundamental Theorem of Arithmetic*, that every positive natural number factors uniquely into primes. We split the proof of this into two separate parts.

Lemma 2

For any positive natural number m, there is a list P consisting only of primes so that $m = \prod P$.

Proof: Here we use strong induction.

- [Strong Inductive Hypothesis] Assume that for some k, it is the case that for every $0 < j < k$, there is a list of primes P so that $j = \prod P$.
- [Strong Inductive Step] There are three cases to consider. Either $k = 1$, or $k = i \cdot j$ for some positive i and j strictly less than k, or neither of these holds.

  Suppose $k = 1$. Then we let $P = []$. This is a (trivial) list of primes whose product is k.

  Suppose $k = i \cdot j$ where i and j are both positive and strictly less than k. By the inductive hypothesis, there are lists of primes Q and R so that $i = \prod Q$ and $j = \prod R$. Hence $k = \prod Q \cdot \prod R = \prod(Q \otimes R)$.

  Suppose k is neither equal to 1, nor equal to $i \cdot j$ for any two natural numbers strictly less than k. Then k itself is prime. So $k = \prod[k]$.

  These are the only possible cases.

□

The list of primes we constructed in the Lemma 2 is called a **prime factorization of** m. Generally, a composite has more than one prime factorization for trivial reasons. For example, [2, 3] and [3, 2] both are prime factorizations of 6. But the *only* way two prime factorizations of the same number can differ is by the order in which the factors are listed. If we insist that our lists are sorted in increasing order, this ambiguity is avoided. That is, say P is a *sorted* list of natural numbers if $P_i \leq P_j$ whenever $i \leq j < \mathsf{len}(P)$.

Lemma 3

For every positive natural number m, there is at most one sorted prime factorization of m.

Proof: Suppose P and Q are sorted prime factorizations of m. We need to show that they are equal. We do this by structural induction on P. That is, we show that for all sorted lists of

Lemma 3 (cont.)

primes Q, if $\prod P = \prod Q$, then $P = Q$.

- [Basis] If $\prod [] = \prod Q$, then Q must also be the empty list because no non-empty list of primes has a product of 1.
- [Inductive hypothesis] Assume that, for some fixed sorted list of primes K, it is the case that for all sorted lists of primes R, if $\prod K = \prod R$, then $K = R$.
- [Inductive Step] Suppose $\prod p : K = \prod Q$ where $p : K$ and Q are sorted lists of primes. Then $p \mid \prod Q$. But p is prime, so p must appear somewhere in the list Q. There are two cases: either p is the initial item of Q, or not.

  If p is the initial element of Q, then we can write $Q = p : R$ for some list R. By cancellativity, $\prod K = \prod R$. So by the inductive hypothesis, $K = R$. So $p : K = Q$.

  On the other hand, suppose p is not the initial item on the list Q. We show that this leads to a contradiction, so the previous case is the only possibility. Since p is prime, it is somewhere on the list Q. So Q is not empty. That is, Q can be written as $q : Q'$ where q is a prime strictly less than p. But q is also prime, so it must appear somewhere on the list P. But that violates the assumption that $p : P$ is sorted, for it means that the smaller value q appears later in the list that p.

□

The two preceeding lemmas show that every positive m has a unique sorted prime factorization, typically called *the* prime factorization.

Theorem 4: Fundamental Theorem of Arithmetic

Every positive natural number has a unique sorted prime factorization.

Proof: All that remains is to the remark that if P is a prime factorization of m, then P can be sorted into increasing order. The result has the same product P because of commutativity. □

For example, 24 is factored as $[2, 2, 2, 3]$ and 800 is factored as $[2, 2, 2, 2, 2, 5, 5]$. We can get a simpler representation by listing the number of times each prime is repeated. So we can represent 800 by $[5, 0, 2]$, signifying that $800 = 2^5 3^0 5^2$. Notice that in this notation, we need the middle 0 as a "place holder" to indicate that our number does not have any 3 factors. Also notice that "trailing zeros" in this notation do not make a difference. $[5, 0, 3, 0]$ also represents 800. The extra 0 at the end simply tells us that 800 is not divisible by the next prime (7). We will investigate this notation after establishing that we have a plentiful supply of primes.

Theorem 5

There are infinitely many primes.

Proof: We prove this by showing that no finite list of primes exhausts all the possible primes.

Suppose L is a non-empty list consisting of primes. To show that L is missing a prime, consider the number $m = 1 + \prod L$. Since $\prod L \geq 1$, $m > 1$. So m has a non-empty prime factorization, say M. Clearly M does not have any item in common with L (this could be proved explicitly by induction on L). So we have found a prime number (namely, any item of M) that is missing from L. Thus L can not be an exhaustive list of all primes. □

Definition 6

We can enumerate the primes: $2, 3, 5, 7, \ldots$ in increasing order. For every natural number k, let p_k be the k^{th} prime. That is, the numbers p_k satisfy

$$\begin{aligned} p_0 &= 2 \\ p_{k^\frown)} &= \text{the smallest prime } q \text{ so that } p_k < q \end{aligned}$$

Notice that this is well-defined because for any m there is a prime greater than m. We would not know this if we did not know there are infinitely many primes.

Exercises for Lecture 7

1. What is $p_1 0$?
2. What is the prime factorization of 1440?

Using p_k, we can succinctly represent any positive natural number by a list of exponents of primes. Namely, for a list R of natural numbers, define R_{pe} (for "prime representation") to be

$$R_{\mathsf{pe}} \coloneqq \prod_{i<\mathsf{len} R} p_i^{R_i}.$$

For example, $[1, 2, 0, 2]_{pr} = 2^1 3^2 5^0 7^2 = 882$.

For a positive natural number n, let $\mathsf{PR}(n)$ denote the unique list so that (i) $\mathsf{PE}(n)_{\neq} = n$ and (ii) $\mathsf{PE}(n)$ does not contain any trailing zeroes.

Recall that we defined $P \overline{+} Q$ for two lists of natural numbers by adding the items of the two lists itemwise.

$$\begin{aligned} P \overline{+} [] &= P \\ [] \overline{+} Q &= Q \\ m : P \overline{+} n : Q &= (m+n) : (P \overline{+} Q) \end{aligned}$$

Then it is clear (we will not give a proof) that $P_{\mathsf{pe}} \cdot Q_{\mathsf{pe}} = (P \overline{+} Q)_{\mathsf{pe}}$. Also define a relation $P \preceq Q$ on lists of natural numbers by $[] \preceq Q$ always, $m : P \preceq []$ never, and $m : P \preceq n : Q$ if $m \leq n$ and $P \preceq Q$. Then $P_{\mathsf{pe}} \mid Q_{\mathsf{pe}}$ if and only if $P \preceq Q$. In other words, if we list the exponents of primes that constitute given numbers m and n, we can compare them for divisibility simply by comparing the exponents by $\leq$.

7.1 Greatest Common Divisor and Least Common Multiple

Recall that min and max are defined in terms of $\leq$ (which is defined in terms of addition). They have analogues defined in terms of $|$ (which is defined in terms of multiplication).

Definition 7

For natural numbers m and n, a *common divisor of* m *and* n is a natural number c satisfying $c \mid m$ and $c \mid n$; a *greatest common divisor of* m *and* n is a natural number g so that

- g is a common divisor of m and n, and
- if p is a common divisor of m and n then $p \mid g$.

For natural numbers m and n, a *common multiple of* m *and* n is a natural number c satisfying $m \mid c$ and $n \mid c$; a *least common multiple of* m *and* n is a natural number ℓ so that

- ℓ is a common mulitple of m and n; and
- if p is a common multiple of m and n, then $\ell \mid p$.

For now, let us at least see that *if* a greatest common divisor or a least common multiple exists, then it is unique. The pattern of the proof is important because it sows up in many other places in mathematics. So it is worth noting here.

Lemma 8

For any natural numbers m and n, there is at most one greatest common divisor and at most one least common multiple.

Lemma 8 (cont.)

Proof: Suppose g and g' are both greatest common divisors of m and n. Then $g \mid m$ and $g \mid n$. Since g' is a greatest common divisor, $g \mid g'$ according to the second requirement in the definition. Similarly, g' is a common divisor of m and n, so $g' \mid g$. But divisibility is anti-symmetric, so $g = g'$. The proof for least common multiples is *dually similar* meaning the role of "divides" is replaced by the opposite, "is divided by". □

This justifies writing $\gcd(m, n)$ for *the* greast common divisor and $\mathrm{lcm}(m, n)$ for *the* least common multiple, because if some gretest common divisor or least common multiple exists, it is unique. So when does a greatest common divisor exist? The following proof (due essentially to the Greek mathematician Euclid) for calculating $gcd(m, n)$ for any m and n answers the question.

Theorem 9: Euclid's Algorithm (modern version)

For any two natural numbers m and n, $\gcd(m, n)$ exists.

Proof: Without loss of generality, we may assume that $m \leq n$ because the requirements for $\gcd(m, n)$ to exist are the same as for $\gcd(n, m)$.

To simplify the notation, we write $P(i)$ to mean that for all $j \geq i$, the greatest common divisor $\gcd(i, j)$ exists. We proceed by strong induction on m to show that for all m, $P(m)$ holds. The result follows immediately from that.

- [Strong Inductive Hypothesis] Assume that for some natural number a it the case that for all natural numbers $k < a$, $P(k)$ holds.
- [Strong Inductive step] We must show that $P(a)$ holds. That is, for every $b \geq a$, the a greatest common divisor of a and b exists. Consider any $b \geq a$. We consider two cases depending on whether $a = 0$ or not. In case $a = 0$, then b could be any natural number. Evidently b divides both 0 and b. And since any natural number is a divisor of 0, b is the greatest common divisor of 0 and b.

  In case $a > 0$, then natural number division (Theorem 3) provides a remainder $b \bmod a$. Since $b \bmod a < a$, the Strong Inductive Hypothesis ensures $P(b \bmod a)$ is true. In particular, $\gcd(b \bmod a, a)$ exists. Let $g = \gcd(b \bmod a, b)$. We claim that g is also a greatest common divisor of a and b.

  Clearly, $g \mid a$. And since $b = a \cdot (b /\!/ a) + (b \bmod a)$, and g divides both summands separately, $g \mid b$. So g is a common divisor of a and b. Now, suppose c divides both a and b, then c also divides $b \bmod a$. Hence $c \mid g$, because $g = \gcd(b \bmod a, a)$.

□

The proof of this lemma gives us an algorithm (a method) for calculating $\mathrm{gcd}(a, b)$. Namely,

$$\begin{aligned} \mathrm{gcd}(0, b) &:= b \\ \mathrm{gcd}(a, b) &:= \mathrm{gcd}(b \bmod a, a) && \text{for } a > 0 \end{aligned}$$

This translates to Python.

Algorithm 10

```
def gcd(a,b):
    if a == 0:
        return b
    else:
        return gcd(b%a, a)
```

Note that in python `%` is the symbol used for the remainder operation.

Variants of this algorithm have many important applications, including in cryptography. We look at some of those applications in Lecture **??**.

Another way to understand gcd and lcm for positive natural numbers is via prime factorizations. This is not a practical way to *calculate* them because it requires a lot of work simply to find all of number's prime factors – a fact that may not be obvious from looking at small examples. Nevertheless, it is useful in another way. We can very easily prove facts about gcd and lcm when we shift our thinking to regard a positive natural number n as being represented by its prime factorization.

Recall that we can express any positive natural number n as a product of primes: $n = 2^{k_0} 2^{k_1} 3^{k_2} \dots p_{m-1}^{k_{m-1}}$ where the primes here are used in their standard order. The list of exponents $\mathsf{PE}(n) = [k_0, k_1, k_2, \dots, k_{m-1}]$ is all the information we need to reconstruct n. Recall that for any list of natural numbers, we let R_{pe} denote the product $\prod_{i<\mathsf{len}(R)} p_i^{R_i}$, where p_i is the i^{th} prime number.

Consider writing $\mathrm{gcd}(42, 180) = 6$ in prime factorized form. This is

$$\mathrm{gcd}(2^1 3^1 5^0 7^1, 2^2 3^2 5^1 7^0) = 2^1 3^1 5^0 7^0.$$

Each exponent in the result is the minimum of the corresponding exponents in the arguments. This makes sense because $m \mid n$ is true if and only if each entry i in $\mathsf{PE}(n)$ is less than or equal to entry i in $\mathsf{PE}(m)$. By taking the minimum in each position, the result is the greatest common divisor.

A similar observation indicates that $\mathrm{lcm}(a, b)$ can be obtained from prime factorization by taking the maximumm in each position instead of the minimum. In our example, $\mathrm{lcm}(42, 180) = 2^2 3^2 5^1 7^1 = 1260$.

This leads to simple algorithms computing gcd and lcm if we are given two positive natural numbers in prime exponent form. That is, define the following operations on lists of natural numbers.

$$\begin{aligned}
\overline{\min}(P, [\,]) &= [\,] \\
\overline{\min}([\,], n : Q) &= [\,] \\
\overline{\min}(m : P, n : Q) &= \min(m, n) : \overline{\gcd}(P, Q) \\
\overline{\max}(P, [\,]) &= P \\
\overline{\max}([\,], n : Q) &= n : Q \\
\overline{\max}(m : P, n : Q) &= \max(m, n) : \overline{\max}(P, Q)
\end{aligned}$$

Now it is straightforward to show that the following equations are true for any two lists of natural numbers. The proofs (left as an exercise) are by induction on the lengths of the lists.

Lemma 11

$$\begin{aligned}
P \overline{+} Q &= \overline{\min}(P, Q) \overline{+} \overline{\max}(P, Q) \\
\gcd(P_{\mathsf{pe}}, Q_{\mathsf{pe}}) &= \overline{\min}(P, Q) \\
\mathrm{lcm}(P_{\mathsf{pe}}, Q_{\mathsf{pe}}) &= \overline{\max}(P, Q)
\end{aligned}$$

Proof: Exercise. □

Exercises for Lecture 7

1. Prove that $P \overline{+} Q = \overline{\min}(P, Q) \overline{+} \overline{\max}(P, Q)$ holds for any lists of natural numbers.

2. Prove that $\gcd(P_{\mathsf{pe}}, Q_{\mathsf{pe}}) = \overline{\min}(P, Q)$ holds for any lists of natural numbers.

3. Prove that $\mathrm{lcm}(P_{\mathsf{pe}}, Q_{\mathsf{pe}}) = \overline{\max}(P, Q)$ holds for any lists of natural numbers.

Recall Laws 5.7. There, we list important ways in which min, max and addition interact. Because of Lemma 11, we can justify an exactly analogous list of laws for how gcd, lcm and multiplication interact. For example,

$$\gcd(\gcd(m, n), p) = \gcd(m, \gcd(n, p))$$

because

$$\min(\min(m, n), p) = \min(m, \min(n, p)).$$

Exercises for Lecture 7

1. Rewrite the entire table of Laws 5.7 for the analgous laws pertaining to gcd, lcm and multiplication.

LECTURE **8**

Other Number Systems

Discrete Mathematics is not only concerned with natural numbers. Integers, rational numbers, real numbers and even, on a few occasions, complex numbers, all play their roles. In this lecture, we discuss, with much less formality, integers and rational numbers, leaving real and complex numbers to a courses on analysis.

Goals

Lecture

- Introduce the standard notation for these number systems.
- Discuss briefly the extensions of arithmetic to integers, rationals.
- Wave hands and mumble unconvincingly about real and complex numbers.

Study

- Learn the meanings of the symbols for standard number systems.
- Prove some simple facts about arithmetic on integers and rational numbers.

8.1 Integers

Our intention here is to give a general overview of the integers in light of our detailed study of natural numbers. Though we could, we will not go into the same level of detail.

Integers, we understand, include the natural numbers but also include negatives. They constitute the simplest possible extension of the natural numbers in such a way so that for any integers a and b, there is a solution of the equation $a = b + x$. So the fundamental structure of integers is that (a) the natural numbers are integers and (b) addition on integers is invertible, i.e., subtraction works correctly.

Postulate 1

The integers have the following basic structure.

- Every natural number is an integer. For this discussion, we will use letters $a, b, c, \ldots$ for integers, and $m, n, \ldots$ for natural numbers. To emphasize when a natural number m plays the role of an integer we write ${}^{+}m$.
- For every two integers, a and b, there is a *sum* and a *difference*, denoted by $a + b$ and $a - b$.

We must insist that addition on integers behaves as we should expect.

Postulate 2

For natural numbers m and n,

$$ {}^{+}m + {}^{+}n = {}^{+}(m + n). $$

That is, adding m and n *as integers* is the same as adding them *as natural numbers*.

Furthermore, addition on integers is associative and commutative, and $a + {}^{+}0 = a$ for all integers a.

Next, sum and difference are related.

Postulate 3

For any integers a, b and c,

$$ a = b + c \quad \text{if and only if} \quad a - b = c $$

Finally, we require minimality.

Postulate 4

No integers can be removed without violating Postulate 1.

We do not go through the details, but these axioms determine the structure of the integers similar to the way our earlier axioms determine the structure of the natural numbers. Moreover, all the familiar laws of addition and subtraction follow. We illustrate with two simple laws. As usual, we can abbreviate ${}^{+}0 - a$ by $-a$.

Lemma 5

For any integer a, ${}^{+}0 = a + -a$.

Proof: Since ${}^{+}0 - a = {}^{+}0 - a$ (trivially), Postulate 3 yields ${}^{+}0 = a + ({}^{+}0 - a) = a + (-a)$. □

Lemma 6

For integers a and b, $a - b = a + (-b)$.

Proof: $a = a + {}^{+}0 = a + (-b + b) = b + (a + (-b))$ by the basic laws of addition (identity, commutativity and associativity) and the previous lemma. So by Postulate 3, $a - b = a + (-b)$. □

In the axiomatization of integers, we did not mention multiplication. This is because multiplication is definable from addition. But we need an aditional fact.

Say that an integer a is "simple" if either $a = {}^{+}m$ or $-a = {}^{+}m$ for some natural number m.

Lemma 7

Every integer is simple.

Proof: Exercise □

Definition 8

For any two integers a and b, the *product* $a \cdot b$ defined as follows:

$$a \cdot {}^{+}0 = {}^{+}0$$
$$a \cdot {}^{+}(k^\frown) = a + a \cdot {}^{+}k$$
$$a \cdot -({}^{+}m) = -(a \cdot {}^{+}m)$$

These requirements are enough to ensure that multiplication is the operation we expect it to be. We will not verify that fact, but it involves an inductive argument similar to the arguments that we saw in previous lectures.

Exercises for Lecture 8

1. Show that every integer is simple, using only the postulates given here.

   Hint: First observe that, trivially, every integer of the form ${}^{+}m$ is simple. This acts as the basis for an inductive proof. Now suppose a and b are simple (the inductive hypothesis). Show that both $a + b$ and $a - b$ are simple (the inductive step). Since the integers are minimal for including the natural numbers and being closed under $+$ and $-$, this shows that all integers are simple.

2. Using only the axioms for integers, the foregoing two lemmas and the definition of multiplication, prove the following (remember that $-a$ abbreviates ${}^{+}0 - a$). You should prove these in the order presented here, as later exercises may depend on earlier ones. For all integers a, b and c show that the following are true.

   1. Addition is cancellative: $a + c = b + c$ implies $a = b$.
   2. $a - b = -(b - a)$
   3. $(a + b) - c = a + (b - c)$
   4. $a - (b + c) = (a - b) - c$
   5. $-(-a) = a$
   6. ${}^{+}0 \cdot a = {}^{+}0$
   7. $-a \cdot -b = a \cdot b$
   8. Multiplication distributes over addition.

8.2 Rational Numbers

Rational numbers, roughly, are minimal with respect to the integers and inverting multiplication, as integers are minimal with respect to natural numbers and inverting addition. Care must be taken, however, because ${}^{+}0 \cdot a = {}^{+}0 \cdot b$ is always true, this does not tell us anything about a and b. So "division by 0" can not make sense.

Postulate 9

The rational numbers have the following basic structure.

- Every integer is a rational number. For this discussion, to emphasize when an integer a

Postulate 9 (cont.)

plays the role of a rational number, we write $\frac{a}{1}$.

- For every two rational numbers p and q, there is a *product*, denoted by $p \cdot q$. Moreover, if $q \neq \frac{0}{1}$, then there is a *quotient*, denoted by p/q.

In analogy with integers, where addition and subtraction are the main operations, we insist that multiplication behaves correctly.

Postulate 10

For integers a and b,

$$\frac{a}{1} \cdot \frac{b}{1} = \frac{ab}{1}$$

That is, multiplying integers a and b *as rational numbers* is the same as multiplying them *as integers*.

Furthermore, multiplication is associative and commutative, and $p \cdot \frac{1}{1} = p$ for all rational numbers p.

Again, continuing the analogy with integers, multiplication and division must cooperate.

Postulate 11

For any rationals p, q and r, if $q \neq \frac{0}{1}$, then

$$p = q \cdot r \quad \text{if and only if} \quad p/q = r$$

Finally, we require minimality.

Postulate 12

No rational numbers can be removed without violating Postulate 9.

Like the axioms for integers, these axioms completely determine the rational numbers along with multiplication and division. To extend addition and subtraction to the rational numbers we can require the following.

For integers, we introduced the idea of *simplicity* to capture the fact that every integer is either a natural number or a negated natural number. The analogous idea for rational numbers is that every rational number is expressible as a fraction.

Definition 13

Say that a rational number p is **fractional** if it is the case that $p = a/b$ for some integer a and non-zero integer b.

Lemma 14

Every rational number is fractional.

Proof: Like the proof that every integer is simple, this involves an inductive argument. Namely, every integer is, according to the basic structure of rationals, fractional. Namely, $a/+1 = \frac{a}{1}$. Indeed the difference between the two notations is almost trivial. For the sake of completeness though, note that an integer a is written as $\frac{a}{1}$ only to emphasize that it is also a rational number. So $a/{}^{+}1 = a$ because $a = {}^{+}1 \cdot a$.

Suppose p and q are fractional. So $p = a/b$ and $q = c/d$ for some integers a, b, c and d with $b \neq 0$ and $d \neq 0$. Note that $a/b = p$ means that $a = b \cdot p$. Likewise $c = d \cdot q$. We claim that $p \cdot q = (a \cdot c)/(b \cdot d)$. This is true if $p \cdot q \cdot b \cdot d = a \cdot c$. But this follows from the foregoing. Since b and d are not zer, $b \cdot d$ is not zero. Hence $p \cdot q$ is fractional. We also claim that p/q is fractional. Here we need two cases. Suppose $c = 0$. Then $p = 0/1$ as needed. Suppose $c \neq 0$. Then $p/q = (a \cdot d)/(b \cdot c)$ because $a \cdot d \cdot q = pb \cdot c \cdot p$. We leave it to the reader to figure out why this last equation is true. □

Notice that this lemma does not tell us anything about simplified fractions. It merely says that every rational number can be written as a fraction. But over course, $2/4 = 4/8 = 6/12$ and so on.

Definition 15

For any two rational numbers $p = a/b$ and $q = c/d$, their *sum*, denoted by $p + q$ is the rational number $(ad + cb)/bd$.

This definition is not entirely legitimate yet. Because it could be the case that the result $(ad + cb)/bd$ depends on how we write p and q as fractions. In fact, though, this is not the case. For suppose $a/b = a'/b'$ and $c/d = c'/d'$. Then $(ad + cb)/bd = (a'd' + c'b')/b'd'$. So the way we add *fractions* is correct for adding rational numbers. These requirements are enough to ensure that addition is the operation we expect it to be.

Now suppose a is an integer and m is a positive natural number. Then we can write $\frac{a}{m}$ as an abbreviation of $\frac{a}{1}/\frac{+m}{1}$. This is the usual *fraction* notation for rational numbers. For a non-zero rational number p, we also usually write p^{-1} to abbreviate $\frac{+1}{1}/p$.

Exercises for Lecture 8

1. Using the proof of Lemma 5 as a prototype, show that if $p \neq 0$, then $p \cdot p^{-1} = 1$.
2. Using the proof of Lemma 6 as a prototype, show that if $q \neq \frac{+0}{1}$ then $p/q = p \cdot q^{-1}$.
3. Just using our axioms and the fraction notation (and the previous two exercises), show that $\frac{4}{6} = \frac{2}{3}$.

8.3 Real and Complex Numbers

We will not deal with rigorous characterizations of the real numbers and complex numbers in this course. A first course in real analysis is the appropriate place for such considerations. We simply take for granted that $\mathbb{R}$ constitutes a suitable extension of $\mathbb{Q}$ to account for *limits* (the mathematics you know from calculus). Likewise, $\mathbb{C}$ constitutes a suitable extension of $\mathbb{R}$ to account for algebraic closure (all polynomials having roots). The main things to know are that every rational number is real, every real number is complex, and arithmetic operations behave as they should for real and complex numbers.

Exercises for Lecture 8

1. Relax. Take a walk.

Lecture 9

Modular Arithmetic

Goals

Lecture

- Introduce counting “on a dial”
- Extend to arithemetic.
- Introduce the idea of an *equivalence relation* via modular equivalence.
- Introduce the extended Euclidean algorithm.

Study

- Learn to calculate within a modulus.
- Learn to use the extended Euclidean algorithm to determine multipicative inverses.

The positions on a dial such as a clock have some of the features of a number system. There is a “top of the dial”. Each position has a next position (clock-wise) and a previous position (counter-clockwise). As it happens, arithmetic on a dial, which mathematicians call *modular arithmetic*, is fundamental to how we actually write numbers, for example as base 10 numerals, and how we perform arithmetic on those numerals. Cryptography is also an important area of application.

Recall that the postulates of the natural numbers included the requirement that 0 does not have a predecessor (Postulate 1.2). Exercise 1.3. asked you to draw depictions in which Postulate 1.2 is relaxed. Assuming your response to Exercise 1.3. is correct, you may have drawn something similar to the dial-like structures in Figure **??**.

In the integers, 0 (as well as all other integers) has a precedessor. Suppose we *replace* Postulate 1.2 with the requirement that every value has a predecessor. Then the integers satisfy all postulates, including the requirement that nothing can be removed. But it is also true that the cyclical pictures of Figure **??** do the job.

For now, in order to avoid confusing pictures like this with the natural numbers, let us use alternative notation for “the start” and “successor”, which we have been denoting as 0 and $n^{\curvearrowright}$ in the natural numbers. For structures like those in **??**, let us write $\owedge$ for the “top of the dial” and $\mathsf{nxt}(a)$ for the “next” value after a in the clock-wise direction. So we are now considering structures of the following sort.

Definition 1

A **dial** is a structure satisfying the following conditions.

1. There is a value $\owedge$ and every value a has a successor $\mathsf{nxt}(a)$.
2. Every value has a predecessor: for every b, there is an a so that $\mathsf{nxt}(a) = b$.
3. Predecessors are unique: $\mathsf{nxt}(a) = \mathsf{nxt}(b)$ implies $a = b$.
4. No values can be removed without violating 1 and 2.

For our purposes, we may refer to the values of a dial as **positions**. So think of $\owedge$ as the name for the *top the dial*; think of $\mathsf{nxt}(a)$ as the next position *clock-wise from* a.

Example 2

The integers form a dial by letting $\owedge = {}^{+}0$ and $\mathsf{nxt}(a) = a + 1$. Every integer has a successor. Every integer has a unique predecessor. Moreover, if we were to remove any integers, then eith er we would remove 0, or some other remaining integer would either not have a successor or not have a predecessor.

Any actual cyclic structure like those in Figure **??** is a dial.

Consider a disc with labels "A" through "Z" printed evenly spaced on its edge. Then the label clockwise after "Z" is "A". So letting $\owedge = A$, this is a physical model of finite dial.

Almost the same definitions for addition and multiplication work in dial as they did in the natural numbers:

Definition 3: Addition on a Dial

Suppose $\mathbb{D}$ is a dial (it satisfies the four postulates above). For any a in $\mathbb{D}$ and any natural number n, the **sum** $a +_{\mathbb{D}} n$ is given by the equations:

$$\begin{aligned} a +_{\mathbb{D}} 0 &= a \\ a +_{\mathbb{D}} k^{\curvearrowright} &= \mathsf{nxt}(a +_{\mathbb{D}} k) \end{aligned}$$

This definition extends to all integers.

Lemma 4

Let $\mathbb{D}$ be a dial. For any natural number n, there is a position a on the dial so that $a +_{\mathbb{D}} n = \owedge$.

Proof: First, we proof that $\mathsf{nxt}(a +_{\mathbb{D}} n) = \mathsf{nxt}(a) +_{\mathbb{D}} n$ for all a by induction on n. The basis is obvious from the definition of $+_{\mathbb{D}}$. Assume that $\mathsf{nxt}(b +_{\mathbb{D}} k) = \mathsf{nxt}(b) +_{\mathbb{D}} k$ for all b. Then

$$\begin{aligned}\mathsf{nxt}(a +_{\mathbb{D}} k^{\frown}) &= \mathsf{nxt}(\mathsf{nxt}(a +_{\mathbb{D}} k))\\ &= \mathsf{nxt}(\mathsf{nxt}(a) +_{\mathbb{D}} k)\\ &= \mathsf{nxt}(a) +_{\mathbb{D}} k^{\frown}.\end{aligned}$$

Now we proof the lemma by induction on n,

- [Basis] $\owedge +_{\mathbb{D}} 0 = \owedge$. If $a +_{\mathbb{D}} 0 = \owedge$, then obviously $a = \owedge$.
- [Inductive Hypothesis] Assume that there is a unique b so that $b +_{\mathbb{D}} k = \owedge$.
- [Inductive Step] Let a be the unique predecessor of the position b given by the inductive hypothesis. Then $\mathsf{nxt}(a) = b$. Hence $a +_{\mathbb{D}} k^{\frown} == \mathsf{nxt}(a) +_{\mathbb{D}} k = \owedge$.

□

Lemma 5

Let $\mathbb{D}$ be a dial. For any a in $\mathbb{D}$ and any natural numbers n and p, $a +_{\mathbb{D}} (n + p) = (a +_{\mathbb{D}} n) +_{\mathbb{D}} p$.

Proof: Exercise □

So now we can define $a -_{\mathbb{D}} n$ as the unique element so that $(a -_{\mathbb{D}} n) +_{\mathbb{D}} n = \owedge$. So every integer i determines an element $\owedge +_{\mathbb{D}} i$.

The elements on a dial that can be written as $\owedge +_{\mathbb{D}} i$ for some integer i evidently satisfy the conditions 1. and 2 in the definition of dials, so according to condition 4, they constitute all elements of the dial. In other words, when we wish to consider a general position a on a dial, we may say "pick an integer i so that $a = \owedge +_{\mathbb{D}} i$. This i may not be unique, but it exists.

Exercises for Lecture 9

1. Prove Lemma 5.

Now that we know we can describe any position as $\owedge +_{\mathbb{D}} i$, we are tempted to define addition of positions: $a +_{\mathbb{D}} b$ by saying "pick some i so that $b = \owedge +_m i$. Then define $a +_{\mathbb{D}} b$ to be equal to $a +_m i$." That, however, is *not* a definition unless we can show that the result does not depend on which i we pick. In other words, we need to show that if $b = \owedge +_m i = \owedge +_m j$, then $a +_{\mathbb{D}} i = a +_{\mathbb{D}} j$.

Lemma 6

On any dial $\mathbb{D}$, for any positions a and b and any integers i and j, if $b +_{\mathbb{D}} i = b +_{\mathbb{D}} j$, then $a +_{\mathbb{D}} i = a +_{\mathbb{D}} j$.

Proof: Suppose $b +_{\mathbb{D}} i = b +_{\mathbb{D}} j$. Pick an integers k and h so that $a = \owedge +_{\mathbb{D}} k$ and $b = \owedge +_{\mathbb{D}} h$. Thus $a +_{\mathbb{D}} i = \owedge +_{\mathbb{D}} (k - h + h + i) = (b +_{\mathbb{D}} i) +_{\mathbb{D}} (k - h)$. Likewise $a +_{\mathbb{D}} j = (b +_{\mathbb{D}} j) +_{\mathbb{D}} (k - h)$. So the two are equal. □

This justifies writing $a +_{\mathbb{D}} b$ for adding positions on the dial. Namely, we can pick any i so that $b = \owedge +_{\mathbb{D}} i$ and define $a +_{\mathbb{D}} b$ to be equal to $a +_{\mathbb{D}} i$.

Exercises for Lecture 9

Fix a dial $\mathbb{D}$. [All of these are easy exercises. Do not over think them.]

1. Show that addition of positions is associative.
2. Show that addition of positions is commutative.
3. Show that $\owedge$ is an additive identity.
4. Show that every position has a unique additive inverse. That is, for any a there is exactly one position ^{-}a so that $a +_{\mathbb{D}} {^{-}a} = \owedge$. [Hint: if $a = \owedge +_{\mathbb{D}} i$, then let $^{-}a = \owedge +_{\mathbb{D}} (-i)$. Now show that if $a +_{\mathbb{D}} b = \owedge$, then $b = {^{-}a}$.]

We may now also define multiplication in a dial.

Definition 7

For position $\mathfrak{a}$ of a dial $\mathbb{D}$ and integer $\mathfrak{i}$, define $\mathfrak{a} \cdot_{\mathbb{D}} \mathfrak{i}$ by the equations

$$\begin{aligned} \mathfrak{a} \cdot_{\mathbb{D}} 0 &= \owedge \\ \mathfrak{a} \cdot_{\mathbb{D}} \mathfrak{j}^{\curvearrowright} &= \mathfrak{a} +_{\mathbb{D}} (\mathfrak{a} \cdot_{\mathbb{D}} \mathfrak{j}) \end{aligned}$$

This defines addition of natural numbers in the usual way, but it also extends uniquely

Definition 7 (cont.)

to negative integers. For example, $a \cdot_{\mathbb{D}} {}^{-}1$ must satisfy the equation $a \cdot_{\mathbb{D}} {}^{-}1^{\curvearrowright} = a +_{\mathbb{D}} a \cdot {}^{-}1$. Therefore, $\owedge = a +_{\mathbb{D}} a \cdot {}^{-}1$. So $a \cdot_{\mathbb{D}} {}^{-}1 = {}^{-}a$.

Lemma 8

Fix a dial $\mathbb{D}$. For any position a and integers i and j, it is the case that

$$\begin{aligned}(a \cdot_{\mathbb{D}} i) +_{\mathbb{D}} (a \cdot_{\mathbb{D}} j) &= a \cdot_{\mathbb{D}} (i + j)\\ (a \cdot_{\mathbb{D}} i) \cdot_{\mathbb{D}} j &= a \cdot_{\mathbb{D}} (i \cdot j)\\ (a \cdot_{\mathbb{D}} i) +_{\mathbb{D}} (b \cdot_{\mathbb{D}} i) &= (a +_{\mathbb{D}} b) \cdot_{\mathbb{D}} i\\ a \cdot_{\mathbb{D}} {}^{-}i &= {}^{-}a \cdot_{\mathbb{D}} i\end{aligned}$$

Proof: Exercise. □

These results tell us that multiplication of a dial position by an integer"behaves" as we would expect it to. Thus we can also make sense of multiplying two positions: $a \cdot_{\mathbb{D}} b$ is defined by picking some i so that $b = \owedge +_{\mathbb{D}} i$ and letting $a \cdot_{\mathbb{D}} b = a \cdot_{\mathbb{D}} i$. As with addition, this is meaningful only if the result does not depend on the choice of i. That is, we need to show that if $\owedge +_{\mathbb{D}} i = \owedge +_{\mathbb{D}} j$, then $a \cdot_{\mathbb{D}} i = a \cdot_{\mathbb{D}} j$. We leave that as an unassigned exercise.

Exercises for Lecture 9

1. Prove Lemma 8.
2. Refer to the Arithmetic Laws in Lecture 3. For each law, determine whether or not the same law holds for arithmetic on all dials.

Moduli

The cyclical dials (see Figure **??**) differ from the dial consisting of all integers, for in a cyclical dail $\owedge = \owedge +_{\mathbb{D}} k$ for some positive integer k. We refer to the smallest positive integer m satisfying $\owedge = \owedge +_{\mathbb{D}} m$ as the **modulus** of $\mathbb{D}$. In that case, we write $\mathbb{Z}_m$ instead of $\mathbb{D}$. If no modulus exists for a dial, then it is a model of the integers. In that case, we write $\mathbb{Z}$. Thus for example, $\mathbb{Z}_{12}$ models a standard twelve hour clock. $\mathbb{Z}_{10}$ models a dial with labels 0 through 9, $\mathbb{Z}_{26}$ models a dial with the letters of the alphabet. An important example is $\mathbb{Z}_2$. This is a dial with exactly two labels, say one at the top and one at the bottom. We might as well think of this as a switch with on and off positions. $\mathbb{Z}_1$ is a trivial dial with only one position.

For a dial $\mathbb{Z}_m$, we will write $a +_m b$ instead of $a +_{\mathbb{Z}_m} b$. Rememer though that $i + j$ means "honest" integer arithmetic. Now $i +_m j$ means arithmetic in the dial $\mathbb{Z}_m$: $(\textcircled{\wedge} +_m i) +_m (\textcircled{\wedge} +_m j)$.

Definition 9: Modular Equivalence

Fix a modulus $m > 0$. Two integers i and j are *equivalent modulo* m if it is the case that $\textcircled{\wedge} +_m i = \textcircled{\wedge} +_m j$. In that case, we write $i \equiv j \mod m$.

Recall that for any positive m and any integer i, there is a unique pair of integers (q, r) so that $i = qm + r$ and $0 \leq r < m$. We used the notation $i /\!/ m$ and $i \operatorname{mod} m$ to denote these. It is easy to see that $i \equiv j \mod m$ if and only if $(i \operatorname{mod} m) = (j \operatorname{mod} m)$ if and only if $m \mid (i - j)$.

The lemmas of the previous section show that integer arithmetic is compatible with modular equivalence in the sense that if $a_0 \equiv a_1 \mod m$ then $a_0 + b \equiv a_1 + b \mod m$ and $a_0 b \equiv a_1 b \mod m$. This means we can safely perform arithmetic in the familiar way and maintain equivalence modulus m.

Exercises for Lecture 9

For each of the following, find an integer x making the statement true, or show that no such x exists.

1. $5 + x \equiv 3 \mod 7$
2. $x \cdot 5 \equiv 1 \mod 11$
3. $3 + 4x \equiv 2 \mod 9$
4. $x \cdot 4 \equiv 1 \mod 12$
5. $4 \equiv 7 \mod x$
6. $x \equiv 2 \mod 7$ and $x \equiv 3 \mod 5$

Two integers a and b act as reciprocals if $a \cdot b \equiv 1 \mod m$. That is, in rational numbers $q \cdot \frac{1}{q} = 1$ for any non-zero rational q. So $a \cdot b \equiv 1 \mod m$ is analogous. But, for example, $x \cdot 5 \equiv 1 \mod 15$ can not be solved for x. We need a criterion for determining when an integer a has a reciprocal modulo m. Even better, a method for determining reciprocals when they exist, would allow us to make modular arithmetic effective.

Lemma 10: Extended Euclid's Algorithm

For any two integers a and b, there exist integers s and t so that $\gcd(a, b) = as + bt$.

Proof: The proof is by strong induction on a. Suppose $a = 0$. Then let $s = 0$ and $t = 1$. Otherwise, we know that $\gcd(a, b) = \gcd(b \bmod a, a)$. By the inductive hypothesis, there are integers s' and t' so that $\gcd(b \bmod a, a) = (b \bmod a)s' + at'$. But also $b = a \cdot (b \mathbin{//} a) + (b \bmod a)$. So $\gcd(a, b) = (b - a \cdot (b \mathbin{//} a)) \cdot s' + at' = a(t' - (b \mathbin{//} a)s') + bs'$. Thus we let $s = t' - (b \mathbin{//} a)s'$ and $t = s'$. □

Theorem 11

For integer $a \neq 0$ and positive integer m, there is an integer x so that $x \cdot a \equiv 1 \mod m$ if and only if $\gcd(a, m) = 1$.

Proof: content □

Part II

Sets and Functions

The mathematical universe consists of various things: numbers, functions, graphs, lists and so on. A *set* is, in the formulation of Cantor, a "many which can be comprehended as a one". For example, many playing cards taken together can be understood to be one deck of cards. For a mathematical example, the infinitely many natural numbers can be regarded as single thing, the *set of natural numbers*. Thus a set is essentially a collection of elements. A *function* is a correlation of the members of one set with members of another set. These two abstract concepts (sets and functions) form a framework in which virtually all of mathematics can be built. So an understanding of sets and functions is key to a rigorous approach to most other parts of mathematics. This conceptual framework can itself be put on a formal, precise footing called the Category of Sets and Functions. In these lectures, we investigate the Category of Sets and Functions, so that we can use these things as the basic building blocks of everything else we do.

LECTURE 10

Sets

Goals

Lecture

- Describe the basic structure of sets.
- Define list set notation.
- Introduce the idea of a subset.
- Introduce the axiom of extensionality for sets and some of its consequences.

Study

- Demonstrate ability to determine equality of sets.
- Develop facility in basic set theoretic notation.

Sets are the mathematician's way of thinking about *collections* of objects. Mathematical examples will be the set of natural numbers, the set of pairs of natural numbers, the set of integers, and so on.

An simple example is a set representing poker cards. We may denote it by

$$\begin{aligned}\mathsf{Deck} = \{&A\clubsuit, 2\clubsuit, 3\clubsuit, 4\clubsuit, 5\clubsuit, 6\clubsuit, 7\clubsuit, 8\clubsuit, 9\clubsuit, 10\clubsuit, J\clubsuit, Q\clubsuit, K\clubsuit,\\ &A\diamondsuit, 2\diamondsuit, 3\diamondsuit, 4\diamondsuit, 5\diamondsuit, 6\diamondsuit, 7\diamondsuit, 8\diamondsuit, 9\diamondsuit, 10\diamondsuit, J\diamondsuit, Q\diamondsuit, K\diamondsuit,\\ &A\spadesuit, 2\spadesuit, 3\spadesuit, 4\spadesuit, 5\spadesuit, 6\spadesuit, 7\spadesuit, 8\spadesuit, 9\spadesuit, 10\spadesuit, J\spadesuit, Q\spadesuit, K\spadesuit,\\ &A\heartsuit, 2\heartsuit, 3\heartsuit, 4\heartsuit, 5\heartsuit, 6\heartsuit, 7\heartsuit, 8\heartsuit, 9\heartsuit, 10\heartsuit, J\heartsuit, Q\heartsuit, K\heartsuit\}\end{aligned}$$

The elements are arranged here conveniently, but we could just as well have listed the cards in any "shuffled" order. The set of them would be the same.

Functions, introduced formally in Lecture 11 are the mathematicians way of thinking about attributes of the things in a collection, like "the color of", "the mass of", "the location of", "the father of", "the favorite book of the person to the left of" and so on. For our example of cards in a poker deck, "rank of" or "suit of" are two attributes. So we might write $\mathsf{rank}(A\diamondsuit) = A$ and $\mathsf{suit}(A\diamondsuit) = \diamondsuit$. In general, $\mathsf{rank}(x)$ and $\mathsf{suit}(x)$ pick out these attributes of a card x. The values these atrributes can take are also set $\mathsf{Rank} = \{A, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K\}$ and $\mathsf{Suit} = \{\clubsuit, \diamondsuit, \spadesuit, \heartsuit\}$.

The functions rank and suit capture some structure of the elements of Deck. For any rank r and any suit s, there is exactly one card c so that $\mathsf{rank}(c) = r$ and $\mathsf{suit}(c) = s$. For example, if $\mathsf{rank}(c) = 4$ and $\mathsf{suit}(c) = \spadesuit$, then we know exactly what c must be. So the two functions, in a sense, explain what a card is. We will use functions and sets to discuss more complicated structures, but the idea will be very similar to this simple example.

Taken together, sets and functions constitute a fundamental structure in contemporary mathematics called the *Category of Sets and Functions*. This is a slight lie. Actually, there are many different categories of sets and functions that differ in subtle ways. But for most mathematics, the differences are irrelevant. So in practice, it is safe to talk as if there is just one. The Category of Sets and Functions sometimes abbreviated as **Set**.

To understand sets and functions as they are used in every day mathematics, we need to answer some questions:

- What do we mean by saying that a set is a collection?
- What do we mean by saying that two sets are equal?
- What do we mean by saying that a function behaves like an attribute?
- What do we mean by saying that two functions are equal?
- How do we construct sets and functions?

The answers to these questions lead to some basic principles governing the ways that mathematicians actually use sets and functions. We could be more formal and present these principles as *axioms*, but the word "axiom" is used in mathematics to suggest that we intend the principles to be formal and to express everything in a minimal way. We do not need either the formality or the minimality here. Nevertheless, everything we say in these lectures could be presented that way.

10.1 Set Basics

A set consists of things that are "in" the set. All other things are "not in" the set. In our running example, A$\spadesuit$ is in the set Deck, but 25 is not in Deck. Let us make the idea precise.

Principle 1: Basic Structure of Sets

A **set** is a mathematical entity X with the following feature. For any mathematical entity x, either x **is in** X or x **is not in** X. We write $x \in X$ if x is in X and $x \notin X$ if x is not in X.

The symbol $\in$ is used in mathematics exclusively to indicate membership in a set. You will not see it used in any other way.

For variety, all of the following phrases mean the same thing:

- x is in X
- x is an **element of** X

- x is a **member of** X
- X **contains** x
- x **belongs to** X

Principle 1 describes how we can talk about sets and elements, and how to use the notation of membership, but does not tell us that any sets actually exist. Our first remedy for this is to make room for finite sets.

Principle 2: Finite Sets

For any list $L = [a_0, \ldots, a_{n-1}]$, there is a set, denoted by $\{a_0, \ldots, a_{n-1}\}$, so that $x \in \{a_0, \ldots, a_{n-1}\}$ if and only if $x = a_i$ for some $i < n$. More precisely,

- $x \notin \{\}$ for any x ($\{\}$ is said to be **empty**);
- $x \in \{a_0, \ldots, a_n\}$ if and only if $x = a_0$ or $x \in \{a_1, \ldots, a_n\}$.

Example 3

Here are some examples of sets built from finite lists:

- $\{\}$ – an empty set;
- $\{1, 2, 5\}$ – a set consisting of three elements;
- $\{\{\}\}$ – a set consisting of one element, which is $\{\}$;
- $\{1, 2, 4, \{1, 2\}\}$ – a set consisting of four elements, 1, 2, 4 and the set $\{1, 2\}$.
- $\{4, 5, \{\}, []\}$ – a set consisting of four elements. Note that the set $\{\}$ and the list $[]$ are not the same things.
- $\{1, 2, 3, 4, 3, 2, 1\}$ – a set consisting of four elements, listing an element twice is redundant.
- The sets Deck, Rank and Suit from the introduction.

The study of finite sets is surprisingly complex, and comprises a large part of the branch of mathematics called *combinatorics*. We will touch on the basics of combinatorics later in the course.

Various infinite sets of numbers also exist, but these follow from general principles we have not discussed yet. We do not try to justify anything for now, but introduce them informally along with the standard symbols used to denote them.

Definition 4

The following sets are denoted by the special symbols:

$$\begin{aligned}\mathbb{N} &= \text{the set of natural numbers}\\ \mathbb{Z} &= \text{the set of integers}\\ \mathbb{Q} &= \text{the set of rational numbers}\\ \mathbb{R} &= \text{the set of real numbers}\\ \mathbb{C} &= \text{the set of complex numbers}\end{aligned}$$

Exercises for Lecture 10

1. Let $A = \{1, \{2, 3\}, 4\}$. Determine which of the following assertions are true.

    1. $1 \in A$

Exercises for Lecture 10 (cont.)

2. $2 \in A$
3. $\{\} \in A$
4. $\{2, 3\} \in A$
5. $A \in A$

2. In the following examples of sets with elements following a pattern, write an expression for the same set that makes the pattern clearer.

    1. $\{0, 2, 4, \ldots, 100\}$
    2. $\{1, 2, 4, 8, \ldots, 256\}$
    3. $\{0, 1, 3, 6, 10, \ldots, 55\}$

3. $\bullet \in \mathsf{Suit}$ (the finite set defined above).

10.2 Subsets and Extensionality

A set is meant to be a collection: some things are in, some are not. That's all we can say. Unlike a list, a set has no "initial" element. For example, the set $\{1, 2, 3\}$ is the same as the set $\{2, 3, 1\}$, because both have the same elements. This is one important difference between lists and sets: $[1, 2, 3]$ and $[2, 3, 1]$ are *not* the same list because order matters in lists. To make this precise, we need to be clear about when sets are equal. To help, we introduce an important definition.

Definition 5

For sets X and Y, we say that X **is a subset of** Y provided that every element of X is an element of Y. We write this as $X \subseteq Y$, and say that X **is included in** Y. We may also write $Y \supseteq X$ to mean the same thing, and say that Y **is a superset of** B.

If X is *not* a subset of Y, we write $X \not\subseteq Y$. If $X \subseteq Y$ and $Y \not\subseteq X$, then X is called a **proper subset of** Y. To indicate that X is a proper subset of Y, we may write $X \subsetneq Y$. Some people write $X \subset Y$ for proper subsets, but we will never use that symbol in this course.

To say $X \subseteq Y$ is exactly to say that for any x, if $x \in X$ then $x \in Y$. In plain English, we may translate it informally as "all Xs are Ys." For example, suppose P is the set of all professors, and H is the set of all human beings. Then $P \subseteq H$ is the (dubious) assertion that "all professors are human beings".

Example 6

Here are some examples and counter-examples of the subset relation.

- $\{1, 2, 3\} \subseteq \{0, 1, 2, 3\}$
- $\{\} \subseteq \{0\}$
- $X \subseteq X$ for any set X because, trivially, every element of X is an element of X
- $\{\} \subseteq X$ for any set X because every element of $\{\}$ (there are none) is an element of X
- $\{1, 2, 3\} \not\subseteq \{0, 2, 3\}$ because $1 \in \{1, 2, 3\}$ but $1 \notin \{0, 2, 3\}$
- $\{1, 2, 3\} \subseteq \{2, 3, 1\}$
- $\{\spadesuit\} \subseteq S$

Exercises for Lecture 10

For each of the following pairs of sets, determine whether or not the first is a subset of the second. Explain your answer.

1. $\{0, 1\}$ and $\{1, 0\}$
2. $\{a, b, c, d\}$ and $\{a, b, d, e, c\}$
3. $\{\}$ and $\{\{\}\}$
4. $\{0, 3, 6, 10\}$ and $\{10, 9, 8, 7, 6, 5, 4, 2, 1, 0\}$

We can summarize two useful properties of $\subseteq$ as follows.

- [Reflexivity] For any set X, $X \subseteq X$. We say $\subseteq$ is *reflexive*.
- [Transitivity] For any sets X, Y and Z, if $X \subseteq Y$ and $Y \subseteq Z$, then $X \subseteq Z$. We say $\subseteq$ is *transitive*.

Another familiar example of a reflexive, transitive relation is $\leq$ on the natural numbers. In fact there are many examples of reflexive, transitive relations throughout mathematics. The relation $\leq$ is also *anti-symmetric*, meaning that if $m \leq n$ and $n \leq m$ then $m = n$. Suppose $X \subseteq Y$ and $Y \subseteq X$. Then, by definition X and Y have exactly the same elements. By our understanding of sets as collections, X and Y must be equal. So we state this as another princple.

Principle 7: Set Extensionality

For sets X and Y, if $X \subseteq Y$ and $Y \subseteq X$, then $X = Y$. In other words, $\subseteq$ is anti-symmetric.

Based on this, we can already establish a useful fact: there is exactly one empty set. To set the tone for what follows, we make this a formal claim.

Lemma 8

There is exactly one empty set.

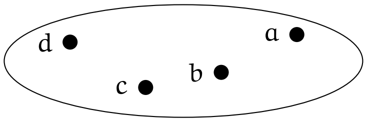
Proof: We have already noted that the set built from an empty list $\{\}$ has no elements. So there is at least one empty set.

Suppose E is a set with no elements. Then $E \subseteq \{\}$ because every element of E (there are none) is an element of $\{\}$. Similarly, $\{\} \subseteq E$ because every element of $\{\}$ (again, there are none) is an element of E. So by Principle **??** $E = \{\}$. □

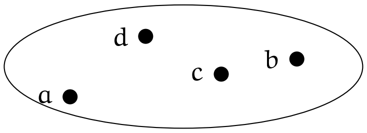
Definition 9

The set $\{\}$ is also denoted by $\emptyset$.

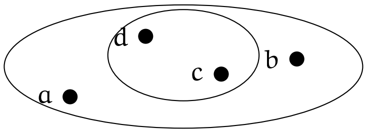
Set extensionality makes precise the idea that a set by itself does not have any structure other than what members it possesses. To emphasize this, sometimes it is useful to depict a set with elements scattered about something like



with the elements scattered about. Evidently, a re-arrangement of the elements does not change the depicted set. So



is the same set. Depicting a subset of a set is a simple matter of drawing a smaller boundary around some of the elements as in the following.



Exercises for Lecture 10

Draw depictions of the following sets

1. $\{1, 4, 5, 2, 3\}$
2. $\{1, 2, 3, \ldots, 23\}$
3. $\{a, b, c, d, e\}$ and $\{c, e, f, g\}$ on the same diagram
4. $\{a, e, b, c, e\}$ [sic]
5. $\{1, 3, 6, 7\}$ and $\{1, 3, 5, 6, 7, 9\}$ on the same diagram
6. $\{\textsc{f}, \textsc{t}$
7. $\{\bullet\}$
8. $\{\textsc{t}, \textsc{f}, 3, 5, 1, \bullet\}$

LECTURE 11

Functions

Goals

Lecture

- Introduce basic structure of functions
- Define the identity functions and function composition
- Introduce internal diagrams of functions.

Study

- Be able to determine equality of functions
- Use internal diagrams to depict function composition

Functions (perhaps in your calculus courses) are often talked about as *operations*. For example,

$$f(x) = x^2$$

can be seen as an operation that transforms a number x into its square. But it can also be seen as an attribute (the "square of x"). The "operational" view is informal, and often useful. As we will see, though, it gets an important aspect of functions wrong because two entirely different operations may define the same function.

Informally, a function "takes" an element of a given set as input and "produces" an element of a given set as output. So the function f defined by $f(x) = x^2 + 2x + 1$ might "take" the natural number 2 and "produce" the natural number 10. That is, $f(3) = 3^2 + 1 = 10$. We begin by making this idea formal, introducing the vocabulary of functions.

Principle 1: Basic Structure of Functions

- For a set X and a set Y, there are things called **functions from** X **to** Y. We write $f\colon X \to Y$ or $A \xrightarrow{f} B$ to indicate that f is a function from X to Y.
- For $f\colon X \to Y$, the set X is called the **domain** of f and Y is called the **codomain** of f.

Principle 1 (cont.)

- For any function $f\colon X \to Y$ and any element $a \in X$, f and a determine an element of Y, written $f(a)$, and read "f of a".

A function may sometimes also be called a **map**, a **transformation**, or an **operation**. As we will see, however, **operation** is somewhat misleading, so we usually avoid it.

Often, a function $f\colon X \to Y$ is *defined* by a rule. We typically write such rules by giving the function a name (very often f because we are lazy) and spelling out the rule at the same time. So we write things like

$$f(x) = x^2 + 4x + 2$$

to define a function $f\colon \mathbb{R} \to \mathbb{R}$ (recall that $\mathbb{R}$ is the set of real numbers). But sometimes it is useful to have a rule without giving it a name. To do that, we will use the "maps to" arrow $\mapsto$. So we may define the same function f by saying that f **is given by the rule**

$$x \mapsto x^2 + 4x + 2.$$

The rule $x \mapsto x^2 4x + 2$ is the same as the rule $y \mapsto y^2 + 4y + 2$. The variable only serves as a place holder, so its particular name does not matter.

There are two fundamental (trivial) types of rules that can be used to build functions.

Principle 2: Identities and Function Composition

- For any set X, there is a function $\mathsf{id}_X\colon X \to X$ defined by the rule $x \mapsto x$. This is called the **identity** function on A.

- For any two functions $f\colon X \to Y$ and $g\colon Y \to Z$, there is a function $g \circ f\colon X \to Z$ defined by the rule $x \mapsto g(f(x))$. This is called the **composition of** g **and** f (or sometimes g **following** f).

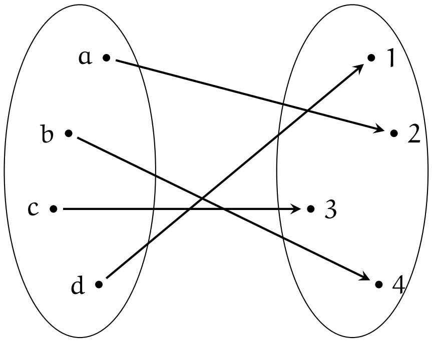
Notice that $g \circ f$ is only defined when the *domain* of g matches exactly the *codomain* of f.

Exercises for Lecture 11

1. Suppose $f\colon W \to X$, $g\colon X \to Y$ and $h\colon Y \to Z$ are functions. Then $h \circ (g \circ f)$ and $(h \circ g) \circ f$ are functions from W to Z. Do you think they are equal? Explain your answer in a few clearly written sentences.

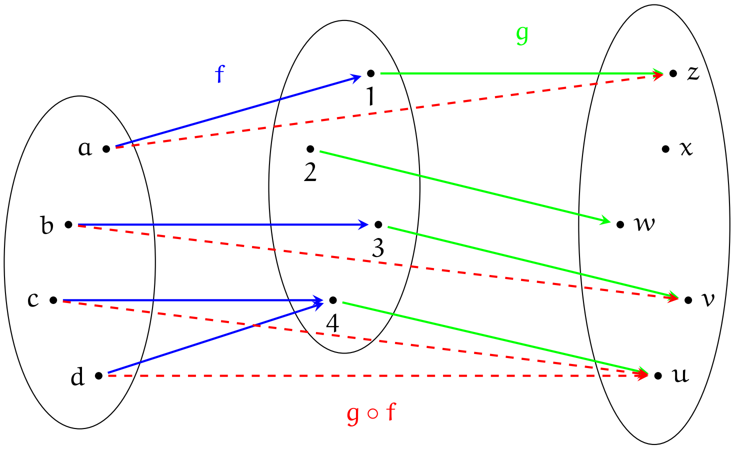
11.1 Internal Diagrams

To depict a function on small sets, we can use the internal diagrams of the last section with arrows indicating the input/output relationship. For example,



depicts a function from the set $\{a, b, c, d\}$ to the set $\{1, 2, 3, 4\}$.

Composition can also be illustrated using internal diagrams. For example,



Exercises for Lecture 11

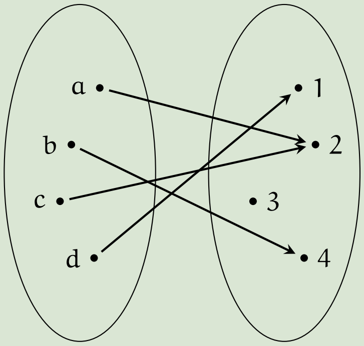
Use internal diagrams for the following exercises.

1. Depict four different functions from the set $\{1, 2, 3\}$ to the set $\{\mathrm{F}, \mathrm{T}\}$. [Draw four different diagrams.]
2. Depict all of the functions from $\{\bullet\}$ to $\{a, b, c\}$
3. Depict all of the functions from $\{a, b, c\}$ to $\{\bullet\}$
4. Are there any functions from $\{a, b\}$ to $\emptyset$?

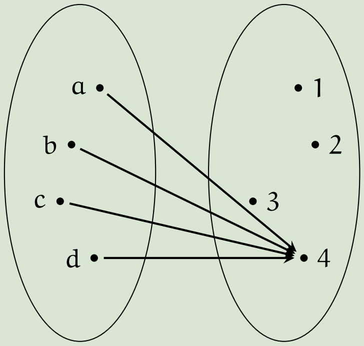
Exercises for Lecture 11 (cont.)

5. Are there any functions from $\emptyset$ to $\{a, b\}$? If there are, how many?

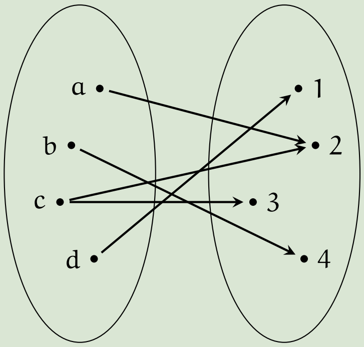
6. For each of the following diagrams, determine whether or not the diagram depicts a function. If not, explain why not.



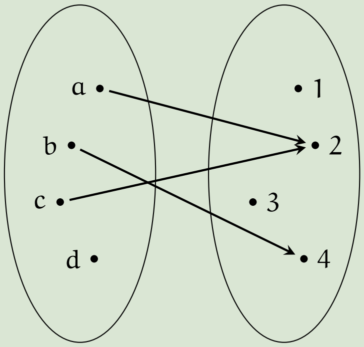
1.



3.



2.



4.

7. Let $A = \{1, 2, 3\}$. Let $B = \{a, b, c, e\}$ and let $C = \{\textsc{f}, \textsc{t}\}$. Depict some functions $f\colon A \to B$, $g\colon B \to C$, and $g \circ f$.

8. Think about how you might depict a function $h\colon A \to A$ using only one picture of the set A. Describe what you would do, and provide an example.

9. Suppose $f\colon \mathbb{R} \to \mathbb{R}$ is given by the rule $x \mapsto x^2$, suppose $g\colon \mathbb{R} \to \mathbb{R}$ is given by the rule $x \mapsto x - 1$. Write rules for $f \circ g$ and $g \circ f$ without using the symbols f and g. Explain whether or not it is the case that $f \circ g = g \circ f$.

11.2 Extensionality

As with sets, we need a way to say when two functions are equal. Consider an example. Recall that $\mathbb{N}$ denotes the set of natural numbers. Then define $f\colon \mathbb{N} \longrightarrow \mathbb{N}$ and $g\colon \mathbb{N} \longrightarrow \mathbb{N}$ by

$$\begin{aligned} f(n) &= n^2 + 2n + 1 \\ g(n) &= (n+1)^2 \end{aligned}$$

Evidently, for each $n \in \mathbb{N}$, it is true that $f(n) = g(n)$. So even though f and g are defined by different *operations*, the two functions yield the same results. This leads to a principle for equality of functions.

Principle 3: Equality of Functions

For functions $f\colon X \longrightarrow Y$ and $g\colon X \longrightarrow Y$, if it is the case that $f(x) = g(x)$ for all $x \in X$, then $f = g$. Note that equality of functions only makes sense when the two functions share the same domain and the same codomain.

When we are not concerned about the detailed internals of sets, but only with how functions interact, then an individual function can be depicted very simply as $X \xrightarrow{f} Y$. So a composition of functions can be depicted as in

$$\begin{tikzcd} X \arrow[r, "f"] \arrow[dr, "g \circ f"'] & Y \arrow[d, "g"] \\ & Z \end{tikzcd}$$

We do not really need to draw $g \circ f$ as a separate arrow because the *path* from X to Y to Z is already an implicit depiction of $g \circ f$. So the simpler diagram

$$\begin{tikzcd} X \arrow[r, "f"] & Y \arrow[d, "g"] \\ & Z \end{tikzcd}$$

shows the same information, namely, that $f\colon X \longrightarrow Y$ and $g\colon Y \longrightarrow Z$ are functions and therefore, $g \circ f\colon X \longrightarrow Z$ is too.

Now a diagram such as this

$$\begin{tikzcd} W \arrow[r, "f"] \arrow[d, "h"'] & X \arrow[d, "g"] \\ Y \arrow[r, "k"'] & Z \end{tikzcd}$$

depicts two composite functions $g \circ f$ and $k \circ h$, but $g \circ f$ and $k \circ k$ may not be equal. We say that the diagram *commutes* or that it is a *commutative diagram* if $g \circ f = k \circ h$. In other words, saying that a certain diagram commutes *is* an assertion that certain functions are equal.

Exercises for Lecture 11

1. For each of the following pairs of functions $\mathbb{N} \longrightarrow \mathbb{N}$, determine whether they are equal and explain why or why not.
    1. $f(n) = 2n + 3$ and $g(m) = 2m + 3$
    2. $f(n) = 2^{n+1} - 1$ and $g(n) = \sum_{i=0}^{n} 2^i$
    3. $f(n) = n^2 + 5n + 6$ and $g(n) = (n + 3)(n + 2)$
    4. $f(n) = n^4 - 10n^3 + 35n^2 + 50n + 24$ and $g(n) = 24$
2. Let $\mathbb{R}$ denote the set of all real numbers. Let $f(x) = \tan(x)$. Explain why this does *not* define a function from $\mathbb{R}$ to $\mathbb{R}$.
3. Suppose the following functions exist: $f\colon W \longrightarrow X$, $g\colon X \longrightarrow Y$, $a\colon W \longrightarrow Z$, $b\colon Y \longrightarrow Z$. Draw a commutative diagram asserting that $b \circ g \circ f = a$.
4. Suppose the following functions exist: $f\colon C \longrightarrow A$, $g\colon C \longrightarrow B$, $h\colon C \longrightarrow P$, $p\colon P \longrightarrow A$ and $q\colon P \longrightarrow B$. Draw a commutative diagram asserting that $f = p \circ h$ and $g = q \circ h$.

Lecture 12

Constructing Sets and Functions

Goals

Lecture

- Characterize and define
  - Pointer and constant functions
  - Solution sets
  - Characteristic functions
  - Products of sets
  - Exponents of sets

Study

- Be able to calculate membership in various constructed sets
- Learn to use universal constructions to define functions.

So far, we have thought mainly about informally defined sets and functions. To fill out our understanding of sets, we need to be able to build sets and functions for specific purposes and with specific structure in mind.

Three finite sets will play particularly important roles in this. We have already discussed the set $\emptyset$, consisting of no elements. We also need a designated set with one element and a designated set with two elements. We denote these by $\mathbb{1}$ and $\mathbb{2}$. It does not matter at all *what* the elements are because, as we will soon see, sets of the same size are interchangable.

For the time being, we merely need to agree on a fixed reference set with one element and a fixed reference set with two elements. The particular choices we make here will be clearer as we put them to use.

Definition 1

Let •, F and T be fixed symbols. Then define

$$\mathbb{1} = \{\bullet\}$$
$$\mathbb{2} = \{\mathrm{F}, \mathrm{T}\}$$

The single element of $\mathbb{1}$ is intended to look like a generic point in an internal diagram. The element T is meant to indicate 'True' and F, 'False'.

12.1 Elements, Pointers and Constant Functions

Suppose we are told that $p\colon \mathbb{1} \to X$ is a function. Since $\bullet \in \mathbb{1}$, this function determines an element of X, namely $p(\bullet)$. A picture of the situation might be this:

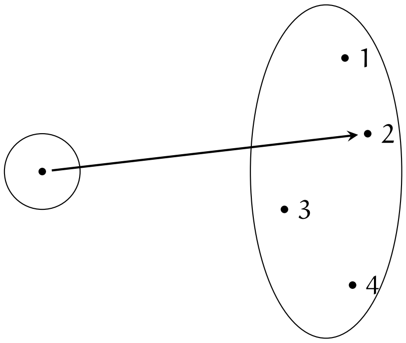


Figure 12.1: A function from $\mathbb{1}$ "points" to an element

Since • is the only element of $\mathbb{1}$, p can "point" only to a single element of X. So we might refer to a function $\mathbb{1} \xrightarrow{p} X$ as a **pointer** into X. Each pointer determines an element of X. And conversely, it should be possible to point to any element of X. This leads to a principle guaranteeing that certain (nearly trivial) functions exist.

Principle 2: Elements Determine Pointers

For any set X, and any $a \in X$, there is a function $\hat{a}\colon \mathbb{1} \to X$ given by the rule $x \mapsto a$.

Thus the function depicted in Figure 12.1 is $\hat{2}$. This principle simply asserts that elements of a set X and functions $\mathbb{1} \to X$ are interchangible: from $a \in X$ we get $\hat{a}\colon \mathbb{1} \to X$; from $p\colon \mathbb{1} \to X$. we get the element $p(\bullet)$.

Suppose $f\colon X \to \mathbb{1}$ and $g\colon X \to \mathbb{1}$ are functions, that is, their *codomain* is $\mathbb{1}$ instead their *domain*. Then $f(a) = \bullet = g(a)$ is true for every $a \in X$ because $\bullet$ is the only possible value. So $f = g$ by the Principle of Function Extensionality. In other words, there is at most one function from X to $\mathbb{1}$. But the rule $x \mapsto \bullet$ is as simple a rule as one can imagine. This leads to another definition and another principle.

Definition 3

A set T is **terminal** if it is the case that for any set X there is exactly one function from X to T.

Principle 4: A Terminal Set Exists

The set $\mathbb{1}$ is a terminal set. We denote the unique function from X to $\mathbb{1}$ by $\Diamond_X\colon X \to \mathbb{1}$.

The rule defining $\Diamond_X$ is

$$x \mapsto \bullet.$$

Using $\Diamond_X$ and $\hat{b}$ for an element $b \in Y$, we can now define a constant function. That is $\hat{b} \circ \Diamond_X$ is a function from X to Y given by the rule $x \mapsto \hat{b}(\Diamond_X(x)) = \hat{b}(\bullet) = b$. In short, this is the function sending all elements of X to the constant b. It will be convenient to have a standard name for this function.

Definition 5

For a sets X and Y, and element $b \in Y$, let $c_{X,b} = \hat{b} \circ \Diamond X$. When X is obvious form context, we omit it.

Exercises for Lecture 12

1. Show that for any pointer $p\colon \mathbb{1} \to X$, it is the case that $\widehat{p(\bullet)} = p$.

2. Show that any set with exactly one element is a terminal set.

3. Show that any terminal set has exactly one element.

4. Suppose that $f\colon X \to Y$ is a function. Show that for every $a \in X$, $\widehat{f(a)} = f \circ \hat{a}$.

12.2 The Empty Set

For trivial reasons, there is at most one function from $\emptyset$ to X, for any set X. That is, if $f, g \colon \emptyset \to X$ are functions, then for each $x \in \emptyset$, $f(x) = g(x)$ because there are no x's to concern us. Hence by Principle **??**, $f = g$. The empty "rule" that tells us to do nothing specifies a function from $\emptyset$ to X. So for any set X, there is exactly one function from $\emptyset$ to X. Let's make that official.

Definition 6

An **initial set** is a set I so that for any set X, there is exactly one function from I to X.

Principle 7: An Initial Set Exists

The emptyset $\emptyset$ is an initial set. For a set X, the unique function from $\emptyset$ to X (given by the empty rule) may be denoted by $\Box_A \colon \emptyset \to X$.

Notice that a function $X \to \emptyset$ is impossible unless X is also empty. So $\emptyset$ is the only initial set.

Exercises for Lecture 12

1. How many functions are there from $\emptyset$ to $\{a, b, c, d\}$? Explain.
2. How many functions are there from $\{a, b, c, d\}$ to $emptyset$? Explain.
3. How many functions are there from $\mathbb{1}$ to $\{a, b, c, d\}$? Explain.
4. How many functions are there from $\{a, b, c, d\}$ to $\mathbb{1}$? Explain.

12.3 Solution Sets, Subsets, Characteristic Functions

Suppose we are given two functions that are "parallel": $f \colon X \to Y$ and $g \colon X \to Y$. To aid readability, we will write this as $X \overset{f}{\underset{g}{\rightrightarrows}} Y$. For some values $a \in X$, it might be the case that $f(a) = g(a)$. Let us call such a value a **particular solution to the equation** $f(x) = g(x)$.

It might be the case that there are no particular solutions to an equation $f(x) = g(x)$. For example, there are no natural numbers n such that $n + 1 = n$. On the other hand, there might be many particular solutions. For example, let $f(x) = x^3$ and let $g(x) = 6x^2 - 11x + 6$ both regarded as functions on the natural numbers. Then it is easy to check that 1, 2 and 3 solve the equation $f(x) = g(x)$. In fact, these three are the only particular solutions. We generalize as follows.

Definition 8

For two functions $X \overset{f}{\underset{g}{\rightrightarrows}} Y$, a **solution** is a function $S \xrightarrow{s} X$ so that $f \circ s = g \circ s$. Thus for example, if $a \in A$ is a particular solution then the pointer $\hat{a}$ is a solution.

For functions $A \overset{f}{\underset{g}{\rightrightarrows}} B$, an **equalizer** is a solution $E \xrightarrow{e} X$ so that for any solution $S \xrightarrow{s} X$, there is exactly one function $S \xrightarrow{h} E$ so that $e \circ h = k$.

Principle 9: Equalizers Exist

For functions $X \overset{f}{\underset{g}{\rightrightarrows}} Y$, the collection of all particular solutions to the equation $f(x) = g(x)$ form a set, denoted by $\{x \in X \mid f(x) = g(x)\}$. The function $\{x \in X \mid f(x) = g(x)\} \xrightarrow{i} A$ given by the rule $x \mapsto x$ (called an **inclusion map**) is an equalizer for f and g.

If $S \xrightarrow{s} X$ is a solution (that is, $f \circ s = g \circ s$), then the function $C \xrightarrow{\check{s}} \{x \in A \mid f(x) = g(x)\}$ given by the rule

$$x \mapsto s(x)$$

is the unique function for which $s = i \circ \check{s}$.

This axiom tells us three main things. First, we can form a subset of X by specifying an equation $f(x) = g(x)$ for two functions $X \overset{f}{\underset{g}{\rightrightarrows}} Y$, and picking out the particular solutions. Second, a subset formed in this way "embeds" in the given set X by its inclusion map i. Third, for any solution s, the function into the set of particular solutions is defined by the same rule as s.

External diagrams help us understand equalizers. An equalizer is a solution

$$\begin{tikzcd} E \arrow[r, "e"'] & X \arrow[r, shift left, "f"] \arrow[r, shift right, "g"'] & Y \end{tikzcd}$$

so that if

$$\begin{tikzcd}
S \arrow[dr, "s"] & & \\
E \arrow[r, "e"'] & X \arrow[r, shift left, "f"] \arrow[r, shift right, "g"'] & Y
\end{tikzcd}$$

is also a solution ($f \circ s = g \circ s$), then there is exactly one function making

$$\begin{tikzcd}
S \arrow[d, "\check{s}"'] \arrow[dr, "s"] & & \\
E \arrow[r, "e"'] & X \arrow[r, shift left, "f"] \arrow[r, shift right, "g"'] & Y
\end{tikzcd}$$

commute.

Inverse Image of an Element

Suppose $c \in Y$ and $X \xrightarrow{f} Y$ is a function, then we can form the equalizer of f and the constant function $\hat{c} \circ \Diamond_X$. This is more easily written we $\{x \in X \mid f(x) = c\}$. Since it is common to pick out sets like this, special notation is in order.

Definition 10

For a function $X \xrightarrow{f} Y$, and an element $c \in Y$,

$$f^{-}(c) = \{x \in X \mid f(x) = c\}.$$

In this case, $f^{-}(c)$ is called the **inverse image of** c **with respect to** f.

A diagram can help us understand inverse images as well. Suppose $f\colon X \to Y$ and $c \in C$, then we can arrange a diagram

$$\begin{tikzcd}
 & \mathbb{1} \arrow[d, "\hat{c}"'] \\
X \arrow[r, "f"'] & Y
\end{tikzcd}$$

The inverse image is a subset of X with an inclusion map that makes the following diagram commute:

$$\begin{tikzcd}
f^{-}(c) \arrow[r, "\Diamond_{f^{-}(c)}"] \arrow[d, "i"'] & \mathbb{1} \arrow[d, "\hat{c}"] \\
A \arrow[r, "f"'] & C
\end{tikzcd}$$

For any other function $W \xrightarrow{g} X$ that makes following similar diagram commute:

$$\begin{tikzcd}
W \arrow[drr, "\Diamond_W"] \arrow[ddr, "g"'] & & \\
& f^{-}(c) \arrow[r, "\Diamond_{f^{-}(c)}"'] \arrow[d, "i"'] & \mathbb{1} \arrow[d, "\hat{c}"] \\
& A \arrow[r, "f"'] & C
\end{tikzcd}$$

there is a unique function $B \xrightarrow{\check{g}} f^{-}(c)$ making

$$\begin{tikzcd}
W \arrow[drr, "\Diamond_W"] \arrow[dr, "\check{g}"] \arrow[ddr, "g"'] & & \\
& f^{-}(c) \arrow[r, "\Diamond_{f^{-}(c)}"'] \arrow[d, "i"'] & \mathbb{1} \arrow[d, "\hat{c}"] \\
& X \arrow[r, "f"'] & Y
\end{tikzcd}$$

commute.

In Definition 10, the set $f^{-}(c)$ is a subset of X. It would be good to know that any subset of X can be described as an inverse image. This is where the set $\mathbb{2}$ plays a role.

Definition 11

Subset Classifier

A **pointed set** is a set P with a distinguished element $p \in P$.

A **subset classifier** is a set T with a distinguished element $t \in T$ so that for any set X and any subset $A \subseteq X$, there is exactly one function $k\colon X \to T$ for which $A = k^{-}(t)$. That is, A is

Definition 11 (cont.)

uniquely defined as the inverse image of t with respect to a function into T.

Principle 12: $\mathbb{2}$ and Characteristic Functions

The set $\mathbb{2}$ with the distinguished element T is a subset classifier. For subset $A \subseteq X$, the function corresponding to A, called the **characteristic function of** A, is denoted by κ_A. In other words, κ_A is the unique function for which $A = \kappa_A^{-}(\mathrm{T})$.

For $A \subseteq X$, the characteristic function is defined by the rule

$$x \mapsto \begin{cases} \mathrm{T} & \text{if } x \in A \\ \mathrm{F} & \text{otherwise} \end{cases}$$

Just as Principle **??** asserts that elements of X and functions $\mathbb{1} \to X$ are interchangeable, Principle 12 asserts that the subsets of X and the functions $X \to \mathbb{2}$ are interchangeable.

Exercises for Lecture 12

1. Draw a depiction of $A = \{a, b, c, d, e, f, g\}$ and its subset $B = \{a, c, e, g\}$ in the same internal diagram. Now depict the characteristic map for B as a subset of A.

2. Define two functions $\mathbb{N} \underset{g}{\overset{f}{\rightrightarrows}} \mathbb{N}$ so that the set of particular solutions of $f(x) = g(x)$ is $\{1, 5\}$.

3. Consider the functions $f\colon \mathbb{R} \to \mathbb{R}$ defined by $f(x) = \sin(x) + \cos(x)$, and $s\colon \mathbb{N} \to \mathbb{R}$ defined by $s(n) = 2\pi n^2$. Is s a solution for the equation $f(x) = -1$? What is the set of all particular solutions?

4. Describe what a subset classifier is, using diagrams similar to the diagrams we have used to describe equalizers and inverse images. That is, for a start, we have a diagram

$$\begin{tikzcd} \mathbb{1} \arrow[d, "\hat{t}"] \\ T \end{tikzcd}$$

Exercises for Lecture 12 (cont.)

Now suppose we are given a subset $A \subseteq X$ with its inclusion map:

$$\begin{tikzcd}
A \arrow[r, "\Diamond_A"] \arrow[d, "i"'] & \mathbb{1} \arrow[d, "\hat{t}"] \\
X & S
\end{tikzcd}$$

What additional function is required to exist? What properties is it required to have? [Hint: the result should be that A is an inverse image.]

12.4 Product Sets and Functions of Two Arguments

We should be able to deal with functions of more than one argument, such as a function $f(x, y) = x^2 + y^2$. To account for these, we take our cue from Descartes.

Descartes studied the geometric plane in terms of a coordinate system consisting of the so-called x-axis and y-axis (what we call cartesian coordinates in his honor). Once we have decided where to place the axes (as long as they do not run in parallel), a pair such as $(2, 3)$ determines a point on the plane, and any point p in the plane determines a pair. So Descartes realized that we might as well just say that the plane actually *is* the collection of all pairs of real numbers. What makes this work is that points in the plane **project** onto the two axes in a universal way. Products of sets generalize this idea.

Definition 13

For sets X and Y, a **table** consists of two functions $X \xleftarrow{f} T \xrightarrow{g} Y$. Note that the two functions have the same domain. We may call the two functions **legs** of the table.

For sets X and Y, a **product of** X **and** Y is a table $X \xleftarrow{p} P \xrightarrow{q} Y$ so that for any table $X \xleftarrow{f} T \xrightarrow{g} Y$ there is exactly one function $T \xrightarrow{h} P$ for which $f = p \circ h$ and $g = q \circ h$. For a product, the legs p and q are called the **projections**.

Principle 14: princ:Products

For sets X and Y, the collection of all pairs (x, y) where $x \in X$ and $y \in Y$ is a set, denoted by $X \times Y$. The functions $X \xleftarrow{\pi_0} X \times Y \xrightarrow{\pi_1} Y$ given by the rules $(x, y) \longmapsto x$ and $(x, y) \longmapsto y$ are

Principle 14 (cont.)

projections. For $X \xleftarrow{f} T \xrightarrow{g} Y$, the unique function required by the product may be denoted by $\langle f, g\rangle$.

For $X \xleftarrow{f} T \xrightarrow{g} Y$, the function $T \xrightarrow{\langle f,g\rangle} X \times Y$ is given by the rule $t \mapsto (f(t), g(t))$.

As with equalizers and inverse images, products can be described in terms of diagrams. A product of X and Y is depicted as an external diagram

$$\begin{tikzcd}
X & P \arrow[l, "p"'] \arrow[r, "q"'] & Y
\end{tikzcd}$$

so that for any other table over X and Y:

$$\begin{tikzcd}
 & T \arrow[dl, "f"'] \arrow[dr, "g"] & \\
X & P \arrow[l, "p"] \arrow[r, "q"'] & Y
\end{tikzcd},$$

there is a unique function making

$$\begin{tikzcd}
 & T \arrow[dl, "f"'] \arrow[dr, "g"] \arrow[d, "{\langle f, g\rangle}"'] & \\
X & P \arrow[l, "p"] \arrow[r, "q"'] & Y
\end{tikzcd},$$

commute.

Suppose we are given two unrelated functions $X \xrightarrow{f} Y$ and $A \xrightarrow{g} B$. We can form a single function from $X \times A \xrightarrow{f \times g} Y \times B$ by combining f and g "independently". That is, define $f \times g = \langle f \circ \pi_0, g \circ \pi_1 \rangle$. Calculating concretely in terms of elements $(f \times g)(x, y) = (f(x), g(y))$. So $f \times g$ acts on a pair (x, y) by applying f to x and unrelatedly applying g to y.

Products can by generalized to three, four or more sets. For example, given sets X, Y and Z, we might write $X \times Y \times Z$ for the set of triples (x, y, z) where $x \in X$, $y \in Y$ and $z \in Z$. Instead of two projections, this would have three projections $(x, y, z) \mapsto x$, and so on. It turns out, however, that binary products are enough because $X \times (Y \times Z)$ behaves just like $X \times Y \times Z$.

Exercises for Lecture 12

1. For the sets $A = \{a, b, c\}$ and $B = \{1, 2, 3, 4\}$, write out $A \times B$ and $B \times A$

2. What is $\emptyset \times A$?

Exercises for Lecture 12 (cont.)

3. Write out $\{4, a, 0\} \times 2$.
4. Describe in plain English what are the elements of $\mathbb{N} \times \mathbb{N}$.
5. Suppose A is a finite set with m elements and B is a finite set with n elements. How many elements are in $A \times B$?
6. Describe in plain English why it makes sense to refer to $A \times B$ as a "product."
7. For sets $A = \{a, b\}$, $B = \{0, 1, 2\}$ and $C = \{c, d\}$, calculate $A \times B \times C$, $A \times (B \times C)$ and $(A \times B) \times C$. Are these sets equal? If not, how do they differ.
8. Describe a model of the standard fifty-two card poker deck as a product of two sets.

12.5 Function Sets and Parametric Functions

A function from X to Y might depend on a parameter from another set P. For example, the function $\mathbb{R} \xrightarrow{f} \mathbb{R}$ given by the rule $x \mapsto \sin(x + c)$ depends on the constant c. There is a related function $\mathbb{R} \times \mathbb{R} \xrightarrow{g} \mathbb{R}$ given by $(c, x) \mapsto \sin(x + c)$. Though g describes the same behavior as f, it makes the parameter c explicit as another argument. The relation between f and g leads to the following definition.

Definition 15

For sets P, X and Y, a **parametric function from** X **to** Y is a function $P \times X \xrightarrow{g} Y$ for some set P. The set P will be called the **parameter set**.

Suppose $Q \times X \xrightarrow{f} Y$ is a parametric function with parameter set Q and $P \xrightarrow{k} Q$ is a function. Then another parametric function with parameter set P by composing: $f \circ (k \times \mathsf{id}_X)$. The function k acts like a **change of parameters** because it transforms the parametric function with parameters in Q into a parametric function with parameters in P. Specifically, $f \circ (k \times \mathsf{id}_X)$ is given by the rule $(c, a) \mapsto f(k(c), a)$.

An **evaluation map** for X and Y is a parametric function $F \times X \xrightarrow{a} Y$ with parameter set F, so that for any parametric function $P \times X \xrightarrow{g} Y$ there is exactly one change of parameters $h \colon P \to F$ so that $g = a \circ (h \times \mathsf{id}_X)$. In that case, F is called an **exponential with base** Y **and exponent** X.

Principle 16: Function Sets Exist

For sets X and Y, the collection of all functions from X to Y, denoted by Y^X, is a set.

The rule $(f, x) \mapsto f(x)$ defines an evaluation map $Y^X \times X \xrightarrow{\mathsf{appl}} Y$.

For a parametric function $f\colon P \times X \to Y$, the unique function from P to Y^X determined by f does not have a completely standard name. Increasingly, mathematicians honor the twentieth century logician, Haskell Curry, by referring to this as 'currying'. For these lectures, we follow that tradition and write $\mathsf{curry}[f]$ for the unique function satisfying $f = \mathsf{appl} \circ (\mathsf{curry}[f] \times \mathsf{id}_X)$.

Calculating how $P \xrightarrow{\mathsf{curry}[f]} Y^X$ must behave, we see that for any parameter $p \in P$, $\mathsf{curry}[f](p) \in Y^X$ is the function from X to Y given by the rule $x \mapsto f(p, x)$.

λ Notation

In defining $\mathsf{curry}[f]$, we needed to describe certain elements of Y^X. But elements of Y^X are functions. And a function typically is described by a rule. So it would be convnient to have a notation that permits us to describe the behavior of a function without giving the function a name. The logician Alonzo Church was interested in the fundamental idea of just what *is* a function. He proposed a notation for describing functions, writing things like $\lambda x.x^2$ to describe the function that squares its input. The Greek letter λ means nothing. It is used only as a marker to introduce a function. The "λ" notation is widely adopted in computer science. Indeed, it appears even in languages such as Python. We could make the λ notation formal (as did Church), but for our purposes informality is enough. We use this notation to describe elements of Y^X. Several examples will help to explain this.

Example 17

- For $f \in Y^X$ and $g \in Z^Y$, the composite function $g \circ f \in Z^X$ is $\lambda x.g(f(x))$.
- The element of $\mathbb{N}^{\mathbb{N}}$ defined by $\lambda x.x^{\curvearrowright}$ is the successor function.
- For any $a \in X$, the function $\hat{a} \in X^{\mathbb{1}}$ is $\lambda x.a$.
- For $X \xleftarrow{f} T \xrightarrow{g} Y$, the function $\langle f, g\rangle X \times Y^T$ is $\lambda x.(f(x), g(x))$.
- For a parametric function $P \times X \xrightarrow{f} Y$, the function $P \xrightarrow{\mathsf{curry}[f]} Y^X$ can be defined by the rule $p \mapsto \lambda x.f(p, x)$.
- For any $f \in Y^X$, $f = \lambda x.f(x) = \lambda y.f(y)$.
- The only element of $\mathbb{1}^X$ is $\lambda x.\bullet$.

Exercises for Lecture 12

1. For set $A = \{1, 2, 3\}$ and $B = \{a, b\}$

    1. draw internal diagrams corresponding to each element of B^A (there are eight of them);
    2. draw internal diagrams corresponding to each element of A^B (there are nine of them).

2. If X is a finite set with k elements, Y is a finite set with j elements, how many elements are there in the set Y^X?

3. Consider the function $\mathbb{N} \times \mathbb{N} \xrightarrow{f} \mathbb{N}$ defined by $f(m, n) = m^n$. What element of $\mathbb{N}^{\mathbb{N}}$ is $\mathsf{curry}[f](3)$? Use λ notation to describe it.

4. For the function min from $\mathbb{R} \times \mathbb{R}$ to $\mathbb{R}$ defined to mean the minimum of x and y, define $\mathsf{curry}[\mathrm{min}]$.

5. Use λ notation to describe the element of $\mathbb{N}^{\mathbb{N}}$ that quadruples the square of the input.

LECTURE **13**

The Set of Natural Numbers

Goals

Lecture:

- Re-introduce the natural numbers as a set
- Introduce sequences and recursively defined sequences
- Relate recursion to proofs by induction

Study:

- Be able to define simple functions by recursion
- Be able to explain how induction and recursion are related

We have used $\mathbb{N}$ informally to denote the set of natural numbers. It is time that we make the structure of $\mathbb{N}$ explicit within our theory of sets and functions. It will turn out that $\mathbb{N}$ is also a universal construction.

Natural numbers provide a precise picture of counting and of putting things in an order. Now that we have sets and functions we can consider a function $a\colon \mathbb{N} \to A$ to be an **infinite sequence**: $a(0), a(1), a(2), \ldots$. When we do that, we sometimes write $a_0, a_1, a_2, \ldots$ instead. Still a itself is just function from $\mathbb{N}$ to A. To emphasize the notation that a represents an infinite sequence, we sometimes also write $(a_i)_{i \in I}$.

Much of what we discuss in this lecture has the feel of computer programming. This is partly because natural numbers are the main objects of calculation. We want to understand, for example, how to define a function like $n \mapsto n!$ (n factorial) as a function from $\mathbb{N}$ to $\mathbb{N}$ by specifying how it is calculated. In particular, $0! = 1$ and $(n^{\curvearrowright})! = n^{\curvearrowright} \cdot n!$ characterize factorial by spelling out how to calculate it. For example,

$$\begin{aligned} 4! &= 4 \cdot 3! \\ &= 4 \cdot (3 \cdot 2!) \\ &= 4 \cdot (3 \cdot (2 \cdot 1!)) \\ &= 4 \cdot (3 \cdot (2 \cdot (1 \cdot 0!))) \\ &= 4 \cdot (3 \cdot (2 \cdot (1 \cdot 1))) & &= 24 \end{aligned}$$

It is quite common to think about a sequence in which a_{n+1} is functionally related to a_n. For example, in the sequence $1, 2, 4, 8, \ldots$, the initial entry is 1 and each successive entry is double its predecessor.

Indeed, if we know just those two facts – the initial entry is 1, and each subsequent entry is double its predecessor – then we know how the entire sequence behaves. Also, we know how to calculate the n^{th} entry, by recursion just like factorial.

The most basic sequence, of course, is $0, 1, 2, \ldots$. Its initial entry is 0 and each subsequent entry is the successor of its predecessor. So we think of the sequence that comprises $\mathbb{N}$ as a universal recursively defined sequence.

13.1 Sequences and Simple Recurrences

Let us make the informal word *sequence* official.

Definition 1

A **sequence in set** X is a function $a \colon \mathbb{N} \to X$.

As we studied in previous lectures, the basic vocabulary of natural numbers is that (i) there is a starting natural number, 0, and (ii) for each natural number n there is a next one, $n^{\curvearrowright}$. To discuss successor in the language of sets and functions, we stipulate that successor is a function $\mathsf{suc} \colon \mathbb{N} \to \mathbb{N}$ given by the rule $n \mapsto n^{\curvearrowright}$. So $\mathbb{N}$ is not just a set. It comes with functions $\mathbb{1} \xrightarrow{\hat{0}} \mathbb{N} \xleftarrow{\mathsf{suc}} \mathbb{N}$.

Suppose $\mathbb{1} \xrightarrow{\hat{b}} X \xleftarrow{r} X$ is a similar structure. Then we ought to be able to define a a sequence in X, so that $a_0 = b$, $a_1 = r(b)$, $a_2 = r(r(b))$, and so on. In general, a_k should be determined by starting with b and repeatedly applying r a total of k times.

Definition 2

A **simple recurrence** is a set with functions $\mathbb{1} \xrightarrow{\hat{b}} X \xleftarrow{r} X$. We will say the two functions $\hat{b}$ and r form a **recurrence on** X.

A **natural number set** is a simple recurrence $\mathbb{1} \xrightarrow{\hat{z}} N \xleftarrow{s} N$ so that for any other simple recurrence $\mathbb{1} \xrightarrow{\hat{b}} X \xleftarrow{r} X$, there is exactly one function $N \xrightarrow{f} X$ so that $f \circ \hat{z} = \hat{b}$ and

Definition 2 (cont.)

$f \circ s = r \circ f$.

The principle we are interested in here is that simple recurrences determine sequences.

Principle 3: Natural Numbers for a Set

The collection of natural numbers is a set, denoted by $\mathbb{N}$. Moreover, $\mathbb{1} \xrightarrow{\hat{0}} \mathbb{N} \xleftarrow{\mathsf{suc}} \mathbb{N}$. makes $\mathbb{N}$ a natural numbers set.

From a simple recurrence $\mathbb{1} \xrightarrow{\hat{b}} X \xleftarrow{r} X$, the corresponding unique sequence in X may be denoted by $\mathsf{s\text{-}rec}[b, r]$. So $\mathbb{N} \xrightarrow{\mathsf{s\text{-}rec}[b,r]} X$ is characterized by

$$\begin{aligned} \mathsf{s\text{-}rec}[b, r]_0 &= b \\ \mathsf{s\text{-}rec}[b, r]_{n^\frown} &= r(\mathsf{s\text{-}rec}[b, r]_n) \end{aligned}$$

So every simple recurrence on a set X determines a sequence in X. On the other hand, it is not the case that every sequence is determined by a simple recurrence. Take for example, the sequence $0, 1, 0, 2, 0, 3, \ldots$. This can not be defined (at least not directly) by giving an initial entry and specifying successive entries based only on the predecessors. After all, the entries 1, 2, 3 and so on all have the same preceding entry.

13.2 Primitive Recursion

Evidently, addition, multiplication, factorial, and other familiar functions should be definable using Principle **??**. But there are problems to overcome: Addition is not a sequence, at least not in an obvious way. And factorial is not obviously definable by a simple recurrence because we would need a function $r \colon \mathbb{N} \to \mathbb{N}$ so that $n^\frown! = r(n!)$ for all n. If, in place of r, we could use a function that depends on n as well as on $n!$, we could define factorial recursively the usual way.

Putting things together, we consider a scheme that generalizes simple recursion to permit (i) dependence on a parameter not directly involved in the recursion, and (ii) dependence on n at each stage of the recursion.

Definition 4

A **primitive recurrence in** A **(with parameters in** C**)** consists of two functions $C \xrightarrow{b} A \xleftarrow{r} \mathbb{N} \times C \times A$.

A **parametric sequence in** A **with parameters in** C is a function $a\colon C \times \mathbb{N} \to A$. For a parametric sequence, we may write $a_{c,n}$ instead of $a(c,n)$.

Theorem 5

For any primitive recurrence $C \xrightarrow{b} A \xleftarrow{r} C \times \mathbb{N} \times A$, there is a unique function $\mathsf{p\text{-}rec}[b,r]\colon C \times \mathbb{N} \to A$ satisfying:

$$\begin{aligned} \mathsf{p\text{-}rec}[b,r]_{c,0} &= b(c) \\ \mathsf{p\text{-}rec}[b,r]_{c,k^\frown} &= r(c,k,\mathsf{p\text{-}rec}[b,r]_{c,k}) \end{aligned}$$

Proof: The proof of this requires additional ideas that are implict in the way subset classification works. We have all the principles we need, but a proof now would involve quite a long development. So we only sketch the rough idea.

We use simple recursion to define a sequence of "approximations" of the desired function. First, let $D_n = C \times \{0, \dots, n\}$ for each $n \in \mathbb{N}$. Then define $f_n\colon D_n \to A$ for each $n \in \mathbb{N}$ by

$$\begin{aligned} f_0(c,0) &= b(c) \\ f_{k^\frown}(c,i) &= f_k(c,i) && \text{if } i \leq k \\ f_{k^\frown}(c,k^\frown) &= r(c,k,f_k(c,k))\text{otherwise.} \end{aligned}$$

Then define $f\colon C \times \mathbb{N} \to A$ by $f(c,n) = f_n(c,n)$. It can be checked that f satisfies the desired equations, and that no other function does.

The technical part of the proof is mainly concerned with ensuring that the sequence of functions f_n is indeed definable and that f can be defined from those. □

Example 6

The "predecessor" function is defined by the scheme $\mathsf{pred}(0) = 0$ and $\mathsf{pred}(n^\frown) = n$. The primitive recurrence $\mathbb{1} \xrightarrow{\hat{0}} \mathbb{N} \xleftarrow{\pi_1} \mathbb{1} \times \mathbb{N} \times \mathbb{N}$ where π_1 is the projection $(x,y,z) \mapsto y$ determines a function $g\colon \mathbb{1} \times \mathbb{N} \to \mathbb{N}$ given by $g(\bullet, 0) = 0$ and $g(\bullet, n^\frown) = \pi_2(\bullet, n, g(\bullet, n)) = n$. Hence,

Example 6 (cont.)

we may define $\mathsf{pred}(n) = g(\bullet, n)$

Exercises for Lecture 13

1. Define addition $\mathbb{N} \times \mathbb{N} \xrightarrow{+} \mathbb{N}$ by primitive recursion. That is, find functions $\mathbb{N} \xrightarrow{b} \mathbb{N}$ and $\mathbb{N} \times \mathbb{N} \times \mathbb{N} \xrightarrow{r} \mathbb{N}$ so that

$$\begin{aligned} m + 0 &= b(m) \\ m + n^{\frown} &= r(m, n, m + n) \end{aligned}$$

   That way, $\mathsf{p\text{-}rec}[b, r]$ is addition.

2. Define multiplication $\mathbb{N} \times \mathbb{N} \xrightarrow{\cdot} \mathbb{N}$ by primitive recursion. You may use addition in defining the primitive recurrence.

3. Define the factorial function by primitive recursion. You may use multiplication in defining the primitive recurrence.

4. For a given function $f \colon \mathbb{N} \to \mathbb{N}$, find a primitive recurrence that defines the function $n \mapsto \sum_{i=0}^{n-1} f(i)$. That is, result will be the sequence 0, $f(0)$, $f(0) + f(1)$, $f(0) + f(1) + f(2)$, $\ldots$.

5. The operation of *monus* $m \mathbin{\dot{-}} n$ is defined to be $m - n$ when $m \geq n$ and to be 0 otherwise. Define monus by primitive recursion.

6. Let $\mathbb{N} \xrightarrow{p} 2$ be a characteristic function. Show that bounded existential quantification is primitive recursive. That is, define the function $\exists_p^< \colon \mathbb{N} \to 2$ by $\exists_<^p(n) = \mathrm{T}$ iff there is at least one $k < n$ for which $p(k) = \mathrm{T}$. Show that this can be defined by a primitive recurrence. [Hint: $\exists_<^p(0) = \mathrm{F}$ because there are no natural numbers below 0. Now ask what value is $\exists_<^p(n^{\frown})$ in terms of p and $\exists_<^p(n)$.]

13.3 Primitive Recursion and “for” loops

It is a theorem (we will not prove here) that there is a technical sense in which primitive recursive functions exactly the functions implementable, say in Python, by nested “for” loops. To get a taste of what this means, let us pretend that $P \xrightarrow{b} X$ and $P \times \mathbb{N} \times X \xrightarrow{r} X$ are somehow defined by Python functions. Then the following code implements $\mathsf{p\text{-}rec}[b, r]$.

```
def p_rec($b \in X^P$ ,$r \in X^{P\times\mathbb{N}\times X}$)  $\in X^{P\times\mathbb{N}}$
def f(p ∈ P, n ∈ ℕ): ∈ X
    x = b(p)
    for i in range(n):
        x = r(p, i, x)
    return x
return f
```

Conversely, suppose a function in Python is defined using only natural number variables and the following parts of Python:

- Assignment statements
- Increment statements: "n+=1"
- loops of the form "**for** i **in** range(n): ..."
- evaluations of functions that are defined similarly

The theorem we allude to claims that such a function is definable by primitive recursion. The theorem is not especially difficult, but it does involve a lot of careful checking. The point is that primitive recursion allows us to define a lot of the functions that we expect to be able to program in a standard programming language.

A question arises, however. If "for" loops are sufficient to program any primitive recursive function, why bother with other more complicated loops? One answer is that programming languages are not merely for computing functions. They are used to implement lots of behavior that is not so easily cast in terms of sets and functions. Another answer, though, is internal to set theory. Namely, there are computable functions on $\mathbb{N}$ which are not primitive recursive.

For each $n \in \mathbb{N}$, define the sets $P^n \subseteq \mathbb{N}^{\mathbb{N}^n}$ of n-*ary number theoretic primitive recursive functions* to be the smallest sets satisfying:

- $\hat{0} \in P^0$
- $\mathsf{suc} \in P^1$
- for each $k < n$, $\pi^n_k \in P^n$ where $\pi^n_k \in \mathbb{N}^{\mathbb{N}^n}$ is defined by $\pi^n_k(x_0, \dots, x_{n-1}) = x_k$
- If $g \in P^n$ and for each $k < n$, $f_k \in P^m$, then $g \circ \langle f_0, \dots, f_{n-1} \rangle \in P^m$
- if $b \in P^m$ and $h \in P^{m+2}$, then $\mathsf{p\text{-}rec}[b, h] \in P^{m+1}$

Consider the following sequence $A_0, A_1, \dots$ in $\mathbb{N}^{\mathbb{N}}$ defined by

$$A_0 \mathsf{suc}$$
$$A_{k^\frown} = \mathsf{p\text{-}rec}[\hat{A_k(1)}, A_k \circ \pi^2_1]$$

Putting this in terms of elements

$$\begin{aligned} A_0(m) &= m + 1 \\ A_{k^\frown}(0) &= A_k(1) \\ A_{k^\frown}(m^\frown) &= A_k(A_{k^\frown}(m)). \end{aligned}$$

Evidently, each individual function A_k is a number theoretic primitive recursive unary function. But the function $\mathsf{Ack}\colon \mathbb{N} \to \mathbb{N}$ defined by $\mathsf{Ack}(n) = A_n(n)$ is not. The proof is quite ingeneous. Roughly, one shows that Ack grows faster than any number theoretic primitive recursive function.

To get an idea of how fast this function grows, $\mathsf{Ack}(0) = 1$, $\mathsf{Ack}(1) = 3$, $\mathsf{Ack}(3) = 61$, $\mathsf{Ack}(4) = 2^{2^{65536}} - 3$ a number vastly larger than the number of electrons in the visible universe.

13.4 Lists

For a set X, the lists consisting of elements from X also form a set. The structure of this set is similar to $\mathbb{N}$.

Definition 7

For a set X, a **simple X-recurrrence** is a set with two functions $\mathbb{1} \xrightarrow{\hat{b}} A \xleftarrow{r} X \times A$.

For a set X, an **X-list set** is a simple X-recurrence $\mathbb{1} \xrightarrow{\hat{e}} L \xleftarrow{c} X \times L$ so that for any simple A-recurrence $\mathbb{1} \xrightarrow{\hat{b}} A \xleftarrow{r} X \times A$ there is exactly one function $h\colon L \to A$ so that

$$h \circ \hat{e} = \hat{b}$$
$$h \circ c = r \circ (\mathsf{id}_X \times h).$$

Principle 8: List(X) is a Set

The collection of lists with items drawn from X is a set, denoted by List[X]. The functions $\mathbb{1} \xrightarrow{\widehat{[]}} \mathsf{List}[X] \xleftarrow{:} X \times \mathsf{List}[X]$ constitute an X-list set, where the function $X \times \mathsf{List}[X] \xrightarrow{:} \mathsf{List}[X]$ is the function given by the rule $(x, L) \mapsto x : L$.

For $\mathbb{1} \xrightarrow{\hat{b}} A \xleftarrow{r} X \times A$, we reuse our notation for simle recursive functions on $\mathbb{N}$ and write $\mathsf{s\text{-}rec}[\hat{b}, r]$ for the unique function satisfying

$$\mathsf{s\text{-}rec}[\hat{b}, r]([\,]) = b$$
$$\mathsf{s\text{-}rec}[\hat{b}, r](x : L) = r(x, \mathsf{s\text{-}rec}[\hat{b}, r](L))$$

Example 9

$\mathbb{1} \xrightarrow{\hat{0}} \mathbb{N} \xleftarrow{+} \mathbb{N} \times \mathbb{N}$ determine a function $\mathsf{s\text{-}rec}[\hat{0}, +]$ from $\mathsf{List}(\mathbb{N})$ to $\mathbb{N}$. The function satifies $\mathsf{s\text{-}rec}[\hat{0}, +]([\,]) = 0$ and $\mathsf{s\text{-}rec}[\hat{0}, +](n : L) = n + \mathsf{s\text{-}rec}[\hat{0}, +](L)$. So $\mathsf{s\text{-}rec}[0, +]$ returns the sum

Example 9 (cont.)

of items in a list natural numbers. Earlier, we wrote this as $\sum L$. In other words, we *defined* $\Sigma = \mathsf{s\text{-}rec}[\hat{0}, +]$.

Like $\mathbb{N}$, $\mathsf{List}[A]$ supports a form of primitive recursion. The details can be worked out by the reader.

Exercises for Lecture 13

1. For a function $f \colon X \longrightarrow Y$, specify a simple X- recurrence that defines the function
$$\mathsf{map}[f] \colon \mathsf{List}[X] \longrightarrow \mathsf{List}[Y]$$
so that
$$\mathsf{map}[f]([a_0, a_1, \ldots, a_{n-1}]) = [f(a_0), f(a_1), \ldots, f(a_{n-1})]$$

2. For $\mathsf{map}[\cdot]$ as defined in the previous exercise, show that for any $f \colon X \longrightarrow Y$ and $g \colon Y \longrightarrow Z$, $\mathsf{map}[g \circ f] = \mathsf{map}[g] \circ \mathsf{map}[f]$.

3. Define concatenation in $\mathsf{List}[A]$ by a simple A recurrence. [Hint: I am asking for a function from $\mathsf{List}[A] \times \mathsf{List}[A]$ to $\mathsf{List}[A]$, but simple A recurrence alone will not do the job, because that can only define a function $\mathsf{List}[A] \longrightarrow X$ for some X. Do this instead: (i) specify a simple A-recurrence to define a function $c \colon \mathsf{List}[A] \longrightarrow \mathsf{List}[A]^{\mathsf{List}[A]}$ for which $c(M)(L)$ is the concatetation of L followed by M, (ii) define $L \star M$ to be $c(M)(L)$.

4. Show that $\mathbb{N}$ constitutes a $\mathbb{1}$-list set. That is, describe two functions $\mathbb{1} \xrightarrow{\hat{e}} \mathbb{N} \xleftarrow{c} \mathbb{1} \times \mathbb{N}$ so that for any two functions $\mathbb{1} \xrightarrow{\hat{b}} A \xleftarrow{r} \mathbb{1} \times A$ there is a unique function $\mathbb{N} \xrightarrow{f} A$ satisfying
$$f(e) = b$$
$$f(c(x, y)) = s(x, f(y))$$

5. Write a scheme for “primitive X-recursion”, specified by functions $P \xrightarrow{b} A \xleftarrow{r} P \times \mathsf{List}[X] \times A$. Write the equations involving b and r that should determine a unique function $P \times \mathsf{List}[X] \longrightarrow A$. You do not need to prove that your scheme actually defines a function.

LECTURE 14

Powersets

Goals

In this lecture we look at the structure of the collection of all subsets of a given set. This structure is intimately related to how we *describe* subsets. The result of our investigation leads to *logic* as a means of reasoning about descriptions.

For this lecture, we concentrate on a fixed set U, which we will think of as our "universe". For example, if we are currently mainly interested in natural numbers, U could be taken to be $\mathbb{N}$. If we are interested in poker, it could be Deck. If we are interested in functions on the reals, it could be $\mathbb{R}^{\mathbb{R}}$.

The exponential 2^U is the set of all characteristic maps on U. Each $k \in 2^U$ corresponds to a subset of U by $k^{-1}(\mathrm{T}) = \{x \in U \mid k(x) = \mathrm{T}\}$. Vice versa, a subset $A \subseteq U$ determines a characteristic function $\kappa_A : U \to 2$ by the rule

$$x \mapsto \begin{cases} \mathrm{T} & \text{if } x \in A \\ \mathrm{F} & \text{otherwise} \end{cases}$$

So 2^U is *representative* of the collection of all subsets of U. It is convenient also to suppose that the actual collection of subsets of a set U forms a set.

Principle 1: Powersets Exist

For any set U, the collection of subsets of U is a set, denoted by $\mathcal{P}(U)$ and called the **power set of** U. Moreover, there is a function $\ni_X : \mathcal{P}(U) \times U \to 2$ defined by

$$\ni_U(A, x) = \begin{cases} \mathrm{T}, & \text{if } x \in A \\ \mathrm{F} & \text{otherwise} \end{cases}$$

The function $\ni_U$ is an evaluation map, meaning that for any function $f : P \times U \to 2$, there is a unique function $f^\dagger : P \to \mathcal{P}(U)$ for which $f = \ni_U \circ (f^\dagger \times \mathsf{id}_U)$. The function $f^\dagger$ is given by the rule $p \mapsto \{x \in U \mid f(p, x) = \mathrm{T}\}$.

The function $\mathsf{curry}[\ni_U]\colon \mathcal{P}(U) \to 2^U$ is the unique function satisfying $\ni_U = \mathsf{appl}_{U,2} \circ (\mathsf{curry}[\ni_U] \times \mathsf{id}_U)$. In particular, this must send $A \in \mathcal{P}(U)$ to its characteristic function. So this is none other than the rule $A \mapsto \kappa_A$. Also, $\mathsf{appl}^{\dagger}_{U,2}\colon 2^U \to \mathcal{P}(U)$ is the unique function satisfying $\mathsf{appl}_{U,2} = \ni_U \circ (\mathsf{appl}^{\dagger}_{U,2} \times \mathsf{id}_U)$. This must send $k \in 2^U$ to the subset $k^{-}(\top)$. So it is given by the rule $k \mapsto k^{-}(\top)$. In this sense, elements of $\mathcal{P}(U)$ and elements of 2^U are interchangible.

For a function $f\colon U \to V$, $\ni_V \circ (\mathsf{id}_{\mathcal{P}(U)} \times f)$ is a function from $\mathcal{P}(V) \times U$ to 2. So there is a unique function from $\mathcal{P}(V)$ to $\mathcal{P}(U)$ determined by f according to the above princple. This is called *inverse image*. We introduce notation for this.

Definition 2

For $U \xrightarrow{f} V$, let $\mathcal{P}(V) \xrightarrow{f^{-}} \mathcal{P}(U)$ denote the unique function given by the rule $B \mapsto \{x \in X \mid f(x) \in B\}$. That is, f^{-} is the unique function for which $\ni_V \circ (\mathsf{id}_{\mathcal{P}(V)} \times f) = \ni_U \circ (f^{-} \times \mathsf{id}_U)$. In terms of elements, f^{-} satisfies

$$x \in f^{-}(B) \text{ if and only if } f(x) \in B$$

for every $B \subseteq Y$.

This notation clashes slightly with our earlier definition of inverse image of an element, $f^{-}(b) = \{x \in U \mid f(x) = b\}$. But this is harmless because $f^{-}(b)$ in the earlier usage is the same as $f^{-}(\{b\})$ in the new usage.

Exercises for Lecture 14

1. For $f\colon \mathbb{N} \to \mathbb{N}$ defined by $f(n) = x^2$, what is $f^{-}(\{2, 3, 4, 5, 6, 7, 8, 9\})$?
2. For $\sin\colon \mathbb{R} \to \mathbb{R}$, what is $\sin^{-}(\{-1, 1\})$?
3. Show that for any two functions $f\colon X \to Y$ and $g\colon Y \to Z$, $(g \circ f)^{-} = f^{-} \circ g^{-}$.

14.1 Finitary Structure of $\mathcal{P}(U)$

Suppose A and B are subsets of U. Then it makes sense to consider the elements that A and B have in common. For example, for $A \subseteq \mathbb{N}$ being the set of even natural numbers and $B \subseteq \mathbb{N}$ the set of perfect squares, we might want to concentrate on the set of even, perfect squares, another subset of $\mathbb{N}$. In general, for $A \subseteq U$ and $B \subseteq U$, the elements in common constitute another subset of U. This is called the **intersection** and is denoted by $A \cap B$.

Likewise, we might consider merging A and B into a single set (in our example, the set of numbers that are either even or perfect squares). This is called the **union**.

Intersection is defined by "and", union is defined by "or". That is, $x \in A \cap B$ if and only if $x \in A$ *and* $x \in B$; $x \in A \cup B$ if and only if $x \in A$ *or* $x \in B$.

Related to intersection, there is a largest set C so that $C \cap A \subseteq B$. This is called the *residual*, and is denoted by $A \Rightarrow B$. This is defined by "implies". That is, $x \in A \Rightarrow B$ if it is the case that $x \in A$ implies $x \in B$. Flipping things the other way, there is also a smallest set C so that $A \subseteq B \cup C$. This is called the **set difference** and is denoted by $A \setminus B$. This is defined by "but not". So $x \in A \setminus B$ if and only if $x \in A$ but not $x \in B$.

These operations on subsets of U are closely related to the logic of propositions. Imagine that U consists of a "universe" of possible worlds. Then subsets of U are collections of worlds where certain things are true. For example, perhaps $P \subseteq U$ is the set of worlds in which pigs fly; $K \subseteq X$ is the set of worlds in which kittens smoke cigars. So $P \cap K$ is the set of worlds in which pigs fly *and* kittens smoke cigars. Likewise, $P \cup K$ is the set of worlds in which *Either* pigs fly *or* kittens smoke cigars. And $P \Rightarrow K$ is the set of worlds in which it is true that if pigs fly, *then* kittens smoke cigars. Finally, $P \setminus K$ is the set of worlds in which pigs fly, but kittens don't smoke cigars.

Understanding the interaction of $\cap$, $\cup$ and $\Rightarrow$ is closely related to the logic of "and", "or" and "implies". If we add a sentence "False" that is never true and another one "True" that is always true, then the logic of "and", "or" and "implies" form what is called a *Heyting algebra*. The other operation $\setminus$ is less commonly studied (perhaps because most languages to not have a single word meaning "but not").

In fact, "implies" interacts with "False" in a useful way. It turns out that "P implies False" is essentially the same as saying "P is not true". And if "P is not true" is not true, then "P" must be true. This oservation indicates that the Heyting algebra of subsets is actually a **Boolean algebra**.

The operation of set difference is not as familiar in a logical setting. It corresponds to"but not", so $P \setminus K$ is the set of worlds in which pigs fly, but kittens do not smoke cigars.

Lemma 3

In the following, let U be a set, and $A, B \subseteq U$ be subsets.

- There is a subset of U, denoted by $A \cap B$, so that for every $C \subseteq U$, $C \subseteq A \cap B$ if and only if $C \subseteq A$ and $C \subseteq B$.
- There is a subset of U, denoted by $A \cup B$, so that for every $C \subseteq U$, $A \cup B \subseteq C$ if and only if $A \subseteq C$ and $B \subseteq C$.
- There is a subset of U, denoted by $A \Rightarrow B$, so that for every $C \subseteq U$, $C \subseteq A \Rightarrow B$ if and only if $C \cap A \subseteq B$.
- There is a subset of U, denoted by $A \setminus B$, so that for every $C \subseteq U$, $A \setminus B \subseteq C$ if and only if $A \subseteq B \cup C$.

Proof: A and B are determined by characteristic functions $\kappa_A \colon X \to 2$ and $\kappa_B \colon X \to 2$. So $\rangle \kappa_A, \kappa_B \langle$ is a function from X to 2×2. If we compose with a function $h \colon 2 \times 2 \to 2$, we have

Lemma 3 (cont.)

another characteristic function on X. So this determines another subset of X. So all of the above constructions amount to defining suitable functions $2 \times 2 \longrightarrow 2$.

The four functions corresponding to the subset operations can be given by tables. In these tables, we read the first argument on the left, and the second argument on the top.

$\wedge$	F	T
F	F	F
T	F	T

$\vee$	F	T
F	F	T
T	T	T

$\longrightarrow$	F	T
F	T	T
T	F	T

$-$	F	T
F	F	F
T	T	F

So for each $h \in \{\wedge, \vee, \longrightarrow, -\}$, there is a subset $(h \circ \langle \kappa_A, \kappa_B \rangle)^{-}(\mathrm{T})$.

It is routine to check that for $h = \wedge$, the result is $A \cap B$; for $h = \vee$, the result is $A \cup B$; for $h = \longrightarrow$, the result is $A \Rightarrow B$; and for $h = -$, the result is $A \setminus B$. □

This lemma, together with the fact that $\emptyset$ is the smallest element of $\mathcal{P}(X)$ and X is the largest, can be summarized by saying that $\cap$, $\cup$, $\Rightarrow$, $\emptyset$ and X make $\mathcal{P}(X)$ into what is known as a **Heyting algebra**. We spell out the axioms for Heyting algebras in the next section, but generally speaking these are the structures that correspond to a sort of minimal version of propositional logic in which "and", "or", "implies", "true" and "false" interact in natural ways. Likewise, "or", "and", "but not", "false" and "true" interact in natural ways to determine a co-Heyting algebra.

Additionally, "implies" and "false" interact in a stronger way, as do "but not" and "true". In particular, $\mathcal{P}(U)$ is a *Boolean algebra*. Let us abbreviate $A \Rightarrow \emptyset$ by writing A^* (read this informally as "not A). Then the Law of Double Negation which characterizes Boolean algebras, asserts that double negations do not change anything: $A^{**} = A$.

Lemma 4

In $\mathcal{P}(X)$, $A^{**} = A$.

Proof: Calculating the members of A^*, it is easy to check that for every element $x \in X$, either $x \in A$ or $x \in A^*$, but not both. So $x \in A^{**}$ if and only if $x \notin A^*$ if and only if $x \in A$.

Put differently, define the function $\neg\colon 2 \longrightarrow 2$ by $\neg\mathrm{T} = \mathrm{F}$ and $\neg\mathrm{F} = \mathrm{T}$. Then A^* is defined by $(\neg \circ \kappa_A)^{-}(\mathrm{T})$. Obviously, $\neg \circ \neg = \mathsf{id}_2$. As with the other operations, it is now routine to check that $A^{**} = A$. □

One can define the term "Boolean algebra" to mean "Heyting algebra that satisfies the Law of Double Negation." So $\mathcal{P}(X)$ is indeed a Boolean algebra.

Exercises for Lecture 14

1. An easy consequence of Double Negation is that $A \Rightarrow \emptyset = \mathbb{U} \setminus A$. Prove it.

14.2 Laws of Finitary Set Operations

As we mentioned, $\mathcal{P}(X)$ forms a Boolean algebra. This means that $\cup$, cup, $\Rightarrow$, $\emptyset$ and X satisfy various laws. We spell the most important out here.

Laws 5

For any set U and any subsets A, B and C:

	Semilattice Laws
Associativity	$A \cap (B \cap C) = (A \cap B) \cap C$
	$A \cup (B \cup C) = (A \cup B) \cup C$
Commutativity	$A \cap B = B \cap A$
	$A \cup B = B \cup A$
Idempotency	$A \cap A = A$
	$A \cup A = A$
	Lattice Laws
Absorptivity	$A = (A \cap B) \cup A$
	$A = (A \cup B) \cap A)$
Ordering	$A = B \cap A$ if and only if $A \cup B = A$
	Bounded Lattice Laws
Identity	$A \subseteq A \cap U$
	$A \cup \emptyset \subseteq A$
	Heyting Algebra Law
Residuation	$A \cap B \subseteq C$ if and only if $A \subseteq B \Rightarrow C$
	co-Heyting Algebra Law
Co-Residuation	$A \setminus B \subseteq C$ if and only if $A \subseteq B \cup C$
	Boolean Algebra Law
Double Negation	$A^{**} \subseteq A$
	Distributive Lattice Laws
Distributivity	$A \cap (B \cup C) \subseteq (A \cap B) \cup (A \cap C)$
	$(A \cup B) \cap (A \cup C) \subseteq A \cup (B \cap C)$
	Other Boolean Laws
de Morgan's Laws	$(A \cap B)^* = A^* \cup B^*$
	$(A \cup B)^* = A^* \cap B^*$

Several remarks are in order.

- The Semilattice Laws describe how $\cap$ and $\cup$ behave without any interaction between the two. In fact, any binary operation that is associative, commutative and idempotent is called a **semilattice operation**.
- The Lattice Laws describe how $\cap$ and $\cup$ interact. The two Absorption Laws together are equivalent to the Ordering Law. For suppose $A = A \cup (A \cap B)$ holds for all A and B. Now suppose $X = Y \cap X$. Then $Y \cup X = Y \cup (X \cap Y) = Y$. Conversely, suppose $A = B \cap A$ implies $B \cup A = B$ for all A and B. Then $(X \cap Y) = (X \cap Y) \cap X$. So $X \cup (X \cap Y) = X$.
- The remaining laws are stated in terms of $\subseteq$ instead of equality. Becaause of the Ordering Laws, $A \subseteq B$ is equivalent to $A = B \cap A$ and also equivalent to $B = A \cup B$.
- The Bounded Lattice Laws indicate that U is the unit element for $\cap$ and $\emptyset$ is the unit element for $\cup$. It follows that $\emptyset$ is the smallest element of $\mathcal{P}(U)$ and U is the largest.
- The Residuation and Co-residuation Laws show that $A \Rightarrow B$ and $A \setminus B$ are defined as *duals* of one another.
- In the Double Negation Law recall that A^* is defined to be $A \Rightarrow \emptyset$. Since $A \cap \emptyset \subseteq A \Rightarrow \emptyset$, it follows from Residuation that $A \subseteq A^{**}$. So in fact, $A = A^* *$.
- If we defined $A^\dagger = U \setminus A$, then in any co-Heyting algebra, $A^{\dagger\dagger} \subseteq A$. In a Boolean algebra $A^\dagger = A^*$.
- With respect to Distributivity, in any lattice, the opposite inclusions hold: $A \cap B \subseteq A \cap (B \cup C)$ and $A \cap B \subseteq A \cap (B \cup C)$, so $(A \cap B) \cup (A \cap C) \subseteq A \cap (B \cup C)$. Similarly, for the opposite inclusion for the second Distributive Law.
- The Distributivity laws are equivalent to each other. Suppose $A \cap (B \cup C) \subseteq (A \cap B) \cup (A \cap C)$ is true for all A, B and C. Then $(X \cup Y) \cap (X \cup Z) \subseteq ((X \cup Y) \cap X) \cup ((X \cup Y) \cap Z) = X \cup ((X \cup Y) \cap Z) \subseteq X \cup ((X \cap Z) \cup (Y \cap Z)) = X \cup (Y \cap Z)$
- The Heyting (or co-Heyting) Law implies Distributivity. Obviously, $A \cap B \subseteq (A \cap B) \cup (A \cap C)$, so $B \subseteq A \Rightarrow ((A \cap B) \cup (A \cap C))$. Likewise $C \subseteq A \Rightarrow ((A \cap B) \cup (A \cap C))$. So $B \cup C \subseteq A \Rightarrow ((A \cap B) \cup (A \cap C))$. And again using Residuation, $A \cap (B \cup C) \subseteq (A \cap B) \cup (A \cap C)$.

Exercises for Lecture 14

In the following, assume that U is a set, and all other sets are subsets of U.

1. Prove, using only the Semilattice Laws, the Absorption Laws and the Bounded Lattice Laws, that $A \cap \emptyset = \emptyset$. Likewise, show that $A \cup U = U$.
2. Prove, using only the Semilattice and Lattice Laws, that the two Distribution Laws are equivalent. That is, show that if $A \cap (B \cup C) \subseteq (A \cap B) \cup (A \cap C)$ holds for all A, B, C, then so does $(A \cup B) \cap (A \cup C) \subseteq A \cup (B \cap C)$. Hint: Let $X = A \cup B$, $Y = A$ and $Z = C$.

Exercises for Lecture 14 (cont.)

So $(A \cup B) \cap (A \cup C) = X \cap (Y \cup Z)$. Now use the first distributivity law, followed by absorption, then a second use of distributivity, and association and finally a second use of absorption.

3. Prove, using only the Semilattice, Lattice and Heyting Algebra Laws, the Distributivity Law (either one).

4. Prove, using only the Semilattice, Lattice Laws, Heyting Algebra and Boolean Algebra Laws, that the two de Morgan's Laws hold.

5. Prove that $A \Rightarrow B = A^* \cup B$, using any method.

6. Prove that $A \setminus B = A \cap B^*$, using any method.

14.3 Binary Relations

In many situations, we wish to consider how elements of a set are related to elements of another. For example, if S is set modelling the students at Chapman and M is a set modelling majors the university offers, then "majors in" is a relation between elements of S and elements of M. Perhaps "Jethro majors in Phrenology" is true, while "Aczel majors in Flim-Flam Studies" is not. For a more mathematical example, "is less than" is a relation between natural numbers and natural numbers. So for example "5 is less than 6" is true while "9 is less than 6" is not.

Notice that "majors in" can not be modelled obviously by a function because some students may not have a declared major and some students may actually be double majors. So we need a new idea. A binary relation between X and Y determines, for each $x \in X$ and each $y \in Y$, whether or not x is in the relation to y. There are several equivalent ways to model the general notion. But traditionally, the following is usually taken to be the official version.

Definition 6

A **relation between** X **and** Y is a subset $R \subseteq X \times Y$. We can write $x \mathrel{R} y$ instead of $(x, y) \in R$ to mimic familiar examples.

A **relation on** X is a relation between X and X.

Here are two common equivalent formulations.

- $R \subseteq X \times Y$ determines a characteristic function $\kappa_R \colon X \times Y \to 2$ so that $R = \{(x, y) \in X \times Y \mid \kappa_R(x, y) = \top\}$.

- R determines a function $R[-] \colon X \to \mathcal{P}(Y)$ so that $R = \{(x, y) \in X \times Y \mid y \in R[x]\}$.

The ability to move between these can be helpful. So it is worth practicing.

Exercises for Lecture 14

Let $r\colon \mathbb{N}\times\mathbb{N}\to \mathbb{2}$ be the function defined by

$$r(m,n)=\begin{cases}\mathrm{T} & \text{if } m\geq n^2\\ \mathrm{F} & \text{otherwise.}\end{cases}$$

Consider the relation $R = r^{-}(\mathrm{T})$. Explain your answers for the following.

1. Is it the case that $3 \mathrel{R} 2$?

2. Is it the case that $9 \mathrel{R} 3$?

3. What is $R[5]$?

4. What is $\mathsf{curry}[r]$?

We will discuss the structure of relations in more detail in Lecture **??**.

14.4 Quantifiers and Completeness

Consider how we might try to define the set of perfect square natural numbers. These are the natural numbers of the form m^2 for some natural number m. So the first few are $0, 1, 4, 9, 16$ and so on. We would be right to define this set by $\{n \in \mathbb{N} \mid \text{for some } m \in \mathbb{N}, m^2 = n\}$. To make sense of this, we need to understand what "for some $m \in \mathbb{N}, \ldots$" means formally. Just as we abbreviated "and" with symbol $\wedge$ and "or" with $\vee$, we will abbreviate "for some $m \in \mathbb{N}, \ldots$" as $\exists m \in \mathbb{N}, \ldots$.

Lemma 7

For sets W and U and relation $R \subseteq W \times U$, there is a subset of U, denoted by $\{x \in U \mid \exists w \in W . w \mathrel{R} x\}$ so that for all $C \subseteq U$,

$$\{x \in U \mid \exists w \in W . w \mathrel{R} x\} \subseteq C \text{ if and only if } R \subseteq W \times C.$$

Also, there is a subset of U, denoted by $\{x \in U \mid \forall w \in W, w \mathrel{R} x\}$ so that for all $C \subseteq U$,

$$C \subseteq \{x \in U \mid \forall w \in W, w \mathrel{R} x\} \text{ if and only if } W \times C \subseteq R.$$

Lemma 7 (cont.)

Proof: The relation R determines a function r from U to $\mathcal{P}(W)$ by $x \mapsto \{w \in W \mid w \mathrel{R} x\}$. Then

$$\begin{aligned}\{x \in U \mid \exists w \in W, w \mathrel{R} x\} &= r^{-}(\mathcal{P}(W) \setminus \{\emptyset\})\\ \{x \in U \mid \forall w \in W, w \mathrel{R} x\} &= r^{-}(\{W\})\end{aligned}$$

Proving that these sets satisfy the desired conditions is technical, but routine.

Concretely, $\{x \in U \mid \exists w \in W, w \mathrel{R} x\}$ consists of those $x \in U$ so that $w \mathrel{R} x$ for some $w \in W$; $\{x \in U \mid \forall w \in W, w \mathrel{R} x\}$ consists of those $x \in U$ so that $w \mathrel{R} x$ for all $w \in W$. This justifies our notation: $\exists w \in W, w \mathrel{R} x$ is read as "there *exists* $w \in W$ so that $w \mathrel{R} x$;" $\forall w \in W, w \mathrel{R} x$" is read as "for *all* $w \in W$, $w \mathrel{R} x$." □

Expressions using ∃ and ∀ can be nested, so for example suppose we wish to define the set of all functions $\mathbb{R} \to \mathbb{R}$ that are continuous at a. We can write

$$\{f \in \mathbb{R}^{\mathbb{R}} \mid \forall \epsilon \in \mathbb{R}^{+} \exists \delta \in \mathbb{R}^{+}, B(a, \delta) \subseteq f^{-}(B(f(a), \epsilon))\}$$

where $\mathbb{R}^{+} = \{x \in \mathbb{R} \mid 0 < x\}$ – the set of positive real numbers – and $B(b, \gamma) = \{x \in \mathbb{R} \mid b - \gamma < x < b + \gamma\}$ – the set of real numbers in the interval $(b - \gamma, b + \gamma)$. So a function f belong to this set if and only if for every $\epsilon > 0$ there is a $\delta > 0$ so that for all $x \in \mathbb{R}$, if $a - \delta < x < a + \delta$ then $f(a) - \epsilon < f(x) < f(a) + \epsilon$. The reader who is familiar with Calculus will recognize that this is the precise definition of *continuity at* a.

We can use ∃ and ∀ to generalize union and intersection to arbitrary collections.

Definition 8

Let $A\colon I \to \mathcal{P}(U)$ be a function into the powerset of U. We write A_i instead of $A(i)$ to emphasize that each A_i is a set. Then define

$$\bigcup_{i \in I} A_i = \{x \in X \mid \exists i \in I. x \in A_i\}$$

According to Lemma **??**, $\bigcup_{i \in I} A_i$ is again a subset of U, generalizing $\cup$ to the union of arbitrary families of subsets of U, rather than just two. In this sense, $\mathcal{P}(U)$ is **complete**. That is, the union of any family of subsets of U exists. This justifes saying that $\mathcal{P}(U)$ is a **complete** Boolean algebra.

Exercises for Lecture 14

In the following, consider the family $A\colon I \to \mathcal{P}(U)$.

Exercises for Lecture 14 (cont.)

1. Show that $\bigcup_{i \in I} A_i \subseteq C$ if and only if $A_k \subseteq C$ holds every $k \in I$.
2. Define $\bigcap_{i \in I} A_i$ in analogy with $\bigcup_{i \in I} A_i$.
3. For $I = \emptyset$, what is $\bigcup_{i \in I} A_i$?
4. For $I = \emptyset$, what is $\bigcap_{i \in I} A_i$?

14.5 Atomicity

The complete Boolean algebras $\mathcal{P}(U)$ have one more feature that characterizes powersets. They are **atomic**. This means, roughly, that all subsets are built from the simplest ones.

An **atom** of a Boolean algebra is an element with the property that there is nothing strictly between the smallest element and it. A singleton subset $\{x\} \subseteq U$ is an atom of $\mathcal{P}(U)$ because there are no other subsets lying between $\emptyset$ and $\{x\}$

Lemma 9

For any set U, the rule $x \mapsto \{x\}$ determines a function from U to $\mathcal{P}(U)$.

Proof: Recall that the *diagonal* relation on U is $\Delta_U = \{(x, y) \in U \times U \mid x = y\}$. Let $s\colon U \to \mathcal{P}(U)$ be the unique function for which $\in_U \circ (s \circ \mathsf{id}_U) = \kappa_{\Delta_U}$. In other words $\ni_U(s(x), y) = \top$ if and only if $x = y$. Since $\ni_U(s(x), y) = \top$ if and only if $y \in s(x)$, it is the case that $y \in s(x)$ if and only if $x = y$. So $s(x) = \{x\}$. □

Now every subset of U is obtained as a union of singletons: $A = \bigcup_{x \in A}\{x\}$. For a complete Boolean algebra, this is what is meant by saying that $\mathcal{P}(U)$ is a complete **atomic** Boolean algebra. Although we do not investigate this here, any complete atomic Boolean algebra has the same structure as $\mathcal{P}(U)$ for some U.

The structure of $\mathcal{P}(U)$ is even preserved by inverse images.

Lemma 10

For any function $f\colon X \to Y$, any $B\colon I \to \mathcal{P}(Y)$, it is the case that $f^{-}(\bigcup_{i \in I} B_i) = \bigcup_{i \in I} f^{-1}(B_i)$ and $f^{-1}(\bigcap_{i \in I} B_i) = \bigcap_{i \in I} f^{-1}(B_i)$.

Lemma 10 (cont.)

Proof: $x \in f^{-}(\bigcup_{i \in I} B_i)$ if and only if $f(x) \in \bigcup_{i \in I} B_i$ if and only if $f(x) \in B_k$ for some $k \in I$ if and only if $x \in f^{-}(B_k)$ for some $k \in I$ if and only if $x \in \bigcup_{i \in I} f^{-}(B_i)$. The proof for $\bigcap$ is similar with “for some …” replaced by “for all …". □

Exercises for Lecture 14

1. Write out $\mathcal{P}(\{a, b, c\})$
2. Write out $\mathcal{P}(\emptyset)$
3. Is it the case that $\emptyset \in \mathcal{P}(A)$ for any set A? Explain.
4. Write out $\mathcal{P}(2 \times 2)$ and $\mathcal{P}(\mathcal{P}(2))$. Pay attention to writing them in a systematic way, so that it is clear you have actually listed everything.
5. I claim that $\mathcal{P}(\emptyset)$ is a terminal set (Definition 3). Justify the claim.
6. I claim that $\mathcal{P}(\emptyset) \in \mathcal{P}(\mathcal{P}(\emptyset))$ is a subset classifier (Definition 11). Justify the claim.

14.6 Forward images

For a function $f \colon A \to B$ the function $f^{-} \colon \mathcal{P}(C) \to \mathcal{P}(B)$ has what is known as an *upper adjoint*.

Definition 11

For $f \colon X \to Y$, define $f^{+} \colon \mathcal{P}(X) \to \mathcal{P}(Y)$ by the rule $A \mapsto \{y \in B \mid \exists x \in A . f(x) = y\}$. The subset $f^{+}(A) \subseteq Y$ is called the **forward image of** A **with respect to** f.

The important fact about f^{+} is that is related to f^{-}.

Lemma 12

For any $A \subseteq X$ and $B \subseteq Y$, $f^{+}(A) \subseteq B$ if and only if $A \subseteq f^{-}(B)$.

Proof: Suppose $f^{+}(A) \subseteq B$. For $x \in A$, $f(x) \in f^{+}(A)$. So $f(x) \in B$. By definition, this means $x \in f^{-}(B)$. This shows that $A \subseteq f^{-}(B)$. Conversely, suppose $A \subseteq f^{-}(B)$. For $y \in f^{+}(A)$, there

Lemma 12 (cont.)

is some $x \in A$ so that $f(x) = y$. So there is some $x \in f^{-}(B)$ so that $f(x) = y$. Thus $y = f(x) \in B$. This shows that $f^{+}(A) \subseteq B$ □

There is also a lower adjoint of f^{-}, but as it is less commonly used, we do not investigate it here.

The important features of f^{+} are easily checked. First, f^{+} preserves atoms. That is, $f^{+}(\{x\}) = \{f(x)\}$. Second, f^{+} preserves all unions So $f^{+}(\bigcup_{i \in I} A_i) = \bigcup_{i \in I} f^{+}(A_i)$. Note that f^{+} does not necessarily preserve intersections.

Exercises for Lecture 14

1. Define a function $f \colon X \to Y$ and two subsets $A, B \subseteq X$ so that $f(X) \cap f(Y) \neq f(X \cap Y)$. Try to find the smallest example you can.

2. Prove that for any function $f \colon X \to Y$ and any $A \subseteq X$, the inclusion $A \subseteq f^{-}(f^{+}(A))$ holds.

3. Prove that for any function $f \colon X \to Y$ and any $B \subseteq Y$, the inclusion $f^{+}(f^{-}(B)) \subseteq B$ holds.

Lecture 15

Additional Constructions

Goals

Many other constructions can be built up using the principles we have discussed. In this lecture, we consider some of the most useful and most general.

15.1 Induction

We now can justify how inductive proofs on natural numbers work from a set theoretic perspective. Consider some $\mathsf{P} \subseteq \mathbb{N}$. Define $\mathsf{P}' = \{n \in \mathbb{N} \mid n^{\curvearrowright} \in \mathsf{P}\}$. Now suppose $0 \in \mathsf{P}$ and $k \in \mathsf{P}$ implies $k^{\curvearrowright} \in \mathsf{P}$. Then the simple recurrrence $\mathbb{1} \xrightarrow{\hat{0}} \mathsf{P} \xleftarrow{\succ} \mathsf{P}$ determines a function $\mathsf{s\text{-}rec}[0, \succ] \colon \mathbb{N} \to \mathsf{P}$. This is evidently one-to-one:

15.2 Unions

Although not strictly needed for most purposes, mathematicians generally agree that sets, no matter how they are related, can be merged into a single set with elements taken from the originals. That is, union ($\bigcup$) is meaningful for any set of sets. We take this as an additional principle.

Principle 1: Unions Exist

Suppose $\mathcal{X}$ is a set and each element of $\mathcal{X}$ is a set (so $\mathcal{X}$ is a set of sets). Then there is a set U so that $\mathcal{X} \subseteq \mathcal{P}(\mathsf{U})$ and for any set Z, if $\mathcal{X} \subseteq \mathcal{P}(\mathsf{Z})$, then $\mathsf{U} \subseteq \mathsf{Z}$.

Using this principle, the set U is actually the union $\bigcup_{X \in \mathcal{X}} X$. That is, $x \in \bigcup_{X \in \mathcal{X}} X$ holds if any only if $x \in X$ for some $X \in \mathcal{X}$.

15.3 Monomorphisms and Epimorphisms

As we know, the fact that addition is cancellative ($m + p = n + p$ implies $m = n$) is quite useful. Likewise, multiplication is amost always cancellative ($m \cdot p = n \cdot p$ implies $m = n$ provided $p \neq 0$). Even list concatenation is cancellative. But function composition is not. At least not generally. It will turn out though that certain functions can indeed by cancelled on the left and others can be cancelled on the right. The role of these special functions is akin to the role that non-0's play in multiplication.

Definition 2

A function $f\colon X \to Y$ is called

- a *monomorphism* if it is the case that for every $W \overset{h}{\underset{k}{\rightrightarrows}} X$, if $f \circ h = f \circ k$ then $h = k$;
- an *epimorphism* if it is the case that for every $Y \overset{h}{\underset{k}{\rightrightarrows}} Z$, if $h \circ f = k \circ f$ then $h = k$.

The terms *mono-* and *epi-* refer to behavior that we explore in this Lecture. For now, just commit them to memory: monomorphisms *cancel on the left* of $\circ$, epimorphisms *cancel on the right*.

Example 3

- For any subset $A \subseteq X$, the inclusion function $A \xrightarrow{i} X$ is a monomorphism.
- For sets X and Y, the projection functions $X \times Y \xrightarrow{\pi_0} X$ and $X \times Y \xrightarrow{\pi_1} Y$ are epimorphisms.
- The functions $X \xrightarrow{\Diamond_X} \mathbb{1}$ are almost always epimorphisms. The only except is when $X = \emptyset$.
- All pointers $\mathbb{1} \xrightarrow{\hat{a}} X$ for $a \in X$ are monomorphisms.
- For any set X, id_X is both a monomorphism and an epimorphism.

Exercises for Lecture 15

1. Show that if $X \xrightarrow{f} Y$ and $Y \xrightarrow{g} Z$ are epimorphisms, then so is $g \circ f$.
2. Show that if $X \xrightarrow{f} Y$ and $Y \xrightarrow{g} Z$ are monomorphisms, then so is $g \circ f$.

Exercises for Lecture 15 (cont.)

3. Show that if $X \xrightarrow{f} Y$ and $X \xrightarrow{Y}$ satisfy $g \circ f = \mathsf{id}_X$, then g is an epimorphism and f is a monomorphism.

The two lemmas spell out equivalent internal behavior.

Definition 4

A function $X \xrightarrow{f} Y$ is called

- *one-to-one* or *injective* or *an injection* if it is the case that for every $b \in Y$ there is at most one $a \in X$ so that $f(a) = b$;
- *onto* or *surjective* or *a surjection* if it is the case that for every $b \in Y$ there is at least one $a \in X$ so that $f(a) = b$.

The property of being injective can also be stated by saying that if $f(a_0) = f(a_1)$ then $a_0 = a_1$.

Lemma 5

A function is a monomorphism if and only if it is an injection.

Proof: Assume $X \xrightarrow{m} Y$ is a monomorphism. Consider $a_0, a_1 \in X$ so that $m(a_0) = m(a_1)$. Then $m \circ \hat{a}_0 = m \circ \hat{a}_1$. Since m is a monomorphism $\hat{a}_0 = \hat{a}_1$. Hence $a_0 = a_1$.

Assume $X \xrightarrow{m} Y$ is not a monomorphism. So there must be some set W and functions $W \overset{h}{\underset{k}{\rightrightarrows}} X$ so that $m \circ h = m \circ k$ but $h \neq k$. Since $h \neq k$, there must be an element $w \in W$ so that $h(w) \neq k(w)$. Let $a_0 = h(w)$ and $a_1 = k(w)$. These witness that m is not an injection because $m(a_0) = m(a_1)$ but $a_0 \neq a_1$. □

Lemma 6

A function is an epimorphism if and only if it is a surjection.

Proof: Assume $X \xrightarrow{e} Y$ is not onto. So there is some $b \in Y$ for which $e(x) \neq y$ is the case for

Lemma 6 (cont.)

all $x \in X$. Define $Y \underset{k}{\overset{h}{\rightrightarrows}} \{0, 1, 2\}$ by

$$h(y) = \begin{cases} 0 & \text{if } y = b \\ 2 & \text{otherwise} \end{cases}$$

$$k(y) = \begin{cases} 1 & \text{if } y = b \\ 2 & \text{otherwise.} \end{cases}$$

It is easy to verify that $h \circ e = k \circ e$, but obviosly $h \neq k$. So e is not an epimorphism.

Assume $X \xrightarrow{e} Y$ is onto. Suppose $Y \underset{k}{\overset{h}{\rightrightarrows}} Z$ satisfy $h \circ e = k \circ e$. For any $y \in Y$, there is some $x \in X$ so that $e(x) = y$. So $h(y) = h(e(x)) = k(e(x)) = k(y)$. By function extensionality, $h = k$. □

Subsets permit us to concentrate attention on elements with special properties. For an informal example, let $[-1, 1] \subseteq \mathbb{R}$ be the set of real numbers satisfying $-1 \leq x \leq 1$. Then the trigonometric function $\sin\colon \mathbb{R} \to \mathbb{R}$ has the property that $\sin(x) \in [-1, 1]$ is true for every $x \in \mathbb{R}$. So this suggests that sin can be "co-restricted" to the smaller codomain of $[-1, 1]$. In fact, Principle 12 guarantees this. We need, however, to generalize the idea to situations not covered by the principle as stated.

Consider the two functions depicted in Figure **??**. The function m is a monomorphism. The two functions are related by having the same codomain. Moreover, the elements of the codomain that are reached by the function f are also reached by m ($f(x) = m(d)$, $f(y) = m(a)$ and $f(z) = m(a)$). So it makes sense to define a function as indicated in 15.2. The situations in which this should be possible can be characterized internally via the behavior of functions on elements, or externally via their behavior with other functions.

Monomorphisms act like subsets, in the sense that if $A \subseteq X$, then the inclusion function $A \xrightarrow{i} X$ is a monomorphism. Our next principle stipulates that monomorphisms behave like subsets, also in the sense that they determine characteristic functions.

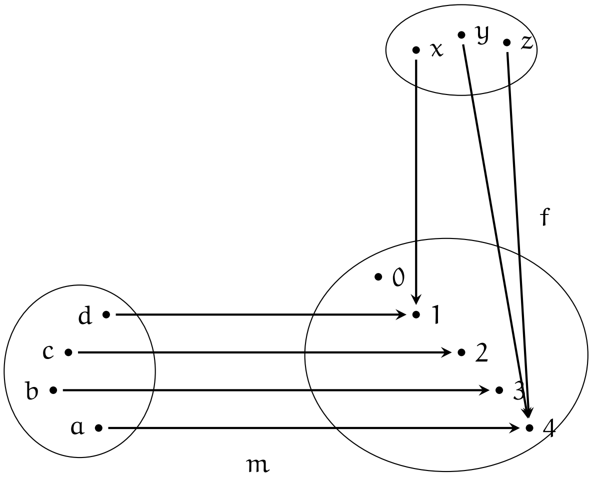


Figure 15.1: Function f and m with a common codomain

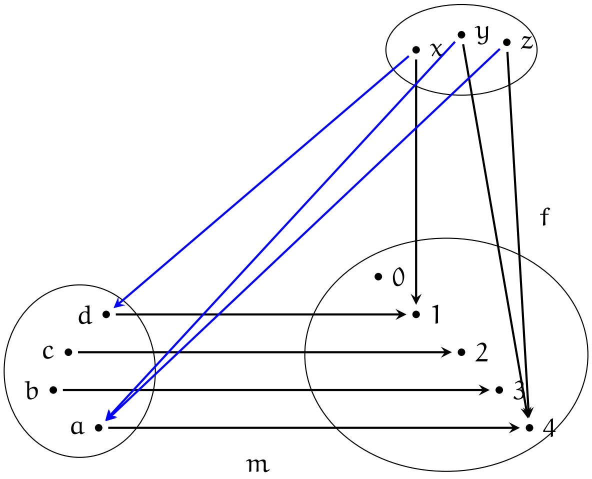


Figure 15.2: Function f *factors uniquely through* m

Principle 7: Monomorphisms have Characteristic Functions

For a monomorphism $X \xrightarrow{m} Y$ there is a unique function $Y \xrightarrow{\kappa_m} 2$ so that m is an equalizer for the equation $\kappa_m(y) = \top$. In other words, $\kappa_m \circ m = c_\top$ and if $W \xrightarrow{f} Y$ is a function satisfying $\kappa_m \circ f = c_\top$, then there is a unique function $W \xrightarrow{m\backslash f} X$ so that $f = (m\backslash f) \circ m$. We say that f *factors through* m *uniquely*.

For a monomorphism $X \xrightarrow{m} Y$, the characteristic function κ_m is defined by

$$\kappa_m(y) = \begin{cases} \top & \text{if } y = m(x) \text{ for some } x \in X \\ F & \text{otherwise.} \end{cases}$$

Principle 7 (cont.)

Suppose $W \xrightarrow{f} Y$ is a function satisfying $\kappa_m \circ f = c_T$. Then $(m\backslash f)$ satisfies $(m\backslash f)(w) = x$ if and only if $f(w) = m(x)$.

In Figure 15.1, the function f satisfies $\kappa_m \circ f = c_T$, so there is indeed a unique function $(m\backslash f)$ as indicated in Figure 15.2.

To detect that f is a solution of the equation $\kappa_m(y) = T$ is a simple matter. An internal characterization is this: for every $w \in W$, there is some $x \in X$ so that $f(w) = m(x)$. An external characterization is also useful.

Definition 8

For functions $X \xrightarrow{m} Y \xleftarrow{f} W$, say that f is m-**coinvariant** if it is that case that for all $Y \overset{h}{\underset{k}{\rightrightarrows}} Z$, $h \circ m = k \circ m$ implies $k \circ f = h \circ f$.

Lemma 9

For function $X \xrightarrow{m} Y \xleftarrow{f} W$ where m is a monomorphism, f is m-coinvariant if and only if $\kappa_m \circ f = c_\top$.

Proof: First, note that $\kappa_m \circ f = c_\top$ if and only if $\kappa_m(f(w)) = \top$ for every $w \in W$. But $\kappa_m(y) = \top$ if and only if there is some $x \in X$ so that $y = m(x)$. Therefore, $\kappa_m \circ f = \top$ if and only if for every $w \in W$ there exists some $x \in X$ for which $f(w) = m(x)$.

Assume $\kappa_m \circ f = c_\top$, and $h \circ m = k \circ m$. For any $w \in W$, there is some $x \in X$ so that $f(w) = m(x)$. Hence $h(f(w)) = h(m(x)) = k(m(x)) = k(f(w))$ for a suitable choice of x. Since this works for any w, $h \circ f = k \circ f$.

Assume there is some $w \in W$ for which $f(w) \neq m(x)$ for all $x \in X$. Define functions $Y \underset{k}{\overset{h}{\rightrightarrows}} \{0, 1, 2\}$ by

$$h(y) = \begin{cases} 0 & \text{if } y = f(w) \\ 2 & \text{otherwise} \end{cases}$$

$$k(y) = \begin{cases} 1 & \text{if } y = f(w) \\ 2 & \text{otherwise.} \end{cases}$$

It is easy to verify that $h \circ m = k \circ m$, but $h \circ f \neq k \circ k$. So f is not m-coinvariant. □

Lemma 10

For functions $W \xrightarrow{m} Y \xleftarrow{f} X$ where m is an monomorphism and f is m-coinvariant, there is a unique function $X \xrightarrow{m \backslash f} W$ so that $f = m \circ (m \backslash f)$.

Proof: There is nothing new to prove. Lemma **??** ensures if f is m-coinvariant, then $\kappa_m \circ f = c_\top$. Principle 7 ensures that $m \backslash f$ exists. □

Epimorphisms are dual to monomorphisms (they cancel on the left instead of on the right). The concept of m-coinvariance also has a useful dual.

Definition 11

For functions $W \xleftarrow{f} Y \xrightarrow{e} X$, say that f is e-coinvariant if it is that case that for all $Z \overset{h}{\underset{k}{\rightrightarrows}} Y$, $e \circ h = e \circ k$ implies $f \circ h = f \circ k$.

Lemma 12

For functions $W \xleftarrow{f} Y \xrightarrow{e} X$ where e is an epimorphism and f is e-invariant, there is a unique function $X \xrightarrow{f/e} W$ so that $f = (f/e) \circ e$.

Proof: A full proof is technical and not especially informative. The idea is that f/e can be defined by a "rule": $e(y) \mapsto f(y)$. Obviously, this does not look like a rule sending elements of X to elements of W, but in fact it is. This is because e is onto, each $x \in X$ takes the form $e(y)$ for some $y \in Y$. And because f is e-invariant, the choice of y does not matter. That is, $e(y_0) = x$ and $e(y_1) = x$ implies $f(y_0) = f(y_1)$. So the rule does indeed pick a single value $(f/e)(x)$ for each $x \in X$. □

Comparing Lemmas 10 and 12, we note that monomorphisms and epimorphisms play dual roles. Also a monomorphism behaves in most respects like the inclusion of a subset. So a question arises as to what should be the dual concept corresponding to subset. That is, what completes the formal analogy "Subsets are to monomorphisms as ___s are to epimorphisms".

Definition 13

For a set X, a *partition of* X is a subset $P \subseteq \mathcal{P}(X)$ so that

- for each $B \in P$, there is at least one $x \in X$ so that $x \in B$;
- for each $x \in X$, there is exactly one $B \in P$ so that $x \in B$.

The subsets $B \in P$ are called *blocks*.

Because each x is in exactly one block, we write $[x]_P$ for that block. This means that $x \mapsto [x]_P$ defines a function $X \xrightarrow{[-]} P$. Furthermore, since each $B \in P$ is inhabited ($x \in B$ for some x), the function $[-]_P$ is clearly onto. Suppose $e \colon X \to Y$ is an epimorphism. Then the collection $\{e^{-}(y) \mid y \in Y\}$ is a partition of X. That is, each $e^{-}(y)$ is non-empty because e is onto. Clearly, $x \in e^{-}(e(x))$. And if $x \in e^{-}(y)$, then $e(x) = y$, so x belongs only to one block. Let X/e denote the partition $\{e^{-}(y) \mid y \in Y\}$. Then $x \mapsto [x]_{X/e}$ simply picks out the set of all x' so that $e(x) = e(x')$. Putting these observations together, we see that every epimorphism e with domain X determines X/e, a partition of X. And every partition P of X determines an epimorphism $x \mapsto [x]_P$ of X onto P.

A monomorphism also determines characteristic function: $A \xrightarrow{m} X$ determines $X \xrightarrow{\kappa_m} 2$. What is the analogue of a characteristic function for epimorphisms? The answer is an *equivalence* relation.

For a function $f\colon X \to Y$ (in practice, usually we expect f to be an epimorphism), define $\equiv_f \subseteq X \times X$ by $x_0 \equiv_f x_1$ if and only if $f(x_0) = f(x_1)$. This relation has three characteristic properties:

- $\equiv_f$ is *reflexive*: $x \equiv_f x$ for all $x \in X$;
- $\equiv_f$ is *transitive*: if $x_0 \equiv_f x_1$ and $x_1 \equiv_f x_2$, then $x_0 \equiv_f x_1$; and
- $\equiv_F$ is *symmetric*: if $x_0 \equiv_f x_1$, then $x_1 \equiv_f x_0$.

Definition 14

A relation E on X is an *equivalence relation on* X if it is reflexive, transitive and symmetric.

Lemma 15

Suppose E is an equivalence relation on X. Then there is a set X/E and epimorphism $[-]\colon X \to X/E$ so that $E = \equiv_{[-]}$.

Proof: For each $x \in X$, let $[x]_E = \{x' \in X \mid x \mathrel{E} x'\}$. Since E is reflexive, $x \in [x]_E$ for each $x \in X$. Also, if $[x]_E \cap [x']_E \neq \emptyset$, then $x \mathrel{E} x'$ because E is transitive and symmetric. So the collection $\{[x]_E \mid x \in X\}$ is a partition of X. The function $x \mapsto [x]_E$ is an epimorphism by construction. And evidently, $[x]_E = [x']_E$ holds if and only if $x \mathrel{E} x'$. So $E = \equiv_{[-]}$. □

A function $f\colon X \to Y$ is invariant with respect to $[-]_E$ if it is the case that $x_0 \mathrel{E} x_1$ implies $f(x_0) = f(x_1)$. Any function with this property determines a function from X/E to Y defined by $[x]_E \mapsto f(x)$. This is defined on all blocks of X/E because each block is $[x]_E$ for some x. Moreover, if $[x_0]_E = [x_1]_E$, then $x_0 \mathrel{E} x_1$. So $f(x_0) = f(x_1)$. That is, our rule is unambiguous.

Example 16

We might wish to define a ‘clock’. Recall that $\mathbb{Z}$ denotes the set of all integers. On a clock 1 and 13 are the same because one hour after midnight and thirteen hours after midnight read the same on the clock. In fact, 1, 13, 25, and so on all read the same. So let is define the relation $\equiv_{12}$ on $\mathbb{Z}$ by saying $a \equiv_{12} b$ if and only if for some integer m, it is the case that

Example 16 (cont.)

$a + 12 \cdot m = b$. In other words, a and b differ by some even multiple of twelve.

Clearly, $a \equiv_{12} a$ is true for all integers a (take $m = 0$). And if $a \equiv_{12} b$ and $b \equiv_{12} c$, then $a + 12 \cdot m = b$ and $b + 12 \cdot n = c$ for some m and n. Hence $a + 12 \cdot (m + n) = c$. So $a \equiv_{12} c$. And finally, if $a \equiv_{12} b$, then $a + 12 \cdot m = b$, so $b + 12 \cdot -m = a$. So $b \equiv_{12} a$. This shows that $\equiv_{12}$ is an equivalence relation. Now, $\mathbb{Z}/\equiv_{12}$ consists of exactly twelve equivalence blocks: $[0]_{\equiv_{12}}, \ldots, [11]_{\equiv_{12}}$. Each block corresponds to an hour on the clock dial.

The quotient set $\mathbb{Z}/\equiv_{12}$ is usually denoted by $\mathbb{Z}_{12}$. Clearly, the same idea works for any positive integer *modulus* k in place of 12. We investigate this idea in depth in Lecture **??**, where applications in cryptography arise.

15.4 Co-products

Definition 17

For sets X and Y, a *co-table* is a set with a pair of functions $X \xrightarrow{f} C \xleftarrow{g} Y$.

A *co-product* is a co-table $X \xrightarrow{i} S \xleftarrow{j} Y$ so that for any co-table $X \xrightarrow{f} C \xleftarrow{g} Y$ there is exactly one function $c\colon S \to C$ so that $f = c \circ i$ and $g = c \circ j$.

Lemma 18

Any two sets X and Y have a co-product.

Proof: Define $X \uplus Y \subseteq \mathcal{P}(X) \times \mathcal{P}(Y)$ to consist only of pairs of the form $(\{x\}, \emptyset)$ for $x \in X$ and $(\emptyset, \{y\})$ for $y \in Y$.

Define the two functions $\mathsf{inj}_0\colon X \to X \uplus Y$ and $\mathsf{inj}_1\colon Y \to X \uplus Y$ by $\mathsf{inj}_0(x) = (\{x\}, \emptyset)$ and $\mathsf{inj}_1(y) = (\emptyset, \{y\})$. Now, for any pair of functions $X \xrightarrow{f} C \xleftarrow{g} Y$ define $[f, g]'\colon X \uplus Y \to \mathcal{P}(C)$ by the rule $(A, B) \mapsto f^{+}(A) \cup g^{+}(B)$. Since either $A = \emptyset$ and B is a singleton or A is a singleton and $B = \emptyset$, it is always the case that $[f, g]'(A, B)$ is a singleton. So it factors uniquely through the function sending $c \in C$ to $\{c\}$. That is, there is a unique function $[f, g]\colon X \uplus Y \to C$ so that $\{[f, g](A, B)\} = [f, g]'(A, B)$. Now it is easily checked that $[f, g] \circ \mathsf{inj}_0 = f$ and $[f, g] \circ \mathsf{inj}_1 = g$. And no other function has this property. □

The set $X \uplus Y$ is sometimes called the *disjoint union* of X and Y. It is a union, preceded by "marking" each $x \in X$ by putting it into a pair $(\{x\}, \emptyset)$ and marking each $y \in Y$ differently by putting it into a pair $(\emptyset, \{y\})$. Hence it is the union of disjoint copies of X and Y.

Exercises for Lecture 15

Let $A = \{a, b, c, d, e\}$, $B = \{w, x, y, z, a, b, c, \}$ and $C = \{2, 3\}$

1. Calculate $A \uplus B$
2. Calculate $A \uplus \emptyset$
3. Calculate $C \times (A \uplus B)$ and $(C \times A) \uplus (C \times B)$.

15.5 Quotients

Sometimes, elements of a set X will be classified into "like" kinds. For example, we might classify the natural numbers into even and odd. We might classify poker cards according to suit, ignoring the rank – or by rank, ignoring suit. Or, if C is set modelling a Discrete Mathematics class, we might classify the elements (students) according to their grade: A, B, etc. Situations like this are modelled by what is known as a *partion* of a set. If we also wish to think of all A students as being "equivalent", all B students as being "equivalent", we model this by what is known as an *equivalence relation*.

Definition 19

For a set X, a *partition of* X is a set $P \subseteq \mathcal{P}(X) \setminus \{\emptyset\}$ so that $\bigcup_{A \in P} A = X$ and for every $A, B \in P$, if $A \neq B$ then $A \cap B = \emptyset$. The sets in P are called *blocks*.

For a set X, an *equivalence relation on* X is a binary relation satisfying

- $x \mathrel{E} x$ for all $x \in X$,
- $x \mathrel{E} y$ and $y \mathrel{E} z$ implies $z \mathrel{E} z$ for all $x, y, z \in X$, and
- $x \mathrel{E} y$ imples $y \mathrel{E} x$

For a partition P define the relation $\equiv_P \subseteq X \times X$ by $x \equiv_P y$ if and only if for some $A \in P$, $x \in A$ and $y \in A$.

For an equivalence relation E and an element $x \in X$, let $[x]_E = \{y \in X \mid x \mathrel{E} y\}$. So $x \mapsto [x]_E$ defines a function from X to $\mathcal{P}(X)$. Let $X/E = \{[x]_E \mid x \in X\}$. That is, X/E is the range of the function $x \mapsto [x]_E$.

The two notions, partition and equivalence relation, are essentially the same in the sense that there is a natural way to pass from a partitions to an equivalence relation and vice versa.

Lemma 20

Let X be any set. Then

- For any partition P of X, the relation $\equiv_P$ is an equivalence relation on X;
- For any equivalence relation E on X, the collection $\mathcal{P}_E$ forms a partition of X;
- For any partition P of X, $P = X/\equiv_P$
- For any equivalence relation E on X, $E == \equiv_{X/E}$;
- the rule $x \mapsto [x]_E$ determines an onto function from X to X/E.

Proof: Exercise. □

Suppose $E \subseteq X \times X$ is an equivalance relation and $f\colon X \to Y$ is a function so that $x \; E \; x'$ implies $f(x) = f(x')$. Then we can define a function from X/E to Y by the "rule" $[x]_E \mapsto f(x)$. We can not call this a rule in the usual way because the left side is not a variable or a tuple of variables. But we can define a relation $F \subseteq (X/E) \times Y$ by stipulating that for $B \in X/E$ and $y \in Y$, $B \; F \; y$ if and only if $f(x) = y$ for some $x \in B$. This relation is total because every block $B \in X/E$ is non-empty. So there is some $x \in B$, and thus $B \; F \; f(x)$. The relation F is also deterministic because if $B \; F \; y$ and $B \; F \; y'$, then there is some $x \in B$ for which $f(x) = y$ and there is some $x' \in B$ for which $f(x') = y'$. But $x, x' \in B$ implies $x \; E \; x'$. Hence $f(x) = f(x')$.

Definition 21

Suppose $E \subseteq X \times X$ is an equivalence relation. A function $f\colon X \to Y$ is E-*invariant* if $x \; E \; x'$ implies $f(x) = f(x')$. For an E-invariant function $f\colon X \to Y$, let f/E denote the function from X/E to Y defined by $(f/E)([x]_E) = f(x)$.

Example 22

Let $E \subseteq \mathbb{R} \times \mathbb{R}$ be the relation xEy if and only if $x - y = k2\pi$ for some integer k. This is an equivalence relation: xEx is true because $x - x = 0\dot{2}\pi$. If $x - y = k2\pi$ then $y - x = -k2\pi$, so E is symmetric. And if $x - y = k2\pi$ and $y - z = j2\pi$, then $x - z = (x - y) + (y - z) = (k + j)2\pi$. So E is transitive. The functions sin and cos are E invariant because $\sin(x + k2pi) = \sin(x)$ and $\cos(x + k2\pi) = \cos(x)$ for any x and any k.

Exercises for Lecture 15

On the integers $\mathbb{Z}$, define a relation $\equiv_7$ by $i \equiv_7 j$ if and only if there is some integer k so that $i + 7k = j$.

1. Show that $\equiv_7$ is an equivalence relation.
2. Describe the set $[5]_{\equiv_7}$.
3. Show that the function $f(n) = n + 3$ is $\equiv_7$-invariant.
4. Show that the function $g(n) = 2n$ is $\equiv_7$-invariant.
5. Determine whether the function $h(n) = 2^n$ is $\equiv_7$-invariant.

15.6 Function Graphs

Suppose S is a set modelling the students at Chapman and M is a set modelling the academic majors the university offers: Mathematics, Philosophy, HeadScratching, and so on. Then elements of S (students) can be related to elements of M (majors) by 'is majoring in' as in "Jethro is majoring in Phrenology." Because a student might have a double major, we can not model the situation as a function, at least not in the most obvious way. Instead, we introduce the notion of a *(binary) relation*. The same idea, generalized to higher dimensional relations, is at the heart of what we call *relational databases*.

Definition 23

A *binary relation from* X *to* Y is a subset $R \subseteq X \times Y$. A *relation on* X is a binary relation from X to X. Since we will only be concerned with binary relations, from now on we refer them simply as *relations*

For a relation R from X to Y, we will say "x is R-related to y" and write $R(x, y)$ when $(x, y) \in R$. In many situations, we use "infix" notation, writing $x \mathrel{R} y$ instead of $R(x, y)$.

Note that there are other equivalent ways to think about relations.

- $R \subseteq X \times Y$ determines a characteristic function $\kappa_R \colon X \times Y \to 2$ so that $R = \{(x, y) \in X \times Y \mid \kappa_R(x, y) = \top\}$
- R determines a function $R[-] \colon X \to \mathcal{P}(Y)$ so that $R = \{(x, y) \in X \times Y \mid y \in R[x]\}$.

The ability to move between these can be helpful. So it is worth practicing.

Exercises for Lecture 15

Let $r\colon \mathbb{N} \times \mathbb{N} \longrightarrow 2$ be the function defined by

$$r(m,n) = \begin{cases} \mathrm{T} & \text{if } m \geq n^2 \\ \mathrm{F} & \text{otherwise.} \end{cases}$$

Consider the relation $R = r^{-}(\mathrm{T})$.

1. Is it the case that $3 \mathrel{R} 2$?
2. Is it the case that $9 \mathrel{R} 3$?
3. What is $R[5]$?
4. What is $\mathsf{curry}[r]$?

Like functions, relations allow a kind of composition and every set has an identity relation.

Definition 24

Suppose $R \subseteq X \times Y$ and $S \subseteq Y \times Z$ are relations. Define $R;S \subseteq X \times Z$ by $x \mathrel{R;S} z$ if and only if there is some $y \in Y$ so that $x \mathrel{R} y$ and $y \mathrel{S} z$. For set X, define $\Delta_X = \{(x,y) \in X \times X \mid x = y\}$.

It will be useful to know that this form of composition behaves similar to function composition.

Lemma 25

For a relation $R \subseteq X \times Y$,

$$\Delta_X; R = R = R; \Delta_Y.$$

For relations $R \subseteq W \times X$, $S \subseteq X \times Y$ and $T \subseteq Y \times Z$,

$$R; (S; T) = (R; S); T.$$

Proof: Exercise. □

Definition 26

A function $f \colon X \to Y$ determines a relation called the *graph of* f defined by $\Gamma_f = \{(x, y) \in X \times Y \mid f(x) = y\}$.

Note: Γ_f is an equalizer $f \circ \pi_0$ and π_1.

The composition of functions and identity functions are essentially the same as composition of graphs and identity relations.

Lemma 27

For functions $X \xrightarrow{f} Y$ and $Y \xrightarrow{g} Z$, $\Gamma_{g \circ f} = \Gamma_f ; \Gamma_g$. For a set X,

$$\mathsf{id}_X = \Gamma_{\Delta_X}$$

.

Proof: Exercise. □

This lemma tells us that Γ is a *functor*. This is an operation that, in some suitable sense, preserves composition and identities. You will encounter functors in further studies. We don't need to investigate the general idea now.

From these simple observations, we might ask which relations arise as graphs of functions. The obvious conditions to consider are these:

Definition 28

A relation $R \subseteq X \times Y$ is

- *total* if for every $x \in X$, there is at least one $y \in Y$ for which $x \mathrel{R} y$;
- *deterministic* if for every $x \in X$, there is at most one $y \in Y$ for which $x \mathrel{R} y$;
- *functional* if it is total and deterministic.

Lemma 29

For any function $X \xrightarrow{f} Y$, the relation $\Gamma_f \subseteq X \times Y$ is functional.

Moreover, if $R \subseteq X \times Y$ is functional, then there is a unique function $X \xrightarrow{f} Y$ so that $R = \Gamma_f$.

Proof: The first claim is easy: Γ_f is total because for each $x \in X$, $x \; \Gamma_f \; f(x)$. It is deterministic because if $x \; \Gamma_f \; y_0$ and $x \; \Gamma_f \; y_1$, then $y_0 = f(x) = y_1$.

Suppose R is a functional relation. Recall that $Y \xrightarrow{\{-\}} \mathcal{P}(Y)$ is the "singleton" function. Let

$$S(Y) = \{A \in \mathcal{P}(Y) \mid \exists y \in Y. A = \{y\}\}.$$

In other words, $S(Y)$ consists of the collection of singleton subsets of Y. Let $\kappa_{S(Y)}$ be the characteristic function of $S(Y)$. So $S(Y)$ is an equalizer $S(Y) = \{A \in \mathcal{P}(Y) \mid \kappa_{S(Y)}(A) = \top\}$. But in fact, Now consider the function $R[-]$ from X to $\mathcal{P}(Y)$ given by $R[x] = \{y \in Y \mid x \; R \; y\}$. Because R is functional, $\kappa_{S(Y)}(R[x]) = \top$ for each x in the domain. So $R[-]$ factors through a function from $f \colon X \to Y$ so that $R[x] = \{f(x)\}$ for all $x \in X$. □

Example 30

The relation of "less than or equal to" $\leq$ on real numbers can be regarded as a subset $\leq \subseteq \mathbb{R} \times \mathbb{R}$ defined by $(x, y) \in \leq$ when x is actually less than or equal to y and $(x, y) \notin \leq$ otherwise. This is a good example of why we prefer to write $x \leq y$ instead of $\leq(x, y)$ or $(x, y) \in \leq(x, y)$. Note that $\leq$ is a total relation, because for any $x \in \mathbb{R}$ there is a $y \in \mathbb{R}$ for which $x \leq y$.

The relation $=$ on any set is functional because, for any $x \in X$ there is exactly one $y \in X$ so that $x = y$.

Define the relation S on $\mathbb{R}$ by $x \; S \; y$ if and only if $x = y^2$. Then S is not deterministic because $1 \; S \; 1$ and $1 \; S \; -1$. It is not total because there is no y for which $-1 \; S \; y$.

Lemma 31

For any function $f \colon Y \to X$, the graph Γ_f is functional.

Proof: This is pretty obvious from the basic properties of functions. That is, for each $x \in X$, $f(x) \in Y$ and obviously $f(x) = f(x)$. So Γ_f is total. On the other hand if $f(x) = y$ and $f(x) = y'$, then $y = y'$. So Γ_f is deterministic. □

Suppose we have a functional relation $R \subseteq X \times Y$. Then it is reasonable to suppose it actually determines a function.

Principle 32: Functional Relations Determine Functions

Suppose $R \subseteq X \times Y$ is a functional relation. Then there is a function $F_R \colon X \to Y$ so that $\Gamma_{F_R} = R$.

It is easy enough to check that $F_{\Gamma_f} = f$ for any function f. That is, $F_{\Gamma_f}(x) = y$ if and only if $x \, \Gamma_f \, y$ if and only if $f(x) = y$. So functions from $X \xrightarrow{f} Y$ correspond exactly to functional relations from $R \subseteq X \times Y$.

Exercises for Lecture 15

1. Define $T \subseteq \mathbb{R} \times \mathbb{R}$ by $x \; T \; y$ if and only if $\tan x = y$. Is T deterministic? Is it total?
2. Show that for any relations $R \subseteq X \times Y$, $\Delta_X; R = R = R; \Delta_Y$.
3. Show that for any relations $R \subseteq W \times X$, $S \subseteq X \times Y$ and $T \subseteq Y \times Z$, $R;(S;T) = (R;S);T$.
4. Show that for any functions $X \xrightarrow{f} Y$ and $Y \xrightarrow{g} Z$, $\Gamma_f; \Gamma_g = \Gamma_{g \circ f}$. Show that for any set X, $\Gamma_{\mathrm{id}_X} = \Delta_X$.

15.7 Splitting and the Axiom of Choice

The us start the section with exercises.

Exercises for Lecture 15

Suppose functions $s \colon Y \to X$ and $r \colon X \to Y$ satisfy $r \circ s = \mathrm{id}_Y$.

1. Show that r is an epimorphism.
2. Show that s is a monomorphism.
3. Show that $s \circ r$ is *idempotent*: $(s \circ r) \circ (s \circ r) = s \circ r$.

Definition 33

Two functions $s\colon Y \to X$ and $r\colon X \to Y$ satisfying $r \circ s = \mathsf{id}_Y$ are called a *section-retract pair*. An idempotent function fXX ($f \circ f = f$) is called a *retraction*.

The latest exercise shows that every section-retract pair gives rise to a retraction. Conversely, if $f\colon X \to X$ is a retraction, we may define $Y = f^+(X)$ – the forward image of X. Then f restricted to Y is a function $s\colon Y \to X$ and corestricted to Y is a function $r\colon X \to Y$. Evidently, s and r form a section-retract pair for which $f = r \circ s$. So retractions and section-retract pairs are essentially the same things.

In a section-retract pair, the section is a monomorphism. In fact, nearly all monomorphisms are sections.

Lemma 34

Suppose $Y \neq \emptyset$. For any monomorphism $m\colon Y \to X$, there is a function $r\colon X \to Y$ so that $r \circ m = \mathsf{id}_Y$.

Proof: Since $Y \neq \emptyset$, we may pick some $b_0 \in Y$. Now define a relation $R \subseteq X \times Y$ by $a\ R\ b$ if and only if either $a = m(b)$ or it is the case that $\forall \in Y.a \neq m(y)$ and $b = b_0$. This is a total relation clearly. And it is deterministic because m is one-to-one. Hence R determines a function r from X to Y. Apparently, $r(m(b)) = b$ because $m(b)\ R\ b$. □

This tells us that every monomorphism from a non-empty domain is *split*: it is the section of a section-retract pair.

The dual of this would be that every epimorphism is also split. But it turns out that the dual is not provable without additional assumptions about Y. We won't go what additional assumptions would help in general, but a hint is given by the natural numbers. Suppose $e\colon \mathbb{N} \to X$ is an epimorphism. Then each subset $e^-(x) \subseteq \mathbb{N}$ is not empty. Every non-empty subset of $\mathbb{N}$ has a least element. So we may define a section corresponding to e by setting $s(x) =$ the least $n \in \mathbb{N}$ so that $e(n) = x$. The point is that $\mathbb{N}$ has enough structure (every non-empty subset has a least element) to allow us to *define* a section.

Without additional structure, it seems reasonable that a section might *exist* for an epimorphism, even though we may not be able to *construct* one explicitly. The Axiom of Choice addresses this situation. Roughly is says that some how, one can make arbitrarily many choices all at once, without having an explicit construction of those choices. Though we will not need it very often in this course, the Axiom of Choice (AC) is quite useful in other areas of mathematics, sometimes to prove something that genuinely needs the axiom, other times to simplify a proof that could have been proved without it.

Principle 35: Axiom of Choice

Axiom of Choice

For every epimorphism $X \xrightarrow{e} Y$ there is a function $Y \xrightarrow{c} X$ for which $e \circ c = \mathsf{id}_Y$

We will rarely use the Axiom of Choice in this course and will draw attention to it when we do. It is a mainstay, however, of most approaches to analysis and topology. There are many equivalent formulations of the axiom that are useful to those areas of mathematics.

Other strong axioms of sets are possible. There are even axioms that inconsistent with AC in the sense that they implies that there actually are epimorphisms that can not be split. For these lectures, though, our goal is to have machinery to do "ordinary" mathematics. The Axiom of Choice falls into that category. The interested student should follow up with a course in Set Theory.