ABRAHIM LADHA

DISCRETE MATHEMATICS (DRAFT)

Contents

List of Figures

What is Discrete Math?

TBD and all that.

Introduction to Logic

Why Logic?

Logic is a formalization of pure thought. Before we go to logic, lets look at arithmetic. Arithmetic is the basic calculator math we all do. In the before times, we[1] did arithmetic on our fingers. Or we did it by making hatch marks, or maybe using a abacus. Eventually, we found a much more efficient way to do arithmetic, symbolically. We made up names for the numbers, and a syntax and list of rules you could apply to these equations. For example:

$$\begin{aligned} & 5(3+2) \\ 5 \cdot 3 + 5 \cdot 2 & \\ 15 + 10 & \\ 25 & \end{aligned}$$

First note, that quantity is intangible, nonphysical. "3" is nothing. You cannot have "3". You can have three of something, three hats, three apples, three fingers, but you cannot have a "3". We use the curly symbol "3" to reprent an ideal concept of a quantity. Using symbols for the numbers, we made up symbols for the operations, addition, multiplication.

Then we made up rules on how to manipulate these symbols. From there, arithmetic is performed by moving the symbols around, instead of counting one by one. This is really efficient and we have stuck with this system since. From there, we could develop so much more. The quadratic formula, trigonometry, physics, calculus and so on. We went to the moon.

Performing arithmetic symbolically was a huge revolution. What else could be advanced by symbolic representation? What other mental procedures could be formalized? What about thought itself? The end of this result is logic; **Logic is the formalization of thought**. Leibniz was one of the first to think in this direction. He believed that thought was compounded from some kind of alphabet of ideas, and new ideas were produced in a process similar to multiplication of numbers.[2]

Your brain has ideas, but they are trapped in your head. In order to express those ideas, you must use language. Unfortunately, since language is a natural product, it is imperfect. Since the meaning of words cannot be fixed, it is not possible to exclude all possible misunderstandings. Consider the sentence "The dog ate the cat who scratched the fence. " We may say this sentence is ambiguous. There are many kinds of ambiguity.

Rather than expressing assertions, inferences, and deductions in natural language, we may express them in our formal language in order to limit ambiguity and fallacy. Mathematicians have created a "formal language", in that the meaning of these symbols cannot be misinterpretted. Deduction and reasoning

[1] as in, humanity

Figure 1: Calculating-Table by Gregor Reisch: Margarita Philosophica, 1503.

[2] "*It is obvious that if we could find characters or signs suited for expressing all our thoughts as clearly and as exactly as arithmetic expresses numbers or geometry expresses lines, we could do in all matters insofar as they are subject to reasoning all that we can do in arithmetic and geometry. For all investigations which depend on reasoning would be carried out by transposing these characters and by a species of calculus.*" - Leibniz

can be performed in this formal language as one might perform arithmetic. The following is an example

$$\forall x \forall y (\forall z (z \in x \iff z \in y) \implies x = y)$$

This sequence of symbols to you now is unreadable and meaningless, but it has one advantage over natural language. It is precise and unambiguous. It is impossible to misinterpret it, like it is impossible to misinterpret a well formed mathematical formula. You should treat this subject like learning a new language.

We will not extensively discuss why logic appears to emulate our own laws of thought, but rather the use of logic as a tool. As we define parts of this formal language, for each step, convince yourself like "yea, thats how I think".

Figure 2: One kind of the ambiguity of language.

Truth

What is truth? Hard to say, and hard to define. Without further inquiry, we may be confident in the following two assertions:

- Truth exists
- Not everything is true.

Given a collection of truths, you may derive new truths. But where do those truths come from? To prevent a chicken-egg problem, we make use of *axioms*. An axiom is a true statement which needs no proof. It is given to you by god. No more questions. It is the basic truths, usually too simple to demonstrate. For example with numbers, it is true that for any number n that $n = n$ is true. This is an axiom. The rules of deduction are themselves axioms.

Note that intuitively, propositions may either be true or false. If a proposition is not true, it must be false. If a proposition is not false, it must be true. This is called *law of excluded middle*. We take the law of excluded middle to be true because it is how we interact and use truth. For any proposition, it must be true, or it must not be true. Everything in this world either is, or isn't. There is no intermediate between contradictories. [3]

[3] "*It will not be possible to be and not to be the same thing, except in virtue of an ambiguity, just as if one whom we call "man", and others were to call "not-man"; but the point in question is not this, whether the same thing can at the same time be and not be a man in name, but whether it can be in fact.*" - Aristotle, in *Metaphysics*

Proposition

A proposition is a sentence which may be understood to have an objective truth value, one which can be demonstrated and asserted. The following are examples of propositions.

- Socrates is a man.
- The sun will rise tomorrow
- The sum of two numbers is a number.
- $10 > 7$
- $1 + 1 = 3$

Even though some of these sentences are false, they are propositions because they may be understood to have an objective truth value. The following sentences are not propositions:

- What time is it?
- Subscribe to me on youtube.
- Team X is the best.
- $n > 7$

Examples of sentences which are not propositions include questions, commands, and sentences with subjective truth value. For a truth to be objective, intuitively, you must understand that it is demonstrable. The fourth one cannot be assigned an objective truth value, as it is true or false for different values of n. This is a predicate, and not a proposition. It can be made into a proposition and assigned a truth value by evaluation of n by a number.

Rather than pure english, we with to use a calculus of symbols so that the meaning of well formed formulas may be understood more objectively than natural language. To that extent, we use *propositional variables*.[4] A propositional variable is a letter, such as p, q, r, s which represents a proposition. For example:

$$p := \text{“Socrates is a man”}$$

Now whenever p is referenced, it is in context of the proposition which asserts that Socrates is a man.

[4] You may recall a moment in your earliest math classes when they added letters to arithmetic and never went back. This is an analogous moment.

Negation ($\neg$)

What is the negation of a proposition? The negation of a proposition p is written as $\neg p$. It is understood to be the logical “opposite” of proposition p. If p is truth, then the opposite of truth is certainly and can only be falsehood. So if p is true, then $\neg p$ is false, and if p is false, then $\neg p$ is true. In fact, we define false to mean “not true”. What is the opposite of white? If you said “black”, thats incorrect. The opposite of “white” is “not white”. For p the proposition that asserts “Socrates is a man”. We may write $\neg p$ either as:

- It is not the case that Socrates is a man
- “Socrates is a man” is false.
- “Socrates is a man. Not.[5]

[5] https://www.youtube.com/watch?v=fhIdbRp6xeg

Although these are correct representations, they are cumbersome in english. The proposition $\neg p$ may be equivalently rewritten as “Socrates is not a man”. Note again that since p was true, $\neg p$ must be false.

What is the proposition which occurs when we negate a negated statement? Consider $\neg\neg p$. If p is true, then $\neg p$ is false, so then $\neg\neg p$ must be true. If p is true, then it is not the case that it is not the case that p is true, so we see that $\neg\neg p$ has identically the same truth value as p. The logical negations cancel each other out in a way similar to arithmetic: $(-1) \cdot (-1) \cdot x = x$

Combining Propositions

You must observe that ideas may be formed from smaller, more atomic ideas. There are few words in english, reserved for making such combinations.

Conjunction ($\wedge$)

Consider the proposition "Socrates is a man and all men are mortal". We use the english word "and" here to represent a combination of two smaller propositions. Let p be the proposition to assert that "Socrates is a man", and q be the proposition to assert that "all men are mortal". We may write the original proposition as the *conjunction* $p \wedge q$. How does the truth value of this conjunction depend upon its pieces? If you conjunct truths together, you can only get truth. If you conjunct several truths together, but then any falsehood, then the conjuncted proposition in whole must be false. For example, "p is true and p is false" is false. The proposition p must be either true or false, but it cannot be both. As another example, consider "Socrates is a man and I am Socrates". This proposition is false. Although Socrates is a man, I am not Socrates. As a whole, the statement is false, even if it contains some truth.

The word "and" is one english word we represent with conjunction, but others which are equivalent may include:

- p and q
- p but q
- p plus q
- p in addition to q

Natural language may distinguish between "and" and "but", but notice that logically, there is no difference with respect to the truth value of the established proposition. The word "but" may provide some connotation, a subliminal meaning, but the truthfulness or nontruthfulness of the sentence is unperturbed.

Disjunction ($\vee$)

Ideas may not need to be composed in such a way which requires all pieces, to be true, but perhaps just some, or any of the pieces. Consider the proposition "Socrates is a man or I am Socrates". Even though I am not Socrates, the fact that Socrates is a man makes this proposition true on the whole. We may write the disjunction of propositions $p \vee q$ to represent a logical "or". A disjunction is true if any of its propositions are true. For example, the disjunction "p is true or p is false" is a true proposition, as p must definitely be either true or false. In a disjunction, if both are true, then the proposition is true.

Some other logically equivalent words to "or" include

- p otherwise q
- p rather q

Exclusive Or ($\oplus$)

Note the english use of the word "or" is not understood identically with the logical definition of "or". Sometimes, the use of "or" is exclusionary, in that "p or q" means "either p or q but not both". A logician goes to olive garden and the waiter asks, "would you like soup or salad with that?" The logician replies "yes". and is unfortunately charged extra for both.[6] This kind of "or" we denote as an exclusionary or, xor. We use the symbol $p \oplus q$ to mean xor.

[6] true story

In language, it is often ambiguous if an "or" is meant inclusionary or exclusionary. In our formalization, we remove this ambiguity by having two different distinct symbols.

Truth Tables

A truth table is a list where the first columns represent propositions and all possible truth values they may take on, and the later columns represent combinations of those propositions, and their respective truth values as well. Of the logical primitives we have seen so far, convince yourself that the following truth table is correct.

p	q	$p \vee q$	$p \wedge q$	$\neg p$	$\neg\neg p$	$p \oplus q$
T	T	T	T	F	T	F
T	F	T	F	F	T	T
F	T	T	F	T	F	T
F	F	F	F	T	F	F

if there are n distinct propositions, then a truth table will have 2^n rows. When filling out a truth table, please write all rows and columns. The leftmost columns are reserved for the propositional variables.

Logical Consequence

Some propositions entail each other, in a causal manner. Consider the proposition "If Socrates is a man, then all men are mortal". We may represent this as "If p then q", and use the notation $p \implies q$. This is called an implication and is read as "p implies q". We refer to p as the premise, or hypothesis. Then q is the conclusion, or consequence.

What is the truth value of an implication? Certainly if p is true, and q is false, then the implication $p \implies q$ should be false. But what about the other cases?

Consider the implication "If you study, you will pass". To evaluate its truth, suppose someone tells you this, and we will determine if they were lying or not.

- If p is true, and q is true, then $p \implies q$ should be true
- If p is true, and q is false, then $p \implies q$ should be false. If you study and fail, then the implication was not true.
- If p is false, and q is true, then $p \implies q$ is true. If you don't study and you pass, whoever told you this was not lying, so it must have been true.
- If p is false, and q is false, then $p \implies q$ is true. If you don't study and you fail, the person who told you this was not lying, so they were telling the truth.

We may represent this as the following truth table: [7]

p	q	$p \vee q$	$p \wedge q$	$\neg p$	$\neg\neg p$	$p \oplus q$	$p \implies q$
T	T	T	T	F	T	F	T
T	F	T	F	F	T	T	F
F	T	T	F	T	F	T	T
F	F	F	F	T	F	F	T

There are many english equivalents:

[7] Pay special attention to the truth of $p \implies q$ in the case that p is false. It is perhaps unnatural. There is no notion of "undefined" or "untested", these are temporal. It is as plain as if it was not a lie, then it must be a truth, by excluded middle. Everything in this world either is, or isn't.

- p implies q
- If p then q
- A necessary condition for p is q
- A sufficient condition for p is q
- q given p
- q when p
- If p, q
- p only if q
- q follows from p

Related Implications

There are three related implications to $p \implies q$

- Given implication $p \implies q$, its converse is defined as $q \implies p$.
- The inverse of the implication $p \implies q$ is defined as $(\neg p) \implies (\neg q)$
- The contrapositive of the implication $p \implies q$ is defined to be $(\neg q) \implies (\neg p)$.

Lets add these to our truth table. Rather than compute these by understanding sentences, we may simply compute them using the previous entries in the truth table.

p	q	$p \vee q$	$p \wedge q$	$\neg p$	$\neg\neg p$	$p \oplus q$	$p \implies q$	$q \implies p$	$(\neg p) \implies (\neg q)$	$(\neg q) \implies (\neg p)$
T	T	T	T	F	T	F	T	T	T	T
T	F	T	F	F	T	T	F	T	T	F
F	T	T	F	T	F	T	T	F	F	T
F	F	F	F	T	F	F	T	T	T	T

Notice the $p \implies q$ and its contrapositive have identical columns in the truth table. That means that these two logical statements are equivalent. This is true, even in english:

- If it is raining, then the bus is late.
- If the bus is not late, then it is not raining.

From here on, if we say two propositions are equivalent, we mean their truth tables have identical columns. We will detail equivalence of propositions later.

Biconditionals

If two propositions are equivalent, we may write them as $p \iff q$ and say p if and only if q. This is called a biconditional, or a characterization. This is similar to the definition of equality in arithmetic. For example, "It rains if and only if it pours". We may write the equivalence of an implication with its contrapositive as

$$(p \implies q) \iff (\neg q \implies \neg p)$$

Also note that since $p \implies q$ and $q \implies p$ are necessary and sufficient, we can represent $p \iff q$ as

$$(p \implies q) \wedge (q \implies p)$$

In english, its equivalents are:

- p is necessary and sufficient for q
- p implies q and q implies p
- p iff[8] q
- p exactly and only when q

[8] This is a mathematical short hand to mean "if and only if".

Note in english, it is not common for people to say "if and only if" when they mean it. The use of language in this case is imprecise.

Precedence

In order for our formal language of propositional logic to remove ambiguity, we must define an ordering of the operators. Is $\neg p \implies q$ to mean $(\neg p) \implies q$ or $\neg(p \implies q)$? Much like PEMDAS for arithmetic, we have an equivalent for propositional calculus. In order of priority first:

- parenthesis
- $\neg$
- $\wedge$
- $\vee$
- $\implies$
- $\iff$

Propositions you write should be sufficiently paranthesized so that one reading it may not have to delegate to the rules of precedence. Avoid use of other symbols, such as xor ($\oplus$) and "implied by" ($\impliedby$). Most formally, a proposition is correct sequence of symbols only over the symbols $(,), \neg, \wedge, \vee, \implies, \iff$ and propositional variables. Other symbols can be defined in terms of these.

Propositional Logic

Logic Puzzles

If propositional logic is supposed to be a formalization of thought, then we should be able to apply it to solve some problems. There are often small gotcha's, or confusing paradoxes. But many of these are not true paradoxes. Although they may seem contradictory, this vanishes when we can express the problem in the clear language that is the propositional calculus.

Figure 3: The dictionary definition of a knave is "An unprincipled, crafty fellow."

Consider the Knights and Knaves problem. You are in some sort of monty python skit, in which everyone is either a knight or a knave, not niether and not both. Knights always tell the truth, and knaves always lie. You come across two travelers, we may denote as A, B.

- Person A says "B is a knight".
- Person B says "The two of us are opposite types"

You do not know if they telling the truth or not. But you not know nothing. You know something conditionally: *If they are telling the truth*, then what they say is true. Moreso, if they are lying, then you know the negation of what they are saying is true. What are the types of A and B? If A is telling the truth, then B is a knight, so B is telling the truth, so then A must be a knave, and must be lying? On the surface it seems paradoxical, muddied. Lets try again by applying the propositional calculus.

Let p, q denote the propositions that "A is a knight" and "B is a knight" respectively. Then $\neg p, \neg q$ denote the propositions that "A is a knave" and "B is a knave" respectively. If you are not a knight, then you must be a knave.

We may represent the statements given then as

- Person A: $(p \implies q) \wedge (\neg p \implies \neg q)$
- Person B: $q \iff [(p \wedge \neg q) \vee (\neg p \wedge q)]$

If A is a knight, then p is true, so q must also be true, so p, q are both true. But q asserts that p, q must be different, so A cannot be a knight. If p is false, and B is a knave, then we know that B must also be a knave. Since B asserts that A, B must be different types, if they are lying then A, B must be the same type. So they actually were both knaves.

We may determine the answer even more mechanically with a truth table. Observe that the information A provides can be simplified to $p \iff q$.

p	q	$\neg p$	$\neg q$	$p \iff q$	$p \wedge \neg q$	$\neg p \wedge q$	$(p \wedge \neg q) \vee (\neg p \wedge q)$	$q \iff (p \wedge \neg q) \vee (\neg p \wedge q)$	
T	T	F	F	T	F	F	F	F	
T	F	F	T	F	T	F	T	F	
F	T	T	F	F	F	T	T	T	
F	F	T	T	T	F	F	F	T	

Lets look at the columns that are important, and take the conjunction of the statements given by both parties

p	q	A: $p \iff q$	B: $q \iff (p \wedge \neg q) \vee (\neg p \wedge q)$	A and B: $[p \iff q] \wedge [q \iff (p \wedge \neg q) \vee (\neg p \wedge q)]$
T	T	T	F	F
T	F	F	F	F
F	T	F	T	F
F	F	T	T	**T**

There is only one valid solution, both p, q are false, so they must both be knaves. Note that if there if the last column has no solution, then there is no solution to the puzzle. If the last column has more than one truth value, then there may be more than one solution to the problem.

Equivalence

We may use the symbol $\equiv$ to denote that two propositions are equivalent, in a truth table sense. For example, we will show that

$$p \implies q \equiv \neg p \vee q$$

p	q	$\neg p$	$\neg p \vee q$	$p \implies q$
T	T	F	**T**	**T**
T	F	F	**F**	**F**
F	T	T	**T**	**T**
F	F	T	**T**	**T**

Notice that the two last bolded columns are the same. Their respective propositions are said to be equivalent. [9]

[9] There is a relationship between $\equiv$ and $\iff$. Let Φ_1, Φ_2 be any two propositions. Then $\Phi_1 \equiv \Phi_2$ if and only if $\Phi_1 \iff \Phi_2$ is always true. The difference between them is that $\equiv$ is a relation among two different propositions, while $\iff$ is a symbol which can be used in a proposition. $\Phi_1 \iff \Phi_2$ is one proposition.

Laws of Thought

Truth tables are useful, but actually quite annoying. If a proposition has n propositional variables, it will have a truth table of 2^n rows, and who knows how many columns. Why don't we define a set of laws by which you can manipulate, simplify, expand a proposition? If two propositions are equivalent, then you may replace one for the other in some larger proposition syntactically. We can modify the proposition into a smaller or simpler one this way, and determine its truth with a smaller truth table. We will demonstrate the equivalence of all our rules using truth table, but afterwards, you may take them like you take the laws of arithmetic. You use $a + b = b + a$ for example without prejudice. Some of these laws we can show by a truth table equivalence, others are so simple, we can only take them as laws. Don't take the following truth tables too much in depth. They simply are there to demonstrate the laws are good ones.

Identity

- $p \wedge T \equiv p$
- $p \vee F \equiv p$

Domination

- $p \vee T \equiv T$

- $p \wedge F \equiv F$

Idempotent

- $p \wedge p \equiv p$
- $p \vee p \equiv p$

Double Negation

- $\neg\neg p \equiv p$

p	$\neg p$	$\neg\neg p$
T	F	T
F	T	F

Communitivity

- $p \wedge q \equiv q \wedge p$
- $p \vee q \equiv q \vee p$

Associativity

- $(p \wedge q) \wedge r \equiv p \wedge (q \wedge r)$
- $(p \vee q) \vee r \equiv p \vee (q \vee r)$

p	q	r	$(p \wedge q)$	$(q \wedge r)$	$(p \wedge q) \wedge r$	$p \wedge (q \wedge r)$	$(p \vee q)$	$(q \vee r)$	$(p \vee q) \vee r$	$p \vee (q \vee r)$	
T	T	T	T	T	**T**	**T**	T	T	**T**	**T**	
T	T	F	T	F	**F**	**F**	T	T	**T**	**T**	
T	F	T	F	F	**F**	**F**	T	T	**T**	**T**	
T	F	F	F	F	**F**	**F**	T	F	**T**	**T**	
F	T	T	F	T	**F**	**F**	T	T	**T**	**T**	
F	T	F	F	F	**F**	**F**	T	T	**T**	**T**	
F	F	T	F	F	**F**	**F**	F	T	**T**	**T**	
F	F	F	F	F	**F**	**F**	F	F	**F**	**F**	

Distributive Laws

- $p \wedge (q \vee r) \equiv (p \wedge q) \vee (p \wedge r)$
- $p \vee (q \wedge r) \equiv (p \vee q) \wedge (p \vee r)$

p	q	r	$(p \wedge q)$	$(p \wedge r)$	$(p \wedge q) \vee (p \wedge r)$	$(q \vee r)$	$p \wedge (q \vee r)$	
T	T	T	T	T	**T**	T	**T**	
T	T	F	T	F	**T**	T	**T**	
T	F	T	F	T	**T**	T	**T**	
T	F	F	F	F	**F**	F	**F**	
F	T	T	F	F	**F**	T	**F**	
F	T	F	F	F	**F**	T	**F**	
F	F	T	F	F	**F**	T	**F**	
F	F	F	F	F	**F**	F	**F**	

p	q	r	$(p \vee q)$	$(p \vee r)$	$(p \vee q) \wedge (p \vee r)$	$(q \wedge r)$	$p \vee (q \wedge r)$	
T	T	T	T	T	**T**	T	**T**	
T	T	F	T	T	**T**	F	**T**	
T	F	T	T	T	**T**	F	**T**	
T	F	F	T	T	**T**	F	**T**	
F	T	T	T	T	**T**	T	**T**	
F	T	F	T	F	**F**	F	**F**	
F	F	T	F	T	**F**	F	**F**	
F	F	F	F	F	**F**	F	**F**	

We have a distribute law for propositional logic, it may or may not be surprising to you. Lets try to explain why it should be true. The truth table asserts its correctness, but we can give intution. Consider a proposition which is a specification for possible restarant orders. Let b, f, c be propositional variables representing if you order a burger/fries/coleslaw respectively. You must choose one main and one side, so we may write this as $b \wedge (f \vee c)$. Choose the burger, then choose a side (here, perhaps both). What are the possible meal configurations? You could have had a burger with fries, or a burger with coleslaw. We may write this as $(b \wedge f) \vee (b \wedge c)$.

DeMorgan's Laws

- $\neg(p \wedge q) \equiv (\neg p \vee \neg q)$
- $\neg(p \vee q) \equiv (\neg p \wedge \neg q)$

speech on demorgans law

p	q	$\neg p$	$\neg q$	$p \wedge q$	$\neg(p \wedge q)$	$\neg p \vee \neg q$	$p \vee q$	$\neg(p \vee q)$	$(\neg p \wedge \neg q)$	
T	T	F	F	T	**F**	**F**	T	**F**	**F**	
T	F	F	T	F	**T**	**T**	T	**F**	**F**	
F	T	T	F	F	**T**	**T**	T	**F**	**F**	
F	F	T	T	F	**T**	**T**	F	**T**	**T**	

Absorption

- $p \vee (p \wedge q) \equiv p$
- $p \wedge (p \vee q) \equiv p$

p	q	$p \wedge q$	$p \vee q$	$p \vee (p \wedge q)$	$p \wedge (p \vee q)$	
T	T	T	T	**T**	**T**	
T	F	F	T	**T**	**T**	
F	T	F	T	**F**	**F**	
F	F	F	F	**F**	**F**	

Implications

- $p \implies q \equiv \neg p \vee q$ (conditional disjunction equivalence)
- $p \implies q \equiv \neg q \implies \neg p$ (law of contraposition)
- $p \vee q \equiv \neg p \implies q$
- $p \wedge q \equiv \neg(p \implies \neg q)$

- $\neg(p \implies q) \equiv p \wedge \neg q$

p	q	$\neg p$	$\neg q$	$p \vee q$	$\neg p \implies q$	$p \implies \neg q$	$\neg(p \implies \neg q)$	$p \wedge q$	$p \implies q$	$\neg(p \implies q)$	$p \wedge \neg q$
T	T	F	F	**T**	**T**	F	**T**	**T**	T	**F**	**F**
T	F	F	T	**T**	**T**	T	**F**	**F**	F	**T**	**T**
F	T	T	F	**T**	**T**	T	**F**	**F**	T	**F**	**F**
F	F	T	T	**F**	**F**	T	**F**	**F**	T	**F**	**F**

- $(p \implies q) \wedge (p \implies r) \equiv p \implies (q \wedge r)$
- $(p \implies r) \wedge (q \implies r) \equiv (p \vee q) \implies r$
- $(p \implies q) \vee (p \implies r) \equiv p \implies (q \vee r)$
- $(p \implies r) \vee (q \implies r) \equiv (p \wedge q) \implies r$

p	q	r	$(p \implies q)$	$(p \implies r)$	$(q \wedge r)$	$p \implies (q \wedge r)$	$(p \implies q) \wedge (p \implies r)$	...
T	T	T	T	T	T	**T**	**T**	
T	T	F	T	F	F	**F**	**F**	
T	F	T	F	T	F	**F**	**F**	
T	F	F	F	F	F	**F**	**F**	
F	T	T	T	T	T	**T**	**T**	
F	T	F	T	T	F	**T**	**T**	
F	F	T	T	T	F	**T**	**T**	
F	F	F	T	T	F	**T**	**T**	

$(q \implies r)$	$p \vee q$	$(p \implies r) \wedge (q \implies r)$	$(p \vee q) \implies r$	...
T	T	**T**	**T**	
F	T	**F**	**F**	
T	T	**T**	**T**	
T	T	**F**	**F**	
T	T	**T**	**T**	
F	T	**F**	**F**	
T	F	**T**	**T**	
T	F	**T**	**T**	

$(p \implies q) \vee (p \implies r)$	$p \implies (q \vee r)$	$(p \implies r) \vee (q \implies r)$	$(p \wedge q) \implies r$
T	**T**	**T**	**T**
T	**T**	**F**	**F**
T	**T**	**T**	**T**
F	**F**	**T**	**T**
T	**T**	**T**	**T**
T	**T**	**T**	**T**
T	**T**	**T**	**T**
T	**T**	**T**	**T**

Biconditionals

- $p \iff q \equiv (p \implies q) \wedge (q \implies p)$
- $p \iff q \equiv \neg p \iff \neg q$
- $p \iff q \equiv (p \wedge q) \vee (\neg p \wedge \neg q)$
- $\neg(p \iff q) \equiv (p \iff \neg q)$

p	q	$\neg p$	$\neg q$	$p \implies q$	$q \implies p$	$p \wedge q$	$\neg p \wedge \neg q$	...
T	T	F	F	T	T	T	F	
T	F	F	T	F	T	F	F	
F	T	T	F	T	F	F	F	
F	F	T	T	T	T	F	T	

$p \iff q$	$\neg p \iff \neg q$	$(p \implies q) \wedge (q \implies p)$	$(p \wedge q) \vee (\neg p \wedge \neg q)$	$\neg(p \iff q)$	$(p \iff \neg q)$
T	**T**	**T**	**T**	**F**	**F**
F	**F**	**F**	**F**	**T**	**T**
F	**F**	**F**	**F**	**T**	**T**
T	**T**	**T**	**T**	**F**	**F**

Tautologies and Contradictions

A tautology is a proposition which is always true. A contradiction is a proposition which is always false. $p \vee \neg p$ is a canonical example of a tautology and $p \wedge \neg p$ is a canonical example of a contradiction.

p	$\neg p$	$p \vee \neg p$	$p \wedge \neg p$
T	F	T	F
F	T	T	F

This is how we define excluded middle as a law of thought. It is the case that $p \wedge \neg p$ is always true.

Examples

So far, we know we can demonstrate two propositions to be equivalent by computing their truth tables and observing they have the same columns. But we don't need to do this, we can do without this by simply applying our previously demonstrated laws.

Suppose we want to demonstrate that $\neg(p \implies q)$ is equivalent to $p \wedge \neg q$. We can do this without a truth table as follows

$$\neg(p \implies q) \equiv \quad \text{conditional disjunction equivalence} \tag{1}$$
$$\neg(\neg p \vee q) \equiv \quad \text{DeMorgan's} \tag{2}$$
$$\neg(\neg p) \wedge \neg q \equiv \quad \text{Double Negation} \tag{3}$$
$$p \wedge \neg q \equiv \tag{4}$$

Suppose we want to demonstrate that $(p \wedge q) \implies (p \vee q)$ is always true, it is a tautology.

$$(p \wedge q) \implies (p \vee q) \equiv \quad \text{conditional disjunction equivalence} \tag{5}$$
$$\neg(p \wedge q) \vee (p \vee q) \equiv \quad \text{DeMorgan's} \tag{6}$$
$$(\neg p \vee \neg q) \vee (p \vee q) \equiv \quad \text{Associativity} \tag{7}$$
$$(\neg p \vee p) \vee (\neg q \vee q) \equiv \quad \text{Negation} \tag{8}$$
$$T \vee T \equiv \quad \text{Domination} \tag{9}$$
$$T \tag{10}$$

Rather than doing very complicated and tedious truth tables, we can demonstrate equivalence by applying laws. We demonstrate the absorption law $p \vee (p \wedge q) \equiv p$

$$p \vee (p \wedge q) \equiv \quad \text{Identity} \tag{11}$$
$$(p \wedge T) \vee (p \wedge q) \equiv \quad \text{Distributive} \tag{12}$$
$$p \wedge (T \vee q) \equiv \quad \text{Domination} \tag{13}$$
$$p \wedge T \equiv \quad \text{Identity} \tag{14}$$
$$p \tag{15}$$

All Laws

Here is a small cheatsheet of exactly and only the laws you can use.

- $p \wedge T \equiv p$ **Identity**
- $p \vee F \equiv p$

- $p \vee T \equiv T$ **Domination**
- $p \wedge F \equiv F$

- $p \wedge p \equiv p$ **Idempotent**
- $p \vee p \equiv p$

- $\neg\neg p \equiv p$ **Double Negation**

- $p \wedge q \equiv q \wedge p$ **Communitivity**
- $p \vee q \equiv q \vee p$

- $(p \wedge q) \wedge r \equiv p \wedge (q \wedge r)$ **Associativity**
- $(p \vee q) \vee r \equiv p \vee (q \vee r)$

- $p \wedge (q \vee r) \equiv (p \wedge q) \vee (p \wedge r)$ **Distributive Laws**
- $p \vee (q \wedge r) \equiv (p \vee q) \wedge (p \vee r)$

- $\neg(p \wedge q) \equiv (\neg p \vee \neg q)$ **DeMorgan's Laws**
- $\neg(p \vee q) \equiv (\neg p \wedge \neg q)$

- $p \vee (p \wedge q) \equiv p$ **Absorption**
- $p \wedge (p \vee q) \equiv p$

- $p \vee \neg p \equiv T$ **Negation**
- $p \wedge \neg p \equiv F$

Implication

- $p \implies q \equiv \neg q \implies \neg p$ contrapositive
- $p \implies q \equiv \neg p \vee q$ conditional disjunction equivalence

---

- $p \iff q \equiv (p \implies q) \wedge (q \implies p)$ **Biconditional**

Other Laws

These are some laws which may be demonstrated from those previous. You may not apply these, but you should know them.

- $p \vee q \equiv \neg p \implies q$
- $p \wedge q \equiv \neg(p \implies \neg q)$
- $\neg(p \implies q) \equiv p \wedge \neg q$
- $(p \implies q) \wedge (p \implies r) \equiv p \implies (q \wedge r)$
- $(p \implies r) \wedge (q \implies r) \equiv (p \vee q) \implies r$
- $(p \implies q) \vee (p \implies r) \equiv p \implies (q \vee r)$
- $(p \implies q) \vee (p \implies r) \equiv (p \wedge q) \implies r$
- $p \iff q \equiv (p \implies q) \wedge (q \implies p)$
- $p \iff q \equiv \neg p \iff \neg q$
- $p \iff q \equiv (p \wedge q) \vee (\neg p \wedge \neg q)$
- $\neg(p \iff q) \equiv (p \iff \neg q)$

Predicates and Quantification

Predicates

Recall that we had the example of a declarative sentence "$n > 7$" and that it was not a proposition. Its truth value relied on a "free variable" called n. Were we to fix this variable, then the truth value could be determined, and it would be a proposition. This is called a *predicate*, or a *propositional function*.

We may write this as

$$P(n) : n > 7$$

We will traditionally use capital letters for predicates. Notice that $P(10) \equiv T$, but $P(2) \equiv F$. A predicate may have multiple arguments, such as $P(x, y, z) : x + y = z$. For a predicate to become a proposition, it must be the case it has no free variables. $P(1, 2, z)$ is a predicate, but not a proposition. $P(1, 2, 3)$ is a proposition, and is true.

The predicate $P(n) : T$ is still technically a predicate and not a proposition, even though its truth value does not vary. The constant function $f(x) = 3$ is still a function and not a number.

The *universe of discourse*[10] must be defined for the variables of predicates. There is an understood set of possible values each free variable of a predicate can take on. Without this, the predicate is simply undefined.

[10] Also called the universe, the domain of discourse, or the domain

Consider the predicate

$$P(x, y) : \text{If } x > 0 \text{ then } x + y = 10$$

Observe that

- $P(-1, 100)$ is true
- $P(4, 6)$ is true
- $P(4, 5)$ is false

Again, a predicate is not assigned a truth value until all its free variables have been assigned.

Quantification

Some words used in declarative english sentences include "Any, all, some, none, few," and so on. Consider the english

A quantifier specifies how a variable of a predicate interacts logically with the universe of discourse the variable ranges over.

Existential Quantification

The existential quantifier corresponds to the english words "there exists, some, atleast one" and so on. We may express this quantification by using a *quantifier*. We *bind* a quantifier to a free variable of a predicate. For example, given a

predicate of three free variables $P(x, y, z)$. We may bind an existential quantifier to x and write $\exists x P(x, y, z)$. Here, a backwards "E" is used to mean "there exists x such that:".

Consider the predicate

$$P(x, y) : \text{elephant } x \text{ is heavier than duck } y$$

The universe of discourse of x is all possible elephants, and the universe of discourse of y is all possible ducks. What is a predicate for the sentence 'Some elephant is heavier than duck y"? This is a predicate, but we determine that the elephant must exist, so its truth is dependent only upon y. We may express this using the symbol "$\exists$" and write three examples.

- $\exists x P(x, y)$: There exists an elephant which is heavier than duck y
- $\exists y P(x, y)$: elephant x is heavier than some duck
- $\exists x \exists y P(x, y)$: Some elephant is heavier than some duck.[11]

[11] Most formally, this would be read as "There exists an elephant and there exists a duck such that the elephant is heavier than the duck

Note that the third example not a predicate, but is a proposition! The quantifier $\exists$ *binds* to the variable, making it no longer free. If predicate has all its variables bound, it is a proposition. We can demonstrate the truth of this proposition by finding just one elephant and just one duck such that the elephant weighs more than the duck. Again, the universe of discourse must be defined for the quantifier to make sense. In our previous example, the universe of x is all possible elephants, and the universe of y is all possible ducks.

Consider the proposition $\exists x[x \text{ is even}]$. This is true even if $P(x)$ is false for some values of x. Since we know there is atleast one even number, then we know this is true. There is no claim as to which numbers are even, or how to find them, simply that an even number exists.

Uniqueness

In language, we often want to denote not only that an item exists, but does so uniquely. To this extent, we use the quantifier $\exists!$ to denote this. For example "exactly one number x is positive", we may represent as $\exists! P(x)$. This is not a real quantifier, but you should know the notation if you come across it.

Universal Quantification

While the existential quantifier logically captures the meaning of words like "there is, atleast one, some, there exists" and so on, what about words like "every, for all, for each"? For these, we use the universal quantifier. Let $P(x)$ be a predicate. We write

$$\forall x P(x)$$

to mean that for every single possible value that x could take on from its universe of discourse, the predicate $P(x)$ is true. For example consider the sentence "every elephant is heavier than duck y". We may write this as

$$\forall x P(x, y) : \text{Every elephant is heavier than duck y}$$

We may also quantify over the variable y to get

$$\forall x \forall y P(x, y) : \text{Every elephant is heavier than every duck}$$

.

Consider the Predicate $(x^2 \geq 0)$. We can bind its free variable to get the proposition $\forall x(x^2 \geq 0)$. Note how important the universe of discourse is. If the universe is a restriction of real numbers, then its true. If the universe involves complex numbers, its false.

Equivalence

We may say two statements[12] are equivalent if and only if they have the same truth values, regardless of propositions used, or universes of discourse allowed for the variables of the predicates. We denote $S \equiv T$ to mean these two statements are equivalent.

[12] involving both predicates and quantifiers and propositions

Let $P(x), Q(x)$ be two predicates, with the same variable over some universe of discourse. We demonstrate

$$\forall x(P(x) \wedge Q(x)) \equiv (\forall x P(x)) \wedge (\forall x Q(x))$$

We must perform such a demonstration since a quantification can occur over a universe of discourse which is infinite. We cannot do an infinitely large truth table to show equivalence. We know that $\Phi_1 \equiv \Phi_2$ is true exactly and only when $\Phi_1 \iff \Phi_2$ is a tautology. So we will demonstrate that $\Phi_1 \implies \Phi_2$ and $\Phi_2 \implies \Phi_1$.

Suppose that $\forall x(P(x) \wedge Q(x))$ is true. Then for all a in the universe, we know that $P(a) \wedge Q(a)$ is true. So both $P(a)$, $Q(a)$ are true. Since $P(a)$ is true and $Q(a)$ is true for any a in the univese, we know that $\forall x P(x)$ is true, and $\forall x Q(x)$ is true. So $(\forall x P(x)) \wedge (\forall x Q(x))$ is true.

Suppose that $(\forall x P(x)) \wedge (\forall x Q(x))$ is true. Then $(\forall x P(x))$ is true, and $(\forall x Q(x))$ is true. Then since they share the universe, we know that for every a that $P(a)$ is true and $Q(a)$ is true. So $P(a) \wedge Q(a)$ is true. Since a is any element in the universe of x, we see that $\forall x(P(x) \wedge Q(x))$ is true.

We can write the uniqueness quantifier equivalently as just an existential one

$$\exists! P(x) \equiv \exists x[P(x) \wedge \forall y(y \neq x \implies \neg P(y))]$$

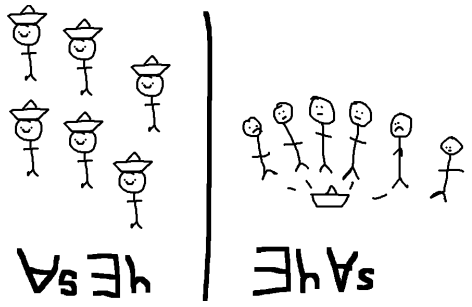
To interpret this back in natural language, it states that $\exists x P(x)$, for every other distinct value y, that $\neg P(x)$. This is the definition of uniqueness.

Multiple Quantifiers

Note that the order of quantifiers does matter. **They do not commute.** Suppose that x, y have universes of booleans. Observe that the following are not equivalent

$$\forall x \exists y[(x \vee y) \wedge (\neg x \vee \neg y)]$$
$$\exists y \forall x[(x \vee y) \wedge (\neg x \vee \neg y)]$$

Consider the semantic difference between "every sailor has a hat" and "some hat has every sailor".



Negation of Quantifiers

How do you compute the negation of a quantifier? For example let $P(x)$ be the predicate that $P(x)$: "man x is mortal" and the proposition $\forall x P(x)$ to

mean that "all men are mortal". The logical opposite of "all men are mortal" may be interpreted as "there is a man who is not mortal". The negation of a universal quantification can be understood as an existential quantification of the opposite!

$$\neg(\forall x P(x)) \equiv \exists x[\neg P(x)]$$

Similarly, the negation of the statement "some man is not mortal" could be understood as "all men are mortal. So

$$\neg(\exists x Q(x)) \equiv \forall x[\neg Q(x)]$$

One way to think about this is as a generalization of DeMorgan's law. You cannot have an infinitely long proposition, but if you could, a quantifier could be expressed this way.

$$\neg(\exists x P(x)) \equiv \neg(P(0) \vee P(1) \vee ...) \equiv (\neg P(0) \wedge \neg P(1) \wedge ...) \equiv \forall x(\neg P(x))$$

Again, you cannot have an infinitely long proposition, but DeMorgan's law carries over this way. Consider the propsition $\forall x[x^2 \geq x]$ where the universe of x is integers. Note that this proposition is true. If we compute the negation of it, then

$$\neg(\forall x[x^2 \geq x]) \equiv \exists x \neg[x^2 \geq x] \equiv \exists x[x^2 < x]$$

If your proposition has multiple quantifiers, then we may represent this in a nested manner. For example, $\forall x \exists y P(x, y)$ may really mean $\forall x[\exists y P(x, y)]$, where there is an internal predicate. Negation is handled recursively so that

$$\neg(\forall x \exists y \forall z ... P(x, y, z, ...)) \equiv \exists x \forall y \exists z[\neg P(x, y, z, ...)]$$

Lets compute the negation of the unique existential quantifier. Since

$$\exists! x P(x) \equiv \exists x[P(x) \wedge \forall y(y \neq x \implies \neg P(y))]$$

Then

$$\neg(\exists! P(x)) \equiv \forall x[\neg P(x) \vee \exists y(y \neq x \wedge P(y))]$$

We can read this back in english as either its false for all possible x, or if its true for one x, there exists a distinct y which its also true for. So either it doesn't exist, or if it exists, its not unique.

Prenex Normal Form

A quantified proposition is said to be written in prenex normal form if it can be written as all quantifications coming first. For example

$$\forall x \exists y ... \Phi(x, y, ...)$$

Every quantified predicate or proposition can be rewritten into an equivalent one which is in prenex normal form.

Table

- $\forall x P(x)$ is true when $P(x)$ is true for every x, and is false when there is an x such that $P(x)$ is false.
- $\exists x P(x)$ is true, when there is an x such that $P(x)$ is true, and is false when for every x that $P(x)$ is false.
- $\neg\exists x P(x)$ is understood as "there does not exist an x where $P(x)$ is true", so it is equivalent to "for every x that $P(x)$ is false, or $\forall x \neg P(x)$.
- $\neg\forall x P(x)$ is understood as "It is false that for every x that $P(x)$ is true" which is equivalent to "There is an x such that $P(x)$ is false, or $\exists x \neg P(x)$.

Inference

Argument

We have seen that propositional logic is good at removing ambiguity from many parts of natural language, but we have not seen how it can be used to establish truth. The act of deduction is done sequentially, as a sequence of steps. An argument is a sequence of statements which establishes the total, undeniable truth and validity of some statement.

The form of an argument usually begins with a presumed body of knowledge $p_1, p_2, ..., p_k$. Each of these statements consists of the facts, and are presumed true. They are called premises. You wish to *deduce* a *conclusion*, called q. We may represent this as

$$(p_1 \wedge p_2 \wedge ... \wedge p_k) \implies q$$

We conjunct the body of knowledge together because they all must be true. An argument is said to be valid if $(p_1 \wedge ... \wedge p_k) \implies q$ is a tautology. We could demonstrate the validity of an argument by writing out a truth table. But note, we actually do not care about situations when any premise $p_1, ..., p_k$ is false. We need to only demonstrate that q follows from when $p_1, ..., p_k$ are all true. Observe that if any premise is false, then the deduction trivially becomes true. Many people[13] do not act illogically, they act logically from wrong premises. The steps of their argument appear correct, but since they assume an invalid premise, then they could "argue" the "truth" of any statement. Recall that an implication is trivially true if its premise is false. For an argument to be correct, its premises must also be true.

[13] Debate bro's, flat earthers, etc

burden of proof

The Rules of Inference

A rule of inference is like a law of thought, in that we may apply it to deduce some statement from a given collection of premises or other deductions. When we use a law of thought to manipulate a proposition, such as $\neg(p \vee q) \equiv (\neg p \wedge \neg q)$, these laws preserve truth. If its true before, its true after applying the law, and if it was false before, it will remain false after applying the law.

The rules of inference are less general and more specific, but this allows them more power in presenting an argument. Combination of premises may assume the combined premises are true.

We construct a rule of inference in the following syntax. Let $p_1, ..., p_k$ be the premises, and let q be the conclusion. Then we may write

$$\begin{array}{c} p_1 \\ ... \\ p_k \\ \hline \therefore q \end{array}$$

To mean that from premises $p_1, ..., p_k$ we may deduce q. The symbol $\therefore$ means "therefore".

Modus Ponens

$$\begin{array}{c} p \\ p \implies q \\ \hline \therefore q \end{array}$$

If p, and if $p \implies q$, then we may deduce that q is true. It is not too hard to show this is a tautology with a truth table as well.

p	q	$p \implies q$	$(p \implies q) \wedge p$	$((p \implies q) \wedge p) \implies q$	
T	T	T	T	T	
T	F	F	F	T	
F	T	T	F	T	
F	F	T	F	T	

Let p be the proposition corresponding to "Today is Thursday". Let $p \implies q$ correspond to "If today is Thursday, then you will go to class". We may deduce then that q is true, that "you will go to class". This is one of the most basic deductive tools we have as humans. You may interpret $p \implies q$ as "actions have consequences", and you may interpret p as "actions occur". So the deduction of q is that "consequences occur". It is a rule of inference because the truth table is a tautology, sure. But it is a good rule of inference because it is often used in our day-to-day cognitive problem solving and existence.

Modus Tollens

$$\begin{array}{c} \neg q \\ p \implies q \\ \hline \therefore \neg p \end{array}$$

Loosely, if $p \implies q$, but q never happened, then p didn't happen. If $p \implies q$ corresponds to the proposition "If it is Friday then students wear blue", and $\neg q$ corresponds to "students are not wearing blue", then we may deduce that "It is not Friday" is true. You may again interpret $p \implies q$ as "actions have consequences", and you may interpret $\neg q$ as "consequences didn't occur", so the deduction of $\neg p$ is "actions must not have occured".

Hypothetical Syllogism

$$\begin{array}{c} p \implies q \\ q \implies r \\ \hline \therefore p \implies r \end{array}$$

If $p \implies q$ corresponds to the proposition "If you make an A on the final, you will pass the class" and $q \implies r$ corresponds to "If you pass the class, then you will graduate on time". We may deduce $p \implies r$, that "If you make an A on the final, you will graduate on time." Actions have consequences sure, but those consequences may be actions for even more consequences. This also displays that the implication in propositional logic is a transitive relation.

Disjunctive Syllogism

$$\frac{\begin{array}{c} p \vee q \\ \neg p \end{array}}{\therefore q}$$

If we take $p \vee q$ to mean "Bob brought cake or Alice brought cake" and $\neg p$ to mean "Bob did not bring cake". Then it must be the case that q: "Alice brought cake".

Addition

$$\frac{p}{\therefore p \vee q}$$

If p : "I like dogs", then $p \vee q$:"I like dogs or cats" is certainly true.

Simplification

$$\frac{p \wedge q}{\therefore p}$$

If $p \wedge q$ means "I like both dogs and cats" then p : "I like dogs" is true.

Conjunction

$$\frac{\begin{array}{c} p \\ q \end{array}}{\therefore p \wedge q}$$

If p :"I like dogs" and if q :"I like cats". Then $p \wedge q$:"I like dogs and cats" is true.

Resolution

$$\frac{\begin{array}{c} p \vee q \\ \neg p \vee r \end{array}}{\therefore q \vee r}$$

It is the case that p is always true or always false. So in the case that p is true, then r must be true, and in the case that p is false, then q must be true. Either way, $q \vee r$ must be true since $p \vee \neg p$ is true.

Examples

Lets give an example of an argument with several steps and several applications of the rules of inference. We demonstrate

$$\begin{array}{c} (\neg p \vee \neg q) \implies (r \wedge s) \\ (r \implies t) \\ \neg t \\ \hline \therefore p \end{array}$$

1. $r \implies t$ (Premise)
2. $\neg t$ (Premise)
3. $\neg r$ (Modus Tollens of 1,2)
4. $\neg r \vee \neg s$ (Addition of 3)
5. $\neg(r \wedge s)$ (DeMorgan's Law)
6. $(\neg p \vee \neg q) \implies (r \wedge s)$ (Premise)
7. $\neg(\neg p \vee \neg q)$ (Modus Tollens of 5,6)
8. $(\neg\neg p \wedge \neg\neg q)$ DeMorgan's law
9. $(p \wedge \neg\neg q)$ Double Negation
10. p Simplification

Principle of Explosion

Given the laws of arithmetic, you should not be able to correctly deduce that $0 = 1$, or anything else incorrect. The laws are harmonious with each other, and preserve truth. Similarly, the laws given for propositional calculus also preserve truth.

We demonstrate the *Principle of Explosion.*[14] If there exists any statement which is both true and false simultaneously, then every statement is both true and false simultaneously. Truth is then meaningless. [15]

Let p be some single statement such that both p and $\neg p$ are true. Let q be any possible statement. We will demonstrate that q is true. It could be anything, that the sun won't rise tomorrow, that $1+1 = 3$, that zebras are and are not blue.

1. p (Premise)
2. $\neg p$ (Premise)
3. $p \vee q$ (Addition of 1)
4. q (Disjunctive Syllogism of 2,3)

Important here is that nothing about q was referenced other than it exists. You could repeat this similarly for $\neg q$.

[14] https://xkcd.com/704

[15] In a psychological context, cognitive dissonance refers to the mental disturbance people experience when they realize their thoughts, cognitions, or actions may be contradictory. The principle of explosion could be understood as an application of the propositional calculus to explain this.

Fallacies

Fallacies can include an incorrect application of correct laws of thought. For example, $((p \implies q) \wedge q) \implies p$ is not a tautology because it may be false when p is false. If $p \implies q$ is to mean "If you get into a car accident, you will die" and q is to mean "You die". You cannot conclude you got into a car accident. You may have died from other methods (perhaps a meteor). This is called the fallacy of affirming the conclusion.

Similarly $((p \implies q) \wedge \neg p) \implies \neg q$ is not a tautology. If you do not get into a car crash, you are not immortal, as you may die of other methods. This is called the fallacy of denying the hypothesis.

Quantified Statements

The rules of inference may also apply to those statements which are quantified.

Universal Instantiation

$$\frac{\forall x P(x)}{\therefore P(c)}$$

If all men are mortal, then Socrates is mortal.

Universal Generalization

$$\frac{P(c) \text{ for any } c}{\therefore \forall x P(x)}$$

Existential Instantiation

$$\frac{\exists x P(x)}{\therefore P(c) \text{ for some } c}$$

We do not know which c this is true for, only that it is true for some c.

Existential Generalization

$$\frac{P(c) \text{ for some } c}{\therefore \exists x P(x)}$$

These rules may seem redundant, but they are necessary when you may syntactically need a quantifier or not.

speech on this more

One of the classic examples of logic is All men are mortal Socrates is a man, therefore Socrates is mortal

We may formalize this, and express this in formal language. Let $Man(\cdot)$ and $Mortal(\cdot)$ be two predicates whos variables range over the same universe of discourse of all objects of being.

- All men are mortal: $\forall x(Man(x) \implies Mortal(x))$

- Socrates is a man: $Man(Socrates)$

1. $\forall x(Man(x) \implies Mortal(x))$ (Premise)
2. $Man(Socrates) \implies Mortal(Socrates)$ (Universal Instantiation)
3. $Man(Socrates)$ (Premise)
4. $Mortal(Socrates)$ (Modus Ponens of 2,3)

Introduction to Proof

Why Prove Things?

Our goal with mathematics is to seek truth in all forms. The purpose of proof is to establish the total and *convincing* truth. It is evident that truth may only be derived and established from other truths. If we wish to demonstrate total certainty of a mathematical statement, then we must make some basic assumptions. These are called *Axioms*

Definition 0.0.1 (Axiom)**.** *An axiom is a statement which may be assumed true without proof.*

Different fields of math use different sets of axioms, and the set of axioms you use characterizes the math you are working in. In real numbers, we have axioms like $ab = ba$, the communitivity of multiplication. Or $a(b+c) = ab+ac$, distributivity. Usually an axiom is so simple, it is impossible to prove it, and there is little debate whether or not an axiom is true. It is so simple that it must be true. The axiomatic method[16] is the standard method we use to establish truth. You begin with only axioms and previously established truths. You perform a sequence of true deductions, and concludes with the statement to be proven.

Definition 0.0.2 (Theorem)**.** *A theorem is a statement which is not an axiom, but has been proven true.*

A proof from the axioms involves combining axioms with the laws of thought (themselves axioms) and other proven theorems. A corollary is a theorem which follows some more general theorem. A lemma is a tiny helper theorem used to prove some main theorem. A conjecture is a statement which is unproven. It may be hard to prove, but a mathematician states it hoping someone else may prove it some day in the future. There is also a connotation that a theorem should be interesting. The fact that $1 + 1 = 2$ is true, and not an axiom, so it must be a theorem, but few would call it that.

Universes of Discourse

Without defining what a set is, we implement some common notation for the universes of discourse.

- The Naturals $\mathbb{N} = 0, 1, 2, 3, ...$
- The Integers $\mathbb{Z} = ..., -2, -1, 0, 1, 2, ...$
- The Rationals $\mathbb{Q} = a/b$ where a, b are any integer but b isn't zero.

[16] The axiomatic method began with Euclid over two millenia ago. He gave five simple axioms for what we now call Euclidean geometry. For example the fourth axiom is just "all right angles equal each other". From just these five axioms, he was able fill several volumes with proofs. Some theorems he proved include the interior angle sum of any triangle is 180 degrees, the Pythagorean theorem, and much more. His thirteen volumes of *The Elements* has been translated and studied across civilizations for over two thousand years.

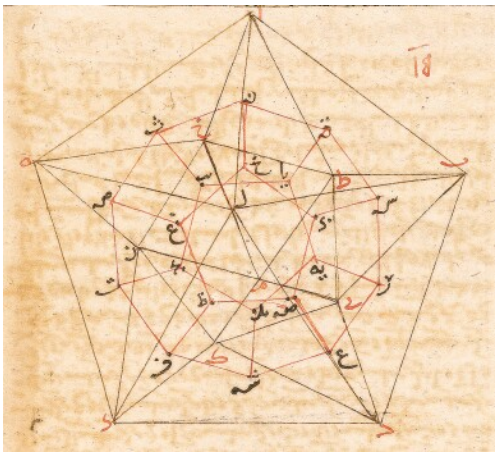


Figure 4: Euclid's construction of the dodecahedron, from Nasir al-Din al-Tusi's translated copy of The Elements, 1280 AD.

- The Irrationals $\mathbb{I}$ any quantity which isn't rational
- The Reals $\mathbb{R}$ = any number with any decimal expansion. Every real is either rational or irrational.
- The Complex Numbers $\mathbb{C} = a + bi$ where a, b are any reals and $i^2 = -1$.

Different theorems require different techniques. Given a statement, it may beg for you to use this or that specific technique. You should keep these techniques like tools in your toolbox, and knowing when to use what tool is a skill to develop.

Direct Proof

For this section, let the universe of discourse be the natural numbers.

Definition 0.0.3 (Even Number)**.** *A number is even if it satisfies the predicate*

$$Even(n) := \exists k[n = 2k]$$

This is a definition. A number is even if it can be written as two times a number. It is even if it can be split in two wholes equally. A number is even if two divides it.

Definition 0.0.4 (Odd Number)**.** *We can correspondingly define the predicate*

$$Odd(n) := \exists k[n = 2k + 1]$$

Note that a number is odd if and only if it is not even. $Odd(n) \equiv \neg Even(n)$.

Theorem 1. *The product of two even numbers is even.*

We could write this using the predicate calculus as

$$\forall a \forall b[(Even(a)) \wedge (Even(b)) \implies (Even(ab))]$$

We do not often wish to over detail a theorem in terms of predicates and quantifiers. It can become too cumbersome. Rather, we express them in terms of natural language. This is a relatively simple statement, but already involves two quantifiers, a logical and, and an implication. Statements we wish to prove may be far more complex if written this way. If asked to rewrite a statement into the propositional and predicate calculus, you should be able to. Otherwise, just know it is going on in the background. Now let us prove the theorem.

Proof. Let a be an even number. Then there exists a number k such that $a = 2k$. Let b be an even number. Then there exists a number l such that $a = 2l$. Then $ab = (2k)(2l) = 2(2kl)$. Since we may write ab as two times a number, it is even. □

It is polite that the beginning and end of your proof are denoted in some way. In a larger body of text, which may contain more rambling thoughts, you want to make it clear and explicit to the reader where the argument begins and where the argument ends. Note that this proof actually shows more. It shows that the product of two even numbers is actually divisible by four. Its like, twice as even as normal even number. Doesn't matter. We are tasked with proving that a product of even numbers was even. Were we to conclude that a

Why even bother to prove things? Why not simply observe that $0 \cdot 0, 0 \cdot 2, 2 \cdot 2, 4 \cdot 2$ are all even and call it a day. A few examples of a theorem does not constitute a proof, and does not imply a theorem is true for all a, b. This is especially true for a statement universally quantified over an infinite universe of discourse.

Pierre de Fermat was a 17th century French lawyer and hobbyist mathematician. He noticed the following pattern:

$$2^{2^0} + 1 = 3$$
$$2^{2^1} + 1 = 5$$
$$2^{2^2} + 1 = 17$$
$$2^{2^3} + 1 = 257$$
$$2^{2^4} + 1 = 65537$$

They are all prime. He reasonably conjectured that for all n that $2^{2^n} + 1$ is prime. But Euler showed the very next number in the sequence was not prime.

$$2^{2^5} + 1 = 4294967297 = 641 \cdot 6700417$$

In fact, the only known values in which $2^{2^n} + 1$ is prime are $n = 0, 1, 2, 3, 4$. We may observe a pattern, and predict that such a pattern may continue, but there are infinitely many numbers, and we may make only finitely many observations. Our proof demonstrates truth for any possible a, b, in the way a few examples could never.

"*The man who has fed the chicken every day throughout its life at last wrings its neck instead, showing that more refined views as to the uniformity of nature would have been useful to the chicken.*" - Bertrand Russell

product of even numbers was divisble by four, it may not be immediate and clear to the reader that is sufficient for it to be even. The proof should conclude exactly with the theorem to be proved, for clarity. Lets do some more simple examples.

Theorem 2. *The product of an odd number and an even number is even.*

Proof. Let a be an even number. Then $a = 2k$ for some number k. Let b be an odd number. Then $b = 2l + 1$ for some number l. Then $ab = (2k)(2l + 1) = 2(k(2l + 1))$. Since we may write ab as two times a number, it is even. □

Theorem 3. *The product of an odd number and an odd number is odd.*

Proof. Let a be an odd number. Then $a = 2k + 1$ for some number k. Let b be an odd number. Then $b = 2l+1$ for some number l. Then $ab = (2k+1)(2l+1) = 4kl + 2k + 2l + 1 = 2(2kl + k + l) + 1$. Since we may write ab as two times a number plus one, it is odd. □

Corollary 4. *If n is a number and n^2 is odd then n is odd.*

Recall a corollary is a tiny theorem following some main one. This actually doesn't directly follow from theorem 3, but from theorem 3 and the contrapositive of theorem 1. The product of even numbers is even and the product of odd numbers is odd. So n^2 being odd means that n cannot be even, so it must be odd.

One of the most famous examples of a counterexample involves the dialogue of Diogenes and Plato. Plato, great man and great mind, had a school in Athens. He had many students and much recognition. Diogenes was an eccentric character who lived in a barrel on the outskirts of the city. Plato, to his school, attempts to establish the definition of a man (as in humanity). Plato asserts that

$$\text{Man}(x) \iff \neg\text{Feathered}(x) \wedge \text{Biped}(x)$$

All that are humanity are featherless bipeds, and all that are featherless bipeds are man. Plato was interested in a dichotomy and hierarchy of all objects, real or otherwise. To an ancient greek man, the only things he may have seen include some sheep, a mountain, a cloud, etc. Everything is or isn't a biped, and is or isn't featherless. All examples of a biped he may have known had feathers, except man. As the myth goes, Diogenes busts into the amphitheatre, raises a plucked chicken and yells *"Behold! A Man!"*. This is a counterexample. Is a plucked chicken a featherless biped? Yes. Is it a man? Certainly not. Then

$$\text{man} \not\Longleftrightarrow \text{featherless biped}$$

Diogenes displays this counterexample, and proves Plato wrong.

Proof by Counter Example

If trying to prove a universally quantified statement to be false, such as $\forall x P(x)$, you can simply find one example c in which $P(c)$ is false, since

$$\neg(\forall x(P(x))) \equiv \exists x(\neg P(x))$$

An example in which a statement is false is called a counterexample.

Theorem 5. *It is false that every positive number is the sum of two squares.*

We could represent this as the negation of $\forall n \exists a \exists b[(n > 0) \implies (n = a^2 + b^2)]$, but we shouldn't. To prove that it is false, you simply need to demonstrate a counterexample where it is false.

Proof. Consider $n = 3$. For what values a, b could it be the case that $3 = a^2 + b^2$? Since $2^2 = 4$, we know that $a < 2$ and $b < 2$. So a, b can only be 0 or 1. Lets try all possible combinations.

$$0^2 + 0^2 = 0$$
$$1^2 + 0^2 = 1$$
$$0^2 + 1^2 = 1$$
$$1^2 + 1^2 = 2$$

We only get the possible values of $0, 1, 2$. So 3 is a counter example to the statement, and it is thus, proven false. □

Figure 5: Diogenes seated with his barrel behind him, and reading a book while holding a stick that rests on a geometry book to his right, Giovanni Jacopo Caraglio, 1526-27

Proof by Contraposition

Recall that we proved using a truth table that the contrapositive of an implication was equivalent to it.

$$p \implies q \equiv \neg q \implies \neg p$$

To prove an implication. It may then be easier to prove its contrapositive.

Theorem 6. *If $5n + 4$ is odd, then n is odd.*

Let us try to prove it directly first. Assume $5n + 4$ is odd, and we will try to prove n is odd. If $5n + 4$ is odd then $5n + 4 = 2k + 1$ for some k. Moving terms around, we see that $n = (2k - 3)/5$. Here we get stuck. Its not even clear if n is a natural number, let alone an odd one. Lets instead prove the contrapositive.

Proof. We prove the equivalent statement that if n is even, then $5n + 4$ is even. if n is even, then $n = 2k$ for some number k. If we substitute it into $5n + 4$, we get $5n + 4 = 5(2k) + 4 = 2(5k + 2)$. Since $5n + 4$ can be written as two times a number, it is even. □

Here, observe that the contrapositive was easier to prove. A direct proof of the theorem may exist, but you want the shortest, clearest proof possible. If you prove the contrapositive, please write the contrapositive of the theorem for the reader. Lets do another example.

Theorem 7. *If $n = ab$ then $a \leq \sqrt{n}$ or $b \leq \sqrt{n}$*

This one is also difficult to prove directly, but its a very useful property of composite numbers. Since n is so general, we don't really have good information about a or b to work with, except that they exist. Lets instead prove the contrapositive.

Proof. Assume that $a > \sqrt{n}$ and $b > \sqrt{n}$. We prove that $ab \neq n$. If $a > \sqrt{n}$ and $b > \sqrt{n}$ then $ab > \sqrt{n}\sqrt{n} = n$. So since $ab > n$, we know $ab \neq n$. □

We get an interesting corollary. The smallest prime factor of a composite number is less than or equal to its square root.

Proof By Contradiction

A proof by contradiction is one of the most versatile techniques, and also may involve some creativity. If you wish to demonstrate some proposition p is true, you can show the negation of the proposition must be absurd. That $\neg p \implies (0 = 1)$. For this reason, it is also called *Reductio Ad Absurdum*[17]

Your proof should always begin soon after stating the theorem. The first sentence of your proof should be an acknowledgement that you are about to perform a proof by contradiction. Traditionally, if you want to prove p, you may begin with "Assume to the contrary $\neg p$". Or sometimes simply "Suppose not". It must be made explicit in some way. You should proceed with deduction applying laws of thought, until you produce *the absurdity*. The absurdity is a statement derived as a consequence of $\neg p$. It aught to be so absurd that the reader will have no choice but to accept that $\neg p$ must be false. The absurdity can take on the form of a negation of a premise or the negation of an axiom. It can take on the form that there is some statement that $p \wedge \neg p$ is true. The

[17] "*Reductio ad absurdum, which Euclid loved so much, is one of a mathematician's finest weapons. It is a far finer gambit than any chess play: a chess player may offer the sacrifice of a pawn or even a piece, but a mathematician offers the game.*" - G. H. Hardy, *A Mathematician's Apology*

concluding absurdity should be so absurd, that the reader should gawk. If its not absurd enough, proceed further in deduction.

demonstrate proof by contradiction works Let c be a contradictory statement such that $\neg c$ is true. A proof by contradiction of p would then be written as $\neg p \implies c$. Why is proof by contradiction a valid proof technique? Modus Tollens:

$$\frac{\begin{array}{c}\neg c \\ \neg p \implies c\end{array}}{\therefore p}$$

Lets do a few examples.

Theorem 8. *There is no largest number*

Proof. Assume to the contrary there was a largest number n. Consider the number $n+1$. We know $n+1$ is a number when n is a number, but $n < n+1$, so n was not the largest number, contradiction. □

The statement of the theorem is obvious, but take note of the setup and syntax. The fact it is a proof by contradiction is declared at the beginning. The absurdity is the negation of a premise.

Theorem 9. *If x, y are positive real numbers, then $\sqrt{x+y} \neq \sqrt{x} + \sqrt{y}$*

Proof. Assume to the contrary that there exists positive real numbers x, y such that $\sqrt{x+y} = \sqrt{x} + \sqrt{y}$. Then

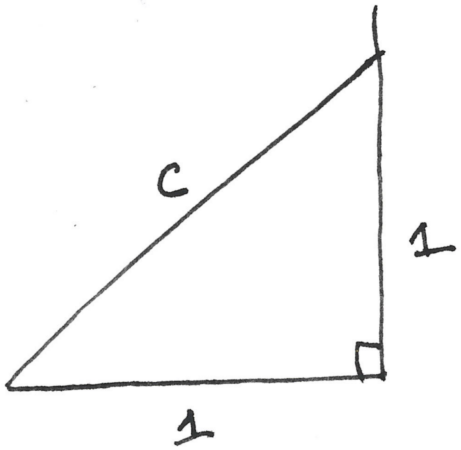
$$\sqrt{x+y} = \sqrt{x} + \sqrt{y} \tag{16}$$
$$x + y = (\sqrt{x} + \sqrt{y})^2 \tag{17}$$
$$x + y = x + 2\sqrt{xy} + y \tag{18}$$
$$0 = 2\sqrt{xy} \tag{19}$$
$$0 = xy \tag{20}$$

By the zero product property, if $xy = 0$, then one of x, y must be zero. This contradicts our assumption that x, y are both positive. □

Note again how we negate the implication here. Recall that $\neg(p \implies q) \equiv p \wedge \neg q$. We phrase this negation as there do exist positive real numbers (p), but $\sqrt{x+y} = \sqrt{x} + \sqrt{y} \quad (\neg q)$.

We finish with one more proof and its legend. Pythagoras is well known for many advancements in mathematics, including the Pythagorean theorem[18] He led a society, a cult maybe, which believed in numerology. They believed that all of nature could be explained by either numbers, or ratio of whole numbers. Today we write $\frac{2}{3}$ and understand it as a "part". They did not. They would have only interpretted this as $2:3$, as in two wholes to three wholes, as a ratio. We may eat $\frac{2}{3}$rds of a pie. They would have understood it as two wholes to three wholes. Two pies of three pies. Every number they believed was either whole, or a ratio. The concept of an irrational number was unfathomable to them. Following the Pythagorean theorem grew an essential question. What ratio was the hypotenuse of a right triangle with unit side lengths? How long was the diagonal of a square of side lengths 1?

Following the Pythagorean theorem, we see that $1^2 + 1^2 = c^2$. For what ratio c could $c^2 = 2$? Today we know that $\sqrt{2}$ can not be rational, it cannot



[18] Even though it had been discovered by others, a few thousand years before him.

be represented as a ratio of whole numbers. Ancient civilizations thought it might be 577/408 or even 305470/216000, but these are simply approximations. Pythagoras could not comprehend that an irrational number could exist, since it contradicted his view of nature. A student[19] of his, was able to demonstrate that not only do irrational quantities exist, but $c = \sqrt{2}$ must be irrational.

Theorem 10. *The number $\sqrt{2}$ is irrational.*

Proof. Assume to the contrary that $\sqrt{2} = m/n$ for m, n numbers in reduced form. The numbers m, n do not share any factors, the ratio has been simplified. Certainly every rational number can be written in such a reduced form. Since it is reduced, we know both m, n cannot both be even, so atleast one must be odd. We may write

$$\sqrt{2} = \frac{m}{n} \tag{21}$$
$$\sqrt{2}n = m \tag{22}$$
$$(\sqrt{2}n)^2 = m^2 \tag{23}$$
$$2n^2 = m^2 \tag{24}$$

Since we may write m^2 as two times a number, it must be that m^2 is even. Since the square of an odd number is always odd, then m must also be even. So $m = 2k$ for some k. Then

$$2n^2 = m^2 \tag{25}$$
$$2n^2 = (2k)^2 \tag{26}$$
$$2n^2 = 4k^2 \tag{27}$$
$$n^2 = 2k^2 \tag{28}$$

Since we can write n^2 as two times something, n^2 is also even, so we know that n must also be even. But how can both m, n be even? We assumed they were both reduced. If they are both even, they are not reduced, as they share the common factor of two. A contradiction. □

wording on the foreshadowing

A number is prime if the only numbers which divide it are 1 and itself. The first few prime numbers are $2, 3, 5, 7, 11, 13, 17...$ As another famous proof by contradiction, we present Euclid's proof of the infinitude of primes.

Theorem 11 (Euclid's Theorem)**.** *There are infinitely many primes.*

Proof. Assume to the contrary there are only finitely many prime numbers. Let them be denoted as $p_1, p_2, ..., p_k$ where p_i denotes the ith prime number. Consider the number

$$n = (p_1 \cdot p_2 \cdot ... \cdot p_k) + 1$$

Note that n is not equal to any of the finitely many primes, so by assumption, it must not be prime. Since it is composite, it has some prime divisor p, which must be one of $p_1, ..., p_k$. But then p divides $P = (p_1 \cdot ... \cdot p_k)$ and p divides $n = (p_1 \cdot p_2 \cdot ... \cdot p_k) + 1$. So p divides $n - P = (p_1 \cdot p_2 \cdot ... \cdot p_k) + 1 - (p_1 \cdot p_2 \cdot ... \cdot p_k) = 1$. But no prime number divides into 1, a contradiction. □

[19] "*For Pythagoras, the beauty of mathematics was the idea that rational numbers (whole numbers and fractions) could explain all natural phenomena. This guiding philosophy blinded Pythagoras to the existence of irrational numbers and may even have led to the execution of one of his pupils. One story claims that a young student by the name of Hippasus was idly toying with the number $\sqrt{2}$, attempting to find the equivalent fraction. Eventually he came to realise that no such fraction existed, i.e. that $\sqrt{2}$ is an irrational number. Hippasus must have been overjoyed by his discovery, but his master was not. Pythagoras had defined the universe in terms of rational numbers, and the existence of irrational numbers brought his ideal into question. The consequence of Hippasus' insight should have been a period of discussion and contemplation during which Pythagoras ought to have come to terms with this new source of numbers. However, Pythagoras was unwilling to accept that he was wrong, but at the same time he was unable to destroy Hippasus' argument by the power of logic. To his eternal shame he sentenced Hippasus to death by drowning. The father of logic and the mathematical method had resorted to force rather than admit he was wrong. Pythagoras' denial of irrational numbers is his most disgraceful act and perhaps the greatest tragedy of Greek mathematics. It was only after his death that irrationals could be safely resurrected.*" - Simon Singh, *Fermat's Enigma*

Proving If and Only If Statements

Exhaustive Proof

Suppose we want to prove a statement of the form $\forall x P(x)$. If we are lucky enough that the universe of discourse of x is finite, then we may simply prove it for each x. If x can only be one of a, b, c, d, then $\forall x P(x) = P(a) \wedge P(b) \wedge P(c) \wedge P(d)$ Lets do a simple example

Theorem 12. *If n is a number between two and four, then $n^2 > n$*

Proof. We confirm that $2^2 = 4 > 2$ and $3^2 = 9 > 3$ and $4^2 = 16 > 4$. □

This obviously doesn't work in the case that the universe of discourse is infinite. You are not allowed to have an infinitely long proof. A proof of $\forall x P(x)$ must itself be of finite length, but assert something which is true for infinitely many values of x.

Proof by Cases

A theorem may require a trickier proof, in that it may need to be decomposed into *cases*. If you wish to prove a statement of the form $(p_1 \vee ... \vee p_k) \implies q$, it is equivalent to prove $(p_1 \implies q) \wedge ... \wedge (p_k \implies q)$. For example, if you wish to prove a statement about all numbers, you may do it into cases, one case with the assumption that your number is even, and another case with the assumption that your number is odd. Each case may imply the conclusion for very different reasons, and each case may be proven with different techniques even.

Theorem 13. *If n is any number, then $n^2 + n$ is even.*

Proof. Let n be any number. Then we have two cases if n is even or odd.

- Case 1: If n is even, then $n = 2k$ for some number k. Then $n^2 + n = 4k^2 + 2k = 2(2k^2 + 1)$ which is even.
- Case 2: If n is odd, then $n = 2k + 1$ for some number k. Then $n^2 + n = 4k^2 + 4k + 1 + 2k + 1 = 2(2k^2 + 3k + 1)$ which is even.

□

Note that when you break your problem into cases, they must cover the entire universe of discourse. If you wish to prove something is true for any integer of $\mathbb{Z}$, it is not sufficient to prove it in the cases that $x > 0$ and $x < 0$, since you have not covered the case that $x = 0$. When presenting your cases to the reader, it has to be obvious that the cases cover all possibilities. Famously, the four color theorem was proved by checking nearly two thousand cases. Along with a proof of each case, they have to provide a proof that the those were the only cases which needed to be proven.

Proving Uniqueness

Recall we characterized the uniqueness quantifier as

$$\exists! x P(x) \equiv \exists x (P(x) \wedge \forall y (x \neq y \implies \neg P(y)))$$

Proving uniqueness is a two step process. First you show the object exists. Then you show the object is unique. Usually for the second step, it is done in the contrapositive. Suppose that there are two objects, both have the property a, b. Conclude that it must be the case that $a = b$, and that they are not actually distinct. It is just one object with two names.

Theorem 14. *If x is any nonzero rational number, then there exists a unique rational number y such that $xy = 1$.*

We are proving that the reciprocal of a nonzero fraction is unique.

Proof. First we show that y exists, and it does so uniquely. Let x be any nonzero rational number. Then $x = \frac{a}{b}$ with a, b both not zero. Consider $y = b/a$. Since x is not zero, its reciprocal is not zero. That $xy = \frac{a}{b}\frac{b}{a} = 1$.

Now we show that the reciprocal of a nonzero rational is unique. Suppose that there are two rational numbers $y = \frac{c}{d}$ and $y' = \frac{c'}{d'}$ such that $xy = 1$ and $xy' = 1$. We show that $y = y'$. Since $1 = xy = xy' = 1$ we see that $xy = xy$.

$$xy = \frac{a}{b}\frac{c}{d} = \frac{a}{b}\frac{c'}{d'} \tag{29}$$

$$acbd' = ac'bd \tag{30}$$

$$cd' = c'd \tag{31}$$

$$\frac{c}{d} = \frac{c'}{d'} \tag{32}$$

$$y = y' \tag{33}$$

Since y, y' are not distinct, the reciprocal is unique.

□

Without Loss of Generality

Sometimes, a theorem doesn't need multiple cases if the cases are all the logically similar. For example, suppose you were to prove "If x, y have opposite parity then xy is even". You don't need to split this into the two cases that x even y odd and x odd y even. Since $xy = yx$, you may simply say "without loss of generality [20], suppose x is even and y is odd". Each case is simply a relabeling of the other where you swap the names of x and y. This is a powerful proof tool, and you should be careful that you are applying it correctly.

[20] "Without loss of generality" is often appreviated as "WLOG" and pronounced *wuh-log*.

Vacuous Proof

Recall that $p \implies q$ is true when p is false or q is true. A proof of an implication is said to vacuous if it demonstrates p is false. A proof of an implication is said to be trivial if it demonstrates that q is true.

Theorem 15. *Let x be a real number. If $x^2 + 2x + 2 \leq 0$ then x^{100} is even.*

Proof. We may factor $x^2 + 2x + 2 = x^2 + 2x + 1 + 1 = (x + 1)^2 + 1$. A square is always greater than or equal to zero, so a square plus one is alwasy greater than or equal to one. Therefore, the implication is vacuously true. □

Non-constructive Proof

Suppose we want to prove an existential statement of the form $\exists x P(x)$. We may simply find a value x from its universe of discourse which satisfies the

predicate P. As it turns out, this is not necessary. You can prove something to exist *without knowing what it is*. This is called a nonconstructive proof. It demonstrates something must exist without any indication of where it is or how to find it. This is a correct proof strategy because an existential quantifier only asserts the existence of something, not that you may know specifically what it is. Again, we witness the power of proof.

Theorem 16. *Some digit of $\pi = 3.14...$ appears infinitely often.*

Proof. Suppose not. Then every digit of π appears only finitely many times. Then the decimal expansion of π must terminate, which would imply that π is a rational, contradiction. □

Observe how we used the fact that decimal numbers which terminate must be rational. Any terminating decimal of the form 1.23 may be written as $1 + \frac{23}{100}$. Next, note that this proof established that a digit of π does appear infinitely often. It didn't establish which digit, or how often, or where it appears. It simply established exactly and only what it stated. It didn't give us any method to even determine what digit appears infinitely often. This is why we may denote the proof as *non-constructive*.

Theorem 17. *There exists irrational numbers a, b such that a^b is rational.*

This result should surprise you. If a, b are irrationals, it turns out, you would be wrong to expect that a^b is also irrational. The proof should surprise you even more. It doesn't tell us for which irrational numbers a, b is the theorem true, or even one example. But it does simply assert such a pair of irrationals must exist.

Proof. Consider $\sqrt{2}^{\sqrt{2}}$. We have two cases, whether or not that $\sqrt{2}^{\sqrt{2}}$ is rational or irrational.

- Case 1: If $\sqrt{2}^{\sqrt{2}}$ is rational, then let $a = b = \sqrt{2}$ and we are done.
- If $\sqrt{2}^{\sqrt{2}}$ is irrational, then let $a = \sqrt{2}^{\sqrt{2}}$ and $b = \sqrt{2}$

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

In either case, we have asserted that there exist irrational a, b such that a^b is rational. □

For the proof, we don't even know if $\sqrt{2}^{\sqrt{2}}$ is rational or irrational! Yet in either case, we may assert the existence of irrationals a, b with the property that a^b is rational. One of the pairs $(a, b) = (\sqrt{2}, \sqrt{2})$ or $(a, b) = (\sqrt{2}^{\sqrt{2}}, \sqrt{2})$ must work. We don't know which, but we know it must be one of them.

Mathematical Writing

Now that we have seen a few proofs, let us detail exactly what is a proof, and what is required of it. A proof establishes the mathematical correctness of a theorem. It is a paragraph, and not a calculation. To write a good proof, you must employ very clear, and precise technical writing.

There are many more great resources on mathematical writing,[21] but here are few important things to keep in mind.

[21] Sections 1,2 of Mathematical Writing by Don Knuth et al, and How to Write Mathematics by Paul Halmos.

- **Avoid use of symbols and write precisely**. Recall that we proved "the product of even numbers is even". This theorem is represented as a concise, clear, english sentence. It could also be represented by the more formal proposition

$$\forall a \forall b[(Even(a)) \wedge (Even(b)) \implies (Even(ab))]$$

Although this representation is more technical, it is worse. It contains two quantifiers, three uses of a predicate, and an implication. The examples we have done so far are simple, but they may get so complicated, that it could take a page or more to represent them purely logically. A good analogy is using pseudocode to describe an algorithm, rather than using assembly. Every proof could be formalized all the way down to pure axioms of logic, but such a proof would be unreadable and extremely long.[22] We did spend a great deal of time formulating logic just so we could remove ambiguity from natural language, but a proof is prose. Your proof should be clear enough that someone else could choose to formalize it if they wanted to spend the time. Much like how a compiler may turn your readable code into unreadable machine instruction. This does not mean don't use equations or formulas, but don't *only* use equations or formulas. Don't begin a sentence with a symbol, and use complete sentences. Its okay to use symbols in your proof drafts. The reader, on skimming a proof will often only glance at equations, so make sure the remaining part of your proof is intelligible. Avoid using vague terms, like "that" or "it". Avoid the use of the pronoun "I". Even when a proof is authored by a single writer, the pronoun should always be "we", in reference to the writer and the reader. It is seductive for you to want to skip certain tedious steps and use phrases such as "clearly" and "obviously". Avoid this whenever possible.

[22] Russell and Whitehead in *Principia Mathematica* gave a set of axioms for mathematics. Its over five thousand pages long, and the proof that 1+1 = 2 takes over 350 pages. Their proof is done fully symbolically.

- **The Syntax** Proofs are commonly templated in the following manner.

Theorem XYZ. The theorem comes before the proof, and is explicitly called a theorem. The theorem can also be numbered, so reference can be made to it later (for example, "Hence by theorem XYZ we see that...")

In between a theorem and a proof, there may contain minimal commentary on the proof strategy, or the proof idea. If there is too much to say on the proof strategy, don't be afraid to repeat yourself, and restate the theorem. The point is that it is clear to the reader what is about to be demonstrated.

Proof. The beginning of the proof should be denoted in some way. A safe way is to simply begin with "Proof.". The first sentence of the proof aught to detail the proof method. If you are doing a proof by contraction, declare "Assume to the contrary...". If you are doing a proof of the contrapositive, explicitly state the contrapositive for the reader. Finally, denote the end of your proof in some way, such as in the following box. □

A proof must be clearly terminated in some way. This can be with a small box, or symbol, or by declaring QED.[23]

[23] Quod Erat Demonstrandum, meaning "that which was to be demonstrated" in Latin

- **The Flow** A proof is not a calculation, but a demonstration. In prior mathematics courses, the problems are motivated by determining a solution to something like the volume of a shape, or root of a polynomial or something.

Here, we are dealing with a totally different setting. Our goal is not to discover a solution, but to explain or convince someone of a solution; We are establishing truth.

The point of the proof is not the theorem, it is about the show. It is the journey, and not the destination. The proof is a small game between a reader and a writer. As you read and write proofs, you will wear either of these two hats. The writer must convince the reader that the theorem is true. The theorem is declared at the beginning, what is being proved is very explicit. The proof begins with its outline, and first principles. In an obvious order, the writer makes one step at a time. For each step, the writer proposes a deduction to be true, and the reader should convince themselves it is true. The steps should not be so much of a stretch to lose the reader. If you skip too many steps, you will quickly lose and upset the reader, rendering the proof incorrect. You should be doing the proof for them, and not rely on them to do too much calculation to verify the correctness of your proof. The steps of the proof should be chronological and obvious. The flow only goes one way. You should never try to begin with the theorem and work towards something. Its quite likely you will accidentally assume the theorem to be proved as a premise, making your proof incorrect.The proof should conclude with the theorem that was to be proved, but you don't assert to the reader that the theorem is true, they should be forced to conclude exactly and only what you wanted them to. *They convince themselves of its truth.* If each step of the proof is correct, then the end of the proof must be correct. The proof should conclude no sooner and no later than immediately after the truth of the theorem has been established. If your proof goes on, put this further commentary or discussion outside the proof. If your proof ends too soon, then the reader won't be able to establish the truth of the theorem for themselves.

When a mathematician wishes to prove a theorem, they write up the proof and circulate it among the community. Usually, everyone is in consensus if the proof is correct (or not), but there are two interesting case studies of when this didn't happen.

In 1976, Appel and Haken proved the four color theorem, but with a catch. There were 1,834 cases and they were only checked by a computer over thousands of hours. Initially, many argued this did not constitute a proof, because they had to trust the computers calculation. The dust has since settled, and their proof has been widely accepted, even greatly improved upon.

More recently, there has been interest in the *abc* conjecture. In over five hundred pages, Shinichi Mochizuki had developed something he calls *Inter-universal Teichmüller theory.* He estimates it would take a mathematics grad student ten years to understand his work. In another hundred pages, he used his theory to "prove" the *abc* conjecture. Despite many attempts by many people, only a few can claim to understand his work. Of those who do, some argue there is an irreparable error in corrollary 3.12. He disagrees, and accepted his own papers at a journal in which he is editor-in-chief. His proof is not widely accepted or rejected, but decisively in limbo, where it may remain for decades. Even if his proof might be correct, the root of the controversy lies in his inability to communicate his ideas effectively.

Elegance and clarity is very important for a proof. A good proof can be moving, like poetry, but very much unlike poetry, it is not open to interpretation. If mathematics is a language, the elegance factor in proofs is part of the dialect.[24] I can't tell you how to develop this skill, you can only develop it for yourself with lots of practice.

[24] Paul Erdös, the modern father of discrete mathematics, used to keep a running joke about "The Book", one which God keeps which contains the most elegant proof of each mathematical theorem. "*You need not believe in God but, as a mathematician, you should believe in The Book.*" - Paul Erdös

There are several common errors in writing proofs.

- Do not begin a proof like"Theorem XYZ is true, this is because...." It is the readers job to establish the truth of the theorem. Don't tell them its true at the beginning. The proof should be a sequence of undeniably true statements which ends with the theorem proved. More generally, don't state or reference things which are not known to be true (yet).

- While a proof in formal logic makes which axioms and premises it uses explicit, this is not necesarily true in the intuitive use of proof. Often when a proof is incorrect, the error may not be directly pointed to on the page. This is what can make hard problems hard to prove.[25] Be careful on what mathematical tools you assume you may use, and try to make their use as explicit as possible.

- This isn't so much a mistake, but a common negative trait. Many proofs by contradiction are simply a direct proof in a shell. This is especially true

[25] I can give you an example. There are maybe fourty wrong proofs a year that attempt to resolve $P \neq NP$, a very big and famous open problem in computer science. There was this (incorrect) paper. The proof relied the practical, statistical properties of SHA256, a hash function. But hash functions only *provably* have those properties if and only if $P \neq NP$. It was never stated in the proof, but by the fact the author assumed the hash function had certain properties, he had assumed $P \neq NP$, then used that to prove $P \neq NP$.

when the absurdity reached is just the negation of the base premise. It would be cleaner to simply remove the shell, and present a direct proof.

Mathematical Induction

Induction is a proof technique which can find itself useful for proving ordered or recursive properties, especially those which are universally quantified. It is not simply a way to prove things, but it is a way of thinking.

What is Induction?

Like other proof techniques, it is somewhat templated. But this template is a little more complex than the others. The reason it is a valid technique is also a little more complex.

Definition 0.0.5 (Principle of Mathematical Induction)**.** *Let* $\Phi(\cdot)$ *be any predicate over the natural numbers. The Principle of Mathematical Induction states*

$$(\Phi(0) \wedge \forall k(\Phi(k) \implies \Phi(k+1))) \implies \forall n \Phi(n)$$

Every induction proof has two parts.

- The **base case** must be proven true. For a statement over the naturals, you would usually prove $\Phi(0)$ to be true. What the base case is varies on what is being proven, and can vary wildly. Some examples we shall see will have base cases of $n = 1, 2, 3$ and so on. If you prove a base case of $n = 3$, then your statement will only be true for $n \geq 3$.

- In the **induction step**, You assume the induction hypothesis. Let k be a fixed number, and assume $\Phi(k)$ is true. This is then used to prove that $\Phi(k+1)$ must be true. You would usually try to phrase $\Phi(k+1)$ as a function of $\Phi(k)$ and other things to deduce its truth.

We will convince you induction is a valid proof technique in two ways. First, appeal to your intuition. Second, that it is equivalent to a more obvious axiom.

Intuition

"*Induction makes you feel guilty for getting something out of nothing, and it is artificial, but it is one of the greatest ideas of civilization*" - Herbert Wilf

The classic analogy is an to imagine an infinite row of dominoes. The base case can be thought of as "the first domino falls over". The induction step can be thought of as "If the k'th domino falls over, then the $(k+1)$'th domino falls over". From there, you should deduce "every domino must fall over".

Another way to think it that it proves that *there is a proof* for each n. If someone wants you to prove to them that $\Phi(6)$ is true, you can prove $\Phi(0)$ is true, then you prove $\Phi(0) \implies \Phi(1)$, then $\Phi(1) \implies \Phi(2)$, and so on until you conclude with $\Phi(5) \implies \Phi(6)$. Since you can do this for any n, then it must be the case that $\forall n \Phi(n)$ is true.

The Well-Ordering Principle

The Well-Ordering Principle is a fairly basic axiom of set theory. The axioms are usually chosen to be so simple that no one could imagine to wage an objection. They are written by lawyers. The Well-Ordering Principle is such an axiom.

Definition 0.0.6 (Well Ordering Principle)**.** *Every non-empty subset $S \subseteq \mathbb{N}$ has a least element.*

Even if your intuition does not directly convince you that induction is a valid technique, Your intuition should convince you that the Well-Ordering Principle has to be true. We prove that they are logically equivalent.

Theorem 18. *Mathematical Induction is a valid proof technique if and only if you believe the Well-Ordering Principle.*

Since we take the Well-Ordering Principle as an axiom, this proof should convince you that induction is valid.

Proof. We will prove both implications of our if and only if.
($\Longrightarrow$) Assume Mathematical Induction is true. Let $S \subseteq \mathbb{N}$ be non-empty and assume to the contrary that S has no least element. Let $\Phi(n) : \{0, ..., n\} \cap S = \emptyset$. Since 0 is the least element of $\mathbb{N}$, it must be the case that $0 \notin S$, so $\Phi(0)$ is true. Let $k \in \mathbb{N}$ and suppose $\Phi(k)$ is true. Then $\{0, ..., k\} \cap S = \emptyset$. If $k+1 \in S$ then it would be the least element of S, so we know $\{0, ..., k+1\} \cap S = \emptyset$, so $\Phi(k+1)$ must be true. By induction, since we have proven $\Phi(0)$ and $\forall k(\Phi(k) \implies \Phi(k+1))$ then it must be true that $\forall n \Phi(n)$. But if $\Phi(n)$ is true for all n, then $S = \emptyset$, contradicting the assumption that S is non-empty.
($\Longleftarrow$) Assume the Well-Ordering Principle is true, and assume to the contrary that induction is not a valid proof technique for establishing truth. Then there is some Φ such that $\Phi(0)$ is true and $\forall k(\Phi(k) \implies \Phi(k+1))$ is true but $\forall n(\Phi(n))$ is false. Let $S \subseteq \mathbb{N}$ be the set of numbers of which Φ is false for. By the Well-Ordering Principle, there is a least element we shall denote as e. Since $\Phi(0)$ is true, we know $e \geq 1$ and $e - 1 \in \mathbb{N}$. Since e is the least element which $\Phi(e)$ is false, then $\Phi(e-1)$ must be true. But by our assumption that $\forall k(\Phi(k) \implies \Phi(k+1))$, If $\Phi(e-1)$ is true, then $\Phi(e)$ would also be true. Contradiction, $\Phi(e)$ cannot be both true and false. □

The more important direction is of course, the converse. The Well-Ordering Principle is an established and intuitive axiom of set theory. If you believe it, you must believe in induction. [26]

[26] The Well-Ordering Principle does not hold over other universes of discourse, such as the integers, rationals, and reals.

Examples

The best way to learn a technique like induction is not to talk about it, but to do it. To this end, we will do many diverse examples.

Summations

Theorem 19. *Let n be a natural number greater than or equal to one. Then the sum of the first n numbers is*

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

Proof. We proceed by induction on n. Our base case is $n = 1$, and we verify that $\sum_{i=1}^{1} i = 1 = \frac{1\cdot 2}{2}$. Now assume the induction hypothesis, that

$$\sum_{i=1}^{k} i = \frac{k(k+1)}{2}$$

We prove that $\sum_{i=1}^{k+1} i = \frac{(k+1)(k+2)}{2}$.

$$\sum_{i=1}^{k+1} i = \sum_{i=1}^{k} i + (k+1)$$

By our induction hypothesis, $\sum_{i=1}^{k} i = \frac{k(k+1)}{2}$ so

$$\begin{aligned}\sum_{i=1}^{k+1} i &= \frac{k(k+1)}{2} + (k+1)\\ &= \frac{k(k+1)}{2} + \frac{2(k+1)}{2}\\ &= \frac{k(k+1)+2(k+1)}{2}\\ &= \frac{(k+1)(k+2)}{2}\end{aligned}$$

Since we have shown that $\sum_{i=1}^{k+1} i = \frac{(k+1)(k+2)}{2}$, we have proven for all $n \geq 1$ that $\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$. □

This is a classic first induction problem, but it actually need not be proven by induction.[27] Observe the proof structure. We outline the base case with $n = 1$, since $\Phi(0)$ makes no sense here. We also assume for a *specific* k that $\Phi(k)$. We don't assume $\forall n \Phi(n)$, because this is what we are trying to prove.

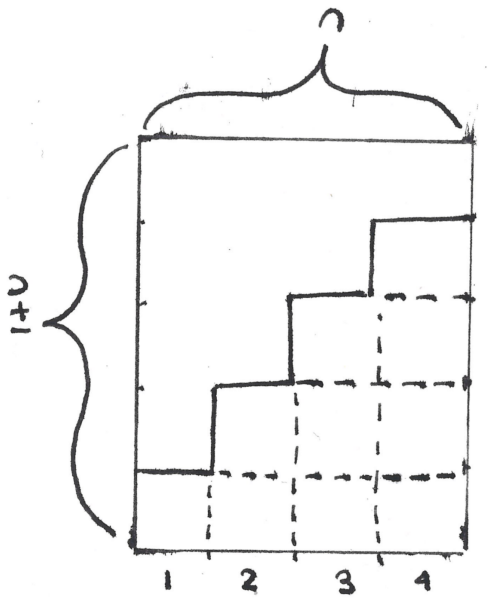
Theorem 20. *Let n be a natural number greater than or equal to one. Then the sum of the first n perfect squares is*

$$\sum_{i=1}^{n} i^2 = 1^2 + 2^2 + 3^2 + ... + n^2 = \frac{n(n+1)(2n+1)}{6}$$

Proof. We proceed by induction on n. Our base case is $n = 1$ and we verify that $\sum_{i=1}^{1} i^2 = 1^2 = 1 = \frac{1(2)(3)}{6}$. Now assume the induction hypothesis, that

$$\sum_{i=1}^{k} i^2 = \frac{k(k+1)(2k+1)}{6}$$

[27] This formula by multiplication was known to Diophantus in around 120BC. It was probably known to Pythagoras even earlier. They most likely had observed the following visual "proof":



Twice the summation is the area of a rectangle of dimensions $n \times (n+1)$.

A second proof, likely a folktale, credits Carl Friedrich Gauss: "*In the 1780s a provincial German schoolmaster gave his class the tedious assignment of summing the first 100 integers. The teacher's aim was to keep the kids quiet for half an hour, but one young pupil almost immediately produced an answer: 1 + 2 + 3 + ... + 98 + 99 + 100 = 5,050. The smart aleck was Carl Friedrich Gauss, who would go on to join the short list of candidates for greatest mathematician ever. Gauss was not a calculating prodigy who added up all those numbers in his head. He had a deeper insight: If you "fold" the series of numbers in the middle and add them in pairs,1 + 100, 2 + 99, 3 + 98, and so on, all the pairs sum to 101. There are 50 such pairs, and so the grand total is simply* 50×101. *The more general formula, for a list of consecutive numbers from 1 through n, is* $n(n+1)/2$. " - Brian Hayes in *Gauss's Day of Reckoning*

We prove that $\sum_{i=1}^{k+1} i^2 = \frac{(k+1)(k+2)(2k+3)}{6}$.

$$\sum_{i=1}^{k+1} i^2 = \sum_{i=1}^{k} i^2 + (k+1)^2$$

By our induction hypothesis, $\sum_{i=1}^{k} i^2 = \frac{k(k+1)(2k+1)}{6}$ so

$$\begin{aligned}
\sum_{i=1}^{k+1} i^2 &= \frac{k(k+1)(2k+1)}{6} + (k+1)^2 \\
&= \frac{k(k+1)(2k+1)}{6} + \frac{6(k+1)^2}{6} \\
&= \frac{k(k+1)(2k+1) + 6(k+1)^2}{6} \\
&= \frac{(k+1)(k(2k+1) + 6(k+1))}{6} \\
&= \frac{(k+1)(2k^2 + k + 6k + 6)}{6} \\
&= \frac{(k+1)(2k^2 + 7k + 6)}{6} \\
&= \frac{(k+1)(k+2)(2k+3)}{6}
\end{aligned}$$

Since we have shown that $\sum_{i=1}^{k+1} i^2 = \frac{(k+1)(k+2)(2k+3)}{6}$, we have proven for $n \geq 1$ that $\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}$. □

Lets do another summation example.

Theorem 21. *Let n be a natural number greater than or equal to one. Then the sum of the first n cubes is equal to the sum of the first n numbers, squared.*

$$\sum_{i=1}^{n} i^3 = \left(\sum_{i=1}^{n} i\right)^2$$

Proof. We proceed by induction. By a previous theorem, we have a closed form for the sum of the first n numbers. It is sufficient then for us to show that the sum of the first n cubes is

$$\left(\sum_{i=1}^{n} i\right)^2 = \left(\frac{n(n+1)}{2}\right)^2 = \frac{n^2(n+1)^2}{4}$$

Our base case is $n = 1$ and we verify that $\sum_{i=1}^{1} i^3 = 1^3 = 1 = \frac{1^2(2)^2}{4}$. Now assume the induction hypothesis, that

$$\sum_{i=1}^{k} i^3 = \frac{k^2(k+1)^2}{4}$$

We prove that $\sum_{i=1}^{k+1} i^3 = \frac{(k+1)^2(k+2)^2}{4}$.

$$\sum_{i=1}^{k+1} i^3 = \sum_{i=1}^{k} i^3 + (k+1)^3$$

By our induction hypothesis, $\sum_{i=1}^{k} i^3 = \frac{k^2(k+1)^2}{4}$ so

$$\begin{aligned}
\sum_{i=1}^{k+1} i^3 &= \frac{k^2(k+1)^2}{4} + (k+1)^3 \\
&= \frac{k^2(k+1)^2}{4} + \frac{4(k+1)^3}{4} \\
&= \frac{k^2(k+1)^2 + 4(k+1)^3}{4} \\
&= \frac{(k+1)^2(k^2 + 4(k+1))}{4} \\
&= \frac{(k+1)^2(k^2 + 4k + 4)}{4} \\
&= \frac{(k+1)^2(k+2)^2}{4}
\end{aligned}$$

Since we have shown that $\sum_{i=1}^{k+1} i^3 = \frac{(k+1)^2(k+2)^2}{4}$, we have proven for $n \geq 1$ that $\sum_{i=1}^{n} i^3 = \left(\sum_{i=1}^{n} i\right)^2$. □

Recursive Algorithms

A proof of correctness of an algorithm is a proof that what the code outputs is exactly what the mathematical problem it claims to solve is. For example, consider the following recursive pseudocode on input n to print the factorial $n! = n \cdot (n-1) \cdot ... \cdot 2 \cdot 1$.

```
def factorial(n):
if n == 0: return 1
else return n * factorial(n-1)
```

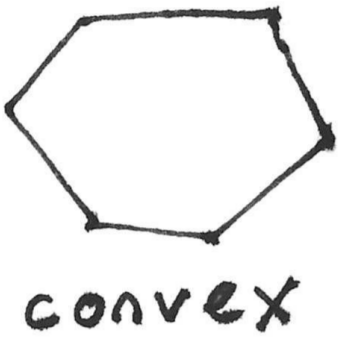
A lot of algorithms are clear, and their proofs of correctness are trivial. Recursive algorithms pose an exception to this. Induction is a technique that works extremely well for proving the correctness of recursive algorithms, because if the algorithm only calls itself on smaller inputs, you may assume those are correct by induction.

Theorem 22. *Let n be a natural number. The program on input n correctly returns $n!$.*

Proof. We proceed by induction on n. Our base case is $n = 0$, and we see that the algorithm does correctly return 1. Assume the induction hypothesis, that on input k that our algorithm outputs $k!$. We argue on input $k+1$ that our algorithm outputs $(k+1)!$. Consider the algorithm on input $k+1$. Notice the recursive call parameter is $k + 1 - 1 = k$. The recursive call made is then `factorial(k)`. By the induction hypothesis, this correctly returns $k!$. Our program then returns $(k+1) \cdot k! = (k+1)!$ as desired. □

Convex Polygons

Definition 0.0.7 (Convexity of a Polygon)**.** *A polygon is convex if none of its interior angles have degree greater than 180 degrees.*



More generally, a shape is convex if given any two points in the interior, the shortest distance between them is also contained within the shape.

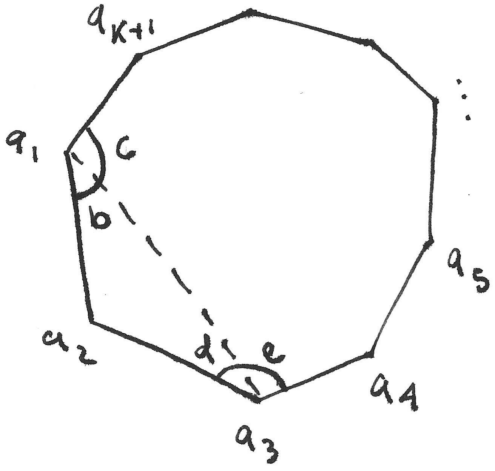
Theorem 23. *Given a convex n-gon with $n \geq 3$, the interior angle sum is equal to $(n-2) \cdot 180$ degrees.*

Proof. We proceed by induction on n. Our base case is that $n = 3$, which is a triangle. It is well known that a triangle has interior angle sum of $(3-2) \cdot 180 = 180$ degrees.[28] Now assume our induction hypothesis, that any convex k-gon has interior angle sum of $(k-2) \cdot 180$ degrees. Consider any convex $k+1$-gon. We will argue it has interior angle sum of $(k-1) \cdot (180)$ degrees. Let the interior angles of our convex $k+1$-gon be denoted as $a_1, ..., a_{k+1}$. We wish to compute $a_1 + a_2 + a_3 + ... + a_{k+1}$. Draw a line segment between a_1, a_3. This will partition the convex $k+1$-gon into a convex k-gon and a triangle. Our newly formed line segment divides angles a_1 and a_3. Let these splits be denoted as $a_1 = b + c$ and $a_3 = d + e$. Notice that a_2, b, d are the interior angles of our triangle, so $a_2 + b + d = 180$. Similarly, notice $c, e, a_4, ..., a_{k+1}$ are the interior angle sum of our convex k-gon. By the induction hypothesis, the interior angle sum of our convex k-gon is $(k-2) \cdot 180$ degrees, so the interior angle sum of our convex $k+1$-gon is

$$\begin{aligned} a_1 + a_2 + a_3 + a_4 + ... + a_{k+1} = \\ (b+c) + a_2 + (d+e) + a_4 + ... + a_{k+1} = \\ (a_2 + b + d) + (c + e + a_4 + ... + a_{k+1}) = \\ 180 + ((k-2) \cdot 180) = \\ (k-1) \cdot 180 \end{aligned}$$

By induction, we may now conclude that the interior angle sum of a convex n-gon is $(n-2) \cdot 180$ degrees. □

[28] This can be proven from the Axioms of Euclidean geometry, and is known to be equivalent to Euclid's fifth postulate.



Tilings

A triominoe is a tetris piece of three unit squares in the shape of an L. We may say a board is "tiled by triominoes" if there exists a way to place only triominoes to cover every space of the board and no two pieces overlap.

Theorem 24. *Let there be a board of dimensions $2^n \times 2^n$ with one quadrant removed. For all n, any such board can be tiled by triominoes.*

Proof. We proceed by induction on n. Our base case is $n = 1$. Our board is a 2×2 square with one square removed, and may be tiled perfectly by one triomino. Now assume that a board of dimension $2^k \times 2^k$ with one quadrant removed may be tiled by triominoes. We will prove there is a tiling of a board of dimension $2^{k+1} \times 2^{k+1}$ with one quadrant removed.

Consider a board of dimension $2^{k+1} \times 2^{k+1}$ with one quadrant removed. We may split this board into four disjoint areas as per our diagram. Observe that each area is exactly a board of dimension $2^k \times 2^k$ with one quadrant removed. By the induction hypothesis, each of these four areas has a tiling, so we may

simply compose our tilings to create one larger tiling. Then this is a tiling of a board of dimension $2^{k+1} \times 2^{k+1}$ with one quadrant removed. We have proven for all n that a board of dimensions $2^n \times 2^n$ with one quadrant removed can be tiled by triominoes. □

We don't really understand theorems unless we can understand their proofs. The proofs explain why the theorems should be true. Here, the proof gives us a little bit of explanatory power. Suppose someone requested you tile a 16×16 board with one quadrant removed using 4^3 triominoes. Solving just this one problem may be quite tedious and annoying. The proof by induction solves infinitely many boards, and from the proof, we may extract specific answers. Again, we witness the power of proof, and induction specifically. One problem may be hard, infinitely many problems may be easy. The proof motivates the solution.

Theorem 25. *Let there be a board of dimensions $2^n \times 2^n$ with any one square removed. For all n, any such board can be tiled by triominoes.*

Proof. We proceed by induction on n. Our base case is $n = 1$. Our board is a 2×2 square with one square removed, and may be tiled perfectly by one triomino. Now assume that a board of dimension $2^k \times 2^k$ with any possible one square removed may be tiled by triominoes. We will prove there is a tiling of a board of dimension $2^{k+1} \times 2^{k+1}$ with any one square removed.

Consider a board of dimension $2^{k+1} \times 2^{k+1}$ with some one square removed. We may split this board into four disjoint areas as per our diagram. This missing square must fall into one of the four disjoint areas. By the induction hypothesis, the quadrant with this missing square is a board of dimension $2^k \times 2^k$ with one square removed, and thus there exists a tiling of it. The remaining three quadrants are a board of dimension $2^k \times 2^k$ with one quadrant removed, and thus by our previous theorem, there exists a tiling of it. We may compose these tilings to tile the entire board of dimension $2^{k+1} \times 2^{k+1}$ with some one square removed. We have proven for all n that a board of dimensions $2^n \times 2^n$ with any one square removed can be tiled by triominoes. □

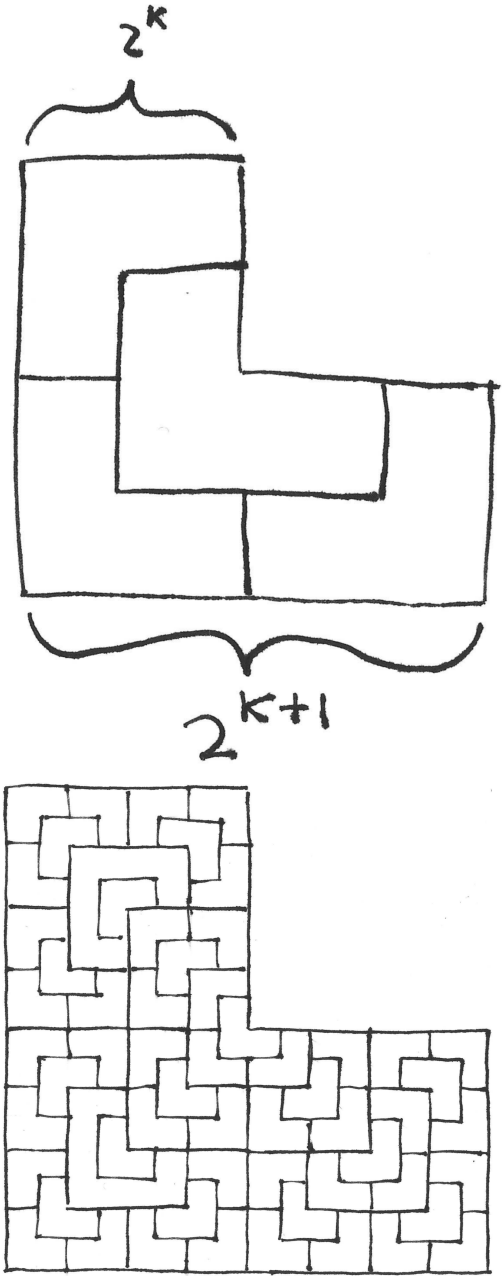
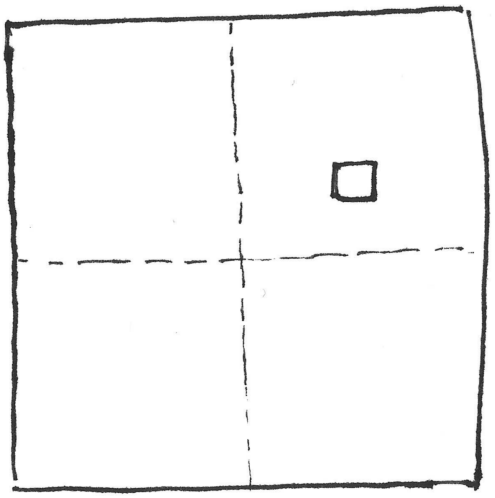


Figure 6: a 16×16 board with one quadrant removed tiled by 64 triominoes



Strong Induction

Definition 0.0.8 (Strong Principle of Mathematical Induction)**.** *Let $\Phi(\cdot)$ be any predicate over the natural numbers. The Strong Principle of Mathematical Induction states*

$$(\Phi(0) \wedge \forall k((\Phi(0) \wedge ... \wedge \Phi(k)) \implies \Phi(k+1))) \implies \forall n \Phi(n)$$

Strong induction is like a bigger hammer. In our induction step, instead of assuming an induction hypothesis $\Phi(k)$ and using only that to conclude $\Phi(k+1)$, with strong induction, you get a strong induction hypothesis. For i a natural number with $0 \leq i \leq k$ you may assume that $\Phi(i)$ is true. If you think of induction like one domino knocking over the next, strong induction is analogously several dominoes in the past knocking over the next. Strong induction may be necessary when the dominoes are not arraigned in a straight line, and the ninth domino may require force from the third, fifth, and eighth. Let us proceed with some examples. We emphasize the diversity of ways you can apply strong induction.

Theorem 26. *Let $\{a_n\}$ be a recursively defined sequence. Let $a_0 = 0, a_1 = 1$, and for $n \geq 2$ let*

$$a_n = 2a_{n-1} - a_{n-2} + 2$$

We prove for all natural numbers n that

$$a_n = n^2$$

Proof. We proceed by strong induction on n. We verify our base cases $n = 0, 1$ with $0^2 = 0 = a_0$ and $1^2 = 1 = a_1$. Now assume our induction hypothesis. For all natural numbers i with $0 \leq i \leq k$ that $a_i = i^2$. We prove that $a_{k+1} = (k+1)^2$.

$$\begin{aligned}
a_{k+1} &= 2a_k - a_{k-1} + 2 \\
&\quad \text{By our strong inductive hypothesis, } a_k = k^2 \text{ and } a_{k-1} = (k-1)^2 \\
a_{k+1} &= 2(k^2) - (k-1)^2 + 2 \\
a_{k+1} &= 2k^2 - (k^2 - 2k + 1) + 2 \\
a_{k+1} &= k^2 + 2k + 1 \\
a_{k+1} &= (k+1)^2
\end{aligned}$$

as desired.

□

Theorem 27 (Fundamental Theorem of Arithmetic)**.** *Every positive number $n \geq 2$ has a unique prime factorization.*

To prove uniqueness, we need to first show that every number $n \geq 2$ can be written as a product of primes. Then we must show that this factorization is unique.

Proof. We first prove that every number may be written as a product of primes. We proceed by induction on n. Our base case is $n = 2$. As 2 is prime, n may be written as a product of primes, and the base case is complete. Now assume for all natural numbers i with $2 \leq i \leq k$ that i can be written as a product of primes. We show $k + 1$ can be written as a product of primes. We have two cases.

- If $k + 1$ is prime, then we are done, as every prime number is a product of primes.
- If $k + 1$ is not prime, then k is composite, so there exists a, b such that $k + 1 = ab$ with $a \neq 1$ and $b \neq 1$. We see this implies that $a < k + 1$ and $b < k + 1$. By our strong induction hypothesis, a, b may each be written as a product of primes, so $ab = k + 1$ is a product of products of primes.

Next we will show that the prime factorizations must be unique. Let n be a number with $n \geq 2$, and suppose to the contrary that n has two distinct prime factorizations. Then there exists primes $p_1, ..., p_k, q_1, ..., q_k$ such that $n = p_1 \cdot ... \cdot p_k$ and $n = q_1 \cdot ... \cdot q_l$. Since these prime factorizations must differ by assumption, there is a p_i which does not equal any q_j. Observe that since $p_i \mid n$, then $p_i \mid (q_1 \cdot ... \cdot q_n)$. So there must exist some j such that $p_i \mid q_j$. Since q_j is prime, the only numbers which may divide into it are one and itself, and since p_i is not one, it must be the case that $p_i = q_j$, contradiction. □

Fibonacci Numbers

Theorem 28. *Let n be a natural number. Then the nth Fibonacci number F_n is strictly less than 2^n.*

Proof. We proceed by induction on n. Our base case is $n = 0$, and we see that $F_0 = 0 < 1 = 2^0$. Now assume our strong induction hypothesis: For all natural numbers i with $0 \leq i \leq k$ that $F_k < 2^k$. We prove that $F_{k+1} < 2^{k+1}$.

$$
\begin{aligned}
&F_{k+1} = F_k + F_{k-1} \\
&\qquad \text{By our strong inductive hypothesis, } F_k < 2^k \text{ and } F_{k-1} < 2^{k-1} \\
&F_{k+1} < 2^k + 2^{k-1} \\
&F_{k+1} < 2^{k-1}(2+1) \\
&F_{k+1} < 2^{k-1}(2+2) \\
&F_{k+1} < 2^{k-1}(4) \\
&F_{k+1} < 2^{k+1}
\end{aligned}
$$

We may then conclude for all natural numbers n that $F_n < 2^n$. □

A proof by induction is very much unlike a calculation, in which you are trying to determine the answer, usually a quantity. In order to do a proof by induction, you must have something to take as an induction hypothesis. Suppose we wish to know a closed formula for the Fibonacci numbers. We start with an upper bound, and try to prove it. Then we examine the proof, and see how much "slack" we have in our overestimate. Could a proof by induction succeed to show $F_n < 2^{0.9n}$? What about $F_n < 2^{0.5n}$? What about $F_n < 2^{n/2}$? The first two proofs will succeed, the third will fail. Just because a proof of a theorem fails does not imply that its negation must be true. But it can guide your intuition, with truth discovery as a procedure.

We have shown previously you can use strong induction to prove closed formulas of recursively defined sequences. Today we prove a closed formula of one of the most popular recurrences, the Fibonacci numbers. Defined via the recurrence

$$F_n = F_{n-1} + F_{n-2}$$

with two base cases $F_0 = 0, F_1 = 1$. You may surprised to know that it has a closed form. The reason you've never been taught the closed form is because it is very complicated looking.

Theorem 29 (Binet's Formula)**.** *Let n be a natural number and F_n the n'th Fibonacci number. Then*

$$F_n = \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^n - \frac{1}{\sqrt{5}}\left(\frac{1-\sqrt{5}}{2}\right)^n \tag{34}$$

Before we prove the correctness of this formula, we will need a small lemma.

Lemma 30. *For $\varphi = \frac{1+\sqrt{5}}{2}$ and $\psi = \frac{1-\sqrt{5}}{2}$ These two numbers satisfy $\varphi^2 = \varphi + 1$ and $\psi^2 = \psi + 1$*

Proof. Notice that φ, ψ are the two roots of $x^2 - x - 1 = 0$ by the quadratic formula, and thus satisfy $x^2 = x + 1$. □

Proof. We have two bases for $n = 0, 1$ which we now verify. We want to prove $F_0 = 0$ and $F_1 = 1$.

$$F_0 = \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^0 - \frac{1}{\sqrt{5}}\left(\frac{1-\sqrt{5}}{2}\right)^0 = \frac{1}{\sqrt{5}} - \frac{1}{\sqrt{5}} = 0$$

$$F_1 = \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^1 - \frac{1}{\sqrt{5}}\left(\frac{1-\sqrt{5}}{2}\right)^1 = \frac{1+\sqrt{5}-1+\sqrt{5}}{2\sqrt{5}} = \frac{2\sqrt{5}}{2\sqrt{5}} = 1$$

Now assume by strong induction that the formula is correct for $0, 1, 2, ..., k$. We prove that it is true for $k+1$. We prove that $F_{k+1} = \frac{1}{\sqrt{5}}\varphi^{k+1} - \frac{1}{\sqrt{5}}\psi^{k+1}$

$$\begin{aligned}
F_{k+1} = F_k + F_{k-1} = \text{by our strong induction hypothesis} \\
\left(\frac{1}{\sqrt{5}}\varphi^k - \frac{1}{\sqrt{5}}\psi^k\right) + \left(\frac{1}{\sqrt{5}}\varphi^{k-1} - \frac{1}{\sqrt{5}}\psi^{k-1}\right) = \\
\left(\frac{1}{\sqrt{5}}\varphi^k + \frac{1}{\sqrt{5}}\varphi^{k-1}\right) + \left(-\frac{1}{\sqrt{5}}\psi^k - \frac{1}{\sqrt{5}}\psi^{k-1}\right) = \\
\frac{1}{\sqrt{5}}\varphi^{k-1}(\varphi+1) - \frac{1}{\sqrt{5}}\psi^{k-1}(\psi+1) = \\
\frac{1}{\sqrt{5}}\varphi^{k-1}(\varphi^2) - \frac{1}{\sqrt{5}}\psi^{k-1}(\psi^2) = \\
\frac{1}{\sqrt{5}}\varphi^{k+1} - \frac{1}{\sqrt{5}}\psi^{k+1} =
\end{aligned}$$

We have therefore proved that

$$F_n = \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^n - \frac{1}{\sqrt{5}}\left(\frac{1-\sqrt{5}}{2}\right)^n$$

□

This is the power of proof by induction, proof in general. We have absolute certainty of the correctness of the formula. It doesn't matter if we know how to get the formula, we can prove that we know for certain it is right. This is a very weird formula, yet we know it is correct, even if we can't explain where it comes from![29]

[29] Proving the correctness of this formula is one thing, but determining what this formula was first is also not too hard, if you know a little linear algebra.

Structural Induction

Induction is a really powerful technique. We have so far applied it obviously over sequences denoted by the natural numbers, but you can apply it in many other cases where induction might not be your first guess. You can perform induction, structurally. You just need to find a way to induct over the objects you want.

Propositional Logic

Theorem 31. *The truth value of a proposition is only dependent upon the propositional variables which appear in it.*

In order to prove such a thing by induction, we need a more formal definition of what a proposition is.

Definition 0.0.9. *A proposition is defined recursively as:*

- *Propositional variables are propositions* $p_1, p_2, p_3, ...$
- *If* B *is a proposition then* $A \equiv (\neg B)$ *is a proposition.*
- *If* B, C *are propositions then* $A \equiv (B \wedge C)$ *is a proposition.*
- *If* B, C *are propositions then* $A \equiv (B \vee C)$ *is a proposition.*
- *noting else is a proposition*

This makes propositions "well-formed" for us to do math on them. Recall we do not need implications or biconditionals, as we may represent these with and, or, and not still. Formalizing the propositions to only look like those which are well formed allows us to define the complexity of a proposition as follows.

Definition 0.0.10. *The complexity of a proposition is the the maximum number of steps 1-4 which must be applied in order to write it. Let* A *be a proposition, and let* $c(A)$ *denote the complexity of the proprosition.*

- *If* A *is a propositional variable then* $c(A) = 0$
- *If* A *is a proposition of the form* $A \equiv (B \wedge C)$ *or* $A \equiv (B \vee C)$ *then* $c(A) = \max(c(B), c(C)) + 1$

Now that we have something nice and natural number like, the Well-Ordering Principle applies, and we have something to induct over. Now we proceed with proving that the truth of a proposition is dependent only on the propositional variables which appear in it.

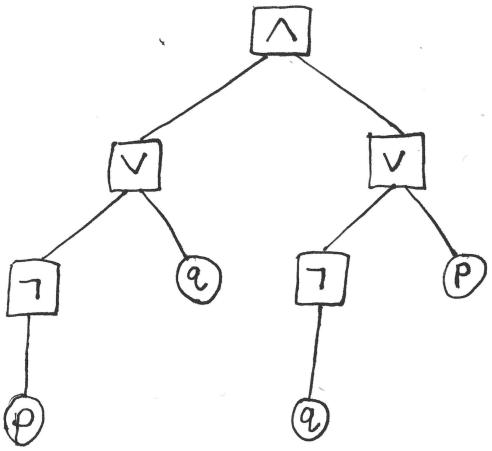


Figure 7: The complexity of a proposition can be thought of as the maximum depth of the tree to parse it. Since $p \iff q$ may be written well-formed as $(((\neg p) \vee q) \wedge ((\neg q) \vee p))$, the complexity of this formula is 3.

Proof. We induct on the complexity of the propositions. Our base case is when a proposition has complexity zero, and is only a propositional variable. A proposition which is only a propositional variable is true when its propositional variable is true, and is false when its propositional variable is false, so its truth value is only dependent on its propsitional variable. Assume the induction hypothesis, any proposition of complexity k has its truth value dependent only its propositional variables. Let A be a proposition such that $c(A) = k + 1$. We have a few cases:

- Case 1: If the last step in construction of A was a negation. Then there is a proposition B with $c(B) = k$ such that $A \equiv \neg B$. By the induction hypothesis, B has its truth only determined by its propositional variables. By our definition of negation, A is true when B is false and A is false, when B is true, so the truth of A is entirely determined by the truth of B. Since the propositional variables of A are the propositional variables of B. The truth value of A is determined only by its own propositional variables.
- Case 2: If the last step in the construction of A was a conjunction. Then there are propositions B, C with $c(B) \leq k$ and $c(C) \leq k$ such that $A \equiv (B \wedge C)$. By our strong induction hypothesis, the propositions B, C have their truth value determined only by their propositional variables. By definition of conjunction, A is true when B and C is true, and A is false when B or C is false. Therefore, the truth of A is entirely determined by the truth of B and C. Since the propositional variables of A are the propositional variables of B and C. The truth value of A is determined only by its own propositional variables.

- Case 3: If the last step in the construction of A was a disjunction. Then there are propositions B, C with $c(B) \leq k$ and $c(C) \leq k$ such that $A \equiv (B \vee C)$. By our strong induction hypothesis, B, C have their truth value determined only by their propositional variables. By definition of disjunction, A is true when B or C is true, and A is false when B is false and C is false. Therefore, the truth of A is entirely determined by the truth of B and C. Since the propositional variables of A are the propositional variables of B and C. The truth value of A is determined only by its own propositional variables.

In all cases, the truth value of A is entirely determined by its propositional variables. □

Binary Trees

Lets do just one more simple example

Definition 0.0.11. *A binary tree is full if every node has two or zero children*

Definition 0.0.12. *A binary tree is complete if all leaves are at the same level.*

A binary tree is full and complete if it looks like the textbook picture of a generic binary tree, all leaves at the same, last level and no leaves are missing.

Theorem 32. *A full complete binary tree of depth k has 2^k leaves*

Although this can be proved by a simple counting argument, for demonstration, we proof it by induction

Proof. We proceed by induction on the depth of the binary tree. Our base case is on binary trees of depth zero. A depth zero full and complete binary tree has one leaf, which is also the root. We see that $2^0 = 1$ and the base case is proved.

Now assume it is true for binary trees of depth $k-1$. We prove it is true for binary trees of depth k. Consider a full and complete binary tree of depth k. If you delete the root, you have two full and complete binary trees of depth $k-1$.

Notice that we haven't touched a leaf, so the number of leaves are the same. By the inductive hypothesis, These two binary trees of depth $k-1$ have 2^{k-1} leaves each, so the number of leaves of our depth k tree was $2^{k-1} + 2^{k-1} = 2(2^{k-1}) = 2^k$. □

Some Induction Mistakes

We consider a wrong example of a proof by induction.

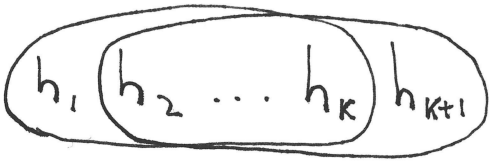
Theorem 33 (Not a real theorem)**.** *All horses are the same color*

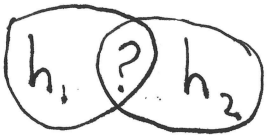
Not a Proof. We proceed by induction on the number of horses. Our base case is when there is one horse. Certainly, this one horse is the same color as itself. Now assume the induction hypothesis, any collection of k horses must all have the same color. Consider a collection of $k+1$ horses, and suppose they are denoted like $h_1, h_2, ..., h_k, h_{k+1}$. Consider the group of horses $h_1, ..., h_k$. By our induction hypothesis, all these horses have the same color as each other. Consider the group of horses $h_2, ..., h_{k+1}$. By our induction hypothesis, all these

horses have the same color. Since h_2 has the same color as all of $h_1, ..., h_k$, and also has the same color as $h_2, ..., h_{k+1}$, then $h_1, ..., h_{k+1}$ all have the same color. □

The error is subtle. Before you read on to what the bug in the proof is, try to find it for yourself.

The proof actually fails for $n = 2$. By splitting $h_1, ..., h_{k+1}$ into two collections of size k, we impicitly assume there was overlap, but there is no overlap when $n = 2$. Often times, the errors which make a proof incorrect will not be spoken aloud to be pointed at, but will be hidden by the presentation choices the writer has made. The assumption this overlap exists is incorrect, but it was also never stated.





Both the base cases and induction step must be present for the proof to be correct. Consider the following incorrect proof.

Theorem 34 (Also not a real theorem)**.** *Let n be a natural number. Then $n(n+1)$ is odd.*

Also Not a Proof. Our base case $n = 0$ is trivial so assume the induction hypothesis, that $k(k+1)$ is odd. So there exists an l such that $k(k+1) = 2l+1$. We prove $(k+1)(k+2)$ is odd.

$$(k+1)(k+2) = k(k+1) + 2(k+2) \overset{IH}{=} (2l+1) + 2(k+2) = 2(l+k+2) + 1$$

Since we may write $(k+1)(k+2)$ as two times a number plus one, it is odd. We have proven for all n that $n(n+1)$ is odd □

The error lies in the fact the base case is obviously false! We were still able to correctly prove the induction hypothesis, but it was vacuous anyway, so it doesn't even matter.

Writing Proofs by Induction

Writing base cases is usually uninteresting and procedural, but please do not skip them. Do not be tempted to declare it to be trivial and unworthy of your words. As you have seen, if you have a successful induction step, but no base case, you have nothing. In proofs by strong induction, determining the number of necessary base cases can be nontrivial. The induction hypothesis should "kick in" right after the last base case.

Proof by induction only applies to infinite universes of discourse in which the infinite goes "one way". The Well-Ordering Principle should give you this intuition. The integers, rationals, and reals do not have a Well-Ordering Principle. You can still perform induction on theorems over these universes of discourse, but it must be in a well-ordered way. Induction can be used to prove limits and convergence, theorems about taylor series of functions and more. Induction could in theory be done over the integers if done twice in both directions.

When to choose between induction and strong induction is a skill to develop. The fibonacci and recurrence problems should obviously jump out that you need strong induction, because the problem is naturally defined not from only the previous term, from from the last two previous terms. In the rare case you can apply both induction and strong induction, it is usually more elegant to use the simple form of induction. Consider the convex polygon example. We used

simple induction, but we could have actually done strong induction. Instead of decomposing a $k+1$-gon into a k-gon and a triangle, we could have decomposed it into a $(k-m+2)$-gon and an m-gon, for any $3 \leq m \leq k-1$. By a strong induction hypothesis, these would both be convex and have the required interior angle sums, but the arithmetic following would be extremely ugly and tedious, as well as the explanation of correctness.

When writing a proof by induction, the template is a little more specific and other proof strategies. Here is some well rounded advice:

- You may denote the base cases and induction step separately. Do the base case before you do the inductive step.
- Specify that you are doing induction, and if you are doing strong induction. This can be done by beginning the proof with "We proceed by induction on n" or "We proceed by strong induction on n".
- Specify what you are actually inducting over. Sometimes a problems will have many variables, and you should specify in which direction is the induction going. It won't always be over the variable denoted by n. Sometimes, its the number of steps of a computation, or dimensions of space, or complexity of a formula.
- Be careful, especially in strong induction, that you do not assume as your induction hypothesis that $\forall k \Phi(k)$. This is what you are trying to prove. Your induction hypothesis is assumed *for some* k. To emphasize this, in all our examples, we have inducted over the variable k, and left n to be used in the last statement.
- In the induction step, when demonstrating $\Phi(k) \implies \Phi(k+1)$, explicitly say what the induction hypothesis $\Phi(k)$ actually is in english. Then explicitly say what $\Phi(k+1)$ is, and that you will prove it.
- It is impolite to end a proof on a calculation. Most other kinds of proofs won't have this problem but a lot of induction proofs do end this way. What you can do is simply repeat the theorem for the reader, perhaps like "Therefore we have shown $\forall n \Phi(n)$ for whatever $\Phi(n)$ is in english.

Set Theory

There exists many kinds of mathematical structures, numbers are a common example. Today we will describe sets. Sets are so essential, all other mathematical theories can be derived from sets.

Basic Set Notation

Sets are simply a collection of other mathematical objects from some understood universe of discourse Ω. If the universe of discourse is itself sets, then sets may contain other sets even. We denote the beginning and ending of a set by "{" and "}". An example of a set is $\{2, 3\}$. This set contains two *elements*, which are the numbers 2, 3. We may write $2 \in \{2, 3\}$ to mean "two is an element of the set". Here $\in$ is read as "is an element of" or simply "in".
[30]

Sets do not contain duplicates of elements. $\{2, 2, 3\}$ is not a set, but a multi set, which we will not discuss. A set is also not ordered or even necessarily orderable. A computer is a highly structured object, but a set is simply an association between objects.

Sometimes we use the ellipses to denote continuation of a pattern. For example $\{1, 2, 3, ..., 99, 100\}$. Here, this set has 100 elements, but we describe it simpler. As another example, $\{2, 4, 6, ..., 98, 100\}$ has only the even numbers between 2 and 100 inclusive. Note that we don't have to have a termination, and sets are allowed to be infinite. We describe $\mathbb{N}$ as $0, 1, 2, ...$ but it really is a set $\{0, 1, 2, 3, ...\}$. Lets define a few large sets:

- $\mathbb{N} = \{0, 1, 2, ...\}$
- $\mathbb{Z} = \{..., -2, -1, 0, 1, 2, ...\}$
- $\mathbb{N}^+ = \mathbb{Z}^+ = \{1, 2, 3, 4, ...\}$

These are some common universes of discourse, and in fact, every universe of discourse is a set. Previously, when we wrote

$$\forall x \exists y [\Phi(x, y)]$$

we may specify what possible values x, y may quantify over using this set notation as

$$\forall x \in X \ \exists y \in Y [\Phi(x, y)]$$

where X, Y are some sets.

[30] Today we understand $x \in A$ to mean that "x is an element of the set A", as in A is some container or collection, and x is "in" this box. The original history of this symbol ($\in$) is a stylized e from the greek word $\grave{\varepsilon}\sigma\tau\acute{\iota}$, which most literally translates to "is". The original meaning of $x \in A$ was not that "x in A", rather "x is A". It was much more like a type declaration in a programming language! We use set theory much more than to declare universes of discourses of variables now.

Figure 8: A set which contains two elements, a cat, and a set containing a cat.

Set Builder Notation

We were able to define $\mathbb{N}, \mathbb{Z}$ so far, but why not $\mathbb{Q}$? Turns out it is not so easy to describe the rationals by listing out a few elements. Which one comes first? What sequence could possibily enumerate all of them? We instead take a different strategy.

Definition 0.0.13 (Axiom of Comprehension)**.** *Let Φ be any predicate and let Ω be any universe of discourse. Then*

$$\{x \in \Omega \mid \Phi(x)\ \}$$

is a set.

Where $\Phi(x)$ is a predicate to enforce conditions on x. This is called set builder notation. Instead of listing elements in some possibily ambiguous pattern, we define a set to contain exactly and only the elements which satisfy some predicate. Given this, we can easily define the rationals:

$$\mathbb{Q} = \left\{ \frac{a}{b} \mid a, b \in \mathbb{Z} \text{ and } b \neq 0 \right\}$$

The left, before the bar may be understood as the syntax and representation of objects in this set, as well as the universe of discourse. Right of the bar are the conditions to define this set. Here are some other examples of set builder notation.

- $\{n \in \mathbb{N} \mid 1 < n < 10\} = \{2, ..., 9\}$
- $\{n \in \mathbb{Z} \mid n \geq 0\} = \mathbb{N}$
- $\{n \mid n \in \mathbb{N} \text{ or } -n \in \mathbb{N}\} = \mathbb{Z}$
- $\mathbb{R}$ You can suppose that the definition of a real number is one with any possible decimal expansion. [31]
- $\mathbb{C} = \{a + bi \mid a, b \in \mathbb{R} \text{ and } i^2 = -1\}$

[31] Construction of the reals from axiomatic set theory is very difficult, but technically possible. It is far out of scope of what we want to study, which is the discrete. You could try to form a restriction of the complex numbers like $\mathbb{R} = \{a + bi \in \mathbb{C} \mid b = 0\}$. Unfortunately, this would be circular, as the complex numbers are defined as an extension of the reals.

[32] This notation was coined by Bourbaki, a secretive pseudoanonymous collective of mathematicians who have influenced much of the way mathematics is presented today. They had high standards for proof and rigor, and avoided illustrations.

Empty Set

There exists a set with no elements, we write it as $\{\}$ or $\emptyset$.[32] This is called the empty set. The empty set is not defined, but you can use comprehension with a predicate which is unsatisfiable to construct it. Let $\Phi(x) = \neg(x = x)$. Then $\{x \in \Omega \mid \neg(x = x)\} = \emptyset$.

Intervals

You may be familiar with interval notation in calculus, like $(a, b]$. We can define these using set builder notation

- $(a, b) = \{x \in \mathbb{R} \mid a < x < b\}$
- $[a, b) = \{x \in \mathbb{R} \mid a \leq x < b\}$
- $(a, b] = \{x \in \mathbb{R} \mid a < x \leq b\}$
- $[a, b] = \{x \in \mathbb{R} \mid a \leq x \leq b\}$

Equality of Sets

Definition 0.0.14 (Axiom of Extensionality)**.** *For any two sets $A, B \subseteq \Omega$*

$$A = B \iff \forall x(x \in A \iff x \in B)$$

Two sets are equal if and only if they have the same elements.

This "sameness" of their elements are defined up to the definition of equality used in the universe of discourse. A set is defined by its elements, it is only its elements.

Subsets and Supersets

Definition 0.0.15. *For two sets A, B, we say $A \subseteq B$ (A is a subset of B) if*

$$\forall x(x \in A \implies x \in B)$$

We may also define the relations $\subsetneq$ or $\subset$ to mean

$$A \subsetneq B \iff (A \subseteq B) \wedge (A \neq B)$$

This is read as "A is a proper subset of B" or "A is a strict subset of B". We may also say that A is a superset of B ($A \supseteq B$) if $B \subseteq A$. We can describe our common universes of discourse as subsets of each other.

$$\mathbb{N}^+ \subsetneq \mathbb{N} \subsetneq \mathbb{Z} \subsetneq \mathbb{Q} \subsetneq \mathbb{R} \subsetneq \mathbb{C}$$

As another example, given the way we have defined intervals, every interval is a subset of $\mathbb{R}$. Some intervals are subsets of each other.

$$(a, b) \subsetneq [a, b] \subsetneq (a - 1, b + 1)$$

Notice that not all sets are comparable. It is true for every number that exactly one of $(x < y), (x > y), (x = y)$ is true, the numbers are totally ordered. Not true for sets. Consider the sets $A = \{1, 2, 3\}, B = \{3, 4, 5\}$. They certainly are not equal, and both $A \subseteq B$ and $B \subseteq A$ are false. Numbers are totally ordered, sets are partially ordered. This generality gives set theory a lot of expressive power. Number theory can only express the parts of mathematics which happen to be totally ordered. Sets can express that, and also much more.

Cardinality

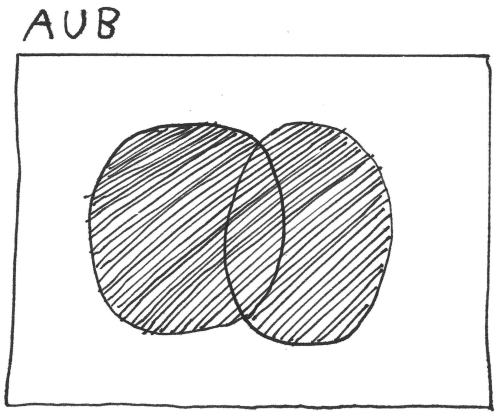
The cardinality of a set is the number of elements it has. It is either a natural number or infinity. The set $\{1, 2, 3, 4\}$ has only four elements. Sometimes sets have an infinite number of elements, so the cardinality of $\mathbb{N}$ is denoted as infinity.[33] The empty set has no elements, so its cardinality is zero. For S any set, we write $|S|$ to mean its cardinality. Although $|\mathbb{N}| = \infty$, we know $|\{\mathbb{N}\}| = 1$. The set $\{\mathbb{N}\}$ is a set containing one element. That element is itself a set, containing infinitely many elements. A set may itself be a collection of elements, or an element to be collected.

[33] All three element sets have the same "threeness", but not all infinite sets have the same "infiniteness". Some infinities are greater than others. We will not discuss this further. Discrete mathematics is the study of discrete objects. We will have heavy attention on finite sets.

Set Operations

A set operation takes two sets and produces a third. Given sets, you can combine them in many diverse and creative way to create even more interesting sets. We detail several of the most basic set operations.

Union



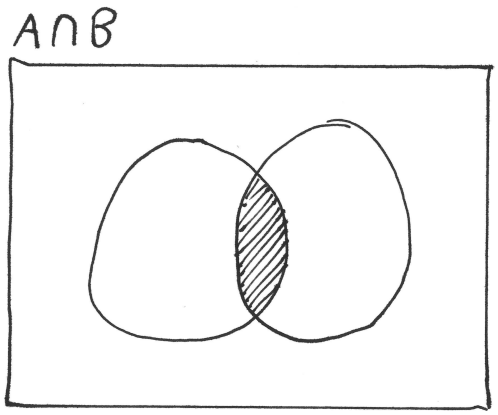
Definition 0.0.16. *Let $A, B \subseteq \Omega$. We define $A \cup B$ to be*

$$A \cup B = \{x \in \Omega \mid x \in A \text{ or } x \in B\}$$

The union of two sets is a new set containing all the elements of both parts.

- $\{1,2,3,4\} \cup \{1,5,7,8\} = \{1,2,3,4,5,7,8\}$
- $\mathbb{N} \cup \mathbb{Z} = \mathbb{Z}$
- $\mathbb{Q} \cup \mathbb{I} = \mathbb{R}$

It doesn't matter that A, B may contain some of the same elements, one will be in $A \cup B$. Also, observe that if $A = B$ then $A \cup B = A = B$. And if $A \subsetneq B$ then $A \cup B = B$. The union is analogous to a logical disjunction.



Intersection

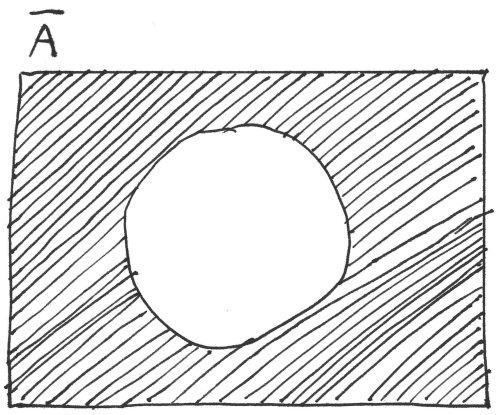
Definition 0.0.17. *Let $A, B \subseteq \Omega$. We define $A \cap B$ to be*

$$A \cap B = \{x \in \Omega \mid x \in A \text{ and } x \in B\}$$

The intersection of two sets contains exactly and only the elements that are in both.

- $\{1,2,3,4\} \cup \{1,5,7,8\} = \{1\}$
- $\mathbb{R} \cap \mathbb{Q} = \mathbb{Q}$
- $\mathbb{Q} \cap \mathbb{I} = \emptyset$

Note that if $A = B$ then $A \cap B = A = B$ and if $A \subseteq B$ then $A \cap B = A$. Also note that $|A \cap B| \leq \min(|A|, |B|)$. The intersection is analogous to a logical conjunction.



Complement

Definition 0.0.18. *If $A \subseteq \Omega$ some set, then we define the complement of A written as $\overline{A}$ as*

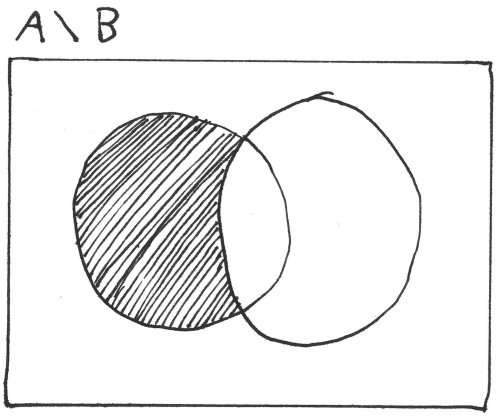
$$\overline{A} = \{x \in \Omega \mid x \notin A\}$$

You may also see this written as $A^{\complement}$ or A^c.

- $\overline{Q} = \mathbb{I}$
- If $\Omega = \mathbb{N}$ and $A = \{2,3,4\}$ then $\overline{A} = \{0,1\} \cup \{5,6,...\}$

The complement is defined with respect to the universe of discourse, and its always important to keep in mind what it is at all times. Observe that $\overline{\overline{A}} = A$ and $x \in A \iff x \notin \overline{A}$. The complement is analogous to the logical negation.

Difference



Definition 0.0.19. *The difference of two sets A, B, written $A \setminus B$ (or even $A - B$) is defined as*

$$A \setminus B = \{x \in \Omega \mid x \in A \text{ and } x \notin B\}$$

.

For $A \setminus B$, you keep everything in A but take out everything B, if there is anything to take out. Think of it like A excluding B.

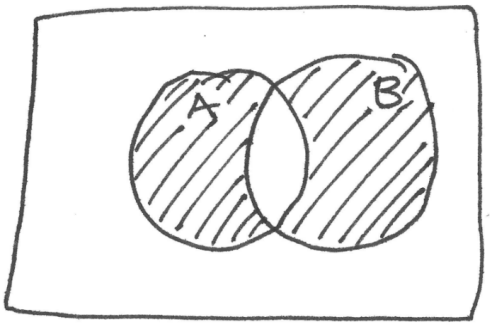
- $\mathbb{Z} \setminus \mathbb{N} = \{..., -2, -1, 0\}$
- $\mathbb{R} \setminus \mathbb{Q} = \mathbb{I}$

Also note that $A \setminus B = A \cap \overline{B}$.

Symmetric Difference

Definition 0.0.20. *The symmetric difference of two sets A, B, usually written $A \oplus B$ or $A \triangle B$ is defined as*

$$A \triangle B = \{x \in \Omega \mid x \in A \text{ orx} \in B \text{ but not both}\}$$



It can be written in many equivalent ways.

$$A \triangle B = (A \setminus B) \cup (B \setminus A) = (A \cap \overline{B}) \cup (\overline{A} \cap B)$$

Cartesian Product

Definition 0.0.21. *Given two sets A, B, we write the cartesian product*

$$A \times B = \{(x, y) \mid \forall x \in A, \forall y \in B\}$$

The cartesian product is simply a new set consisting of all possible pairs, in which the pairs are ordered.For example

$$\{0, 1, 2\} \times \{2, 3\} = \{(0, 2), (1, 2), (2, 2), (0, 3), (1, 3), (2, 3)\}$$

The pair $(1, 2)$ is different than the pair $(2, 1)$. A set is unordered, but a tuple is ordered.[34] For any set A, $A \times \emptyset = \emptyset$ There are no elements in $\emptyset$, and every cartesian product is a set containing elements which satisfy something. Since nothing satisfies it the set is empty.

The graphs of functions can be represented as subsets of cartesitan products. Consider the function $f(x) = x^2$. We consider this function as something which takes input and brings it to output. We could technically consider it as a subset of the cartesian plane:

$$\{(x, y) \in \mathbb{R} \times \mathbb{R} \mid y = x^2\}$$

The graph of this function is really a subset of $\mathbb{R} \times \mathbb{R}$.

[34] You may notice that $(A \times B) \times C \neq A \times (B \times C)$, since the fact that tuples are ordered implies elements of the first are of the form $((a, b), c)$, and elements of the second are of the form $(a, (b, c))$. They may look different, but they behave the same. Mathematicians have not found it useful to distinguish between these, and consider elements of $A \times B \times C$ to be of the form (a, b, c).

Indexed Collections

An indexed collection of sets is simply a way to notate many sets when venn diagrams get too big. You could even denote infinitely many sets.

Let $i \in \mathbb{N}$ and let $A_i = \{i, i+1\}$. Then

$$\bigcup_{i=1}^{5} A_i = \{1,2\} \cup \{2,3\} \cup \{3,4\} \cup \{4,5\} \cup \{5,6\} = \{1,2,3,4,5,6\}$$

Lets do an example with intervals

$$\bigcap_{i=n}^{\infty} \left[0, \frac{1}{n}\right) = \{0\}$$

[35] This is an infinite intersection of intervals, each containing infinitely many elements. While in propositional calculus, you can only conjunct or disjunct finitely many propositions together, in set theory, you may take the intersection or union of infinitely many sets. This is an infinite intersection of infinite sets which only contains one element. Lets prove it.

[35] Like how we have giant sigmas to denote sums, we have giant unions and intersections to denote the operation over a collection of sets.

Proof. Observe first that zero is in the intersection. It is also the least element of each interval, and is then the least element of the intersection. Assume to the contrary then that there is a second element in this intersection x, and further suppose that this element is positive, $x > 0$. Since the sequence $\frac{1}{1}, \frac{1}{2}, \frac{1}{3}, ...$ tens to zero, there must exist a k such that $\frac{1}{k+1} \leq x < \frac{1}{k}$. So $x \notin [0, \frac{1}{k+1})$. But then x could not have been in the intersection, contradiction. □

Power Set

Definition 0.0.22. *For S some set, We define $\mathcal{P}(S)$ to be the set containing all possible subsets of S.*

For example,

$$\mathcal{P}(\{1,2\}) = \{\emptyset, \{1\}\{2\}, \{1,2\}\}$$

Additionally, $\mathcal{P}(\emptyset) = \{\emptyset\}$ and $\mathcal{P}(\{\emptyset\}) = \{\emptyset, \{\emptyset\}\}$. For any set A, it is true that $\emptyset \subseteq A$ and $A \subseteq A$, so the empty set and the set itself will always be elements of the power set. Think of a subset like a possible choice, or selection of elements. For any set A, you always have the options of choosing nothing ($\emptyset$) or choosing everything (A). The power set is itself a set, which contains as elements all possible subsets, even if the original set had a universe of discourse of something else. There are certainly more ways to choose subsets than there are to choose elements. The power set of an infinite set will obviously be infinite, but how many sets are in the power set of a finite set?

Theorem 35. *Let S be a finite set. If $|S| = n$ then $|\mathcal{P}(S)| = 2^n$*

Proof. We proceed by induction on n, our base case is $n = 0$. There is only one set of zero elements, and it is the empty set $\emptyset$. We know $\mathcal{P}(\emptyset) = \{\emptyset\}$, so $|\mathcal{P}(\emptyset)| = 1$ and $2^0 = 1$.

Now assume our induction hypothesis, that If $|S| = k$ then $|\mathcal{P}(S)| = 2^k$. Let S' be any set such that $|S'| = k + 1$. We prove $|\mathcal{P}(S')| = 2^{k+1}$. Since S' is nonempty, it has some element $x \in S'$. Consider the set $S = S' \setminus \{x\}$. Consider all possible subsets of S'. There are two cases:

- If $A \in \mathcal{P}(S')$ and $x \notin A$, then $A \in \mathcal{P}(S)$.
- If $A \in \mathcal{P}(S')$ and $x \in A$, then there is a unique set B with $A = B \cup \{x\}$ such that $B \in \mathcal{P}(S)$.

Lets now count the number of subsets of S', the number of elements of $\mathcal{P}(S')$. Each subset is either in $\mathcal{P}(S)$, or it is a set in $\mathcal{P}(S)$ with the element x. We then see that $|\mathcal{P}(S')| = |\mathcal{P}(S)| + |\mathcal{P}(S)| = 2^k + 2^k = 2(2^k) = 2^{k+1}$. We have proven that all n element sets have 2^n possible subsets. □

Sometimes you may see the notation $\mathcal{P}(S) = 2^S$ for this reason.

Partitions

Definition 0.0.23. *A partition of some set S is a collection of sets A, B such that $A \cup B = S$ and $A \cap B = \emptyset$*

A set can be partitioned into multiple sets. A parition of some set S is a a collection of sets $A_1, A_2, ...$ such that

$$S = \bigcup_{i=1}^{\infty} A_i$$

and if $i \neq j$ then $A_i \cap A_j = \emptyset$

For example, we may partition $\mathbb{N}$ into two sets, one containing evens, and one containing odds. $\mathbb{N} = E \cup O$. We may also partition it into infinitely many subsets each containing one element.

$$\mathbb{N} = \bigcup_{i=0}^{\infty} \{i\} = \{0\} \cup \{1\} \cup \{2\} \cup \{3\} \cup ...$$

Proof by Double Set Containment

Rather than mechanically applying a set of laws, proving equality of sets is more practically done with the technique of *double set containment*

Theorem 36.

$$A = B \iff (A \subseteq B) \wedge (B \subseteq A)$$

Proof. We take the Axiom of Existentionality, apply a logical equivalence to the if and only if, and observe the definition of subset.

$$\begin{aligned}
&A = B \iff \forall x(x \in A \iff x \in B) \\
&A = B \iff \forall x((x \in A \implies x \in B) \wedge (x \in B \implies x \in A)) \\
&A = B \iff \forall x(x \in A \implies x \in B) \wedge \forall x(x \in B \implies x \in A) \\
&A = B \iff (A \subseteq B) \wedge (B \subseteq A)
\end{aligned}$$

□

To prove two sets A, B are equal, in two steps, simply prove $A \subseteq B$ and $B \subseteq A$. Often times, these two proofs may follow for different reasons. The proof syntax should usually look like "Let $x \in A$. ... Therefore, $x \in B$. Let $y \in B$. ... Therefore, $y \in A$." This is the standard form, but like with any proof, there will be variance in why the deduction follows correctly. Let us do some examples.

Theorem 37. *Let $A, B \subseteq \Omega$. Then*

$$A \setminus B = A \cap \overline{B}$$

Proof. We proceed by a double set containment. We first prove that $A \setminus B = A \cap \overline{B}$. Let $x \in A \setminus B$. Then $x \in A$ and $x \notin B$. Therefore, $x \in \overline{B}$. Since $x \in A$ and $x \in \overline{B}$, it follows that $x \in A \cap \overline{B}$.

Let $x \in A \cap \overline{B}$. Then $x \in A$ and $x \in \overline{B}$. So $x \notin B$. Since $x \in A$ and $x \notin B$, it follows that $x \in A \setminus B$. □

We reuse the same variable here x, because they belong to different parts of the proof and are scoped differently.

Theorem 38. *Let $A, B, C \subseteq \Omega$. Then*

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

Proof. We proceed by a double set containment. We first prove that $A \cap (B \cup C) \subseteq (A \cap B) \cup (A \cap C)$. Let $x \in A \cap (B \cup C)$. Then $x \in A$ and $x \in B \cup C$. Then $x \in B$ or $x \in C$. We have two cases:

- If $x \in B$, then $x \in A \cap B$, so $x \in (A \cap B) \cup (A \cap C)$
- If $x \in C$, then $x \in A \cap C$, so $x \in (A \cap B) \cup (A \cap C)$

in all cases, we have $x \in (A \cap B) \cup (A \cap C)$

Now let $x \in (A \cap B) \cup (A \cap C)$. Then $x \in (A \cap B)$ or $(A \cap C)$ We have two cases.

- If $x \in A \cap B$ then $x \in A$ and $x \in B$. Since $x \in B$, we know $x \in B \cup C$.
- If $x \in A \cap C$ then $x \in A$ and $x \in C$. Since $x \in C$, we know $x \in B \cup C$.

Since $x \in A$ and $x \in B \cup C$, then it follows that $x \in A \cap (B \cup C)$. □

Theorem 39. *Let*

$$S = \{3s + 2t \mid s, t \in \mathbb{Z}\}$$

Then $S = \mathbb{Z}$

Proof. We proceed by a double set containment. Let $x \in S$. Then there exists $s, t \in \mathbb{Z}$ such that $x = 3s + 2t$. By closure, we see $x \in \mathbb{Z}$.

Now let $x \in \mathbb{Z}$. We prove that $x \in S$. Consider $s = -3x$ and $t = 5x$. By closure $s, t \in \mathbb{Z}$ so

$$3s + 2t = 3(-3x) + 2(5x) = -9x + 10x = x$$

Since there exists $s, t \in \mathbb{Z}$ such that $3s + 2t = x$, we see that $x \in S$. □

DeMorgan's Law

Theorem 40.

$$\overline{A \cup B} = \overline{A} \cap \overline{B}$$

Proof. We proceed by a double set containment.

- ($\subseteq$)

$$\begin{aligned} x &\in \overline{A \cup B} \\ x &\notin A \cup B \\ x \notin A &\text{ and } x \notin B \\ x \in \overline{A} &\text{ and } x \in \overline{B} \\ x &\in \overline{A} \cap \overline{B} \end{aligned}$$

- ($\supseteq$)

$$\begin{aligned} x &\in \overline{A} \cap \overline{B} \\ x \in \overline{A} &\text{ and } x \in \overline{B} \\ x \notin A &\text{ and } x \notin B \\ x &\notin A \cup B \\ x &\in \overline{A \cup B} \end{aligned}$$

□

Theorem 41. *Let $n \geq 2$ and $A_1, ..., A_n \subseteq \Omega$. Then*

$$\overline{\bigcup_{i=1}^{n} A_i} = \bigcap_{i=1}^{n} \overline{A_i}$$

Proof. We proceed by induction on n. Our base case is $n = 2$ and $\overline{A_1 \cup A_2} = \overline{A_1} \cap \overline{A_2}$ is true by our previous proof. Now assume our induction hypothesis, for any k sets that

$$\overline{\bigcup_{i=1}^{k} A_i} = \bigcap_{i=1}^{k} \overline{A_i}$$

We prove

$$\overline{\bigcup_{i=1}^{k+1} A_i} = \bigcap_{i=1}^{k+1} \overline{A_i}$$

Let $T = A_1 \cup ... \cup A_k$. Then

$$\overline{\bigcup_{i=1}^{k+1} A_i} = \overline{T \cup A_{k+1}} = \overline{T} \cap \overline{A_{k+1}}$$

By DeMorgan's law on two sets. Since T is a union of k sets, we may apply our induction hypothesis

$$\overline{T} \cap \overline{A_{k+1}} = \overline{\bigcup_{i=1}^{k} A_i} \cap \overline{A_{k+1}} = \bigcap_{i=1}^{k} \overline{A_i} \cap \overline{A_{k+1}} = \bigcap_{i=1}^{k+1} \overline{A_i}$$

Thus for any Thus, we have proven

$$\overline{\bigcup_{i=1}^{n} A_i} = \bigcap_{i=1}^{n} \overline{A_i}$$

□

On the Power of Set Theory

There is quite a zoo of mathematical structures. These include numbers, functions, matricies, vectors, tensors, polygons, graphs, polynomials, groups, rings, fields, points, lines, planes, propositions, predicates, surfaces, manifolds, knots, limits, derivatives, integrals, I could go on. It is not the case that the study of each of these objects has their own sets of axioms and rules defined. Rather, each of these objects may be constructed from the axioms of pure set theory. Sets are such a basic, atomic object, they can serve as a foundation for all of mathematics. Given axioms for sets, you can construct numbers, and the properties of numbers, but given axioms for numbers, you cannot derive properties of sets. We give a few examples of such constructions.

The Natural Numbers from Set Theory

The natural numbers are not defined, rather, they are constructed inductively.

- zero is a natural number.

$$(0 \in \mathbb{N})$$

- If n is a natural number, then $S(n)$ is a natural number.

$$\forall n(n \in \mathbb{N} \implies S(n) \in \mathbb{N})$$

  Here $S(n)$ is the successor function $S(n) = n + 1$.

This is why it is so important that zero is a natural number. It is not only a natural number, it is the only natural number, it is the most natural number. If we want to construct the natural numbers in set theory, we only need something that looks like a zero, and something that looks like a successor function. An ordinal is a special kind of set, which pretends its a number. The properties that numbers have with each other, we will simulate with these ordinals.

In the beginning, there was nothing. No universes of discourse, no objects to associate with each other. Nothing. Yet, nothing is itself, a kind of something. It may represent nothing, but it is not nothing. Zero is itself not nothing, it is a something which represents nothing. We have a set which contains nothing, but it is still something. We have our zeroth ordinal:

$$T_0 := \emptyset$$

What better candidate for the zeroth ordinal than the empty set, since it has no elements. We need to now construct a another set, one which is distinct from the first, and therefore, non empty. Since it is not empty, it must contain something. Why not have this something be what we already know is something. We have our first ordinal:

$$T_1 := \{\emptyset\}$$

This set is different than the emptyset, just like how $0 \neq 1$. It contains one element, and is a good candidate for the number one. What about the number two? We need a set with two distinct elements. Coincidentally, we have just contructed two somethings. We have our second ordinal.

$$T_2 := \{\emptyset, \{\emptyset\}\}$$

We can repeat this process to realize our successor function. For any ordinal set T_n, let

$$T_{n+1} = S(T_n) = T_n \cup \{T_n\}$$

We can construct the first few ordinals then as follows.

- $T_0 = \emptyset$
- $T_1 = \{\emptyset\}$
- $T_2 = \{\emptyset, \{\emptyset\}\}$
- $T_3 = \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}$
- $T_4 = \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}, \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}\}$
- $T_5 = \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}, \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}, \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}, \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}\}\}$
- ...

Observe that $S(T_n) = \{T_0, T_1, ..., T_n\}$, and that $|T_n| = n$. We now show that these ordinals do have all the useful properties of numbers.

The first desirable property of numbers would be equality. By the Axiom of Existenionality, we have a definition of equality of sets to be that two sets are equal if and only if they have the same elements. We can define equality of numbers to be if their ordinals are equal.

$$n = m \iff T_n = T_m$$

The first equality is an equality of numbers, the second one is an equality of sets. Equality of numbers is defined on the equality of sets.

The next desirable property of numbers would be well-ordering. Although in general, sets are partially ordered, the ordinals are a special case which are totally ordered. Notice here there isn't a universe of discourse. The elements of sets can be only other sets! Set theory, and therefore, all of mathematics, can be done totally atomless. Of course, this is not always useful, only cool that it could be done at all. We much rather usefully use set theory with a well defined universe of discourse.

$$n < m \iff T_n \in T_m$$

How useful for us was that the definition of an ordinal to be the set containing all previous ordinals.

Identities

- $A \cap \Omega = A$ **Identity**
- $A \cup \emptyset = A$

- $A \cup \Omega = \Omega$ **Domination**
- $A \cap \emptyset = \emptyset$

Figure 9: Cool Lamp I found on Tenth Street

- $A \cup A = A$ **Idempotent**
- $A \cap A = A$

- $\overline{\overline{A}}$ **Complementation**

- $A \cup B = B \cup A$ **Communitivity**
- $A \cap B = B \cap A$

- $A \cup (B \cup C) = (A \cup B) \cup C$ **Associativity**
- $A \cap (B \cap C) = (A \cap B) \cap C$

- $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ **Distributive**
- $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

- $\overline{A \cup B} = \overline{A} \cap \overline{B}$ **DeMorgan's**
- $\overline{A \cap B} = \overline{A} \cup \overline{B}$

- $A \cup (A \cap B) = A$ **Absorption**
- $A \cap (A \cup B) = A$

- $A \cup \overline{A} = \Omega$ **Complement**
- $A \cap \overline{A} = \emptyset$

Functions

You should have come across functions at some point in your mathematical career. They are the most studied mathematical object of all time. But what exactly is a function to begin with?

Relations

Definition 0.0.24 (Relation)**.** *For any sets A, B we define a relation R to be a subset of $R \subseteq A \times B$*

We write aRb for $a \in A$ and $b \in B$ to mean a relates to b. The symbol R here is generic, and is usually meant to be replaced with some operator.

Consider $A = \{0, 1, 2\}$ and $B = \{a, b\}$, then the following are some possible relations $R \subseteq A \times B$

- $R = \{(0, b)\}$
- $R = \{(0, a), (1, a), (2, a), (1, b), (2, b)\}$[36]
- $R = A \times B$
- $R = \emptyset$

[36] table

Relations among finite sets exist, but the richness occurs when sets are infinite. We may define the inequality of real numbers as a relation:

$$\text{“} \leq \text{”} = \{(a, b) \in \mathbb{R} \times \mathbb{R} \mid a \leq b\}$$

[37] There are many more interesting properties of relations we will into later. For now we focus on the rich case of when a relation is function.

[37] non circularity

Functions

A function is a kind of relation with exactly two special properties.

Definition 0.0.25. *Let $f \subseteq A \times B$ be a relation. A relation is a function when for all $a \in A$ there exists a unique $b \in B$ such that the pair $(a, b) \in f$.*

[38] Instead of writing a function as $f \subseteq A \times B$, we usually write it as $f : A \to B$, in the sense it is a map from the set A to the set B.They are more interesting to think of like a map, a transformation from elements in the set A to elements in the set B Implicit in the definition of a function is actually two properties:

[38] vertical line test

- First, every element of A is mapped to some element of B. The function is *total*[39]
- Second, an element of A is mapped to at most one element of B.[40]

[39] partial

[40] vertical line test.

We can describe some functions you may have seen before using set builder notation.

- $\{(x, y) \in \mathbb{R} \times \mathbb{R} \mid y = e^x\}$ is usually written as $f(x) = e^x$
- $\{(x, y) \in \mathbb{R} \times \mathbb{R} \mid y = x^2\}$ is usually written as $f(x) = x^2$

For $f : A \to B$, instead of writing $(a, b) \in f$, we usually write $f(a) = b$. Functions can be defined using set theory, but we need not do so for conveinence. These you have associated as an input and output, but you can formally describe a function as a subset of a cartesian product. Using sets, we can define functions as a kind of set.

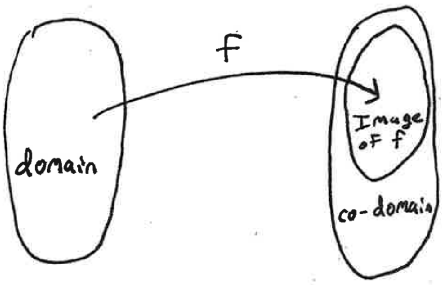
- The domain of a function $f : A \to B$ is A, the set where its input is defined.
- The co-domain of a function $f : A \to B$ is B, the set where its output is defined.
- The image of a function is defined as a subset of the co-domain which has values mapped to.
- For example, if $f(x) = x^2$, then the domain of f is $\mathbb{R}$, the codomain is $\mathbb{R}$, but the image is $\{x \in \mathbb{R} \mid x \geq 0\}$. Not every element of the codomain gets mapped to.

We discuss a function $f : A \to B$ which for $f(a) = b$ maps the element $a \in A$ to the element $b \in B$. We can also discuss how f maps a subset of the elements of A.

Definition 0.0.26. *Let $f : A \to B$. Let $S \subseteq A$ and define*

$$f(S) = \{f(a) \in B \mid \forall a \in A\}$$

Note that if $S \subseteq A$ then $f(S) \subseteq B$. It is not mapping the set, but the set of elements of S which are mapped to in B.



Equality of two functions

Definition 0.0.27. *Two functions $f : A \to B$ and $g : A \to B$ are said to be equal if*

$$\forall a \in A[f(a) = g(a)]$$

They map the same elements to the same elements. Two functions are not equal if they only have the same domain and co-domain. They must map the same elements to the same elements. Note that two functions could be "equal" on all values, yet have different co-domains. Consider $f(n) = n^2$ for $f : \mathbb{N} \to \mathbb{Z}$ versus $g : \mathbb{N} \to \mathbb{N}$. Since the image of this function is the naturals anyway, why not restrict the co-domain to the image. These functions are technically not equal because they do not have the same co-domain, but they are practically equivalent.

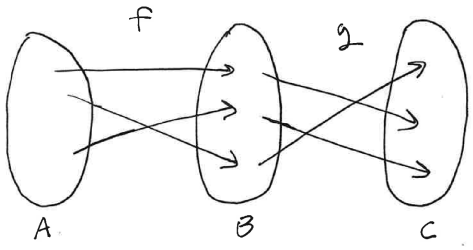
This formulation of what is or isn't a function is relatively modern. Functions used to be too restrictive in their definition, but the generality of this definition of function allows for some interesting examples. For example Let $f : \mathbb{R} \to \mathbb{R}$ such that

$$f(x) = \begin{cases} 1 & x \in \mathbb{Q} \\ 0 & x \in \mathbb{R} \setminus \mathbb{Q} \end{cases}$$

Does this function have a derivative? Is it even continuous? Does it even have an area under its curve? Or an average of its values within some interval? Nice and clean and simple functions do have all kinds of interesting properties, but the generality in the definition of what kinds of functions you can make allows for a lot of horrors. The Descartian idea of a function was something you could plot in the cartesian plane. I claim that you could not even attempt to plot this function. Even though this is not a nice function, $|2f(x) - 1| = 1$, a constant function. Extremely nice.

Combining functions

Definition 0.0.28. *Let f, g be two functions $f : A \to B$, $g : B \to C$. We define the composition of f, g as $= g(f(x))$.*



This can also be written as $g \circ f$.

f takes an element of A and returns an element of B, which is part of the domain of g, so g then maps this element to some element of C. The domain of $g \circ f$ is A but the codomain of $g \circ f$ is C.

The set definition of a function allows you to define multiple arguments. For example, if you want to define a function that takes two numbers and outputs a third, you may write $f(a, b) = c$. Rather than defining two domains of this function somehow, it only has one domain, $\mathbb{N} \times \mathbb{N}$. The input is not two distinct numbers, rather an element of the cartesian product $(a, b) \in \mathbb{N} \times \mathbb{N}$.

Bijectivity

A bijection is an ideal function. For $f : A \to B$, if f is a bijection, it is a perfect pairing between A, B. We detail how to prove when a function is bijective, and properties of bijections.

Definition 0.0.29. *We say a function f is injective, or one-to-one if $a \neq b \implies f(a) \neq f(b)$. Element distinctness in the domain implies distinctness after mapping into the co-domain.*

To prove a function to be injective, you usually use the contrapositive. $f(a) = f(b) \implies a = b$.

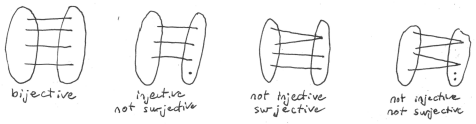
Definition 0.0.30. *We say a function f is surjective, or onto if $\forall b \in B, \exists a$ such that $f(b) = a$.*

A function is surjective[41] if there is no element in the co-domain which remains unmapped. Nothing is left behind. A function is surjective if its image is equal to its co-domain.

[41] french people

Definition 0.0.31. *A function is bijective if and only if it is injective and surjective.*

When you think of a bijection, you should think of the following picture. A bijection is a perfect correspondence between the domain and co-domain. In order to prove that a function is a bijection, you must prove it is injective and surjective. We do some examples.



Theorem 42. *Let $a, b \in \mathbb{R}$ and $a \neq 0$. Define the function $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = ax + b$. The function f is a bijection*

Proof. We prove it is injective and surjective. To first prove it is injective, we let $f(x) = f(x')$ and prove $x = x'$.

$$\begin{aligned} f(x) &= f(x') \\ ax + b &= ax' + b \\ ax &= ax' \text{ and since } a \neq 0 \\ x &= x' \end{aligned}$$

Next we prove it is surjective. Let $r \in \mathbb{R}$. We prove there exists an $x \in \mathbb{R}$ such that $f(x) = r$. Consider $x = \frac{r-b}{a}$. Then

$$f(x) = f\left(\frac{r-b}{a}\right) = a\left(\frac{r-b}{a}\right) + b = (r-b) + b = r$$

□

proof commentary

You usually have to do a proof three times. First to figure out if its true, second to work out some of the details and structure of the proof, and a third time formally. This proof is the third, final version, and it does not show how we were able to get the proof itself. To show surjectivity, behind the scenes, we computed the inverse of the function to get $x = 2r/(r-3)$. Once we had this, we could procede with the proof of surjectivity normally.

Theorem 43. *Let $f : \mathbb{R} \setminus \{1\} \to \mathbb{R} \setminus \{1\}$ by $f(x) = \frac{x}{x-1}$. Then f is a bijection.*

Proof. We prove it is injective and surjective. To first prove it is injective, we let $f(x) = f(x')$ and prove $x = y$

$$\begin{aligned} f(x) &= f(y) \\ \frac{x}{x-1} &= \frac{y}{y-1} \\ x(y-1) &= y(x-1) \\ xy - x &= yx - y \\ -x &= -y \\ x &= y \end{aligned}$$

Next we prove it is surjective. Let $r \in \mathbb{R} \setminus \{1\}$. We prove there is an $x \in \mathbb{R} \setminus \{1\}$ such that $f(x) = r$. Consider $x = \frac{r}{r-1}$. Then

$$\begin{aligned} f(x) &= f\left(\frac{r}{r-1}\right) \\ &= \frac{\frac{r}{r-1}}{\frac{r}{r-1} - 1} \\ &= \frac{r}{r-1}\left(\frac{r}{r-1} - 1\right)^{-1} \\ &= \frac{r}{r-1}\left(\frac{r-(r-1)}{r-1}\right)^{-1} \\ &= \frac{r}{r-1}\left(\frac{1}{r-1}\right)^{-1} \\ &= \frac{r}{r-1}\left(\frac{r-1}{1}\right) \\ &= r \end{aligned}$$

Since we have proven f is injective and surjective, it is bijective. □

Theorem 44. *Let $S^2 = \{(x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 = 1\}$ be the unit sphere in three dimensions and let $N = (0, 0, 1)$ be the "north pole". Let the stereographic projection*

$$f : S^2 \setminus \{N\} \to \mathbb{R}^2$$

be defined by

$$f(x, y, z) = \left(\frac{x}{1-z}, \frac{y}{1-z}\right)$$

The stereographic projection in three dimensions is a bijection

Proof. We first prove f is injective. Let $f(x, y, z) = f(x_1, y_1, z_1)$. We prove that $(x, y, z) = (x_1, y_1, z_1)$

$$f(x, y, z) = f(x_1, y_1, z_1)$$
$$\left(\frac{x}{1-z}, \frac{y}{1-z}\right) = \left(\frac{x_1}{1-z_1}, \frac{y_1}{1-z_1}\right)$$

Since both of these points in $\mathbb{R}^2$ are equal, their norms are equal. For $(a, b) \in R^2$, the norm[42] is $a^2 + b^2$. We set the norms of these points equal to each other.

[42] Credit to Melvin Gao for figuring out this trick with the norms!

$$\left(\frac{x}{1-z}\right)^2 + \left(\frac{y}{1-z}\right)^2 = \left(\frac{x_1}{1-z_1}\right)^2 + \left(\frac{y_1}{1-z_1}\right)^2$$
$$\frac{x^2 + y^2}{(1-z)^2} = \frac{x_1^2 + y_1^2}{(1-z_1)^2}$$

Since $(x, y, z) \in S^2 \setminus \{N\}$, they satisfy $x^2 + y^2 + z^2 = 1$, so $x^2 + y^2 = 1 - z^2$

$$\frac{x^2 + y^2}{(1-z)^2} = \frac{x_1^2 + y_1^2}{(1-z_1)^2}$$
$$\frac{1 - z^2}{(1-z)^2} = \frac{1 - z_1^2}{(1-z_1)^2}$$
$$\frac{(1-z)(1+z)}{(1-z)^2} = \frac{(1-z_1)(1+z_1)}{(1-z_1)^2}$$
$$\frac{(1+z)}{(1-z)} = \frac{(1+z_1)}{(1-z_1)}$$
$$(1+z)(1-z_1) = (1+z_1)(1-z)$$
$$1 + z - z_1 - zz_1 = 1 - z + z_1 - zz_1$$
$$z - z_1 = z_1 - z$$
$$2z = 2z_1$$
$$z = z_1$$

Since $z = z_1$, we may return to our original expression and replace z_1 with z

$$\left(\frac{x}{1-z}, \frac{y}{1-z}\right) = \left(\frac{x_1}{1-z_1}, \frac{y_1}{1-z_1}\right)$$
$$\left(\frac{x}{1-z}, \frac{y}{1-z}\right) = \left(\frac{x_1}{1-z}, \frac{y_1}{1-z}\right)$$
$$(x, y) = (x_1, y_1)$$

Since $(x, y, z) = (x_1, y_1, z_1)$ we see that f is injective.

Next we prove it is surjective. Let $(s, t) \in \mathbb{R}^2$. We prove there exists a point $(x, y, z) \in S^2 \setminus \{N\}$ with $f(x, y, z) = (s, t)$. Consider

$$(x, y, z) = \frac{1}{s^2 + t^2 + 1}\left(2s, 2t, s^2 + t^2 - 1\right)$$

Since the domain is non trivial, we first prove $(x, y, z) \in S^2 \setminus \{N\}$. First suppose to the contrary $z = 1$. Then the numerator and denominator would be equal

$$\begin{aligned} s^2 + t^2 - 1 &= s^2 + t^2 + 1 \\ -1 &= 1 \end{aligned}$$

Contradiction. Next we verify $(x, y, z) \in S^2$

$$\begin{aligned} x^2 + y^2 + z^2 = \\ \left(\frac{2s}{s^2+t^2+1}\right)^2 + \left(\frac{2s}{s^2+t^2+1}\right)^2 + \left(\frac{s^2+t^2-1}{s^2+t^2+1}\right)^2 = \\ \frac{4s^2 + 4t^2 + (s^2+t^2-1)^2}{(s^2+t^2+1)^2} = \\ \frac{4s^2 + 4t^2 + (s^4 + 2s^2t^2 - 2s^2 + t^4 - 2t^2 + 1)}{s^4 + 2s^2t^2 + 2s^2 + t^4 + 2t^2 + 1} = \\ \frac{s^4 + 2s^2t^2 + 2s^2 + t^4 + 2t^2 + 1}{s^4 + 2s^2t^2 + 2s^2 + t^4 + 2t^2 + 1} = 1 \end{aligned}$$

We now verify $f(x, y, z) = (s, t)$

$$f(x, y, z) = \left(\frac{x}{1-z}, \frac{y}{1-z}\right)$$

We first verify $\frac{x}{1-z} = s$

$$\begin{aligned} \frac{x}{1-z} &= \left(\frac{2s}{s^2+t^2+1}\right)\left(\frac{1}{1-\frac{s^2+t^2-1}{s^2+t^2+1}}\right) \\ &= \left(\frac{2s}{s^2+t^2+1}\right)\left(\frac{1}{\frac{s^2+t^2+1}{s^2+t^2+1} - \frac{s^2+t^2-1}{s^2+t^2+1}}\right) \\ &= \left(\frac{2s}{s^2+t^2+1}\right)\left(\frac{1}{\frac{s^2+t^2+1-s^2-t^2+1}{s^2+t^2+1}}\right) \\ &= \left(\frac{2s}{s^2+t^2+1}\right)\left(\frac{1}{\frac{2}{s^2+t^2+1}}\right) \\ &= \left(\frac{2s}{s^2+t^2+1}\right)\left(\frac{s^2+t^2+1}{2}\right) = s \end{aligned}$$

By a symmetrical argument, we see $\frac{y}{1-z} = t$. Therefore, $f(x, y, z) = (s, t)$, and f is surjective. □

Properties of Bijections

Theorem 45. *For any function f, The image of f is equal to the co-domain of f if and only if f is surjective.*

Proof. Let $f : A \to B$ be a function and let $f(A) = \{f(a) \in B \mid \forall a \in A\}$ be the image of f.
($\implies$) We first prove that if $f(A) = B$ then f must be surjective. The definition of surjectivity is that for every $b \in B$ there exists some $a \in A$ such that $f(a) = b$. Every element of B is mapped to by some element A. The image $f(A)$ is defined to be the set of elements which are mapped to by elements of B. So if $f(A) = B$, then every element of B is mapped to by some element of A, implying that f must be surjective.

($\Longleftarrow$) We now prove that if f is surjective, then $f(A) = B$ By definition, the image is certainly a subset of the co-domain, so $f(A) \subseteq B$. We only need to prove then that $B \subseteq f(A)$. Let $b \in B$. Since f is surjective, we know that for every $b \in B$ there exists an $a \in A$ such that $f(a) = b$. Then this $b \in f(A)$. Since this is for every $b \in B$, we see that $B \subseteq f(A)$. □

Theorem 46. *The composition of bijections is a bijection.*

Proof. Let $f : A \to B$ and $g : B \to C$ be bijections. We prove that $g \circ f : A \to C$ is a bijection.

We first prove it is injective. Let $g(f(a)) = g(f(a'))$. We prove that $a = a'$. Since g is injective then

$$\begin{aligned} g(f(a)) &= g(f(a')) && \text{Since } g \text{ is injective} \\ f(a) &= f(a') && \text{Since } f \text{ is injective} \\ a &= a' \end{aligned}$$

Next, we prove it is surjective, we do so by proving that the image of $g \circ f$ is equal to the codomain. Let the image of g be denoted by

$$g(f(A)) = \{g(f(a)) \in C \mid \forall a \in A\}$$

Since f is surjective, the image of f is equal to the co-domain of f, so $f(A) = B$

$$g(f(A)) = g(B)$$

Since g is surjective, the image of g is equal to the co-domain of g, so $g(B) = C$. Therefore

$$g(f(A)) = g(B) = C$$

□

Theorem 47. *The inverse of a bijection is a bijection.*

Proof. Let $f : A \to B$ be a bijection. Let $f^{-1} : B \to A$ be a function such that for $f(a) = b$ then $f^{-1}(b) = a$. We prove f^{-1} is a bijection.

We first prove f^{-1} is injective. Let $f^{-1}(b) = f^{-1}(b')$. We prove that $b = b'$. Note that $f^{-1}(b)$ is an element of A, and since f is a function, for each element in A, it maps it to exactly one element of B. Therefore, $f(f^{-1}(b))$ is exactly one element of B, namely, $b \in B$.

$$\begin{aligned} f^{-1}(b) &= f^{-1}(b') \\ f(f^{-1}(b)) &= f(f^{-1}(b')) \\ b &= b' \end{aligned}$$

Next we prove that f^{-1} is a surjection. Let $a \in A$. We prove that there exists $b \in B$ such that $f^{-1}(b) = a$. Consider $a = f(b)$. Then

$$f^{-1}(b) = f^{-1}(f(a)) = a$$

. □

Theorem 48. *If $A, B \subseteq \Omega$ are finite sets then*

$$|A| = |B| \text{ if and only if there exists a bijection } f : A \to B$$

Proof. ($\implies$) Let A and B be finite sets with $|A| = |B|$. We prove there exists a bijection from $f : A \to B$. Without loss of generality, let $A = \{a_1, ..., a_n\}$ and let $B = \{b_1, ..., b_n\}$. Each set contains exactly n distinct elements. Consider the function $f(a_i) = b_i$. It will map the ith element of A to the ith element of B. We prove that f is a bijection by proving it is injective and surjective.

To prove it is injective, Let $f(a_i) = f(a_j)$ with $1 \leq i, j \leq n$. We prove that $a_i = a_j$.

$$f(a_i) = f(a_j)$$
$$b_i = b_j$$

By our assumption of distinctness, if $b_i = b_j$ then $i = j$, so

$$b_i = b_j$$
$$i = j$$
$$a_i = a_j$$

Next, we prove it is surjective. Let $b \in B$. Then there is an i with $1 \leq i \leq n$ such that $b = b_i$. There is an element of the domain which maps to b_i, namely a_i since $f(a_i) = b_i$.
($\impliedby$) Let A, B be any finite sets such that there exists a bijection $f : A \to B$. We prove that $|A| = |B|$. Assume to the contrary that such a bijection exists but $|A| \neq |B|$. Since f is a function, each element of A maps to exactly one element of B, so the maximum size of the image is then bounded by the domain $|f(A)| \leq |A|$. Then we have two cases

- Case 1: If $|A| > |B|$ Then there are more elements in A to map than elements in B to map to, so there must exist distinct $a_i \neq a_j$ with $f(a_i) = f(a_j)$. This implies f would not be injective.
- Case 2: If $|A| < |B|$ then $|f(A)| < |B|$, implying that the image is a strict subset of the co-domain. This implies that f would not be surjective.

In either case, we see either f is not injective or not surjective, contradicting the fact that f must be bijective. □

Monotonic Functions

Definition 0.0.32. *A function $f : \mathbb{R} \to \mathbb{R}$ is monotonically increasing if*

$$x \leq y \implies f(x) \leq f(y)$$

A function is said to be strictly monotonically increasing if

$$x < y \implies f(x) < f(y)$$

Think of a monotonically increasing function as one whos plot does not go back down, it is always going up, or staying flat.

Theorem 49. *Let $f : \mathbb{N} \to \mathbb{N}$ be defined as $f(x) = x^2$. Then f is strictly monotonically increasing.*

Proof. Let $x, y \in \mathbb{N}$ with $x < y$. Then $x^2 < xy$ and $xy < y^2$. So $x^2 < y^2$ as desired. □

Theorem 50. $\log_2(x)$ *is strictly monotonically increasing. Let* $x, y \in \mathbb{R}^+$ *with* $x < y$. *Then there is a constant* $c > 1$ *such that* $y = cx$. *Then*

$$\log_2(y) = \log_2(cx) = \log_2(c) + \log_2(x) > \log_2(x)$$

Asymptotic Analysis

speech why we care about big o

Definition 0.0.33. *Let $f : \mathbb{N} \to \mathbb{R}^+$ and $g : \mathbb{N} \to \mathbb{R}^+$ be two functions. We say*

$$f(n) \text{ is } O(g(n))$$

if there exists positive constants $c, n_0 \in \mathbb{N}$ if for all $n \geq n_0$ that

$$f(n) \leq cg(n)$$

speech moores law

Witness Proofs

Given two functions f, g, suppose you want to prove f is $O(g)$? The definition of big O is existential, you only need to give two constants c, n_0 (called witnesses). How can you easily convince your reader that these are correct and valid witnesses? The bad thing to do would be to choose $c = 1000$ and $n_0 = 1000$. Functions can have all kinds of weird, interesting, and diverse properties. A *witness proof* not only provides constants, but convinces the reader such constants are correct.

Theorem 51.

$$(n+1)^2 \text{ is } O(n^2)$$

Proof. Observe that $(n+1)^2 = n^2 + 2n + 1$. We find witnesses for each term

$$\begin{aligned} n^2 &\leq n^2 && \forall n \geq 1 \\ 2n &\leq n^2 && \forall n \geq 2 \\ 1 &\leq n^2 && \forall n \geq 1 \end{aligned}$$

If we sum the inequalities, we see

$$n^2 + 2n + 1 = (n+1)^2 \leq 3n^2 \qquad \forall n \geq 2$$

Consider witnesses $c = 3$ and $n_0 = 2$. Since for all $n \geq n_0$ we see that $(n+1)^2 \leq cn^2$, we conclude that $(n+1)^2$ is $O(n^2)$ □

Proof commentary

Theorem 52. *Let f_1, g_1, f_2, g_2 be functions which are strictly monotonically increasing. Futher suppose that that $f_1(n)$ is $O(g_1(n))$ and $f_2(n)$ is $O(g_2(n))$. Then*

$$(f_1(n) \cdot f_2(n)) \text{ is } O((g_1(n) \cdot g_2(n)))$$

Proof. Since $f_1(n)$ is $O(g_1(n))$, there exists positive constants c_1, n_1 such that for all $n \geq n_1$ that $f_1(n) \leq c_1 g_1(n)$. Since $f_2(n)$ is $O(g_2(n))$, there exists positive constants c_2, n_2 such that for all $n \geq n_2$ that $f_2(n) \leq c_2 g_2(n)$. If we multiply these inequalities together, we see

$$f_1(n)f_2(n) \leq c_1 g_1(n) f_2(n) \leq c_1 g_1(n) c_2 g_2(n) = (c_1 c_2) g_1(n) g_2(n)$$

Let $c = c_1 \cdot c_2$ and let $n_0 = \max(n_1, n_2)$. Then $\forall n \geq n_0$, we see that $f_1(n)f_2(n) \leq cg_1(n)g_2(n)$. Therefore $(f_1(n) \cdot f_2(n))$ is $O((g_1(n) \cdot g_2(n)))$. □

Corollary 53. *Notice we actually get the following corollaries as a consequence of the previous theorem*

- c *is* $O(\log n)$ *for any constant* c
- $\sqrt{n}$ *is* $O(n)$
- n *is* $O(n^2)$
- n^2 *is* $O(n^3)$
- n^k *is* $O(n^{k+1})$ *for any constant* k

Theorem 54.

$$\log n \text{ is } O(n)$$

Proof. Consider $c = n_0 = 1$. We prove $\log_2(n) \leq n$ by induction on n. Our base case is $n = 1$. $\log_2(1) = 0 < 1$. Now assume our induction hypothesis, that $\log_2(n) \leq n$. We prove $\log_2(n+1) \leq n + 1$. Since $\log_2(n)$ is strictly monotonically increasing, then

$$\log_2(n+1) < \log_2(2n) = \log_2(n) + \log_2(2) = \log_2(n) + 1$$

By the induction hypothesis, since $\log_2(n) \leq n$, then $\log_2(n) + 1 \leq n + 1$. Together, these imply $\log_2(n+1) <\leq n + 1$. Therefore, $\log n$ is $O(n)$. □

Corollary 55. *Notice we actually get the following corollaries as a consequence of the previous theorem*

- $\log n$ *is* $O(n)$
- $n \log n$ *is* $O(n^2)$

Theorem 56.

Computer scientists don't write the base of logs because they are lazy, they don't write the base because it doesn't matter.

Theorem 57. *Let* $k, l \in \mathbb{R}$ *with* $k, l > 1$. *Then*

$$\log_k n \text{ is } O(\log_l n)$$

Proof. Recall the □

Contradiction

Theorem 58. *For any positive real numbers k, l*

$$\log(n)^k \text{ is } o(n^l)$$

Proof. Assume to the contrary that there is some constants c, n_0 for all $n \geq n_0$ that $\log(n)^k \geq cn^l$. Then since $cn^l > n^{l-1}$ for large enough n, we get

$$\log(n)^k \geq cn^l \tag{35}$$
$$\log(n)^k \geq n^{l-1} \tag{36}$$
$$k \log\log n \geq (l-1)\log(n) \tag{37}$$
$$\frac{k}{l-1} \geq \frac{\log n}{\log\log n} \tag{38}$$
$$\frac{k}{l-1} \geq \log n(\log\log n)^{-1} \tag{39}$$
$$\frac{k}{l-1} \geq \log n\left(\log \frac{1}{\log n}\right) \tag{40}$$
$$\frac{k}{l-1} \geq \log\left(n + \frac{1}{\log n}\right) \tag{41}$$

Contradiction, a strictly monotonically increasing function cannot be bounded by a constant. □

The an arbitrary small possible amount of a polynomial, even $n^{0.00001}$ eventually will out grow an arbitrarily large about of a logarithm, even $\log(n)^{10000}$

Theorem 59. *Let $\epsilon > 0$ be a positive real number. Then $2^{(\log n)^{1+\epsilon}}$ grows faster than any polynomial and slower than every exponential.*

Proof. aa □

Counting arguments

You can over apply over counting and under counting arguments to asymptotic relationships among monotonically increasing functions.

Theorem 60. 2^n *is* $O(n!)$

Proof. We proceed by an overcounting argument

$$2^n =$$
$$\underbrace{2 \cdot 2 \cdot 2 \cdot 2 \cdot \ldots \cdot 2}_{n} \leq$$
$$2 \cdot 2 \cdot 3 \cdot 4 \cdot \ldots \cdot n =$$
$$2 \cdot n!$$

So $2^n \leq 2n!$ and we conclude 2^n is $O(n!)$. □

Theorem 61. $n!$ *is* $O(n^n)$

Proof. We proceed by an overcounting argument

$$n! =$$
$$1 \cdot 2 \cdot 3 \cdot 4 \cdot \ldots \cdot n \leq$$
$$\underbrace{n \cdot n \cdot n \cdot n \cdot \ldots \cdot n}_{n} =$$
$$= n^n$$

So $n! \leq n^n$ and we conclude $n!$ is $O(n^)$. □

Recall several of our closed forms of summations of powers.

- $$1 + 2 + 3 + ... + n = \sum_{i=1}^{n} i = \frac{n(n+1)}{2} = \Theta(n^2)$$
- $$1 + 4 + 9 + ... + n = \sum_{i=1}^{n} i^2 = \frac{n(n+1)}{2} = \Theta(n^3)$$
- $$1 + 2 + 3 + ... + n = \sum_{i=1}^{n} i^3 = \frac{n(n+1)}{2} = \Theta(n^4)$$

The only three numbers are $0, 1, n$, so given such a pattern, rather than try to prove a sum of powers to 4, you should try sums of powers to some variable, say t. What could $\sum_{i=1}^{n} i^t$ be in terms of n and t? The closed form is called Faulhaber's formula:

$$\sum_{i=1}^{n} i^t = 1^t + 2^t + ... + n^t = \frac{1}{t+1} \sum_{j=0}^{t} \frac{(t+1)!}{j!(t-j+1)!} B_j n^{t-j+1}$$

where B_j is the jth Bernoulli number. Proving such a formula is extremely difficult for us now. Even defining the Bernoulli numbers is beyond us. In such cases, you need not give up. Where ever you need a theorem, you may not need it in its totality, and should consider settling for proving something sufficient still, but weaker.

Theorem 62.

$$\sum_{i=1}^{n} i^t = \Theta(n^{t+1})$$

Proof. We give overcounting and undercounting arguments.

$$\begin{aligned} \sum_{i=1}^{n} i^t = \\ 1^t + 2^t + ... + n^t \leq \\ \underbrace{n^t + n^t + ... + n^t}_{n} = \\ n(n^t) = n^{t+1} \end{aligned}$$

Thus $\sum_{i=1}^{n} i^t$ is $O(n^{t+1})$. Next we do an undercounting argument

$$\begin{aligned} \sum_{i=1}^{n} i^t = \\ 1^t + 2^t + ... + n^t \geq \\ (n/2)^t + (n/2+1)^t + ... + n^t \geq \\ \underbrace{(n/2)^t + (n/2+1)^t + ... + n^t}_{n/2} = \\ (n/2)(n/2)^t = \\ (n/2)^{t+1} = \frac{1}{2^{(t+1)}} n^{t+1} \end{aligned}$$

Since $\frac{1}{2^{(t+1)}}$ is a constant, we see that $\sum_{i=1}^{n} i^t$ is $\Omega(n^{t+1})$, so we conclude $\sum_{i=1}^{n} i^t$ is $\Theta(n^{t+1})$ □

Relations

Definition 0.0.34. *A relation R is any subset of a cartesian product of sets* $R \subseteq A \times B$

When $(a, b) \in R$, we mean "a relates to b", and write aRb.

For example, consider the relation $R \subseteq A \times B$ with $A = \{0, 1, 2\}$ and $B = \{a, b\}$ with $R = \{(0, a), (1, b), (2, b), (2, a)\}$. [43]

[43] picture here with arrows

Definition 0.0.35. *We say "R is a relation over a set A" to mean* $R \subseteq A \times A$.

The following are some more examples of relations

- The relation "$\leq$" over $\mathbb{R}$ denotes the set $\{(x, y) \in \mathbb{R} \times \mathbb{R} \mid x \leq y\}$ [44]
- $\emptyset$
- The relation R over $\mathbb{N}$ with aRb if $a = b + 1$ denotes pairs of the form $(1, 0), (2, 1), (3, 2), ...$

[44] why not circular

Recall the definition of when a relation is a function

Definition 0.0.36. *We say a relation is a function if whenever ...*

note that if A is a finite set of cardinality n, then the number of relations over A is equal to the number of possible subsets of $A \times A$, so there are 2^{n^2} possible relations.

Definition 0.0.37. *For a relation R over A, we say that R is reflexive if*

$$\forall a \in A[aRa]$$

why we care

Definition 0.0.38. *for a relation R over A, we say that R is symmetric if*

$$\forall a, b \in A[aRb \implies bRa]$$

commentary on symmetric

Definition 0.0.39. *for a relation R over A, we say that R is transitive if*

$$\forall a, b, c \in A[aRb \text{ and } bRc \implies aRc]$$

commentary on transitive

Definition 0.0.40. *We say a relation is an equivalence relation if it is reflexive, symmetric, and transitive.*

Theorem 63. *Let "=" be a relation over* $\mathbb{R}$ *of the axiomatically defined equality of real numbers. Then "* $=$ *" is an equivalence relation.*

Proof. We cannot prove it, as the fact that it is reflexive, symmetric, and transitive are axioms. Convince yourself these are true.

- (reflexive) $\forall x \in \mathbb{R}(x = x)$
- (symmetric) $\forall x, y \in R(x = y \implies y = x)$
- (transitive) $\forall x, y, z \in R(x = y \text{ and } y = z \implies x = z)$

□

Theorem 64. *Let "$\leq$" be a relation over $\mathbb{R}$ of the axiomatically defined equality of real numbers. Then "$\leq$" is reflexive, not symmetric, and transitive.*

Proof. • Let $x \in \mathbb{R}$, then $x \leq x$ since $x = x$

- Consider $x = 2$ and $y = 3$. Notice that $2 \leq 3$ but $3 \not\leq 2$
- This is technically something so basic, it would be acceptable to assume it, but let us try to prove it. Let $x, y, z \in \mathbb{R}$ with $x \leq y$ and $y \leq z$. Since $x \leq y$ there is $\epsilon_1 \in \mathbb{R}$ with $\epsilon_1 \geq 0$ such that $x + \epsilon_1 = y$. Since $y \leq z$ there is $\epsilon_2 \in \mathbb{R}$ with $\epsilon_2 \geq 0$ such that $y + \epsilon_2 = z$. Then

$$\begin{aligned} y + \epsilon_2 &= z \\ (x + \epsilon_1) + \epsilon_2 &= z \\ x + (\epsilon_1 + \epsilon_2) &= z && \text{since } \epsilon_1 + \epsilon_2 \geq 0 \\ x &\leq z \end{aligned}$$

□

Theorem 65. $<$

Theorem 66. *Let R be a relation over the set of all people such that aRb if person a and person b have the same birthday. Then R is an equivalence relation*

Proof. • R is reflexive since everyone has the same birthday as themselves

- R is symmetric, if a has the same birthday as b, then b must have the same birthday as a.
- R is transitive. If aRb and bRc, then all three must have the same birthday, so aRc.

□

Theorem 67. *The relation "$|$" over $\mathbb{Z}$ defined by $a \mid b$ if $\exists c[ac = b]$ is reflexive, not symmetric, and transitive.*

Proof. • (reflexive) Let $a \in \mathbb{Z}$. Since $a \cdot 1 = a$, then $a \mid a$

- (not symmetric) As a counter example, consider $a = 2$ and $b = 6$. Notice that $2 \mid 6$ but $6 \nmid 2$
- (transitive) Let $a, b, c \in \mathbb{Z}$ and let $a \mid b$ and $b \mid c$. Then there exists integers $d, e \in \mathbb{Z}$ such that $ad = b$ and $be = c$. Then

$$c = be = (ad)e = a(de)$$

Since $de \in \mathbb{Z}$, we see that $a \mid c$.

□

Theorem 68. *Let the relation $\sim$ be a relation defined over $\mathbb{Q}$ such that $\frac{a}{b} \sim \frac{c}{d}$ if $\frac{a}{b}$ and $\frac{c}{d}$ have the same reduced form. Then $\sim$ is an equivalence relation.*

Proof. We prove $\sim$ is reflexive, symmetric, and transitive.

- (reflexive) Let $\frac{a}{b} \in \mathbb{Q}$. $\frac{a}{b}$ has the same reduced form as itself so certainly $\frac{a}{b} \sim \frac{a}{b}$
- (symmetric) Let $\frac{a}{b}, \frac{c}{d} \in \mathbb{Q}$, and assume $\frac{a}{b} \sim \frac{c}{d}$. Then we know that $\frac{a}{b}$ has the same reduced form as $\frac{c}{d}$, so certainly $\frac{c}{d}$ has the same reduced form as $\frac{a}{b}$, so we observe that $\frac{c}{d} \sim \frac{a}{b}$.
- (transitive) Let $\frac{a}{b}, \frac{c}{d}, \frac{e}{f} \in \mathbb{Q}$. If $\frac{a}{b}$ has the same reduced form as $\frac{c}{d}$ and $\frac{c}{d}$ has the same reduced form as $\frac{e}{f}$, then certainly $\frac{a}{b}$ has the same reduced form as $\frac{e}{f}$

□

Some of these proofs are so basic, they need not need proof. The previous relation on $\mathbb{Q}$ is really what we use to denote the equality of rationals. Two rationals are not equal if they look the same, but if they have the same reduced form! An equivalence relation is a generalization of the notion of equality of mathematical objects. There are many diverse mathematical objects, and each of these has a notion of when two objects are equal.Two functions are equal if they map all the same elements to the same elements. Two matricies are equal if their elements component wise. Two sets are equal if they contain the same elements. This "sameness" of the elements is really defined up to equality of those elements. Each mathematical object has a slightly different notion of what equality means. Every notion of equality is an equivalence relation. Every equivalence relation can be thought of as a kind of equality,

Theorem 69. *Consider the relation $\equiv$ over the set of all well formed propositions where $\Phi_1 \equiv \Phi_2$ when Φ_1 and Φ_2 have the same truth table*[45]*. Then $\equiv$ is an equivalence relation*

[45] equality of propositions is defined up to the equivalence of their truth tables

Proof. We prove $\equiv$ is reflexive, symmetric, and transitive.

- (reflexive) Every proposition has the same truth table as itself, so every propositions truth table is equal to itself, thus $\Phi_1 \equiv \Phi_1$
- (symmetric) Let Φ_1, Φ_2 be propositions with $\Phi_1 \equiv \Phi_2$. Then they have the same truth table, so certainly $\Phi_2 \equiv \Phi_1$
- (transitive) Let Φ_1, Φ_2, Φ_3 be propositions with $\Phi_1 \equiv \Phi_2$ and $\Phi_2 \equiv \Phi_3$. Then Φ_1, Φ_2, Φ_3 all have the same truth table so $\Phi_1 \equiv \Phi_3$.

□

Theorem 70. $|x - y| < 1$ *over* $\mathbb{R}$

Proof. content... □

Theorem 71. $|x - y| < 1$ *over* $\mathbb{N}$

Proof. content... □

Theorem 72. *linear recurrence like a + 3b is even*

Proof. content... □

Theorem 73. *Let Ω be any universe of discourse and consider a relation $\sim$ over $\mathcal{P}(\Omega)$ with $A \sim B$ if and only if there exists a bijection $f : A \to B$. Then $\sim$ is an equivalence relation.*

Proof. We prove $\sim$ is reflexive, symmetric, and transitive.

- (reflexive) For any set $A \in \mathcal{P}(\Omega)$, there always exists a bijection $f : A \to A$ with f as the identity function, $\forall a \in A(f(a) = a)$. Then $A \sim A$
- (symmetric). For any sets $A, B \in PS(\Omega)$, let $A \sim B$. Then there exists a bijection $f : A \to B$. We previously proved the inverse of a bijection is also a bijection, so there exist a bijection $f^{-1} : B \to A$, so $B \sim A$.
- Let $A, B, C \in \mathcal{P}(\Omega)$ and let $A \sim B$ and $B \sim C$. Then there exists bijections $f : A \to B$ and $g : B \to C$. We previously proved that the composition of bijections is a bijection so $g \circ f : A \to C$ is a bijection from A to C, so $A \sim C$.

□

our most important one

Theorem 74. *Let the relation* $(\text{mod } n)$ *be defined over $\mathbb{Z}$ denoted by $a \equiv b \pmod{n}$*[46] *if $n \mid (a - b)$. Then modular equivalence is an equivalence relation*

[46] Sometimes, also denoted as $a \equiv_n b$

Proof. We prove $a \equiv b \pmod{n}$ is an equivalence relation

- Since $n \cdot 0 = 0$, we know $n \mid 0$ for any integer n, so $n \mid (a - a)$ and $a \equiv a \pmod{n}$.
- Since

□

Equivalence Classes

Definition 0.0.41. *equivalence classes*

examples

Theorem 75. *Let R be an equivalence relation on A. Then the equivalence classes form a partition of the set A.*

Proof. We need to prove the equivalence classes cover the set A, and are all pairwise disjoint. □

examples

Modular Arithmetic

Definition 0.0.42. *base representation as a sum of powers*

Theorem 76. *Let $n, b \in \mathbb{N}$ and $b \geq 2$. Then n can be uniquely represented in base b.*

Proof. We first prove the basis representation exists, then we show its is unique

We proceed by induction on n. Our base case is $n = 0$ and we see that 0 can be written in base b as "0". Now suppose our induction hypthesis, that k can be written in base b. We prove $k + 1$ can be written in base b. Since k can be written in base b, there exists $d_0, ..., d_l \in \{0, ..., b-1\}$ such that

$$k = d_l b^l + ... + d_1 b + d_0$$

Consider

$$k + 1 = d_l b^l + ... + d_1 b + d_0 + 1$$

Let i be the smallest number such that d_i is the first digit of k such that $d_i < b - 1$. Then let $d_i' = d_i + 1$ and $d_i - 1 = ... = d_0 = 0$. Then

$$\begin{aligned} k + 1 &= (d_l b^l + ... + d_i b^i + ... + d_1 b + d_0) + 1 \\ &= d_l b^l + ... + (d_i + 1) b^i + 0 \cdot b^{i-1} + ... + 0 \cdot b + 0 \\ &= d_l b^l + ... + d_i' b^i + ... + d_1' \cdot b + d_0 \end{aligned}$$

which is a basis representation of $k + 1$, thus by induction we have proven all $n \in \mathbb{N}$ has a representation in base b.

Now we prove uniqueness. Suppose n has two representations

$$\begin{aligned} n &= d_l b^l + ... + d_2 b^2 + d_1 b + d_0 \\ n &= d_l' b^l + ... + d_2' b^2 + d_1' b + d_0' \end{aligned}$$

Without loss of generality, we may assume these representations have the same length (for if they don't, add leading zeroes to the shorter one). We prove these representations are the same, and that $0 \leq i \leq l$ that $d_i = d_i'$. Since these two representations are for the same number, they are equal. What happens if we take both equations mod b?

$$\begin{aligned} d_l b^l + ... + d_2 b^2 + d_1 b + d_0 &= d_l' b^l + ... + d_2' b^2 + d_1' b + d_0' \\ b(d_l b^{l-1} + ... + d_2 b + d_1) + d_0 &= b(d_l' b^{l-1} + ... + d_2' b + d_1') + d_0' \\ d_0 &\equiv d_0' \pmod{b} \end{aligned}$$

Since $d_0, d_0' \in \{0, ..., b-1\}$, not only are they congruent mod b, they are equal. Thus $d_0 = d_0'$.

$$b(d_l b^{l-1} + ... + d_2 b + d_1) + d_0 = b(d_l' b^{l-1} + ... + d_2' b + d_1') + d_0'$$
$$b(d_l b^{l-1} + ... + d_2 b + d_1) = b(d_l' b^{l-1} + ... + d_2' b + d_1')$$
$$d_l b^{l-1} + ... + d_2 b + d_1 = d_l' b^{l-1} + ... + d_2' b + d_1'$$

Were we to mod by b now we would observe $d_1 = d_1$. We repeat this argument $l-1$ additional times and see that for all $0 \leq i \leq l$ that $d_i = d_i'$, and thus, every numbers representation is unique. $\square$

Cardinalities of Sets

Recall the main motivation of the class. We wish to understand the correspondence between infinite objects and their finite descriptions. Previously, we have studied the finite objects quite well. Let us now undertake a rigorous understanding of the infinite.

A quote often misattributed to Aristotle is that "*The whole is greater than any part*". Ancient Greek philosophy is often too vague to argue with. It certainly appears true. We may formalize this notion with sets.

Definition 0.0.43 (Aristotelian Property)**.** *If $A \subsetneq B$ then $|A| \lneq |B|$*

Despite this, Galileo discovered what he called a paradox. The squares of numbers could be put into correspondence with the numbers.

1	2	3	4	5	...	n	...
1	4	9	16	25	...	n^2	...

The squares are part of the numbers, yet by putting them in a 1-1 correspondence, it was apparent that there are as many squares as there are numbers. A strict part could be equal to the whole. Was this a counter-example to the Aristotelian property? The Aristotelian property is generalized from the intuition about *finite* sets. Galileo here, assumes that a "whole" could be an infinite set.[47] The Aristotelian property is not true for infinite sets. But can a set even be infinite?

[47] Both Aristotle and Galileo did not know what a set was.

Today, without hesitation, we may consider sets to contain infinitely many elements, but this was not always the case historically. Infinity used to be just a figure of speech, or perhaps a useful abstraction, not a real thing. It was understood that you could not discuss $\mathbb{N}$ as a set, only as the outcome of a iterative, never ending process. The natural numbers are constructed by induction.

- $0 \in \mathbb{N}$
- $\forall n \in \mathbb{N} \implies S(n) \in \mathbb{N}$

Because this process is ceaseless, it does not make sense to discuss $|\mathbb{N}|$, much like it doesn't make sense to discuss $f(x) = 1/x$ evaluated at $x = 0$. The sequence 0,1,2,... could be discussed, but not the set $\{0, 1, 2, ...\}$ The infinite could only be discussed in terms of limits??, and never addressed directly. Georg Cantor disagreed. In the late 19th century, he undertook a serious attempt to formalize and understand the infinite, generalizing ideas from finite sets to infinite ones.

Generalizing our Intuition

We denote the cardinality of the set S as $|S|$. If S is finite then $|S|$ is just the size, the number of elements. But what is the cardinality of the natural

numbers $|\mathbb{N}|$? Certainly for all finite sets F, it is true that $|F| < |\mathbb{N}|$. When we talk about the cardinality of infinite sets, we want to preserve our intuition as much as possible. If A is a subset of B then $A \subsetneq B \implies |A| \leq |B|$. We have observed for infinite sets that we do not have the Aristotelian property: $A \subsetneq B \not\Rightarrow |A| \lneq |B|$ when A, B are infinite.

Definition 0.0.44. *We say a set S is* countable *if $|S| \leq |\mathbb{N}|$. All finite sets are countable. We say a set is* countably infinite *if $|S| = |\mathbb{N}|$.*

How can we show that a set has the same cardinality as natural numbers?

Definition 0.0.45 (Injection)**.** *Recall $f : A \to B$ is one to one (injective) if $f(a) = f(b) \implies a = b$.*

Definition 0.0.46 (Surjection)**.** *Recall $f : A \to B$ is onto (surjective) if $\forall y \ \exists x$ such that $f(x) = y$. There do not exist any unmapped elements in the co-domain.*

Definition 0.0.47 (Bijection)**.** *A function is said to be bijective if it is both injective and surjective.*

See how both 3 and 4 map to the same element? That makes this function **not** injective. See how A is unmapped? That makes this function also **not** surjective. Bijections gives us a natural "same size-ness" because if there is a bijection between two sets, the elements seem to pair up nicely, meaning they should intuitively have the same size.

Definition 0.0.48. *We say a set S is countably infinite if $\exists\ f : \mathbb{N} \to S$ which is a bijection. Recall the inverse of a bijection is also a bijection so equivalently if $\exists\ g : S \to \mathbb{N}$ which is bijective.*

To prove that a set is countably infinite, you need only to put it in correspondence with the naturals, like Galileo did.

Examples of Countably Infinite Sets

Those "other naturals"

Other groups may not consider zero to be a natural. Lets prove it doesn't really matter, $|\mathbb{N}| = |\mathbb{N} \setminus \{0\}|$. Recall that $\mathbb{N} = \{0, 1, 2, 3, ...\}$ and $\mathbb{N} \setminus \{0\} = \{1, 2, 3, ...\}$. To prove these sets have the same cardinality, we give an obvious bijection. Namely $f : \mathbb{N} \to \mathbb{N} \setminus \{0\}$ by $f(n) = n + 1$. The elements pair up obviously like $0 \to 1, 1 \to 2$ and so on, so our function is certainly bijective. This shows that if you add or remove a constant amount of elements from a countably infinite set, its still countably infinite.

Theorem 77. *The function $f : \mathbb{N} \to \mathbb{N} \setminus \{0\}$ defined by $f(n) = n + 1$ is a bijection*

Proof. We first prove f is injective. Let $f(n) = f(m)$. We prove $n = m$

$$f(n) = f(m)$$
$$n + 1 = m + 1$$
$$n = m$$

Now we prove f is surjective. Let $r \in \mathbb{N} \setminus \{0\}$. We prove there exists $N \in \mathbb{N}$ such that $f(n) = r$. Consider $n = r - 1$. Since $r > 0$, we know $r - 1 \in \mathbb{N}$. Then

$$f(n) = f(r - 1) = (r - 1) + 1 = r$$

Since we have proven f is injective and surjective, it is bijective. □

The Evens

What is the cardinality of the even numbers? Define $2\mathbb{N} = \{0, 2, 4, 8, ...\}$. Our bijection is again obviously $f(n) = 2n$. This shows that "half" of a countably infinite set is still countably infinite.

The Integers

Recall $\mathbb{Z} = \{-2, -1, 0, 1, 2, ...\}$. When you are asked to give a bijection, it is equivalent to showing that you can order the elements of a set in some way. Intuitively, if you can "count" them. A bad idea is to first order the elements like 0,1,2,... because then we will never reach the negative numbers. Since -1 never appears in this ordering, the map is not surjective. A better idea is to dovetail the negative and positive integers in the following way.

If you were to actually work out what this bijection would be like functionally, you would get

$$f(n) = \begin{cases} \frac{-n}{2} & n \text{ is even} \\ \frac{n+1}{2} & n \text{ is odd} \end{cases}$$

You should convince yourself it is a bijection. Since every number appears atleast once in the ordering, it is surjective. Since each number appears no more than once, it is injective, so it must be bijective.

The integers feel like "twice as many" of the naturals so this can show that if you "double" a countably infinite set, it remains countably infinite. A countably infinite set also need not be well-ordered, it need not have a least element.

The Rationals

We define the rational numbers[48] as $\mathbb{Q} = \{ \frac{a}{b} \mid a, b \in \mathbb{N}; a, b \neq 0\}$. They do not contain repetitions, so $\frac{1}{1}, \frac{2}{2}$ are not distinct. Rational numbers have some very different properties than the previous examples. For example for the natural numbers, there is only a finite number of naturals between any two naturals, but this isn't true for the rationals.

[48] The rationals can actually contain negatives and zero but suppose that we are only in consideration of the positive non-zero rationals

- $\exists x \in \mathbb{N}$ where $0 < x < 1$? no
- $\exists x \in \mathbb{Q}$ where $\frac{a}{b} < x < \frac{c}{d}$? yes

The naturals appear in discrete steps, but between any two rationals, there exists an infinite amount of rationals. Why? The average of any two rationals is a rational, so the midpoint between any two, you will find a rational[49]. Recursively applying this idea will give you an infinite amount between any two! The mathematically correct term for this is "dense". Could there be more rationals than naturals? It feels like there is a lot more of them. It turns out even despite this, the rationals are still only countably infinite, that $|\mathbb{N}| = |\mathbb{Q}|$. This bijection is a little less obvious. Put all the representations of rationals into a table with columns and rows ordered by numerators and denominators. This table contains duplicates since 1/1 and 2/2 are the same rational. A bad idea would be to try to go left to right row by row. You would never reach the second row. The idea is then to compose the anti-diagonals ignoring duplicates!

[49] If you wanted to work it out, the rational between $\frac{a}{b}, \frac{c}{d}$ is $\frac{a}{b} + (\frac{c}{d} - \frac{a}{b})/2$. You could simplify that with arithmetic it into numerator/denominator form.

This certainly is a bijection. Its surjective since every element is hit somewhere in this criss-crossing, since every element is on some anti-diagonal. Its injective as every element only can appear once in this ordering since we define it to ignore duplicates.

Here's another solution. Consider the function $f(a/b) = 2^a 3^b$. This function is bijective to some set $S = \{2^i 3^j \mid i, j \in \mathbb{N}_{\geq 1}\}$. Notice that $|\mathbb{Q}| = |S|$. Also notice that since $S \subseteq \mathbb{N}$ then $|S| \leq |\mathbb{N}|$. So by transitivity $|\mathbb{Q}| = |S| \leq |\mathbb{N}| \implies |\mathbb{Q}| \leq |\mathbb{N}|$. We also know that $|\mathbb{N}| \leq |\mathbb{Q}|$ by the injection $f(a) = \frac{a}{1}$ so combined we see that $|\mathbb{Q}| = |\mathbb{N}|$. We could have also just observed that since $|\mathbb{Q}| \leq |\mathbb{N}|$, we know $\mathbb{Q}$ is countable, as subsets of countable sets are countable. Observing that $\mathbb{Q}$ is infinite is enough to show it must be countably infinite.

Cartesian Products

The rationals are really just like, pairs of numbers. If we are tasked with finding a bijection $f : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$, we can immediately apply the same argument with the table and anti-diagonals. This is enough to prove that the cartesian product of two countable sets is countable. We can also immediately induct this argument to get that finitely many cartesian products of countable sets is countable. Notice that $\mathbb{N} \times \mathbb{N} \times \mathbb{N} = (\mathbb{N} \times \mathbb{N}) \times \mathbb{N}$. We know that $\mathbb{N} \times \mathbb{N}$ is countable. It remains countable if we perform one more cartesian product, and so on.

Hillbert's Hotel

Suppose we have an infinitely tall hotel of countably infinite rooms. Each room already has a guest, so the hotel is full.

- A single new guest arrives. Although every room already has a guest, the hotel staff aren't worried. They make each old guest move from room n into the next room, room $n + 1$. Now room zero is empty for the new guest.
- Suppose an infinitely long bus arrives with a countably infinite number of new guests. Even though the hotel seemingly has no space, the new guests can still be accommodated. Tell each old guest to move from room n to room $2n$, then each of the new guests to move into the now empty odd-numbered rooms.
- What if a countably infinite number of infinitely long busses arrive, each with countably infinite number of guests? I claim they can still be accommodated, and I leave it to you as an exercise to figure out how.

Cantor's Theorem

It would seem that you can play with infinity in most ways and remain countably infinite. If we were to say that $|\mathbb{N}| = \infty$, then it would seem that $\infty + 3, 3 \cdot \infty, \infty^3$ all equal to ∞. These are all polynomially related. Could it be the case that $2^\infty = \infty$? It turns out, no. Lets denote $|\mathbb{N}| = \aleph_0$ and $2^{\aleph_0} = \aleph_1$. We will show these are two very different infinites. We do not use ∞, as we need a way to

distinguish between the kinds of infinite. We represent these as cardinals, $\aleph_0, \aleph_1$, and so on. These are not numbers, they are cardinals. Cantor's theorem tells us that there is no bijection between any set, and its power set[50].

[50] Recall that a power set is the set of all subsets of a set

Theorem 78 (Cantor)**.** *If A is any set and $\mathcal{P}(A)$ is the set containing all subsets of A then*

$$|A| \lneq |\mathcal{P}(A)|$$

Note that its obviously true for finite sets. if $A = \{x, y\}$ then $\mathcal{P}(A) = \{\emptyset, \{x\}, \{y\}, \{x, y\}\}$ and $|\mathcal{P}(A)| = 2^{|A|}$.

Diagonalization

Diagonalization is an extremely important proof technique. Perhaps one of the most important in history. We will prove Cantor's theorem several times to emphasize diagonalization. First we prove Cantor's theorem for the special case of the naturals.

Theorem 79.

$$|\mathbb{N}| \lneq |\mathcal{P}(\mathbb{N})|$$

Proof. Assume to the contrary that there does exist a bijection $f : \mathbb{N} \to \mathcal{P}(\mathbb{N})$. Then there exists a way to totally order the subsets of $\mathbb{N}$ like $S_0, S_1, S_2, ...$ where every possible subset of the naturals appear in this ordering exactly once. Consider the set D defined such that for all $i \in \mathbb{N}$

$$i \in D \iff i \notin S_i \tag{42}$$
$$i \notin D \iff i \in S_i \tag{43}$$

We go to the ith set in the ordering, see if it contains the ith number, and if it does, we define it not to be in D, and if it doesn't, we define D to include it. Notice that D is a set of numbers, so $D \subseteq \mathbb{N}$, or that $D \in \mathcal{P}(\mathbb{N})$. Then there exists a spot for it in line in our total ordering. There exists a number j such that $D = S_j$. Two sets are equal if they contain the same elements, so we know that

$$j \in D \iff j \in S_j$$

But by the definition of how we defined D, we know that

$$j \in D \iff j \notin S_j$$

Combining these, we see

$$j \in S_j \iff j \notin S_j$$

A contradiction. □

Let us remark on the proof. It is important that you understand that the proof is not circular. A circular proof is one which assumes its conclusion as a premise, so it demonstrates nothing. Nothing circular is going on here, but something else definitely is; self reference. The set D is defined against the ordering of subsets $S_0, S_1...$ and for a different ordering, there must be a different D. To further illustrate the mechanics of this proof technique, lets do it again. We will prove Cantor's theorem for the countably infinite case using diagonalization.

Theorem 80. *Let A be any countably infinite set. Then*

$$|A| \lneq |\mathcal{P}(A)|$$

Proof. First we define the characterisitic sequence of a subset. Assume that A is countably infinite. Then its elements may be ordered like $a_0, a_1, ...$ If $S \subseteq A$, to S we associate the infinitely long binary sequence $\chi \in \Sigma^\infty$ such that

$$\chi[i] = \begin{cases} 1 & a_i \in S \\ 0 & a_i \notin S \end{cases}$$

For example

- if $S = \{0, 3, 4\}$ then $\chi = 10011000000...$
- if $S = 2\mathbb{N}$ then $\chi = 10101010...$
- if $S = \mathbb{N}$ then $\chi = 11111...$
- if $S = \emptyset$ then $\chi = 00000...$

Notice immediately that to each subset, corresponds a unique characteristic sequence. There is a bijection between the set of infinitely long binary sequences, and the subsets of a countably infinite set. The infinite sequence of bits exactly characterizes which elements are and aren't in a subset. What is a subset if not just a selection of the elements? It is important to remember that these sequences are infinitely long.

Let us proceed with the proof. Assume to the contrary that there exists some bijection $f : A \to \mathcal{P}(A)$ with A countably infinite. The elements of $\mathcal{P}(A)$ are exactly the subsets of A. So then there exists an ordering of the elements of $\mathcal{P}(A)$ like $S_0, S_1, S_2, ...$, where every element is in this ordering. Let $\chi_0, \chi_1, \chi_2, ...$ be the characteristic sequences of $S_0, S_1, S_2, ...$ ordered in the same way. We define "the diagonal" D to be the infinite binary sequence with digits defined as

$$D[i] = 1 - \chi_i[i] = \overline{\chi_i[i]}$$

We take our ordering of characteristic sequences, find the ith one, find its ith digit, and then set the ith digit of D to be the exact opposite of that. D certainly is an infinite binary sequence, so it must be the characteristic sequence of some subset. Since f is bijective, D exists somewhere in our ordering $\chi_0, \chi_1, \chi_2, ...$ There exists a number j such that $D = \chi_j$ What is the jth bit of D? $D[j]$? Well, since $D = \chi_j$ then $D[j] = \chi_j[j]$. But recall how we originally defined D, where $D[j] = 1 - \chi_j[j]$. Together, these imply that

$$\chi_j[j] = \overline{\chi_j[j]}$$

A digit cannot be zero and one simultaneously! Therefore, we see that we have reached a contradiction, and $|\mathcal{P}(A)|$ is not countable. □

Why is it called diagonalization? Well suppose you listed $\chi_0, \chi_1, ...$ into a table with each χ_i as a row:

χ_0	**0**	1	1	0	1	1	0	...
χ_1	0	**1**	0	0	0	0	0	...
χ_2	0	0	**0**	0	1	1	1	...
χ_3	0	0	1	**1**	0	0	1	...
χ_4	0	1	0	1	**1**	0	1	...
χ_5	0	0	1	1	1	**0**	0	...
...	...							

Then $D = 101001...$ is the opposite of the diagonal of the table. Since D is different than any row of the table, it exists no where in the table. For each row, it is defined to be different in atleast one place, namely the diagonal (i, i) but maybe more. Could it be χ_3? No because $\chi_3[3] = 1$ and $D[3] = 0$. Could it be χ_4? no, and so on. We assumed to the contrary that these sequences were countable and that we can order them, but no matter how we order them, we can always construct an element not in the ordering. So there can never exist a bijection $f : A \to \mathcal{P}(A)$.

This is where the term diagonalization comes from. The element you are constructing is the negation of the diagonal of this implicit table. The diagonal entries are at coordinate (i, i) and the elements of D are defined considering if $i \in S_i$ or not. It is important that you understand that the diagonalization technique does have this nice visualization, but the technique goes far beyond this. When doing a proof by diagonalization, do not draw a table and define its elements. The table is completely implicit. This is important. See the first proof we did of diagonalization, in the proof by contradiction, the absurdity we derived was completely based on logic, and made no reference to a table. You will apply diagonalization to things which cannot be nicely visualized as a table. Let us now prove Cantor's theorem in the general case, so it may apply even when A is uncountable.

Theorem 81 (Cantor's Theorem)**.** *If A is any set, then*

$$|A| \lneq |\mathcal{P}(A)|$$

Proof. Assume to the contrary that there is a bijection between A and $\mathcal{P}(A)$ Consider the subset of A such that the bijection f doesn't map x to a set containing x.

$$D = \{x \in A \mid x \notin f(x)\}$$

Since $D \subseteq A$, we know $D \in \mathcal{P}(A)$. Since f is a bijection, it is also surjective, so there must exist some j such that $D = f(j)$. Is $j \in D$? Two sets are equal if they have the same elements

$$j \in D \iff j \in f(j)$$

But by definition of D, we have that

$$j \in D \iff j \notin f(j)$$

contradiction.

□

For any set A, there always exists an injection $f : A \to \mathcal{P}(A)$, namely $f(x) = \{x\}$. We proved there cannot exist a bijection, but if there is always an injection, that means there is never a surjection. If every map from a set to its powerset is not surjective, so there is an element of the codomain always goes unmapped for every map. If cardinality is supposed to be an extension of the intuition about size, $\mathcal{P}(A)$ is "bigger" than A, even when A is infinite. There are atleast two infinities! In fact, we can remark that there are atleast a countably infinite number of infinities.

$$|\mathbb{N}| \lneq |\mathcal{P}(\mathbb{N})| \lneq |\mathcal{P}(\mathcal{P}(\mathbb{N}))| \lneq |\mathcal{P}(\mathcal{P}(\mathcal{P}(\mathbb{N})))| \lneq ...$$

Uncountability

We have now shown that $\mathcal{P}(A)$ is not countably infinite when A is countably infinite, it is something greater. We call these sets uncountable. Intuitively, a countably infinite set is one in which you can "count". It feels infinite in a discrete sense. At some element, you can choose a next one. Conversely, an uncountable set is literally "uncountable". Imagine a stream of water. What are the units? What is the "next" water?[51]. It feels infinite in a continuous sense. By a similar diagonalization argument, you can prove the real interval $(0, 1]$ is uncountable, by diagonalizing over the decimal expansions beginning with zero[52]. Given that $(0, 1]$ is uncountable, you can prove that $\mathbb{R}_{\geq 0}$ is uncountable by the bijection $f(r) = 1/r - 1$. Essentially you can stretch the unit sized interval over the entire real positive line.

[51] If you recall that water is atoms then technically water is discrete and countable but the intuition is there even if the science isn't

[52] recall that $0.\overline{9} = 1$. There are a few proofs of this. One is that $1 - 0.999... = 0.0000...$ and another is to notice that $0.999... = 0.333... + 0.333... + 0.333... = 3(0.333...) = 3\frac{1}{3} = 1$

How to Prove Countability

Union of Two Countable Sets

Let A, B be countably infinite. Then there exists bijections $f : A \to \mathbb{N}$, $g : B \to \mathbb{N}$. We give a bijection for $A \cup B$ as

$$h(x) = \begin{cases} 2f(x) & \text{if } x \in A \\ 2g(x) + 1 & \text{if } x \in B \end{cases}$$

We leave it to you as an exercise to show its bijective, reducing to the bijectivity of f, g.

Countable Union of Countable Sets

A countable union of countable sets is countable. Most unions you have ever seen have been countable. They index over $\mathbb{N}$ with $i = 0, 1, 2, 3, ...$ but the index set of a union need not be countable in general. Consider

$$\bigcup_{x \in \mathbb{R}} \{x\} = \mathbb{R}$$

Here we index over $\mathbb{R}$, an uncountable set. Each element is a singleton containing just x, it is finite and therefore countable. But our union is over $\mathbb{R}$, uncountable. We have an uncountable union of countable sets, yet, it is uncountable.

Lets prove that a countable union of countable sets is countable. Let A be countable and each S_i be countable. Consider the map

$$f : A \times \mathbb{N} \to \bigcup_{i \in A} S_i$$

Such that $f((i, j))$ maps to the j'th element of S_i. Note that this map is surjective so $|A \times \mathbb{N}| \geq |\cup_{i \in A} S_i|$ and since $A \times \mathbb{N}$ is countable, so is our countable union.

Three solutions

Let's do a problem. Let $\mathbb{N}^*_{\geq 1}$ be the set of finite sequences of natural numbers greater than one. It may contain things like $[1, 11, 1]$ or $[23, 100, 18]$ and so on. We give three solutions to showing this set is countably infinite.

- Let $A_i =$ sequences which sum to i, for example A_3 would contain $[1,1,1],[3],[2,1]$ and so on. Since each sequence sums to something, The A_i's partition $\mathbb{N}^*_{\geq 1}$

$$\mathbb{N}^*_{\geq 1} = \bigcup_{i=1}^{\infty} A_i$$

  Notice that each A_i is finite, so countable. Then $\mathbb{N}^*_{\geq 1}$ is a countable union of countable sets, so its countable.

- consider the map: $F([x_1, x_2, x_3, ..., x_k]) = 2^{x_1}3^{x_2}5^{x_3}...p_k^{x_k}$ or more generally

$$F([x_1, ..., x_k]) = \prod_{i=1}^{k} p_i^{x_i}$$

  where p_i is the i'th prime. By the fundamental theorem of arithmetic, every number has a unique prime factorization, and this immediately gives us that our map is injective. Suppose two sequences exist a, b with $F(a) = F(b)$. Then they are divisible by exactly the same power of two, so then they share the same first element, x_1. Repeating this argument we see that $a = b$. Therefore, we have an injection $F : \mathbb{N}^*_{\geq 1} \to \mathbb{N}$ which implies that it is countable.

- There is a injection hiding in us all along. What is the difference between the two sequences $[1,1]$ and $[2,3,4]$? Is it the length? Is it the number of elements? I put these on the board, you immediately know that the sequences are different. You didn't check the lengths or the elements, so how you did you know? The answer is that the two sequences are different because they look different! Define our injection $f(a) =$ "a". That is, it is the string casting function. Now its certainly true that "$[1,1]$" $\neq$ "$[2,3,4]$". We observe that $f(\mathbb{N}^*_{\geq 1}) \subseteq \Sigma^*$ and subsets of countable sets are countable. Why is Σ^* countable? It is the countable union of countable sets. Recall

$$\Sigma^* = \bigcup_{i=0}^{\infty} \Sigma^i$$

This last point leads us to a powerful theorem called the **Typewriter principle**:

Theorem 82. *If some set S has elements $a \in S$ where every element can be* uniquely *described by a string. Then S is countable.*

Proof. If every element of S can uniquely be described by a string, then $f : S \to \Sigma^*$ is injective. The image $f(S)$ is a set of strings, so $f(S) \subseteq \Sigma^*$ and f is certainly bijective to $f(S)$ so we see that S is bijective to a subset of a countable set, and is therefore, countable. □

This is not sufficient to show uncountability. Showing some elements of a set have some infinite encoding isn't enough, since you must also show that there does not exist a unique finite encoding. This turns out to be as hard as finding a bijection. Please only use it to show countability. Also take naturals of the contrapositive. If a set is uncountable, its elements may not all be able to be uniquely described by strings.

We now have an entire toolbox to show a set is countable. Let C be any countable set, and we want to show S is countable. We can do any of the following

- Give a bijection $f : C \to S$
- Give a bijection $f : S \to C$
- Give an injection $f : S \to C$
- Give a surjection $f : C \to S$
- Give an ordering of every element where no element appears twice
- Show that S is a subset of some countable set, since

$$S \subseteq C \implies |S| \leq |C| \implies |S| \leq \aleph_0$$

- Show that S is representable as a countable union of countable sets
- Arrange the elements of S into a grid and compose the anti-diagonals, or some other pattern to implicitly give a bijection
- Show it is the closed under operations we know do not change the cardinality of the set, for example $S = (\{C \times C\} \cup \{0, 1\})^k$.
- Show that its elements can be uniquely represented as finite length strings and apply the typewriter principle.

Common known countable sets include $\mathbb{N}, \mathbb{Z}, \Sigma^*, \mathbb{N}^*$, and so on. Every language is also countable, as it is a subset of a countable set.

How to prove Uncountability

Let U be some known uncountable set. We give several ways to show a set S is uncountable.

- Diagonalization
- Find a bijection $f : S \to U$
- Show that S contains some uncountable subset. Since if $U \subseteq S$ is uncountable then $|U| \leq |S| \implies \aleph_1 \leq |S|$
- Find an injection $f : U \to S$
- Apply Cantor's theorem, show that it is the powerset $\mathcal{P}(A)$ of some infinite A

We do have far fewer ways to show a set is uncountable than to show a set is countable. Which tool you use depends on ease of use. Common uncountable sets include $\mathcal{P}(\Sigma^*), \mathcal{P}(\mathbb{N}), \mathbb{R}, \mathbb{C}$, and others.

Disivibility of Integers

In number theory, we usually don't like division as an operation. The integers $\mathbb{Z}$ are closed under addition,multiplication, and subtraction, but not division. $10, 7 \in \mathbb{Z}$ but $\frac{10}{7} \notin \mathbb{Z}$. Rather than speak of division as a some fractional piece, we speak of division as a whole and a remainder. To a number theorist, you divide a number into a pair of numbers, a part and a remainder. For example

$$\frac{30}{7} = \frac{4 \cdot 7 + 2}{7} = 4 + \frac{2}{7}$$

Rather than discuss $\frac{30}{7}$ as a piece of something, we would rather say that 30 divided by 7 is the pair of numbers $(4, 2)$. Lets prove such a pair always uniquely exists.[53]

[53] Why do we care about such a form? Well if $n = dq + r$ then

$$\frac{n}{d} = \frac{dq + r}{d} = q + \frac{r}{d}$$

We refer to d as the divisor, q as the quotient, and r as the remainder. Note that this theorem is often called "the division algorithm", but its not an algorithm.

Theorem 83 (The Division Theorem)**.** *Let n, d be positive numbers. Then there exists unique positive numbers q, r such that $n = qd + r$ and $0 \leq r < d$*

Proof. Let

$$S = \{n - dx \mid x \in \mathbb{Z} \text{ and } b - ax \geq 0\}$$

First observe that S is not empty. Choose $x = 0$ and observe that $n \in S$. Next, observe that S only contains natural numbers. By closure, it only contains integers, and it is conditioned on containing non-negative integers. Therefore, $S \subseteq \mathbb{N}$. By the Well-Ordering Principle, every non-empty subset of the natural numbers contains a least element. Let this least element be r. Since $r \in S$, there exists $q \in \mathbb{Z}$ such that $r = n - dq$, or that $n = qd + r$. Now we need to prove that $0 \leq r < d$. We know that since $S \subseteq \mathbb{N}$, and $r \in S$ that $r \in \mathbb{N}$, so $r \geq 0$.

We only need to show that $r < d$. Assume to the contrary that $r \geq d$. Then there is some integer $t \geq 0$ such that $r = d + t$, where t balances the inequality to become an equality. Then $t = r - d$. By assumption that $r \geq d$, and that d is positive, we observe that $t < r$. But then

$$t = r - d = n - dq - d = n - d(q + 1)$$

Since $q + 1 \in \mathbb{Z}$ and $t \geq 0$, this implies that $t \in S$, but if $t < r$, this contradicts definition of r as the least element of S.

We have proven existence. Now we prove uniqueness. Suppose that for positive numbers n, d there exists distinct pairs (q, r) and (q', r') such that $n = dq + r$ and $n = dq' + r'$. We prove $q = q'$ and $r = r'$. Without loss of generality, suppose $r' \geq r$. Since these are two divisor forms of n, they are equal.

$$n = dq + r = dq' + r' \tag{44}$$
$$dq - dq' = r' - r \tag{45}$$
$$d(q - q') = r' - r \tag{46}$$
$$\tag{47}$$

Then $d \mid (r' - r)$, but since $d > (r' - r)$ it must be that $r' - r = 0$, or $r = r'$. Then

$$d(q - q') = r' - r \tag{48}$$
$$d(q - q') = 0 \tag{49}$$
$$q - q' = 0 \tag{50}$$
$$q = q' \tag{51}$$

Thus, $q = q'$ and $r = r'$ and we see our pair (q, r) is unique. □

Greatest Common Divisor

Definition 0.0.49. *Let a, b be two integers both not zero. We define the greatest common divisor of a, b as a number $d \geq 1$ such that $d \mid a$ and $d \mid b$. We denote the greatest common divisor of a, b as* $\gcd(a, b)$

$\gcd(a, b)$ computes the greatest common divisor of a, b. For example, $gcd(105, 30) = gcd(3 \cdot 5 \cdot 7, 2 \cdot 3 \cdot 5) = 3 \cdot 5 = 15$. You can analogously think of gcd like a set intersection. What number is greatest to divide into both a, b? If $a = 2^2$ and $b = 2^3$, then their greatest common divisor must be 2^2.

The following is an easy divide and conquer algorithm discovered long ago by Euclid to calculate gcd of any two numbers.

```
function gcd(a, b)
    if b = 0
        return a
    else
        return gcd(b, a mod b)
```

We can prove correctness by proving that $gcd(a, b) = gcd(a, a - b)$. Repeatedly subtracting b from a will give you $a \pmod b$. We will show these two numbers to be equal by proving that they divide each other. If two numbers divide each other, they must be equal, as a number is greater than or equal to any of its factors.

Theorem 84.

$$\gcd(a, b) = \gcd(b, a - b)$$

Proof. Let $d = gcd(a, b)$. If $d \mid a$, and $d \mid b$, then there exists integers k, l such that $a = dk$ and $b = dl$. So, $a - b = dk - dl = d(k - l)$. Since $a - b$ is an integer, then $d \mid (a - b)$. So, $d \mid gcd(b, a - b)$.

Let $gcd(b, a - b) = d'$. If $d' \mid b$, $d' \mid (a - b)$, then there exists integers s, t such that $d's = b$ and $d't = a - b$. Then $a = (a - b) + b = d't + d's = d'(t + s)$. Since $t + s$ is an integer, then $d'|a$. Since $d'|a$ and $d'|b$, then $d'| \gcd(a, b)$. Since d, d' divide each other and are nonzero, it can only be the case that $d = d'$. □

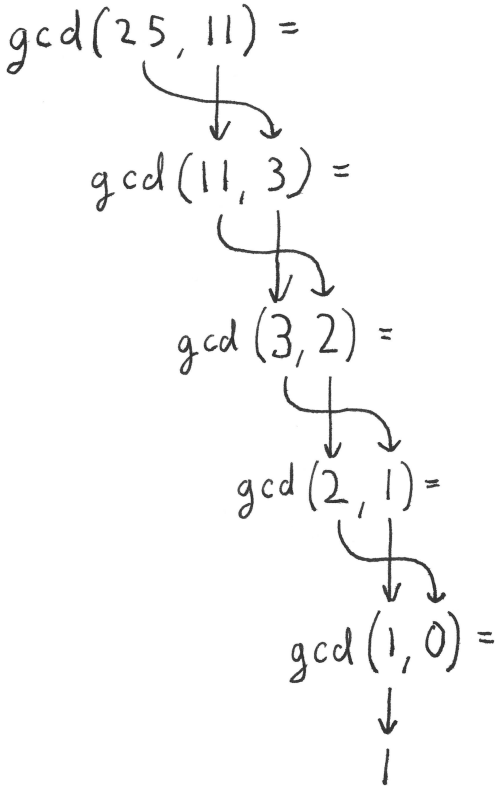


Figure 10: You should think of the execution of the euclidean algorithm as a swapping of pairs for a smaller pair of numbers with the same gcd.

To write the execution of the algorithm, put the larger number on the left hand side, and represent it in division form $a = bq + r$. Then repeatedly chain down.

$$\begin{aligned} \gcd(25, 11) = \\ 25 &= 2 \cdot 11 + 3 \\ 11 &= 3 \cdot 3 + 2 \\ 3 &= 1 \cdot 2 + 1 \\ 2 &= 2 \cdot 1 + 0 \end{aligned}$$

When the last remainder is zero, then you have your greatest common divisor.

Extended Euclidean Algorithm

Theorem 85 (Bezout's)**.** *For any numbers a, b there exists integers s, t such that*

$$gcd(a, b) = as + bt$$

proof by WOP
[54]

[54] kuttaka

Bezout's theorem is incredibly important. We will be able to show you how to calculate s, t given $a, b, \gcd(a, b)$. The calculation simply takes the execution of the euclidean algorithm, and uses it to find s, t. As we computed the euclidean algorithm, we went through a sequence of pairs

$$(25, 11) \to (11, 3) \to (3, 2) \to (2, 1) \to (1, 0)$$

We will work backwards through the pairs, until we are left with a linear combination of the first pair. For example, we will replace 3 with a linear combination of 25 and 11. The way we will do this, is by taking the steps of the euclidean algorithm and substituting them back into each other back up until the first one. First, take your equations, and rewrite them with the remainder on one side

$$\begin{aligned} \gcd(25, 11) = \\ 25 &= 2 \cdot 11 + 3 & 3 &= (1)25 + (-2)11 \\ 11 &= 3 \cdot 3 + 2 & 2 &= (1)11 + (-3)3 \\ 3 &= 1 \cdot 2 + 1 & 1 &= (1)3 + (-1)2 \\ 2 &= 2 \cdot 1 + 0 & 0 &= (1)2 + (-2)1 \end{aligned}$$

Our pairs are (1,0), (2,1), (3,2), (11,3) and (25,11). Let us compute s, t such that $25s + 11t = 1$ working backwards First, write the $gcd(a, b) = 1$ as a linear combination of the first pair, (1,0)

$$1 = 1 + 0$$

Next, we want to go from pair (1,0) to pair (2,1) so we will use the last equation of $0 = ...$ and substitute this in.

$$1 = 1 + \mathbf{0} = \tag{52}$$

$$1 + [(1)2 + (-2)1] = \tag{53}$$

$$(1)2 + (-1)1 \tag{54}$$

Note how we have written the linear combination of (1,0) as a linear combination of (2,1). Let us substitute out the 1 for a linear combination of 3 and 2. Since we leave the 2 unchanged, we will have a linear combination of 3 and 2.

$$\begin{aligned} 1 &= (1)2 + (-1)\mathbf{1} = \\ &\quad (1)2 + (-1)[(1)3 + (-1)2] = \\ &\quad (-1)3 + (2)2 \end{aligned}$$

Now we replace the 2 with the linear combination of 11 and 3.

$$\begin{aligned} 1 &= (-1)3 + (2)\mathbf{2} = \\ &\quad (-1)3 + (2)[(1)11 + (-3)3] = \\ &\quad (2)11 + (-7)3 \end{aligned}$$

Replace 3 with a linear combination of 25 and 11 and we will be complete

$$\begin{aligned} &(2)11 + (-7)\mathbf{3} = \\ &(2)11 + (-7)[(1)25 + (-2)11] = \\ &(-7)25 + (16)11 \end{aligned}$$

So we may conclude that

$$1 = 25(-7) + 11(16)$$

For $a, b = 25, 11$ our values of $s, t = -7, 16$. These are not guaranteed to be unique or minimal, and you may find other numbers which work, but the extended Eucldean algorithm is guaranteed to give you a pair of numbers s, t to satisfy Bezout's theorem.

Least Common Multiple

Definition 0.0.50. *Let a, b be numbers, both not zero. Then the least common multiple of a, b is the least number l such that $a \mid l$ and $b \mid l$.*

- $\text{lcm}(24, 10) = 120$
- $\text{lcm}(2^a, 2^b) = 2^{\max(a,b)}$
- $\text{lcm}(p, q) = pq$ if p, q are primes.

One way to compute the lcm is to write out two lists of multiples, and find the first one to appear in both. Suppose we wanted to compute $\text{lcm}(6, 15)$.

$$6, 12, 18, 24, \mathbf{30}, ... \tag{55}$$

$$15, \mathbf{30}, ... \tag{56}$$

LCM and GCD share a kind of duality. Multiplies are bigger than their numbers, and the LCM is the least multiple. Divisors are smaller than their numbers, and the GCD is the greatest divisor. The LCM is analogous to a set union, and the GCD is analogous to a set intersection. We can take advantage of this duality for an efficient way to compute the LCM of two numbers.

Theorem 86. *For any numbers $a, b \geq 1$ it is true that*

$$\gcd(a,b)lcm(a,b) = ab$$

Proof. By the fundamental theorem of arithmetic, a, b have unique prime factorizations. Without loss of generality, suppose that a has p_k as its largest prime divisor and b has p_l as its largest prime divisor with $k \geq l$. Suppose the factorizations are

$$a = 2^{a_1}3^{a_2}...p_k^{a_k} = \prod_{i=1}^{k} p_i^{a_i}$$

$$b = 2^{b_1}3^{b_2}...p_l^{b_l} = \prod_{i=1}^{l} p_i^{b_i}$$

Since $\gcd(a,b) \mid a$ and $\gcd(a,b) \mid b$ we know that

$$\gcd(a,b) = 2^{\min(a_1,b_1)}3^{\min(a_2,b_2)} \cdot ... \cdot p_k^{\min(a_k,b_k)} = \prod_{i=1}^{k} p_i^{\min(a_i,b_i)}$$

Similarly, the least common multiple must be large enough to accomodate all the prime divisors of both a, b so

$$\operatorname{lcm}(a,b) = 2^{\max(a_1,b_1)}3^{\max(a_2,b_2)} \cdot ... \cdot p_k^{\max(a_k,b_k)} = \prod_{i=1}^{k} p_i^{\max(a_i,b_i)}$$

Since $k \geq l$, the values for b_i with $i > l$ may be zero. Then

$$\gcd(a,b) \cdot \operatorname{lcm}(a,b) =$$
$$\Big(\prod_{i=1}^{k} p_i^{\min(a_i,b_i)}\Big) \cdot \Big(\prod_{i=1}^{k} p_i^{\max(a_i,b_i)}\Big) =$$
$$\Big(\prod_{i=1}^{k} p_i^{\min(a_i,b_i)+\max(a_i,b_i)}\Big)$$

For any two numbers, x, y it is true that $\max(x,y) + \min(x,y) = x + y$. One will be the min, the other must be the max. So

$$\Big(\prod_{i=1}^{k} p_i^{\min(a_i,b_i)+\max(a_i,b_i)}\Big) = \Big(\prod_{i=1}^{k} p_i^{a_i+b_i}\Big) =$$
$$\Big(\prod_{i=1}^{k} p_i^{a_i} p_i^{b_i}\Big) =$$
$$\Big(\prod_{i=1}^{k} p_i^{a_i}\Big) \cdot \Big(\prod_{i=1}^{k} p_i^{b_i}\Big) = ab$$

□

Corollary 87.

$$lcm(a,b) = \frac{ab}{\gcd(a,b)}$$

This gives us a useful way to compute the least common multiple of two numbers. Simply compute their product, use the euclidean algorithm and perform an integer division.

Group Theory

Definition 0.0.51. *A group is a set and operation $(G, \cdot)$ with the following four properties*

- ***identity*** $\exists e \in G \forall a \in G ae = ea = a$
- ***associativity*** $\forall a, b, c \in G a(bc) = (ab)c$
- ***closure*** $\forall a, b \in G ab \in G$
- *inverses* $\forall a \in G \exists a^{-1} \in G aa^{-1} = a^{-1}a = e$

why groups

Theorem 88. *a^{-1} exists* $\mod n$ *if and only if* $\gcd(a, n) = 1$

Proof. ($\implies$) Suppose that a^-1 exists $\mod n$. By the definition of the inverse $aa^{-1} \equiv 1 \pmod{n}$, so $n \mid aa^{-1} - 1$. Thus, there exists a c such that $nc = aa^{-1} - 1$, or that $nc + aa^{-1} = 1$. By Bezout's theorem, we know that the least positive linear combination of a, n is the *gcd*, so we may conclude that $\gcd(a, n) \leq 1$, but since 1 divides all integers $1 \geq \gcd(a, b)$. We may then only conclude that $\gcd(a, n) = 1$

($\impliedby$) Let $\gcd(a, n) = 1$. By Bezout's theorem, there exists integers s, t such that $as + nt = 1$. If we mod by n, we observe that $as \equiv 1 \mod n$. Since Bezout's asserts this s exists, then the inverse of a is s so the inverse exists. □

actually a way to calculate the inverse

Corollary 89. *$(\mathbb{Z}_n \setminus \{0\}, \cdot \pmod{n})$ is a group if and only if n is prime.*

Proof. we prove the contrapositive for the reverse implication. Suppose that n is composite. Then there exists a number a such that $a < n$ but $a \mid n$. Then $\gcd(a, n) \neq 1$ and a^{-1} does not exist. □

Just because certain elements don't have inverses with respect to certain moduli doesn't mean we don't study group theory with respect to these moduli. We just throw out all the elements without inverses.

Definition 0.0.52. *ZZp*

The Chinese Remainder Theorem

Theorem 90. *Let $n_1, n_2, ..., n_k$ be pairwise relatively prime, that is $i \neq j \implies \gcd(n_i, n_j) = 1$. Let $r_1, ..., r_k$ be numbers. Then the system of equations*

$$\begin{aligned} x &\equiv r_1 \pmod{n_1} \\ x &\equiv r_2 \pmod{n_2} \\ &\ddots \\ x &\equiv r_1 \pmod{n_k} \end{aligned}$$

has a unique solution $x \pmod{n_1 \cdot ... \cdot n_k}$

Proof. We first prove existence, then we prove uniqueness

Suppose there are two distinct solutions, we prove they are equal. Suppose there is x, x' that are solutions to the same system of equations

$$\begin{aligned} x &\equiv r_1 \pmod{n_1} \\ x &\equiv r_2 \pmod{n_2} \\ &\ddots \\ x &\equiv r_k \pmod{n_k} \end{aligned}$$

$$\begin{aligned} x' &\equiv r_1 \pmod{n_1} \\ x' &\equiv r_2 \pmod{n_2} \\ &\ddots \\ x' &\equiv r_k \pmod{n_k} \end{aligned}$$

For all i, since $x \equiv r_i \pmod{n_i}$ and $x' \equiv r_i \pmod{n_i}$, by transitivity, we see that $x \equiv x' \pmod{n_i}$. Thus

$$\begin{aligned} x &\equiv x' \pmod{n_1} \\ x &\equiv x' \pmod{n_2} \\ &\ddots \\ x &\equiv x' \pmod{n_k} \end{aligned}$$

□

Calculation

Suppose you would actually like to calculate a solution to a system of linear congruences. There are two ways

Tabulation

Since our proof of the Chinese remainder theorem was constructive, Lets just apply the construction. Let $N = n_1 \cdot ... \cdot n_k$, let $N_i = N/n_i$, and let $N_i' = N_i^{-1} \pmod{n}_i$. We compute $x = \sum_{i=1}^{k} N_i N_i' r_i$. Consider the following problem[55]

[55] This is an original problem from the Sunzi Suanjing, 400AD: *"There are certain things whose number is unknown. If we count them by threes, we have two left over; by fives, we have three left over; and by sevens, two are left over. How many things are there?"*

$$\begin{aligned} x &\equiv 2 \pmod{3} \\ x &\equiv 3 \pmod{5} \\ x &\equiv 2 \pmod{7} \end{aligned}$$

n_i	r_i	N_i	N_i'	$N_i N_i' r_i$
3	2	35	2	140
5	3	21	1	63
7	2	15	1	30

Then $x \equiv 140 + 63 + 30 \equiv 233 \equiv 23 \pmod{105}$. So we see that that 23 is the unique solution[56] to the system of linear congruences.

[56] Sunzi also included the solution: *"Multiply the number of units left over when counting in threes by 70, add to the product of the number of units left over when counting in fives by 21, and then add the product of the number of units left over when counting in sevens by 15. If the answer is 106 or more then subtract multiples of 105"*. As you can see, some information is left as exercise to the reader.

Gaussian Elimination

Let us do the same problem and ensure we get the same answer.

$$\begin{aligned} x &\equiv 2 \pmod{3} \\ x &\equiv 3 \pmod{5} \\ x &\equiv 2 \pmod{7} \end{aligned}$$

First we check the moduli are all pairwise relatively prime. Since they are all prime, this follows. Thus, we know a solution exists. We will compute a solution $x \pmod{105}$. We take equation with the largest modulus, rewrite it, and substitute it in to the one above it. Chaining these together, we will

eventually compute our solution.

$$
\begin{aligned}
x &\equiv 2 \pmod 7 \\
x &= 7k + 2 \text{ for some } k \\
x &\equiv 3 \pmod 5 \\
7k + 2 &\equiv 3 \pmod 5 \\
2k &\equiv 1 \pmod 5 \\
k &\equiv 2^{-1} \pmod 5 \\
k &\equiv 3 \pmod 5 \\
k &= 5l + 3 \text{ for some } l \\
x &= 7(5l + 3) + 2 \\
x &= (5 \cdot 7)l + 23 \\
x &\equiv 2 \pmod 3 \\
(5 \cdot 7)l + 23 &\equiv 2 \pmod 3 \\
2l + 2 &\equiv 2 \pmod 3 \\
2l &\equiv 0 \pmod 3 \\
l &\equiv 0 \pmod 3 \\
l &= 3r + 0 \text{ for some } r \\
x &= (5 \cdot 7)(3r) + 23 \\
x &= (3 \cdot 5 \cdot 7)r + 23 \\
x &\equiv 23 \pmod{105}
\end{aligned}
$$

Both methods work equally well, the second method is slightly easier, the first method is more mechanical. The only difficulty in the first method is you have to compute k inverses. For the second method, this isn't always necessary.

Fermat and Euler

Lemma 91. *If* $\gcd(a,n)=1$ *and* $\gcd(b,n)=1$ *then* $\gcd(ab,n)=1$.

Proof. Since a,b are relatively prime to n then a^{-1},b^{-1} exist $\pmod n$. Observe that $ab(b^{-1}a^{-1})\equiv 1 \pmod n$, so the inverse of ab exists, and is $(b^{-1}a^{-1})$, We know an inverse of ab exists if and only if it is relatively prime so we conclude that $\gcd(ab,n)=1$. □

Theorem 92. *If* $\gcd(m,n)=1$ *then* $\varphi(mn)=\varphi(m)\varphi(n)$

Proof. Let m,n be relatively prime. We prove there exists a bijection

$$f:(\mathbb{Z}/mn\mathbb{Z})^\times \to (\mathbb{Z}/m\mathbb{Z})^\times \times (\mathbb{Z}/n\mathbb{Z})^\times$$

Let $c\in(\mathbb{Z}/mn\mathbb{Z})^\times$ and consider the function $f(c)=(a,b)$ where $a\equiv c \pmod m$ and $b\equiv c \pmod n$. By our lemma, since $\gcd(c,mn)=1$ we know $\gcd(c,m)=1$ and $\gcd(c,n)=1$. So (a,b) is in the co-domain we say it is in. Next we prove that f is a bijection.

We first prove surjectivity. Let $(a,b)\in(\mathbb{Z}/m\mathbb{Z})^\times\times(\mathbb{Z}/n\mathbb{Z})^\times$. We prove there exists a $c\in(\mathbb{Z}/mn\mathbb{Z})^\times$ such that $f(c)=(a,b)$. By the chinese remainder theorem, we know if $c\equiv a \pmod m$ and $c\equiv b \pmod n$, then there exists a solution $c \pmod{mn}$. We apply the chinese remainder theorem to find this c. The surjectivity of this function follows from the existence part of the CRT.

Next, we prove injectivity. Let $f(c)=f(c')$. We prove $c\equiv c' \pmod{mn}$.

$$\begin{aligned} f(c)&=f(c')\\ (c \pmod m, c \pmod n)&=(c' \pmod m, c' \pmod n)\\ c&\equiv c' \mod m\\ c&\equiv c' \mod n \end{aligned}$$

By the chinese remainder theorem, we know that if $x\equiv r_1 \pmod m$ and $x\equiv r_2 \pmod n$ and if m,n are relatively prime, then there exists a unique solution $x \pmod{mn}$. Ifc $c\equiv c' \mod m$ and $c\equiv c' \mod n$ then $c=c' \pmod{mn}$. The injectivity of this function follows from the uniquess part of the CRT.

We have now proven that f is a bijection. By a previous theorem, this implies that the domain and co-domain have the same cardinality, thus

$$\begin{aligned} |(\mathbb{Z}/mn\mathbb{Z})^\times| &= |(\mathbb{Z}/m\mathbb{Z})^\times\times(\mathbb{Z}/n\mathbb{Z})^\times|\\ |(\mathbb{Z}/mn\mathbb{Z})^\times| &= |(\mathbb{Z}/m\mathbb{Z})^\times|\cdot|(\mathbb{Z}/n\mathbb{Z})^\times|\\ \varphi(mn)&=\varphi(m)\varphi(n) \end{aligned}$$

□

Theorem 93. *Let p be a prime and $k \geq 1$. Then $\varphi(p^k) = p^k - 1(p-1)$*

Proof. Consider the numbers in sequence

$$1, 2, 3, ..., p^k$$

What numbers appear in this sequence which are not relatively prime to p^k? Notice that $gcd(m, p^k) \neq 1 \iff m \in p, 2p, 3p, ..., p^{k-1}p$. There are exactly p^{k-1} multiples of p, so there are $p^k - p^{k-1} = p^{k-1}(p-1)$ numbers between $1, p^k$ which are relatively prime to p^k. □

We now have enough to give a formula for $\varphi(n)$.

Theorem 94. *Let p be a prime number which divides n. Then*

$$\varphi(n) = n \prod_{p \mid n} \left(1 - \frac{1}{p}\right)$$

Proof. By the fundamental theorem of arithmetic, n can be uniquely written as a product of prime powers $n = p_1^{a_1} \cdot ... \cdot p_k^{a_k} = \prod_{i=1}^{k} p_i^{a_i}$. Without loss of generality, suppose the powers $a_i \geq 1$, so n is only written as a product of the powers which divide it non trivially.

$$\begin{aligned}
\varphi(n) = \\
\varphi\left(\prod_{i=1}^{k} p_i^{a_i}\right) = \\
\prod_{i=1}^{k} \varphi(p_i^{a_i}) = \\
\prod_{i=1}^{k} p_i^{a_i - 1}(p_i - 1) = \\
\prod_{i=1}^{k} p_i^{a_i}\left(1 - \frac{1}{p_i}\right) = \\
\left(\prod_{i=`1}^{k} p_i^{a_i}\right)\left(\prod_{i=1}^{k}\left(1 - \frac{1}{p_i}\right)\right) = \\
n \prod_{p \mid n} \left(1 - \frac{1}{p}\right)
\end{aligned}$$

□

some sample computations go here

Modular Exponentiation

We have dealt with equality, addition, multiplication, in modular arithmetic. Now we deal with exponentiation. Given something of the form $a^b \pmod n$, you can always simplify the base, since

$$a^b \equiv \underbrace{a \cdot a \cdot ... \cdot a}_{b} \equiv (a \pmod n)^b$$

For example, $12^{100} \pmod{11} \equiv 1^{100} \equiv 1 \pmod{11}$.

repeated squaring

We can always simplify the base this easily. Simplifying in the exponent requires a little more subtlety. We first need to prove a small and useful lemma.

Theorem 95 (Fermat's Little Theorem)**.** *If $1 \leq a < p$ and p is a prime number then*

$$a^{p-1} \equiv 1 \pmod{p}$$

Proof. Consider the function $f(i) = ai \pmod{p}$. We prove that f is a bijection from $(\mathbb{Z}/p\mathbb{Z})^\times$ to $a(\mathbb{Z}/p\mathbb{Z})^\times = \{a \pmod{p}, 2a \pmod{p}, ..., (p-1)a \pmod{p}\}$.

We first prove f is injective. Let $f(i) = f(j)$. We prove $i \equiv j \pmod{p}$.

$$\begin{aligned} f(i) &= f(j) \\ ai &\equiv aj \pmod{p} \end{aligned}$$

Since $\gcd(a, n) = 1$, we know a^{-1} exists, thus

$$\begin{aligned} ai &\equiv aj \pmod{p} \\ a^{-1}(ai) &\equiv a^{-1}(aj) \pmod{p} \\ i &\equiv j \pmod{p} \end{aligned}$$

Next we prove f is surjective. Let $y \in \{a, 2a, 3a, ..., (p-1)a\}$ all $\pmod{p}$. We prove there exists a x such that $f(x) = y$. Consider $x \equiv a^{-1}y$. Then

$$f(x) = f(a^{-1}y) \equiv a(a^{-1}y) \equiv y \pmod{p}$$

We know that two finite sets have the same cardinality if and only if there is a bijection between them, thus $|\{1, 2, ..., p-1\}| = |\{a \pmod{p}, 2a \pmod{p}, ..., p-1 \pmod{p}\}| = p - 1$. The function $\pmod{p}$ will return values 0 to p. If we prove that $0 \notin a(\mathbb{Z}/p\mathbb{Z})^\times$, then the remaining $p-1$ elements are exactly the $p-1$ elements, of $a(\mathbb{Z}/p\mathbb{Z})^\times$, which happen to exactly be the elements of $(\mathbb{Z}/p\mathbb{Z})^\times$. So if $0 \notin a(\mathbb{Z}/p\mathbb{Z})^\times$ then $(\mathbb{Z}/p\mathbb{Z})^\times = a(\mathbb{Z}/p\mathbb{Z})^\times$.

Assume to the contrary $0 \in a(\mathbb{Z}/p\mathbb{Z})^\times$. Then there exists an $i \in (\mathbb{Z}/p\mathbb{Z})^\times$ such that $ai \equiv 0 \pmod{p}$. Then

$$ai \equiv 0 \pmod{p} \tag{57}$$

$$a^{-1}(ai) \equiv a^{-1}(0) \pmod{p} \tag{58}$$

$$i \equiv 0 \pmod{p} \tag{59}$$

Contradiction, as $0 \notin (\mathbb{Z}/p\mathbb{Z})^\times$.

Now that we have proven $(\mathbb{Z}/p\mathbb{Z})^\times = a(\mathbb{Z}/p\mathbb{Z})^\times$, we may prove Fermat's little theorem. Note that if two sets are equal, then the product of their elements are equal.

$$1 \cdot 2 \cdot 3 \cdot ... \cdot p - 1 \equiv (a)(2a)(3a)...((p-1)a) \pmod{p} \tag{60}$$

$$(p-1)! \equiv a^{p-1}(p-1)! \pmod{p} \tag{61}$$

Since $\gcd((p-1)!, p) =$, then $((p-1)!)^{-1}$ exists, so

$$(p-1)! \equiv a^{p-1}(p-1)! \pmod{p} \tag{62}$$

$$1 \equiv a^{p-1} \pmod{p} \tag{63}$$

□

We wish to extend Fermat's little theorem for the case that the modulus is not prime. Where did $p - 1$ actually come from here? It was the cardinality of the group $|(\mathbb{Z}/p\mathbb{Z})^\times| = \varphi(p) = p - 1$. We also made good use of the fact that a^{-1} exists. With these conditions, we may generalize to proof Euler's theorem.

Theorem 96 (Euler's Theorem)**.** *Let $n \geq 2$ and $1 \leq a < n$ and $\gcd(a, n) = 1$. Then*

$$a^{\varphi(n)} \equiv 1 \pmod{n}$$

Proof. Let $n \geq 2$ and let $1 \leq a < n$ with $\gcd(a, n) = 1$. Let $(\mathbb{Z}/n\mathbb{Z})^\times = \{k_1, k_2, ..., k_{\varphi(n)}\}$. Consider the function $f : (\mathbb{Z}/n\mathbb{Z})^\times \to a(\mathbb{Z}/n\mathbb{Z})^\times$ by $f(k_i) = ak_i \pmod{n}$. We prove that f is a bijection and that zero is not an element of the image of f.

To prove injectivity, let $f(k_i) = f(k_j)$. We prove $k_i \equiv k_j \pmod{n}$.

$$\begin{aligned} f(k_i) &\equiv f(k_j) && \pmod{n} \\ ak_i &\equiv ak_j && \pmod{n} \end{aligned}$$

Since $\gcd(a, n) = 1$ then a^{-1} exists in $(\mathbb{Z}/n\mathbb{Z})^\times$ so

$$\begin{aligned} ak_i &\equiv ak_j && \pmod{n} \\ a^{-1}(ak_i) &\equiv a^{-1}(ak_j) && \pmod{n} \\ k_i &\equiv k_j && \pmod{n} \end{aligned}$$

Next we prove it is surjective. For $y \in a(\mathbb{Z}/n\mathbb{Z})^\times$, we prove there exists $x \in (\mathbb{Z}/n\mathbb{Z})^\times$ such that $f(x) = y$. Consider $x \equiv a^{-1}y$. Then

$$f(x) \equiv f(a^{-1}y) \equiv a(a^{-1}y) \equiv y \pmod{n}$$

Finally, we prove that zero is not an element of $a(\mathbb{Z}/n\mathbb{Z})^\times$. Suppose to the contrary that it was. Then there would exist $k_i \in (\mathbb{Z}/n\mathbb{Z})^\times$ such that $0 \equiv ak_i \pmod{n}$.

$$ak_i \equiv 0 \pmod{n} \tag{64}$$
$$a^{-1}(ak_i) \equiv a^{-1}(0) \pmod{n} \tag{65}$$
$$k_i \equiv 0 \pmod{n} \tag{66}$$

Contradiction, as $0 \notin (\mathbb{Z}/n\mathbb{Z})^\times$.

We can deduce that since $|(\mathbb{Z}/n\mathbb{Z})^\times| = |a(\mathbb{Z}/n\mathbb{Z})^\times| = \varphi(n)$, and their least element is 1, they must containt the same elements, and therefore, be equal. Thus $(\mathbb{Z}/n\mathbb{Z})^\times = a(\mathbb{Z}/n\mathbb{Z})^\times$. If two sets are equal, then the products of their elements are equal.

$$\prod_{i=1}^{\varphi(n)} k_i \equiv \prod_{i=1}^{\varphi(n)} ak_i \pmod{n} \tag{67}$$
$$\prod_{i=1}^{\varphi(n)} k_i \equiv a^{\varphi(n)} \prod_{i=1}^{\varphi(n)} k_i \pmod{n} \tag{68}$$
$$\tag{69}$$

Since $\prod_{i=1}^{\varphi(n)} k_i$ is a product of elements all relatively prime to n, then $\prod_{i=1}^{\varphi(n)} k_i$ is relatively prime to n and its inverse exists $\pmod{n}$. We multiply both sides by this inverse to conclude

$$a^{\varphi(n)} \equiv 1 \pmod{n}$$

□

Corollary 97. *If a is relatively prime to n and $k = q\varphi(n) + r$ then $a^k \equiv a^r \pmod{n}$*

Proof. By Euler's theorem

$$a^k \equiv a^{q\varphi(n)+r} \equiv (a^{\varphi(n)})^d a^r \equiv 1^d a^r \equiv a^r \pmod{n}$$

□

observe that since $r < n$ this simplifies the exponent for us.

These theorems are pretty powerful, lets do an example. We will compute

$$3^{3^{3^3}} \pmod{100}$$

This exponent is quite complex. Lets apply Euler's theorem to the exponent, solve a subproblem, and plug it back in.

$$3^{3^3} \pmod{\varphi(100)}$$

We see a similar problem, and repeat

$$3^3 \pmod{\varphi(\varphi(100))}$$

With the factorization of $100 = 2^2 5^2$, we get $\varphi(100) = 40 = 2^3 5^1$, so $\varphi(\varphi(n)) = 16$. Then

$$3^3 \equiv 27 \equiv 11 \pmod{16}$$

Lets substitute this back in to get

$$3^{3^3} \equiv 3^{27} \equiv 3^{11} \equiv 3^{4\cdot 2+3} \equiv (3^4)^2 3^3 \equiv (81^2)(3^3)(1^2)(27) \equiv 27 \pmod{40}$$

We now substitute this all the way back into the original

$$3^{3^{3^3}} \equiv 3^{3^{27}} \equiv 3^{27} \pmod{100}$$

We could to try to divide out the exponent so that we may mod out by something close to 100, but there are no easy candidates. For example

$$3^{27} \equiv (3^5)^5 3^2 \equiv (243)^5 3^2 \equiv (43)^5 3^2$$

You can't simply that any further without some complex work. Instead, we bring in a second powerful tool: the chinese remainder theorem. The factorization of $100 = 2^2 5^2$. We compute $3^{27} \pmod 4$ and $3^{27} \pmod{25}$ and then recombine back up.

$$3^{27} \equiv (3^2)^{13} 3 \equiv (9)^{13} 3 \equiv (1)^{13} 3 \equiv 3 \pmod 4$$
$$3^{27} \equiv (3^3)^9 \equiv (27)^9 \equiv 2^9 \equiv 512 \equiv 12 \pmod{25}$$

Now we can apply a chinese remainder theorem algorithm to compute a unique solution $\pmod{100}$, which must be then equivalent to $3^{27} \pmod{100}$. Lets

use Gaussian elimination.

$$\begin{aligned}
x &\equiv 3 \pmod 4 \\
x &\equiv 12 \pmod{25} \\
x &= 25k + 12 \text{ for some } k \\
25k + 12 &\equiv 3 \pmod 4 \\
1 \cdot k + 0 &\equiv 3 \pmod 4 \\
k &\equiv 3 \pmod 4 \\
k &= 4l + 3 \text{ for some } l \\
x &= 25(4l + 3) + 12 \\
x &= 100l + 87 \\
x &\equiv 87 \pmod{100}
\end{aligned}$$

Thus $3^{3^{3^3}} \equiv 87 \pmod{100}$.

Pigeonhole Principle

Theorem 98. *If you have n pigeons and k holes, then there exists a hole with ≥ 2 pigeons.*

The Pigeonhole Principle is an extremely simple and atomic theorem, but as we will see, it has great power.[57] To demonstrate the diversity in which you may apply it, we do several examples.

Theorem 99. *Every run on sentence of 27 words or more must contain two words which begin with the same letter.*

Proof. There are 26 letters, and any for any collection of 27 words, by the pigeonhole principle, there must exist two words that begin with the same letter. □

Notice it doesn't really matter that the 27 words form a run on sentence. It only matters that there are more pigeons than holes. Implicit in the definition of the pigeonhole principle is that every pigeon must be placed in a hole. No one waits in line or doesn't get place. Also implicit is that no pigeon can be in two holes at once.

The pigeonhole principle is also nonconstructive. You know nothing about the hole. Suppose you have n red and n blue socks, and you pull socks out of your sock bucket one at a time, blindly. If you pull just three socks, since there are only two colors, then by the pigeonhole principle you must have pulled out a pair of socks. You don't know what color it is though. Suppose you specifically wanted to pull out a red pair of socks. You must pull out $n + 2$ socks to guarantee you eventually get your red pair, since its possible[58] that you would pull all n blue socks first.

Theorem 100. *Let A, B be finite sets with $|A| > |B|$. Then there do not exist any injections $f : A \to B$.*

Proof. Let $f : A \to B$ be any such function. Since f is a function, everything in the domain is mapped to exactly one element in the co-domain. By the pigeonhole principle, since $|A| > |B|$, there exists distinct $a_i, a_j \in A$ such that they map to the same element. So $a_i \neq a_j$ but $f(a_i) = f(a_j)$. This implies that f cannot be injective. □

Theorem 101. *For each length n, some file of length n which zipped does not get smaller.*

Proof. Suppose there exists a perfect compression algorithm $f : A \to B$ with $f(x) = x.zip$ Suppose that f takes on input files of size n and outputs files that are strictly smaller than size n. There are 2^n files of size n, so $|A| = 2^n$. How many possible files are there of size $< n$? 1 file of size 0, 2 files of size 1, and 4 files

[57] Every theorem in your arsenal is a spell. You may compose and combine spells together to prove powerful theorems. If a spell is of the form $p \implies q$, you should think of p like a mana cost, the conditions you need to set up first to apply the spell, and then q is the outcome. The pigeonhole principle is among the most unique theorems you can every apply. The conditions and costs it needs are very weak. You simply need to have some pigeons and holes to put them in, and ensure there are more pigeons than holes. Often you can construct the pigeons and holes yourself, and always ensure you have more pigeons then holes. The cost to apply the spell is minimal, but the outcome is great. The pigeonhole principle conjures, or *wills into existence* an object with a specific property. Although this seems powerful, you know nothing else about the summoned object other than it exists. You don't know where it is, how to find it, or anything else. The pigeonhole principle is inherently nonconstructive. Nothing is really conjured, or created or summoned. The pigeonhole principle proves that the object always had to exist, and only this existence is asserted. Although it is nonconstructive, in certain cases, you may combine it with other theorems to make it constructive.

[58] but we concede, unlikely

of size 2. Continuing in this way, we see $|B| = 1+2+4+8+...+2^{n-1} = 2^n - 1$. Since $|A| > |B|$ by the pigeonhole principle, two files of size n when compressed must map to the same zipped file. Contradiction, as either f is not a correct compression algorithm, or must make one of those files not smaller. □

Theorem 102 (Posa's Problem)**.** [59] *Let $S \subseteq \{1, 2, ..., 2n\}$. If $|S| \geq n+1$ then there exists $a, b \in S$ such that* $\gcd(a, b) = 1$.

Proof. First observe that consecutive numbers are relatively prime. If $k \geq 1$ then

$$\gcd(k, k+1) = \gcd(k+1, (k+1)-k) = \gcd(k+1, 1) = 1$$

Suppose we partition $\{1, 2, ..., 2n\}$ into two element sets

$$\{1, 2\}, \{3, 4\}, \{5, 6\}, ..., \{2n-1, 2n\}$$

There are n such sets. If $|S| \geq n+1$ then by the pigeonhole principle, both elements of some two element set must be in S, and these are consecutive. □

Theorem 103. *Let $S \subseteq \{1, 2, ..., 2n\}$. If $|S| \geq n+1$ then there exists $a, b \in S$ such that $a \mid b$*

Proof. Let $S = \{a_1, ..., a_{n+1}\}$. Notice that every number may be written as a power of two times an odd number, so there exists $b1, ..., b_{n+1}$ and odd $o_1, ..., o_{n+1}$ with $a_i = 2^{b_i} o_i$. Since each $a_i \leq 2n$, we know each $o_i \leq 2n-1$. There are $n+1$ odd components $o_1, ..., o_{n+1}$, but the set $\{1, 3, ..., 2n-1\}$ has size n. By the pigeonhole principle, there exists distinct i, j such that $o_i = o_j$.

Consider a_i, a_j, the numbers corresponding to these equal odd components. Recall $a_i = 2^{a_i} o_i$ and $a_j = 2^{a_j} o_j 2^{a_j} o_i$. We have two cases.

- If $a_i = a_j$ then certainly $a_i \mid a_j$.
- If $a_i \neq a_j$, then one must be greater than the other. Without loss of generality, suppose $a_i > a_j$. Since they have the same odd component, the power of two of a_i must be greater than the power of two of a_j. So $b_i > b_j$. Thus $2^{b_j} \mid 2^{b_i}$. Since o_i is odd then $2^{b_j} o_i \mid 2^{b_i} o_i \implies 2^{b_j} o_j \mid 2^{b_i} o_i \implies a_j \mid a_i$.

In either case, we see S must contain two numbers such that one divides the other. □

Theorem 104. *for all $n \geq 2$, there exists a multiple of n which is written in base ten only using 1's followed by 0's.*

Proof. Consider the sequence of numbers in base ten

$$1, 11, 111, ..., \underbrace{1...1}_{n+1}$$

where $\underbrace{1...1}_{k}$ is a number in base 10 written as k 1s. Suppose we take these $n+1$ numbers and mod them all by n. By the pigeonhole principle, two of these numbers must be congruent mod n. There exists distinct i, j such that

$$\underbrace{1...1}_{i} \equiv \underbrace{1...1}_{j} \pmod{n}$$

[59] "When I returned from the United States in the summer of 1959," recalled Erdos, "I was told that there is a little boy whose mother is a mathematician and who knows all there is to be known in high school. I was very interested and next day I had lunch with him... While we had lunch and [Louis] Pósa was eating his soup, I told him the following problem. Prove that if you have $n+1$ integers less than or equal to $2n$ there are always two of them which are relatively prime. I discovered this simple result some years ago but it took me about ten minutes until I found the very simple proof." As an example, choose n to be 5. Then the conjecture is that if you take any six integers from the set 1, 2, 3, 4, 5, 6, 7, 8, 9, and 10, you can't avoid choosing two that are relatively prime (meaning, remember, that they have no common divisor greater than 1). The conjecture would fail if you were allowed to choose just five of these integers: there are five even numbers in this set, namely, 2, 4, 6, 8, and 10, all of which obviously share the divisor 2. Louis Pósa finished his soup and announced, "The two are neighbors." In other words, the two are consecutive. "If you have $n+1$ integers less than or equal to $2n$," said Erdos, "two of them are consecutive and therefore they are relatively prime."

Without loss of generality, suppose $i > j$. Then

$$\underbrace{1...1}_{i} \equiv \underbrace{1...1}_{j} \pmod{n}$$
$$\underbrace{1...1}_{i} - \underbrace{1...1}_{j} \equiv 0 \pmod{n}$$
$$\underbrace{1...1}_{i-j}\underbrace{0...0}_{j} \equiv 0 \pmod{n}$$

Thus $\underbrace{1...1}_{i-j}\underbrace{0...0}_{j}$ is a multiple of n written in base ten as 1's followed by 0's. □

Theorem 105. *Let $a_1, a_2, ..., a_n$ be any n numbers. Then there exists a subsequence $a_k, a_{k+1}, ..., a_{k+l}$ such that $a_k + a_{k+1} + ... + a_{k+l}$ sums to a multiple of n.*

Proof. Consider the following numbers $s_1, ..., s_n$ where

$$s_1 = a_1$$
$$s_2 = a_1 + a_2$$
$$s_3 = a_1 + a_2 + a_3$$
$$...$$
$$s_n = a_1 + a_2 + a_3 + ... + a_n$$

Suppose we take these n numbers and mod them all by n. We have two cases

- If there is an s_i such that $s_i \equiv 0 \pmod{n}$ then $a_1 + ... + a_i$ is a subsequence which sums to a multiple of n.
- If there is not an s_i such that $s_i \equiv 0 \pmod{n}$, then modding $s_1, ..., s_n$ each by n must map them into the $n-1$ equivalence classes $[1], ..., [n-1]$. By the pigeonhole principle, there is distinct i, j such that s_i, s_j fall into the same equivalence class. Therefore $s_i \equiv s_j \pmod{n}$. Without loss of generality, suppose $i > j$.

$$s_i \equiv s_j \pmod{n}$$
$$s_i - s_j \equiv 0 \pmod{n}$$
$$(a_1 + a_i) - (a_1 + ... + a_j) \equiv 0 \pmod{n}$$
$$a_{j+1} + ... + a_i \equiv 0 \pmod{n}$$

Thus, $a_{j+1} + ... + a_i$ is a subsequence of $a_1, ..., a_n$ which sums to a multiple of n.

□

rubix cube

Theorem 106. *If G is a finite group, then for all $a \in G$ there exists a number $k > 1$ such that $a^k = e$.*

Proof. Let G be a finite group of n elements. Consider the sequence

$$a, a^2, a^3, ..., a^{n+1}$$

By closure, these are all in G. Since there are $n+1$ such products but only n elements in G, then by the pigeonhole principle we know there is distinct i, j such that $a^i = a^j$. Without loss of generality, suppose $i > j$. Then $a^{i-j} = e$. □

Theorem 107. *For any way to place five points in a 2×2 square, there must exist two points whose distance apart is less than or equal to $\sqrt{2}$.*

Proof. Divide the 2×2 square into four 1×1 squares sections. Since there are five points and four sections, by the pigeonhole principle, there must exist two points in the same section. Within the section, the farthest these points could be apart is the hypotenuse. By the pythagorean theorem, we know the hypotenuse length is $\sqrt{1^2 + 1^2} = \sqrt{2}$. Thus the distance of these two points is $\leq \sqrt{2}$. □

Notice how we had to construct the pigeonholes ourselves. They were not served to us. A skilled construction can get a stronger theorem.

Theorem 108. *For any way to place 101 points in a 6×8 rectangle, there must exist two points whose distance apart is less than or equal to 1.*

Proof. We give three constructions and improve each. First, consider that we divide the 6×8 rectangle into 48 1×1 squares. By the pigeonhole principle, we know that there must exist some square with atleast two points, which are of maximum distanct $\leq \sqrt{2} \approx 1.4$ apart.

Next, consider that we divide the 6×8 rectangle into 96 0.5×1 rectangles. By the pigeonhole principle, we know that there must exist some rectangle with atleast two points, which are of maximum distanct $\leq \sqrt{(0.5)^2 + 1^2} \approx 1.11$ apart.

Finally, consider that we divide the 6×8 rectangle into 100 0.6×0.8 rectangles. By the pigeonhole principle, we know that there must exist some rectangle with atleast two points, which are of maximum distanct $\leq \sqrt{(0.6)^2 + (0.8)^2} = 1$ apart.

□

Theorem 109. *Suppose you colored every point of the cartesian plane $\mathbb{R}^2$ either red of blue. Then for each distance $d > 0$, there exists a pair of points exactly d apart which are the same color.*

crazy

Proof. Form an equilateral triangle anywhere of sidelengths d. Consider the coloring of the endpoints of the triangle. There are three points and two colors. By the pigeonhole principle, there exists two points of the same color. These points are also exactly d apart. □

Observe how the pigeonhole principle was used.

Theorem 110. *If you have n pigeons and k holes, then there exists a hole with $\lceil \frac{n}{k} \rceil$ pigeons.*

Proof. Assume to the contrary you have n pigeons placed into holes $S_1, ...S_k$ and each hole has $|S_i| < \lceil \frac{n}{k} \rceil$ pigeons in it. Then $|S_i| \leq \lceil \frac{n}{k} \rceil - 1 < \frac{n}{k}$, so each hole must have strictly less than $\frac{n}{k}$ pigeons in it. But

$$n = |S_1 \cup ... \cup S_k| = |S_1| + ... + |S_k| < \frac{n}{k} + ... + \frac{n}{k} = k\frac{n}{k} = n$$

which would imply $n < n$, contradiction. □

The idea is if no pigeonhole was above $\lceil \frac{n}{k} \rceil$, then there was never n pigeons to begin with.

Theorem 111. *As of writing, three people in class made the same exam grade.*

Proof. There are 263 students in class, and 101 possible exam grades. By the generalized pigeonhole principle, $\lceil \frac{263}{101} \rceil = 3$ people must have made the same score. □

commentary on distribution.

Theorem 112. *No matter how five points are placed on a sphere, there exists a way to draw a great circle*[60] *such that 3 points lie on one side.*

[60] A great circle is like the equator, any line which separates a sphere into two equally sized hemispheres.

Proof. Draw a great circle anywhere. This will partition the sphere into two separate hemispheres. Since there are five points but two hemispheres, by the generalized pigeonhole principle, $\lceil \frac{5}{2} \rceil = 3$ points must lie on the same hemisphere. □

By choosing a great circle carefully, and taking advantage of the geometry of the sphere, we can actually strengthen this.

Theorem 113. *No matter how five points are placed on a sphere, there exists a way to draw a great circle such that 4 points lie on one side.*

Proof. content... □

Theorem 114. *six mutual friends or six mutual enemies*

[61]

Erdos/Szekeres Theorem

[61] Suppose aliens invade the earth and threaten to obliterate it in a year's time unless human beings can find the Ramsey number for red five and blue five. We could marshal the world's best minds and fastest computers, and within a year we could probably calculate the value. If the aliens demanded the Ramsey number for red six and blue six, however, we would have no choice but to launch a preemptive attack.