Basic Logic and Mathematical Structures for COMP 330 Fall 2018

Prakash Panangaden
McGill University

27th August 2018

These notes are not meant as a substitute for learning the subject of the title properly. They are meant to make sure we have some basic vocabulary in place. The section on "Logical Connectives", for example, is not an *ab initio* explanation; rather, it is a brief reminder. Other sections are more expository in character.

Prolegomenon to any future mathematics

One of the greatest difficulties students have in this course is difficulty with *reading* mathematical notation. They are used to natural language with its usual ambiguities and conventions. This level of ambiguity is not appropriate for mathematics. It is **essential** that **you**[1] exercise care and *discipline* in reading and writing mathematics. Even the surrounding text is important; don't look at a mathematical formula in isolation. Use complete sentences. **Be careful about which adjectives apply to which nouns.**

1 Logical Connectives

The basic logical connectives are: **and** $\wedge$, **or** $\vee$, **not** $\neg$ and **implication** $\Rightarrow$, usually rendered in English with the word "if".

[1]Yes, you!

People have the greatest problem with implication. First of all they seem to have a problem with the fact that implication is *not* symmetric. The word "if" signifies a one-directional implication. For bi-directional implication (equivalence) one uses the phrase "if and only if," in mathematical notation we write $p \iff q$. The statement $p \Rightarrow q$ is *not* the same as $q \Rightarrow p$. For example, "if I turn on the light, I will be able to see" is not the same as "If I can see, I turned on the light." However, in ordinary English one often does say "if" when one means "if and only if" (something we almost never say in natural conversation). For example, "If I get hungry, I will eat." Usually, one understands that "If I am not hungry, I will not eat." In mathematical writing, however, we *always* mean the one-directional implication when we say "if."

People have problems with implication and its converse and contrapositive. The **converse** of $p \Rightarrow q$ is $q \Rightarrow p$. In the preceding paragraph, I have emphasized the difference between an implication and its converse: *they are not the same*. The contrapositive of $p \Rightarrow q$ is $\neg q \Rightarrow \neg p$: this **has the same meaning** as the original implication. One common way of rendering $p \iff q$ is $p \Rightarrow q$ and $\neg p \Rightarrow \neg q$; for the second statement we have used the converse of the contrapositive. What is the negation of an implication? $\neg(p \Rightarrow q)$ is most usefully thought of as $p \wedge \neg q$.

How does one prove an implication like $p \Rightarrow q$? Here one **assumes** p and (using this assumption, if necessary) one proves q. The important point to realize is that one is under **no obligation to prove** p. I have proved statements of the form $p \Rightarrow q$ in class only to have students ask me afterwords "But how do you know p is true?" I don't know p is true and I am making no promises about p. I am saying that if p turns out to be true then I can promise that q will be true; if p happens not to be true no promises are made. An implication is a **contract**; if something else makes p hold then q will hold. If p does not hold the contract is off and no promises are made.

How does one prove a statement like $\neg p$? I like to think of $\neg p$ as shorthand for $p \Rightarrow$ **false**. So one way it to assume p and derive a contradiction. This is where "proof by contradiction" is *essential*. If you want to prove $p \Rightarrow q$ one often assumes p *and* $\neg q$ and then derives a contradiction. Essentially they are proving $\neg\neg q$ which is equivalent to q[2] This is fine, but sometimes people are "trigger-happy" to use this approach. Look for a direct proof if you can.

[2]This is not true in intuitionistic logic.

People generally understand what *and* and *or* mean but sometimes they are so sloppy they confuse the two. Usually students do not need to have this explained to them, but, on occasion, they need to have it pointed to them that they have confounded[3] the two.

The basic rule for using implication is given the fancy name *modus ponens*[4]:

$$(p \Rightarrow q) \wedge p \Rightarrow q.$$

If I have proved $p \Rightarrow q$ and also proved p then I am entitled to conclude q. Unfortunately, many people use incorrect reasoning principles like this one:

$$(p \Rightarrow q) \wedge q \Rightarrow p.$$

I think this stems from the subconscious confusion of implication and equivalence. The classic example of this is from the film *Monty Python and the Holy Grail*[5]: If X is made of wood then X floats in water hence if X floats in water X is made of wood.

2 Set notation

I assume familiarity with set notation $\{\cdot \mid \cdot\}$. In order to avoid confusion about the "universe" that we are talking about one often writes something like

$$\{x \in A \mid \phi(x)\}.$$

This stands for the collection of all things *within* the set A that satisfy the condition $\phi(\cdot)$. Note the implicit universal quantifier in the set notation. I am not going to belabour the well-known set operations: $\cup, \cap, \setminus$ or the notation $\emptyset$ for the empty set. Sometimes I use the notation $\overline{X}$ for the complement of a set X. One source of confusion (arising from careless use of language) is the distinction between containment and membership. We write $A \subset X$ or $A \subseteq X$ for A is a subset of X and we say X *contains* A. This is a relation between two sets. By contrast, we write $a \in X$ to mean that a is one of the elements of X. We say "a is a member of X."

[3] The word "confounded" means mixing up two different things.

[4] Using this name only makes you sound pompous, it is the logical principle that is important, not the name.

[5] Search on YouTube for "The Witch Scene by Monty Python".

Often we talk about sets whose elements are themselves sets. The phrase "set of sets" sounds clumsy to the ear so we use phrases like "family of sets" or "collection of sets" instead.

One very important operation on sets is the cartesian product. Given two elements a and b we define a new element, written $\langle a,\ b\rangle$, called the *ordered pair* of a and b. It is simply a new element consisting of two pieces in a particular order. We define two pairs to be equal if and only if the corresponding components are equal, symbolically:

$$\langle a,\ b\rangle = \langle c,\ d\rangle \text{ iff } (a = c) \wedge (b = d).$$

We introduce the *projections*, π_1 and π_2, that take pairs apart:

$$\pi_1(\langle a,\ b\rangle) = a \text{ and } \pi_2(\langle a,\ b\rangle) = b.$$

Given two sets A and B we define their cartesian product (or just product) to be the collection of all ordered pairs:

$$A \times B \overset{\text{rel}}{=} \{\langle a,\ b\rangle \mid a \in A,\ b \in B\}.$$

3 Quantifiers and games

The meaning of quantifiers is subtle: one understands that $\forall$ means "for all" and that $\exists$ means "there exists" but reading them *in context* can be complicated. There are two crucial things to keep in mind: (a) the *order* of the quantifiers is essential,(b) the conceptual complexity arises from *alternation* of quantifiers. In this section I describe a way of understanding quantifiers by using games.

The first point can be explained by the following example. Which of the two following formulas – the variables all refer to positive integers – is correct? The first: $\forall n\ \exists m\ n < m$ or the second: $\exists m\ \forall n\ n < m$. We read these as follows: the first is read "for all n there is some m such that $n < m$". The second one is read "there is some m such that for every n, $n < m$." Intuitively the first says, given any number I can find a bigger number: clearly true. The second says, I can find a number bigger than all other numbers, obviously false. But what is the difference in these two formulas? It is only the *order* in which the quantifiers are written, apart from that they are exactly the same.

Notice the phrase "I can find"; it suggests an *interactive reading* of these formulas and the most common activity where people interact is playing games. I will formalize the quantifiers in terms of games. I will restrict to the case where all the quantifiers appear in front of the formula. So things like the formulas above are acceptable but I will not attempt to explain formulas like $\forall x\ \exists y\ [Q(x,y) \Rightarrow \exists z\ R(x,y,z)]$.

The game consists of taking a formula with variables, constants and other symbols – but no quantifiers – and putting explicit values for the variables. A *move* corresponds to choosing a value for a variable. Who are the players? Here is where the quantifiers come in. We have two players called **Abelard** and **Eloise**[6]: Abelard's turn corresponds to the $\forall$ quantifiers and Eloise's turn corresponds to the $\exists$ quantifiers. The order in which the quantifiers are written determines the order in which they play. What is the point of the game? Winning of course! We will say that Abelard wins if the formula that you get with all the values filled in is false and Eloise wins if the formula with all the values filled in is true. Now here is the main point: **The original quantified formula is true if and only if Eloise has a winning strategy**. Note the phrase "winning strategy"; she must be able to respond to *anything* that Abelard does. She cannot rely on him making a stupid move. It is perfectly possible for there to be a situation where Eloise wins because Abelard has made a dumb move. This does not count, Eloise has to be able to win *no matter what Abelard does.* If she cannot do that she does not have a winning strategy and the original quantified formula is false. Why is the order important? Because if you are playing second *you know your opponent's move* and can respond accordingly.

Consider the two formulas above. Let us play the game with $\forall n\ \exists m, n < m$. Abelard plays first, he picks some number, call it k, it could be 17 or 273 or 1729 or whatever. Since Eloise goes next, *she knows what Abelard has played*, and she can choose her move accordingly. If Abelard chose 17 she can choose 18, if he chose 273 she can choose 274, if he chose 1729 she can choose 1730, in each case she wins because $17 < 18$, $273 < 274$ and $1729 < 1730$. In general, if he chooses k she can choose $k+1$. She has a *strategy*, which is "whatever Abelard picks, I will pick the next bigger number." She did not win just because Abelard made a dumb choice.

Now what happens with the other formula? This formula reads $\exists m\ \forall n, n < m$. Now Eloise has to go first and Abelard gets to see what her move is before he has to play. Can she pick a number that he cannot beat? Obviously not!

[6]Use web search to find the origin of these names.

Here Abelard has a winning strategy. This quantified formula is false.

Sometimes we say "player" and "adversary" instead of "Eloise" and "Abelard." You can think of the upside-down A as a mnemonic for "adversary." The word "satan" in Hebrew means "adversary" so we sometimes call the $\forall$ player the "demon." We will use this language when we discuss the pumping lemma later.

4 Relations

A relation is simply a correspondence between the elements of a number of sets. If we have just two sets we get a special case called a *binary relation.* I will focus on binary relations in this note. Intuitively you should think of a relation as being given by some kind of rule or description.

How can a relation in general be represented? We cannot write down a general pattern of all possible rules. Instead we think of the relation as being given somehow and the *instances* of the relation as *pairs.* Think of the related elements as being in a table. For example, consider the relation brother-of in a particular family:

Person	Brother-of
Prakash	Prema
Pradeep	Prema
Prakash	Pradeep
Pradeep	Prakash

Note that if x is a brother of y it does not mean that y is a brother of x; y might be a girl as in the example above with Prema.

Formally a **binary relation** between sets A and B is just a subset of $A \times B$. A special case is when $A = B$, in this case we say that we have a binary relation on A. There is no need for the sets in question to be finite. For example, the relation predecessor-of defined on natural numbers consists of the pairs $\{\langle 0,\ 1\rangle, \langle 1,\ 2\rangle, \langle 2,\ 3\rangle, \ldots\}$. Sometimes we define a relation through a rule or description and we refer to the set of ordered pairs as the **graph** of the relation.

Given two sets A and B we commonly use the letter R to stand for a generic (binary) relation. Thus we write $R \subseteq A \times B$. There are two kinds of notations commonly used: $\langle a,\ b\rangle \in R$ or, more commonly, aRb. We can visualize R as a table or as a graph with edges going from A to B.

Note that it is not required that every element of A is related to some element of B nor that an element of A be related to only one element of B. There are all kinds of special relations that are of interest.

Definition 1. A relation f from A to B is called a **function** if for every element $a \in A$ there *is* a *unique* element $b \in B$ such that afb or $\langle a,\ b\rangle \in f$.

The relation brother-of is certainly not a function. I expect you know all about functions and special cases of them like surjections (onto functions), injections (one-to-one functions) and bijections (one-to-one and onto).

Definition 2. Given a relation $R \subseteq A \times B$ we define the **converse** relation, written R^c to be the collection of pairs $\langle b,\ a\rangle \in B \times A$ such that $\langle a,\ b\rangle \in R$.

For any relation R the converse is a well-defined relation. Given a function f the converse is not necessarily a function though it is always defined as a relation. In the very special case when the function is a bijection the converse is also a function.

A crucial operation on relations is *composition*

Definition 3. Given relations $R \subset A \times B$ and $S \subset B \times C$ we define the relation $R \circ S \subset A \times C$ by

$$a(R \circ S)c \text{ iff } \exists b \in B, aRb \wedge bSc.$$

We use the special symbol I_A (or just I if A is understood from context) for the identity relation on A:

$$aI_Aa' \text{ iff } a = a'.$$

You should check that $I_A \circ R = R$. It is common to call I *equality.*

4.1 Equivalence relations

An extremely important class of relations are the equivalence relations. They are defined on a single set, say X, so they are subsets of $X \times X$. They abstract the notion of equality so they are defined to have the crucial properties of equality.

Definition 4. A binary relation R on X is said to be **reflexive** if $\forall x \in X,\ xRx$; equivalently $I_X \subset R$. R is said to be **symmetric** if xRy implies

yRx; equivalently $R^c = R$. R is said to be **transitive** if

$$\forall x, y, z \in X,\ xRy \wedge yRz \Rightarrow xRz,$$

or, equivalently $R \circ R \subset R$. A relation that is reflexive, symmetric and transitive is called an **equivalence relation**.

Common symbols for equivalence relations are $\sim, \simeq, \approx, \equiv$. They all look like variations of the basic equality symbol.

Definition 5. Given a set X and an equivalence relation R defined on it, we define, for each element $x \in X$ the set $[x] \overset{\text{def}}{=} \{y \in X \mid xRy\}$, called the **equivalence class of** x.

The proof of the following crucial fact is left as an exercise[7]:

Proposition 6. Given a set X and an equivalence relation R defined on it and given $x, y \in X$ we have **either** $[x] = [y]$ **or** $[x] \cap [y] = \emptyset$.

The collection of equivalence classes is a set in its own right: it is often called the **quotient** of X by R and we write X/R for this set.

Here is another important theorem about equivalence relations, please prove it.

Proposition 7. Given two equivalence relations R_1, R_2 on the same set X, the relation $R_1 \cap R_2$ is also an equivalence relation. In fact given *any* family $\{R_i \mid i \in \mathcal{I}\}$ of equivalence relations on X, the intersection $\bigcap_{i \in \mathcal{I}} R_i$ is an equivalence relation.

From this it follows that given any relation R there is a *smallest* equivalence relation containing R. To see this we just take the intersection of all equivalence relations containing R. The proposition above assures us that it is an equivalence relation, it clearly contains R and it is obviously the smallest such equivalence relation. How do we know that there are any equivalence relations containing R? Well the universal relation $U_X \overset{\text{def}}{=} X \times X$ is one such.

Note that the union of two equivalence relations need not be an equivalence relation. Why not?

[7] Do it!

4.2 Partial orders

Just as equivalence relations are an abstraction of equality, there is another class of relations that abstract inequality. These are the relations that abstract the notion of "ordering." However, unlike the usual notion of ordering of real numbers or integers we will **not** insist that every pair of elements are related.

Definition 8. Let S be any set. A **partial order** on S is a binary relation, usually written $\leq$, satisfying

1. $\forall x \in S.x \leq x$ [reflexivity]
2. $\forall x, y \in S.(x \leq y) \wedge (y \leq x) \Rightarrow (x = y)$ [antisymmetry].
3. $\forall x, y, z.(x \leq y) \wedge (y \leq z) \Rightarrow (x \leq z)$ [transitivity].

We are **not** requiring

$$\forall x, yS,\ x \leq y \text{ or } y \leq x.$$

If a partial order does satisfy this additional condition we call it a **total order** or a **linear order**.

Typical examples are the set of integers ordered by the usual notion of less than or equal, the set of subsets of a given set ordered by inclusion, or the set of complex numbers ordered by their magnitude. Henceforth we shall say *poset* rather than "set together with a partial order". Of course a given set may have many partial orders defined on it. Note that this notion does not include the usual notion of "strictly less than". One can introduce the symbol $x < y$ as an abbreviation for $x \leq y$ and $x \neq y$. When I refer to this concept I will use the phrase "strictly smaller" or "strictly less than" or "strictly decreasing". When I just say "less than" I mean what one would call "less than or equal to" in ordinary language.

Given a subset $X \subset S$ of a poset S we say that the element x_0 is the *least* element of X if it is less than every other element of X. In symbols

$$\forall x \in X,\ x_0 \leq x.$$

A given set may or may not have a least element. For example if we look at all the negative integers there is no least element. If we look at all the positive fractions there is no least element. In both these examples we had a total order. If we have a partially ordered set we can have the following situation. We say that an element $m \in X$ is *minimal* if it is not strictly greater than anything else in X. In symbols

$$\forall x \in X,\ \neg(x < m).$$

The element m could be less than x *or unrelated to it.* Consider the non-empty subsets of the $\{a, b, c\}$ ordered by set inclusion. The singleton sets $\{a\}, \{b\}, \{c\}$ are all minimal but there is no least element in the set. If a set does have a least element then it is obviously also (the only) minimal element.

5 Well-founded orders and Induction

In this section I will prove the principle of induction for general well-founded orders. The note is a bit terse and is intended to be a supplement to the lecture. It is, however, complete and you can - in principle - learn everything that you need to know for this class and beyond, about induction, from this note.

I assume that you have studied the principle of mathematical induction and used it to prove identities like $\Sigma_{i=1}^{n} i^2 = \frac{1}{6}n(n+1)(2n+1)$.

For many students this principle appears as a magic wand. My experience is that many are unsure of the status of this principle; "is it an assumption?", "a definition?" or a "theorem?". The short answer is that it is a theorem. Its range of applicability is much wider than the exercises that you had to do in discrete mathematics classes may have suggested.

The theory of induction is deep and difficult as can be seen by glancing at a book like "Elementary Induction on Abstract Structures" by Y. Moschovakis. What we do in this note is a small exercise in chapter 1 of that book. It will, however, suffice for the applications that we have in mind. The main theorem that we prove rests on the crucial notion of well-founded order.
Definition 9. Let $(S, \leq)$ be a poset. The order relation $\leq$ is said to be **well founded** if every nonempty subset has a minimal element.

Note, I did not say "minimum", just "minimal". The set of subsets of a *finite set* form a well-founded order. The set of positive integers forms a well-founded order, but the set of all integers do not. The set of subsets of an infinite set, ordered by inclusion, does not form a well-founded subset. The set of positive rationals does not form a well-founded order. But almost every structure that arises in computer science does form a well-founded order.

Proposition 10. A poset S is a well-founded order iff there are no infinite descending sequences in S.

Proof Immediate from the definition. ∎

One can see why the negative integers (and hence all integers) do not form a well-founded order.

5.1 Induction

So who cares about well-founded orders anyway?

Definition 11. Suppose that $(S, \leq)$ is a poset, we say that the **principle of induction** holds for S if for any predicate P, we have

$$(\forall x \in S.((\forall y \in S.(y < x \Rightarrow P(y)) \Rightarrow P(x))) \Rightarrow (\forall x \in S.P(x)).$$

This generalizes the usual notion of induction. You may notice that the usual notion of "base case" appears to be missing. This is implicitly included in the above formula since, if u is a minimal element, the (empty) antecedent above implies that $P(u)$ is true.

The reason for studying well-founded orders is contained in the following theorem.

Theorem 12. For a poset $(S, \leq)$ the principle of induction holds if and only if the order relation $\leq$ is well-founded.

Proof Suppose that $\leq$ is indeed a well-founded order. Let P be any predicate. We need to show that the principle of induction holds; since this has the form of an implication we need to show that the consequent holds whenever we assume the antecedent. Accordingly we assume the antecedent. Suppose that $\forall x \in S.(\forall y \in S.y < x \Rightarrow P(y)) \Rightarrow P(x)$. We must show that $\forall x \in S.P(x)$. Consider the set $U = \{u \in S | \neg P(u)\}$. What we need to show is that U is empty. Suppose that U is not empty then, because U is a subset of S, which we have assumed is a well-founded order, there must be a minimal element u_0. Thus $\forall y \in S.y < u_0 \Rightarrow y \notin U$, in other words if y is less than u_0 it must satisfy P. But according to our assumption, whenever everything less than u_0 satisfies P, u_0 itself must satisfy P. But now we have a contradiction because $u_0 \in U$ which by definition consists of elements that *do not* satisfy P. Thus our original assumption that U was not empty must be false, which is to say that $\forall x \in S.P(x)$.

Now for the reverse direction. Suppose that the principle of induction holds for $(S, \leq)$. Consider the predicate $F(x)$ defined as "there are no infinitely decreasing chains of elements of S that are all less that x". Now suppose that x is some fixed element of S and that somehow we know that for all elements y that are strictly less than x, $F(y)$ holds. It must be the case that $F(x)$ holds because suppose that there is a decreasing sequence $x_1 > x_2 > x_3 \ldots$ all below x. Then x_2 is strictly less than x and there is an infinite, strictly decreasing sequence below it; in other words $\neg F(x_2)$. But we assumed that for all y strictly less than x – and x_2 is certainly in this collection – that $F(y)$ does hold. Thus such a chain cannot exist. Thus we have shown that

$$\forall x \in S.(\forall y \in S.y < x \Rightarrow F(y)) \Rightarrow F(x)$$

which means we can apply the principle of induction to $F(x)$ to conclude that $\forall x.F(x)$. By our proposition above, this means that the poset is a well-founded order.

Here is another, slicker, argument for the reverse direction. Suppose that there is a set $U \subset S$ which has no minimal element. This means that $\forall u \in U\ \exists v, v < u$ and $v \in U$. Now consider the predicate $P(x) \overset{\text{def}}{=} x \notin U$. Suppose that for any $x \in S$ we know that $\forall y < x\ P(y)$. I claim that $P(x)$ must hold. If $P(x)$ does not hold then $x \in U$ but our assumption about x is that every element $y < x$ is not in U (this is what $P(y)$ means) so, in other words, x is a minimal element of U. But this contradicts the assumption about U. Thus we have proved

$$\forall x \in S\ [\forall y < x\ P(y)] \Rightarrow P(x).$$

We are assuming that the principle of induction holds and the formula above is exactly the premise of the principle of induction. Thus, we have proved $\forall x\ P(x)$. But what is P? We have just proved $\forall x\ x \notin U$, i.e. U is the empty set. In short every non-empty set must have a minimal element. ∎

Now we can check that the so called principle of mathematical induction is a special case of this. The structure of interest here is the natural numbers i.e. the set $N = \{0, 1, 2, 3, \ldots\}$. This is clearly a well-founded order so the principle of induction is true. It only remains to put this in the familiar form. Let P be any predicate. There is a unique minimum element namely 0. Thus if we choose x to be 0 in the statement of the principle of induction above we have to show that $(\forall y \in N.y < 0 \Rightarrow P(y)) \Rightarrow P(0)$. The antecedent is vacuously true thus we must show that $P(0)$ is true. Now if we choose x to be $n+1$ we get

$$(\forall y \in N.y < (n+1) \Rightarrow P(y)) \Rightarrow P(n+1)$$

or rewriting this less clumsily we can say $\forall y \leq n.P(y) \Rightarrow P(n+1)$. Putting the pieces together we get the usual statement i.e.

$$(P(0) \wedge (\forall n.(\forall y \leq n.P(y) \Rightarrow P(n+1)))) \Rightarrow \forall n \in N.P(n).$$

The other principle that interests us in the principle of structural induction. Let S be any inductively defined set. Associated with the elements of S is the stage at which they enter the set S. We can define a well-founded order on S by saying that $x \leq y$ if x enters S at the same stage or before y in the inductive definition. Thus we can apply induction to this collection. An example will illustrate the idea.

Suppose that we define the set of *binary trees* as follows. The empty tree, written [] is a tree. If t_1 and t_2 are trees then $maketree[t_1, t_2]$ is also a tree. This is a typical inductive definition. Now the principle of induction applied to these tree structures says the following. Let P be any predicate on trees. If you can prove $P([])$ and if you can prove that for any trees t_1 and t_2 that whenever $P(t_1)$ and $P(t_2)$ holds then $P(maketree(t_1, t_2))$ also holds you can invoke the principle of structural induction to conclude that for all trees, t, $P(t)$ must hold. Thus we can do induction on trees and a multitude of other structures as well.

5.2 Zermelo's theorem

Please do not read this section. It will torment you needlessly.
Definition 13. A **well-order** is a well-founded total order.

The usual order on the natural numbers is a well-order. Given a well-order every element, except the largest one if such an element exists, has a unique next element. Many examples of well-ordered sets can be given. On the other hand try to find a well-order on the real numbers. After trying for a while one tends to think that this is not possible.

However, using the Axiom of Choice, Zermelo proved
Theorem 14. Every set can be well ordered.

This created a sensation. The axiom of choice seems like an “obvious” principle but Zermelo’s theorem seems unbelievable. As of yet no one can write down an explicit example of a well-order on the reals. This leads many logicians to reject the axiom of choice. However, the axiom of choice is equivalent to Zermelo’s theorem and also to Zorn’s lemma. The latter is essential to prove all kinds of theorems in mathematics: for example, every vector space has a basis. Thus most mathematicians are loath to give up the axiom of choice.

6 Infinite vs finite

Many students are confused by finite vs infinite numbers. It is not my intention to discuss cardinality here or to give a detailed discussion of infinity. I just want to recall some quick facts. I assume you know the words *injection* (one-to-one function), *surjection* (onto function) and *bijection*. A set X is infinite if there is a bijection between X and a proper subset of X. Thus, the natural numbers constitute an infinite set because they are in bijective correspondence with the set of even numbers, which is a proper subset of the set of all natural numbers.

There is no such thing as an infinite natural number. The natural numbers $\{0, 1, 2, 3, 4, \ldots\}$ are *each* finite; the **collection** of all of them taken together is an infinite set. Please use the adjectives with the appropriate nouns. If you speak sloppily you will be confused; if you speak clearly you will find that you are not confused. Unfortunately, even mathematicians have become used to a certain sloppiness in their language. For example, one hears the statement, “There is an infinite number of foobars such that...” Of course they mean, “the set of foobars is an infinite set” but it would be tedious and sound pedantic to say it like that. As long as you understand what is really meant, it is acceptable to use such idioms.

If you do not know that there are different infinities and have never seen a diagonalization argument you should learn these things now.

7 Alphabets, words and languages

The basic structures that we will deal with are called *languages*. An *alphabet* is just an arbitrary set, in this course we will always consider finite alphabets. We view the elements of an alphabet as *uninterpreted* symbols. Typically, we use the letters $a, b, c, \ldots$ or perhaps the set of bits $\{0, 1\}$ or some such thing. We call the elements of the alphabet *letters.* It is traditional to use the capital greek letter Σ to represent an alphabet.

From letters we build *words.* A word is just a sequence, or string if you prefer, of letters. Letters and words are just like characters and strings in a programming language. Letters may be repeated and may occur in any order. We will always take strings to be finite, thus they have a well defined *length.* If our alphabet is $\{a, b\}$, for example, then examples of words constructed from this alphabet are $abba, bab, aaaa, bbaabba$ and b; they have length $4, 3, 4, 7$ and 1 respectively[8] One very important string is the *empty string* or *empty word* with no letters whatsoever. We need a symbol for this; if we were just to leave a blank space like this   we cannot[9] tell whether this is the empty word or just an anomaly in the typesetting. We write ε for the empty word. We typically use symbols like u, v, w, x, y, z to stand for words. We write $\mid x \mid$ for the length of the word x. Note that a single letter by itself is also regarded as a word. The empty word has, of course, length zero, in symbols, $\mid \varepsilon \mid= 0$.

The set of all possible words over the alphabet Σ is written Σ^*. This set explicitly includes ε. The set Σ is (usually) finite but the set Σ^* is (almost) always infinite. The only time it is not infinite is when $\Sigma = \emptyset$. Then only one word can possibly be constructed, namely the empty word. Thus we have $\emptyset^* = \{\varepsilon\}$. Note $\emptyset^* \neq \emptyset$. Even if Σ has only one element, say $\Sigma = \{a\}$, then $\Sigma^* \simeq \mathbb{N}$ in an obvious way.

The set Σ^* is equipped with a *binary operation* written $\cdot$, often just indicated by juxtaposition, and called **concatenation**. It is the simple operation of sticking two words together. For example $(abba) \cdot (baba) = abbababa$. This operation is not commutative, for example $(ab) \cdot (ba) = abba \neq (ba) \cdot (ab) = baab$. It is, however, *associative*: $(x \cdot y) \cdot z = x \cdot (y \cdot z)$. This means that we do not have to be too scrupulous about bracketing. It becomes tedious to write $\cdot$ all the time[10], so *almost all the time* we just use juxtaposition. For example, the associative rule is written $\forall x, v, z \in \Sigma^*, (xy)z = x(yz)$. This concatenation operation has an identity element, namely ε. In symbols

$$\forall x \in \Sigma^*, \varepsilon x = x\varepsilon = x.$$

[8]It is possible to imagine infinite words and there is a rich theory of infinite words, but we will not consider them in this class.

[9]Please use the word "cannot", people who write "can not" are using it incorrectly.

[10]Especially in LaTeX!

A set with a binary associative operation is called a *semi-group.* If a semi-group operation has an identity element it is called a *monoid.* Other examples of monoids are the set of non-negative integers (the natural numbers), the set of all $n \times n$ matrices with integers entries and with matrix multiplication as the operation and the set of all functions from a fixed set to itself with function composition as the operation[11].

Finally we define a **language** as a subset of Σ^*. A language can be empty, it can be all of σ^* or it can be something in between. Here are examples of languages over the alphabet $\{a, b\}$:

1. The set $\{abba, baba, ab, baa, baabaa, b\}$.
2. The set of all words that begin with a: $\{ax \mid x \in \Sigma^*\}$ $= \{a, aa, ab, aab, aba, aaa, abb, \ldots\}$.
3. The set of all words that contain a sequence of a's followed by a sequence of b's: $\{a^n b^m \mid n, m \geq 0\}$.

In the first example I have just listed all the words; this works for finite languages. Some languages are finite, as in the first example but others are not. We have to come up with a better scheme for describing them than listing the elements. In the second example I have used mathematical notation to give a description. I have listed a few examples of words in this language but that is only illustrative; it is definitely not a formal definition. In the third example I have used a new mathematical notation that is used in language theory. When I write a^n, I mean a repeated n times. Thus, for example, $a^3b^5 = aaabbbbb$. It has **nothing to do with** exponentiation. Remember letters in the alphabet are *uninterpreted* symbols, they do not stand for numbers, functions, goats or anything else. The form $a^n b^m$ then just means a sequence of n a's followed by a sequence of m b's. Note that the n and m are allowed to be arbitrary natural numbers so this is how we say "any bunch of a's followed by any bunch of b's." What do we do if we want to say "some bunch of a's followed by the *same number* of b's?" We write $a^n b^n$; here it is understood that n is an indeterminate number but whatever it is, all its occurrences mean the same thing. Note that writing a^{-3} is plain nonsense, it does not mean anything.

How do we specify languages more systematically? How do we *algorithmically* check whether a given word is in a language? How difficult is it to

[11]If the operation has, in addition, an inverse for every element, it is called a *group*. None of the examples that I have given are groups.

describe a language? These are questions that will occupy us the rest of the term. Language recognition is not the only thing computers do but essentially every aspect of theoretical analysis of computer science — including complexity theory — can be cast in terms of language recognition.

Concluding remarks

I will not write detailed notes like this again. This was written to make sure we are on the same page. Speaking bluntly, if all this seemed way over your head, you should drop the class. If all this seemed childish be patient. Future notes, which will **only be for material not in the book** will be more terse and to the point. The content will also be more mature.