Chapter 1:
From integers to monoids

Algebra was created to solve problems. Like other branches of mathematics, it started off solving very applied problems of a certain type; that is, polynomial equations. When studying algebra the last few years, you have focused on techniques necessary for solving the simplest examples of polynomial equations.

These techniques do not scale well to larger problems. Because of this, algebraists typically take a different turn, one that develops not just techniques, but structures and viewpoints that can be used to solve a vast array of problems, many of which are surprisingly different.

This chapter serves two purposes. First, we re-present ideas you have seen before, but state them in fairly precise terms, which we will then use repeatedly, and require you to use, so as to encourage you to reason with precise meanings of words. This is motivated by a desire for clarity and reproducibility; too often, people speak vaguely to each other, and words contain different meanings for different people.

On the other hand, we also try to introduce some very important algebraic ideas, but intuitively. We will use very concrete examples. True, these examples are probably not as concrete as you might like, but believe me when I tell you that the examples I will use are more concrete than the usual presentation. One goal is to get you to use these examples when thinking about the more general ideas later on. It will be important not only that you reproduce what you read here, but that you explore and play with the ideas and examples, specializing or generalizing them as needed to attack new problems.

Success in this course will require you to balance these inductive and deductive approaches.

1.1: A recapitulation of the past

This chapter focuses on two familiar objects of study: the integers and the monomials. They share a number of important parallels that lay the foundation for the first algebraic structure that we will study. Before we investigate that in detail, let's turn to some general tools of mathematics that you should have seen before now.

Sets

The most fundamental object in mathematics is the **set**. Sets can possess a property called **inclusion** when all the elements of one set are also members of the other. More commonly, people say that the set A is a **subset** of the set B if every element of A is also an element of B. If A is a subset of B but not equal to B, we say that A is a **proper subset** of B. All sets have the **empty set** $\emptyset$ as a subset.

Notation 1.1. If A is a subset of B, we write $A \subseteq B$. If A is a proper subset, we can still write $A \subsetneq B$, but if we want to emphasize that they are not equal, we write $A \subsetneq B$.

You should recognize these sets:

- the **positive integers**, $\mathbb{N}^+ = \{1, 2, 3, \ldots\}$, also called the **counting numbers**, and
- the **integers**, $\mathbb{Z} = \{\ldots, -2, 1, 0, 1, 2, \ldots\}$, which extend $\mathbb{N}^+$ to "complete" subtraction.

You are already familiar with the intuitive motivation for these numbers and also how they are applied, so we won't waste time rehashing that. Instead, let's spend time re-presenting some basic ideas of sets, especially the integers.

Notation 1.2. Notice that both $\mathbb{N} \subseteq \mathbb{Z}$ and $\mathbb{N} \subsetneq \mathbb{Z}$ are true.

We can put sets together in several ways.

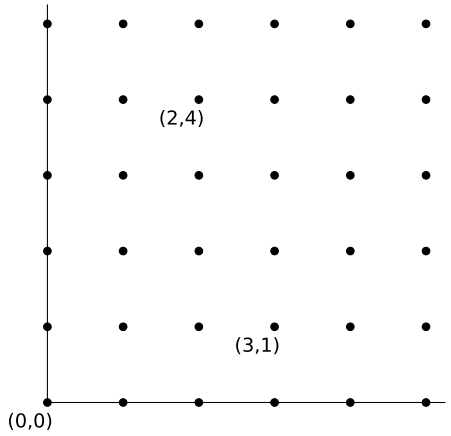
Definition 1.3. Let S and T be two sets. The **Cartesian product of S and T** is the set of ordered pairs

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

Example 1.4. Suppose $S = \{a, b\}$ and $T = \{x+1, y-1\}$. By definition

$$S \times T = \{(a, x+1), (a, y-1), (b, x+1), (b, y-1)\}.$$

Example 1.5. If we let $S = T = \mathbb{N}$, then $S \times T = \mathbb{N} \times \mathbb{N}$, the set of all ordered pairs whose entries are natural numbers. We can visualize this as a lattice, where points must have integer co-ordinates:



Relations

We often want to describe a relationship between two elements of two or more sets. It turns out that this relationship is also a set. Defining it this way can seem unnatural at first, but in the long run, the benefits far outweigh the costs.

Definition 1.6. Any subset of $S \times T$ is **relation on the sets S and T**. A **function** is any relation f such that $(a,b) \in f$ implies $(a,c) \notin f$ for any $c \neq b$.

An **equivalence relation on** S is a subset R of $S \times S$ that satisfies the properties

reflexive:	for all $a \in S$, $(a,a) \in R$;
symmetric:	for all $a, b \in S$, if $(a,b) \in R$ then $(b,a) \in R$; and
transitive:	for all $a, b, c \in S$, if $(a,b) \in R$ and $(b,c) \in R$ then $(a,c) \in R$.

Notation 1.7. Even though relations and functions are sets, we usually write them in the manner to which you are accustomed.

- We typically denote relations that are not functions by symbols such as $<$ or $\subseteq$. If we want a generic symbol for a relation, we usually write $\sim$.
- If $\sim$ is a relation, and we want to say that a and b are members of the relation, we write not $(a,b) \in \sim$, but $a \sim b$, instead. For example, in a moment we will discuss the subset relation $\subseteq$, and we always write $a \subseteq b$ instead of "$(a,b) \in \subseteq$".
- We typically denote functions by letters, typically f, g, or h, or sometimes the Greek letters, η, φ, ψ, or μ. Instead of writing $f \subseteq S \times T$, we write $f : S \to T$. If f is a function and $(a,b) \in f$, we write $f(a) = b$.
- The definition and notation of relations and sets imply that we can write $a \sim b$ and $a \sim c$ for a relation $\sim$, but we cannot write $f(a) = b$ and $f(a) = c$ for a function f.

Example 1.8. Let's define a relation $\sim$ on $\mathbb{R}$, the set of real numbers, in the following way:

$$a \sim b \text{ if and only if } a - b \in \mathbb{Z}.$$

We write this symbolically as

$$a \sim b \iff a - b \in \mathbb{Z}.$$

What kind of elements are related in this fashion?

- $1 \sim 3$ since $1 - 3 = -2 \in \mathbb{Z}$;
- $3 \sim 1$ since $3 - 1 = 2 \in \mathbb{Z}$;
- $4.3 \sim 8.3$ since $4.3 - 8.3 \in \mathbb{Z}$; but
- $0 \not\sim \frac{1}{2}$ since $0 - \frac{1}{2} \notin \mathbb{Z}$.

If you think about it, you should see that $a \sim b$ if they have the same *fractional part*: if we write a and b in decimal form, we see exactly the same numbers on the right hand side of the decimal point, in exactly the same order.

Is $\sim$ an equivalence relation?

reflexive: For any $a \in \mathbb{R}$, we know that $a - a = 0 \in \mathbb{Z}$, so $a \sim a$. Yes, the relation is reflexive.

symmetric: For any $a, b \in \mathbb{R}$, assume that $a \sim b$. By definition of the relation, $a - b \in \mathbb{Z}$. A property of the integers is that if $x \in \mathbb{Z}$, so is $-x$. Thus, since $a - b \in \mathbb{Z}$, we know that $-(a-b) \in \mathbb{Z}$. But $-(a-b) = -a + b = b - a$, so $b - a \in \mathbb{Z}$, too. By definition of the relation, $b \sim a$. Yes, the relation is symmetric.

transitive: For any $a, b, c \in \mathbb{R}$, assume that $a \sim b$ and $b \sim c$. By definition of the relation, $a - b, b - c \in \mathbb{Z}$. We want to show that $a \sim c$, which we could get if we could show that $a - c \in \mathbb{Z}$. How can we get information about $a - c$ from the information given? Notice that $a - c = (a - b) + (b - c)$. Since $a - b$ and $b - c$ are both integers, we can get $a - c$ by adding integers. Since adding integers gives us an integer, we know that $a - c$ is also an integer. Thus, $a \sim c$. Yes, the relation is transitive.

We conclude that, yes, $\sim$ is an equivalence relation.

Another important relation is defined by an operation.

Definition 1.9. Let S and T be sets. An **binary operation from S to T** is a function $f : S \times S \to T$. If $S = T$, we say that f is a binary operation **on** S. A binary operation f on S is **closed** if $f(a,b)$ is defined for all $a, b \in S$.

Example 1.10. Addition of the natural numbers is a function, $+ : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$; the sentence, $2+3=5$ can be thought of as $+(2,3)=5$. Hence, addition is a binary operation on $\mathbb{N}$. Addition is defined for all natural numbers, so it is closed.

Subtraction of natural numbers can be viewed as a function, as well: $- : \mathbb{N} \times \mathbb{N} \to \mathbb{Z}$. However, while subtraction is a binary operation, it is not closed, since it is not "on $\mathbb{N}$": the range ($\mathbb{Z}$) is not the same as the domain ($\mathbb{N}$). This is the reason we need the integers: they "close" subtraction of natural numbers.

In each set described above, you can perform arithmetic: add, subtract, multiply, and (in most cases) divide. We need to make the meaning of these operations precise.[6]

Addition of positive integers is defined in the usual way: it counts the number of objects in the union of two sets with no common element. To obtain the integers $\mathbb{Z}$, we extend $\mathbb{N}^+$ with two kinds of new objects.

- 0 is an object such that $a+0=a$ for all $a \in \mathbb{N}^+$ (the *additive identity*). This models the union of a set of a objects and an empty set.
- For any $a \in \mathbb{N}^+$, we define its *additive inverse*, $-a$, as an object with the property that $a+(-a)=0$. This models *removing* a objects from a set of a objects, so that an empty set remains.

Since $0+0=0$, we are comfortable deciding that $-0=0$. To add with negative integers, let $a, b \in \mathbb{N}^+$ and consider $a+(-b)$:

- If $a=b$, then substitution implies that $a+(-b)=b+(-b)=0$.
- Otherwise, let A be any set with a objects.
  - If I can remove a set with b objects from A, and have at least one object left over, let $c \in \mathbb{N}^+$ be the number of objects left over; then we define $a+(-b)=c$.
  - If I *cannot* remove a set with b objects from A, then let $c \in \mathbb{N}^+$ be the smallest number of objects I would need to add to A so that I could remove b objects. This satisfies the equation $a+c=b$; we then define $a+(-b)=-c$.

For multiplication, let $a \in \mathbb{N}^+$ and $b \in \mathbb{Z}$.

- $0 \cdot b = 0$ and $b \cdot 0 = 0$;
- $a \cdot b$ is the result of adding a copies of b, or

$$\underbrace{(((b+b)+b)+\cdots b)}_{a};$$

  and
- $(-a) \cdot b = -(a \cdot b)$.

We won't bother with a proof, but we assert that such an addition and multiplication are defined for all integers, and satisfy the following properties:

- $a+b=b+a$ and $ab=ba$ for all $a, b \in \mathbb{N}^+$ (the *commutative property*).
- $a+(b+c)=(a+b)+c$ and $(ab)c=a(bc)$ for all $a, b, c \in \mathbb{N}^+$ (the *associative property*).
- $a(b+c)=ab+ac$ for all $a, b, c \in \mathbb{Z}$ (the *distributive property*).

Notation 1.11. For convenience, we usually write $a-b$ instead of $a+(-b)$.

[6] We will not make the meanings as precise as possible; at this level, some things are better left to intuition. For example, I will write later, "If I can remove a set with b objects from [a set with a objects]…" What does this mean? We will not define this, but leave it to your intuition.

We have not yet talked about the additive inverses of additive inverses. Suppose $b \in \mathbb{Z}\backslash\mathbb{N}$; by definition, b is an additive inverse of some $a \in \mathbb{N}^+$, $a + b = 0$, and $b = -a$. Since we want addition to satisfy the commutative property, we *must* have $b + a = 0$, which suggests that we can think of a as the additive inverse of b, as well! That is, $-b = a$. Written another way, $-(-a) = a$. This also allows us to define the **absolute value** of an integer,

$$|a| = \begin{cases} a, & a \in \mathbb{N}, \\ -a, & a \notin \mathbb{N}. \end{cases}$$

Orderings

Notice that we say nothing about the "ordering" of these numbers; that is, we do not "know" yet whether 1 comes before 2 or vice versa. However, our definition of adding negatives has imposed a natural ordering.

Definition 1.12. For any two elements $a, b \in \mathbb{Z}$, we say that:
- $a \leq b$ if $b - a \in \mathbb{N}$;
- $a > b$ if $b - a \notin \mathbb{N}$;
- $a < b$ if $b - a \in \mathbb{N}^+$;
- $a \geq b$ if $b - a \notin \mathbb{N}^+$.

So $3 < 5$ because $5 - 3 \in \mathbb{N}^+$. Notice how the negations work: the negation of $<$ is *not* $>$.

Remark 1.13. You should not yet assume certain "natural" properties of these orderings. For example, it is true that if $a \leq b$, then either $a < b$ or $a = b$. But why? You can reason to it from the definitions given here, so you should do so.

More importantly, you cannot yet assume that if $a \leq b$, then $a + c \leq b + c$. You can reason to this property from the definitions, and you will do so in the exercises.

Some relations enjoy a special status.

Definition 1.14. Let S be any set. A **linear ordering** on S is a relation $\sim$ where for any $a, b \in S$ one of the following holds:

$$a \sim b, a = b, \text{ or } b \sim a.$$

Suppose we define a relation on the subsets of a set S by inclusion; that is, $A \sim B$ if and only if $A \subseteq B$. This relation is *not* a linear ordering, since

$$\{a, b\} \nsubseteq \{c, d\}, \{a, b\} \neq \{c, d\}, \text{ and } \{c, d\} \nsubseteq \{a, b\}.$$

By contrast, the orderings of $\mathbb{Z}$ *are* linear.

Theorem 1.15. The relations $<$, $>$, $\leq$, and $\geq$ are linear orderings of $\mathbb{Z}$.

Our "proof" relies on some unspoken assumptions: in particular, the arithmetic on $\mathbb{Z}$ that we described before. Try to identify where these assumptions are used, because when you write your own proofs, you have to ask yourself constantly: Where am I using unspoken assumptions? In such places, either the assertion must be something accepted by the audience,[7] or you have to cite

[7] In your case, the *instructor* is the audience.

a reference your audience accepts, or you have to prove it explicitly. It's beyond the scope of this course to explain the holes in this proof, but you should at least try to find them.

Proof. We show that $<$ is linear; the rest are proved similarly.

Let $a, b \in \mathbb{Z}$. Subtraction is closed for $\mathbb{Z}$, so $b-a \in \mathbb{Z}$. By definition, $\mathbb{Z} = \mathbb{N}^+ \cup \{0\} \cup \{-1, -2, \ldots\}$. Since $b-a$ must be in one of those three subsets, let's consider each possibility.

- If $b-a \in \mathbb{N}^+$, then $a < b$.
- If $b-a = 0$, then recall that our definition of subtraction was that $b-a = b+(-a)$. Since $b+(-b) = 0$, reasoning on the meaning of natural numbers tells us that $-a = -b$, and thus $a = b$.
- Otherwise, $b-a \in \{-1, -2, \ldots\}$. By definition, $-(b-a) \in \mathbb{N}^+$. We know that $(b-a) + [-(b-a)] = 0$. It is not hard to show that $(b-a)+(a-b) = 0$, and reasoning on the meaning of natural numbers tells us again that $a-b = -(b-a)$. In other words, and thus $b < a$.

We have shown that $a < b$, $a = b$, or $b < a$. Since a and b were arbitrary in $\mathbb{Z}$, $<$ is a linear ordering. □

It should be easy to see that the orderings and their linear property apply to all subsets of $\mathbb{Z}$, in particular $\mathbb{N}^+$ and $\mathbb{N}$. Likewise, we can generalize these orderings to the sets $\mathbb{Q}$ and $\mathbb{R}$ in the way that you are accustomed, and you will do so for $\mathbb{Q}$ in the exercises. That said, this relation behaves differently in $\mathbb{N}$ than it does in $\mathbb{Z}$.

Linear orderings are already special, but some are *extra* special.

Definition 1.16. Let S be a set and $\prec$ a linear ordering on S. We say that $\prec$ is a **well-ordering** if

Every nonempty subset T of S has a **smallest element** a;
that is, there exists $a \in T$ such that for all $b \in T$, $a \prec b$ or $a = b$.

Example 1.17. The relation $<$ is *not* a well-ordering of $\mathbb{Z}$, because $\mathbb{Z}$ itself has no smallest element under the ordering.

Why not? Proceed by way of contradiction. Assume that $\mathbb{Z}$ has a smallest element, and call it a. Certainly $a-1 \in \mathbb{Z}$ also, but

$$(a-1)-a = -1 \notin \mathbb{N}^+,$$

so $a \not< a-1$. Likewise $a \neq a-1$. This contradicts the definition of a smallest element, so $\mathbb{Z}$ is not well-ordered by $<$.

We now assume, *without proof*, the following principle.

The relations $<$ and $\leq$ are well-orderings of $\mathbb{N}$.

That is, any subset of $\mathbb{N}$, ordered by these orderings, has a smallest element. This may sound obvious, but it is very important, and what is remarkable is that *no one can prove it.*[8] It is an assumption about the natural numbers. This is why we state it as a principle (or axiom, if you prefer). In the future, if we talk about the well-ordering of $\mathbb{N}$, we mean the well-ordering $<$.

[8] You might try to prove the well-ordering of $\mathbb{N}$ using induction. But you can't, because it is equivalent to induction. Whenever you have one, you get the other.

One consequence of the well-ordering property is the following fact.

Theorem 1.18. Let $a_1 \geq a_2 \geq \cdots$ be a nonincreasing sequence of natural numbers. The sequence eventually stabilizes; that is, at some index i, $a_i = a_{i+1} = \cdots$.

Proof. Let $T = \{a_1, a_2, \ldots\}$. By definition, $T \subseteq \mathbb{N}$. By the well-ordering principle, T has a least element; call it b. Let $i \in \mathbb{N}^+$ such that $a_i = b$. The definition of the sequence tells us that $b = a_i \geq a_{i+1} \geq \cdots$. Thus, $b \geq a_{i+k}$ for all $k \in \mathbb{N}$. Since b is the *smallest* element of T, we know that $a_{i+k} \geq b$ for all $k \in \mathbb{N}$. We have $b \geq a_{i+k} \geq b$, which is possible only if $b = a_{i+k}$. Thus, $a_i = a_{i+1} = \cdots$, as claimed. □

Another consequence of the well-ordering property is the principle of:

Theorem 1.19 (Mathematical Induction)**.** Let P be a subset of $\mathbb{N}^+$. If P satisfies (IB) and (IS) where
(IB) $1 \in P$;
(IS) for every $i \in P$, we know that $i+1$ is also in P;
then $P = \mathbb{N}^+$.

Proof. Let $S = \mathbb{N}^+ \backslash P$. We will prove the contrapositive, so assume that $P \neq \mathbb{N}^+$. Thus $S \neq \emptyset$. Note that $S \subseteq \mathbb{N}^+$. By the well-ordering principle, S has a smallest element; call it n.

- If $n = 1$, then $1 \in S$, so $1 \notin P$. Thus P does not satisfy (IB).
- If $n \neq 1$, then $n > 1$ by the properties of arithmetic. Since n is the smallest element of S and $n - 1 < n$, we deduce that $n - 1 \notin S$. Thus $n - 1 \in P$. Let $i = n - 1$; then $i \in P$ and $i + 1 = n \notin P$. Thus P does not satisfy (IS).

We have shown that if $P \neq \mathbb{N}^+$, then P fails to satisfy at least one of (IB) or (IS). This is the contrapositive of the theorem. □

Induction is an enormously useful tool, and we will make use of it from time to time. You may have seen induction stated differently, and that's okay. There are several kinds of induction which are all equivalent. We use the form given here for convenience.

Division

Before closing this long beginning, we need one more property of the integers.

Theorem 1.20 (The Division Theorem for Integers)**.** Let $n, d \in \mathbb{Z}$ with $d \neq 0$. There exist unique $q \in \mathbb{Z}$ and $r \in \mathbb{Z}$ satisfying (D1) and (D2) where
(D1) $n = qd + r$;
(D2) $r < |d|$.

Proof. We consider two cases: $d \in \mathbb{N}^+$, and $d \in \mathbb{Z} \backslash \mathbb{N}$. First we consider $d \in \mathbb{N}^+$. We must show two things: first, that q and r exist; second, that r is unique.

Existence of q and r: First we show the existence of q and r that satisfy (D1). Let $S = \{n - qd : q \in \mathbb{Z}\}$ and $M = S \cap \mathbb{N}$. You will show in Exercise 1.31 that M is non-empty. By the well-ordering of $\mathbb{N}$, M has a smallest element; call it r. By definition of S, there exists $q \in \mathbb{Z}$ such that $n - qd = r$. Properties of arithmetic imply that $n = qd + r$.

Does r satisfy (D2)? By way of contradiction, assume that it does not; then $|d| \leq r$. We had assumed that $d \in \mathbb{N}^+$, so Exercises 1.22 and 1.27 implies that $0 \leq r - d < r$. Rewrite property (D1) using properties of arithmetic:

$$\begin{aligned} n &= qd + r \\ &= qd + d + (r - d) \\ &= (q+1)\,d + (r - d)\,. \end{aligned}$$

Rewrite this as $r - d = n - (q+1)\,d$, which shows that $r - d \in S$. Recall $0 \leq r - d$; by definition, $r - d \in \mathbb{N}$. We have $r - d \in S$ and $r - d \in \mathbb{N}$, so $r - d \in S \cap \mathbb{N} = M$. But recall that $r - d < r$, which contradicts the choice of r as the *smallest* element of M. This contradiction implies that r satisfies (D2).

Hence $n = qd + r$ and $0 \leq r < d$; q and r satisfy (D1) and (D2).

Uniqueness of q and r: Suppose that there exist $q', r' \in \mathbb{Z}$ such that $n = q'd + r'$ and $0 \leq r' < d$. By definition of S, $r' = n - q'd \in S$; by assumption, $r' \in \mathbb{N}$, so $r' \in S \cap \mathbb{N} = M$. We chose r to be minimal in M, so $0 \leq r \leq r' < d$. By substitution,

$$\begin{aligned} r' - r &= (n - q'd) - (n - qd) \\ &= (q - q')\,d. \end{aligned}$$

Moreover, $r \leq r'$ implies that $r' - r \in \mathbb{N}$, so by substitution, $(q - q')\,d \in \mathbb{N}$. Similarly, $0 \leq r \leq r'$ implies that $0 \leq r' - r \leq r'$. By substitution, $0 \leq (q - q')\,d \leq r'$. Since $d \in \mathbb{N}^+$, it must be that $q - q' \in \mathbb{N}$ also (repeated addition of a negative giving a negative), so $0 \leq q - q'$. If $0 \neq q - q'$, then $1 \leq q - q'$. By Exercise 1.28, $d \leq (q - q')\,d$. By Exercise 1.26, we see that $d \leq (q - q')\,d \leq r' < d$. This states that $d < d$, a contradiction. Hence $q - q' = 0$, and by substitution, $r - r' = 0$.

We have shown that if $0 < d$, then there exist unique $q, r \in \mathbb{Z}$ satisfying (D1) and (D2). We still have to show that this is true for $d < 0$. In this case, $0 < |d|$, so we can find unique $q, r \in \mathbb{Z}$ such that $n = q\,|d| + r$ and $0 \leq r < |d|$. By properties of arithmetic, $q\,|d| = q\,(-d) = (-q)\,d$, so $n = (-q)\,d + r$. □

Definition 1.21 (terms associated with division)**.** Let $n, d \in \mathbb{Z}$ and suppose that $q, r \in \mathbb{Z}$ satisfy the Division Theorem. We call n the **dividend**, d the **divisor**, q the **quotient**, and r the **remainder**.

Moreover, if $r = 0$, then $n = qd$. In this case, we say that d **divides** n, and write $d \mid n$. We also say that n is **divisible by** d. If on the other hand $r \neq 0$, then d **does not divide** n, and we write $d \nmid n$.

Exercises.

In this first set of exercises, we assume that you are not terribly familiar with creating and writing proofs, so we provide a few outlines, leaving blanks for you to fill in. As we proceed through

Claim: Explain precisely why $0 < a$ for any $a \in \mathbb{N}^+$, and $0 \leq a$ for any $a \in \mathbb{N}$.
Proof:

1. Let $a \in \mathbb{N}^+$ be arbitrary.
2. By definition of _____, $a + 0 = a$.
3. By _____, $0 = -0$.
4. By _____, $a + (-0) = a$.
5. By definition of _____, $a - 0 = a$.
6. By _____, $a - 0 \in \mathbb{N}^+$.
7. By definition of _____, $0 < a$.
8. A similar argument tells us that if $a \in \mathbb{N}$, then $0 \leq a$.

Figure 1.1. Material for Exercise 1.22

Claim: We can order any subset of $\mathbb{Z}$ linearly by $<$.
Proof:

1. Let $S \subseteq \mathbb{Z}$.
2. Let $a, b \in$_____.
3. We consider two cases.
4. If $a - b \in \mathbb{N}^+$, then by $a < b$ by _____.
5. If $a - b = 0$, then simple arithmetic shows that _____.
6. Otherwise, $a - b \in \mathbb{Z} \backslash \mathbb{N}$. By definition, $b - a \in$_____.
7. Then $a < b$ by _____.
8. We have shown that we can order a and b linearly.
9. Since a and b were arbitrary in _____, we can order *any* two elements of that set linearly.

Figure 1.2. Material for Exercise 1.23

the material, we expect you to grow more familiar and comfortable with thinking, so we provide fewer outlines, and in the outlines that we do provide, we require you to fill in the blanks with more than one or two words.

Exercise 1.22.
(a) Fill in each blank of Figure 1.1 with the justification.
(b) Why would someone writing a proof of the claim think to look at $a - 0$?
(c) Why would that person start with $a + 0$ instead?

Exercise 1.23.
(a) Fill in each blank of Figure 1.2 with the justification.
(b) Why would someone writing a proof of this claim think to look at the values of $a - b$ and $b - a$?
(c) Why is the introduction of S essential, rather than a distraction?

Exercise 1.24. Identify the quotient and remainder when dividing:
(a) 10 by -5;
(b) -5 by 10;
(c) -10 by -4.

Exercise 1.25. Let $a \in \mathbb{Z}$. Show that:
(a) $a \leq a+1$;
(b) if $a \in \mathbb{N}$, then $0 \leq a$; and
(c) if $a \in \mathbb{N}^+$, then $1 \leq a$.

Exercise 1.26. Let $a, b, c \in \mathbb{Z}$.
(a) Prove that if $a \leq b$, then $a = b$ or $a < b$.
(b) Prove that if both $a \leq b$ and $b \leq a$, then $a = b$.
(c) Prove that if $a \leq b$ and $b \leq c$, then $a \leq c$.

Exercise 1.27. Let $a, b \in \mathbb{N}$ and assume that $0 < a < b$. Let $d = b - a$. Show that $d < b$.

Exercise 1.28. Let $a, b, c \in \mathbb{Z}$ and assume that $a \leq b$. Prove that
(a) $a + c \leq b + c$;
(b) if $a, c \in \mathbb{N}^+$, $a \leq ac$; and
(c) if $c \in \mathbb{N}^+$, then $ac \leq bc$.

Note: You may henceforth assume this for *all* the inequalities given in Definition 1.12.

Exercise 1.29. Prove that if $a \in \mathbb{Z}$, $b \in \mathbb{N}^+$, and $a \mid b$, then $a \leq b$.

Exercise 1.30. Let $S \subseteq \mathbb{N}$. We know from the well-ordering property that S has a smallest element. Prove that this smallest element is unique.

Exercise 1.31.
(a) Fill in each blank of Figure 1.3 with the justification.
(b) Why would someone writing a proof of the claim think to look at $n - qd$?
(c) Why would this person want to find a value of q?

Exercise 1.32. Show that $>$ is not a well-ordering of $\mathbb{N}$.

Exercise 1.33. Show that the ordering $<$ of $\mathbb{Z}$ generalizes "naturally" to an ordering $<$ of $\mathbb{Q}$ that is also a linear ordering.

Exercise 1.34. Show that divisibility is transitive for the integers; that is, if $a, b, c \in \mathbb{Z}$, $a \mid b$, and $b \mid c$, then $a \mid c$.

Exercise 1.35. By definition, a function is a relation. Can a function be an equivalence relation?

Exercise 1.36.
(a) Fill in each blank of Figure 1.4 with the justification.
(b) Why would someone writing a proof of the claim think to write that $a_i > a_{i+1}$?
(c) Why would someone want to look at the smallest element of A?

Let $n, d \in \mathbb{Z}$, where $d \in \mathbb{N}^+$. Define $M = \{n - qd : q \in \mathbb{Z}\}$.
Claim: $M \cap \mathbb{N} \neq \emptyset$.
Proof: We consider two cases.

1. First suppose $n \in \mathbb{N}$.
   (a) Let $q =$_____. By definition of $\mathbb{Z}$, $q \in \mathbb{Z}$.
   (You can reverse-engineer this answer if you look down a little.)
   (b) By properties of arithmetic, $qd =$_____.
   (c) By _____, $n - qd = n$.
   (d) Recall that $n \in$_____.
   (e) By _____, $n - qd \in \mathbb{N}$.
2. It's possible that $n \notin \mathbb{N}$, so now let's assume that, instead.
   (a) Let $q =$_____. By definition of $\mathbb{Z}$, $q \in \mathbb{Z}$.
   (Again, you can reverse-engineer this answer if you look down a little.)
   (b) By substitution, $n - qd =$_____.
   (c) By _____, $n - qd = -n(d-1)$.
   (d) By _____, $n \notin \mathbb{N}$, but it is in $\mathbb{Z}$. Hence, $-n \in \mathbb{N}^+$.
   (e) Also by _____, $d \in \mathbb{N}^+$, so arithmetic tells us that $d - 1 \in \mathbb{N}$.
   (f) Arithmetic now tells us that $-n(d-1) \in \mathbb{N}$. (pos×natural=natural)
   (g) By _____, $n - qd \in \mathbb{N}$.
3. In both cases, we showed that $n - qd \in \mathbb{N}$. By definition of _____, $n - qd \in M$.
4. By definition of _____, $n - qd \in M \cap \mathbb{N}$.
5. By definition of _____, $M \cap \mathbb{N} \neq \emptyset$.

Figure 1.3. Material for Exercise 1.31

Definition 1.37. We define lcm, the **least common multiple** of two integers, as

$$\operatorname{lcm}(a,b) = \min\left\{n \in \mathbb{N}^+ : a \mid n \text{ and } b \mid n\right\}.$$

This is a precise definition of the least common multiple that you should already be familiar with: it's the smallest (min) nonnegative ($n \in \mathbb{N}$) multiple of a and b ($a \mid n$, and $b \mid n$).

Exercise 1.38.
(a) Fill in each blank of Figure 1.5 with the justification.
(b) One part of the proof claims that "A similar argument shows that $b \mid r$." State this argument in detail.

Exercise 1.39. Let X and Y on the lattice $L = \mathbb{Z} \times \mathbb{Z}$. Let's say that addition is performed as with vectors:

$$X + Y = (x_1 + y_1, x_2 + y_2),$$

multiplication is performed by this *very odd* definition:

$$XY = (x_1 y_1 - x_2 y_2, x_1 y_2 + x_2 y_1),$$

Let S be a well-ordered set.
Claim: Every strictly decreasing sequence of elements of S is finite.
Proof:

1. Let $a_1, a_2, \ldots \in$_____.
   (a) Assume that the sequence is _____.
   (b) In other words, $a_i > a_{i+1}$ for all $i \in$_____.
2. By way of contradiction, suppose the sequence is _____.
   (a) Let $A = \{a_1, a_2, \ldots\}$.
   (b) By definition of _____, A has a smallest element. Let's call that smallest element b.
   (c) By definition of _____, $b = a_i$ for some $i \in \mathbb{N}^+$.
   (d) By _____, $a_i > a_{i+1}$.
   (e) By definition of _____, $a_{i+1} \in A$.
   (f) This contradicts the choice of b as the _____.
3. The assumption that the sequence is _____ is therefore not consistent with the assumption that the sequence is _____.
4. As claimed, then, _____.

Figure 1.4. Material for Exercise 1.36

and the magnitude of a point is devided by the usual Euclidean metric,

$$||X|| = \sqrt{x_1^2 + x_2^2}.$$

(a) Suppose $D = (3,1)$. Calculate $(c,0)\,D$ for several different values of c. How would you describe the results geometrically?
(b) With the same value of D, calculate $(0,c)\,D$ for several different values of c. How would you describe the results geometrically?
(c) Suppose $N = (10,4)$, $D = (3,1)$, and $R = N - (3,0)\,D$. Show that $||R|| < ||D||$.
(d) Suppose $N = (10,4)$, $D = (1,3)$, and $R = N - (3,3)\,D$. Show that $||R|| < ||D||$.
(e) Use the results of (a) and (b) to provide a geometric description of how N, D, and R are related in (c) and (d).
(f) Suppose $N = (10,4)$ and $D = (2,2)$. Find Q such that if $R = N - QD$, then $||R|| < ||D||$. Try to build on the geometric ideas you gave in (e).
(g) Show that for any $N, D \in L$ with $D \neq (0,0)$, you can find $Q, R \in L$ such that $N = QD + R$ and $||R|| < ||D||$. Again, try to build on the geometric ideas you gave in (e).

1.2: Monomials and monoids

We now move from one set that you may consider to be "arithmetical" to another that you will definitely recognize as "algebraic". In doing so, we will notice a similarity in the mathematical structure. That similarity will motivate our first steps into modern algebra, with monoids.

Monomials

Let x represent an unknown quantity. The set of "univariate monomials in x" is

$$\mathbb{M} = \{x^a : a \in \mathbb{N}\}, \tag{1}$$

Let $a, b, c \in \mathbb{Z}$.
Claim: If a and b both divide c, then $\operatorname{lcm}(a,b)$ also divides c.
Proof:

1. Let $d = \operatorname{lcm}(a,b)$. By _____, we can choose q, r such that $c = qd + r$ and $0 \leq r < d$.
2. By definition of _____, d is a multiple of both a and b.
3. By definition of _____, we can find $x, y \in \mathbb{Z}$ such that $d = ax$.
4. By _____, $ax = q(ax) + r$.
5. By _____, $r = a(x - q\widehat{a})$.
6. By definition of _____, $a \mid r$. A similar argument shows that $b \mid r$.
7. We have shown that r is a common multiple of a and b. Recall that $0 \leq r < d$, and _____. Thus, $r = 0$.
8. By _____, $c = qd = q\operatorname{lcm}(a,b)$.
9. By definition of _____, $\operatorname{lcm}(a,b)$ divides c.

Figure 1.5. Material for Exercise 1.38

where x^a, a "monomial", represents precisely what you'd think: the product of a copies of x. In other words,

$$x^a \quad = \quad \prod_{i=1}^{a} x \quad = \quad \underbrace{x \cdot x \cdot \cdots \cdot x}_{n \text{ times}}.$$

We can extend this notion. Let x_1, x_2, …, x_n represent unknown quantities. The set of "multivariate monomials in x_1, x_2, …, x_n" is

$$\mathbb{M}_n = \left\{ \prod_{i=1}^{m} \left(x_1^{a_{i1}} x_2^{a_{i2}} \cdots x_n^{a_{in}} \right) : m, a_{ij} \in \mathbb{N} \right\}. \tag{2}$$

("Univariate" means "one variable"; "multivariate" means "many variables".) For monomials, we allow neither coefficients nor negative exponents. The definition of $\mathbb{M}_n$ indicates that any of its elements is a "product of products".

Example 1.40. The following are monomials:

$$x^2, \quad 1 = x^0 = x_1^0 x_2^0 \cdots x_n^0, \quad x^2 y^3 x y^4.$$

Notice from the last product that the variables need not commute under multiplication; that depends on what they represent. This is consistent with the definition of $\mathbb{M}_n$, each of whose elements is a product of products. We could write $x^2y^3xy^4$ in those terms as

$$\left(x^2y^3\right)\left(xy^4\right) = \prod_{i=1}^{m} \left(x_1^{a_{i1}} x_2^{a_{i2}} \right)$$

with $m = 2$, $a_{11} = 2$, $a_{12} = 3$, $a_{21} = 1$, and $a_{22} = 4$.

The following are not monomials:

$$x^{-1} = \frac{1}{x}, \quad \sqrt{x} = x^{\frac{1}{2}}, \quad \sqrt[3]{x^2} = x^{\frac{2}{3}}.$$

Similarities between $\mathbb{M}$ *and* $\mathbb{N}$

We are interested in similarities between $\mathbb{N}$ and $\mathbb{M}$. Why? Suppose that we can identify a structure common to the two sets. If we make the obvious properties of this structure precise, we can determine non-obvious properties that must be true about $\mathbb{N}$, $\mathbb{M}$, and any other set that adheres to the structure.

If we can prove a fact about a structure,
then we don't have to re-prove that fact for all its elements.
This ***saves*** *time and* ***increases*** *understanding.*

It is harder at first to think about general structures rather than concrete objects, but time, effort, and determination bring agility.

To begin with, what operation(s) should we normally associate with $\mathbb{M}$? We normally associate addition and multiplication with the natural numbers, but the monomials are *not* closed under addition. After all, $x^2 + x^4$ is a *polynomial*, not a monomial. On the other hand, $x^2 \cdot x^4$ is a monomial, and in fact $x^a x^b \in \mathbb{M}$ for any choice of $a, b \in \mathbb{N}$. This is true about monomials in any number of variables.

Lemma 1.41. Let $n \in \mathbb{N}^+$. Both $\mathbb{M}$ and $\mathbb{M}_n$ are closed under multiplication.

Proof for $\mathbb{M}$. Let $t, u \in \mathbb{M}$. By definition, there exist $a, b \in \mathbb{N}$ such that $t = x^a$ and $u = x^b$. By definition of monomial multiplication, we see that

$$tu = x^{a+b}.$$

Since addition is closed in $\mathbb{N}$, the expression $a + b$ simplifies to a natural number. Call this number c. By substitution, $tu = x^c$. This has the form of a univariate monomial; compare it with the description of a monomial in equation (1). So, $tu \in \mathbb{M}$. Since we chose t and u to be arbitrary elements of $\mathbb{M}$, and found their product to be an element of $\mathbb{M}$, we conclude that $\mathbb{M}$ is closed under multiplication. □

Easy, right? We won't usually state all those steps explicitly, but we want to do so at least once.

What about $\mathbb{M}_n$? The lemma claims that multiplication is closed there, too, but we haven't proved that yet. I wanted to separate the two, to show how operations you take for granted in the univariate case don't work so well in the multivariate case. The problem here is that the variables might not commute under multiplication. If we knew that they did, we could write something like,

$$tu = x_1^{a_1+b_1} \cdots x_n^{a_n+b_n},$$

so long as the a's and the b's were defined correctly. Unfortunately, if we assume that the vairables *are* commutative, then we don't prove the statement for everything that we would like. This requires a little more care in developing the argument. Sometimes, it's just a game of notation, as it will be here.

Proof for $\mathbb{M}_n$. Let $t, u \in \mathbb{M}_n$. By definition, we can write

$$t = \prod_{i=1}^{m_t} \left(x_1^{a_{i1}} \cdots x_n^{a_{in}} \right) \quad \text{and} \quad u = \prod_{i=1}^{m_u} \left(x_1^{b_{i1}} \cdots x_n^{b_{in}} \right).$$

(We give subscripts to m_t and m_u because t and u might have a different number of elements in their product. Since m_t and m_u are not the same symbol, it's possible they have a different value.) By substitution,

$$tu = \left(\prod_{i=1}^{m_t}\left(x_1^{a_{i1}}\cdots x_n^{a_{in}}\right)\right)\left(\prod_{i=1}^{m_u}\left(x_1^{b_{i1}}\cdots x_n^{b_{in}}\right)\right).$$

Intuitively, you want to declare victory; we've written tu as a product of variables, right? All we see are variables, organized into two products.

Unfortunately, we're not quite there yet. To show that $tu \in \mathbb{M}_n$, we must show that we can write it as *one* product of a list of products, rather than two. This turns out to be as easy as making the symbols do what your head is telling you: two lists of products of variables, placed side by side, make one list of products of variables. To show that it's one list, we must identify explicitly how many "small products" are in the "big product". There are m_t in the first, and m_u in the second, which makes $m_t + m_u$ in all. So we know that we should be able to write

$$tu = \prod_{i=1}^{m_t+m_u}\left(x_1^{c_{i1}}\cdots x_n^{c_{in}}\right) \tag{3}$$

for appropriate choices of c_{ij}. The hard part now is identifying the correct values of c_{ij}.

In the list of products, the first few products come from t. How many? There are m_t from t. The rest are from u. We can specify this precisely using a piecewise function:

$$c_{ij} = \begin{cases} a_{ij}, & 1 \le i \le m_t \\ b_{ij}, & m_t < i. \end{cases}$$

Specifying c_{ij} this way justifies our claim that tu has the form shown in equation (3). That satisfies the requirements of $\mathbb{M}_n$, so we can say that $tu \in \mathbb{M}_n$. Since t and u were chosen arbitrarily from $\mathbb{M}_n$, it is closed under multiplication. □

You can see that life is a little harder when we don't have all the assumptions we would like to make; it's easier to prove that $\mathbb{M}_n$ is closed under multiplication if the variables commute under multiplication; we can simply imitate the proof for $\mathbb{M}$. You will do this in one of the exercises.

As with the proof for $\mathbb{M}$, we were somewhat pedantic here; don't expect this level of detail all the time. Pedantry has the benefit that you don't have to read between the lines. That means you don't have to think much, only recall previous facts and apply very basic logic. However, pedantry also makes proofs long and boring. While you could shut down much of your brain while reading a pedantic proof, that would be counterproductive. Ideally, you want to reader to *think* while reading a proof, so shutting down the brain is bad. Thus, a good proof does not recount every basic definition or result for the reader, but requires her to make basic recollections and inferences.

Let's look at two more properties.

Lemma 1.42. Let $n \in \mathbb{N}^+$. Multiplication in $\mathbb{M}$ satifies the commutative property. Multiplication in both $\mathbb{M}$ and $\mathbb{M}_n$ satisfies the associative property.

Proof. We show this to be true for $\mathbb{M}$; the proof for $\mathbb{M}_n$ we will omit (but it can be done as it was above). Let $t, u, v \in \mathbb{M}$. By definition, there exist $a, b, c \in \mathbb{N}$ such that $t = x^a$, $u = x^b$, and $v = x^c$. By definition of monomial multiplication and by the commutative property of addition in $\mathbb{M}$, we see that

$$tu = x^{a+b} = x^{b+a} = ut.$$

As t and u were arbitrary, multiplication of univariate monomials is commutative.

By definition of monomial multiplication and by the associative property of addition in $\mathbb{N}$, we see that

$$\begin{aligned} t(uv) &= x^a \left(x^b x^c\right) = x^a x^{b+c} \\ &= x^{a+(b+c)} = x^{(a+b)+c} \\ &= x^{a+b} x^c = (tu)v. \end{aligned}$$

□

You might ask yourself, *Do I have to show* every *step?* That depends on what the reader needs to understand the proof. In the equation above, it *is* essential to show that the commutative and associative properties of multiplication in $\mathbb{M}$ depend strictly on the commutative and associative properties of addition in $\mathbb{N}$. Thus, the steps

$$x^{a+b} = x^{b+a} \quad \text{and} \quad x^{a+(b+c)} = x^{(a+b)+c},$$

with the parentheses as indicated, are absolutely crucial, and cannot be omitted from a good proof.[9]

Another property the natural numbers have is that of an identity: both additive and multiplicative. Since we associate only multiplication with the monomials, we should check whether they have a multiplicative identity. I hope this one doesn't surprise you!

Lemma 1.43. Both $\mathbb{M}$ and $\mathbb{M}_n$ have $1 = x^0 = x_1^0 x_2^0 \cdots x_n^0$ as a multiplicative identity.

We won't bother proving this one, but leave it to the exercises.

Monoids

There are quite a few other properties that the integers and the monomials share, but the three properties we have mentioned here are already quite interesting, and as such are precisely the ones we want to highlight. This motivates the following definition.

[9] Of course, a professional mathematician would not even prove these things in a paper, because they are well-known and easy. On the other hand, a good professional mathematician *would* feel compelled to include in a proof steps that include novel and/or difficult information.

Definition 1.44. Let M be a set, and $\circ$ an operation on M. We say that the pair $(M,\circ)$ is a **monoid** if it satisfies the following properties:
(closed) for any $x,y \in M$, we have $x \circ y \in M$;
(associative) for any $x,y,z \in M$, we have $(x \circ y) \circ z = x \circ (y \circ z)$; and
(identity) there exists an **identity element** $e \in M$ such that for any $x \in M$, we have $e \circ x = x \circ e = x$.
We may also say that M is a **monoid under** $\circ$.

So far, then, we know the following:

Theorem 1.45. $\mathbb{N}$ is a monoid under addition and multiplication, while $\mathbb{M}$ and $\mathbb{M}_n$ are monoids under multiplication.

Proof. For $\mathbb{N}$, this is part of its definition. For $\mathbb{M}$ and $\mathbb{M}_n$, see Lemmas 1.41, 1.42, and 1.43. □

Generally, we don't write the operation in conjunction with the set; we write the set alone, leaving it to the reader to infer the operation. In some cases, this might lead to ambiguity; after all, both $(\mathbb{N},+)$ and $(\mathbb{N},\times)$ are monoids, so which should we prefer? We will prefer $(\mathbb{N},+)$ as the usual monoid associated with $\mathbb{N}$. Thus, we can write that $\mathbb{N}$, $\mathbb{M}$, and $\mathbb{M}_n$ are examples of monoids: the first under addition, the others under multiplication.

What other mathematical objects are examples of monoids?

Example 1.46. You should have seen in linear algebra that the set of square matrices $\mathbb{R}^{m\times m}$ satisfies properties that make it a monoid under both addition and multiplication. That said, your professor almost certainly didn't *call* it a monoid at the time.

Here's an example you probably *haven't* seen before.

Example 1.47. Let S be a set, and let F_S be the set of all functions mapping S to itself, with the proviso that for any $f \in F_S$, $f(s)$ is defined for every $s \in S$. We can show that F_S is a monoid under composition of functions, since

- for any $f,g \in F_S$, we also have $f \circ g \in F_S$, where $f \circ g$ is the function h such that for any $s \in S$,
$$h(s) = (f \circ g)(s) = f(g(s))$$
(notice how important it was that $g(s)$ have a defined value regardless of the value of s);
- for any $f,g,h \in F_S$, we have $(f \circ g) \circ h = f \circ (g \circ h)$, since for any $s \in S$,
$$((f \circ g) \circ h)(s) = (f \circ g)(h(s)) = f(g(h(s)))$$
and
$$(f \circ (g \circ h))(s) = f((g \circ h)(s)) = f(g(h(s)));$$
- if we consider the function $\iota \in F_S$ where $\iota(s) = s$ for all $s \in S$, then for any $f \in F_S$, we have $\iota \circ f = f \circ \iota = f$, since for any $s \in S$,
$$(\iota \circ f)(s) = \iota(f(s)) = f(s)$$
and
$$(f \circ \iota)(s) = f(\iota(s)) = f(s)$$

(we can say that $\iota(f(s)) = f(s)$ because $f(s) \in S$).

Although monoids are useful, they don't capture all the properties that interest us. Not all the properties we found for $\mathbb{N}$ will hold for $\mathbb{M}$, let alone for all monoids. After all, monoids characterize the properties of a set with respect to *only one* operation. Because of this, they cannot describe properties based on two operations.

For example, the Division Theorem requires *two* operations: multiplication (by the quotient) and addition (of the remainder). So, there is no "Division Theorem for Monoids"; it simply doesn't make sense in the context. If we want to generalize the Division Theorem to other sets, we will need a more specialized structure. We will actually meet one later! (in Section 7.4.)

Here is one useful property that we can prove already. A natural question to ask about monoids is whether the identity of a monoid is unique. It isn't hard to show that it is.

Theorem 1.48. Suppose that M is a monoid, and there exist $e, i \in M$ such that $ex = x$ and $xi = x$ for all $x \in M$. Then $e = i$, so that the identity of a monoid is unique.

"Unique" in mathematics means *exactly one*. To prove uniqueness of an object x, you consider a generic object y that shares all the properties of x, then reason to show that $x = y$. This is not a contradiction, because we didn't assume that $x \neq y$ in the first place; we simply wondered about a generic y. We did the same thing with the Division Theorem (Theorem 1.20 on page 11).

Proof. Suppose that e is a left identity, and i is a right identity. Since i is a right identity, we know that

$$e = ei.$$

Since e is a left identity, we know that

$$ei = i.$$

By substitution,

$$e = i.$$

We chose an arbitrary left identity of M and an arbitrary right identity of M, and showed that they were in fact the same element. Hence left identities are also right identities. This implies in turn that there is only one identity: any identity is both a left identity and a right identity, so the argument above shows that any two identities are in fact identical. □

Exercises.

Exercise 1.49. Is $\mathbb{N}$ a monoid under:
(a) subtraction?
(b) division?
Be sure to explain your answer.

Exercise 1.50. Is $\mathbb{Z}$ a monoid under:
(a) addition?
(b) subtraction?
(c) multiplication?

(d) division?

Be sure to explain your answer.

Exercise 1.51. Consider the set $B = \{F, T\}$ with the operation $\vee$ where

$$F \vee F = F$$
$$F \vee T = T$$
$$T \vee F = T$$
$$T \vee T = T.$$

This operation is called **Boolean or**.

Is $(B, \vee)$ a monoid? If so, explain how it justifies each property.

Exercise 1.52. Consider the set $B = \{F, T\}$ with the operation $\oplus$ where

$$F \oplus F = F$$
$$F \oplus T = T$$
$$T \oplus F = T$$
$$T \oplus T = F.$$

This operation is called **Boolean exclusive or**, or **xor** for short.

Is $(B, \oplus)$ a monoid? If so, explain how it justifies each property.

Exercise 1.53. Suppose multiplication of x and y commutes. Show that multiplication in $\mathbb{M}_n$ is both closed and associative.

Exercise 1.54.

(a) Show that $\mathbb{N}[x]$, the ring of polynomials in one variable with integer coefficients, is a monoid under addition.

(b) Show that $\mathbb{N}[x]$ is also a monoid if the operation is multiplication.

(c) Explain why we can replace $\mathbb{N}$ by $\mathbb{Z}$ and the argument would remain valid. (*Hint:* think about the *structure* of these sets.)

Exercise 1.55. Recall the lattice L from Exercise 1.39.

(a) Show that L is a monoid under the addition defined in that exercise.

(b) Show that L is a monoid under the multiplication defined in that exercise.

Exercise 1.56. Let A be a set of symbols, and L the set of all finite sequences that can be constructed using elements of A. Let $\circ$ represent *concatenation of lists.* For example, $(a, b) \circ (c, d, e, f) = (a, b, c, d, e, f)$. Show that $(L, \circ)$ is a monoid.

Definition 1.57. For any set S, let $P(S)$ denote the set of all subsets of S. We call this the **power set** of S.

Exercise 1.58.

(a) Suppose $S = \{a, b\}$. Compute $P(S)$, and show that it is a monoid under $\cup$ (union).

Claim: $(\mathbb{N}, \text{lcm})$ is a monoid. Note that the operation here looks unusual: instead of something like $x \circ y$, you're looking at $\text{lcm}(x, y)$.
Proof:

1. First we show closure.
   (a) Let $a, b \in$_____, and let $c = \text{lcm}(a, b)$.
   (b) By definition of _____, $c \in \mathbb{N}$.
   (c) By definition of _____, $\mathbb{N}$ is closed under lcm.
2. Next, we show the associative property. This is one is a bit tedious...
   (a) Let $a, b, c \in$_____.
   (b) Let $m = \text{lcm}(a, \text{lcm}(b, c))$, $n = \text{lcm}(\text{lcm}(a, b), c)$, and $\ell = \text{lcm}(b, c)$. By _____, we know that $\ell, m, n \in \mathbb{N}$.
   (c) We claim that $\text{lcm}(a, b)$ divides m.
      i. By definition of _____, both a and $\text{lcm}(b, c)$ divide m.
      ii. By definition of _____, we can find x such that $m = ax$.
      iii. By definition of _____, both b and c divide m.
      iv. By definition of _____, we can find y such that $m = by$
      v. By definition of _____, both a and b divide m.
      vi. By Exercise _____, $\text{lcm}(a, b)$ divides m.
   (d) Recall that _____ divides m. Both $\text{lcm}(a, b)$ and _____ divide m. (Both blanks expect the aame answer.)
   (e) By definition of _____, $n \leq m$.
   (f) A similar argument shows that $m \leq n$; by Exercise _____, $m = n$.
   (g) By _____, $\text{lcm}(a, \text{lcm}(b, c)) = \text{lcm}(\text{lcm}(a, b), c)$.
   (h) Since $a, b, c \in \mathbb{N}$ were arbitrary, we have shown that lcm is associative.
3. Now, we show the identity property.
   (a) Let $a \in$_____.
   (b) Let $\iota =$_____.
   (c) By arithmetic, $\text{lcm}(a, \iota) = a$.
   (d) By definition of _____, ι is the identity of $\mathbb{N}$ under lcm.
4. We have shown that $(\mathbb{N}, \text{lcm})$ satisfies the properties of a monoid.

Figure 1.6. Material for Exercise 1.60

(b) Let S be *any* set. Show that $P(S)$ is a monoid under $\cup$ (union).

Exercise 1.59.
(a) Suppose $S = \{a, b\}$. Compute $P(S)$, and show that it is a monoid under $\cap$ (intersection).
(b) Show that $P(S)$ is a monoid under $\cap$ (intersection).

Exercise 1.60.
(a) Fill in each blank of Figure 1.6 with the justification.

Exercise 1.61. Recall the usual ordering $<$ on $\mathbb{M}$: $x^a < x^b$ if $a < b$. Show that this is a well-ordering.

Remark 1.62. While we can define a well-ordering on $\mathbb{M}_n$, it is a much more complicated proposition, which we take up in Section 11.2.

Exercise 1.63. In Exercise 1.34, you showed that divisibility is transitive in the integers.

(a) Show that divisibility is transitive in *any* monoid; that is, if M is a monoid, $a, b, c \in M$, $a \mid b$, and $b \mid c$, then $a \mid c$.

(b) In fact, you don't need all the properties of a monoid for divisibility to be transitive! Which properties *do* you need?

1.3: Isomorphism

We've seen that several important sets share the monoid structure. In particular, $(\mathbb{N}, +)$ and $(\mathbb{M}, \times)$ are very similar. Are they in fact identical *as monoids?* If so, the technical word for this is *isomorphism,* from Greek words meaning "identical shape". How can we determine whether two monoids are isomorphic? We will look for a way to determine whether their operations behave the same way.

Imagine two offices. How would you decide if the offices were equally suitable for a certain job? First, you would need to know what tasks have to be completed, and what materials you need for those tasks. For example, if your job required you to keep books for reference, you would want to find a bookshelf in the office. If it required you to write, you would need a desk, and perhaps a computer. If it required you to communicate with people in other locations, you might need a phone. Having made such a list, you would then want to compare the two offices. If they both had the equipment you needed, you'd think they were both suitable for the job at hand. It wouldn't really matter how the offices satisfied the requirements; if one had a desk by the window, and the other had it on the side opposite the window, that would be okay. If one office lacked a desk, however, it wouldn't be up to the required job.

Deciding whether two sets are isomorphic is really the same idea. First, you decide what structure the sets have, which you want to compare. (So far, we've only studied monoids, so for now, we care only whether the sets have the same monoid structure.) Next, you compare how the sets satisfy those structural properties. If you're looking at finite monoids, an exhaustive comparison might work, but exhaustive methods tend to become exhausting, and don't scale well to large sets. Besides, we deal with infinite sets like $\mathbb{N}$ and $\mathbb{M}$ often enough that we need a non-exhaustive way to compare their structure. Functions turn out to be just the tool we need.

How so? Let S and T be any two sets. Recall that a **function** $f : S \to T$ is a relation that sends every input $x \in S$ to precisely one value in T, the output $f(x)$. You have probably heard the geometric interpretation of this: f passes the "vertical line test." You might suspect at this point that we are going to generalize the notion of function to something more general, just as we generalized $\mathbb{Z}$, $\mathbb{M}$, etc. to monoids. To the contrary; we will *specialize* the notion of a function in a way that tells us important information about a monoid.

Suppose M and N are monoids. If they are isomorphic, their monoid structure is identical, so we ought to be able to build a function that maps elements with a certain behavior in M to elements with the same behavior in N. (Table to table, phone to phone.) What does that mean? Let $x, y, z \in M$ and $a, b, c \in N$. Suppose that $f(x) = a$, $f(y) = b$, $f(z) = c$, and $xy = z$. If M and N have the same structure as monoids, then:

- since $xy = z$,
- we want $ab = c$, or

$$f(x) f(y) = f(z)$$

Substituting xy for z suggests that we want the property

$$f(x)f(y) = f(xy).$$

Of course, we would also want to preserve the identity: f ought to be able to map the identity of M to the identity of N. In addition, just as we only need one table in the office, we want to make sure that there is a one-to-one correspondence between the elements of the monoids. If we're going to reverse the function, it needs to be onto. That more or less explains why we define isomorphism in the following way:

Definition 1.64. Let $(M,\times)$ and $(N,+)$ be monoids. If there exists a function $f : M \longrightarrow N$ such that

- $f(1_M) = 1_N$ *(f preserves the identity)*

and

- $f(xy) = f(x) + f(y)$ for all $x, y \in M$, *(f preserves the operation)*

then we call f a **homomorphism**. If f is also a bijection, then we say that M is **isomorphic** to N, write $M \cong N$, and call f an **isomorphism**. (A **bijection** is a function that is both one-to-one and onto.)

You may not remember the definitions of one-to-one and onto, or you may not understand how to prove them, so here is a precise definition, for reference.

Definition 1.65. Let $f : S \to U$ be a mapping of sets.

- We say that f is **one-to-one** if for every $a, b \in S$ where $f(a) = f(b)$, we have $a = b$.
- We say that f is **onto** if for every $x \in U$, there exists $a \in S$ such that $f(a) = x$.

Another way of saying that a function $f : S \to U$ is onto is to say that $f(S) = U$; that is, the image of S is *all* of U, or that *every* element of U corresponds via f to some element of S.

We used $(M,\times)$ and $(N,+)$ in the definition partly to suggest our goal of showing that $\mathbb{M}$ and $\mathbb{N}$ are isomorphic, but also because they could stand for *any* monoids. You will see in due course that not all monoids are isomorphic, but first let's see about $\mathbb{M}$ and $\mathbb{N}$.

Example 1.66. We claim that $(\mathbb{M},\times)$ is isomorphic to $(\mathbb{N},+)$. To see why, let $f : \mathbb{M} \longrightarrow \mathbb{N}$ by

$$f(x^a) = a.$$

First we show that f is a bijection.

To see that it is one-to-one, let $t, u \in \mathbb{M}$, and assume that $f(t) = f(u)$. By definition of $\mathbb{M}$, $t = x^a$ and $u = x^b$ for $a, b \in \mathbb{N}$. Susbtituting this into $f(t) = f(u)$, we find that $f(x^a) = f\left(x^b\right)$. The definition of f allows us to rewrite this as $a = b$. In this case, $x^a = x^b$, so $t = u$. We assumed that $f(t) = f(u)$ for arbitrary $t, u \in \mathbb{M}$, and showed that $t = u$; that proves f is one-to-one.

To see that f is onto, let $a \in \mathbb{N}$. We need to find $t \in \mathbb{M}$ such that $f(t) = a$. Which t should we choose? We want $f\left(x^?\right) = a$, and $f\left(x^?\right) = ?$, so the "natural" choice seems to be $t = x^a$. That would certainly guarantee $f(t) = a$, but can we actually find such an object t in $\mathbb{M}$? Since

$x^a \in \mathbb{M}$, we can in fact make this choice! We took an arbitrary element $a \in \mathbb{N}$, and showed that f maps some element of $\mathbb{M}$ to a; that proves f is onto.

So f is a bijection. Is it also an isomorphism? First we check that f preserves the operation. Let $t, u \in \mathbb{M}$.[10] By definition of $\mathbb{M}$, $t = x^a$ and $u = x^b$ for $a, b \in \mathbb{N}$. We now manipulate $f(tu)$ using definitions and substitutions to show that the operation is preserved:

$$\begin{aligned} f(tu) &= f\left(x^a x^b\right) = f\left(x^{a+b}\right) \\ &= a + b \\ &= f\left(x^a\right) + f\left(x^b\right) = f(t) + f(u). \end{aligned}$$

Does f also preserve the identity? We usually write the identity of $M = \mathbb{M}$ as the symbol 1, but recall that this is a convenient stand-in for x^0. On the other hand, the identity (under addition) of $N = \mathbb{N}$ is the number 0. We use this fact to verify that f preserves the identity:

$$f(1_M) = f(1) = f\left(x^0\right) = 0 = 1_N.$$

(We don't usually write 1_M and 1_N, but I'm doing it here to show explicitly how this relates to the definition.)

We have shown that there exists a bijection $f : \mathbb{M} \longrightarrow \mathbb{N}$ that preserves the operation and the identity. We conclude that $\mathbb{M} \cong \mathbb{N}$.

On the other hand, is $(\mathbb{N}, \times) \cong (\mathbb{N}, +)$? You might think this is easier to verify, since the sets are the same. Let's see what happens.

Example 1.67. Suppose there *does* exist an isomorphism f between the two monoids. What would have to be true about f?

Suppose the domain of f is $(\mathbb{N}, \times)$, and its range is $(\mathbb{N}, +)$. We know that f preserves the identity; the identity of $(\mathbb{N}, \times)$ is 1, and the identity of $(\mathbb{N}, +)$ is 0, so we need $f(1) = 0$. What would $f(2)$ be, then? Since f must preserve the operation,

$$\begin{aligned} f(2) &= f(1+1) \\ &\underset{\text{iso}}{=} f(1) + f(1) \\ &\underset{\text{sub}}{=} 0 + 0 \\ &= 0 = f(1). \end{aligned}$$

So $f(2) = f(1)$, even though $2 \neq 1$. But this cannot be; it contradicts the definition of isomorphism! Specifically, it contradicts the requirement that f be one-to-one.

Maybe we just chose the domain and range wrong? That's not really possible. Think about it intuitively: if two monoids are isomorphic, it shouldn't matter which one we use for the domain, and which one for the range. Of course, intuition does not suffice; we're talking about the *symmetric* property; that is, if $(\mathbb{N}, \times) \cong (\mathbb{N}, +)$, then $(\mathbb{N}, +) \cong (\mathbb{N}, \times)$. We'll show this in more detail in a moment.

[10] The definition uses the variables x and y, but those are just letters that stand for arbitrary elements of M. Here $M = \mathbb{M}$ and we can likewise choose any two letters we want to stand in place of x and y. It would be a very bad idea to use x when talking about an arbitrary element of $\mathbb{M}$, because there *is* an element of $\mathbb{M}$ called x. So we choose t and u instead.

For the sake of argument, though, suppose against intuition that we merely chose the domain and range wrong; the domain of f should be $(\mathbb{N},+)$, and its range should be $(\mathbb{N},\times)$. Since f must preserve the identity, $f(0)=1$. Let $a\in\mathbb{N}$ such that $f(1)=a$; after all, f has to map 1 to *something!* Then

$$\begin{aligned}
&f(2)=f(1+1)=f(1)\times f(1)=a^2 \text{ and}\\
&f(3)=f(1+(1+1))=f(1)\times f(1+1)=a^3, \text{ so that}\\
&f(n)=\cdots=f(1)^n \text{ for } \textit{any } n\in\mathbb{N}.
\end{aligned}$$

So f sends *every* integer in $(\mathbb{N},+)$to a power of a.

Think about what this implies. For f to be a bijection, it would have to be onto, so *every* element of $(\mathbb{N},\times)$ would *have* to be an integer power of a. ***This is false!*** After all, 2 is not an integer power of 3, and 3 is not an integer power of 2.

The claim was correct: $(\mathbb{N},+)\ncong(\mathbb{N},\times)$.

Exercises.

Exercise 1.68. Show that the monoids "Boolean or" and "Boolean xor" from Exercises 1.51 and 1.52 are *not* isomorphic.

Exercise 1.69. Let $(M,\times)$, $(N,+)$, and $(P,\sqcap)$ be monoids.

(a) Show that the identity function $\iota(x)=x$ is an isomorphism on M.

(b) Suppose that we know $(M,\times)\cong(N,+)$. That means there is an isomorphism $f:M\to N$. One of the requirements of isomorphism is that f be a bijection. Recall from previous classes that this means f has an inverse *function*, $f^{-1}:N\to M$. Show that f^{-1} is an isomorphism.

(c) Suppose that we know $(M,\times)\cong(N,+)$ and $(N,+)\cong(P,\sqcap)$. As above, we know there exist isomorphisms $f:M\to N$ and $g:N\to P$. Let $h=g\circ f$; that is, h is the composition of the functions g and f. Explain why $h:M\to P$, and show that h is also an isomorphism.

(d) Explain how (a), (b), and (c) prove that isomorphism is an equivalence relation.

1.4: Direct products

It might have occurred to you that a multivariate monomial is really a vector of univariate monomials. (Pat yourself on the back if so.) If not, here's an example:

$$x_1^6x_2^3 \text{ looks an awful lot like } \left(x^6,x^3\right).$$

So, we can view any element of $\mathbb{M}_n$ as a list of n elements of $\mathbb{M}$. In fact, if you multiply two multivariate monomials, you would have a corresponding result to multiplying two vectors of univariate monomials componentwise:

$$\left(x_1^6x_2^3\right)\left(x_1^2x_2\right)=x_1^8x_2^4 \quad\text{and}\quad \left(x^6,x^3\right)\times\left(x^2,x\right)=\left(x^8,x^4\right).$$

Last section, we showed that $(\mathbb{M}, \times) \cong (\mathbb{N}, +)$, so it should make sense that we can simplify this idea even further:

$$x_1^6 x_2^3 \text{ looks an awful lot like } (6,3), \text{ and in fact } (6,3) + (2,1) = (8,4).$$

We can do this with other sets, as well.

Definition 1.70. Let $r \in \mathbb{N}^+$ and $S_1, S_2, \ldots, S_r$ be sets. The **Cartesian product** of $S_1, \ldots, S_r$ is the set of all lists of r elements where the ith entry is an element of S_i; that is,

$$S_1 \times \cdots \times S_r = \{(s_1, s_2, \ldots, s_n) : s_i \in S_i\}.$$

Example 1.71. We already mentioned a Cartesian product of two sets in the introduction to this chapter. Another example would be $\mathbb{N} \times \mathbb{M}$; elements of $\mathbb{N} \times \mathbb{M}$ include $\left(2, x^3\right)$ and $\left(0, x^5\right)$. In general, $\mathbb{N} \times \mathbb{M}$ is the set of all ordered pairs where the first entry is a natural number, and the second is a monomial.

If we can preserve the structure of the underlying sets in a Cartesian product, we call it a *direct product*.

Definition 1.72. Let $r \in \mathbb{N}^+$ and $M_1, M_2, \ldots, M_r$ be monoids. The **direct product** of $M_1, \ldots, M_r$ is the pair

$$(M_1 \times \cdots \times M_r, \otimes)$$

where $M_1 \times \cdots \times M_r$ is the usual Cartesian product, and $\otimes$ is the "natural" operation on $M_1 \times \cdots \times M_r$.

What do we mean by the "natural" operation on $M_1 \times \cdots \times M_r$? Let $x, y \in M_1 \times \cdots \times M_r$; by definition, we can write

$$x = (x_1, \ldots, x_r) \quad \text{and} \quad y = (y_1, \ldots, y_r)$$

where each x_i and each y_i is an element of M_i. Then

$$x \otimes y = (x_1 y_1, x_2 y_2, \ldots, x_r y_r)$$

where each product $x_i y_i$ is performed according to the operation that makes the corresponding M_i a monoid.

Example 1.73. Recall that $\mathbb{N} \times \mathbb{M}$ is a Cartesian product; if we consider the monoids $(\mathbb{N}, \times)$ and $(\mathbb{M}, \times)$, we can show that the direct product is a monoid, much like $\mathbb{N}$ and $\mathbb{M}$! To see why, we check each of the properties.

(closure) Let $t, u \in \mathbb{N} \times \mathbb{M}$. By definition, we can write $t = (a, x^\alpha)$ and $u = \left(b, x^\beta\right)$ for

appropriate $a, \alpha, b, \beta \in \mathbb{N}$. Then

$$\begin{aligned} tu &= (a, x^\alpha) \otimes \left(b, x^\beta\right) \\ &= \left(ab, x^\alpha x^\beta\right) && \text{(def. of } \otimes) \\ &= \left(ab, x^{\alpha+\beta}\right) \in \mathbb{N} \times \mathbb{M}. \end{aligned}$$

We took two arbitrary elements of $\mathbb{N} \times \mathbb{M}$, multiplied them according to the new operation, and obtained another element of $\mathbb{N} \times \mathbb{M}$; the operation is therefore closed.

(associativity) Let $t, u, v \in \mathbb{N} \times \mathbb{M}$. By definition, we can write $t = (a, x^\alpha)$, $u = \left(b, x^\beta\right)$, and $v = (c, x^\gamma)$ for appropriate $a, \alpha, b, \beta, c, \gamma \in \mathbb{N}$. Then

$$\begin{aligned} t\,(uv) &= (a, x^\alpha) \otimes \left[\left(b, x^\beta\right) \otimes (c, x^\gamma)\right] \\ &= (a, x^\alpha) \otimes \left(bc, x^\beta x^\gamma\right) \\ &= \left(a\,(bc), x^\alpha \left(x^\beta x^\gamma\right)\right). \end{aligned}$$

To show that this equals $(tu)\,v$, we have to rely on the associative properties of $\mathbb{N}$ and $\mathbb{M}$:

$$\begin{aligned} t\,(uv) &= \left((ab)\,c, \left(x^\alpha x^\beta\right) x^\gamma\right) \\ &= \left(ab, x^\alpha x^\beta\right) \otimes (c, x^\gamma) \\ &= \left[(a, x^\alpha) \otimes \left(b, x^\beta\right)\right] \otimes (c, x^\gamma) \\ &= (tu)\,v. \end{aligned}$$

We took three elements of $\mathbb{N} \times \mathbb{M}$, and showed that the operation was associative for them. Since the elements were arbitrary, the operation is associative.

(identity) We claim that the identity of $\mathbb{N} \times \mathbb{M}$ is $(1,1) = \left(1, x^0\right)$. To see why, let $t \in \mathbb{N} \times \mathbb{M}$. By definition, we can write $t = (a, x^\alpha)$ for appropriate $a, \alpha \in \mathbb{N}$. Then

$$\begin{aligned} (1,1) \otimes t &= (1,1) \otimes (a, x^\alpha) && \text{(subst.)} \\ &= (1 \times a, 1 \times x^\alpha) && \text{(def. of } \otimes) \\ &= (a, x^\alpha) = t \end{aligned}$$

and similarly $t \otimes (1,1) = t$. We took an arbitrary element of $\mathbb{N} \times \mathbb{M}$, and showed that $(1,1)$ acted as an identity under the operation $\otimes$ with that element. Since the element was arbitrary, $(1,1)$ must be *the* identity for $\mathbb{N} \times \mathbb{M}$.

Interestingly, if we had used $(\mathbb{N}, +)$ *instead* of $(\mathbb{N}, \times)$ in the previous example, we *still* would have obtained a direct product! Indeed, the direct product of monoids is *always* a monoid!

Theorem 1.74. The direct product of monoids $M_1, \ldots, M_r$ is itself a monoid. Its identity element is $(e_1, e_2, \ldots, e_r)$, where each e_i denotes the identity of the corresponding monoid M_i.

Proof. You do it! See Exercise 1.77. □

We finally turn our attention the question of whether $\mathbb{M}_n$ and $\mathbb{M}^n$ are the same.

Admittedly, the two are not identical: $\mathbb{M}_n$ is the set of *products* of powers of *n distinct* variables, whereas $\mathbb{M}^n$ is a set of *lists* of powers of *one* variable. In addition, if the variables are *not* commutative (remember that this can occur), then $\mathbb{M}_n$ and $\mathbb{M}^n$ are not at all similar. Think about $(xy)^4 = xyxyxyxy$; if the variables are commutative, we can combine them into x^4y^4, which looks likes $(4,4)$. If the variables are not commutative, however, it is not at *all* clear how we could get $(xy)^4$ to correspond to an element of $\mathbb{N}\times\mathbb{N}$.

That leads to the following result.

Theorem 1.75. The variables of $\mathbb{M}_n$ are commutative if and only if $\mathbb{M}_n \cong \mathbb{M}^n$.

Proof. Assume the variables of $\mathbb{M}_n$ are commutative. Let $f : \mathbb{M}_n \longrightarrow \mathbb{M}^n$ by

$$f\left(x_1^{a_1}x_2^{a_2}\cdots x_n^{a_n}\right) = \left(x^{a_1}, x^{a_2}, \ldots, x^{a_n}\right).$$

The fact that we cannot combine a_i and a_j if $i \neq j$ shows that f is one-to-one, and any element $\left(x^{b_1}, \ldots, x^{b_n}\right)$ of $\mathbb{M}^n$ has a preimage $x_1^{b_1}\cdots x_n^{b_n}$ in $\mathbb{M}_n$; thus f is a bijection.

Is it also an isomorphism? To see that it is, let $t, u \in \mathbb{M}_n$. By definition, we can write $t = x_1^{a_1}\cdots x_n^{a_n}$ and $u = x_1^{b_1}\cdots x_n^{b_n}$ for appropriate $a_1, b_1 \ldots, a_n, b_n \in \mathbb{N}$. Then

$$\begin{aligned}
f(tu) &= f\left(\left(x_1^{a_1}\cdots x_n^{a_n}\right)\left(x_1^{b_1}\cdots x_n^{b_n}\right)\right) && \text{(substitution)}\\
&= f\left(x_1^{a_1+b_1}\cdots x_n^{a_n+b_n}\right) && \text{(commutative)}\\
&= \left(x^{a_1+b_1}, \ldots, x^{a_n+b_n}\right) && \text{(definition of } f\text{)}\\
&= \left(x^{a_1}, \ldots, x^{a_n}\right) \otimes \left(x^{b_1}, \ldots, x^{b_n}\right) && \text{(def. of } \otimes\text{)}\\
&= f(t) \otimes f(u). && \text{(definition of } f\text{)}
\end{aligned}$$

Hence f is an isomorphism, and we conclude that $\mathbb{M}_n \cong \mathbb{M}^n$.

Conversely, suppose $\mathbb{M}_n \cong \mathbb{M}^n$. By Exercise 1.69, $\mathbb{M}^n \cong \mathbb{M}_n$. By definition, there exists a bijection $f : \mathbb{M}^n \longrightarrow \mathbb{M}_n$ satisfying Definition 1.64. Let $t, u \in \mathbb{M}^n$; by definition, we can find $a_i, b_j \in \mathbb{N}$ such that $t = x_1^{a_1}\cdots x_n^{a_n}$ and $u = x_1^{b_1}\cdots x_n^{b_n}$. Since f preserves the operation, $f(tu) = f(t) \otimes f(u)$. Now, $f(t)$ and $f(u)$ are elements of $\mathbb{M}^n$, which is commutative by Exercise 1.78 (with the $S_i = \mathbb{M}$ here). Hence $f(t) \otimes f(u) = f(u) \otimes f(t)$, so that $f(tu) = f(u) \otimes f(t)$. Using the fact that f preserves the operation again, only in reverse, we see that $f(tu) = f(ut)$. Recall that f, as a bijection, is one-to-one! Thus $tu = ut$, and $\mathbb{M}^n$ is commutative. □

Notation 1.76. Although we used $\otimes$ in this section to denote the operation in a direct product, this is not standard; I was trying to emphasize that the product is different for the direct product than for the monoids that created it. In general, the product $x \otimes y$ is written simply as xy. Thus, the last line of the proof above would have $f(t)f(u)$ instead of $f(t) \otimes f(u)$.

Exercises.

Exercise 1.77. Prove Theorem 1.74. Use Example 1.73 as a guide.

Exercise 1.78. Suppose M_1, M_2, ..., and M_n are *commutative* monoids. Show that the direct product $M_1 \times M_2 \times \cdots \times M_n$ is also a commutative monoid.

Exercise 1.79. Show that $\mathbb{M}^n \cong \mathbb{N}^n$. What does this imply about $\mathbb{M}_n$ and $\mathbb{N}^n$?

Exercise 1.80. Recall the lattice L from Exercise 1.39. Exercise 1.55 shows that this is both a monoid under addition and a monoid under multiplication, as defined in that exercise. Is either monoid isomorphic to $\mathbb{N}^2$?

Exercise 1.81. Let $\mathbb{T}_S^n$ denote the set of terms in n variables whose coefficients are elements of the set S. For example, $2xy \in \mathbb{T}_{\mathbb{Z}}^2$ and $\pi x^3 \in \mathbb{T}_{\mathbb{R}}^1$.

(a) Show that if S is a monoid, then so is $\mathbb{T}_S^n$.

(b) Show that if S is a monoid, then $\mathbb{T}_S^n \cong S \times \mathbb{M}_n$.

Exercise 1.82. We define the **kernel** of a monoid homomorphism $\varphi : M \to N$ as

$$\ker \varphi = \{(x, y) \in M \times M : \varphi(x) = \varphi(y)\}.$$

Recall from this section that $M \times M$ is a monoid.

(a) Show that $\ker \varphi$ is a "submonoid" of $M \times M$; that is, it is a subset that is also a monoid.

(b) Fill in each blank of Figure 1.7 with the justification.

(c) Denote $K = \ker \varphi$, and define M/K in the following way.

> A **coset** xK is the set S of all $y \in M$ such that $(x, y) \in K$, and M/K is the set of all such cosets.

Show that

(i) every $x \in M$ appears in at least one coset;

(ii) M/K is a partition of M.

Suppose we try to define an operation on the cosets in a "natural" way:

$$(xK) \circ (yK) = (xy)K.$$

It can happen that two cosets X and Y can each have different representations: $X = xK = wK$, and $Y = yK = zK$. It often happens that $xy \neq wz$, which could open a can of worms:

$$XY = (xK)(yK) = (xy)K \neq (wz)K = (wK)(zK) = XY.$$

Obviously, we'd rather that not happen, so

(iii) Fill in each blank of Figure 1.8 with the justification.

Once you've shown that the operation is well defined, show that

(iv) M/K is a monoid with this operation.

This means that we can use monoid morphisms to create new monoids.

1.5: Absorption and the Ascending Chain Condition

Claim: $\ker\varphi$ is an equivalence relation on M. That is, if we define a relation $\sim$ on M by $x \sim y$ if and only if $(x,y) \in \ker\varphi$, then $\sim$ satisfies the reflective, symmetric, and transitive properties.

1. We prove the three properties in turn.
2. The reflexive property:
   (a) Let $m \in M$.
   (b) By _____, $\varphi(m) = \varphi(m)$.
   (c) By _____, $(m,m) \in \ker\varphi$.
   (d) Since _____, every element of M is related to itself by $\ker\varphi$.
3. The symmetric property:
   (a) Let $a, b \in M$. Assume a and b are related by $\ker\varphi$.
   (b) By _____, $\varphi(a) = \varphi(b)$.
   (c) By _____, $\varphi(b) = \varphi(a)$.
   (d) By _____, b and a are related by $\ker\varphi$.
   (e) Since _____, this holds for all pairs of elements of M.
4. The transitive property:
   (a) Let $a, b, c \in M$. Assume a and b are related by $\ker\varphi$, and b and c are related by $\ker\varphi$.
   (b) By _____, $\varphi(a) = \varphi(b)$ and $\varphi(b) = \varphi(c)$.
   (c) By _____, $\varphi(a) = \varphi(c)$.
   (d) By _____, a and c are related by $\ker\varphi$.
   (e) Since _____, this holds for any selection of three elements of M.
5. We have shown that a relation defined by $\ker\varphi$ satisfies the reflexive, symmetric, and transitive properties. Thus, $\ker\varphi$ is an equivalence relation on M.

Figure 1.7. Material for Exercise 1.82(b)

We conclude our study of monoids by introducing a new object, and a fundamental notion.

Absorption

Definition 1.83. Let M be a monoid, and $A \subseteq M$. If $ma \in A$ for every $m \in M$ and $a \in A$, then A **absorbs from** M. We also say that A is an **absorbing subset**, or that satisfies the **absorption property**.

Notice that if A absorbs from M, then A is closed under multiplication: if $x, y \in A$, then $A \subseteq M$ implies that $x \in M$, so by absorption, $xy \in A$, as well. Unfortunately, that doesn't make A a monoid, as 1_M might not be in A.

Example 1.84. Write $2\mathbb{Z}$ for the set of even integers. By definition, $2\mathbb{Z} \subsetneq \mathbb{Z}$. Notice that $2\mathbb{Z}$ is *not* a monoid, since $1 \notin 2\mathbb{Z}$. On the other hand, any $a \in 2\mathbb{Z}$ has the form $a = 2z$ for some $z \in \mathbb{Z}$. Thus, for any $m \in \mathbb{Z}$, we have

$$ma = m(2z) = 2(mz) \in 2\mathbb{Z}.$$

Since a and m were arbitrary, $2\mathbb{Z}$ absorbs from $\mathbb{Z}$.

The set of integer multiples of an integer is important enough that it inspires notation.

Let M and N be monoids, φ a homomorphism from M to N, and $K = \ker\varphi$.
Claim: The "natural" operation on cosets of K is well defined.
Proof:

1. Let $X, Y \in$_____. That is, X and Y are cosets of K.
2. By _____, there exist $x, y \in M$ such that $X = xK$ and $Y = yK$.
3. Assume there exist $w, z \in$_____ such that $X = wK$ and $Y = zK$. We must show that $(xy)K = (wz)K$.
4. Let $a \in (xy)K$.
5. By definition of coset, _____$\in K$.
6. By _____, $\varphi(xy) = \varphi(a)$.
7. By _____, $\varphi(x)\varphi(y) = \varphi(a)$.
8. We claim that $\varphi(x) = \varphi(w)$ and $\varphi(y) = \varphi(z)$.
   (a) To see why, recall that by _____, $xK = X = wK$ and $yK = Y = zK$.
   (b) By part _____ of this exercise, $(x, x) \in K$ and $(w, w) \in K$.
   (c) By _____, $x \in xK$ and $w \in wK$.
   (d) By _____, $w \in xK$.
   (e) By _____, $(x, w) \in \ker\varphi$.
   (f) By _____, $\varphi(x) = \varphi(w)$. A similar argument shows that $\varphi(y) = \varphi(z)$.
9. By _____, $\varphi(w)\varphi(z) = \varphi(a)$.
10. By _____, $\varphi(wz) = \varphi(a)$.
11. By definition of coset, _____$\in K$.
12. By _____, $a \in (wz)K$.
13. By _____, $(xy)K \subseteq (wz)K$. A similar argument shows that $(xy)K \supseteq (wz)K$.
14. By definition of equality of sets, _____.
15. We have see that the representations of _____ and _____ do not matter; the product is the same regardless. Coset multiplication is well defined.

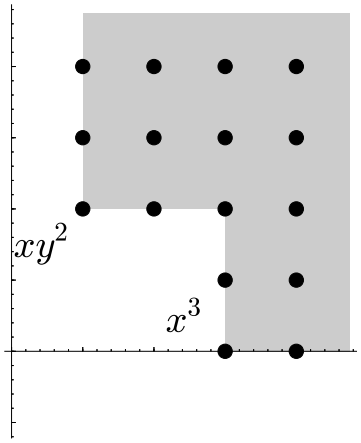
Figure 1.8. Material for Exercise 1.82

Notation 1.85. We write $d\mathbb{Z}$ for the set of integer multiples of d.

So $2\mathbb{Z} = \{\ldots, -2, 0, 2, 4, \ldots\}$ is the set of integer multiples of 2; $5\mathbb{Z}$ is the set of integer multiples of 5; and so forth. You will show in Exercise 1.95 that $d\mathbb{Z}$ absorbs multiplication from $\mathbb{Z}$, but *not* addition.

The monomials provide another important example of absorption.

Example 1.86. Let A be an absorbing subset of $\mathbb{M}_2$. Suppose that $xy^2, x^3 \in A$, but none of their factors is in A. Since A absorbs from $\mathbb{M}_2$, all the monomial multiples of xy^2 and x^3 are also in A. We can illustrate this with a **monomial diagram**:



Every dot represents a monomial in A; the dot at $(1,2)$ represents the monomial xy^2, and the dots above it represent xy^3, xy^4, ... Notice that multiples of xy^2 and x^3 lie *above and to the right* of these monomials.

The diagram suggests that we can identify special elements of subsets that absorb from the monomials.

Definition 1.87. Suppose A is an absorbing subset of $\mathbb{M}_n$, and $t \in A$. If no other $u \in A$ divides t, then we call t a **generator** of A.

In the diagram above, xy^2 and x^3 are the generators of an ideal corresponding to the monomials covered by the shaded region, extending indefinitely upwards and rightwards. The name "generator" is apt, because every monomial multiple of these two xy^2 and x^3 is also in A, but nothing "smaller" is in A, in the sense of divisibility.

This leads us to a remarkable result.

Dickson's Lemma and the Ascending Chain Condition

Theorem 1.88 (Dickson's Lemma)**.** Every absorbing subset of $\mathbb{M}_n$ has a finite number of generators.

(Actually, Dickson proved a similar result for a similar set, but is more or less the same.) The proof is a little complicated, so we'll illustrate it using some monomial diagrams. In Figure 1.9(A), we see an absorbing subset A. (The same as you saw before.) Essentially, the argument *projects* A down one dimension, as in Figure 1.9(B). In this smaller dimension, an argument by induction allows us to choose a finite number of generators, which correspond to elements of A, illustrated in Figure 1.9(C). These corresponding elements of A are always generators of A, but they might not be *all* the generators of A, shown in Figure 1.9(C) by the red circle. In that case, we take the remaining generators of A, use them to construct a new absorbing subset, and project again to obtain new generators, as in Figure 1.9(D). The thing to notice is that, in Figures 1.9(C) and 1.9(D), the y-values of the new generators decrease with each projection. This cannot continue indefinitely, since $\mathbb{N}$ is well-ordered, and we are done.

Proof. Let A be an absorbing subset of $\mathbb{M}_n$. We proceed by induction on the dimension, n.

For the *inductive base,* assume $n = 1$. Let S be the set of exponents of monomials in A. Since $S \subseteq \mathbb{N}$, it has a minimal element; call it a. By definition of S, $x^a \in A$. We claim that x^a is, in fact,

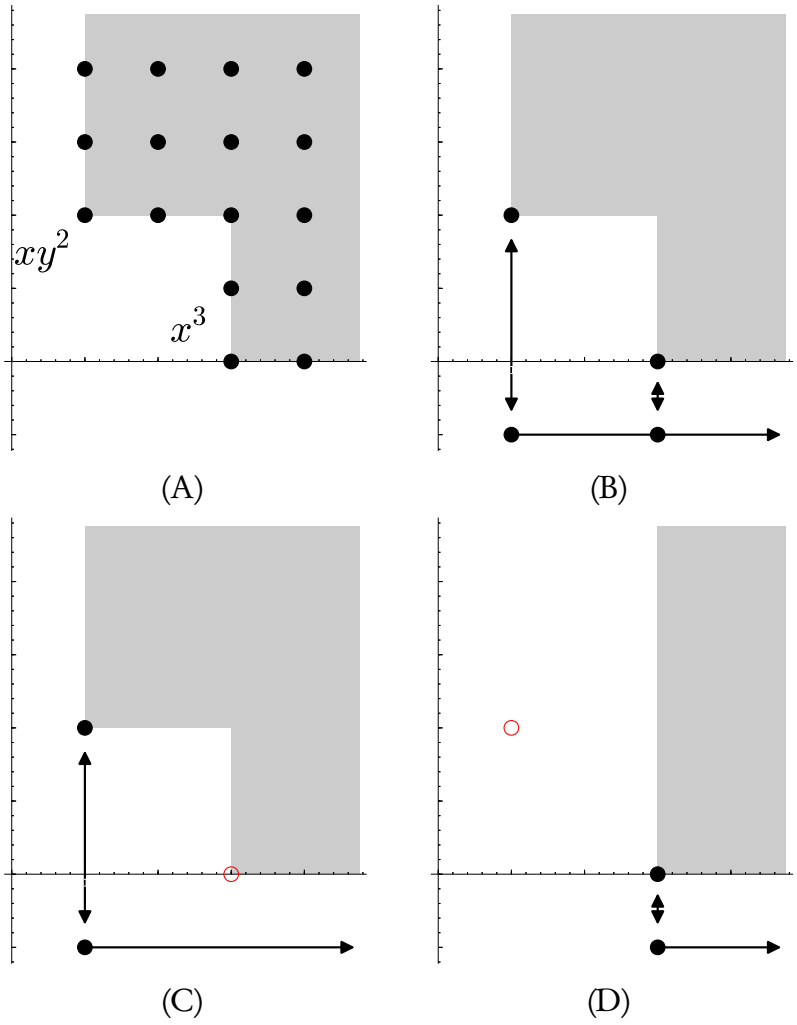


Figure 1.9. Illustration of the proof of Dickson's Lemma.

the one generator of A. To see why, let $u \in A$. Suppose that $u \mid x^a$; by definition of monomial divisibility, $u = x^b$ and $b \leq a$. Since $u \in A$, it follows that $b \in S$. Since a is the *minimal* element of S, $a \leq b$. We already knew that $b \leq a$, so it must be that $a = b$. The claim is proved: no other element of A divides x^a. Thus, x^a is a generator, and since $n = 1$, the generator is unique.

For the *inductive hypothesis*, assume that any absorbing subset of $\mathbb{M}_{n-1}$ has a finite number of generators.

For the *inductive step*, we use A to construct a sequence of absorbing subsets of $\mathbb{M}_{n-1}$ in the following way.

- Let B_1 be the set of all monomials in $\mathbb{M}_{n-1}$ such that $t \in B_1$ implies that $t x_n^a \in A$ for some $a \in \mathbb{N}$. We call this a **projection** of A onto $\mathbb{M}_{n-1}$.
  We claim that B_1 absorbs from $\mathbb{M}_{n-1}$. To see why, let $t \in B_1$, and let $u \in \mathbb{M}_{n-1}$ be any monomial multiple of t. By definition, there exists $a \in \mathbb{N}$ such that $t x_n^a \in A$. Since A absorbs from $\mathbb{M}_n$, and $u \in \mathbb{M}_{n-1} \subsetneq \mathbb{M}_n$, absorption implies that $u\left(t x_n^a\right) \in A$. The associative property tells us that $(ut)\, x_n^a \in A$, and the definition of B_1 tells us that $ut \in B_1$. Since t_1 is an arbitrary element of B_1, u is an arbitrary multiple of t, and we found that $u \in B_1$, we can conclude that B_1 absorbs from $\mathbb{M}_{n-1}$.

This result is important! By the inductive hypothesis, B_1 has a finite number of generators; call them $\{t_1, \ldots, t_m\}$. Each of these generators corresponds to an element of A. Let $T_1 = \left\{t_1 x_n^{a_1}, \ldots, t_m x_n^{a_m}\right\} \subsetneq A$ such that a_1 is the *smallest* element of $\mathbb{N}$ such that $t_1 x_n^{a_1} \in A$, …, a_m is the *smallest* element of $\mathbb{N}$ such that $t_m x_n^{a_m} \in A$. (Such a smallest element must exist on account of the well-ordering of $\mathbb{N}$.)

We now claim that T_1 is a list of some of the generators of A. To see this, assume by way of contradiction that we can find some $u \in T_1$ that is not a generator of A. The definition of a generator means that there exists some other $v \in A$ that divides u. We can write $u = t x_n^a$ and $v = t' x_n^b$ for some $a, b \in \mathbb{N}$; then $t, t' \in B_1$. Here, things fall apart! After all, t' also divides t, contradicting the assumption that t' is a generator of B_1.

- If T_1 is a complete list of the generators of A, then we are done. Otherwise, let $A^{(1)}$ be the absorbing subset whose elements are multiples of the generators of A that are *not* in T_1. Let B_2 be the projection of $A^{(1)}$ onto $\mathbb{M}_{n-1}$. As before, B_2 absorbs from $\mathbb{M}_{n-1}$, and the inductive hypothesis implies that it has a finite number of generators, which correspond to a set T_2 of generators of $A^{(1)}$.
- As long as T_i is not a complete list of the generators of A, we continue building
  - an absorbing subset $A^{(i)}$ whose elements are multiples of the generators of A that are *not* in T_i;
  - an absorbing subset B_{i+1} whose elements are the projections of $A^{(i)}$ onto $\mathbb{M}_{n-1}$, and
  - sets T_{i+1} of generators of A that correspond to generators of B_{i+1}.

Can this process continue indefinitely? No, it cannot. First, if $t \in T_{i+1}$, then write it as $t = t' x_n^a$. On the one hand,

$$t \in A^{(i)} \subsetneq A^{(i-1)} \subsetneq \cdots A^{(1)} \subsetneq A,$$

so t' was an element of every B_j such that $j \leq i$. That means that for each j, t' was divisible by at least one generator u_j' of B_j. However, t was *not* in the absorbing subsets generated by T_1, …, T_i. So the $u_j \in T_j$ corresponding to u_j' does *not* divide t. Write $t = x_1^{a_1} \cdots x_n^{a_1}$ and $u = x_1^{b_1} \cdots x_n^{b_n}$. Since $u' \mid t'$, $b_k \leq a_k$ for each $k = 1, \ldots, n-1$. Since $u \nmid t$, $b_n > a_n$.

In other words, the minimal degree of x_n is decreasing in T_i as i increases. This gives us a strictly decreasing sequence of natural numbers. By the well-ordering property, such a sequence cannot continue indefinitely. Thus, we cannot create sets T_i containing new generators of A indefinitely; there are only finitely many such sets. In other words, A has a finite number of generators. □

This fact leads us to an important concept, that we will exploit greatly, starting in Chapter 8.

Definition 1.89. Let M be a monoid. Suppose that, for any ideals A_1, A_2, … of M, we can guarantee that if $A_1 \subseteq A_2 \subseteq \cdots$, then there is some $n \in \mathbb{N}^+$ such that $A_n = A_{n+1} = \cdots$. In this case, we say that M satisfies the **ascending chain condition**,, or that M is **Noetherian**.

A look back at the Hilbert-Dickson game

We conclude with two results that will, I hope, delight you. There is a technique for counting the number of elements *not* shaded in the monomial diagram.

Definition 1.90. Let A be an absorbing subset of $\mathbb{M}_n$. The **Hilbert Function** $H_A(d)$ counts the number of monomials of total degree d and *not* in A. The **Affine Hilbert Function** $H_A^{\text{aff}}(d)$ is the sum of the Hilbert Function for degree no more than d; that is, $H_A^{\text{aff}}(d) = \sum_{i=0}^{d} H_A(d)$.

Example 1.91. In the diagram of Example 1.86, $H(0) = 1$, $H(1) = 2$, $H(2) = 3$, $H(3) = 2$, and $H(d) = 1$ for all $d \geq 4$. On the other hand, $H^{\text{aff}}(4) = 9$.

The following result is immediate.

Theorem 1.92. Suppose that A is the absorbing subset generated by the moves chosen in a Hilbert-Dickson game, and let $d \in \mathbb{N}$. The number of moves (a, b) possible in a Hilbert-Dickson game with $a + b \leq d$ is $H_A^{\text{aff}}(d)$.

Corollary 1.93. Every Hilbert-Dickson game must end in a finite number of moves.

Proof. Every ith move in a Hilbert-Dickson game corresponds to the creation of a new absorbing subset A_i of $\mathbb{M}_2$. Let A be the union of these A_i; you will show in Exercise 1.96 that A also absorbs from $\mathbb{M}_2$. By Dickson's Lemma, A has finitely many generators; call them $t_1, \ldots, t_m$. Each t_j appears in A, and the definition of union means that each t_j must appear in some A_{i_j}. Let k be the largest such i_j; that is, $k = \max\{i_1, \ldots, i_m\}$. Practically speaking, "largest" means "last chosen", so each t_i has been chosen at this point. Another way of saying this in symbols is that $t_1, \ldots, t_m \in \bigcup_{i=1}^{k} A_i$. All the generators of A are in this union, so no element of A can be absent! So $A = \bigcup_{i=1}^{k} A_i$; in other words, the ideal is generated after finitely many moves. □

Dickson's Lemma is a perfect illustration of the Ascending Chain Condition. It also illustrates a relationship between the Ascending Chain Condition and the well-ordering of the integers: we used the well-ordering of the integers repeatedly to prove that $\mathbb{M}_n$ is Noetherian. You will see this relationship again in the future.

Exercises.

Exercise 1.94. Is $2\mathbb{Z}$ an absorbing subset of $\mathbb{Z}$ under addition? Why or why not?

Exercise 1.95. Let $d \in \mathbb{Z}$ and $A = d\mathbb{Z}$. Show that A is an absorbing subset of $\mathbb{Z}$.

Exercise 1.96. Fill in each blank of Figure 1.10 with its justification.

Exercise 1.97. Let L be the lattice defined in Exercise 1.39. Exercise 1.55 shows that L is a monoid under its strange multiplication. Let $P = (3, 1)$ and A be the absorbing subset generated by P. Sketch L and P, distinguishing the elements of P from those of L using different colors, or an X, or some similar distinguishing mark.

Suppose $A_1, A_2, \ldots$ absorb from a monoid M, and $A_i \subseteq A_{i+1}$ for each $i \in \mathbb{N}^+$.
Claim: Show that $A = \bigcup_{i=1}^{\infty} A_i$ also absorbs from M.

1. Let $m \in M$ and $a \in A$.
2. By _____, there exists $i \in \mathbb{N}^+$ such that $a \in A_i$.
3. By _____, $ma \in A_i$.
4. By _____, $A_i \subseteq A$.
5. By _____, $ma \in A$.
6. Since _____, this is true for all $m \in M$ and all $a \in A$.
7. By _____, A also absorbs from M.

Figure 1.10. Material for Exercise 1.96

Part II

Groups