Lecture Notes for Transition to Advanced Mathematics

James S. Cook
Liberty University
Department of Mathematics and Physics

Spring 2009

introduction and motivations for these notes

These notes are intended to complement your text. I intend to collect all the most important definitions and examples from your text. Additionally, I may add a few examples of my own. I plan to organize these notes lecture by lecture. This semester we should have a total of 23 lectures. Each lecture should last approximately one-hour. Approximately once a week I will take 10 or so minutes to tell you a bit of math history and development, we'll trace out some of the major themes and directions which modern mathematical research includes.

As usual there are many things in lecture which you will not really understand until later. Completing homework in a timely manner is a crucial step in the learning process. Also it may be necessary to read the text more than once to really get it. One size does not fit all, it may be that you need to do more homework than I assign in order to really comprehend the material. I may give a quiz from time to time to help you assess your level of understanding.

In some sense this a survey course. It's a mile wide and an inch deep. However, there are certain issues we will take very seriously in this course. It should be fairly obvious from the homework what is and is not important. I hope you will leave this course with a working knowledge of:

- ✓ conditional and biconditional proofs
- ✓ proof by contradiction
- ✓ proof by contraposition
- ✓ proof by the principle of mathematical induction
- ✓ proper use of set notation and mathematical short-hand
- ✓ given sets A, B, how to prove $A \subseteq B$
- ✓ given sets A, B, how to prove $A = B$
- ✓ equivalence classes and relations
- ✓ proving a function is injective (1-1), surjective(onto)
- ✓ proving a function is bijective(1-1 and onto)

- ✓ basic terminologies in abstract algebra
- ✓ modular arithmetic (how to add subtract and divide in $\mathbb{Z}_m$)

Finally, I hope you can gain some mathematical maturity from this course. Our goal is that this course helps lessen the blow of the "proof" centered courses such as linear algebra, abstract algebra or real analysis.

I believe that in addition to regular homework and tests it is helpful to undertake end of semester projects in advanced courses. This gives you an opportunity to seek out a topic which you find particularly intriguing. This semester I plan for all of you to present a 10 minute proof at the end of the semester. You will be graded both by me and your peers. Part of the grade will be based on questions you ask about other students' presentations. There is a great deal of flexibility in the topic, but you should keep the time limit in mind. I will be quite inflexible about the time. You can present a proof from calculus I,II or III if you wish. We'll say more about this project as the semester progresses.

Contents

1 Logic and Proof

1.1 basic logical operations

A proposition is a sentence which is either true(T) or false(F). We can think of it as a variable which takes just two possible values; True = T or False = F. Incidentally, in engineering it is customary to denote True = 1 while False = 0. If you prefer to use the engineering notation I would not object, just make a note somewhere that 0 = False and 1 = True.

Truth tables are a convenient tool for communicating casewise logic. Traditionally the inputs are listed in the first several columns then the columns to the right of the inputs are formed through various logical operations on the inputs.

Definition 1.1 (negation)**.** *The truth table below defines the logical operation of "negation". The symbol* $\sim P$ *is read "not* P*".*

P	$\sim P$
F	T
T	F

In words, $\sim P$ *takes the opposite truth value of* P*.*

Usually we have propositions that depend on two or more component propositions. Such propositions are called *compound*. Each component proposition has two possible truth values thus truth tables for compound propositions have 2^p rows where $p \in \mathbb{N}$ is the number of component propositions.

Definition 1.2 (conjunction (and))**.** *The truth table below defines the logical operation of "and" which your text calls "conjunction". We read the sentence* $P \wedge Q$ *as "*P *and* Q*".*

P	Q	$P \wedge Q$
F	F	F
F	T	F
T	F	F
T	T	T

In words, the logical operation of "and" is true if and only if both of the input propositions are true.

Definition 1.3 (disjunction (or)). *The truth table below defines the logical operation of "or" which your text calls "disjunction". We read $P \vee Q$ as "P or Q".*

P	Q	$P \vee Q$
F	F	F
F	T	T
T	F	T
T	T	T

Notice this idea of "or" differs from the usual meaning of "or" in common conversation. Mathematical "or" is true if either or both of the members is true.

Definition 1.4 (ex-or). *The truth table below defines the logical operation of "exclusive-or" which is often what is meant by "or" in casual conversation.*

P	Q	P *ex-or* Q
F	F	F
F	T	T
T	F	T
T	T	F

For example: *You come to class or stay in bed.* This sentence is false if you both stay in bed and come to class.

Definition 1.5 (tautology). *A tautology is a propositional form which is true for each possible assignment of its component proposition truth values. A tautology is true always.*

For example: *my wife is correct or she is not.* This is a tautology.

Definition 1.6 (contradiction). *A contradiction is a propositional form which is false for each assignment of truth values for its component propositions. A contradiction is always false.*

For example: *Country music is pleasing to my ears.* This is a contradiction, it's always false. I suspect we will encounter the contradictions such as: "an even number is odd", "an irrational number is rational", "0=1", "a function has two values for a given input".

Definition 1.7 (denial). *A denial of a proposition is logically equivalent to the negation of the proposition.*

The denial of the sentence "$x = 1$" is "$x \neq 1$". The denial of the sentence "The sky is made of cheese" is "the sky is not made of cheese".

Example 1.8. *Let's examine a compound proposition which depends on three given propositions. We'll use the engineering notation for a change of pace,*

P	Q	R	$P \wedge Q$	$Q \wedge R$	$P \wedge Q \wedge R$
0	*0*	*0*	*0*	*0*	*0*
0	*0*	*1*	*0*	*0*	*0*
0	*1*	*0*	*0*	*0*	*0*
0	*1*	*1*	*0*	*1*	*0*
1	*0*	*0*	*0*	*0*	*0*
1	*0*	*1*	*0*	*0*	*0*
1	*1*	*0*	*1*	*0*	*0*
1	*1*	*1*	*1*	*1*	*1*

All three inputs must be true in order for $P \wedge Q \wedge R$ to be true.

Definition 1.9 (implication)**.** *For propositions P and Q, the conditional sentence $P \implies Q$ is the proposition "If P then Q". We call proposition P the antecedent and Q is the consequent. The conditional sentence $P \implies Q$ is true if and only if P is false or Q is true. In other words,*

P	Q	P *implies* Q
F	F	T
F	T	T
T	F	F
T	T	T

Notice that the statement "cats are good implies 1=1" is a true statement. The antecedent P="cats are good" clearly has truth value false, yet the consequent Q="1=1" has truth value true. If the consequent is a tautology then the implication is independendent of the antecedent. In particular, if the consequent is a tautology then the conditional sentence is a tautology. A more useful observation is known as *Modus Ponens*: it states that if we know both P and $P \implies Q$ are true then Q must be true. This is how we typically use implication in our mathematical daily life.

Definition 1.10 (converse)**.** *The converse of $P \implies Q$ is $Q \implies P$*

Example 1.11. *Let S be the conditional sentence "If a function is differentiable then it is a continuous function". The converse of S is "If a function is continuous then it is a differentiable function". The statement S is true while the converse of S is false (take $f(x) = |x|$ for instance).*

Definition 1.12 (contrapositive)**.** *The contrapositive of $P \implies Q$ is $\sim Q \implies \sim P$*

Example 1.13. *"If $n \in \mathbb{N}$ then $n \in \mathbb{Z}$" has the contrapositive "If $\notin \mathbb{Z}$ then $n \notin \mathbb{N}$. The statment and its contrapositive are true.*

We can generalize the patterns seen in the examples just given.

Theorem 1.14 (a.k.a. Theorem 1.1)**.**

a. A conditional sentence and its contrapositive are equivalent

b. A conditional sentence and its converse are not equivalent

Proof: in-class exercise, compare the truth tables for the contrapositive and converse with the truth table of the condtional sentence. The logical equivalence of propositions means those propositions will have matching truth tables. We use this idea to prove theorems such as this one. □

Definition 1.15 (if and only if)**.** *For propositions P and Q, the biconditional sentence $P \Leftrightarrow Q$ is the proposition "P if and only if Q". $P \Leftrightarrow Q$ is true exactly when P and Q have the same truth values.*

I have already used the abbreviation "iff" to replace if and only if. I suppose I should admit there is a certain circularity of this discussion. We are assuming logic to define logic. I'm not sure this can be avoided. I think we can at least agree that we are settling on common definitions for logical operations. I think truth tables are an invaluable help in this regard, sentences in ordinary conversational english leave much to be desired. This is part of the reason people have so much trouble with math, we use common words in uncommon ways.

Theorem 1.16 (this is Theorem 1.2)**.** *For propositions $P, Q,$ and R,*

a. $P \implies Q$ is equivalent to $(\sim P) \vee Q$

b. $P \Leftrightarrow Q$ is equivalent to $(P \implies Q) \wedge (Q \implies P)$

c. $\sim (P \wedge Q)$ is equivalent to $(\sim P) \vee (\sim Q)$

d. $\sim (P \vee Q)$ *is equivalent to* $(\sim P) \wedge (\sim Q)$

e. $\sim (P \implies Q)$ *is equivalent to* $P \wedge (\sim Q)$

f. $\sim (P \wedge Q)$ *is equivalent to* $P \implies \sim Q$

g. $P \wedge (Q \vee R)$ *is equivalent to* $(P \wedge Q) \vee (P \wedge R)$

h. $P \vee (Q \wedge R)$ *is equivalent to* $(P \vee Q) \ \wedge \ (P \vee R)$.

Proof: Exercise 7 of your homework. If you ask nicely I might do part of it in lecture. Proofs for parts c. and d. are given in the worked problems.

Abbreviated Notations: the connectives $\sim, \wedge, \ OR, \implies$,and $\Leftrightarrow$ are applied in the order just listed. This means we can hide certain parenthesis if it is desirable. Generally I tend to use more parentheses then I technically need.

Example 1.17. *The proposition* $P \Rightarrow \sim Q \vee R \Leftrightarrow \sim R \wedge T$ *is an abbreviated form of*

$$\Big(P \Rightarrow \big[(\sim Q) \vee R\big]\Big) \Leftrightarrow \Big((\sim R) \wedge T\Big)$$

1.2 logic with variables; quantifiers

An open sentence depends on one or more variables. To be technical an open sentence is a mapping from $U \to \{F, T\}$. For each point $x \in U$ we assign $P(x) \in \{F, T\}$. The set U is called the **universe of discourse**.

Example 1.18. *Let* $P(x)$ *be the open sentence* $\sin(x) = 0$. *This is an open sentence because we have no value for* x. *The sentence* $P(x)$ *could be true or false. We say that the set of values of* x *which make* $P(x)$ *true are the* **truth set**. *This "truth set" must be taken relative to some* **universe of discourse**. *If the universe is* $[0, 1]$ *then the truth set is just* $\{0\}$. *On the other hand, if the universe is* $\mathbb{R}$ *then the truth set is* $\{n\pi \mid n \in \mathbb{Z}\ \}$. *If the universe of discourse is* $\mathbb{C}$ *then the truth set is still just* $\{n\pi \mid n \in \mathbb{Z}\ \}$, *but that is not something I'd expect you to know unless I gave some supporting theory.*

The big idea here is that the truth set is not allowed to extend past the universe.

Definition 1.19 (equivalence of open sentences)**.** *With a universe specified, two open sentences $P(x)$ and $Q(x)$ are equivalent iff they share the same truth set.*

Example 1.20. *Suppose $U = \mathbb{R}$. Let $P(x)$ be the open sentence* $\sin(x) = 0$ *and* $\cos(x) = 0$*. Furthermore, let $Q(x)$ be the open sentence $e^x = 0$. These are equivalent statements since their truth sets are the same. The truth set of both is the empty set. The empty set is the set which contains no elements. We denote the* **empty set** *by $\emptyset$.*

Definition 1.21 (existential quantifier)**.** *For an open sentence $P(x)$, the sentence $(\exists x)P(x)$ is read "there exists x such that $P(x)$" or "for some x, P(x)" and is true iff the truth set of $P(x)$ is nonempty. The symbol $\exists$ is called the existential quantifier.*

Example 1.22. *If $U = \mathbb{R}$ then $\exists x$ such that $e^x = 1$. The truth set is simply $x = 0$. In contrast, if $U = \mathbb{N}$ then $\sim \exists x$ (read "there does not exist x") such that $e^x = 1$. Sadly, I cannot find the symbol in the standard Latex characters. In lecture I'll write the standard symbol that is shorthand for $\sim \exists$, it is just the $\exists$ symbol with a slash drawn through it.*

Definition 1.23 (universal quantifier)**.** *For and open sentence $P(x)$, the sentence $(\forall x)P(x)$ is read "for all x, $P(x)$" and is true iff the truth set of $P(x)$ is the entire universe. The symbol $\forall$ is called the universal quantifier.*

In calculus, I usually take the universe to be all real numbers so the universe of discourse is set by default in that course. Sometimes the universe is not explicitly listed, this can lead to ambiguity. For example, the statement "$x^2 + 1$ cannot be factored " is true in the universe $\mathbb{R}$ but false in the universe $\mathbb{C}$.

Definition 1.24 (quantified equivalence)**.** *Two quantified sentences are equivalent in a given universe iff they have the same truth value in that universe. Two quantified sentences are equivalent iff the are equivalent in every universe.*

Example 1.25. *Let $S_1(x)$ be the open sentence $P(x) \Rightarrow Q(x)$ then let $S_2(x)$ be the open sentence $\sim Q(x) \Rightarrow \sim P(x)$. Let's prove that $S_1(x)$ and $S_2(x)$ are equivalent.*

Proof: *Let U be any universe. Draw the truth table for $P(x), Q(x), P(x) \Rightarrow Q(x)$ and $(\sim Q(x)) \Rightarrow (\sim P(x))$. For each $x \in U$ we find that the truth values of $S_1(x)$ and $S_2(x)$ match. Thus $S_1(x)$ and $S_2(x)$ are equivalent open sentences in U. But, U was arbitrairy thus we find $S_1(x)$ is equivalent to $S_2(x)$ in all universes. Therefore, $S_1(x)$ and $S_2(x)$ are equivalent quantified sentences.* $\square$

Theorem 1.26 (this is Theorem 1.3 in the text)**.** *If $A(x)$ is an open sentence with variable x, then*

a. $\sim (\forall x)A(x)$ *is equivalent to* $(\exists x) \sim A(x)$.

b. $\sim (\exists x)A(x)$ *is equivalent to* $(\forall x) \sim A(x)$.

This Theorem is useful for constructing "denials" of quantified sentences. Often in a proof by contradiction a first step is find the negation of a proposition. A "denial" is a negation, but when a quantifier is involved we must take care to mind the $\exists$ and $\forall$ that come into play. This particular point may require a little more work than the other points in this section, at least for me it isn't always natural.

Example 1.27. *If we say "there is not a solution which pleases all people" then that is logically equivalent to saying "there exists a person for which this solution is not pleasing."*

Example 1.28. *The following sentences are logically equivalent: "$f(x) \neq 0$ $\forall x$" and "there does not exist x such that $f(x) = 0$"*

Definition 1.29 (unique existence)**.** *For an open sentence $P(x)$, the proposition $(\exists! x)P(x)$ is read "there exists a unique x such that $P(x)$" and is true iff the truth set of $P(x)$ has exactly one element. The symbol $\exists!$ is called the unique existence quantifier.*

Theorem 1.30 (this is Theorem 1.4 of the text)**.** *If $A(x)$ is an open sentence with variable x, then*

a. $(\exists! x)A(x) \implies (\exists x)A(x)$.

b. $(\exists! x)A(x)$ *is equivalent to*

$$\Big[(\exists x)A(x) \wedge (\forall y)(\forall z)\big(A(y) \wedge A(z) \implies y = z\big)\Big].$$

Proof: Let's prove a. Let U be any universe. Assume there exists a unique $x = x_o \in U$ such that $A(x_o)$ is true. Then $A(x_o)$ is true so there exists $x = x_o$ such that $A(x)$ is true. Thus $\exists A(x)$ since the truth set of $A(x)$ is nonempty. This proves the implication. We have shown that when the antecedent is true then the consequent is true. Thus the implication labled a. is true. (*Usually I would not be so verbose, but since we are starting out I'll say a little extra to get us going.*)

Now we prove b. This is an $\Leftrightarrow$ proof. We need to establish that the implications $\Rightarrow$ and $\Leftarrow$ both hold true.

($\Rightarrow$): assume that $(\exists!x)A(x)$ relative to the universe U. We know the truth set of $A(x)$ is nonempty and contains only one element $x_o \in U$. Suppose $\exists A(x)$ and let $y, z \in U$ such that $A(y) \wedge A(z)$ is true. Then both $A(y)$ and $A(z)$ are true hence $y = x_o$ and $z = x_o$. Consequently, $y = z$. We have shown that $(\exists!x)A(x)$ implies that whenever there exists y, z such that $A(y)$ and $A(z)$ are true then $y = z$.

($\Leftarrow$): Assume that $(\exists x)A(x) \wedge (\forall y)(\forall z)\big(A(y) \wedge A(z) \implies y = z\big)$. This means that both $(\exists x)A(x)$ and $(\forall y)(\forall z)\big(A(y) \wedge A(z) \implies y = z\big)$ are true. Suppose that x_0, x_1 are in the truth set of $A(x)$ (*we can do this since* $\exists A(x)$ *is assumed true*). Then $A(x_0) \wedge A(x_1)$ is true, hence by the second portion of the initial assumption $x_0 = x_1$. We have shown that any two arbitrary elements of the truth set are equal, this shows that the truth set has only one element $x_0 = x_1$. Hence $(\exists!x)A(x)$. $\square$.

Remark 1.31. *When a biconditional proof transitions from the $\Rightarrow$ to the $\Leftarrow$ it is customary to drop the assumptions made in the $\Rightarrow$ portion of the proof. It is acceptable to use the same notation in both directions, however it is crucial to order your statements in their proper logical order. Two paragraphs can have all the same statements and yet one is correct and the other is incorrect. The fact that the conditional sentence is not logically equivalent to it's converse means we must be careful to state our assumptions seperate from our desired conclusions.*

1.3 definitions for our toy examples

There are a few pet projects that the text seems to use over and over for proofs. I'll try to throw in some more interesting calculus examples that the

text tends to shy away from. Anywho, let me collect the things which your text thinks you have definitions for. These are scattered here and there in the text, and I sometimes use them without really explaining why in my homework solutions. I pretend that you guys know the definition of "odd", "even" and "divides by". Do try to memorize these:

1.3.1 even integers

Definition 1.32 (even integer)**.** *We say that $n \in \mathbb{Z}$ is* **even** *iff $\exists k \in \mathbb{Z}$ such that $n = 2k$.*

Example 1.33. *Let $n \in \mathbb{Z}$, we can show $2n^2 + 4n + 8$ is even. The proof is simply the observation that $2n^2 + 4n + 8 = 2(n^2 + 2n + 4)$ and it is clear that $k = n^2 + 2n + 4$ is an integer since the sum and products of integers is again an integer. Basically, the standard of proof I require is that to show an integer is even you must show that you can write it as twice another integer. I'm not to fussy about how you "prove" the k is an integer, typically it should be clear since it is almost always just a polynomial of integers.*

1.3.2 odd integers

Definition 1.34 (odd integer)**.** *We say that $n \in \mathbb{Z}$ is* **odd** *iff $\exists k \in \mathbb{Z}$ such that $n = 2k + 1$.*

Example 1.35. *Let $n \in \mathbb{Z}$, we can show $2n^2 + 4n + 7$ is odd. The proof is simply the observation that $2n^2 + 4n + 7 = 2(n^2 + 2n + 3) + 1$ and it is clear that $k = n^2 + 2n + 3$ is an integer.*

1.3.3 "divides by" and "is a factor"

Definition 1.36 (b is a factor of a)**.** *Let $a, b \in \mathbb{Z}$ then we say that b is a factor a with respect to $\mathbb{Z}$ iff $\exists k \in \mathbb{Z}$ such that $a = kb$.*

Definition 1.37 (b divides a)**.** *Let $a, b \in \mathbb{Z}$ then we say that b divides a iff $\exists k \in \mathbb{Z}$ such that $a = kb$. In other words, b divides a iff b is a factor of a. Notation:*

$$\boxed{a \text{ divides } b \quad \Leftrightarrow \quad a \mid b. \quad \Leftrightarrow a \text{ is a factor of } b}$$

Example 1.38. *Let $n \in \mathbb{Z}$, we can show that 3 divides $(n+1)^2+2n^2+n+2$. The proof is simply the observation that:*

$$(n + 1)^2 + 2n^2 + n + 2 = n^2 + 2n + 1 + 2n^2 + n + 2 = 3(n^2 + n + 1).$$

It is clear that n^2+n+1 is an integer. The fact that I can factor out 3 shows that 3 divides $(n+1)^2+2n^2+n+2$.

Remark 1.39. *The idea of factoring extends to polynomials and other number systems besides just the integers. Technically speaking, we ough to say something is divided by something else with respect to the universe of allowed objects. Our universe in this course is often $\mathbb{Z}$. However, for other universes definition of "divides by" is very similar, for example we can say that x^2+x is divided by $(x+1)$ since $x^2+x=x(x+1)$ in the universe $\mathbb{R}[x]$ (this denotes polynomials in x with real coefficients). On the other hand if $r\in\mathbb{R}$ with $r\neq 0$ we can show that 4 divides r with respect to $\mathbb{R}$; notice that $r=4\left(\frac{r}{4}\right)$. In fact any nonzero real number will divide r by the same argument.*

1.4 examples of proofs

It is unlikely we will cover all of these in lecture. I'd rather do a few carefully and let you read the rest.

1.4.1 direct proofs

The following is an example of a "direct proof". See page 32 of your text for a general guideline of how these go.

Example 1.40. If x,y are even integers then xy is even.
Proof: Let $x,y\in\mathbb{Z}$ be even integers. Then, by definition of even integers, $\exists a,b\in\mathbb{Z}$ such that $x=2a$ and $y=2b$. Observe that $xy=(2a)(2b)=2(2ab)$ thus xy is even. $\square$

In the example above, P was the statementthat x,y were even. Then Q was the statement that xy was even. We assumed P and found that Q necessarily followed. Thus, $P\Rightarrow Q$.

Remark 1.41. *The example just given made an argument for arbitrary integers x and y. Therefore, with respect to $U=\mathbb{Z}$, we can conclude $(\forall x)(\forall y)(x \text{ and } y \text{ even } \Rightarrow xy \text{even})$.*

Example 1.42. If $x\in\mathbb{R}$ then $|x|=\sqrt{x^2}$
Proof: Recall that $|x|=x$ if $x\geq 0$ and $|x|=-x$ if $x<0$. If $x\geq 0$ then $|x|=x=\sqrt{x^2}$. If $x<0$ then $-x>0$ and $\sqrt{x^2}=-x$ (since by definition the squareroot function is non-negative) hence $|x|=-x=\sqrt{x^2}$. Therefore, $|x|=\sqrt{x^2}$ for all $x\in\mathbb{R}$. $\square$

Example 1.43. If $a, b \in \mathbb{R}$ **then** $|ab| = |a||b|$
Proof: Let $a, b \in \mathbb{R}$*, and recall* $|x| = \sqrt{x^2}, \forall x$*,*

$$|ab| = \sqrt{(ab)^2} = \sqrt{a^2 b^2} = \sqrt{a^2}\sqrt{b^2} = |a||b|,$$

by the laws of exponents and radicals. $\square$

We take properties of the real numbers such as laws of exponents and radicals as axioms unless otherwise instructed in this course. An axiom is a truth which is basic and does not follow from other more basic truths. One goal of mathematics in general is to find a minimal set of axioms. We prefer to assume as little as possible at the base of things. The more things we can build from base-principles the better. However, you should be aware that it has been proven in mathematics that there are questions which are undecideable. Godel showed that any system of mathematics which contains arithmetic will necessarily have unanswerable "statements". That is they will not fit the narrow-minded definition of proposition we gave the first day in this course. Godel showed it will not be possible to prove them true or false. Thus, even after we choose suitable axioms for a particular realm of mathematics we will not necessarily be able to answer all questions. This would seem to cast aspersion on those rather ambitious physicists who seek a theory of everything. If mathematics is not entirely self-contained then what hope have we that physics will explain itself?

Ok, back to $\mathbb{Z}$,

Example 1.44. The product of an integer and its immediate successor is even for each integer.
Proof: Let $x \in \mathbb{Z}$ *be an integer. It is clear that* x *is either even or odd. Let us proceed casewise:*

1.) **even case:** $x = 2m$ *for* $m \in \mathbb{Z}$*. The immediate successor of* x *is* $x + 1 = 2m + 1$*. Note that* $x(x + 1) = 2m(2m + 1) = 2[m(2m + 1)]$ *and* $m(2m + 1) \in \mathbb{Z}$ *thus* $x(x + 1)$ *is even.*

2.) **odd case:** $x = 2m + 1$ *for* $m \in \mathbb{Z}$*. The immediate successor of* x *is* $x+1 = 2m+2$*. Note that* $x(x+1) = (2m+1)(2m+2) = 2[(2m+1)(m+1)]$ *and* $(2m + 1)(m + 1) \in \mathbb{Z}$ *thus* $x(x + 1)$ *is even.*

Therefore, as $x \in \mathbb{Z}$ *was arbitrary, we find* $x(x + 1)$ *is even* $\forall x \in \mathbb{Z}$*.* $\square$

In the example above, we assumed x was a generic integer. When we wish to prove some claim for all objects it is crucial that we make no additional assumptions beyond what is contained in the given claim. For example, this proof would have been incomplete is we had assumed that the integer was even from the beginning. When we added that even assumption in one case we were obliged to address the other odd case seperately.

The logic of cases follows from the tautology below:

$$[P_1(x) \vee P_2(x)] \Rightarrow Q(x) \;\Leftrightarrow\; [P_1(x) \Rightarrow Q(x)] \vee [P_2(x) \Rightarrow Q(x)]$$

Actually, the tautology is less strict then what we usually mean by breaking up into "cases". The statement above does not assume the cases are non-overlapping. In the example, $P_1(x)$ was "x is even" while $P_2(x)$ was "x is odd". Clearly, $P_1(x) \wedge P_2(x)$ is a contradiction for $x \in \mathbb{Z}$. However, we could even break up into cases which overlapped. So long as the cases cover all possibilities for the quantifier then we can conclude the proof holds for all $x \in U$. The preceding example had $U = \mathbb{Z}$.

In the example 1.43 we would have to have treated 4 cases for the different sign combinations of a, b. Fortunately, we knew $|x| = \sqrt{x^2}$ which is super-nice since it avoids cases. Beware, not all questions about absolute value can be nicely solved without resorting to cases. See your text for a simple proof that $-|x| \leq x \leq |x|$. I think cases are unaviodable for that argument. (also see homework solution page 10, solution for problem 1.4.6e)

Example 1.45. Let $a, b \in \mathbb{Z}$**. If** $a, b > 0$ **and** $a|b$ **and** $b|a$**, then** $a = b$.
Proof: since $a|b$, $\exists x \in \mathbb{Z}$ *such that* $b = ax$. *Moreover, since* $b|a$, $\exists y \in \mathbb{Z}$ *such that* $a = by$. *Since* $a, b > 0$ *it follows that* $x, y > 0$ *thus,*

$$\frac{b}{a} = x \text{ and } \frac{b}{a} = \frac{1}{y} \quad \Rightarrow x = \frac{1}{y}$$

This is a very interesting equation in $\mathbb{Z}$. *It says that* y *has a multiplicative inverse of* x. *There are only two integers with multiplicative inverses,* 1 *and* -1. *We have* $x > 0$ *thus we find the solution must be* $x = 1$. *The claim follows.* $\square$

Example 1.46. Let $a, b, c, d \in \mathbb{Z}$**, if** $a|b$ **and** $c|d$ **then** $ac|bd$**.**
Proof: see homework solutions page 11 for problem 1.4.7j.$\square$

1.4.2 proof by contradiction

The following is an example of "proof by contradiction". We assume the negation of the claim then work towards seeing why it is unreasonable. Then once we see the negation is always wrong we conclude that the claim must be true. The symbol $\rightarrow\leftarrow$ means "contradiction", it points out there is a proposition in the proof which is always false. To be kind to the reader it is nice to announce you are using proof by contradiction at the beginning of the proof. It is also nice to point out explicitly what the contradiction is. Mathematicians are not always nice.

Example 1.47. The circle $x^2 + y^2 = 1$ does not intersect the line $y = 2$.
Proof (by contradiction): Assume that the circle and line intersect. Let (a, b) be a point of intersection then that point must satisfy both the equation of the line and the equation of the circle:

$$(i.)\ a^2 + b^2 = 1 \qquad\qquad (ii.)\ b = 2$$

But then if we substitute $(ii.)$ into $(i.)$ we obtain $a^2+4 = 1$ which is equivalent to $a^2 = -3$, thus $a \notin \mathbb{R}$ $a \rightarrow\leftarrow$. Therefore, using proof by contradiction, we find the circle does not intersect the line. $\square$

The contradiction here was that $a \in \mathbb{R}$ and $a \notin \mathbb{R}$. Your text points out the symbolic structure of proof by contradiction is:

$$P \Leftrightarrow [(\sim P) \implies (Q \wedge \sim Q)].$$

The interesting thing about this pattern is that Q appears just on the RHS. There is a lot of freedom in what Q is seen to be. In practice, for a given problem is does have something to do with P but not directly. Here P was the proposition that the circle intersected the line, however Q was the proposition that a was a real number. You can easily verify that Q is not some trivial logical operation applied to P. The connection between P and Q stems from the theory behind the proof, not from logic alone.

Example 1.48. The $\sqrt{2}$ is irrational.
Proof (by contradiction): Assume that $\sqrt{2}$ is rational. Then there exist $a, b \in \mathbb{Z}$ such that $\sqrt{2} = \frac{a}{b}$. Hence, $2 = \frac{a^2}{b^2}$ which yields $2b^2 = a^2$. By the Fundamental Theorem of Arithmetic both a and b can be factored into a unique(upto ordering) product of primes. Clearly a^2 will have an even number of factors since the total number of factors for the product of aa

is twice that of a. Likewise for b. We argue that a^2 and b^2 have an even number of factors of 2. Consider then $2b^2 = a^2$, since b^2 has an even number of two-factors in its factorization when we multiply by 2 we find $2b^2$ has an odd number of two-factors. Thus, as $2b^2 = a^2$ we find a^2 has an odd-number of two-factors, a $\rightarrow\leftarrow$. Therefore, using proof by contradiction, we find $\sqrt{2} \notin \mathbb{Q}$. $\square$

Notice the contradiction came from something only tangentially related to the main thrust of the theorem here. Many people (myself included) find proof by contradiction less than satisfying. Take the proof above, it leaves us with an obvious question: if $\sqrt{2}$ is not rational then what is it? The proof tells us nothing about that. Proof by contradiction is a useful method despite these comments. To reject this method of proof is to reject much of modern mathematics. I use it as a last resort. When possible, I prefer a proof which is **constructive**. Hilbert agrees, in 1928 he wrote in *Die Grundlagen der Mathematik*,

"Taking the principle of excluded middle from the mathematician would be the same, say, as proscribing the telescope to the astronomer or to the boxer the use of his fists"

There is a camp of mathematicians who follow "mathematical intuitionism". Such mathematicians reject proof by contradiction as a valid method of proof. The founder of this camp was a Dutch mathematician named Luitzen Egbertus Jan Brouwer. At an early age he proved the highly nontrivial topological fixed point theorem. Mathematical intuitionism views mathematics as the formulation of mental constructions that are governed by self-evident laws. Brouwer had interesting opinions (Brouwer was apparently not a fan of applied mathematics):

"The construction itself is an art, its application to the world an evil parasite."

I should mention that the school of "constructivist" mathematics is larger than that of intuitionism. Besides avoiding proof by contradiction, constructivists also eschew the use of the *Axiom of Choice*. If a proof relies on the Axiom of Choice then at some point it chooses objects that could not be constructed explicitly for some reason. Constructive mathematicians do not necessarily disagree with the results the follow from the Axiom of Choice. In fact, one goal is to give new constructive proofs that avoid the Axiom

of Choice. This is not always possible. Certain statements in mathematics have been shown to produce the Axiom of Choice and vice-versa. For example, Sierpinski proved that Zermelo Franklin Set Theory with no Axiom of Choice (ZF) and the Generalized Continuum Hypothesis(GCH) implies the Axiom of Choice(AC), so AC and GCH are not independent in ZF. In short, there are still open questions here and the story is much deeper than I have sketched here.

I should mention that constructivists have mellowed a little since the days of Leopold Kronecker. In a famous quote he argued that arithmetic and analysis must be founded on "whole numbers", saying, "God made the integers; all else is the work of man"

If you were wondering how to define $\mathbb{R}$, a concise explaination is that it is *closure* of $\mathbb{Q}$. We can define the real numbers to be the union of the rational numbers and the accumulation points for all Cauchy-sequences of rational numbers. This is not part of the required material of Math 200, but I figure some of you are interested. You might be able to fill in the details of this idea of "closure" when you have completed the real analysis course here at LU. The structure of the real numbers is more subtle than our early education would have us believe.

1.4.3 proof by contraposition

Example 1.49. Let $x \in \mathbb{Z}$, if 8 does not divide $x^2 - 1$ then x is even. *Proof (by contraposition): Suppose that x is not even, then x is odd. Thus there exists $m \in \mathbb{Z}$ such that $x = 2m + 1$. Consider,*

$$\begin{aligned} x^2 - 1 &= (2m+1)^2 - 1 && (1)\\ &= 4m^2 + 4m + 1 - 1 && (2)\\ &= 4m(m+1) && (3)\\ &= 4(2k) \qquad m(m+1) \text{ is even by a previous example, } k \in \mathbb{Z} && (4)\\ &= 8k. && (5) \end{aligned}$$

Thus $8|(x^2-1)$. Therefore, we find if 8 does not divide x^2-1 then x is even by proof by contraposition.$\square$

The conditional sentence $P \Rightarrow Q$ is equivalent to $\sim Q \Rightarrow \sim P$. In particular, we can identify P as the sentence "$x^2 - 1$ does not divide 8" and Q as the sentence "x is even". We proved $\sim Q \Rightarrow \sim P$ in the natural way: we

assumed $\sim Q$ true then argued that $\sim P$ was also true. That establishes that $\sim Q \Rightarrow \sim P$ is true and consequently $P \Rightarrow Q$ follows.

Example 1.50. Let $x, y \in \mathbb{Z}$, if $x + y$ is even, then either x and y are even or x and y are odd.
Proof: see homework solutions page 13 for problem 1.5.3e.□

Remark 1.51. *Please read your text sections 1.4 up through 1.6. I have made an effort to give new examples in these notes, hopefully there is still much to me gained from the text. In particular, the list of tautologies on page 29 can be a useful guide. Remember that we can trade the given problem for one that is equivalent. This is the idea behind proof by contradiction and also contraposition.*

1.4.4 biconditional proofs

To prove a biconditional statement we can prove a conditional statement and its converse, or we can sometimes string together a list of equivalent statements. Example 1.56 gives an example of the string-type proof.

Example 1.52. Let $a \in \mathbb{Z}$ with $a > 0$, a is odd iff $a + 1$ is even
Proof: Suppose a is odd, then there exists $k \in \mathbb{Z}$ such that $a = 2k + 1$. Observe that $a + 1 = 2k + 1 + 1 = 2(k + 1)$ thus $a + 1$ is even.
Conversely suppose that $a + 1$ is even, then there exists $m \in \mathbb{Z}$ such that $a + 1 = 2m$. Observe that $a = 2m - 1 = 2m - 2 + 1 = 2(m - 1) + 1$ thus a is odd. Therefore a is odd iff $a + 1$ is even. □

1.4.5 proofs involving quantifiers; existence proofs

If we just need to show existence then an example will do.

Example 1.53. If $f(n) = 2^n$ then $\exists n \in \mathbb{N}$ such that $f(n)$ is prime
Proof: Observe that $f(1) = 2^1 = 2$ is prime. Thus, there exists $n \in \mathbb{N}$ such that $f(n)$ is prime. □

If we need to show unique existence then we must find an example and then show that any other example is the same one we already found.

Example 1.54. Suppose $m, b \in \mathbb{R}$ such that $m \neq 0$,there is a unique x-intercept to the line $y = mx + b$.
Proof: The intercept has $y = 0$ thus $0 = mx + b$ yielding $x = -b/m$. Suppose that x_2 is another x-intercept then $0 = mx_1 + b$, but then $x_1 = -b/m$ hence there is just one x-intercept, namely $x = -b/m$.□

Example 1.55. A linear operator from $\mathbb{R}^n$ to $\mathbb{R}^n$ is a function $L : \mathbb{R}^n \to \mathbb{R}^n$ such that $L(x+y) = L(x) + L(y)$ and $L(cx) = cL(x)$ for all $x, y \in \mathbb{R}^n$ and $c \in \mathbb{R}$. Prove that for each such L, there exists a unique matrix A such that $L(x) = Ax$ where x is in column vector notation.
Proof: Let $\{e_i\}$ be the standard basis for $\mathbb{R}^n$ where $i = 1, 2, \ldots n$. Since $\{e_i\}$ is a basis there must exist A_{ij} such that

$$L(e_i) = \sum_{j=1}^{n} A_{ji} e_j$$

Moreover, for each $x \in \mathbb{R}^n$ we can write $x = \sum_{i=1}^{n} x_i e_i$ and note,

$$L(x) = L(\sum_{i=1}^{n} x_i e_i) = \sum_{i=1}^{n} x_i \sum_{j=1}^{n} A_{ji} e_j = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ji} x_i e_j$$

Hence, $L(x)_j = \sum_{i=1}^{n} A_{ji} x_i$. Thus there exists a matrix $A = [A_{ij}]$ such that $L(x) = Ax$. Moreover, this matrix is unique by virtue of it's construction. If $L(x) = Bx$ then $Ax = Bx$ for all $x \in \mathbb{R}^n$ thus $A = B$. **I'm misbehaving a bit here, this belongs in linear algebra where this argument can be made more carefully, I just think it is is a nice example of showing existence through constructive proof** □

Example 1.56. A linear operator from $\mathbb{R}^n$ to $\mathbb{R}^n$ such that $L(x) \cdot L(y) = x \cdot y$ for all $x, y \in \mathbb{R}^n$ is called an orthogonal transformation. Prove that $L : \mathbb{R}^n \to \mathbb{R}^n$ is orthogonal iff the matrix for $L : \mathbb{R}^n \to \mathbb{R}^n$ satisfies $A^t A = I$.
Proof: Recall that we can calculate $x \cdot y$, the dot-product of $x, y \in \mathbb{R}^n$ by the product of the row vector for x and the column vector for y; $x \cdot y = x^t y$. Here x^t is the transpose of x. Furthermore, let A denote the matrix of L; $L(x) = Ax$ for each $x \in \mathbb{R}^n$. Observe,

$$\begin{aligned}
L(x) \cdot L(y) = x \cdot y &\Leftrightarrow (Ax)^t (Ay) = x^t y \qquad \forall x, y \in \mathbb{R}^n \\
&\Leftrightarrow x^t (A^t A) y = x^t I y \qquad \forall x, y \in \mathbb{R}^n \\
&\Leftrightarrow x^t (A^t A - I) y = 0 \qquad \forall x, y \in \mathbb{R}^n \\
&\Leftrightarrow A^t A - I = 0 \\
&\Leftrightarrow A^t A = I.
\end{aligned}$$

It should be no surprise to you by now that I am misbehaving again. Examples of orthogonal transformations are rotations and reflections. □

In truth, most of the proofs we have thus far discuused were "for all" proofs since we argued some property for an arbitrary object or pair of objects. The exceptions to this, prior to this subsection, were examples 1.48 and 1.47. Both of those proofs gave a proof of non-existence:

1.) There does not exist a point on both the circle and the line
2.) There does not exist a rational number which is equal to $\sqrt{2}$

If you examine those proofs you'll see that we proved those assertions by contradiction. In particular, $\sim [(\exists x)P(x)] \Leftrightarrow (\forall x)(\sim P(x))$.
Let's look at a proof that involves both $\exists$ and $\forall$.

Example 1.57. All functions from $\mathbb{R}$ to $\mathbb{R}$ can be written as the sum of an even function and an odd function
Proof: Let me restate the claim: for each function f, there exists an odd function f_o and an even function f_e such that $f = f_o + f_e$. Here I can give a constructive proof. Given the function f, define

$$f_e(x) = \frac{1}{2}\Big(f(x) + f(-x)\Big) \qquad\qquad f_o(x) = \frac{1}{2}\Big(f(x) - f(-x)\Big)$$

Clearly, $f(x) = f_o(x) + f_e(x)$ for each $x \in \mathbb{R}$ thus $f = f_o + f_e$. Moreover, we can check that $f_e(-x) = f_e(x)$ and $f_o(-x) = -f_o(x)$ for each $x \in \mathbb{R}$. $\square$

Notice, that we cannot prove the assertion by just taking a particular example and showing it works. It is true that $f(x) = x + x^2$ has $f_e(x) = x^2$ and $f_o(x) = x$ but that is not enough to prove the claim in the example. It was important we took an arbitrary function. By the way, another fun example is

$$e^x = \cosh(x) + \sinh(x)$$

here $\cosh(x)$ is even while $\sinh(x)$ is odd and e^x is neither even nor odd.

Remark 1.58. *Page 50-51 of the text have useful comments about how $\exists$ and $\forall$ interface with compound propositions. I hope you can avoid 1-4 on page 51 in particular.*

1.4.6 necessary conditions verses sufficient conditions

A condition is said to be "necessary" when it is needed for the conclusion to follow. A condition is said to be "sufficient" when the conclusion follows automatically on the basis of that condition alone. However, every sufficient condition is not necessarily a necessary condition because the sufficient condition might contain extra uneeded assumptions.

Example 1.59. *Consider* $s = \sum_{n=0}^{\infty} c_n$. *The condition* $\lim_{n\to\infty} c_n = 0$ *is a necessary condition for the series* s *to converge. However, it is not sufficient since* $\sum_{n=0}^{\infty} \frac{1}{n}$ *diverges. A sufficient condition for* s *to converge is that the limit of the partial sums of* s *converges. Another sufficient condition would be that the series is alternating and the postive terms go to zero. In calculus II we learned a number of criteria that were sufficient for the series to converge.*

Example 1.60. *Differentiability of a function at a point is a sufficient condition for continuity of a function at that point. However, differentiability is not a necessary condition since* $f(x) = |x|$ *is continuous and not differentiable at zero.*

There are conditions which are both necessary and sufficient conditions. Every good definition has this form. A well-thought out defintion gives the conditions which are necessary and sufficient to encapsulate an idea. For example, later in this course we will learn what conditions are necessary and sufficient conditions on an abstract set in order that it be called a *group*.

2 Set Theory and Induction

Brilliant mathematicians have spent their lifetimes refining the nuanances of axiomatic set theory. It is a story involves Cantor, Russel, Hilbert, Zermelo, Fraenkel, Godel and Cohen and a host of others too numerous to name. Cantor proved many interesting results in the late 19-th century. Particularly, Cantor developed ideas about infinity that boggle the mind. Cantor gave arguments that led to a string of infinities each one larger than the last. However, Cantor built his theory on an intuitive foundation. In short, Cantor failed to rule out the "set of all sets" begin a set. This is known as Russel's paradox, it can be restated as:

"A barber in a certain town has stated he will cut the hair of all those persons and only those persons who do not cut their own hair. Does the barber cut his own hair?"

In response to this problem, Zermelo provided a set of axioms on which to base set theory. Later, Fraenkel refined Zermelo's axioms further. The axiom of choice found a prominent place among these axioms. Moreover, the axiom of choice proved useful in many new branches of mathematics such as infinite dimensional linear algebra, topology, algebraic topology, ring theory and so forth... From a pragmatical perspective, many mathematicians would be reluctant to give up on the axiom of choice. However, another camp of mathematicians had deep misgivings about the axiom and Zermelo's axioms in general.

Hilbert had also brought forth a set of axioms to describe geometry. However, in contrast to Zermelo he proved his axioms were consistent with something else which was known. Zermelo didn't bother, instead he claimed they could probably be proved consistent with further effort. Around 1930 Godel put forth a series of powerful results. Godel showed that it was impossible to prove consistency. However, Godel did prove in 1935 that the axioms were "relatively consistent". Godel also showed that the continuum hypothesis was relatively consistent with Zermelo-Fraenkel axioms. Finally, in 1963 Paul Cohen was able to show that the axiom of choice and the continuum hypothesis are independent relative to Zermelo-Fraenkel axiomatic set theory (without the axiom of choice).

What this means is that there is not just one "set-theory". Rather, there are multiple axiom systems which put forth differing set theories. All of

this said, we will not dwell on questions which cut so finely. I just wanted to give you a little sketch of some of the history here. My source for this discussion was section 18.1 of Victor Katz' text *A History of Mathematics*, 2nd-ed. If you are interested in such things, its a nice source and it actually does have some math in it, not to mention antiquated notations. It's a little dry though. Other books worth perusing for this sort of thing: *Men of Mathematics* by E.T. Bell, *Godel Escher Bach* by Douglas Hofstadter, and *Road to Reality* by Roger Penrose. Penrose's book has deep and insightful discussions of what calculus and advanced mathematics.

2.1 elements and subsets

We say that a set is a collection of elements. If $x \in S$ then we say that x is an element of S. The set $\emptyset$ is the set with no elements. A set with elements a, b, c is denoted { a,b,c }. Unless otherwise stated, no ordering is assumed and we keep only one of each distinct element. If $a = b$ then the set with elements a, b, c would just be $\{a, c\}$ or you could say $\{b, c\}$.

Definition 2.1. *Let A, B be sets, if $x \in A$ implies $x \in B$ for all $x \in A$ then we say that A is a subset of B and write $A \subseteq B$. In other words,*

$$\boxed{A \subseteq B \;\Leftrightarrow (\forall x)(x \in A \Rightarrow x \in B).}$$

Notice that $\emptyset \subseteq A$ for any set A. This follows trivially from the definition since there does not exist $x \in \emptyset$.

Proposition 2.2. *Let A be a set, then $A \subseteq A$.*

Proof: Let $x \in A$ then $x \in A$ therefore, $A \subseteq A$. $\square$

Definition 2.3. *Let A, B be sets, we say that $A = B$ iff $A \subseteq B$ and $B \subseteq A$.*

$$\boxed{A = B \;\Leftrightarrow (A \subseteq B) \wedge (B \subseteq A).}$$

Proposition 2.4 (this is Theorem 2.3 in the text)**.** *Let A, B be sets with no elements, then $A = B$.*

Proof: $(\forall x)(x \in A \Rightarrow x \in B)$ is trivially true since the antcedent $x \in A$ is false. Thus $A \subseteq B$. By symmetry $B \subseteq A$ thus $A = B$. $\square$

The proposition above shows that the empty set is unique. When we say $\emptyset$ there is no ambiguity in what is meant.

Example 2.5 (Constructing Natural Numbers)**.** *The natural numbers are also called counting numbers for the obvious reason. Let me sketch how you can construct* $\mathbb{N}$ *from set theory:*

$$\begin{aligned} 0 &\approx \emptyset \\ 1 &\approx \{\emptyset\} \\ 2 &\approx \{\emptyset, \{\emptyset\}\} \\ 3 &\approx \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\} \end{aligned}$$

Then addition in $\mathbb{N}$ *is defined in a somewhat subtle fashion. We need a few more tools before we can discuss it properly. It is not simple unioning of sets, that is why I got confused in lecture when I mentioned this idea the first time.*

Example 2.6. *Other important sets to know about are:*

$$\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$$

$$\mathbb{Q} = \{\frac{p}{q} \mid p, q \in \mathbb{Z},\ q \neq 0\}$$

$$\mathbb{R} = \{x \mid \ x \text{ is a real number}\}$$

$$\mathbb{C} = \{a + ib \mid a, b \in \mathbb{R}\}$$

Also of interest,

$$\begin{aligned} \textit{open interval} \qquad & (a, b) = \{x | a < x < b\} \\ \textit{half-open interval} \qquad & (a, b] = \{x | a < x \leq b\} \\ \textit{half-open interval} \qquad & [a, b) = \{x | a \leq x < b\} \\ \textit{closed interval} \qquad & [a, b] = \{x | a \leq x \leq b\} \end{aligned}$$

We also have occasion to use $(-\infty, \infty) = \mathbb{R}$*. Fortunately, there is rarely danger of confusion between the point* (a, b) *and the interval* (a, b)*. These are very different objects yet we use the same notation for both. The point* $(a, b) \in \mathbb{R}^2$ *whereas the interval* $(a, b) \subseteq \mathbb{R}$*.*

2.2 subsets and the power set

Given a particular set we can find all possible subsets of that set.

Example 2.7. *Suppose $S = \{a, b, c\}$. We observe that S has subsets:*

$$\emptyset,\ \{a\},\ \{b\},\ \{c\},\ \{a, b\},\ \{a, c\},\ \{b, c\},\ \{a, b, c\}.$$

A set with three elements apparently has $2^3 = 8$ subsets.

Definition 2.8 (Power Set)**.** *Given a set S the power set of S is the set of all subsets of S. We denote the power set of S by $\mathcal{P}(S)$. Symbolically,*

$$\boxed{U \in \mathcal{P}(S) \ \Leftrightarrow\ U \subseteq S}$$

The power set of a given set is bigger than the set. In some sense, finding the power set is like exponentiating the set, crudely speaking.

Example 2.9. *Suppose $S = \{a, b, c\}$. The power set of S is:*

$$\mathcal{P}(S) = \{\emptyset,\ \{a\},\ \{b\},\ \{c\},\ \{a, b\},\ \{a, c\},\ \{b, c\},\ \{a, b, c\}\}.$$

Each element of the power set $\mathcal{P}$ is actually a subset of S.

Example 2.10 (fun with the empty set)**.** *There is a big difference between $\emptyset$ and $\{\emptyset\}$. The empty set has no elements, however the set containing the empty set has one element.*

$$\mathcal{P}(\emptyset) = \{\emptyset\}$$
$$\mathcal{P}(\mathcal{P}(\emptyset)) = \{\emptyset, \{\emptyset\}\}$$
$$\mathcal{P}(\mathcal{P}(\mathcal{P}(\emptyset))) = \{\emptyset, \{\emptyset\}, \{\{\emptyset\}\}, \{\emptyset, \{\emptyset\}\}\}.$$

Notice that the empty set has no elements and the number of elements in $\mathcal{P}(\emptyset)$ is simply $2^0 = 1$; $n(\mathcal{P}(\emptyset)) = 1$ where $N(S)$ denotes the number of elements in the set S. (the text uses $\bar{S}$ for $N(S)$. Likewise,

$$N(\mathcal{P}(\mathcal{P}(\emptyset))) = 2^1 = 2$$
$$N(\mathcal{P}(\mathcal{P}(\mathcal{P}(\emptyset)))) = 2^2 = 4$$

You might guess that if $N(S) = n$ then $N(\mathcal{P}(S)) = 2^n$. You would be correct, I prove this in the homework solutions via induction. (we'll discuss that in a lecture or two from now)

Definition 2.11 (proper subset)**.** *Let S be a set, $U \subseteq S$ is called a proper subset if $U \neq S$.*

One funny case is $S = \emptyset$, if $U \subseteq \emptyset$ then $U = \emptyset$ thus $\emptyset$ has no proper subsets. If $S \neq \emptyset$ then $\emptyset$ is a proper subset of S. The power set has all proper subsets plus the set it self.

2.3 set operations

You are probably already familar with unions and intersections. We use them in calculus in discussing domains and ranges etc... For example,

$$[0,1] \cup [1,2] = [0,2], \qquad (0,2] \cup [1,3) = (0,3), \qquad (0,2) \cap (1,3) = (1,2)$$

A point is in the union if it is in either set being unioned whereas a point is in the intersection if the point is in both sets. These are very general concepts,

Definition 2.12 (union)**.** *Let A, B be sets then we define the* **union** *of A and B as follows:*

$$\boxed{A \cup B = \{x | x \in A \text{ or } x \in B\}}$$

Definition 2.13 (intersection)**.** *Let A, B be sets then we define the* **intersection** *of A and B as follows:*

$$\boxed{A \cap B = \{x | x \in A \text{ and } x \in B\}}$$

Definition 2.14 (difference)**.** *Let A, B be sets then we define the* **difference** *of A and B as follows:*

$$\boxed{A - B = \{x | x \in A \text{ and } x \notin B\}}$$

Definition 2.15 (disjoint)**.** *Sets A and B are* **disjoint** *if $A \cap B = \emptyset$.*

Example 2.16. *Let $A = \{1,2,3\}$, $B = \{0,1,2\}$ and $C = \{4,5\}$ then*

$$\begin{aligned} A \cup B &= \{0,1,2,3\} \\ A \cap B &= \{1,2\} \\ A - B &= \{3\} \\ B - A &= \{0\} \\ A \cap C &= B \cap C = \emptyset \end{aligned}$$

We see that C is disjoint from A and B.

Examples are not nearly as fun as theorems. There are many things we can say in general for set operations.

Theorem 2.17 (set properties: Theorems 2.2 and 2.6 in text)**.** *Let A, B and C be sets,*

2.2. *If* $A \subseteq B$ *and* $B \subseteq C$ *then* $A \subseteq C$.

a. $A \subseteq A \cup B$.

b. $A \cap B \subseteq A$.

c. $A \cap \emptyset = \emptyset$.

d. $A \cup \emptyset = A$.

e. $A \cap A = A$.

f. $A \cup A = A$.

g. $A \cup B = B \cup A$.

h. $A \cap B = B \cap A$.

i. $A - \emptyset = A$.

j. $\emptyset - A = \emptyset$.

k. $A \cup (B \cup C) = (A \cup B) \cup C$.

l. $A \cap (B \cap C) = (A \cap B) \cap C$.

m. $A \cap (B \cup C) = (A \cap B) \cup (A \cap C$.

n. $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.

o. $A \subseteq B$ *iff* $A \cup B = B$.

p. $A \subseteq B$ *iff* $A \cap B = A$.

q. *If* $A \subseteq B$, *then* $A \cup C \subseteq B \cup C$.

r. *If* $A \subseteq B$, *then* $A \cap C \subseteq B \cap C$.

Proof: follows quickly from the definition of subsets, unions, differences and intersection in each case. □.

The concept of set difference is sometimes taken relative to a universal set U, in that case the difference is called the *complement* of a set:

Definition 2.18. *If* U *is the universe and* $B \subseteq U$ *then the* **complement** *of* B *relative to* U *is* $\widetilde{B} = U - B$.

Example 2.19. *Let $U = \mathbb{R}$ then $\widetilde{[0,1]} = (-\infty, 0) \cup (1, \infty)$. Another fun example: $\widetilde{\emptyset} = \mathbb{R}$. The complement of the rational numbers $\mathbb{Q}$ is $\widetilde{\mathbb{Q}} = \mathbb{R} - \mathbb{Q}$ the set of irrational numbers. We proved that $\sqrt{2} \in \widetilde{\mathbb{Q}}$.*

Theorem 2.20 (this is Theorem 2.7 in the text)**.** *Let U be the univers and $A, B \subseteq U$,*

a. $\tilde{\tilde{A}} = A$.

b. $A \cup \tilde{A} = U$.

c. $A \cap \tilde{A} = \emptyset$.

d. $A - B = A \cap \tilde{B}$.

e. $A \subseteq B$ *iff* $\tilde{B} \subseteq \tilde{A}$.

f. $\widetilde{A \cup B} = \tilde{A} \cap \tilde{B}$.

g. $\widetilde{A \cap B} = \tilde{A} \cup \tilde{B}$.

h. $A \cap B = \emptyset$ *iff* $A \subseteq \tilde{B}$.

Parts f, g are known as De Morgan's Laws.

Proof: follows quickly from the definitions of subsets, unions, differences and intersection in each case. We'll work out some in lecture. A Venn diagram is a good place to start some of the proofs, it helps guide the proof $\square$.

2.4 families of sets

A family of sets is simply a set of sets. For example, the power set of S is the family of all subsets of S. Often it is convenient to label the sets in the family by elements from an *indexing* set.

Definition 2.21. *Let $\Delta \neq \emptyset$. If there exists a set A_α for each $\alpha \in \Delta$ then $\mathcal{A} = \{A_\alpha | \alpha \in \Delta\}$ is a family of sets indexed by Δ. The elements of Δ are called* **indices** *in this context.*

Example 2.22. *Let $A_n = (0, n)$ for each $n \in \mathbb{N}$. The family of sets $\{(0,1), (0,2), \dots\} = \{(0, n) | n \in \mathbb{N}\}$ is indexed by the natural numbers $\mathbb{N}$.*

We can define unions and intersections over families of sets:

Definition 2.23. *Let $\mathcal{A}$ be a family of sets, the union over $\mathcal{A}$ is,*

$$\bigcup_{A\in\mathcal{A}} A = \{x \mid \exists B \in \mathcal{A} \text{ such that } x \in B\}$$

If A has index set Δ we can write the union over $\mathcal{A}$ as,

$$\bigcup_{\alpha\in\Delta} A_\alpha = \{x \mid \exists \beta \in \Delta \text{ such that } x \in A_\beta\}$$

Finally, if we have the case $\Delta = \mathbb{N}$ then we can write the union over $\mathcal{A}$ as,

$$\bigcup_{i=1}^{\infty} A_i = \{x \mid \exists j \in \mathbb{N} \text{ such that } x \in A_j\}$$

We also may use the same notation for other subsets of $\mathbb{Z}$. (could run the index to negative values, or starting at zero, two or whatever is convenient to the problem)

Example 2.24. *Let $A_n = (0, n)$ for each $n \in \mathbb{N}$,*

$$\bigcup_{n=1}^{\infty} = (0, \infty)$$

I'll give a number-line explaination in lecture.

Definition 2.25. *Let $\mathcal{A}$ be a family of sets, the intersection over $\mathcal{A}$ is,*

$$\bigcap_{A\in\mathcal{A}} A = \{x \mid x \in B \text{ for each } B \in \mathcal{A}\}$$

If A has index set Δ we can write the intersection over $\mathcal{A}$ as,

$$\bigcap_{\alpha\in\Delta} A_\alpha = \{x \mid x \in A_\beta \text{ for each } \beta \in \Delta\}$$

Finally, if we have the case $\Delta = \mathbb{N}$ then we can write the intersection over $\mathcal{A}$ as,

$$\bigcap_{i=1}^{\infty} A_i = \{x \mid x \in A_j \text{ for each } j \in \mathbb{N}\}$$

We also may use the same notation for other subsets of $\mathbb{Z}$. (could run the index to negative values, or starting at zero, two or whatever is convenient to the problem)

Example 2.26. *Let $A_n = (0, n)$ for each $n \in \mathbb{N}$,*

$$\bigcap_{n=1}^{\infty} = (0, 1)$$

I'll give a number-line explaination in lecture.

2.5 introduction to topology

Topology is the study of continuity. This branch of mathematics was refined in the first half of the twentieth century. Topological concepts lie at the base of most modern geometrical research. In short, a topology tells you what the open sets for a space are. The axiomatic and abstract definition given below is due to Riesz and Hausdorff. Most graduate students in mathematics will be required to take a course or two in topology.

Definition 2.27 (definition of topology)**.** *A topology $\mathcal{T}$ for a set S is a family of subsets of S such that,*

i. $\mathcal{T}$ *must contain* S *and* $\emptyset$*.*

ii. $\mathcal{T}$ *is closed under any union of its elements,*

$$\bigcup_{U \in \mathcal{T}} U \in \mathcal{T}$$

iii. $\mathcal{T}$ *is closed under finite intersections of its elements. If* $U_1, \ldots U_n \in \mathcal{T}$ *then*

$$\bigcap_{i=1}^{n} U_i \in \mathcal{T}$$

The sets in $\mathcal{T}$ *are defined to be* **open**. *Moreover, a set is defined to be* **closed** *if its complement relative to* S *is open. A set* S *paired with a topology* $\mathcal{T}$ *is called a* **topological space**.

I hope you are familar with *open intervals* and *closed intervals* of $\mathbb{R}$. The open intervals are the simplest type of open set in $\mathbb{R}$. We could define the standard topology on $\mathbb{R}$ by letting $\mathcal{T}$ be the empty set together with collection of all open intervals and their unions. Since the finite intersection of open intervals is again a union of open intervals, or the empty set, we will satisfy the three axioms for a topology. Notice that,

$$\mathbb{R} - [a, b] = (-\infty, a) \cup (b, \infty)$$

thus the complement of a closed interval is open. This means that the closed interval $[a, b]$ is in fact a closed set. These bizarre axioms will recover all the ordinary geometric definitions of open and closed with which we are more familar. The definintion above provides the basis for the field of *Point-Set Topology*. The other way to define open and closed sets is in terms of a metric. The concept of a metric space predates topology.

Definition 2.28. *A metric, or distance function, on a space M is a function $d : M \times M \to \mathbb{R}$ such that for $x, y, z \in M$,*

i. *d is postive definite; $d(x, y) \geq 0$ and $d(x, x) = 0$ iff $x = 0$.*

ii. *d is symmetric; $d(x, y) = d(y, x)$.*

iii. *d satisfies triangle inequality; $d(x, y) + d(y, z) \leq d(x, z)$*

A space M together with a distance function is called a **metric space**

You can verify that $\mathbb{R}$ has the distance function $d(a, b) = |b-a|$. This means that $\mathbb{R}$ is a metric space. Every metric space can be given the structure of a topological space via the *metric topology*. You will learn about that in the real analysis course here at LU. The standard topology on $\mathbb{R}$ is the metric topology which is generated from $d(a, b) = |b - a|$.

Metric spaces are quite special. Many sets do not have a natural idea of distance. However, we can still give them a topological structure.

Example 2.29. *Let X be a nonempty set. Define $\mathcal{T} = \{X, \emptyset\}$. This provides a topology on X called the* **discrete topology**. *Axiom i. is satisfied. Then note*

$$X \cup \emptyset = X \in \mathcal{T}$$

$$X \cap \emptyset = \emptyset \in \mathcal{T}$$

Thus axioms ii. and iii. are satisfied. The set X could be most anything. For example, $X = \mathbb{R}$. With respect to the discrete topology, the set $(0, 1)$ is not open since $(0, 1) \notin \mathcal{T}$. There are many topologies available for a given space and they are not always compatible.

The power of topology is that it allows concrete definitions of very intuitive geometric terms such as *compact, disconnected* and *simply connected*. The topological definition of a continuous function states that a function is continuous if the inverse image of open sets is open for the function. We will work with that definition a little later when we discuss functions. If you'd like to learn more about topology or metric spaces then ask me sometime, I can recommend a few books.

2.6 weak induction (PMI)

If we wish to prove a statement based on $n \in \mathbb{N}$ holds for all $n \in \mathbb{N}$ we can use the Principle of Mathematical Induction (PMI). PMI follows from the very properties or construction of the natural numbers.

Theorem 2.30 (Principle of Mathematical Induction)**.** *If S is a subset of $\mathbb{N}$ with the following two properties:*

i. $1 \in S$.

ii. *for all $n \in \mathbb{N}$, if $n \in S$ then $n+1 \in S$.*

then $S = \mathbb{N}$.

A set $S \subseteq \mathbb{N}$ with the properties above is called an **inductive set**. If we wish to prove that a proposition $P(n)$ is true for all $n \in \mathbb{N}$ we can procees as follows:

(1.) Let $S = \{n \in \mathbb{N} \mid P(n) \text{ is true } \}$

(2.) Show $1 \in S$; that is, show $P(1)$ is true.

(3.) For all $n > 1$, show $n \in S$ implies $n+1 \in S$; that is, show $P(n) \Rightarrow P(n+1)$ for arbitrary $n > 1$.

(4.) By PMI, $S = \mathbb{N}$; that is, by PMI $P(n)$ is true for all $n \in \mathbb{N}$

The explicit mention of the inductive set S is not necessary. In practice, I usually just refer to the proposition $P(n)$.

Example 2.31 (child-Gauss' observation)**. Show that**

$$1 + 2 + \cdots + n = \frac{n(n+1)}{2}.$$

Proof: Define the proposition above to be $P(n)$, we seek to show it is true for all $n \in \mathbb{N}$ Proceed by PMI. Observe $P(1) = \frac{1(2)}{2} = 1$ thus the induction hypothesis is true for n=1. Assume $P(n)$ is true for an arbitrary $n > 1$,

$$1 + 2 + \cdots + n = \frac{n(n+1)}{2}.$$

Add $n+1$ to both sides of the above equation,

$$1 + 2 + \cdots + n + (n+1) = \frac{n(n+1)}{2} + n + 1.$$

Now make a common denominator on the rhs,

$$1 + 2 + \cdots + n + (n+1) = \frac{n^2 + 3n + 2}{2} = \frac{(n+1)(n+1+1)}{2}.$$

The equation above shows $P(n+1)$ is true. Hence we have shown $P(n) \Rightarrow P(n+1)$ for any $n > 1$. Therefore, by PMI, $1+2+\cdots+n = \frac{n(n+1)}{2}$ for all $n \in \mathbb{N}$

Induction is also used to give *recursive* defintions.

Definition 2.32 (factorial)**.** *We define $n!$ inductively as follows:*

$$(1.)1! = 1,\ n! = n(n-1)!$$

This is equivalent to the non-inductive definition

$$n! = n(n-1)(n-2)\cdots(3)(2)(1).$$

It is also customary to define $0! = 1$.

It would seem natural to allow the induction to start at $n = 0$ or even $n = -5$ and if we still could prove $P(n) \Rightarrow P(n+1)$ for all $n > -5$ then it sure seems intuitive that the proposition $P(n)$ should be true for all $n \geq -5$. In other words, PMI ought to extend to subsets of $\mathbb{Z}$ provided the subset has a finite boundary.

Theorem 2.33 (General General PMI)**.** *If $Q(z)$ is a proposition where $z \in \mathbb{Z}$ and $s \in \mathbb{Z}$ and (1.) $Q(s)$ is true,*

(2.) for all $n > s$, $Q(n) \Rightarrow Q(n+1)$ for arbitrary $n > s$,

then $Q(n)$ is true for all $n \geq s$.

Proof: Let $P(n-s+1) = Q(n)$. Observe that $1 = n-s+1$ gives $n = s$ thus $P(1) = Q(s)$ hence $P(1)$ is true. Observe that $P(n-s+1) = Q(n)$ gives $P(m) = Q(m+s-1)$ generally. Assume $P(k)$ true for some $k > 1$. Then $P(k) = Q(k+s-1)$ is true and $k > 1$ implies $k+s-1 > s$ which implies by assumption (2.) of the Theorem $Q(k+s-1+1)$ is true (where we substitute $n = k+s-1$ to use (2.)). Hence, $Q(k+1+s-1) = P(k+1)$ is true for all $k > 1$. Therefore, by PMI, $P(n)$ is true for all $n \in \mathbb{N}$. Consequently, $Q(n)$ is true for all $n \geq s$. $\square$

Its not hard to see that induction will also hold for sets which have an upper bound in $\mathbb{Z}$ and go to $-\infty$. I'm disappointed the text doesn't say more about this. The idea I just used in this proof is known as *shifting indices*.

In calculus II when we study series and sequences we use the same idea to move the starting point of the sequence or series around. Although we typically state results for n starting at $n = 0$ or $n = 1$ we can just as well shift the starting point to some other integer. Questions of divergence and convergence really only depend on the tail of the series. More or less the same idea holds here. Induction has to start somewhere, but so long as the sucessor property holds for all integers to the right of the starting point we will find the statement is true for all the integers in question. I am going to call my principle of induction **GGPMI** just to be annoying.

I have a number of easy examples in the homework solution. I am sticking to harder things in this particular stretch of the notes.

Example 2.34. *Work through problem 8i of section 2.4 which is given on page 25 of my homework solutions.*

Theorem 2.35 (binomial theorem)**.** *Let $a, b \in \mathbb{C}$ then*

$$(a+b)^n = a^n + na^{n-1}b + \cdots + \binom{n}{k}a^{n-k}b^k + \cdots + nab^{n-1} + b^n$$

In summation notation,

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k}b^k$$

Where

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Proof: Let $P(n)$ be the binomial theorem. We use induction starting at $n = 0$. Observe that,

$$(a+b)^0 = 1$$

Also,

$$\binom{0}{0} = \frac{0!}{0!(0-0)!} = \frac{1}{1} = 1.$$

Thus $P(0)$ is true. Assume $P(n)$ is true. For a particular $n \in \mathbb{N}$, assume that

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k}b^k. \tag{6}$$

We wish to show,

$$(a+b)^{n+1} = \sum_{k=0}^{n+1} \binom{n+1}{k} a^{n+1-k} b^k$$

A short calculation reveals that for $k = 1, 2, \ldots, n$

$$\binom{n+1}{k} = \binom{n}{k-1} + \binom{n}{k}$$

Whereas for, $k = 0$,

$$\binom{n+1}{0} = \frac{(n+1)!}{0!(n+1)!} = 1$$

Or for $k = n+1$,

$$\binom{n+1}{n+1} = \frac{(n+1)!}{(n+1)!(n+1-(n+1))!} = 1$$

Thus we need to show,

$$\begin{aligned}(a+b)^{n+1} &= a^{n+1} \; + \; \sum_{k=1}^{n} \left[\binom{n}{k-1} + \binom{n}{k} \right] a^{n+1-k} b^k \; + \; b^{n+1} \\ &= a^{n+1} + \sum_{k=1}^{n} \binom{n}{k-1} a^{n+1-k} b^k + \sum_{k=1}^{n} \binom{n}{k} a^{n+1-k} b^k + b^{n+1} \quad (7)\end{aligned}$$

Multiply the induction hypothesis 6 by $(a+b)$,

$$\begin{aligned}(a+b)^{n+1} &= \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k (a+b) \\ &= \sum_{k=0}^{n} \binom{n}{k} a^{n-k+1} b^k + \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^{k+1} \\ &= a^{n+1} + \sum_{k=1}^{n} \binom{n}{k} a^{n-k+1} b^k + \sum_{k=1}^{n-1} \binom{n}{k} a^{n-k} b^{k+1} + b^{n+1} \\ &= a^{n+1} + \sum_{k=1}^{n} \binom{n}{k} a^{n-k+1} b^k + \sum_{p=0}^{n} \binom{n}{p-1} a^{n-(p-1)} b^{p-1+1} + b^{n+1} \\ &= a^{n+1} + \sum_{k=1}^{n} \binom{n}{k} a^{n-k+1} b^k + \sum_{k=0}^{n} \binom{n}{k-1} a^{n+1-k} b^k + b^{n+1} \quad (8)\end{aligned}$$

Comparing Equations 7 and 8 we see that the induction hypothesis for n yields the induction hypothesis for $n+1$. Thus, by PMI, the binomial theorem holds for all $n \in \mathbb{N} \cup \{0\}$. The exceptional cases of $k = 0$ and $k = n$ correspond to the edges of Pascal's triangle. Also the main calculational observation is nothing more than the formula which makes Pascal's triangle work. This proof, is not particularly easy, obviously I expect you be able to do problems like the homework. I included this proof because I think it is healthy to see some harder things from time to time. If time allows we may work through a few additional induction examples from the homework solutions.

2.7 strong induction (PCI)

Strong or complete induction is another equivalent form of induction. Let me state the Principle of Complete Induction(PCI):

(1.) Let $S = \{n \in \mathbb{N} \mid P(n) \text{ is true}\}$

(2.) For all $n \geq 1$, show $\{1, 2, \ldots, n\} \subset S$ implies $n+1 \in S$ that is, show $P(k)$ true for all $k \leq n$ implies $P(n+1)$ true.

(3.) By PCI, $S = \mathbb{N}$; that is, by PCI $P(n)$ is true for all $n \in \mathbb{N}$

The difference between PCI and PMI is that for PCI we get to assume that the induction hypothesis is true for all natural numbers leading up to n. In contrast, PMI says we assume $P(n)$ is true for just n then we had to show $P(n+1)$ was true. PMI also required us to show $P(1)$ was true, in contrast PCI instead requires that we show $1 \in S$ implies $2 \in S$. Although, PCI sometimes requires different arguments for different ranges of n (this is also true for PMI but most examples don't exhibit this subtlety). Both PCI and PMI require the implication to be shown for arbitrary n, so if the first few values are special they have to be treated separately. Finally, it should be understood that PCI can also be applied to inductive sets that start at some integer other than one. Again the key is that all integers to the right of the starting point for the inductive set.

Example 2.36. *(Every natural number n ¿ 1 has a prime factor) To begin recall that $p \in \mathbb{N}$ is a prime iff its only factors are 1 and p. Let S be the set of $n \in \mathbb{N}$ such that n has a prime factor and $n > 1$. Clearly $2 \in S$ since 2 is prime and is obviously a factor of 2. Suppose that $\{2, 3, \ldots, m\} \subset S$. Then $2, 3, \ldots m$ all are assumed to have a prime factor. Consider m. If m is prime then we are done since m would be a prime factor of m. Otherwise, by*

definition of prime, there must exist positive integers r, s such that $m = rs$ with $r, s > 1$ but $r, s < m$. Therefore, $r, s \in \{2, 3, \ldots, m\}$ which implies r, s have prime factors. Consequently, m has prime factors. using the fact that if $p|r$ and $q|s$ then $p|rs$ and $q|rs$, we have $m = rs$

The Fundamental Theorem of Arithmetic states that every positive integer can be factored into a unique (upto ordering) product of primes. The example we just completed goes a long way towards the Theorem. Take a positive integer n, then $\exists p_1 \in \mathbb{N}$ prime, such that $n = n_1 p_1$ and $n_1 < n$. If n_1 is prime then we are done. Otherwise, apply our example to n_1, we can produce another prime $p_2 \in \mathbb{N}$ such that $n = n_2 p_2 p_1$. If n_2 is prime we are done. Otherwise, we can apply our example to $n_2 \in \mathbb{N}$ and so forth. In each case we will have $n_{k+1} < n_k$ as to say otherwise would violate $n_k = n_{k+1} p_{k+1}$. Observe that,

$$n > n_1 > n_2 > \cdots > n_k$$

I argue that this list has length at most $n-1$. Each n_k is at least one smaller than n_{k-1}. If the list had n factors then n_n would be at most zero, but that is a contradiction since $n_n \in \mathbb{N}$. Hence the list has length less than $n - 1$.

This almost proves the Fundamental Theorem of Arithmetic. Time permitting, we may go over a few examples from the homework solution on PCI.

2.8 well ordering principle (WOP)

The Well Ordering Principle (WOP) states:

Every nonempty subset of N has a smallest element.

This is equivalent to PMI and PCI, see your text for half of the proof.

Example 2.37. *(Every natural number $n > 1$ has a prime factor) We proved this already in Example 2.37 via PCI. Lets see how WOP makes it easier. Let $n > 1$ and suppose n is not prime, thus n is composite. If S is the set of all factors of n not equal to one, then it is nonempty since n is composite. We use the WOP to select p, the smallest factor of n.*

We seek to show p is prime. Suppose p is composite, then there exists r, s such that $p = rs$ such that $r, s > 1$ and $r, s < p$. Observe that $r|p$ and $p|n$ implies that $r|n$ thus r is a factor of n smaller than p. But this is a contradiction since the WOP allowed us to select p that was the smallest factor (not equal to one). Hence, p is prime and n has a prime factor.

The example above is typical of a WOP proof. Often the WOP is applied then contradiction is used to show the smallest element possesses some particular property. Here we saw it was prime. Proofs that use the Well Ordering Principle can require a certain amount of artistry, I am most interested in you gaining a throurough understanding of PMI. It is important to know PCI and WOP can help you when PMI fails to be obvious. The Division Algorithm for $\mathbb{N}$ is important, but I am delaying discussion of it and Theorem 2.15 until my Chapter on Modular Arithmetic. I want to gather all the tools in one place, I'll delay discussion until we have more to play with.

3 Relations

The concept of a relation is quite general. Many of the constructions we have experience with for functions will also make sense for relations. For example, we can think about the graph of a relation, the composite of two relations and the inverse of a relation. We will see that functions are a special type of relation. Finally, equivalence relations generalize the idea of equality. We will see how an equivalence relation gives rise to equivalence classes. Equivalence classes partition the set. In my experience, the concept of an equivalence relation was the most important thing I learned in this course as an undergraduate.

3.1 cartesian products

Definition 3.1. *An ordered pair is a set of two elements with an ordering. We denoted an ordered pair of a and b by (a, b). The technical definition is:*

$$(a, b) = \{a, \{a, b\}\}$$

In exercise 15 of the homework you are asked to show that $(a, b) = (x, y)$ if and only if $a = x$ and $b = y$. You should use the technical definition to complete exercise 15 of § 3.1. Notice the idea of the definition is that the element which is listed alone is the first element. That fixes an order for the pair. This means that $(1, 2)$ is distinguished from $(2, 1)$

Remark 3.2. *Beware that as an ordered pair $(1, 2) \neq \{x \mid 1 < x < 2\}$. There is a notational degeneracy that is unavoidable here. In practice the context of the sentence will inform you as to whether (a, b) is an open interval or an ordered pair. These are very different objects. An ordered pair is a set with two objects. An open interval of real numbers is a set with infinitely many elements.*

Generally, we can take a finite number of elements and form an ordered set that is called a tuple. If we have $n \in \mathbb{N}$ then $(x_1, x_2, \ldots, x_n)$ is an n-tuple. Two tuples can only be equal if they have the same length. We say that $\vec{x} = (x_1, x_2, \ldots, x_n)$ has j-th component x_j . Moreover, if $\vec{y} = (y_1, y_2, \ldots, y_n)$ then we say $\vec{x} = \vec{y}$ if and only if $x_j = y_j$ for each $j = 1, 2, \ldots n$.

Definition 3.3. *Let A and B be sets. The set of all ordered pairs with first component in A and second component in B is denoted $A \times B$. We say that $A \times B$ is the Cartesian product of A and B. To be precise,*

$$\boxed{A \times B = \{(a, b) \mid a \in A \text{ and } b \in B\}}$$

Likewise, the defninition for the nth Cartesian product is

$$\boxed{\times_{i=1}^{n} A_i = A_1 \times \cdots \times A_n = \{(a_1, a_2, \ldots a_n) \mid a_j \in A_j \ \forall j = 1, 2, \ldots n\}.}$$

Example 3.4. *(Cartesian Products are Non-Associative). Let A, B, C be sets. Let $a \in A, b \in B$ and $c \in C$. Observe:*

$$(a, b) \in A \times B$$

$$(b, c) \in B \times C$$

$$(a, b, c) \in A \times B \times C$$

$$(a, (b, c)) \in A \times (B \times C)$$

$$((a, b), c) \in (A \times B) \times C$$

$$(a, a, a, a) \in A \times A \times A \times A$$

$$((a, a), (a, a)) \in (A \times A) \times (A \times A)$$

There are of course obvious correspondances between things like $A \times (B \times C)$ and $(A \times B) \times C$. However, it is nice to be able to distinguish these objects if need be.

The xy-plane is the set of all points (x, y) such that $(x, y) \in \mathbb{R}$. The xy-plane can be identified with the Cartesian product of $\mathbb{R}$ with $\mathbb{R}$, it is customary to denote $\mathbb{R} \times \mathbb{R} = \mathbb{R}^2$. In calculus III we deal with three dimensional space which is conviently described by $\mathbb{R} \times \mathbb{R} \times \mathbb{R} = \mathbb{R}^3$.

The concept of a Cartesian product is used throughout calculus and much of higher math as well. For example, $\mathcal{M}$ is manifold if is locally approximated by $\mathbb{R}^n$. Globally a manifold is curved in general. A two-dimensional manifold is a *surface*. For example, a sphere is a two-dimensional manifold because it is locally approximated by a plane. Manifolds in a nutshell are simply spaces on which one can do calculus. In advanced calculus at LU we will learn how to do calculus in $\mathbb{R}^n$. Of course $\mathbb{R}^n$ is the simplest n-dimensional manifold. When you study manifold theory you will learn how to *lift* calculus on $\mathbb{R}^n$ up to a more abstract manifold. For those of you who have taken calculus III or are currently taking it, a two-dimensional manifold is a parametrized surface. If you are interested in understanding Einstein's General Theory of Relativity then you should work towards gaining a good grasp on manifold theory. Ask me for more details if you are interested.

Another example, a *principle fiber bundle* is a space which locally looks like the Cartesian product of a manifold and a *symmetry group*. Fiber bundles are at the heart of modern physical theory because modern physics is by in large based on the idea of a symmetry group. The Standard Model of particle physics is based on a *gauge theory*. Fiber bundles provide the mathematics for gauge theory.

Theorem 3.5. *(Properties of Cartesian Products, Theorem 3.1 in the text) Let A, B, C, D be sets. Then,*

a. $A \times (B \cup C) = (A \cup B) \times (A \cup C)$.

b . $A \times (B \cap C) = (A \times B) \cap (A \times C)$.

c. $A \times \emptyset = \emptyset$.

d. $(A \times B) \cap (C \times D) = (A \cap C) \times (B \cap D)$

e. $(A \times B) \cup (C \times D) \subseteq (A \cup C) \times (B \cup D)$

Proof: See the text and homework problems. We may prove one or two of these in lecture.

3.2 relations

A relation is simply some subset of a Cartesian product

Definition 3.6. *Let A, B be sets. We say that R is a relation from A to B if $R \subseteq A \times B$. Moreover, we say that xRy if $(x, y) \in R$. If xRy then we say that x is related to y. On the other hand if $(x, y) \notin R$ then we say that x is not related to y. When we say xRy, I will call x the input of the relation and y the output of R. The Domain of the relation is the set of all possible inputs for the relation. Whereas the Range of the relation is the set of all possible outputs for the relation.*

$$Domain(R) = \{x \in A \mid y \in B \text{ such that } xRy\}$$

$$Range(R) = \{y \in B \mid x \in A \text{ such that } xRy\}$$

Finally, if $R \subseteq A \times A$ and $dom(R) = A$ then we say R is a relation on A.

Notice that a relation can have more than one output for a given input. This means there are relations which cannot be thought of as a function. Recall that functions have one output for a given input. Let me begin with a silly example:

Example 3.7. *(people) Let S be the set of all living creatures on earth. We can say that x is R-related to y if both x and y are people. In this sense I am R-related to Obama. In contrast, I am not R-related to Napolean because he's dead. I am also not R-related to my mom's dogs. They may be treated like humans but the fact remains they have tails and other dog parts that necessarily disqualify them from the category of people.*

Another silly example:

Example 3.8. *(unrelated relation) Let S be the set of all living creatures on earth. We say that x is NR-related to y if x is not the direct decendent of y. In this sense I am NR-related to Obama. In contrast, my daughter Hannah is not NR-related to me since she is my direct decendent.*

Whenever we have a relation from $\mathbb{R}$ to $\mathbb{R}$ we can picture the relation in the Cartesian plane. (We can also do the same for relations from $\mathbb{N}$ to $\mathbb{N}$ and other subsets of the real numbers)

Example 3.9. *(circle) Define $R = \{(x, y) \mid x^2 + y^2 = 1\}$. This is a relation from $\mathbb{R}$ to $\mathbb{R}$. The graph(R) is clearly a circle.*

Example 3.10. *(disk) Define $R = \{(x, y) \mid x^2 + y^2 \leq 1\}$. This is a relation from $\mathbb{R}$ to $\mathbb{R}$. The graph(R) is clearly a circle shaded in; that is the graph is a disk.*

Example 3.11. *(plane with a hole) Define $R = \{(x, y) \mid x^2 + y^2 > 1\}$. This is a relation from $\mathbb{R}$ to $\mathbb{R}$. The graph(R) is the plane with the disk deleted. Notice, for example, that $1R0$ however, 2 is not related to 0 because $2^2 + 0^2 = 4 \neq 1$ hence $(2, 0) \notin R$.*

Example 3.12. *(positive lattice) Define $R = \{(x, y) \mid x, y \in \mathbb{N}\}$. This is a relation from $\mathbb{R}$ to $\mathbb{R}$. The graph(R) is a grid of points. Notice that is is not a relation on $\mathbb{R}$ since dom$(R) = \mathbb{N} \neq \mathbb{R}$.*

Example 3.13. *(coordinate grid) Define $R = \{(x, y) \mid x \in \mathbb{Z}, y \in \mathbb{R}\} \cup \{(x, y) \mid x \in \mathbb{R}, y \in \mathbb{Z}\}$. This is a relation from $\mathbb{R}$ to $\mathbb{R}$. The graph(R) is a grid of horizontal and vertical lines.*

There is no end to these geometric examples. Let me give a weirder example:

Example 3.14. *(rational numbers) Define $R = \{(x, y) \mid x, y \in \mathbb{Q}\}$. This is a relation from $\mathbb{R}$ to $\mathbb{R}$. For example, $3/4R134/279$. However, π is not related to anything. This means that points in the xy-plane with x-coordinate π will not be included in the graph of R. However, points with $x = 3.1415 =$*

31415/1000 *will be included in the graph so it is hard to see the holes along* $x = \pi$. *In fact, the graph*(R) *looks like the whole plane. However, it has holes infinitely close to any point you pick. This is a consequence of the fact that there are infinitely many irrational numbers between any two distinct rational numbers.*

3.3 composite relations

Definition 3.15. *(composite relation) Let* R *be a relation from* A *to* B *and let* S *be a relation from* B *to* C. *The composite of* R *and* S *is*

$$S \circ R = \{(a, c) \mid b \in B \text{ such that } (a, b) \in R \text{ and } (b, c) \in S\}$$

Example 3.16. *Let* $R = \{(1, 2), (3, 4), (5, 6), (7, 8)\}$ *this is a relation from* $A = \{1, 3, 5, 7\}$ *to* $B = \{2, 4, 6, 8\}$. *Define* $S = \{(2, 1), (4, 3), (5, 4), (7, 6)\}$. *We see that* S *is a relation from* B *to* A. *Notice that* $S \circ R$ *is a relation from* A *to* A; $S \circ R : A \to B \to A$:

$$S \circ R = \{(1, 1), (3, 3), (5, 5), (7, 7)\}$$

Since $1R2$ *and* $2S1$ *we have* $1S \circ R1$ *and so forth... Likewise we can verify that* $R \circ S$ *is a relation from* B *to* B; $R \circ S : B \to A \to B$:

$$R \circ S = \{(2, 2), (4, 4), (6, 6), (8, 8)\}$$

The relations we just exhibited are known as the **identity relations** *on* A *and* B *respective. We denote* $I_A = S \circ R$ *and* $I_B = R \circ S$. *The relations given in this example are inverses of each other.*

Definition 3.17. *(inverse relation) Given a relation* R *from* A *to* B *we define the inverse relation* R^{-1} *to be the relation from* B *to* A *defined by* $R^{-1} = \{(y, x) \mid (x, y) \in R\}$

Proposition 3.18. *(Theorem 3.2 in text) The inverse relation just defined is a relation as claimed in the definition. Moreover,* $domain(R^{-1}) = range(R)$ *and* $range(R^{-1}) = domain(R)$.

Proof: immediate from definition of R^{-1}.

The concept of an inverse relation is nice in that it avoids some of the rather cumbersome restrictions that come with the idea of an inverse function. We'll get into those restrictions in a week or two, but you probably recall from precalculus courses that in order for the inverse function to exist we

need the function's graph satisfy the horizontal line test. Inverse relations have no such restriction. Notice that we can form the inverse of a relation always, no horizontal line test can rain on our parade here.

Example 3.19. *(inverse relation graph goes sideways) Let* $S = \{(x, \sin(x)) \mid x \in \mathbb{R}\}$. *This is a relation from* $\mathbb{R}$ *to* $\mathbb{R}$. *The graph*(S) *is the graph of the sine function in the xy-plane. The inverse of* S *is* $S^{-1} = \{(sin(y), y) \mid y \in \mathbb{R}\}$. *Consider that graph*$(S^{-1})$ *should be the same as the graph of the sine function except that* $x = \sin(y)$ *instead of* $y = \sin(x)$. *If you think about this for a moment or two you'll see that the graph of* S^{-1} *is the same as the graph of* S *just instead of running along the x-axis it runs up the y-axis.*

The fact that $graph(S^{-1})$ fails the vertical line test goes to show it is not a function. The graph of the inverse will not pass the vertical line test unless we restrict S to be smaller so it passes the horizontal line test. Customarily, the inverse sine function is just such an inverse. It is the inverse for the sine function restricted to the interval $[-\frac{\pi}{2}, \frac{\pi}{2}]$. To conclude, relations can be inverted without regard to their particular properties.

Theorem 3.20. *(Theorem 3.3 in the text) Let* A, B, C, D *be sets. Suppose* R, S, T *are relations with* $R \subseteq A \times B, S \subseteq B \times C$ *and* $T \subseteq C \times D$

(a.) $(R^{-1})^{-1} = R$

(b.) $T \circ (S \circ R) = (T \circ S) \circ R$

(c.) $I_B \circ R = R$ *and* $R \circ I_A = R$

(d.) $(S \circ R)^{-1} = R^{-1} \circ S^{-1}$

Proof: See the text.

Item $(b.)$ says that we can write $T \circ S \circ R$ without ambiguity, composition of relations is associative. Item $(d.)$ is sometimes called the "*socks-shoes principle*". Think of it this way, if I put my socks on first and then second my shoes then when I take off my socks and shoes I have to take off my shoes first and then my socks.

3.4 equivalence relations and partitions

Given some set S we can consider various subsets of $S \times S$. Each such subset will form a relation from S to S. Of all those relations there is a particular type which has special structure which is similar to equality.

Definition 3.21. *(equivalence relation) Let A be a set and let R be a relation on A. We say that R is an equivalence relation if R is*

(i.) **reflexive***; for all $x \in A$, $x \; R \; x$.*

(ii.) **symmetric***; for all $x, y \in A$, if $x \; R \; y$ then $y \; R \; x$.*

(iii.) **transitive***;*

$$\text{for all } x, y, z \in A, \text{ if } x \; R \; y \text{ and } y \; R \; z \text{ then } x \; R \; z.$$

Example 3.22. *(equality) Suppose that $S = R$. Let $x, y \in S$, define xRy iff $x = y$. Observe that*

$$x = x, \quad x = y \;\Rightarrow\; x = y, \quad x = y \text{ and } y = z \;\Rightarrow\; x = z$$

Therefore, R is reflexive, symmetric and transitive. Hence equality is an equivalence relation on R. In fact, equality is an equivalence relation wherever one has an idea of equality (the choice of $S = \mathbb{R}$ could be modified and the calculations in this example would still work).

Example 3.23. *(order relation) Suppose that $S = \mathbb{R}$. For $x, y \in \mathbb{R}$, define xRy iff $x < y$. Observe that $x \not< x$ and certainly $x < y$ does not imply $y < x$. Therefore, R is neither reflexive or symmetric. However, R is transitive; $x < y$ and $y < z$ does imply that $x < z$. R is an example of an order relation, section 3.4 of the text discusses these at length. I do not plan to cover section 3.4 this semester.*

Example 3.24. *(even is even or odd is odd) Suppose that $S = \mathbb{Z}$. Suppose $x, y \in \mathbb{Z}$, define xRy iff $x - y$ is even. This is a fancy way of saying that even integers are related to even integers and odd integers are related to odd integers. Clearly R is reflexive since $x - x = 0$ which is even. Let $x, y \in \mathbb{Z}$ and assume xRy thus $x - y = 2k$ for some $k \in \mathbb{Z}$. Observe $y - x = -2k = 2(-k)$ hence yRx which shows R is symmetric. Finally, suppose $x, y, z \in \mathbb{Z}$ such that xRy and yRz. This means there exist $m, k \in \mathbb{Z}$ such that $x - y = 2k$ and $y - z = 2m$. Consider,*

$$x - z = x - y + y - z = 2k + 2m = 2(k + m)$$

Hence $x - z$ is even and we have shown xRz so R is transitive. In total we conclude R is an equivalence relation on $\mathbb{Z}$. Notice this relation is like equality, we just lump all the even integers into the same category and likewise for odd integers. Notice that either an integer is even or it is odd, there is no mixing between these cases.

Definition 3.25. *(equivalence class) Let S be a set with equivalence relation R. The equivalence class containing $x \in S$ is a set defined as follows:*

$$\bar{x} = x/R = \{y \in S \mid xRy\}$$

The set of all equivalence classes of S with respect to the equivalence relation R is denoted $S/R = \{x/R \mid x \in S\}$.

Example 3.26. *(even or odd integer example continued) Let $EVEN = \{x \in \mathbb{Z} |\ x \text{ is even}\}$ and $ODD = \{x \in \mathbb{Z} \mid x \text{ is odd}\}$. It is curious that there are infinitely many different representatives for $EVEN$. In particular,*

$$\bar{0} = \bar{2} = \bar{42} = EVEN$$

and so forth. Likewise, $\bar{1} = \bar{3} = ODD$. Observe that $\mathbb{Z} = ODD \cup EVEN$. Moreover, $ODD \cap EVEN =$. This is called a partition of $\mathbb{Z}$. A partition divides the set up into disjoint pieces.

Definition 3.27. *(Partition) Let S be a nonempty set. A family of subsets $\mathcal{A}$ is called a partition iff the following three conditions hold:*

(i.) Each partition is nontrivial; If $U \in \mathcal{A}$ then $U \neq \emptyset$,

(ii.) $\mathcal{A}$ is made of disjoint subsets; for all $U, V \in \mathcal{A}$, either $X = Y$ or $X \cap Y = \emptyset$.

(iii.) the union of all sets in the partition covers S;

$$\bigcup_{U \in \mathcal{A}} U = S$$

The partition formed by the relation in Example 3.22 is not particularly interesting. The equivalence classes for that example are singletons: $\{x\}$ for each $x \in \mathbb{R}$. The partition formed from the next example is much less trivial.

Example 3.28. *(path connected) A path in S from a to b is a continuous mapping f from $[0, 1]$ into S such that $f(0) = a$ and $f(1) = b$. Let S be some space S. We say that two points $a, b \in S$ are path-connected (let's denote this by R) if there is a path from a to b. Notice that R is reflexive since*

$$f(t) = a, \qquad \text{for each } t \in [0, 1]$$

is a continuous mapping from a to a and we can construct such a map at each point in S.

Furthermore, suppose that $a, b \in S$ such that aRb then there exists $f : [0,1] \to S$ such that $f(0) = a$ and $f(1) = b$. Define $g : [0,1] \to S$ by the formula:

$$g(t) = f(1-t)$$

We see that $g(0) = f(1) = b$ and $g(1) = f(0) = a$ hence g is a path from b to a. Therefore, R is symmetric.

Finally, if a, b, c are points in S such that aRb and bRc then there exist mappings $f : [0,1] \to S$ with $f(0) = a$ and $f(1) = b$, and $g : [0,1] \to S$ with $g(0) = b$ and $g(1) = c$ We can construct a path from a to c as follows:

$$h(t) = \begin{cases} f(2t) & \text{if } 0 \leq t \leq 1/2 \\ g(2t-1) & \text{if } 1/2 \leq t \leq 1 \end{cases}$$

It's easy to check that $h(0) = a$ and $h(1) = c$ hence aRc thus R is transitive. Hence R is an equivalence relation.

It's easy to check that $h(0) = a$ and $h(1) = c$ hence $a \sim c$ thus $\sim$ is transitive. Hence R is an equivalence relation. The subsets of S which are elements of S/R are called path components of S. If S has just one path component then S is said to be path connected. This simply means any two points in S can be connected by some path. For example, $\mathbb{R}^m$ is path connected if $m \in \mathbb{N}$. Other spaces are not. For example, you could think about the set of all invertible n by n square matrices (denoted $GL(n, \mathbb{R})$ if we want matrices with real components). Some of these matrices have $det(A) > 0$ whereas others have $det(A) < 0$. However, there is no invertible matrix with $det(A) = 0$ which means you cannot find a path to connect the two cases. Thus it turns out that $GL(n, \mathbb{R})$ has two equivalence classes with respect to the R of this example. In turn, $GL(n, \mathbb{R})$ is partitioned into two disjoint sets; matrices with positive determinant and matrices with negative determinant.

Yes, I did just talk about a path of matrices in the last example. I dare you to visualize that path. When you do draw me the picture.

Remark 3.29. *(Big Idea of Equivalence Relations) There are two ways to think about S/R. First, we can think about elements of S/R as subsets of S. Second, we can think about elements of S/R as single elements. In essence, the idea of the equivalence relation is to re-define the idea of equality. Points in S are equal if they reside in the same equivalence class. So we can think about the points in S/R relative to a new idea of equality. The new idea of*

equality is given by the equivalence relation. In other contexts S/R is called a quotient space and it is sometimes read "S modulo R.

Example 3.30. *(rational numbers) Let us define an equivalence relation on $S \subseteq Z \times Z$ as follows: suppose $S = \{(a,b) \mid a,b \in \mathbb{Z} \text{ and } b \neq 0\}$, $(a,b),(c,d) \in S$ then $(a,b) \sim (c,d)$ iff $ad = bc$. I claim that $\sim$ is an equivalence relation on S. Let $(a,b),(c,d),(x,y) \in S$,*

$$ab = ba \;\Rightarrow\; (a,b) \sim (a,b)$$

$$(a,b) \sim (c,d) \;\Rightarrow\; ad = bc \;\Rightarrow\; cb = da \Rightarrow\; (c,d) \sim (a,b)$$

The proof that $\sim$ is transitive follows similarly. You should understand what this example is doing better if I use the notation $(a,b) = a/b$ which translates $(a,b) \sim (c,d)$ if $ad = bc$. If you think about it you'll realize you've been using equivalence relations your whole life. We were taught early on that $1/2$ and $2/4$ are the "same" thing. Are they? Yes, of course, but understand that the idea of "same" means that $\mathbb{Q}$ is technically made of equivalence classes. Two fractions are in the same class when they cross-multiply to yield an equality. Rather than write $\sim$ all the time we simply write " = ".

Example 3.31. *(tangent vectors) Let $S \subset \mathbb{R}^3$ be a surface defined as the level-set of a function $F : \mathbb{R}^3 \to \mathbb{R}$; that is S is the set of points $(x,y,z) \in \mathbb{R}^3$ such that $F(x,y,z) = 0$. Parametric curves on S are paths from $\mathbb{R}$ to S. Denote the set of all smooth parametric curves on S that pass through $p \in S$ at time zero to be $C^1(p)$. Let $f,g \in C^1(p)$ then we say $f \sim g$ iff $f(0) = g(0) = p$ and $f'(0) = g'(0)$. I claim $\sim$ is an equivalence relation. Let $f \in C^1(p)$,*

$$f(0) = f(0) \text{ and } f'(0) = f'(0)$$

thus $\sim$ is clearly reflexive. Also, if $f,g \in C^1(p)$ and $f \sim g$ then $f'(0) = g'(0)$ hence $g'(0) = f'(0)$ so $g \sim f$. Moreover, if $f,g,h \in C^1(p)$ and $f \sim g$ and $g \sim h$ it follows that $f'(0) = g'(0)$ and $g'(0) = h'(0)$ hence $f'(0) = h'(0)$ which shows $f \sim h$. Thus $\sim$ is an equivalence relation on $C^1(p)$.

We can identify $\bar{\gamma} = \{f \in C^1(p) \mid f'(0) = \gamma'(0), \gamma(0) = p\}$. You can visualize $\bar{\gamma}$ in $\mathbb{R}^3$, it will be a vector that points in some direction $\vec{v} = \gamma'(0)$. The tangent space to S is formed by taking the union of all such vectors. In view of $\sim$ we can say that each element of the tangent space is an equivalence class of curves.

I have hidden many details in this example. I just wanted to throw in an example or two in these notes that illustrate interesting and non-obvious ap-

plications of equivalence classes. This example reveals one of the four main interpretations of a tangent vector in manifold theory. You might object that you know an easier idea of a vector in $\mathbb{R}^3$ and there is no need for these crazy equivalence classes. I can't argue too much in $\mathbb{R}^3$, however, what if your space was a space of matrices? The concept of tangent vector put forth in this example is annoyingly general. It gives you a way to put your hands on vectors you cannot see.

Besides the abstract, $\sim$ also gives us a tool to calculate the equation for the tangent plane at p. Let $g \in C^1(p)$ be a path on S then

$$F(g(t)) = 0$$

since the path $t \mapsto g(t)$ is assumed to lie on the surface S. Define the components of $g(t)$ as follows $g(t) = (a(t), b(t), c(t))$. The chain rule applied to $F \circ g$ goes as follows:

$$\frac{d}{dt}\Big(F(a,b,c)\Big) = \frac{\partial F}{\partial x}\frac{da}{dt} + \frac{\partial F}{\partial y}\frac{db}{dt} + \frac{\partial F}{\partial z}\frac{dc}{dt}$$

But, $F \circ g = 0$ so the derivative above must equal zero. Moreover, notice that the derivative of the parametric curve $g'(t) =< a'(t), b'(t), c'(t) >$. We can rewrite the chain-rule above as

$$(\nabla F)(g(t)) \cdot g'(t) = 0$$

If we focus our attention to time $t = 0$ where $g(0) = p$ and $\vec{v} = g'(0)$ we find the condition above tells us that

$$(\nabla F)(p) \cdot \vec{v} = 0$$

This holds for each and every tangent vector $\vec{v}$ at $p \in S$. It stands to reason that $(\nabla F)(p)$ is the normal vector to the tangent plane of S at p.

Remark 3.32. *I apologize for those of you not taking calculus III. Don't worry, the homework is more reflective of what I expect you to learn. It would not be fair for me to test the class on calculus III in the test. I'm merely trying to make a connection or two for those who've had or are taking calculus III.*

Remark 3.33. *(Theorems 3.5 and 3.6 in the text) In a nutshell these Theorems state that for every equivalence relation we get a partition. Moreover, we can reverse that, given a partition we can construct a corresponding equiv-*

alence relation. I will likely state and prove Theorem 3.5 in lecture because it is the heart of the whole matter. I have used it implicitly in several of the examples preceding this remark. Theorem 3.5 is how I know that the equivalence relation allows us to divide the set into pieces.

In the Spring 2009 semester we proved this Theorem in lecture. Note to self: add proofs to these notes later.

4 Functions

Functions have found a place in the mathematical lexicon for a few centuries at this time. However, the modern viewpoint of a function is only about a century old. As recently as the mid-nineteenth century mathematicians still had not quite refined the concept of a function. What precisely is a function? How is a function described? In this chapter we take the viewpoint that a function is a special type of relation. This is equivalent to identifying a function with its graph. In practice, this viewpoint is usually adopted but almost never does one think of a function as a subset of a Cartesian product. Instead, we typically envision a function as a rule connecting two sets. In other words, we usually think of a function as a map between two sets.

4.1 domain, range and codomain

Words, words, words, so many words.

Definition 4.1. *(function from A to B) A* **function** *from A to B is a relation f from A to B such that*

(i.) $dom(f) = A$

(ii.) *for each $x \in A$, xfy and xfz implies $y = z$.*

We write $f : A \to B$ to indicate that the function f has **domain** *A and* **codomain** *B. When $A = B$ we say that $f : A \to A$ is a function on A. We also say that $f : A \to B$ is a "B-valued function of A".*

From this point forward, we will exchange the relation notation of xfy for the more familiar notation $f(x) = y$. (this is unambiguous because of ii.)

Definition 4.2. *(arguments and preimages) If $f(x) = y$ then we say that x is an* **argument** *of the function and y is a* **value** *of the function. We also say that if $y = f(x)$ then x is a* **preimage** *of x.*

Let $f : A \to B$. Given $x \in dom(f)$ the value $f(x)$ is uniquely prescribed. In contrast, for a given $y \in B$ there can be many preimages x such that $f(x) = y$.

Example 4.3. *Let $f = \{(x, y) \mid x \in \mathbb{R},\ y = x^2\}$. Notice this is a function from $\mathbb{R}$ to $\mathbb{R}$ and we can write $f(x) = x^2$. The preimages of 3 are $\pm\sqrt{3}$ since $f(\pm\sqrt{3}) = (\pm\sqrt{3})^2 = 3$.*

Definition 4.4. *(images of set, fiber and range) For a function $f : A \to B$ we define,*

(i.) *for $U \subseteq A$, $f(U) = \{y \in B \mid y = f(x) \text{ for some } x \in U\}$*
(ii.) *for $V \subseteq B$, $f^{-1}(V) = \{x \in A \mid f(v) = x \text{ for some } v \in V\}$*

We define the **fiber** *over $y \in B$ to be $f^{-1}(\{y\})$. Furthermore, we define the* **range** *of f to be $range(f) = f(dom(f))$.*

Maybe that was a little too terse for you, let me restate the definition of a fiber and the range:

Definition 4.5. *(inefficient definitions of fiber and range) The set of all preimages of y is called the* **fiber** *of f over y*

$$f^{-1}(\{y\}) = \{x \in dom(f) \mid f(x) = y\}.$$

The **range** *of f is*

$$range(f) = \{y \mid \exists x \in dom(f) \ withf(x) = y\}$$

I think that about covers it for now. Let's look at a few examples.

Example 4.6. *Let $f \subseteq \mathbb{R} \times \mathbb{R}$ be the relation defined by*

$$f = \{(x, 2x^2) \mid x \in [-1, 1]\}$$

This is not a function on $\mathbb{R}$ because it is not well-defined in the following sense: $f(2)$ has no value. Notice we were only given $f(x) = 2x^2$ for $x \in [-1, 1]$. Is f is a function on $[-1, 1]$? No, $f(1) = 2 \notin [-1, 1]$. Remember to say "f is a function on $[-1, 1]$" we need $f : [-1, 1] \to [-1, 1]$. Observe that $f : [-1, 1] \to \mathbb{R}$ is a function. Note, for each input $x \in [-1, 1]$, f outputs the real number x^2. Moreover, you can calculate

$$f([-1, 1]) = [0, 2] = range(f)$$

and while we're at it

$$f^{-1}(\{2\}) = \{1, -1\}.$$

You should know from your prerequisite precalculus knowledge that $f(x) = 2x^2$ has a graph which is a parabola and as such it fails the horizontal line test. The notation $f^{-1}(\{2\})$ does not indicate that f^{-1} is a function. In the present example it is only a relation since $f^{-1}(2) = 1$ and $f^{-1}(2) = -1$. Some people would call such a rule a "double valued" function, we will not use such terminology.

Remark 4.7. *The terminology "f is a function* **of** *A" simply means that $dom(f) = A$. In contrast, the terminology "f is a function* **on** *A" means that $dom(f) = A$ and $range(f) = A$.*

Example 4.8. *Let $f(x) = \frac{x-1}{x-1}$.* **Find the largest domain for which the formula makes sense**. *Observe that if $x \neq 1$ then $f(x)$ is well-defined since $x - 1 \neq 0$. However, if $x = 1$ then the formula for the function is ill-defined. Hence, $dom(f) = \mathbb{R} - \{1\} = (-\infty, 1) \cup (1, \infty)$. The graph of this function is the horizontal line $y = 1$ with a hole at $x = 1$.*

Often a function's domain is not specified in certain contexts. It is customary to take the largest subset of inputs for which the defining formula is well-defined. For typical examples this simply means that we must avoid division by zero or negative inputs to logarithms or even-indiced radical functions. In general the import of the term *well-defined* has many facets.

Given a set S and an equivalence relation R we can construct the set of equivalence classes S/R (which is read "S modulo R" or simply "S mod R"). A typical element of S/R is an equivalence class $\bar{x} = \{s \in S \mid sRx\}$. If $f : S/R \to B$ is to be a function then the rule for the function must be given in such a way that the formula is independent of the representatives. If $f(\bar{x}) = g(x)$ for some function $g : S \to B$ then it must be shown that $g(x) = g(y)$ for any other $y \in \bar{x}$. In other words, g must be constant over the equivalence classes of R.

Example 4.9. *Let $S = \mathbb{R}^2$ and let $\sim$ be the equivalence relation defined by $(x, y) \sim (a, b)$ iff $x^2 + y^2 = a^2 + b^2$. Notice that the equivalence relation $\sim$ partitions the plane into circles about the origin and the origin itself. Let us denote the equivalence classes by $[(x, y)] = \{(a, b) \in \mathbb{R}^2 \mid x^2 + y^2 = a^2 + b^2\}$. Does the following formula decribe a function from $S/\sim$ to $\mathbb{R}$?*

$$f([(x, y)]) = x - y$$

Clearly $x - y \in \mathbb{R}$ is defined for all $x, y \in \mathbb{R}$ and it is obviously a real number. However, this is not a function because if we took a different representative of $[(x, y)]$ we would not get the same output. Notice that $(-x, -y) \in [(x, y)]$ thus $[(x, y)] = [(-x, -y)]$ yet

$$f([(-x, -y)]) = -x + y \neq f([(x, y)])$$

Thus the same equivalence class makes f output to different outputs. This means that f is not a function. We say that f is not "well-defined" when

this happens. In contrast,

$$g([(x,y)]) = \sin(\sqrt{x^2+y^2}) + 3$$

gives a well defined function since if $[(a,b)] = [(x,y)]$ *then* $a^2+b^2 = x^2+y^2$ *and consequently,* $g([(x,y)]) = g([(a,b)])$.

Example 4.10. *Let* A, B *be sets. Then* $\pi_A(a,b) = a$ *is a function from* $A \times B$ *to* A *called the projection onto* A. *Likewise,* $\pi_B(a,b) = b$ *defines* π_B *the projection from* $A \times B$ *to* B. *The fiber of* $b \in B$ *with respect to* π_B *is* $A \times \{b\}$ *since*

$$\pi_B(A \times \{b\}) = b$$

Let A *be the unit circle and* $B = [0,1]$ *then you can visualize* $A \times B$ *as the unit-cylinder. The "fiber" of a point* $p \in [0,1]$ *on the unit interval is circle at the point* p.

Example 4.11. *(identity relation on* A *is a function on* A*) Let* A *be a set. The identity relation on* A *is defined as follows:*

$$I_A = \{(a,a) \mid a \in A\}$$

Observe that $I_A(x) = x$ *for each* $x \in A$ *is a well-defined formula for the function* $I_A : A \to A$. *Moreover,* $dom(I_A) = range(I_A) = A$.

4.2 constructing new functions

Given several functions there are numerous ways to construct new functions by combining the given functions through addition, mulitplication, difference, composition, extension, restriction and so forth... At the heart of each construction is the following basic theorem:

Theorem 4.12. *Two functions* f *and* g *are equal iff*

(i.) $dom(f) = dom(g)$

(ii.) $f(x) = g(x)$ *for each* $x \in dom(f)$.

Proof: Since the value of the function follows uniquely from the argument it follows immediately that x is in the first component of the ordered pair with $f(x)$ in the second slot. Hence $(x, f(x)) \in f$ viewed as a relation. Of course the same is true for g, thus $(x, g(x)) \in g$. Thus, since $f(x) = g(x)$ for each $x \in dom(f)$, $(x, g(x)) \in f$. It follows that $f = g$ since they share all the same elements.

Example 4.13. *Let f, g be functions such that $dom(f) = dom(g)$ and $range(f), range(g) \subseteq \mathbb{R}$. The functions $f + g$, fg, f/g are all defined point-wise: $(f + g)(x) = f(x) + g(x)$, $(fg)(x) = f(x)g(x)$ and $(f/g)(x) = f(x)/g(x)$. The domains of the new functions are $dom(f + g) = dom(fg) = dom(f)$ and $dom(f/g) = \{x \in dom(f) \mid g(x) \neq 0\}$.*

Recall that we defined the composite and inverse of relations in as much was possible. Since functions are a special type of relation we can likewise discuss inverses and composites of functions. However, there is no gaurantee that the inverse of a function will itself be a function.

Theorem 4.14 (Theorem 4.2 of the text)**.** *Let A, B, C be sets and F, G be functions such that $F : A \to B$ and $G : B \to C$. Then $G \circ F$ is a function from A to C with $dom(G \circ F) = A$.*

Proof: Let $x \in A$ then $F(x) \in B$ since $F : A \to B$. Thus $F(x) \in dom(G)$ and $G(F(x)) \in C$ since $G : B \to C$. Hence $dom(G \circ F)$. Notice that $G \circ F$ is single valued since F is single valued at $x \in A$ and G is single valued at $F(x) \in B$. *If you don't find that convincing take a look at the text's verbose proof which is mainly about showing that the output of the composite is single-valued.*

Remark 4.15 (how to find composite of two functions generally)**.** *Often when we compose two functions the range of the inside function will not match the domain of the outside function. In fact, usually we will need to place two restrictions on the domain of $f \circ g$ for $f : B \to C$ and $g : A \to B$. If you think about it we can define $f \circ g$ only for a certain restriction of g. We need two things for $f(g(x))$ to be a sensible formula:*

1. $x \in dom(g)$
2. $g(x) \in dom(f)$

This means that we choose $dom(f \circ g) = dom(g) \cap g^{-1}(dom(f))$. The definition of composite we gave in this section assumes that the given functions have domains and ranges which match up nicely to start with, often this is not the case.

Example 4.16. *Given $f(x) = \sqrt{x}$ and $g(x) = x - 2$ find $f \circ g$ and determine the domain of the composite function. Observe that,*

$$(f \circ g)(x) = f(g(x)) = f(x - 2) = \sqrt{x - 2}$$

Here $range(g) = \mathbb{R}$ *and* $dom(g) = \mathbb{R}$*. However,* $dom(f) = [0, \infty)$ *and you can see that* $g^{-1}([0, \infty)) = \{x \in \mathbb{R} \mid x - 2 \geq 0\} = [2, \infty)$*. Therefore,* $dom(f \circ g) = [2, \infty)$

If you had simply examined the formula for the composite then you would probably have concluded the same domain. However, be careful, formulas can be deceiving, especially if we do any simplifications.

Example 4.17. *Given* $f(x) = x^2$ *and* $g(x) = \sqrt{x}$ *find* $f \circ g$ *and determine the domain of the composite function. Observe that,*

$$(f \circ g)(x) = f(g(x)) = f(\sqrt{x}) = (\sqrt{x})^2 = x$$

Here $range(g) = [0, \infty)$ *and* $dom(g) = [0, \infty)$*, this limits the domain of the composite. However, notice* $dom(f) = \mathbb{R}$ *imposes no restriction in this example. Therefore,* $dom(f \circ g) = [0, \infty)$*. If you just look at the formula* $(f \circ g)(x) = x$ *you might be tempted to say that the domain is all of* $\mathbb{R}$ *(which would be incorrect!)*

Theorem 4.18 (Theorem 4.3 of the text)**.** *Let* A, B, C, D *be sets and* $f : A \to B$*,* $g : B \to C$ *and* $h : C \to D$*. The composition of functions is associative;* $(h \circ g) \circ f = h \circ (g \circ f)$*.*

Proof: Observe that $h \circ g : B \to D$ and $g \circ f : A \to C$ are functions by Theorem 4.2. Thus, the composite of $h \circ g$ and f is a function and likewise, $h \circ (g \circ f)$ is a function. Moreover, these share the same domain, namely A. It suffices to show these are equal at an arbitrary point in A. Consider then,

$$[h \circ (g \circ f)](x) = h((g \circ f)(x)) = h((g(f(x))))$$

and similarly,

$$[(h \circ g) \circ f](x) = (h \circ g)(f(x)) = h((g(f(x)))).$$

Theorem 4.19 (Theorem 4.4 of the text)**.** *Let* $f : A \to B$ *be a function and recall* $I_A : A \to A$ *is the* **identity relation** *on* A*. Then* $f \circ I_A = f$ *and* $I_B \circ f = f$

Proof: Clearly $I_A : A \to A$ is a function since $I_A(x) = x$ is clearly a single-valued formula. Hence, $f \circ I_A, I_B \circ f$ are functions by Theorem 4.2. Consider then,

$$(I_B \circ f)(x) = I_B(f(x)) = f(x)$$

for each $x \in dom(f) = A$ thus $I_B \circ f = f$ by Theorem 4.1. Furthermore,

$$(f \circ I_A)(x) = f(I_A(x)) = f(x)$$

for each $x \in dom(f) = A$ thus $f \circ I_A = f$ by Theorem 4.1.

Theorem 4.20 (this is Theorem 4.5 in the text)**.** *Let $f : A \to B$ be a function with $range(f) = C$. If f^{-1} is a function then $f^{-1} \circ f = I_A$ and $f \circ f^{-1} = I_C$*

Proof: Let $f : A \to B$ is a function and assume that f^{-1} is a function. We know that f^{-1} is also the inverse of the relation $f \subseteq A \times B$; if xfy then $yf^{-1}x$ in the relation notation. But this means that $x(f^{-1} \circ f)x$ for each $x \in A$. Moreover, we can be sure that y is f^{-1} related to only x since f^{-1} is a function. Thus $f^{-1} \circ f = I_A$.

Likewise, if $y \in range(f) = C$ then there exists $x \in dom(f) = A$ such that $f(x) = y$. In other words, xfy which implies $yf^{-1}x$ and since f^{-1} is a function we know that x is the unique element in A which is f^{-1} related to y. To summarize, xfy and $yf^{-1}x$ for each $x \in range(f) = C$. Thus by definition of composite, $x(f \circ f^{-1})x$ for each $x \in C$. Thus by Theorem 4.1, $f \circ f^{-1} = I_C$.

Remark 4.21. *Notice that we can have $B \neq C$ in the preceding Theorem. For example, $f(x) = \sqrt{x}$ defined to be a function from $[0, \infty)$ to $\mathbb{R}$. Then you can calculate $f^{-1}(y) = y^2$ and $dom(f^{-1}) = [0, \infty)$. Clearly, $\mathbb{R} \neq [0, \infty)$. One way to think about this is that the codomain I chose for f was needlessly large. Since $f(x) = \sqrt{x}$ it follows that the range of f is full of non-negative values, we'll never cover the negative half of $\mathbb{R}$. This is the funny thing about codomains, you can always make them bigger.*

You can also make domains smaller. This is known as restricting a function.

Definition 4.22 (restriction of a function)**.** *Let $f : A \to B$ and let $U \subseteq A$ then the* **restriction of f to U** *is the function*

$$f|_U = \{(x, y) \mid x \in U \text{ and } y = f(x)\}$$

Additionally, if g is a restriction of h then we say h is an extension of g.

Restrictions are unique once a subset U is specified. Of course given different U we can form different restrictions. Extensions on the other hand allow for much more imagination.

Example 4.23 (silly extension)**.** *Consider $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = \sin(x)$, I can extend this to*

$$g(x) = \begin{cases} \sin(x) & \text{if } x \in \mathbb{R} \\ \mathbb{R} & \text{if } x \in \{i\} \end{cases}$$

Here clearly $g|_{\mathbb{R}} = f$ whereas $dom(g) = \mathbb{R} \cup \{i\}$ and in my crazy example here $g(i) = \mathbb{R}$. The output in my extension is not even of the same type as f.

The ambiguity in extending a function is pretty much the same ambiguity you have faced in homework problems where you were supposed to extrapolate from a given graph.

Example 4.24 (less silly extension)**.** *How do we extend the sine function to complex numbers? It can be shown that $exp : \mathbb{C} \to \mathbb{C}$ is a well-defined function of $\mathbb{C}$. The definition I am fond of is simply,*

$$\boxed{exp(z) = exp(x + iy) = e^x(\cos(y) + i\sin(y))}$$

It can be shown that this exponential function satisfies the same algebraic properties over $\mathbb{C}$ as the usual exponential function does for $\mathbb{R}$. Given this you can derive that for $x \in \mathbb{R}$

$$\sin(x) = \frac{1}{2i}\left(e^{ix} - e^{-ix}\right).$$

An obvious extension of $\sin : \mathbb{R} \to [-1, 1]$ to $\mathbb{C}$ is given by the formula

$$\sin(z) = \frac{1}{2i}\left(e^{iz} - e^{-iz}\right)$$

for each $z \in \mathbb{C}$. If we define $f(z) = \sin(z)$ for $z \in \mathbb{C}$ (as explained above) then $f|_{\mathbb{R}}$ is simply the ordinary sine function. By the way, $range(f) = \mathbb{C} - \{0\}$. You can take our complex variables course to learn more.

Theorem 4.25 (Theorem 4.6 in the text)**.** *Given two functions with non-overlapping domains, we can form a new function by simply pasting the given functions together. Suppose $f : A \to C$ and $g : B \to D$ are functions and $A \cap B = \emptyset$ then $h = f \cup g \subset (A \cup B) \times (C \cup D)$ is a function. Moreover,*

$$h(x) = \begin{cases} f(x) \text{ if} & x \in A \\ g(x) \text{ if} & x \in B \end{cases}$$

for each $x \in dom(h) = A \cup B$.

Proof: The formula for $h(x)$ is clearly single-valued since $x \in A \cup B$ and $A \cap B = \emptyset$ means either $x \in A$ or $x \in B$ (but not both). So the cases are distinct and in each case the output is single-valued by virtue of the fact that f, g are functions.

Remark 4.26 (Theorem 4.6 is weak)**.** *We can do better. In fact we can paste together two functions f, g which have $dom(f) \cap dom(g) \neq \emptyset$ provide that the functions are equal on the overlap. This is most of what is known as the "***pasting Lemma***".*

Remark 4.27. *Sorry there are not more examples in this particular stretch of the notes. I want to leave time for homework questions. Please ask if you can make these definitions come to life in the homework. I am here to help, just ask.*

4.3 injective functions

Definition 4.28. *Let $f : A \to B$ be a function then we say that f is* **injective** *or* **one-to-one** *iff for all $a, b \in A$, $f(a) = f(b)$ implies $a = b$. In other words, for each output $y \in range(f)$ there is a unique input $x \in dom(f)$ such that $y = f(x)$.*

Example 4.29. *Suppose $f : \mathbb{R} \to [0, \infty)$ is defined by $f(x) = x^2$ then f is not injective since $f(1) = f(-1)$ yet $1 \neq -1$.*

Example 4.30. *Suppose $f : [0, \infty) \to [0, \infty)$ is defined by $f(x) = x^2$ then f is injective. To prove this let $a, b \in [0, infty)$ and suppose $f(a) = f(b)$,*

$$f(a) = f(b) \;\Rightarrow\; a^2 = b^2 \;\Rightarrow\; a = \pm b$$

However, since $a, b > 0$ it follows that only the $(+)$ solution is allowed and thus $a = b$. That little argument proves f is injective.

Remark 4.31. *The horizontal line test is based on the same logic. If a horizontal line crosses the graph at a and b with $a \neq b$ then that means that $f(a) = f(b)$ yet $a \neq b$. A function $f : U \subseteq \mathbb{R} \to V \subseteq \mathbb{R}$ will be injective iff it passes the horizonal line test.*

The last remark is useful, but the definition we gave for injective is far more general than the horizontal line test. How can you apply the horiontal line test in a situation where the graph is 4 or 5 dimensional?

Example 4.32 (determinant is not 1-1)**.** *Let* det $: M_2(\mathbb{R}) \to \mathbb{R}$ *be the determinant function which takes an input of an 2×2 real-entried matrix A and outputs a single number which we denote by $det(A)$. Suppose that*

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

then

$$det(A) = det\begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc.$$

Observe that many different matrices will have a determinant of zero.

$$det\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = det\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = det\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} = 0$$

Clearly if we look at the inverse image of $\{0\}$ *with respect to det we will find that this fiber has lots of stuff in it. Many different matrices have determinant zero. It turns out that many different* 2×2 *matrices also have determinant* 1*, I just picked on zero because it's easy. To conclude, the determinant function is* **not** *injective.*

Example 4.33 (horizontal plane test)**.** *Consider* $f(x, y) = x + y$ *with* $dom(f) = \mathbb{R}^2$*. The graph* $z = f(x, y)$ *is a plane with normal* $< 1, 1, -1 >$ *for those of you in calculus III. Suppose* $f(a_1, a_2) = f(b_1, b_2)$ *then* $a_1 + a_2 = b_1 + b_2$*. This does not imply that* $(a_1, a_2) = (b_1, b_2)$ *take for example* $f(0, 1) = f(1, 0) = 1$*. Thus* $f(a_1, a_2) = f(b_1, b_2)$ *does not imply* $(a_1, a_2) = (b_1, b_2)$*. If you want to think about this graphically, the graph failed the horizontal plane test. Can you think of a function of two variables* $f(x, y)$ *such that its graph* $z = f(x, y)$ *outputs each z-value for just one* (x, y)*-input? Such a function's graph will only be cut by a horizontal plane at one point in the graph* $z = f(x, y)$*? I can only think of rather stupid examples at the moment.*

Theorem 4.34 (this is Theorem 4.11)**.** *If* $f : A \to B$ *be injective and* $g : B \to C$ *be injective, then* $g \circ f : A \to C$ *is injective. The composite of injective functions is injective*

Proof: Suppose $(g \circ f)(a) = (g \circ f)(b)$. Then $g(f(a)) = g(f(b))$ hence $f(a) = f(b)$ using the fact that g is one-one. Likewise, $f(a) = f(b)$ implies $a = b$ by one-one property for f. Therefore, $g \circ f$ is one-one.

4.4 surjective functions

The trouble with codomains is that sometimes they are too big. Surjectivity helps us pin down this ambiguity with functions.

Definition 4.35. *Let* $f : A \to B$ *be a function. We say that* f *is* **onto** $V \subseteq B$ *iff for each* $b \in V$ *there exists* $a \in A$ *such that* $f(a) = b$*. If* f *is onto* B *then we say that* f *is* **surjective** *or* **onto**.

The terms "onto" and "surjective" are the same. Both can be applied to either the function as a whole or to just some subset of the codomain.

Example 4.36. *Let $f(x) = \sin(x)$ then f is not onto $\mathbb{R}$ since there does not exist $x \in \mathbb{R}$ such that $\sin(x) = 2$ (for example). On the other hand, f is onto $[-1, 1]$ since the graph of sine oscillates between $y = 1$ and $y = -1$. We discussed in lecture that $f(x)$ was not one-one. We had to restrict the domain of f to $[-\pi/2, \pi/2]$ in order to gain the injective property for the restriction.*

Theorem 4.37 (this is Theorem 4.7 in the text)**.** *The composite of surjective functions is surjective.*

Proof: Suppose that $f : A \to B$ is a surjective function and $g : B \to C$ is a surjective function. Consider $g \circ f : A \to C$. We seek to show $g \circ f$ is onto C. Let $c \in C$. There exists $b \in B$ such that $g(b) = c$ since g is onto C. Moreover, there exists $a \in A$ such that $f(a) = b$ since f is onto B. Thus, $(g \circ f)(a) = g(f(a)) = g(b) = c$ which demonstrates that the composite is surjective as claimed.

Theorem 4.38 (Theorem 4.8 of the text)**.** *If a composite function is onto then the outside function in the composite is also onto. That is if $f : A \to B$ and $g : B \to C$ are functions such that $g \circ f$ is surjective then g is surjective.*

Proof: Assume f and g as in the Theorem. Suppose $c \in C$ then since $g \circ f : A \to C$ is onto C it follows there exists $a \in A$ such that $(g \circ f)(a) = c$. Consider then that $(g \circ f)(a) = g(f(a)) = c$. Thus we find for each $c \in C$ there exists $f(a) \in B$ such that $g(f(a)) = c$. That shows that g is surjective.

Example 4.39 (determinant is surjective)**.** *Let* $\det : M_2(\mathbb{R}) \to \mathbb{R}$ *be the determinant function. Let $x \in \mathbb{R}$, observe that*

$$det \begin{pmatrix} x & 0 \\ 0 & 1 \end{pmatrix} = x.$$

Thus $\det : M_2(\mathbb{R}) \to \mathbb{R}$ *is* **onto** $\mathbb{R}$.

Example 4.40. *Let $f(x) = \ln(x - 1)$. I claim this function is onto $\mathbb{R}$. I can anticipate this result because I know with confidence what the graph of natural log looks like (I hope the same is true for you). Let $y \in \mathbb{R}$ we wish to find $x \in dom(f) = (1, \infty)$ such that $\ln(x - 1) = y$. This calls for a small calculation,*

$$y = \ln(x - 1) \;\Rightarrow\; e^y = x - 1 \;\Rightarrow\; x = e^y + 1.$$

Now I can prove f is surjective with confidence.
Proof: *Let $y \in \mathbb{R}$ then observe that $x = e^y + 1 > 1$ thus $x \in dom(f)$ and*

$$f(x) = f(e^y + 1) = \ln((e^y + 1) - 1) = \ln(e^y) = y.$$

Therefore, $f : (1, \infty) \to \mathbb{R}$ is onto.

Is this function f also injective? Let $a, b \in (1, \infty)$, assume $f(a) = f(b)$ then $\ln(a - 1) = \ln(b - 1)$ take the exponential of that equation to obtain $a - 1 = b - 1$ hence $a = b$. Thus f is **injective**

4.5 bijective functions

This definition is central to Chapter 5. Basically we will learn that two sets have the same size if you can find a bijection between them.

Definition 4.41 (bijection)**.** *A function $f : A \to B$ is a* **bijection** *iff f is both an injection and a surjection. In other words, f is a bijection iff it is both 1-1 and onto. Finally, we also call a bijection a 1-1 correspondence.*

Example 4.42. *By Example 4.40 we have that $f(x) = \ln(x - 1)$ is a bijection.*

4.6 inverse functions

We characterized how the inverse function interfaces with its function but we have yet to give a general criteria as to when the inverse function exists. Inverse relations were easy to define since we just had to flip the pairs in the Cartesian product. However, for a function, once we flip the pairs then there is nothing that in general assures us that the result is a function. Given $f : A \to B$ a function it will be true that $f^{-1} \subseteq B \times A$ is a relation from B to A. What will make $f^{-1} : B \to A$ a function?

Theorem 4.43 (this is Theorem 4.9)**.** *Let $f : A \to B$ then*

(a.) f^{-1} is a function from $range(f)$ to $dom(f) = A$ iff f is injective.

(b.) If f^{-1} is a function, then f^{-1} is injective.

Proof of (a.): Assume that f^{-1} is a function from $range(f)$ to $dom(f) = A$. Suppose that $f(a) = f(b)$ for $a, b \in dom(f)$. Observe,

$$f^{-1}(f(a)) = f^{-1}(f(b)) \;\Rightarrow\; a = b.$$

Therefore f is one-one.

Conversely suppose that f is injective. Then $f(a) = f(b)$ implies $a = b$. Let $y \in range(f)$ then suppose $f^{-1}(y) = x$ and $f^{-1}(y) = z$, clearly there exists at least one such x or z by definition of $range(f)$. By definition of inverse relation, $f(f^{-1}(y)) = f(x)$ and $f(f^{-1}(y)) = f(z)$ thus $f(x) = f(z)$. Since f is injective it follows $x = z$ thus the relation f^{-1} is single valued with $f^{-1} : range(f) \to dom(f)$

Proof of (b.): Assume that f^{-1} is a function relative to the function $f : A \to B$. Suppose that $f^{-1}(y) = f^{-1}(z)$ then operate on both sides by the function f to get $f(f^{-1}(y)) = f(f^{-1}(z))$ but then by definition of inverse function $y = z$. Therefore f^{-1} is injective.

Corollary 4.44 (this is Corollary 4.10)**.** *If $F : A \to B$ is a bijection then $F^{-1} : B \to A$ is a bijection.*

Proof: left to reader.

Theorem 4.45 (this is Theorem 4.12)**.** *If $f : A \to B$ and $g : B \to C$ are bijections then*

(a.) $g \circ f : A \to C$ *is a one-one correspondence.*

(b.) $f^{-1} : B \to A$ *is a one-one correspondence.*

Proof of (a.): Theorem 4.37 give us that $g \circ f$ is surjective. Likewise, Theorem 4.34 gives us that $g \circ f$ is injective. Hence $g \circ f$ is a one-one correspondence.

Proof of (b.): We know from the relations discussion that $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$ and by Corollary 4.44 $(g \circ f)^{-1}$ is a bijection. Furthermore, Theorem 4.38 shows that f^{-1} is surjective since is the outside function of a surjective composite. Finally, Theorem 4.43 allows us to be sure f^{-1} is injective. Therefore, f^{-1} is a bijection.

Theorem 4.46 (Theorem 4.13 of the text)**.** *If a composite function is injective then the "first function applied" in the composite is also injective. That is if $f : A \to B$ and $g : B \to C$ are functions such that $g \circ f$ is injective then the "first function applied" is injective*

Proof: Left to reader. First you have to figure out what the text means by "first function applied". Is it f or is it g in $g \circ f$? I'm not sure without thinking a minute or two. I leave it to you.

Theorem 4.47 (this is Theorem 4.14)**.** *Let $F : A \to B$ and $G : B \to A$. Then $G = F^{-1}$ iff $G \circ F = I_A$ and $F \circ G = I_B$*

Proof: Suppose $G = F^{-1}$ then $F \circ G = F \circ F^{-1} = I_B$ whereas $G \circ F = F^{-1} \circ F = I_A$ (using Theorem 4.20 twice). Conversely suppose $G \circ F = I_A$ and $F \circ G = I_B$. Observe that F is injective (by $G \circ F = I_A$) and surjective (by $F \circ G = I_B$) thus F^{-1} is a function on B using the Theorems of this section. Consider then,

$$F^{-1} = F^{-1} \circ I_B = F^{-1} \circ F \circ G = I_A \circ G = G$$

Remark 4.48. *There are examples where $F \circ G = I_A$ yet $G \circ F \neq I_B$. In this case we can not say that $F^{-1} = G$. It is necessary in general to check both sides. (certain contexts make one side follow from the other, but that only happens when there is an additional stucture added to the mix)*

Example 4.49. *Let $f(x) = x^2$ and $g(x) = \sqrt{x}$. Clearly $(f \circ g)(x) = x$ for each $x \in dom(g) = [0, \infty)$. However, $(g \circ f)(x) = \sqrt{x^2} = \pm x$ thus we say that g has* **left inverse** *f but has no* **right inverse***. This means that $f^{-1} \neq g$. If we restrict g to $g|_{[o,\infty)}$ then we get both left and right inverse conditions to hold and the preceding Theorem would allow us to conclude that $f^{-1} = g|_{[o,\infty)}$.*

Remark 4.50. *Theorem 4.15 is interesting. Please read it. Essentially it says that restrictions of injective functions are injective. Also if you paste together surjective functions then they are surjective on the union of the ranges, likewise there is a natural thing to say about injectivity for pasted functions.*

4.7 break it down

In this section we consider an function $f : A \to B$. Our mission is to use our knowledge of equivalence classes and partitions together with the function terminology and constructions. Our goal is to understand the interplay between fibers, injectivity, surjectivity, domain, codomain and range.

Proposition 4.51. *If $\mathcal{B}$ is a partition of B then $\mathcal{A} = f^{-1}(\mathcal{B}) - \emptyset$ is partition of A.*

Proof: to be supplied by you for a bonus point. Need to do the proof for an infinite partition since that is often what interests us in the dicussion that

follows. By the way we have to throw out the emptyset which arises from the possibility the codomain is bigger than the range.

Observation 1. We can partition B into singletons. This is a silly partition but it is always available for nonempty B.

$$\mathcal{B} = \{\{b\} \mid b \in B\}$$

The we can define $\mathcal{A}$ to be the inverse image of $\mathcal{B}$ in the following sense,

$$\mathcal{A} = \{f^{-1}(\{b\}) \mid b \in B\} - \emptyset$$

When $range(f) = B$ then there is no need to remove the $\emptyset$. Saying a function is surjective is just a fancy way of saying the codomain is the range. You might complain that we should just use the range and not allow for the codomain to be bigger than the range. I hear you, but that is why we have the term "surjective".

Observation 2. The partition $\mathcal{A}$ of A will yield an corresponding equivalence relation on A. A moments reflection reveals that the equivalence classes are precisely the object we called **fibers** of f,

$$\bar{a} = a/\sim = \{x \in A | f(x) = f(a)\} = f^{-1}(\{f(a)\})$$

This means that $x \sim y$ iff $f(x) = f(y)$; the fibers are subsets of the domain on which the function is constant. It is not hard to explicitly check our logic on the claim that $\sim$ is an equivalence relation on A.

Observation 3. There is a natural map $\pi : A \to A/\sim$ defined as $\pi(a) = \bar{a}$. This map is clearly a surjection onto $A/\sim$ since if $\bar{a} \in A/\sim$ then $\pi(a) = \bar{a}$.

Observation 4. The function $f : A \to B$ induces an injective function $\bar{f}$ from the quotient $A/\sim$ to B. In particular,

$$\bar{f}(\bar{x}) = f(x)$$

Proof: First we should confirm that $\bar{f}$ is actually a function (is it well-defined ?) Let $y \in \bar{x}$ be some other representative of $\bar{x}$. We have $\bar{f}(\bar{x}) = f(x)$ or $\bar{f}(\bar{x}) = f(y)$. Now, we would be in trouble if it were the case that $f(x) \neq f(y)$. However, we know that $f(x) = f(y)$ since $x, y \in \bar{x} = \{z \mid f(z) = f(x)\}$.

Next we verify that $\bar{f}$ is injective. Suppose that $\bar{f}(\bar{x}) = \bar{f}(\bar{y})$, this yields $f(x) = f(y)$ but this means that $x \in \bar{y}$ and hence $\bar{x} = \bar{y}$. Therefore, $\bar{f}$ is injective as claimed.

Observation 5. The function $f : A \to B$ can be modified to yield a surjective function onto the $range(f) = f(A)$.

Observation 6. The induced map $\bar{f}$ will provide a bijection from $A/\approx$ to $range(f)$.

Make me draw the picture of this in lecture: this pattern occurs in many settings. The equivalence class / natural map construction is central to many interesting discussions in advanced mathematics. It has relevance to topology, group theory, ring theory, linear algebra etc... It is construction basic to almost every modern field of mathematics. (That is why I mention it here.)

Summary: in general a function $f : A \to B$ need not be injective or surjective. However, we can always do the following two things to obtain injectivity or surjectivity:

1. Reduce the codomain B to the range to make the map $g : A \to range(f)$ onto.

2. Restrict the domain A to a slice U of A with respect to π to create the map $f|_U : U \to B$ which is one-one.

Definition 4.52. *(section) A* **section** *of $\pi : A \to A/\sim$ is a map $s : A/\sim \to A$ such that $s \circ \pi = id_A$.*

The "slice" U I mentioned in the list is the image of A under a section s. What a section does in this context is to pick out just one reprentative in each class. (I'll draw a picture to help you visualize this).

The idea for this section is largely due to my brother Bill, however these ideas are so *natural* a serious mathematician would find his way to them in time. In other words, while Bill pointed them out to me recently, the math also intrinsically points to these ideas by virtue of its structure.

You should be suspicious of this word *natural*. What do I mean by that? It goes back to the PLA. I'll leave a little mystery here.

5 Cardnality

I should warn you from the outset of this Chapter that I have no intention of proving anything in Chapter 5. The text is quite readable if you are interested and I am purposefully leaving it open as a possible place to pick an end-of-semester proof talk from. My goal for us is to get the big picture about cardinality. In particular I want you to learn the meaning of the terms "finite", "denumerable", "countable" and "infinite". I want you to gain a deeper appreciation for the difference between real and rational numbers. I will only state the most crucial theorems and examples from Chapter 5.

5.1 one-one correspondence and finite sets

Definition 5.1 (equivalent sets)**.** *Two sets A and B are said to be* **equivalent** *iff there exists a one-one correspondence between them. In the case that there exists a bijection from A to B we say that $A \approx B$.*

We can easily show that $\approx$ forms an *equivalence relation* on the "class" of all sets. Notice the text did not say "set of all sets". We discussed why in our conversation about axiomatic set theory. We avoid the tiresome question: "does the set of all sets contain itself?".

Example 5.2 (finite sets)**.** *Consider a set $A = \{1, \emptyset, dora\}$. This is equivalent to the set $\{1, 2, 3\}$. To prove this construct the mapping*

$$\Psi(1) = 1, \quad \Psi(\emptyset) = 2, \quad \Psi(3) = dora$$

it is clear this is both one-one and onto $\{1, 2, 3\}$. You might object that these are not the "same" sets. I agree, but I didn't say they were the same, I said they were **equivalent** *or perhaps it is even better to say that the sets are in* **one-one correspondence**.

Now I repeat the same idea for an arbitrary finite set which has k things in it. Let $\mathbb{N}_k = \{1, 2, \ldots, k\}$. If a set A has k distinct objects in it then it is easy to prove it is equivalent to $\mathbb{N}_k = \{1, 2, \ldots, k\}$. Lable these k objects $A = \{a_1, a_2, \ldots a_k\}$ then there is an obvious bijection,

$$\Psi(a_j) = j \text{ for each } j \in \mathbb{N}_k$$

The mapping Ψ is one-one since for $a_j, a_l \in \mathbb{N}_k$ we find $\Psi(a_j) = \Psi(a_l)$ implies $j = l$ implies $a_j = a_l$.

I claim the mapping Ψ is also onto. Let $y \in \mathbb{N}$ then by definition of $\mathbb{N}_k$ we have $y = j$ for some $j \in \mathbb{N}$ with $1 \leq j \leq k$. Observe that $a_j \in A$ since $1 \leq j \leq k$, and $\Psi(a_j) = j$.

Given the last example, you can appreciate the following definition of **finite**.

Definition 5.3 (finite set, cardnality of finite set)**.** *A set S is said to be* **finite** *iff it is empty $S = \emptyset$ or in one-one correspondence with $\mathbb{N}_k$ for some $k \in \mathbb{N}$. Moreover, if $S \approx \mathbb{N}_k$ we define the* **cardinality** *of A to be $\overline{\overline{A}} = k$. If $S = \emptyset$ then we define $\overline{\overline{A}} = 0$.*

To summarize, the cardality of a finite set is the number of elements it contains. The nice thing about finite sets is that you can just count them.

Definition 5.4 (infinite sets)**.** *A set S is infinite if it is not finite.*

Proposition 5.5 (this is Corollary 5.10 in the text)**.** *A finite set is not equivalent to any of its proper subsets.*

A proper subset $A \subset B$ will be missing something since a "proper subset" A is a subset which is not the whole set B. It follows that B must have more elements and consequently $A \approx \mathbb{N}_a$ and $B \approx \mathbb{N}_b$ where $a < b$. The contrapositive of the Proposition above is more interesting.

Proposition 5.6 (contrapositve of Corollary 5.10 in text)**.** *A set which is equivalent to one or more of its proper subsets is infinite.*

So if you were counting, there are two nice ways to show a set is infinite. First, you could assume it was finite and then work towards a contradiction. Second, you could find a bijection from the set to some proper subset of itself.

Example 5.7 ($\mathbb{N}$ is infinite)**.** *Observe that the mapping $f : \mathbb{N} \to 2\mathbb{N}$ defined by $f(n) = 2n$ is a bijection. First, observe*

$$f(x) = f(y) \quad \Rightarrow \quad 2x = 2y \quad \Rightarrow \quad x = y$$

therefore f is injective. Next $2\mathbb{N} = \{2k \mid \exists k \in \mathbb{N}\}$. Let $y \in 2\mathbb{N}$ then there exists $k \in \mathbb{N}$ such that $y = 2k$. Observe that

$$f(k) = 2k = y$$

thus f is onto $2\mathbb{N}$. Therefore $\mathbb{N} \approx 2\mathbb{N}$ and since $2\mathbb{N}$ is a proper subset of $\mathbb{N}$ it follows that $\mathbb{N}$ is infinite.

5.2 countably infinite sets

Definition 5.8 (denumerable)**.** *Let S be a set, we say S is* **denumerable** *iff $S \approx \mathbb{N}$. The cardnality of $S \approx \mathbb{N}$ is said to be $\aleph_o$. We denote $\overline{\overline{S}} = \aleph_o$ iff $S \approx \mathbb{N}$.*

The following is a list of sets with cardnality $\aleph_o$,

$$\mathbb{N},\ 2\mathbb{N},\ 3\mathbb{N},\ \mathbb{Z},\ 2\mathbb{Z},\ \mathbb{N}\times\mathbb{N},\ \mathbb{N}^{123},\ \{x\in\mathbb{R}\mid \sin(x)=0\}, \{\frac{2}{n}\mid n\in\mathbb{N}\}$$

I don't find any of the examples above too surprising. These are all manifestly discrete sets. If you visualize them there is clearly gaps between adjacent values in the sets. In contrast, think about the rational numbers. Given any two rational numbers we can always find another in between them: given $p/q, m/n \in \mathbb{Q}$ we find

$$\frac{1}{2}\left(\frac{p}{q}+\frac{m}{n}\right)=\frac{1}{2}\left(\frac{pn+qm}{nq}\right)\in\mathbb{Q}$$

at the midpoint between p/q and m/n on the number line. It would seem there are more rational numbers then natural numbers. However, things are not always what they "seem". Cantor gave a bijection between $\mathbb{N}$ and postive rational numbers (see Figure 5.8 in text). Once you have that it's not hard to prove

$$\boxed{\overline{\overline{\mathbb{Q}}}=\aleph_o.}$$

I may have slipped in lecture and referred to a denumerable set as being "countably infinite", here is why:

Definition 5.9 (countable)**.** *A set S is said to be* **countable** *iff S is finite or denumerable. If a set S is not countable then it is said to be* **uncountable**.

5.3 uncountably infinite sets

The title of this section is somewhat superfluous since every uncountable set is necessarily infinite. Uncountable sets are quite common.

Theorem 5.10 (this is Theorem 5.14 in the text)**.** *The open interval $(0,1)$ is uncountable.*

You can see the text for the proof, basically it stems from the decimal expansion of the real numbers. This Theorem assures us the following definition is worthwhile:

Definition 5.11 (continuum c)**.** *We define the cardnality of the open interval* $(0,1)$ *to be* c*.*

The proof $(0,1)$ is uncountable is not too easy, but once you have the unit interval it's easy to get other subsets of $\mathbb{R}$.

Example 5.12. *Show* $(0,1) \approx (5,8)$*. To do this we want a one-one mapping that takes* $(0,1)$ *as its domain and* $(5,8)$ *as its range. A line segment will do quite nicely. Let* $f(x) = mx + b$ *and fit the points*

$$f(0) = 5 = b, \quad f(5) = 5m + 5 = 8$$

Clearly $f(x) = \frac{3}{5}x + 5$ *will provide a bijection of the open intervals. Its not hard to see this construction works just the same for any open interval* (a,b)*. Thus the cardnality of any open interval is* c*.*

There are bijections from the open interval to closed intervals and half-open half-closed intervals not too mention unions of such things. These mappings are not always as easy to find. The text shows how to dodge the construction through a series of insightful theorems which we are skipping.

Example 5.13. *Show* $(0,1) \approx \mathbb{R}$*. First observe that* $(0,1) \approx (-\frac{\pi}{2}, \frac{\pi}{2})$ *thus by transitivity of* $\approx$ *is suffices to show that* $(-\frac{\pi}{2}, \frac{\pi}{2}) \approx \mathbb{R}$*. The graph of inverse tangent comes to mind, it suggests we use*

$$f(x) = \tan^{-1}(x)$$

This mapping has $dom(f) = \mathbb{R}$ *and* $range(f) = (-\frac{\pi}{2}, \frac{\pi}{2})$*. This can be gleaned from the relation between a function and its inverse. The vertical asymptotes of tangent flip to become horizontal tangents of the inverse function. Notice that*

$$f(a) = f(b) \;\Rightarrow\; \tan^{-1}(a) = \tan^{-1}(b) \;\Rightarrow\; a = b$$

by the graph and definition of inverse tangent. Also, if $y \in (-\frac{\pi}{2}, \frac{\pi}{2})$ *then clearly* $f(\tan(y)) = y$ *hence* f *is onto.*

We should skip the next example in lecture. I just wanted to try this example directly for fun.

Example 5.14. *Show that* $[0,1] \approx (0,1)$*. Well, we already have that* $(0,1) \approx \mathbb{R}$ *so if I can find a mapping from* $[0,1]$ *to* $\mathbb{R}$ *which is a bijection then we're done. Let's think. Still thinking. Nothing comes to mind right away, I'll leave this for you. You can earn a bonus point if you can find a bijection which demonstrated* $(0,1) \approx [0,1]$*.*

5.4 Cantor's Theorem and transfinite arithmetic

Definition 5.15. *Let A and B be sets. Then*

1. $\overline{\overline{A}} = \overline{\overline{B}}$ *iff* $A \approx B$, *otherwise* $\overline{\overline{A}} \neq \overline{\overline{B}}$
2. $\overline{\overline{A}} \leq \overline{\overline{B}}$ *iff there exists an injection* $f : A \to B$
3. $\overline{\overline{A}} < \overline{\overline{B}}$ *iff* $\overline{\overline{A}} \leq \overline{\overline{B}}$ *and* $\overline{\overline{A}} \neq \overline{\overline{B}}$

Notice that the injection took A as its domain. The direction is important here, it is not assumed that $f(A) = B$ in part (2.). Thus, while we can form an inverse function from $range(f)$ to A that will not be a bijection from B to A since $range(f)$ may not equal B in general. Transfinite arithmetic enjoys many of the same rules as ordinary arithmetic, see Theorem 5.30 for details.

Theorem 5.16 (Cantor's Theorem)**.** *For every set A, $\overline{\overline{A}} < \overline{\overline{\mathcal{P}(A)}}$.*

It then follows that we have an unending string of infinities:

$$\aleph_o = \overline{\overline{\mathbb{N}}} < \overline{\overline{\mathcal{P}(\mathbb{N})}} < \overline{\overline{\mathcal{P}(\mathcal{P}(\mathbb{N}))}} < \overline{\overline{\mathcal{P}(\mathcal{P}(\mathcal{P}(\mathbb{N})))}} < \cdots$$

An obvious question to ask is "where does the continuum c fit into this picture? It can be shown that $\aleph_o < c$. To see this, note $\aleph_o \leq c$ since $\mathbb{N} \subseteq \mathbb{R}$ we can restrict the identity function to an injection from $\mathbb{N}$ into $\mathbb{R}$ and since $\mathbb{R}$ is not equivalent to $\mathbb{N}$ we have that $\aleph_o < c$.

Theorem 5.17 (Cantor-Schröder-Bernstein Theorem)**.** *If $\overline{\overline{A}} \leq \overline{\overline{B}}$ and $\overline{\overline{B}} \leq \overline{\overline{A}}$, then $\overline{\overline{A}} = \overline{\overline{B}}$.*

This is a non-trivial Theorem despite it's humble appearance.
Application of Theorem: The text gives and argument that shows $\overline{\overline{\mathcal{P}(\mathbb{N})}} \leq \overline{\overline{(0,1)}}$ and $\overline{\overline{\mathcal{P}(\mathbb{N})}} \geq \overline{\overline{(0,1)}}$ (see pg. 250). Thus $\overline{\overline{\mathcal{P}(\mathbb{N})}} = \overline{\overline{(0,1)}} = c$.

Definition 5.18 (trichotomy property of $\mathbb{N}$)**.** *If $m, n \in NN$ then $m > n$, $m = n$, or $m < n$*

Theorem 5.19 (this is Theorem 5.34 in the text, Comparability Theorem)**.** *If A and B are any two sets, then $\overline{\overline{A}} > \overline{\overline{B}}$, $\overline{\overline{A}} = \overline{\overline{B}}$, or $\overline{\overline{A}} < \overline{\overline{B}}$.*

It turns out that it is impossible to prove this Theorem in the Zermelo Fraenkel set theory unless we assume the Axiom of Choice is true.

Theorem 5.20. *If $\mathcal{A}$ is a collection of non-empty sets, then there exists a function F (the* **choice** *function) from $\mathcal{A}$ to $\cup_{A\in\mathcal{A}}A$ such that for every $A \in \mathcal{A}$ we have $F(A) \in A$.*

This axiom does lead to some unusual results. For example, the Banach-Tariski paradox which says that a ball can be cut into pieces and reassembled into two balls such that the total volume is doubled. (don't worry these "cuts" are not physically reasonable). Or the weird result that every subset of $\mathbb{R}$ can be reordered such that it has a smallest element.

Theorem 5.21 (this is Theorem 5.35)**.** *If there exists a function from a set A onto a set B, then $\overline{\overline{B}} \leq \overline{\overline{A}}$.*

Notice surjectivity suggests that there is at least one thing in the domain to map to each element in the range B. It could be the case that more than one thing maps to each element in B, but certainly at least one thing in A maps to a given element in B. If the fibers in A are really "big" then inequality in the Theorem would become a strict $<$. The proof of this Theorem given in the text involves choosing something in the fiber.

Remark 5.22. Confession: *we used the axiom of choice when we constructed the injective function for an arbitrary function $f : A \to B$. In principle there were infinitely many fibers, we claimed that there existed a section that cut through the fibers such that each fiber was intersected just once. The choice function gives us the existence of such a section. Notice the non-constructive nature of that particular corner of the argument. We have no specific mechanism to select an element of the fiber. Now for particular examples the choice function can be explicitly constructed and in such a context we wouldn't really insist we were relying on the Axiom of Choice (like the mobius band example I showed in lecture).*

Remark 5.23 (Continuum Hypothesis)**.** *The* **Continuum Hypothesis** *states that c is the next transfinite number beyond $\aleph_o$. There is no other infinite set between the rationals and the reals. This was conjectured by Cantor, but only later did the work of Godel(1930's) and Cohen(1960's) elucidate the issue. Godel showed that the Continuum Hypothesis was* **undecidable** *but relatively consistent in Zermelo Frankael set-theory. Then later Paul Cohen showed that the Continuum Hypothesis was independent of the Axiom of Choice relative to Zermelo-Frankael set theory modulo the Axiom of Choice. The Continuum Hypothesis and the Axiom of Choice continue to be widely believed since they are important "big guns" for certain crucial steps in hard theorems.*

Well, I hope you don't let this Chapter influence your expectations of my expectations for other Chapters too much. I have taken a very laid-back attitude about proofs in this Chapter. The other Chapters are more important to gain technical prowess in. This material is more about being "well-rounded" mathematically speaking.

6 Modular Arithmetic

The formal study of modular arithmetic is usually attributed to Carl Friedrich Gauss. He introduced the idea in his book *Disquisitiones Arithmeticae* in around 1801. History aside, modular arithmetic is familar to us in many ways. Clocks and calendars have a periodicity that allows us to group certain days, dates or times as being all part of the same class. The rules by which we add hours and minutes are examples of modular arithmatic. We will discuss how $\mathbb{Z}_n$ is formed from equivalences classes of $\mathbb{Z}$. The equivalence classes that comprise $\mathbb{Z}_n$ are quite special because they can be added, subtracted, multiplied and *sometimes* divided. In other words, the equivalence classes that make up $\mathbb{Z}_n$ behave just like numbers in $\mathbb{R}$. However, there is a difference. We will find equations such as $1 + 1 = 0$ and $2 \cdot 3 = 0$ make sense in a certain context.

Let me begin with an example (contrary to my natural tendencies)

Example 6.1. *The following calculations are performed in* $\mathbb{Z}_5$*:*

$$2 + 3 = 5 \equiv 0, \qquad (2)(3) = 6 \equiv 1, \qquad (2)(4) = 8 \equiv 3$$

Here the basic idea is that numbers which differ by 5 *units are equivalent. I'm using the notation* $\equiv$ *is denote when two things are* **congruent**. *Since* $0 \equiv 5 \equiv -100 \equiv 25 \equiv \cdots$ *there is much ambiguity in the presentation of an equation "mod 5".*

Now, I would like to explain what $\mathbb{Z}_5$ really is. We want to construct the object and justify that the mathematics are placed on a solid foundation. I also aim to equip you with a few basic tools from elementary number theory. The example of $\mathbb{Z}_n$ is important for abstract algebra and computer science. Many of the questions we touch on here are at the beginning of lengthy algorthimics discussions. We will not delve into algorithmics, but you should be aware that there is much to learn here about how one actually codes these constructions into the computer.

6.1 Basics of $\mathbb{Z}$

Let's start at the very beginning, it is a good place to start.

Definition 6.2. *The integers* $\mathbb{Z}$ *are the set of natural numbers* $\mathbb{N}$ *together with* 0 *and the negatives of* $\mathbb{N}$*. It is possible to concretely construct (we will not) these from sets and set-operations.*

From the construction of $\mathbb{Z}$ it is clear (we assume these to be true)

1. the sum of integers is an integer
2. the product of integers is an integer
3. the usual rules of arithmetic hold for $\mathbb{Z}$

Much is hidden in (3.): let me elaborate, we assume for all $a, b, c \in \mathbb{Z}$,

$$a + b = b + a \tag{9}$$
$$ab = ba \tag{10}$$
$$a(b + c) = ab + ac \tag{11}$$
$$(a + b)c = ac + bc \tag{12}$$
$$(a + b) + c = a + (b + c) \tag{13}$$
$$(ab)c = a(bc) \tag{14}$$
$$a + 0 = 0 + a = a \tag{15}$$
$$1a = a1 \tag{16}$$

Where we assume the **order of operations** is done multiplication then addition; so, for example, $ab + ac$ means to first multiply a with b and a with c then you add the result. Maybe you haven't thought about this for while, maybe not since the 1-st or 2-nd grade.

6.2 division algorithm

Division is repeated subtraction. For example, consider $11/3$. Notice

$$11 - 3 = 8 \qquad 8 - 3 = 5 \qquad 5 - 3 = 2$$

then we cannot subtract anymore. We were able to subtract 3 copies of 3 from 11. Then we stopped at 2 since $2 < 3$. To summarize,

$$\boxed{11 = 3(3) + 2}$$

Usually this quantity 2 is called the **remainder** and we present the calculation as follows:

$$\frac{11}{3} = 3 + \frac{2}{3}$$

The generalization of the boxed equation for an arbitrary pair of natural numbers is known as the **division algorithm**.

Theorem 6.3. *(The Division Algorithm) If $a, b \in \mathbb{N}$ such that $b \leq a$, then there exists $q \in \mathbb{N}$ (the* **quotient***) and $r \in \mathbb{N} \cup \{0\}$ (the* **remainder***) such that $a = bq + r$ and $0 \leq r < b$.*

Proof: Long division. Or, argue as I do below:

Case 1: If $b|a$ then the Theorem is clearly true. If $b|a$ then there exists $q \in \mathbb{N}$ such that $a = bq$ so we can identify that $r = 0$ and the division algorithm holds true.

Case 2: Suppose b does not divide a. Construct $S \subset \mathbb{N}$ as follows,

$$S = \{r \in \mathbb{N} \mid \exists q \in \mathbb{N} \ s.t. \ r = a - bq\}.$$

Observe that S is nonempty since $a - b > 0$ and the difference of natural numbers is again a natural number, thus $(a - b) \in \mathbb{N}$. Furthermore, $(a - b) \in S$ since $a - b = a - b \cdot 1$ (*think $q = 1 \in \mathbb{N}$ in the definition of S*). Thus S is a non-empty subset of $\mathbb{N}$ so by the WOP there exists a smallest element $r_o \in S$. Since $r_o \in S$ there exists $q_o \in \mathbb{N}$ such that $r_o = a - bq_o$.

Now, suppose (*towards a contradiction*) that $r_o \geq b$. Observe that $r_o - b \geq 0$. If $r_o - b = 0$ then $r_o = b$ and it follows that $a = bq_o + b = b(q_o + 1)$ thus $b|a$ which contradicts our assumption b does not divide a. In addition notice that if $r_o > b$ then we can argue that $r_o - b \in S$. To see this first notice $r_o - b > 0$ so $r_o - b \in \mathbb{N}$, second $r_o = a - bq_o$ and consequently subtracting b from both sides we find $r_o - b = a - bq_o - b = a - b(q_o + 1)$. Clearly $q_o + 1 \in \mathbb{N}$ thus it follows that $r_o - b \in S$. Notice that $r_o - b < r_o$ which is contradicts the fact that r_o is the smallest element of S.

Hence the assmption $r_o \geq b$ forces a contradiction. We conclude that $r_o < b$. Therefore, in all possible cases, there exist $q, r \in \mathbb{N}$ such that $a = qb + r$ with $0 \leq r < b$.

The proof I just gave was rather *non-constructive*. The most critical point was arrived at by contradiction. There is another way to prove the division algorithm. You can set up a loop which subtracts b from a then does it again, and again until the result is less than b. You could write a flow-chart of that algorithm to illustrate the program. You would still need to argue that the loop is not endless, but that is fairly clear since $a < \infty$.

6.3 definition of $\mathbb{Z}_n$ and modular arithmetic

Let $n \in \mathbb{N}$ be a fixed natural number in the discussion that follows. Define a relation R_n for $a, b \in \mathbb{Z}$ by:

$$\boxed{aR_nb \ \Leftrightarrow n|(a-b)}$$

Proposition 6.4. *The relation R_n is an equivalence relation on $\mathbb{Z}$*

Proof: It is clear that $R_n \subseteq \mathbb{Z} \times \mathbb{Z}$. Furthermore,

1. Let $a \in \mathbb{Z}$ then $a - a = 0$ hence aR_na. Thus R_n is reflexive.

2. If aR_nb then $n|(a-b)$ so there exists $k \in \mathbb{Z}$ such that $a - b = nk$. Hence $b - a = n(-k)$ thus $n|(b-a)$ and we find bR_na. This proves R_n is symmetric.

3. If aR_nb and bR_nc then there exist $k, l \in \mathbb{Z}$ such that $a - b = nk$ and $c - b = nl$. Observe $a - c = (b + nk) - (b + nl) = n(k - l)$ hence aR_nc. Thus R_n is transitive.

Therefore R_n is an equivalence relation on $\mathbb{Z}$.

Remark 6.5. *(notation) The equivalence relation R_n is usually denoted $\equiv$. This is called* **congruence modulo** *n. The equivalence classes are defined as before: for $k \in \mathbb{Z}$ define*

$$\boxed{k/R_n = \ k/\equiv = \ \overline{k} \ = [k] = \{m \in \mathbb{Z} \mid n|(m-k)\} = \{np + k \mid p \in \mathbb{Z}\}}$$

Definition 6.6. *$\mathbb{Z}_n$ is defined to be the space of equivalence classes of the congruence modulo n relation on $\mathbb{Z}$. In short, using the notation discussed in Chapter 3 of these notes, $\mathbb{Z}_n = \mathbb{Z}/R_n$.*

Example 6.7. *Let's describe $\mathbb{Z}_3$ in gory detail. This is made of three equivalence classes.*

$$\overline{0} = \{\ldots, -6, -3, 0, 3, 6, \ldots\} = 3\mathbb{Z} + 0$$

$$\overline{1} = \{\ldots, -5, -2, 1, 4, 7, \ldots\} = 3\mathbb{Z} + 1$$

$$\overline{2} = \{\ldots, -4, -1, 2, 5, 8, \ldots\} = 3\mathbb{Z} + 2$$

Hence, $\mathbb{Z} = \{\overline{0}, \overline{1}, \overline{2}\}$.

Remark 6.8. *The* **cannonical representatives** *of $\mathbb{Z}_n$ are naturally $\mathbb{N}_{n-1} \cup \{0\}$. We can use bars or square brackets or nothing at all; that is,*

$$\mathbb{Z}_n = \{0, 1, \ldots, n-1\} = \{\overline{0}, \overline{1}, \ldots, \overline{n-1}\} = \{[0], [1], \ldots, [n-1]\}$$

Sometimes we will use one notation for $\mathbb{Z}_p$ and another for $\mathbb{Z}_q$ in the case $p \neq q$ and we have need to discuss both at once.

Definition 6.9. *(addition and multiplication in $\mathbb{Z}_n$) Addition is a function $+_n : \mathbb{Z}_n \times \mathbb{Z}_n \to \mathbb{Z}_n$ defined by*

$$\boxed{\overline{x} +_n \overline{y} = \overline{x+y}}$$

Multiplication is a function $\cdot_n : \mathbb{Z}_n \times \mathbb{Z}_n \to \mathbb{Z}_n$ defined by

$$\boxed{\overline{x} \cdot_n \overline{y} = \overline{xy}}$$

The addition and multiplication taking place on the RHS of the boxed equations is addition and multiplication in $\mathbb{Z}$ which we assume is well-defined. The definitions above simply state to calculate in $\mathbb{Z}_n$ you can take the representatives of the given equivalence classes and do the corresponding operation in $\mathbb{Z}$ then assign the equivalence class of that resulting integer. It is not at all obvious that these operations are well-defined. We need to check to make sure that the result does not depend on the choice of representative.

For example, in $\mathbb{Z}_3$ we can calculate

$$\overline{2} \cdot_3 \overline{2} = \overline{(2)(2)} = \overline{4}$$

Or we could calculate,

$$\overline{5} \cdot_3 \overline{2} = \overline{(5)(2)} = \overline{10}$$

Notice $\overline{4} = \overline{10} = 3\mathbb{Z} + 1$, these are the same subset of $\mathbb{Z}$. The calculation is independent of the representative chosen to label the class. Of course, this is an example not a proof of the general case.

Theorem 6.10. *$\mathbb{Z}_n$ has a well-defined addition $+_n$ and multiplication $\cdot_n$. In particular, for every $x, y \in \mathbb{Z}$ the operations*

$$\overline{x} +_n \overline{y} = \overline{x+y} \qquad\qquad \overline{x} \cdot_n \overline{y} = \overline{xy}$$

are well-defined. Here $\overline{x}$ is the equivalence class of $x \in \mathbb{Z}$ where the equivalence relation is congruence modulo n and $n \in \mathbb{N}$.

Proof: Let $x, y \in \mathbb{Z}$ then $x + y, xy \in \mathbb{Z}$ thus $\overline{x}, \overline{y}, \overline{x+y}, \overline{xy}$ exist in $\mathbb{Z}_n$. We need only check that the formulas are independent of representative.

Begin with addition modulo n. Let $\overline{a} = \overline{x}$ and $\overline{b} = \overline{y}$ for any $a \in \overline{x}$ and $b \in \overline{y}$. Note that,

1. $\overline{a} = \overline{x}$ and $\overline{b} = \overline{y}$ implies $\overline{a} +_n \overline{b} = \overline{x} +_n \overline{y}$
2. $\overline{a} +_n \overline{b} = \overline{a+b}$ and $\overline{x} +_n \overline{y} = \overline{x+y}$
3. We should find $\overline{a+b} = \overline{x+y}$.

Since $a, x \in \overline{x}$ and $b, y \in \overline{y}$ there exist $k, l \in \mathbb{Z}$ such that $x - a = kn$ and $y - b = ln$ hence $x + y = (kn + a) + (ln + b) = a + b + n(l + k)$. We find

$$(x + y) - (a + b) = n(l + k) \;\Rightarrow\; n|[(x + y) - (a + b)]$$

this means $x + y$ and $a + b$ are congruent mod n. Thus

$$\overline{a+b} = \overline{x+y}$$

and we conclude that the formula is independent of representatives.

Likewise, if $\overline{x} = \overline{a}$ and $\overline{y} = \overline{b}$ then $\overline{xy} = \overline{ab}$ so multiplication modulo n is well-defined. To see this note $\overline{x} = \overline{a}$ and $\overline{y} = \overline{b}$ imply $x - a = nk$ and $y - b = nl$. Observe then that $xy = (a + nk)(b + nl) = ab + n(kb + al + kl)$ which means $n|(xy - ab)$ or simply $xy \equiv ab$. It follows that $\overline{xy} = \overline{ab}$.

To summarize, these calculations show that $+_n$ and $\cdot_n$ are truly **functions** from $\mathbb{Z}_n \times \mathbb{Z}_n$ to $\mathbb{Z}_n$.

Example 6.11. *The addition and multiplication and addition for $\mathbb{Z}_2$ can be written in tabular form as follows: (denoting $\overline{0}$ by 0 and $\overline{1}$ by 1 since there is no danger of confusion here)*

$+_2$	0	1
0	0	1
1	1	0

$\cdot_2$	0	1
0	0	0
1	0	1

Remark 6.12 (concerning the many notations for $\mathbb{Z}_n$)**.** *The table above indicates that $1 + 1 = 0$ in $\mathbb{Z}_2$. In other words, $1 + 1 = 0$ mod 2. Or, we could say that $1 + 1 \equiv 0$. The symbol "$\equiv$" should be read "congruent" in this context. Which notation is best? I think it depends on the context. When I*

am proving things about the structure of $\mathbb{Z}_n$ it helps to have a notation which distinguishes $\overline{1}$ from 1. However, if I am doing a calculation which is just involving $\mathbb{Z}_n$ then the $\overline{1}$-type notation gets old really quick. If I am going to omit the bar notation then I usually will put a sentence at the beginning of the calculation alerting the reader to the fact that "the calculations below are done modulo n". Finally, I should mention that is customary to use $\{0, 1, 2, \ldots n-1\}$ to represent the equivalence classes in $\mathbb{Z}_n$, I will refer to this as the **cannonical representation of** $\mathbb{Z}_n$*. Non-standard representations are also logical but when you do not simplify to the cannonical representatives it may hide certain patterns that would otherwise be obvious. So, it is a good habit to write the final answer in the cannonical representation.*

It is of course fine to write in-between steps in $\mathbb{Z}$ where it is easier to calculate sometimes. For example, in $\mathbb{Z}_5$,

$$4^2 + 3 = 16 + 3 = 19 = \boxed{4}$$

It wouldn't be best to leave the answer as 19. In $\mathbb{Z}_5$ we ought to leave the answer as $0, 1, 2, 3$ or 4. Also sometimes it pays it simplify as you go along, again in $\mathbb{Z}_5$ I calculate,

$$4^{20} = (4^2)^{10} = (16)^{10} = 1^{10} = \boxed{1}.$$

Obviously calculating 4^{20} directly then simplifying would have been a much harder calculation.

Example 6.13. *The addition and multiplication and addition in $\mathbb{Z}_3$ can be written in tabular form as follows:*

$+_3$	0	1	2
0	0	1	2
1	1	2	0
2	2	0	1

$\cdot_3$	0	1	2
0	0	0	0
1	0	1	2
2	0	2	1

Example 6.14. *The addition and multiplication and addition in $\mathbb{Z}_4$ can be written in tabular form as follows:*

$+_4$	0	1	2	3
0	0	1	2	3
1	1	2	3	0
2	2	3	0	1
3	3	0	1	2

$\cdot_4$	0	1	2	3
0	0	0	0	0
1	0	1	2	3
2	0	2	0	2
3	0	3	2	1

Example 6.15. *The addition and multiplication and addition in $\mathbb{Z}_5$ can be written in tabular form as follows:*

$+_5$	0	1	2	3	4
0	0	1	2	3	4
1	1	2	3	4	0
2	2	3	4	0	1
3	3	4	0	1	2
4	4	0	1	2	3

$\cdot_5$	0	1	2	3	4
0	0	0	0	0	0
1	0	1	2	3	4
2	0	2	4	1	3
3	0	3	1	4	2
4	0	4	3	2	1

Example 6.16. *The addition and multiplication and addition in $\mathbb{Z}_6$ can be written in tabular form as follows:*

$+_6$	0	1	2	3	4	5
0	0	1	2	3	4	5
1	1	2	3	4	5	0
2	2	3	4	5	0	1
3	3	4	5	0	1	2
4	4	5	0	1	2	3
5	5	0	1	2	3	4

$\cdot_6$	0	1	2	3	4	5
0	0	0	0	0	0	0
1	0	1	2	3	4	5
2	0	2	4	0	2	4
3	0	3	0	3	0	3
4	0	4	2	0	4	2
5	0	5	4	3	2	1

Example 6.17. *The addition and multiplication and addition in $\mathbb{Z}_7$ can be written in tabular form as follows:*

$+_7$	0	1	2	3	4	5	6
0	0	1	2	3	4	5	6
1	1	2	3	4	5	6	0
2	2	3	4	5	6	0	1
3	3	4	5	6	0	1	2
4	4	5	6	0	1	2	3
5	5	6	0	1	2	3	4
6	6	0	1	2	3	4	5

$\cdot_7$	0	1	2	3	4	5	6
0	0	0	0	0	0	0	6
1	0	1	2	3	4	5	6
2	0	2	4	6	1	3	5
3	0	3	6	2	5	1	4
4	0	4	1	5	2	6	3
5	0	5	3	1	6	4	2
6	0	6	5	4	3	2	1

Example 6.18. *Find solutions(if possible) of $x^2+1=0$ in $\mathbb{Z}_2$. Notice that in $\mathbb{Z}_2$ we have that $2=0$ thus $2x=0$. Consequently, $x^2+1=x^2+2x+1=(x+1)^2=0$. This has solution $x=-1$ which is otherwise known as $x=1$ in $\mathbb{Z}_2$. To summarize, the quadratic equation $x^2+1=0$ has just the $x=1$ solution in $\mathbb{Z}_2$.*

When we try to solve polynomial equations over $\mathbb{Z}_n$ things get weird generally speaking.

Example 6.19. *Does the polynomial $f(x) = x^3+3x^2+7$ have any solutions in $\mathbb{Z}_6$? Well, one solution is just to check all the possiblilities:*

$$\begin{aligned}
f(0) &= 7 \\
f(1) &= 1+3+7 = 10 = 4 \\
f(2) &= 8+3(4)+7 = 27 = 3 \\
f(3) &= 27+3(9)+7 = 3+3+7 = 13 = 1 \\
f(4) &= 4^3+3(16)+7 = (-2)^3+3(4)+7 = -8+12+1 = 5 \\
f(5) &= 5^3+3(25)+7 = (-1)^3+3(1)+1 = 3
\end{aligned}$$

I have made extensive use of identities such as $x = x-6$ in $\mathbb{Z}_6$. For example, I noted $5 = -1$ and $4 = -2$ to simplify the calculation of $f(4)$ and $f(5)$. We conclude that $f(x)$ has no solutions in $\mathbb{Z}_6$.

6.4 euclidean algorithm

Definition 6.20. *(greatest common divisor) Let $a, b \in \mathbb{N}$ and suppose that $d \in \mathbb{N}$ divides both a and b then we say that d is a* **common divisor** *of a and b. If d is a common divisor of a and b such that if c is another common divisor then $c \leq d$ then we say that d is the* **greatest common divisor** *of a, b and we write $gcd(a, b) = d$*

Notice that any two natural numbers will always have the common divisor of 1 thus there is always common divisor. For example,

$$gcd(2, 6) = 2 \qquad gcd(100, 150) = 50 \qquad gcd(3, 5) = 1.$$

The case that 1 is the greatest divisor is so special it get's a name.

Definition 6.21. *(relatively prime or co-prime) Let $a, b \in \mathbb{N}$ then we say that a and b are* **relatively prime** *iff $gcd(a, b) = 1$.*

Relatively prime numbers share no common factors except 1.

Remark 6.22. *I leave the proof of the existence of the $gcd(a, b)$ for arbitrary $a, b \in \mathbb{N}$ to the reader. It will follow easily from the unique factorization of integers into powers of primes upto reordering.*

Example 6.23. *Let me show you how the euclidean algorithm works for a simple example. Consider* $a = 100$ *and* $b = 44$*. Euclid's algorithm will allow us to find* $gcd(100, 44)$*.*

1. $100 = 44(2) + 12$ *divided 100 by 44 got remainder of 12*
2. $44 = 12(3) + 8$ *divided 44 by 12 got remainder of 8*
3. $12 = 8(1) + \boxed{4}$ *divided 12 by 8 got remainder of 4*
4. $4 = 4(1) + 0$ *divided 4 by 1 got remainder of zero*

The last nonzero remainder will always be the gcd when you play the game we just played. Here we find $\boxed{gcd(100, 44) = 4}$*. Moreover, we can write* 4 *as a* $\mathbb{Z}$*-linear combination of* 100 *and* 44*. This can be gleaned from the calculations already presented by working backwards from the gcd:*

3. $4 = 12 - 8$
2. $8 = 44 - 12(3)$ *implies* $4 = 12 - (44 - 12(3)) = 4(12) - 44$
1. $12 = 100 - 44(2)$ *implies* $4 = 4(100 - 44(2)) - 44 = 4(100) - 9(44)$

I call this a "$\mathbb{Z}$*-linear combination of* 100 *and* 44 *since* $4, -9 \in \mathbb{Z}$*. We find* $\boxed{4(100) - 9(44) = 4}$*.*

Remark 6.24. *The fact that we can always work euclid's algorithm backwards to find how the* $gcd(a, b)$ *is written as* $ax + by = gcd(a, b)$ *for some* $x, y \in \mathbb{Z}$ *is remarkable. Personally, I can find the greatest common divisor without euclid's algorithm pretty quick. I just factor* a *and* b *and pick off the largest set of factors which is shared by both. Finding* x, y *to form the linear combination is not nearly as obvious for me.*

Example 6.25. *Find* $gcd(4, 20)$*. This example is a bit silly, but I include it since it is an exceptional case in the algorithm. The algorithm works, you just need to interpret the instructions correctly.*

$$20 = 4(5) + 0$$

Since there is only one row to go from we identify 4 *as playing the same role as the last non-zero remainder in most examples. Clearly,* $gcd(4, 20) = 4$*. Now, what about working backwards? Since we do not have the gcd appearing by itself in the next to last equation (as we did in the last example) we are forced to solve the given equation for the gcd,*

$$20 = 4(4 + 1) = 4(4) + 4 \implies \boxed{20 - 4(4) = 4}$$

Remark 6.26. *(generalizing the preceding example) It is not too hard to believe that we can replicate the idea of the last example for any pair a, b such that $a|b$. If $a|b$ then $b = ka$ for some $k \in \mathbb{N}$ and as $a|a$ and $a|b$ it follows that a is a common divisor. Moreover it is the largest divisor since no number larger than a could ever divide a. Finally, notice $b = ka = (k-1)a + a$ hence $b - (k-1)a = a = gcd(a, b)$.*

Theorem 6.27. *(euclidean algorithm: Theorem 1.6 in the text) Let $a, b \in \mathbb{N}$ such that $b \leq a$. Suppose $d = gcd(a, b)$. Then:*

1. *There exist two finite lists of postive integers $\{q_i\}$ and $\{r_i\}$ such that $b > r_1 > r_2 > \cdots > r_k > r_{k+1} = 0$ and*

$$\begin{aligned} a &= bq_1 + r_1 \\ b &= r_1q_2 + r_2 \\ r_1 &= r_2q_3 + r_3 \\ r_2 &= r_3q_4 + r_4 \\ &\cdots\cdots \\ r_{k-2} &= r_{k-1}q_k + r_k \\ r_{k-1} &= r_kq_{k+1} \end{aligned}$$

   and $d = gcd(a, b) = r_k$

2. *The $gcd(a, b)$ may be written as a $\mathbb{Z}$-linear combination of a and b; that is, there exist $x, y \in \mathbb{Z}$ such that $gcd(a, b) = d = ax + by$.*

Proof: Let $a, b \in \mathbb{N}$ such that $b \leq a$. Then we can apply the division algorithm to find $q_1, r_1 \in \mathbb{N}$ such that $0 \leq r_1 < b$ and

$$a = bq_1 + r_1$$

Note $0 \leq r_1 < b$. If $r_1 = 0$ then we are done. Otherwise, $r_1 < b$ thus I can apply the division algorithm to find $q_2, r_2 \in \mathbb{N}$ such that $0 \leq r_2 < r_1$ and

$$b = r_1q_2 + r_2.$$

Note $0 \leq r_2 < r_1 < b$. If $r_2 = 0$ then we are done. Otherwise, $r_2 < r_1 < b$ and we can apply the division algorithm to r_2, r_1 to find $q_3, r_3 \in \mathbb{N}$ such that $0 \leq r_3 < r_2$ and

$$r_1 = r_2q_3 + r_3.$$

Note $0 \leq r_3 < r_2 < r_1 < b$. If $r_3 = 0$ then we are done. Otherwise continue the pattern until we find $r_{k+1} = 0$. We claim that r_k is the $gcd(a, b)$ by

nature of its construction.

What? why does the pattern have to terminate? How do we know that it will arrive at zero for r_{k+1}? Is this obvious? Actually, its not too hard to see if you focus on the critical data here

$$b > r_1 > r_2 > r_3 > r_4 > \cdots \geq 0$$

We have a list of strictly decreasing positive integers that start at b. The best case scenario is that the r_i's differ by 1. In that case the list has $b+1$ elements and it follows $r_b = 0$. The strictly decreasing condition is the crucial detail here.

Notice that $r_{k-1} = r_k q_{k+1}$ shows that $r_k | r_{k-1}$. Furthermore,

$$r_{k-2} = r_{k-1}q_k + r_k \;\Rightarrow\; r_{k-2} = r_k q_{k+1} q_k + r_k = (q_{k+1}q_k + 1)r_k.$$

Hence, $r_k | r_{k-2}$. But, we also know that

$$r_{k-3} = r_{k-2}q_{k-1} + r_{k-1} \;\Rightarrow\; r_{k-3} = (q_{k+1}q_k + 1)r_k q_{k-1} + r_k q_{k+1} = c_{k-3} r_k$$

where c_{k-3} is defined implicitly by the equation just given. This shows $r_k | r_{k-3}$. We continue in this fashion until we get up to

$$r_1 = r_2 q_3 + r_3$$

which yields $r_1 = c_1 r_k$ where c_1 has been formed by back-substituting the equations in the algorithm from the k-th to $3 - rd$ level. Next

$$b = r_1 q_2 + r_2$$

yields that $b = c_1 r_k q_2 + c_2 r_k = (c_1 q_2 + c_2) r_k = c_0 r_k$ thus $r_k | b$. And finally,

$$a = bq_1 + r_1$$

yields that $a = c_0 r_k q_1 + c_1 r_k = (c_0 q_1 + c_1) r_k$ so $r_k | a$. This proves that r_k is a **common divisor** of a and b.

Let $d = gcd(a, b)$ then by definition of **greatest common divisor** we must have $r_k \leq d$ as r_k is a common divisor of a and b. Notice that $a - bq_1 = r_1$ thus $d | r_1$ (*by a homework you did, if $d|x$ and $d|y$ then $d|mx + ny$, or see Lemma 1.5 in the text*). Likewise, $b - r_1 q_2 = r_2$ thus $d | r_2$. Continuing we find $d | r_k$. This means there exists $m \in \mathbb{N}$ such that $r_k = md$ which implies

$r_k \geq d$. Observe $r_k \leq d$ and $r_k \geq d$, therefore $r_k = d = gcd(a, b)$.

The proof that there exist $x, y \in \mathbb{Z}$ such that $ax + by = gcd(a, b) = r_k$ follows from working backwards on the stack of equations. Start with $r_k = r_{k-2} - r_{k-1}q_k$ then eliminate the r_m's by substitution until all that is left is a, b and r_k. **Proof of euclidean algorithm ends here.**

Remark 6.28. *The proof of the Euclidean algorithm is tedious but not really that difficult if you take the time to understand it. I should mention that the method of proof is not induction. You might be tempted to think that since we work on a pattern which stays the same from iteration to iteration. However, the crucial difference is that there are finitely many steps here. The $\cdots$ does not indicate an infinity of future equations, rather it is just a shorthand for saying that there are a finite, but unknown, set of equations in the algorithm. This does not require induction. (there might be a slick way to do it with induction, it just doesn't pop out at me currently)*

Example 6.29. *Find $gcd(62, 626)$*

$$626 = 10(62) + 6$$

$$62 = 10(6) + 2$$

$$6 = 3(2) + 0$$

From the E.A. I deduce $gcd(62, 626) = 2$. Moreover,

$$2 = 62 - 10(6) = 62 - 10[626 - 10(62)] = 101(62) - 10(626)$$

Example 6.30. *Find $gcd(240, 11)$.*

$$240 = 11(21) + 9$$

$$11 = 9(1) + 2$$

$$9 = 2(4) + 1$$

$$2 = 1(2)$$

Thus, by E.A., $gcd(240, 11) = 1$. Moreover,

$$1 = 9 - 2(4) = 9 - 4(11 - 9) = -4(11) + 5(9) = -4(11) + 5(240 - 11(21))$$

That is,

$$\boxed{1 = -109(11) + 5(240)}$$

So what? Well, there are problems that are exceedingly tedious to solve without the *arithmetic* covered in this section. For example, what is the multiplicative inverse of 11 in $\mathbb{Z}_{240}$? I can tell you without any further calculation it is -109. Which is better written as $11^{-1} = 132 \bmod 240$. Think about this, if you had hunted for 11^{-1} by trial and error it could well have taken you 131 tries before hitting the correct inverse. There are of course patterns and other tricks but this is one of the main reasons that knowing the E.A. is profitable for $\mathbb{Z}_n$ calculations.

6.5 when is it possible to find multiplicative inverses in $\mathbb{Z}_n$?

The following Theorem appeared as part of the Euclidean Algothim.

Theorem 6.31. *Given $a, b \in \mathbb{N}$ there exist $x, y \in \mathbb{Z}$ such that $ax + by = gcd(a, b)$.*

Notice that if we apply this to a, n then we find there exist $x, y \in \mathbb{Z}$ such that:

$$ax + ny = gcd(a, n).$$

We can restate the equation above in terms of the congruence modulo n relation on $\mathbb{Z}$:

Corollary 6.32. *(to the Euclidean Algorithm) Let $a, n \in \mathbb{N}$ then there exists $x \in \mathbb{N}$ such that*

$$\boxed{ax \equiv gcd(a, n)}$$

We will find this observation to be quite useful in what follows below.

Definition 6.33. *The multiplicative identity $1 \in \mathbb{Z}_n$ is $\bar{1}$. We say that b is the* **multiplicative inverse** *of a in $\mathbb{Z}_n$ iff $\bar{a}\bar{b} = \bar{1}$. In other words, $ab \equiv 1$. If b is the multiplicative inverse of a we may denote $b = a^{-1}$. We also say that a is* **invertible** *if a^{-1} exists.*

I would discourage you from using the notation $a^{-1} = \frac{1}{a}$ as it is not customary in $\mathbb{Z}_n$. Instead of writing $2(\frac{1}{2}) = 1$ we write $22^{-1} = 1$. Using the standard fraction notation may tempt you to make steps which are not valid, or a minimum not justified by the Theorems we have so far developed.

Proposition 6.34. *If $a, n \in \mathbb{N}$ then a, n are* **relatively prime** *iff $a^{-1} \in \mathbb{Z}_n$ exists.*

Proof: this proposition follows primarily from Corollary 6.32. Suppose that a, n are relatively prime then by definition $gcd(a, n) = 1$ hence by the Corollary 6.32 there exists $x \in \mathbb{N}$ such that $ax \equiv 1$. Therefore $a^{-1} \equiv x$.

Conversely suppose that a is invertible. Then there exists $a^{-1} \in \mathbb{Z}_n$ such that $a^{-1}a \equiv 1$. Suppose that j is a common factor of a and n then there exist $k, l \in \mathbb{N}$ such that $a = jk$ and $n = jl$. We wish to show $j = 1$. Observe that $a = jk$ implies $j = \frac{a}{k}$. Also, $n = jl$ implies $j = \frac{n}{l}$. Consequently, $\frac{a}{k} = \frac{n}{l}$ (*notice that k couldn't be zero for that would force a to be zero making the equation $aa^{-1} \equiv 1$ nonsense. Also, l cannot be zero since the equation $n = jl$ would force $n = 0$ and we assumed $n \in \mathbb{N}$ from the outset*). Thus, $al = na$ and multiplying by a^{-1} we find $l \equiv n$. If $l \equiv n$ then $l \in \bar{n}$ thus there exists $p \in \mathbb{N}$ such that $l = pn$. But then $n = jl$ becomes $n = jpn$ hence $jp = 1$. But since $j, p \in \mathbb{N}$ it follows $j = 1$ and $p = 1$. Therefore, the only common factor of a and n is one; a and n are relatively prime.

> **Caution:** I can't shake the feeling that the converse portion of the proof above is flawed. If you can point out the flaw and/or fix it then it would be worth some bonus points.

Proposition 6.35. *Let $a \in \mathbb{Z}_n$. Suppose $a \in \mathbb{Z}_n$ is invertible then a^{-1} is invertible with $(a^{-1})^{-1} = a$. Moreover, a is invertible iff a^{-1} is invertible.*

Proof: see the definition of invertible and think for a moment.

Example 6.36. *(finding mulitplicative inverse of 3 in $\mathbb{Z}_{10}$) Notice that $gcd(3, 10) = 1$ thus we know that there exists 3^{-1} in $\mathbb{Z}_{10}$. What is it? You can guess and find it, after all $\mathbb{Z}_{10} = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$ is not too big. We can make an educated guess since you can see that $x = 0, 2, 4, 5, 8$ are ruled out immediately since $gcd(x, 10) \neq 1$ and thus these are not invertible. We need only check those numbers which are relatively prime to 10. In $\mathbb{Z}_{10}$ we calculate:*

$$3 \cdot 1 = 3$$

$$3 \cdot 3 = 9$$

$$3 \cdot 5 = 15 = 5$$

$$3 \cdot 7 = 21 = 1$$

$$3 \cdot 9 = 27 = 7$$

Thus we find that $3^{-1} = 7$ in $\mathbb{Z}_{10}$.

It is sometimes not possible to reasonably implement the idea of the last example. We might need the Euclidean Algorithm to make the calculation tractable.

Example 6.37. *Does* 12^{-1} *exist in* $\mathbb{Z}_{101}$*? Let's employ the E.A. to find the* $gcd(12, 101)$*.*

$$101 = 12(8) + 5$$

$$12 = 5(2) + 2$$

$$5 = 2(2) + 1$$

Thus we deduce that $gcd(101, 12) = 1$ *hence* 101 *is relatively prime to* 12 *and we conclude* 12^{-1} *exists. Moroever, working backwards up the algorithm,*

$$1 = 5 - 2(2) = 5 - 2[12 - 5(2)] = -2(12) + 5(5) = -2(12) + 5[101 - 12(8)]$$

simplifying,

$$1 = -42(12) + 5(101).$$

Hence, $12^{-1} \equiv -42 \equiv 59$*.*

I hope you can see why the Euclidean Algorithm is needed for such questions.

Definition 6.38. *Let* S *be a set with* $0 \in S$*. If* $x \in S$ *with* $x \neq 0$*, then we say that* x *is* **zero divisor** *iff there exists* $y \in S$ *with* $y \neq 0$ *and* $xy = 0$*.*

It is well-known that $\mathbb{R}$ and $\mathbb{C}$ have no zero-divisors. In fact the utility of factoring polynomials to find their zeros hinges upon this critical fact. You should recall that for $a, b \in \mathbb{R}$ we *know* $ab = 0$ implies $a = 0$ or $b = 0$. We solve $2x^2 - x - 3 = (2x - 3)(x + 1) = 0$ over $\mathbb{R}$ by separately solving the equations $2x - 3 = 0$ and $x + 1 = 0$ to find $x = \frac{3}{2}$ and $x = -1$. Factoring still plays an important role in finding solutions for polynomial equations in $\mathbb{Z}_n$, however zero-divisors complicate the logic. (this topic is discussed in Math 422)

Example 6.39. *(no mulitplicative inverse of* 3 *available in* $\mathbb{Z}_6$*) Consider* $\mathbb{Z}_6$*. Notice that* $\overline{2}, \overline{3} \in \mathbb{Z}_6$*. Clearly* $\overline{2}, \overline{3} \neq 0$ *yet* $\overline{23} = \overline{2 \cdot 3} = \overline{6} = \overline{0}$*. We say that* $\mathbb{Z}_6$ *is a set with zero divisors.*

Example 6.40. *The zero divisors of* $\mathbb{Z}_{10}$ *are* $2, 4, 5, 6, 8$ *since in* $\mathbb{Z}_{10}$ *we have:*

$$2 \cdot 5 = 0, \qquad 4 \cdot 5 = 0, \qquad 6 \cdot 5 = 0, \qquad 8 \cdot 5 = 0$$

Key Idea: When calculating in $\mathbb{Z}_n$ we have $\bar{n} = \bar{0}$. This means that if we omit the bar-notation we have $n = 0$. Here the "=" is actually short-hand for congruence "$\equiv$". I always mean to write "the following calculations will be done in $\mathbb{Z}_n$ if I am going to use the shorter notation

6.6 solving equations in $\mathbb{Z}$ using $\mathbb{Z}_n$ tricks

Here we consider a rather impressive application of $\mathbb{Z}_n$ mathematics to the problem of finding **integer solutions** to systems of equations with *integer coefficients*.

Example 6.41. *Does the following system of equations have integer solutions?*

$$\begin{aligned} 3x + 5y - 7z &= 4 \\ x + y + z &= 3 \\ x + 13z &= 42,000,000 \end{aligned}$$

Consider this: if the equalities hold for $x, y, z \in \mathbb{Z}$ then the corresponding congruence relations must also hold in $\mathbb{Z}_n$. That is we ought to have

$$\begin{aligned} 3x + 5y - 7z &\equiv 4 \\ x + y + z &\equiv 3 \\ x + 13z &\equiv 42,000,000 \end{aligned}$$

In particular, consider $\mathbb{Z}_2$. (remember is the partition of $\mathbb{Z}$ into even and odd integers). Simplify the congruences for the case $n = 2$

$$\begin{aligned} x + y + z &\equiv 0 \\ x + y + z &\equiv 1 \\ x + z &\equiv 0 \end{aligned}$$

It is obvious that these equations have no solution since the first two equations are blatantly inconsistent.

There are many reasons to look for *integer solutions* since often in a real-world application we deal with things you can only take in whole-number increments. Like children, dogs, pennies etc... Our trick only shows that

finding integers solutions is sometimes impossible. To actually find the solutions (if there are any) you just solve and substitute. If you had solved

$$\begin{aligned} 3x + 5y - 7z &\equiv 4 \\ x + y + z &\equiv 3 \\ x + 13z &\equiv 42,000,000 \end{aligned}$$

directly you might have found

$$x = -\frac{503999857}{14}, \quad y = \frac{209999955}{7}, \quad z = \frac{83999989}{14}.$$

As you can see it is not an integer solution, rather this is a *rational solution.*

7 Algebra

Algebra is a general term which encompasses a very wide section of modern mathematics. There are hundereds if not thousands of subfields of study in algebra. Almost all of these are born from the desire to solve equations.

Polynomial equations are a central motivating example. Linear equations $ax+b=0$ are easy to solve. Quadratics $ax^2+bx+c=0$ are solved without exception by the quadratic formula (*which you should be able to derive at this point in your mathematical journey, just saying*). Cubics, well those are a little harder, however it is known that $ax^3+bx^2+cx+d=0$ has a solution which is given by a closed form algebraic formula. The same holds for quartics. You might be tempted to hope there is always some "n-order quadratic formula" which gives the general solutions for $p(x)=0$ where $deg(p)=n\in\mathbb{N}$. You would be wrong. It turns out that the 5-th order equation, called a **quintic equation**, does not have a general closed form solution.

Each step in the preceding paragraph has all sorts of algebra tied to it. Linear equations are the first example in study of linear algebra, although I should emphasize that linear algebra is motivated by a myriad of applications. The linear equation also leads you to consider negative numbers; what is the solution to $x+2=0$? If you were born several centuries ago you might have said there was no solution, the concept of negative numbers has only become mainstream relatively recently. Quadratic equations force you to consider numbers beyond fractions of integers, the equation $x^2-2=0$ has a solution which is not rational. Quadratic equations encourage you to think about complex numbers; what is the solution to $x^2+1=0$? The same is true for cubic equations whose solution was given by Cardano in 1545. In 1572 Bombelli pointed out that the cubic $x^3-15x-4=0$ has three real solutions yet the Cardona formula has complex numbers (which somehow cancel out to give a three real solutions). This goes to show that imaginary numbers are not really that "imaginary". In my humble opinion this is why it is high time we taught complex numbers early and often in mathematics, but don't get me started...

The quintic equation was shown to be insolvable by a formula like the quadratic formula by Abel and Ruffini in the early nineteenth century. (this is the Abel for which Abelian groups get there name). However, this is not to say that **all** quintics cannot be solved by a closed-form formula using

a finite series of algebraic operations. For example, $x^5 = 0$ is pretty easy to solve in terms of radicals; $\sqrt[5]{x^5} = \sqrt[5]{0^5}$ thus $x = 0$. There is an obvious question to ask: *When does a quintic have a nice algebraic solution?*. Galois answered this question by studying the *group* of symmetries for roots to a polynomial equations. This group is now known as the Galois group. Incidentally, the history of Galois is rather fascinating, he died at age 20 (in 1832) after getting shot in a duel. He pioneered much of basic group theory and is apparently the first to call a group a group (although, the precise definition of a *group* we consider in these notes wasn't settled until much later). Galois apparently feared his death was likely since he wrote a letter trying to preserve his work just before the duel. There are stories that he worked late into the night so that his mathematical works would not be lost with his death.

It turns out that the very construction of rational numbers, real numbers and complex numbers is provided by in large by techniques in abstract algebra. The natural numbers come from set theory. The integers are constructed by adjoining negatives and zero. Then the rational numbers are constructed as the *field of quotients* of the integers. Then, one can adjoin algebraic numbers using something called an *extension field*. The rational numbers adjoined with all possible algebraic numbers form the set of *algebraic numbers*. For example, $\sqrt{2}$ and $\sqrt{1+\frac{1}{1+\sqrt[5]{3}}}$ are algebraic numbers. It turns out that the cardnality of the algebraic numbers is $\aleph_o$. To obtain the *transcendental* numbers something beyond algebra has to enter the discussion. Analysis, in particular the careful study of sequences and their limits, allows us to adjoin the transcendental numbers. In a nutshell, that is how to construct the real numbers. The construction I just outlined will generate all the properties of real numbers which we have been taught to take for granted since our elementary school days. Algebra (and analysis) is used to construct our number system.

Algebra goes far beyond the history I just outlined. I'll throw in more comments as I develope the material. My goal in this Chapter is simply to equip you with the basic definitions and some familarity with basic examples. Hopefully Math 421 will seem less bizarre if we at see the basics here. Customarily a math major will have two serious courses in abstract algebra as an undergraduate then another pair of graduate abstract algebra courses as a graduate student. For a pure mathematician it is not unusual to take upwards of a dozen courses in algebra.

7.1 algebraic structures

There are many examples of binary operations:

Function Composition: $(f, g) \mapsto f \circ g$

Addition of Numbers: $(a, b) \mapsto a + b$

Multiplication of Numbers $(a, b) \mapsto ab$

Matrix Addition $(A, B) \mapsto A + B$

Matrix Multiplication $(A, B) \mapsto AB$

Cross Product on 3-vectors $(\vec{v}, \vec{w}) \mapsto \vec{v} \times \vec{w}$

A binary operation on a set S is a rule which takes **two** things from the set and outputs another. We can be precise about this:

Definition 7.1. *A* **binary operation** *on A is a function from $A \times A$ to A. We say that a binary relation on A is* **closed on A**. *Moreover, if we denote the operation by $*$ such that $(a, b) \mapsto a * b$ then we say that $(A, *)$ is a set A with an operation $*$.*

The notation $(A, *)$ is nice because it pairs the set with the operation. Clearly there can be many operations on a particular set. When we have one or more operations on a set A it is called an **algebraic system**.

Definition 7.2. *Given a set A and operations $*$ and $\circ$ we say that $(A, *, \circ)$ is a set with two operations $*$ and $\circ$. If a set A has has one or more operations and possibly several relations then we say A is an* **algebraic system**

Example 7.3. *Let $\mathbb{R}$ be the real numbers. We have two natural operations, addition $+$ and multiplication $\cdot$. We also have an order relation on $\mathbb{R}$; we say $(x, y) \in R_< \subset \mathbb{R} \times \mathbb{R}$ iff $x < y$. Collecting these together, $(\mathbb{R}, +, \cdot, <)$ is an algebraic system with operations $+$ and $\cdot$.*

Almost always given a particular set you can add further structure. For example, we could adjoin the relation $R_>$ (where $(x, y) \in R_> \subset \mathbb{R} \times \mathbb{R}$ iff $x > y$) to the algebraic system above; $(\mathbb{R}, +, \cdot, <, >)$. You should understand that when we give a description of a set in algebra we are trying to give a minimal description. The goal is to obtain the desired conclusion with a minimal set of assumptions. As a student this is a change of culture, in calculus we have been in the practice of admonishing you for forgetting

things from years and years ago. In algebra, and abstract math in general, we ask you to forget what you know and work with just what you are given and can prove. This is quite a change in thinking. Almost everyone struggles at first, so if first you don't succeed then don't be too discouraged. Remember, at least for this course the most critical thing is definitions. You have to know the definitions, you should *own* them.

Example 7.4. *Let $\mathbb{Z}$ be the integers. We have two natural operations, addition $+$ and multiplication $\cdot$. Thus $(\mathbb{Z}, +, \cdot)$ is an algebraic system.*

Example 7.5. *Let $\mathbb{Z}_n$ be the integers mod $n \in \mathbb{N}$. We have two natural operations, addition $+_n$ and multiplication $\cdot_n$. Thus $(\mathbb{Z}_n, +_n, \cdot_n)$ is an algebraic system.*

Given an operation on a set A we can sometimes induce an operation on a subset $B \subseteq A$. Let $(A, *)$ be a set with operation $*$ then we say that $B \subseteq A$ is **closed under** $*$ iff $*$ restricted to B is a binary operation.

Example 7.6. *Let $\mathbb{Z}$ be the integers. Notice $2\mathbb{Z} = \{2k \mid k \in \mathbb{Z}\}$ is a subset of $\mathbb{Z}$. We proved that the sum and product of even integers is even thus the two natural operations, addition $+$ and multiplication $\cdot$, are closed on $2\mathbb{Z}$. Therefore $(2\mathbb{Z}, +, \cdot)$ is an algebraic system.*

If you want to be really pedantic (as is sometimes my custom), the $+$ and $\cdot$ on $2\mathbb{Z}$ and those on $\mathbb{Z}$ are technically different functions since they have different domains. Hence, you could say that they are different operations. However, I find it more productive to say that the operations on $2\mathbb{Z}$ are *inherited* or *induced* from $\mathbb{Z}$.

Example 7.7. *Let $\mathbb{Z}$ be the integers. Notice $2\mathbb{Z} + 1 = \{2k + 1 \mid k \in \mathbb{Z}\}$ is a subset of $\mathbb{Z}$. We proved that the product of odd integers is odd thus the natural operation of multiplication $\cdot$ is closed on $2\mathbb{Z}+1$. Therefore $(2\mathbb{Z}+1, \cdot)$ is an algebraic system. Notice, in contrast, $+$ is* **not** *an operation on $2\mathbb{Z}+1$ since $1, 3 \in 2\mathbb{Z} + 1$ yet $1 + 3 = 4 \notin 2\mathbb{Z} + 1$. We see that $\cdot$ is not closed on the set of odd integers.*

The examples that follow go beyond the mainline of the text and homework, but I thought they might help bring context to the earlier discussion. I now give examples of things which are **not** binary operations. But, first a definition:

Definition 7.8. *Suppose A is a set and $n \in \mathbb{N}$. We define an n-ary operation on A to be a function $f : \underbrace{A \times A \times \cdots \times A}_{n \text{ copies}} \to A$. If $n = 1$ we say it is a*

unary *operation on A, if* $n = 2$ *it's a* **binary** *operation on A, if* $n = 3$ *its called a* **ternary** *operation.*

Example 7.9. *Let* $f : \mathbb{R} \to \mathbb{R}$ *be defined by* $f(x) = -x$*. This describes a unary operation on* $\mathbb{R}$ *called the* **additive inverse operation**. *Likewise,* $g(x) : \mathbb{R} - \{0\} \to \mathbb{R} - \{0\}$ *defined by* $g(x) = \frac{1}{x}$ *is the* **multiplicative inverse operation**.

Example 7.10. *The dot product on* $\mathbb{R}^n$ *is defined as follows: let* $n \in \mathbb{N}$ *and* $\vec{x} =< x_1, x_2, \ldots, x_n >$ *and* $\vec{y} =< y_1, y_2, \ldots, y_n >$ *then*

$$\vec{x} \cdot \vec{y} = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n$$

This is a function from $\mathbb{R}^n \times \mathbb{R}^n$ *to* $\mathbb{R}$ *thus the dot-product is* **not** *a binary operation on vectors.*

Example 7.11. *(concerning crossproducts in dimension* $n \neq 3$*) The crossproduct is sometimes said to only exist in* $\mathbb{R}^3$*. In other words, there is no crossproduct in* 2 *or* 4 *dimensions. I usually adopt this viewpoint. However, it does depend on what you mean by "crossproduct in other dimensions". Let me remind you that*

$$(\vec{A} \times \vec{B}) \cdot \vec{A} = 0 \quad and \quad (\vec{A} \times \vec{B}) \cdot \vec{B} = 0$$

essentially describe what the crossproduct does; the crossproduct picks out a vector which is perpendicular to both of the given vectors. If we characterize the crossproduct to simply be the operation which produces a new vector which is orthogonal to its input vectors then you can construct "crossproducts" in any dimension. However, the "crossproduct" on $\mathbb{R}^n$ *will be a* $(n-1)$*-ary operation on* $\mathbb{R}$*. Let me show you the "crossproduct" on* $\mathbb{R}^2$*,*

$$f(< a, b >) =< b, -a >$$

Notice $f(< a, b >) \cdot < a, b >= ab - ba = 0$ *thus* $f(< a, b >)$ *is orthogonal to its input* $< a, b >$*. For* $\mathbb{R}^4$ *you could define the "crossproduct" of* $\vec{x}, \vec{y}, \vec{z}$ *to be*

$$\vec{x} \times \vec{y} \times \vec{z} = det \begin{pmatrix} e_1 & e_2 & e_3 & e_4 \\ x_1 & x_2 & x_3 & x_4 \\ y_1 & y_2 & y_3 & y_4 \\ z_1 & z_2 & z_3 & z_4 \end{pmatrix}$$

where this is a mneumonic just as is the case with $n = 3$ *since the first row is made of vectors* $e_1 =< 1, 0, 0, 0 >$ *and* $e_2 =< 0, 1, 0, 0 >$ *etc.. so technically this is not a determinant. Personally, I'm partial to the formula,*

$$\omega_{\vec{x}} \wedge \omega_{\vec{y}} \wedge \omega_{\vec{z}} = Vol(\vec{x} \wedge \vec{y} \wedge \vec{z})$$

where $Vol_{\vec{x}\wedge\vec{y}\wedge z}$ is given by Hodge duality and $\omega_{\vec{x}}$ is the usual work-form mapping. For example, $Vol(e_1) = e_2 \wedge e_3 \wedge e_4$. In other words, the cross product of three 4-vectors corresponds to the triple wedge product of the vectors through Hodge duality. Hodge duality says that there is an isomorphism between forms of degree p and degree $n-p$. Only in $\mathbb{R}^3$ does it happen there are the same number of one-forms and two-forms. That is where we have the correspondence

$$\Phi(e_1) = e_2 \wedge e_3, \qquad \Phi(e_2) = e_3 \wedge e_1, \qquad \Phi(e_3) = e_1 \wedge e_2.$$

It is not too difficult to (see the extra credit project for calculus III) show that

$$\omega_{\vec{x}} \wedge \omega\vec{y} = \Phi_{\vec{x}\times\vec{y}}.$$

By the way if you'd like to know more about $\wedge$ just ask.

The Point? *only in $\mathbb{R}^3$ is the crossproduct a* **binary** *operation on vectors. So, if you insist that the crossproduct is a binary operation then you have a pretty good case that it only exists for $\mathbb{R}^3$.*

That said, the other "crossproducts" I have described are important to higher dimensional integration and so forth. We use the crossproduct to pick out the normal to a surface. The "crossproduct" in $n = 4$ could pick out the normal to a volume. In fact, the theory of integration in n-dimensions is most elegantly phrased in terms of the wedge product which gives differential forms a natural **exterior algebra**. *At the conclusion of calculus III I always show my students the Generalized Stokes Theorem which unifies all the various versions of the FTC into one overarching principle. I would probably be willing to offer you a course on differential forms if two or three (or more) were interested (we would need linear algebra as a prereq to make the course maximally worthwhile).*

7.2 algebraic properties

Algebraic systems often obey certain rules or properties.

Definition 7.12. *Let $(A, *)$ be an algebraic system. Then*

1. *commutative property* *We say $*$ is* **commutative** *iff for all $a, b \in A$, $a * b = b * a$.*

2. *associative property* *We say $*$ is* **associative** *iff for all $a, b, c \in A$, $(a * b) * c = a * (b * c)$.*

3. unital *We say that A has an* **identity** *$e \in A$ iff for all $a \in A$, $e * a = a$ and $a * e = a$. If A has an identity then it is said to be* **unital**.

4. invertibility *Let A be unital with identity element e. Let $a \in A$, we say a is* **invertible** *iff there exists $b \in A$ such that $a * b = e$ and $b * a = e$. In this case, we say b is the* **inverse of a with respect to $*$**. *If every element of A is invertible then we say that $*$ is* **closed under inverses**.

If a set A is closed under inverses then we could say that the inverse operation is unary on A. Usually $*$ is either addition, multiplication or function composition. In those cases there are standard notations and terminologies for the inverse operation:

1. multiplicative inverse of a: we denote by a^{-1}; we have $a * a^{-1} = 1$ and $a^{-1} * a = 1$.

2. additive inverse of b: we denote by $-b$; we have $b + (-b) = 0$ and $-b + b = 0$.

3. inverse function of f: we denote by f^{-1}; we have $f \circ f^{-1} = id_{range(f)}$ and $f^{-1} \circ f = id_{dom(f)}$.

Notice that e can be many things depending on context. In the list just above we saw that $e = 1$ for multiplication, $e = 0$ for addition and $e = id$ the identity function for function composition.

Example 7.13. *$(\mathbb{R}, +, \cdot)$ is commutative, associative and unital with respect to both $+$ and $\cdot$. Moreover, $\mathbb{R}$ is closed under additive inverses. In contrast, $\mathbb{R}$ is not closed under multiplicative inverses since the additive inverse 0 does not have a multiplicative inverse (cannot divide by zero).*

Example 7.14. *A nice example of a nonassociative operation is the crossproduct. Notice*

$$[(i + j) \times j] \times k = k \times k = 0$$

yet

$$(i + j) \times (j \times k) = (i + j) \times i = -k$$

Example 7.15. *Matrix multiplication is associative: let A, B, C be multipliable matrices then $(AB)C = A(BC)$. This follows directly from the definition of matrix multiplication and the associativity of the numbers which fill the matrices.*

Example 7.16. Lie Algebras *have an operation which is typically nonassociative. The operation on a Lie algebra is called the* **Lie bracket** *and it is denoted by* $[A, B]$*. The bracket has to satisfy the Jacobi Identity which is a sort of weakened associativity condition:*

$$[[A, B], C] + [[B, C], A] + [[C, A], B] = 0$$

In the case that the Lie algebra arises from an associative algebra the Lie bracket is just the **commutator bracket** *which is defined by* $[A, B] = AB - BA$ *(I'll give you a* **bonus point** *if you show this satisfies the Jacobi identity).*

Lie Algebras play an important role in classical and quantum mechanics because the generators of symmetry groups typically obey some sort of Lie algebraic structure. As I understand it, the original goal of Sophius Lie was to find solutions to differential equations. Perhaps you have seen that many differential equations reduce to a problem in algebra. If you study physics you'll learn that many of the hardest physics problems are solved through making a clever change of coordinates. The coordinate change always reflects a deeper symmetry of the problem and usually it makes the differential equations for the problem decouple so that elegant solutions are possible. Lie was trying to generalize this idea from classical mechanics to the solution of arbitrary differential equations, not necessarily stemming from physics. I don't think the complete goal has been realized even at this time, however work on Lie theory continues. I have a textbook which explains the idea of solving differential equations by symmetry methods for one independent variable, I'll show you if you're interested.

Definition 7.17. *A* **Cayley Table** *for an algebraic structure* $(A, *)$ *is a square table which lists all possible operations for the structure.*

Example 7.18. *We have already discussed many Cayley tables for* $\mathbb{Z}_n$*. For example, we wrote all possible additions and multiplications in* $\mathbb{Z}_3$ *in the last chapter. These are the Cayley tables for* $(\mathbb{Z}_3, +_3)$ *and* $(\mathbb{Z}_3, \cdot_3)$ *respectively:*

$+_3$	0	1	2
0	0	1	2
1	1	2	0
2	2	0	1

$\cdot_3$	0	1	2
0	0	0	0
1	0	1	2
2	0	2	1

Example 7.19. *Suppose that* $A = \{x, y\}$ *such that* $(A, *)$ *is an algebraic structure on* A *which is commutative and* $x * y = x$*. Moreover, suppose that*

$a * a = a$ for all $a \in A$. Find the Cayley table and identify the identity if possible. Use the given assumptions to deduce that

$$x * x = x, \qquad y * y = y, \qquad x * y = y * x = x.$$

We can write these results in a Cayley table:

$*$	x	y
x	x	x
y	x	y

Is x is the identity for $*$? Consider

$$x * x = x, \qquad \text{but } y * x = x \neq y.$$

Is y is the identity for $*$? Consider

$$x * y = x, \qquad \text{and } y * y = y$$

It follows $y * a = a * y = a$ for all $a \in A$ hence y is the identity element for A. Moreover, we can see x is not invertible while $y^{-1} = y$. In order for x to be invertible we need there to be some element x^{-1} such that $x * x^{-1} = y$ but there is no such element.

Theorem 7.20 (this is Theorem 6.1 in text)**.** *Let $(A, *)$ be an algebraic structure. Then A has at most one identity element. In addition, if A is associative and unital then each invertible element has a unique inverse.*

Proof: Suppose that $(A, *)$ has two identity elements $e_1, e_2 \in A$. Since e_1 is an identity element we have $a * e_1 = e_1 * a = a$ for all $a \in A$. Notice that $e_2 \in A$ thus $e_2 * e_1 = e_2$. Likewise, since e_2 is an identity element, $a * e_2 = e_2 * a = a$ for all $a \in A$. Hence, as $e_1 \in A$, we also have $e_2 * e_1 = e_1$. Therefore, $e_1 = e_2$. We can speak of **the** identity for an associative algebraic system.

Next, suppose that A is an associative, unital algebra with identity e. Use multiplicative notation and let $a \in A$ such that b_1 and b_2 are both inverses of a. That is, assume $ab_1 = b_1a = e$ and $ab_2 = b_2a = e$. We seek to show that $b_1 = b_2$. Multiply the first equation by b_2 on the left,

$$b_2ab_1 = b_2b_1a = b_2e = b_2$$

Multiply the second equation by b_1 on the right,

$$ab_2b_1 = b_2ab_1 = eb_1 = b_1$$

Comparing these equations we see $b_2ab_1 = b_2 = b_1$. Therefore, the inverse of $a \in A$ (when it exists) is unique. It is hence unambiguous to denote **the** inverse of a by $a^{-1} = b_1 = b_2$.

Question: where did we assume associativity in the proof above?

Remark 7.21. *(fruits of associativity) Let $(A, \cdot)$ be an associative algebraic system. Furthermore, let us use* **juxtaposition** *as a notation for the operation; $a \cdot b = ab$. In this notation, associativity simply means that certain parentheses can be dropped. For all $a, b, c \in A$,*

$$(ab)c = a(bc) = abc,$$

the parentheses can be dropped without ambiguity. In contrast, we cannot drop parentheses in $(\vec{A} \times \vec{B}) \times \vec{C}$ since this not equal to $\vec{A} \times (\vec{B} \times \vec{C})$. If we just write $\vec{A} \times \vec{B} \times \vec{C}$ then what is meant? Associative products allow for nonnegative **power notation**

$$aa = a^2, \qquad aaa = a^3, \qquad a^n = a^{n-1}a$$

Be careful though, generally

$$(ab)^2 \neq a^2b^2.$$

Instead, $(ab)^2 = abab$. If the algebraic stucture is also **commutative** *then we can rearrange those terms to get $(ab)^2 = aabb = a^2b^2$. Matrix multiplication is a popular example of an operation which is associative but* **not** *commutative.*

Theorem 7.22. *(algebraic system of bijections on a set A) Let $A \neq \emptyset$ and let $\mathcal{F}(A)$ be the set of all bijections on A. If $\circ$ denotes function composition then $(\mathcal{F}(A), \circ)$ is an associative, unital algebraic system which is closed under inverses. In particular the identity element of $\mathcal{F}(A)$ is the identity function I_A and the inverse of $f \in \mathcal{F}(A)$ is f^{-1}.*

Proof: See theorems in in Chapter 4. In short, the composite of bijections is again a bijection to the operation of function composition is closed on $\mathcal{F}(A)$. Moreover, we the inverse of a bijection exists and the identity function does satisfy the needed identities.

7.3 groups

Definition 7.23. *(group) We say $(G, *)$ is a* **group** *iff $(G, *)$ is an algebraic system which is associative, unital and closed under inverses. That is a set G is a group iff the following four items hold true:*

1. *if $g, h \in G$ then $*$ assigns one element $g * h \in G$,*
2. *if $a, b, c \in G$ then $(a * b) * c = a * (b * c)$,*
3. *there exists $e \in G$ such that for all $g \in G$, $g * e = e * g = g$.*
4. *for each $g \in G$ there exists $a^{-1} \in G$ such that $a * a^{-1} = a^{-1} * a = e$.*

Here (1.) is to insure that $*$ is a binary operation on G. Many of the examples we have so far discussed are in fact groups. Some are not. For example, $(\mathbb{Z}_n, +_n)$ forms a group with respect to $+$ however $(\mathbb{Z}_n, \cdot_n)$ does not form a group since 0^{-1} does not exist.

Definition 7.24. *(abelian group) An abelian group is a commuative group. That is, G is abelian iff $(G, *)$ is a group and for all $a, b \in G$, $a * b = b * a$.*

$(\mathbb{Z}_n, +_n)$, $(\mathbb{R}, +)$, $(\mathbb{C} - \{0\}, \cdot)$ are all abelian groups. In contrast, $\mathcal{F}(A)$ of Theorem 7.22 is **nonabelian**. In fact, the group $\mathcal{F}(A)$ is so special it gets a name:

Definition 7.25. *($\mathcal{F}(A)$; the group of permutations on A) Let A be a nonempty set. A* **permutation** *on A is a one-one and onto function on A. The set of all one-one and onto functions on A is a group called the* **permutation group on** *A which we denote $\mathcal{F}(A)$ in these notes (not a standard notation).*

When A is finite it turns out that $\mathcal{F}(A)$ is also finite. Counting will reveal that $\overline{\overline{\mathcal{F}(A)}} = (\overline{\overline{A}})!$. If A is infinite then there are more permutations than I can count.

Definition 7.26. *The* **symmetric group** *is the group of permuations on $\{1, 2, 3, \ldots, n\} = \mathbb{N}_n$. We define $\mathcal{F}(\mathbb{N}_n) = S_n$.*

I refer you to the posted homework solutions for examples on how the notation for S_n works in this text. I would like to talk about *cycle notation* if there is time (unlikely). Permutation groups have played an important role in the historical developement of group theory. They are also very useful in the theory of determinants. In short, you can use permutations to encode all those funny signs in the determinant formula.

Theorem 7.27. *(socks-shoes theorem, is Theorem 6.5b in text) Let $(G, *)$ be a group then $(a * b)^{-1} = b^{-1} * a^{-1}$*

Proof: Let $a, b \in G$ notice since G is closed under inverses we know $a^{-1}, b^{-1} \in G$. Observe that

$$(a * b) * (b^{-1} * a^{-1}) = a * b * b^{-1} * a^{-1} = a * e * b^{-1} = b * b^{-1} = e$$

and,

$$(b^{-1} * a^{-1}) * (ab) = b^{-1} * a^{-1} * a * b = b^{-1} * e * b = b^{-1} * b = e.$$

Therefore, $(ab)^{-1} = b^{-1} * a^{-1}$.

Theorem 7.28. *(inverse operation is an involution) Let $(G, *)$ be a group then $(a^{-1})^{-1} = a$*

Proof: Let $a \in G$ notice since G is closed under inverses we know $a^{-1} \in G$ and $aa^{-1} = e$ and $a^{-1}a = e$. But then by definition of inverse we see a is the inverse of a^{-1}. Moreover, since the inverse is unique by Theorem 7.20 we conclude $(a^{-1})^{-1} = a$.

Theorem 7.29. *(cancellation laws, this is Theorem 6.6) Let $x, y, z \in G$ and use juxtaposition to denote the operation then*

(1.) $xy = xz \implies y = z$

(2.) $yx = zx \implies y = z$

We call (1.) the **left cancellation law** *and (2.) the* **right cancellation law**.

Proof: Let $x, y, z \in G$ notice that by definition of group x^{-1} exists. Suppose $xy = xz$ then multiply by x^{-1} on the left to obtain $x^{-1}xy = x^{-1}xz$ which implies $ey = ez$ thus $y = z$. Likewise, multiply by x^{-1} on the right to obtain the right cancellation law.

Let me restate the cancellation laws and their proof in additive notation:

Theorem 7.30. *(cancellation laws, this is Theorem 6.6) Let $x, y, z \in (G, +)$ and use additive notation to denote the operation then*

(1.) $x + y = x + z \implies y = z$

(2.) $y + x = z + x \implies y = z$

We call (1.) the **left cancellation law** *and (2.) the* **right cancellation law**.

Proof: Let $x, y, z \in (G, +)$ notice that by definition of group $-x$ exists. Suppose $x+y = x+z$ then add by $-x$ on the left to obtain $-x+x+y = -x+x+z$ which implies $0 + y = 0 + z$ thus $y = z$. Likewise, add by $-x$ on the right to obtain the right cancellation law.

Remark 7.31. *(additive verses multiplicative group notation) We can define integer powers of $a \in (G, \cdot)$ by repeated multiplication of a or a^{-1}, using the juxtaposition notation: for each $n \in \mathbb{N}$*

$$a^n = \underbrace{aa \cdots a}_{n} \qquad a^{-n} = \underbrace{a^{-1}a^{-1} \cdots a^{-1}}_{n}$$

finally we **define** $a^0 = e$. *It is straightforward to check that $a^s a^t = a^{s+t}$ for all $s, t \in \mathbb{Z}$. In the case the group is commutative we recover the usual laws of exponents.*

If $(G, +)$ is an additive group then we can define **multiples** *of a in terms of repeated addition of a or $-a$: for each $n \in \mathbb{N}$*

$$n \cdot a = \underbrace{a + a + \cdots + a}_{n} \qquad -n \cdot a = \underbrace{-a + (-a) + \cdots + (-a)}_{n}$$

and $0 \cdot a = 0$.

The notation $n \cdot a$ is not necessarily multiplication in the group. Your group might not even have integers in it. It is simply a notation, a shorthand, to express repeated addition.

7.4 subgroups

A subset of a group which is also a group is called a **subgroup**.

Definition 7.32. *(subgroup) Let $(G, *)$ be a group. If $H \subseteq G$ such that $*$ restricted to H gives H the structure of a group then we say that H is a* **subgroup** *of G. Equivalently, we can say H is a subgroup of G if it is a group with respect to the group operation inherited from G. If H is a subgroup of G then we denote this by writing $H < G$.*

Example 7.33. *$G = \mathbb{Z}_6$ is an additive group. We can show $H = \{0, 3\}$ is a subgroup of G. Notice that $0 + 0 = 0$ and $3 + 3 = 0$ thus H contains the identity and each element has an inverse. Moreover, it is clear that addition in H is commutative and associative. Thus $H < G$.*

Remark 7.34. *The Theorem and proof that follows exposes a subtlety that is ignored by the text. The basic issue here is how do we explicitly connect an operation for a set to the operation restricted to a subset. I show how to do that in terms of mappings in the Proposition below. Skip to the "***Adequate Proof***" if this deluge of uncessary notation is too much for you.*

Definition 7.35. *(inclusion mapping) Let $A \subseteq B$ then the identity map on B restricted to A is also called the* **inclusion mapping** *and it is denoted $i : A \hookrightarrow B$ where $i(a) = a$ for each $a \in A$.*

Proposition 7.36. *(associative property and commutative property maintained in subsets) Suppose (B, f) is an associative algebraic system and $A \subseteq B$ then A is an associative algebraic system with respect to the operation $g = f \circ (i \times i)$. In addition, if (B, f) is a commutative algebraic structure and (A, g) is an algebraic structure then (A, g) is commutative.*

Careful Proof: Notice that $f : B \times B \to B$ is a function since (B, f) is an algebraic system. Furthermore, $i \times i : A \times A \to B \times B$ is the Cartesian product of functions and is hence a function (you proved this in § 4.2 # 18). We proved the composite of functions is again a function thus $g = f \circ (i \times i) : A \times A \to B$ is a function. (*if we could show range$(g) \subseteq A$ then that would show the operation restricted to the subset is closed, but that is not our current goal*). We wish to show that the operation g is associative, in our current notation that means for all $a, b, c \in A$

$$g(g(a, b), c) = g(a, g(b, c)).$$

Observe that

$$g(g(a, b), c) = g(f(a, b), c) = f(f(a, b), c)$$

and

$$g(a, g(b, c)) = g(a, f(b, c)) = f(a, f(b, c))$$

Since f is an associative operation on B is follows that $f(f(a, b), c) = f(a, f(b, c))$ therefore $g(g(a, b), c) = g(a, g(b, c))$.

If f is a commutative operation and g is an algebraic system on B then we have that $g : B \times B \to B$ is a function. We seek to show g is commutative. Let $a, b \in B$, notice

$$g(a, b) = f(a, b) = f(b, a) = g(b, a).$$

We see that the commutative property will naturally be induced to any subgroup of an abelian group.

Adequate Proof: If $x, y, z \in H$ then $x, y, z \in G$ thus for all $x, y, z \in H$, $(xy)z = x(yz)$. Likewise, if G is abelian then $x, y \in H$ implies $x, y \in G$ hence $xy = yx$.

Corollary 7.37. *A subgroup of abelian group is abelian.*

Example 7.38. *Let G be a group with identity e. Observe that $H = G$ and $H = \{e\}$ form subgroups of G. The subgroup $H = \{e\}$ is called the* **trivial subgroup**. *A subgroup $H \neq G$ is called a* **proper subgroup**.

I've skipped the two theorems which lead to this theorem. This is the important one for working problems.

Theorem 7.39. *(subgroup test, this is Theorem 6.8 in text) Let $(G, *)$ be a group. A subset $H \subseteq G$ is a subgroup of G iff $H \neq \emptyset$ and for all $a, b \in H$, $a * b^{-1} \in H$.*

Proof: Let $(G, *)$ be a group and $H \subseteq G$. Suppose $H < G$ then $e \in H$ since H is a group. Thus $H \neq \emptyset$. Suppose $a, b \in H$ then since H is a group it follows that $b^{-1} \in H$ and thus $a * b^{-1} \in H$ as $*$ is a binary operation on H.

Conversely, suppose $H \neq \emptyset$ and for all $a, b \in H$, $a * b^{-1} \in H$. Proposition 7.36 shows $(H, *)$ is associative. Let $a \in H$ then $a \in G$ thus $a * a^{-1} = e$. However, by the assumed property we also have $a * a^{-1} \in H$. Therefore, $e \in H$. Let $a \in H$, we have $e * a^{-1} \in H$ and thus $a^{-1} \in H$. Finally, we show that H is closed under $*$. Let $a, b \in H$ then $b^{-1} \in H$ and thus $a * b = a * (b^{-1})^{-1} \in H$ using Theorem 7.28. We see that H satisfies all four group axioms, therefore $H < G$.

Proposition 7.40. *(cyclic subgroup generated by a) Let G be a group with operation denoted by juxtaposition. Let $a \in G$ then*

$$< a >= \{a^n \mid n \in \mathbb{Z}\} = \{..., a^{-2}, a^{-1}, e, a, a^2, \dots\}$$

is an abelian subgroup of G.

Proof: Let G be a group and suppose $a \in G$. Observe that $a \in < a >$ thus $< a > \neq \emptyset$. Suppose that $x, y \in < a >$ then there exist $s, t \in \mathbb{Z}$ such that $x = a^s$ and $y = a^t$. Furthermore, $y^{-1} = a^{-t}$. Observe that $xy^{-1} = a^s a^{-t} = a^{s-t} \in < a >$ thus by the subgroup test we conclude that $< a >$ is a subgroup of G. Notice that $< a >$ is abelian since for all $a^s, a^t \in < a >$, $a^s a^t = a^{s+t} = a^{t+s} = a^t a^s$.

Definition 7.41. *Let G be a group and $a \in G$ then $< a >$ is the* **cyclic subgroup** *generated by a. If there exists $b \in G$ such that $< b >= G$ then we say that G is a* **cyclic group** *with* **generator** *b.*

Example 7.42. *$(\mathbb{Z}_4, +)$ is a cyclic group with generator 1. Additionally, $(\mathbb{Z}_4, +)$ is a cyclic group with generator 3. Notice that, modulo 4 we calculate*

$$\{1, 1+1, 1+1+1, 1+1+1+1\} = \{3, 3+3, 3+3+3, 3+3+3+3\} = \{0, 1, 2, 3\}$$

In contrast, $2 \in \mathbb{Z}_4$ generates the subgroup $< 2 >= \{2, 2+2\} = \{0, 2\}$.

Example 7.43. *$(\mathbb{Z}_5 - \{0\}, \cdot)$ is a cyclic group which can be generatr. Additionally, $(\mathbb{Z}_4, +)$ is a cyclic group with generator 3. Notice that, modulo 4 we calculate*

$$\{1, 1+1, 1+1+1, 1+1+1+1\} = \{3, 3+3, 3+3+3, 3+3+3+3\} = \{0, 1, 2, 3\}$$

In contrast, $2 \in \mathbb{Z}_4$ only generates the subgroup $< 2 >= \{2, 2+2\} = \{0, 2\}$.

Definition 7.44. *The* **order** *of a group G is the cardnality of G. The order of an element $a \in G$ is the order of $< a >$, in particular if $< a >$ is infinite then we say a has* **infinite order**

Example 7.45. *In Example 7.43 we saw that the order of 1 and 3 was 4 since $< 1 >=< 3 >= \mathbb{Z}_4$. On the other hand, the order of 2 was 2 since $< 2 >= 2\mathbb{Z}_4 = \{0, 2\}$ has just two elements.*

Notice the identity has order 1 in any group since $< 0 >= \{0\}$. In other words, the cyclic subgroup generated by the identity element is just the trivial subgroup.

Example 7.46. *Consider $H = \{1, 2, 3, 4\} \subset \mathbb{Z}_5$. You can check that H is a group with respect to multiplication modulo 5. Moreover, $< 1 >= \{1\}$, $< 2 >=< 3 >=< 4 >= H$. In particular,*

$$2^2 = 4, \quad 2^3 = 3 \quad 2^4 = 1.$$

We see notice that the order of 2 is 4 and we also have $2^4 = 1$. This is no accident

Theorem 7.47. *(this is Theorem 6.11 in the text) Let G be a group with identity element e. If $a \in G$ with order r then r is the smallest positive integer such that $a^r = e$.*

Proof: see text. In a nutshell, $< a >= \{e, a, a^2, \ldots a^{r-1}\}$ has r elements thus we must have $a^r = e$ otherwise $< a >$ would have $r + 1$ elements. The real proof requires a bit more thinking. I'll leave it for Math 421.

7.5 operation perserving maps

Definition 7.48. *Let $(A, *)$ and $(B, \circ)$ be algebraic systems. A mapping $\phi : A \to B$ is an* **operation preserving map** *iff for all $x, y \in A$, $\phi(x * y) = \phi(x) \circ \phi(y)$. If $(A, *)$ and $(B, \circ)$ are groups then an operation preserving map is called a* **homomorphism** *and $(B, \circ)$ is said to be a* **homomorphic image** *of $(A, *)$.*

Example 7.49. *Consider the Cayley table for $(\mathbb{Z}_2, \cdot)$ and compare it to the $A = \{x, y\}$ with operation $*$ from Example 7.19. It is pretty clear that 0 is like x and 1 is like y. This suggests that $\phi(0) = x$ and $\phi(1) = y$ will make $\phi : \mathbb{Z}_2 \to \{x, y\}$ an operation preserving map.*

$$\phi(0 \cdot 0) = \phi(0) = x = x * x = \phi(0) * \phi(0)$$

$$\phi(0 \cdot 1) = \phi(0) = x = x * y = \phi(0) * \phi(1)$$

$$\phi(1 \cdot 0) = \phi(0) = x = y * x = \phi(1) * \phi(0)$$

$$\phi(1 \cdot 1) = \phi(1) = y = y * y = \phi(1) * \phi(1)$$

This is not a homomorphism since $(\mathbb{Z}_2, \cdot)$ is not a group.

Example 7.50. *Notice that $((0, \infty), \cdot)$ is a multiplicative group and $(\mathbb{R}, +)$ is and additive group. The exponential function provides a homomorphism of these groups. In particular, $exp : \mathbb{R} \to (0, \infty)$ satisfies*

$$exp(x + y) = exp(x) \cdot exp(y)$$

The operation of addition in the domain of exp is preserved to become multiplication in the range. In fact, exp is an **isomorphism**, *see below:*

Definition 7.51. *If $(A, *)$ and $(B, \circ)$ are groups then an* **isomorphism** *is a homomophism which is a bijection. That is $\phi : A \to B$ is an isomorphism iff it is a bijection and for all $x, y \in A$, $\phi(x * y) = \phi(x) \circ \phi(y)$. If there is an isomorphism from $(A, *)$ to $(B, \circ)$ then the groups are said to be* **isomorphic**.

Example 7.52. *Show that $3\mathbb{Z}_6$ is a homomorphic image of $\mathbb{Z}_6$ as additive groups. Define $f : \mathbb{Z}_6 \to 3\mathbb{Z}_6$ by $f(z) = 3z$ for each $z \in \mathbb{Z}_6$. Let $x, y \in \mathbb{Z}_6$,*

$$f(x + y) = 3(x + y) = 3x + 3y = f(x) + f(y)$$

We can compare the Cayley tables of $3\mathbb{Z}_6 = \{0, 3\}$ and that of $\mathbb{Z}_6$ to see that $3\mathbb{Z}_6$

$+_6$	0	1	2	3	4	5
0	0	1	2	3	4	5
1	1	2	3	4	5	0
2	2	3	4	5	0	1
3	3	4	5	0	1	2
4	4	5	0	1	2	3
5	5	0	1	2	3	4

$+_6$	0	3
0	0	3
3	3	0

The way to understand this is that in $3\mathbb{Z}_6$ the elements $\{0, 2, 4\}$ in $\mathbb{Z}_6$ have been squished to 0 in $3\mathbb{Z}_6$. On the other hand $\{1, 3, 5\}$ in $\mathbb{Z}_6$ have been squished to 3 in $3\mathbb{Z}_6$. Perhaps I can illustrate it better in lecture with some colors. It helps to reorder the rows and columns so that the identified elements are next to each other.

Example 7.53. *Let $(\mathbb{R}^3, +)$ be 3-dimensional vectors with the usual vector addition. Likewise suppose $(\mathbb{R}^2, +)$ are 2-dimensional vectors with the usual addition. The projection $\pi : \mathbb{R}^3 \to \mathbb{R}^2$ defined by*

$$\pi(< a, b, c >) =< a, b >$$

is a homomorphism. Observe that for all $< a, b, c >, < x, y, z > \in \mathbb{R}^3$,

$$\begin{aligned} \pi(< a, b, c > + < x, y, z >) &= \pi(< a + x, b + y, c + z >) \\ &=< a + x, b + y > \\ &=< a, b > + < x, y > \\ &= \pi(< a, b, c >) + \pi(< x, y, z >). \end{aligned}$$

Thus 2-D vectors are a homomorphic image of 3-D vectors. When I think about a homomorphism I think of a sort of shadow of an object with more intricate structure. On the other hand, isomorphic sets are essentially the same thing perhaps cast in differing notation.

Remark 7.54. *If you look at the last example from the perspective of our discussion at the end of the functions chapter you could identify that the*

fibers of π are vertical vectors. Then the space of equivalence classes squishes those vertical vectors to a point. The resulting space of equivalence classes is 2 dimensional and the naturally induced bijection would be an isomophism from the equivalence classes to $\mathbb{R}^2$. Likewise, the sets I mentioned in the last example are the fibers of f and the space of fiber-equivalence classes is isomorphic to the range of f through the natural bijection we discussed at the end of the functions chapter. Theorem 6.12 starts to touch on this idea.

Theorem 7.55. *(fun facts about homomorphisms and isomorphisms) Let $(A, *)$ and $(B, \circ)$ be groups and $\phi : A \to B$ a homomorphism,*

1. *$\phi(e_A) = e_B$ where e_G is the identity in $G = A, B$.*
2. *$\phi(a^{-1}) = (\phi(a))^{-1}$ for each $a \in A$.*
3. *if ϕ is an isomorphism then ϕ^{-1} is an isomorphism.*
4. *The image $\phi(A)$ is a subgroup of B.*

Example 7.56. *(loop groups) Given a space S you can form a group from the closed paths (loops) in the space. The group operation is constructed from simply pasting the paths together. The structure of the loop group will reveal things the topology of the space S. This means that abstract algebra can be used to reveal things about topology. This general type of thinking forms a branch of mathematics called* **algebraic topology**.

Example 7.57. *(Lie groups) A manifold with a smooth group structure is called a* **Lie group**. *The Lie algebra is in one-one correspondence with the tangent space at the identity of the Lie group. For a silly example, $exp : \mathbb{R} \to (0, \infty)$ connects the Lie group $(0, \infty)$ and the Lie algebra $\mathbb{R}$. Lie groups formed from matrices have wide-spread application in theoretical and experimental physics. Rotation groups are Lie groups. The Lorentz group from special relativity contains rotations and velocity boosts on spacetime. I have hardly even touched on the geometrical motivations for group theory in these notes. Look up Klein's Erlanger Programm to see more about this.*

Example 7.58. *(representation theory) One interesting goal of representation theory is to find homomorphic images of certain abstract groups (or algebras) onto matrix groups(or algebras). This is interesting because the homorphic image will naturally act on a vector space. In physics, especially quantum mechanics, the vector space contains the physical states. The symmetry group acts on the states. It "rotates" the states amongst themselves. When a symmetry commutes with the Hamiltonian then the states that rotate*

amongst themselves form a set of equal energy states. Most of those funny magic rules in Chemistry actually can be **derived** *from a careful study of group theory and its application to quantum atomic physics. In mathematics and physics the term* **representation theory** *goes far beyond what I have sketched here. (by the way a "group" in physics is sometimes not a group as mathematicians think of them)*

Example 7.59. *(local symmetry groups) If you want a theory in which a symmetry acts at one point in space at a time then a local symmetry group is what you want. In contrast to a rotation group, the local symmetry group acts locally. Special relativity motivates the desire to use local symmetries. It turns out that a local symmetry group is not just a group. In fact, it is better described by a principle fiber bundle where the symmetry group forms the fiber. This may sound esoteric to you, but you should know that electricity and magnetism is the quintessensial example of a physical model motivated from a local symmetry. Such theories are called* **gauge theories**. *Everytime you turn on a lightswitch or watch TV you reap the benefits of understanding gauge theory. Gauge theory has been a major player in theoretical physics since the late 1950's, although it was pioneered by Weyl in 1929 for Electromagnetism. Einstein discouraged Weyl's original attempt in 1919 because his original theory infringed on the standard view of space and time. Einstein's general theory of relativity can also be phrased as a gauge theory. These are the basics for a theoretical physicist. I think it is amazing how far God has allowed human thought to progress these past few centuries.*

7.6 rings, integral domains and fields

Definition 7.60. *A* **ring** $(R, +, \cdot)$ *is a set* R *togther with two binary operations called addition* $+$ *and multiplication* $\cdot$ *that satisfy the following axioms:*

1. $(R, +)$ *is an abelian group.*
2. $(R, \cdot)$ *is an associative algebraic system*
3. *The multiplication is left and right distributive over additon. That is for all* $a, b, c \in R$ *we have*

$$a \cdot (b + c) = a \cdot b + a \cdot c, \qquad (a + b) \cdot c = a \cdot c + b \cdot c$$

If R *has a multiplicative identity then we say that* R *is a ring with unity. If* R *has a commutative multiplication then we say that* R *is a commutative ring.*

Example 7.61. $\mathbb{C}$, $\mathbb{R}$, $\mathbb{Q}$ *and* $\mathbb{Z}$ *are all rings with respect to the natural addition and multiplication. Also* $(\mathbb{Z}_n, +_n, \cdot_n)$ *is a ring.*

Definition 7.62. *An* **integral domain** *is a commutative ring with unity such that* $(R, +, \cdot)$ *has no* **zero divisors***; that is if* $a \cdot b = 0$ *then either* $a = 0$ *or* $b = 0$.

An integral domain is a set where factoring works.

Example 7.63. $\mathbb{C}$, $\mathbb{R}$, $\mathbb{Q}$ *and* $\mathbb{Z}$ *are all integral domains with respect to the natural addition and multiplication. Also* $(\mathbb{Z}_p, +_p, \cdot_p)$ *for* p *prime is an integral domain.*

Example 7.64. *Notice that* $\mathbb{Z}_4$ *is not an integral domain since* $2 \cdot 2 = 0$ *yet* $2 \neq 0$. *If* n *is not prime you can find zero divisors for* $\mathbb{Z}_n$ *from the factors of* n. *For example, in* $\mathbb{Z}_6$ *we saw that* 2 *and* 3 *are zero divisors.*

Definition 7.65. *A* **field** F *is a integral domain such that the multiplication restricted to* $F - \{0\}$*is closed under inverses. In other words,* F *is an commuative ring with unity such that* $(F - \{0\}, \cdot)$ *is an abelian group.*

Example 7.66. $\mathbb{C}$, $\mathbb{R}$ *and* $\mathbb{Q}$ *are all fields with respect to the natural addition and multiplication. Also* $(\mathbb{Z}_p, +_p, \cdot_p)$ *for* p *prime is a field. Note that* $\mathbb{Z}$ *is not a field since* $2^{-1} \notin \mathbb{Z}$.