Sets, Groups and Knots
Course Notes
Math 101, Harvard University
C. McMullen

Contents

1 Introduction

This course provides an introduction to conceptual and axiomatic mathematics, the writing of proofs, and mathematical culture, with sets, groups and knots as the main topics.

Here is a rapid overview of the three main topics we will consider.

1.1 Set theory

We will start with the rock–bottom foundations of mathematics, and learn how to count, avoid paradoxes and show there are different sizes of infinity.

Notation. Here are some basic sets (of numbers) we are all familiar with:

- $\mathbb{N} = \{0, 1, 2, 3, \ldots\}$ denotes the natural numbers;
- $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$ denotes the integers;
- $\mathbb{Q} = \{p/q \ : \ p, q \in \mathbb{Z}, q \neq 0\}$ denotes the rational numbers;
- $\mathbb{R}$ denotes the real numbers, $x = n + 0.x_1x_2x_3\ldots$, which include π, $\sqrt{2}$, etc.; and
- $\mathbb{C} = \{x + iy \ : \ x, y \in \mathbb{R}\}$ denotes the of complex numbers.

It is also useful to have shorthand notation for the set of positive elements: $\mathbb{N}_+ = \mathbb{Z}_+$ and $\mathbb{R}_+$.

The subset of elements of A that have a *multiplicative inverse* in A are usually denoted by A^*. Thus $\mathbb{C}^*$, $\mathbb{R}^*$ and $\mathbb{Q}^*$ are obtained from $\mathbb{C}$, $\mathbb{R}$ and $\mathbb{Q}$ by discarding the single element $x = 0$. On the other hand, $\mathbb{Z}^* = \{\pm 1\}$.

Counting. But what is 3? One answer: 3 is 3 firetrucks without the firetrucks. In other words, numbers are abstractions of the idea of cardinality. We will see later how to inductively define all these notions using set theory.

Infinite sets. There are many equivalent definitions of an infinite set: A set A is infinite if it is:

(a) Not bijective to a natural number.
(b) Contain a copy of all the natural numbers.
(c) Bijective to a proper subset of itself (Hilbert's Hotel).

The set of all natural numbers, all integers, all even numbers, etc. are all the same size.

Theorem 1.1 $|\mathbb{R}| > |\mathbb{Z}|$.

Is there anything in between? This question has no answer! (It is *independent.*)

Critique. The idea of infinity is now commonplace in mathematics, but it was not always the case. Gauss rejected the idea of a completed infinity like $\mathbb{N}$ (as opposed to a potential infinity, like arbitrarily large numbers).

One might think that the Greeks required infinity, because they worked with lines and lines have infinitely many points. But in fact it is a profoundly modern idea, and a controversial one, to say that a line *consists* of the set of points on it. A more geometric point of view is that lines and points are primitive or atomic notions. A point can be on a line (incident to a line), but it is absurd (?) to imagine a continuous object like a line to be an amalgam of individual points.

1.2 Group theory.

A group is an algebraic structure that captures the idea of symmetry without an object.

Informally, a set G is group if we can form $a \cdot b$, there is an element $1 \in G$ such that $1 \cdot a = a \cdot 1 = a$, and there exist inverses: $a^{-1} \cdot a = 1$.

Examples:

1. $\mathbb{R}^*$, $\mathbb{Z}$. The integers under multiplication are *not* a group.

2. The symmetries of a triangle.

3. The advancing of a clock by one hour. $H^{12} = 1$. If we allow flipping the clock over, then $FH = H^{11} = H^{-1}$.

4. The symmetries of a cube.

5. Permutations of 4 objects, drawn as wiring diagrams. The inverse is the mirror image. There are 24 elements to this group.

6. The cube also has 24 symmetries. It has 6 faces, 12 edges, 8 vertices. Is it an accident that these all divide 24?

7. The rhombic dodecahedron has 24 symmetries too. But it has 12 faces, 24 edges and 14 vertices. How is the 14 possible?

   Cultural asides: The rhombic dodecahedron is the three–dimensional analogue of the hexagon, i.e. it is the Voronoi cell for the densest possible sphere packing; compare honeycombs.

   The Euler characteristic, $V - E + F$, is 2 for these and all other convex polyhedra (and, more generally, subdivisions of the 2-sphere).

8. Further examples of nonabelian groups and subgroups. Parallel parking. Slide puzzle. Quantum mechanics, non–commuting numbers and quantum computing.

1.3 Knot theory

A knot is a closed loop in space, considered as a flexible object that cannot, however, pass through itself. The study of knots is part of the field of *topology*.

It is tricky to show that there is more than one type of knot! (There is a magic trick for tying a trefoil. We will prove that this trick is impossible!)

Are there infinitely many knots? Can the trefoil be converted to a figure eight? How can you tell knots apart?

One of the goals of this course is to show that a group can be associated to a knot, and group theory can be used to tell knots apart. This is the basic idea behind *algebraic topology*.

1.4 Logic, proofs, basic concepts

Before studying set theory we make some remarks about logic and proofs, assuming informally that we already know a certain amount of mathematics.

Logic.

1. First order logic. Truth table for AB, $A+B$, $\overline{A+B}$, $\overline{AB}$, $A \iff B$, $A \implies B$, $\overline{B} \implies \overline{A}$.

2. False $\implies$ anything. If $1+1=3$, then 15 is a prime number. If Lincoln is still alive, then Trump is Clinton's brother, and 21 is prime.

3. Contrapositive: $P \implies Q$ is *equivalent* to $\overline{Q} \implies \overline{P}$. Example: if $n \in \mathbb{Z}$ is a square, then $n \geq 0$. Equivalently, if $n < 0$, then n is not a square.

4. The *converse* of $P \implies Q$ is $Q \implies P$. For example, all square in $\mathbb{Z}$ are positive, but the converse is false: not all positive numbers are squares.

5. Tautology: an NP–complete problem. A tautology is a logical formula that is true no matter what values are assigned to its variables. As an example, we have
$$B + AC + \overline{C} + \overline{AB}C = 1.$$
A nice way to check this is with a Karnaugh map.

   No polynomial–time algorithm is known to determine if a given expression is a tautology. Common belief is that none exists. This is one of the 6 remaining Clay Prize Problems, each of which is worth a million dollars.

6. Quantifiers: $(\forall x \in A)P(x)$; $(\exists x \in A)P(x)$. Examples: $\forall x \in \mathbb{R}, x^2 \geq 0$; $\forall i,j \in \mathbb{Z}, i+j = j+i$. Note: $\exists x \in \mathbb{R} \ : \ x^2 = 2$. Uniqueness is not asserted! (Sometimes people use $\exists! x$ for uniqueness.)

7. Negation of quantifiers: $\sim (\forall x)P(x)$ is the same as $(\exists x) \sim P(x)$; similarly for $\exists$.

8. Example: $(\forall x \in \mathbb{R})(\exists y \in \mathbb{R}) \ : \ xy = 1$. False (correct if we use $\mathbb{R}^*$). The negation is true: $(\exists x) \ : \ (\forall y) \ xy \neq 1$. In fact, just take $x = 0$.

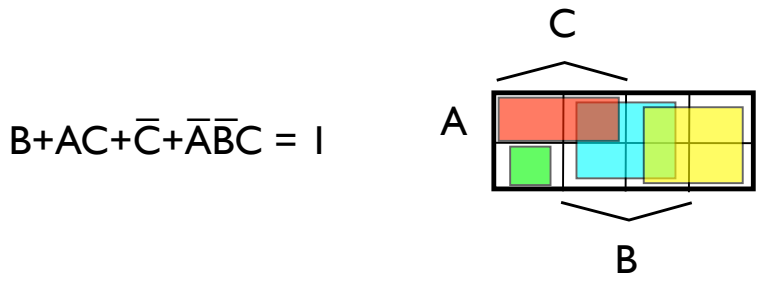

Figure 1. Verifying a tautology. Each of the 8 cells gives one of the possibly assignments of T/F to A, B, C.

9. General principle: to establish "If A then B", or "For all x satisfying A, we have B", you must give a proof. To *disprove* a statement of that type, you must give a *counterexample*. (E.g. $x = 0$ above.)

Linguistic fallacies. It is an interest exercise to listening for misused logic in common langauge, e.g. in advertising.

A typical linguist fallacy is "All aspirin is not alike." What the speaker is trying to do is negate the statement "All aspirin are alike". But they have confused two inequivalent statements:

$$\sim (\forall x) P(x) \quad \text{and} \quad (\forall x)(\sim P(x)).$$

The correct negation of $(\forall x)P(x)$ is $(\exists x) \sim P(x)$. The correct informal statement is, "Some aspirin are different".

Equivalence relations.

1. Ordered pairs and the product $A \times B$. Relations.

2. Equivalence relations: the same thing can have many names. Example on $\mathbb{Z}$: $a \sim b$ if $a - b$ is even.

3. The definition of $\mathbb{Q}$ by an equivalence relation on fractions. Why this is beneficial (e.g. define addition).

4. The definition of $\mathbb{Z}/10$. This forms a group under addition. What about under multiplication?

5. Equivalence relation: what is the rule behind the sequence 8, 5, 4, 9, 1, 7, 6, 10, 3, 2, 0? (Alphabetical order). Consider equivalence relation on the numbers 0 – 10 of having the same first letter. It can be represented as a directed graph: an arrow from a to b means $(a, b) \in R$.

6. Equivalence relations and sets that are not well-defined. Examples:

   $\{x \in \mathbb{Z}/10 \ : \ x$ is prime.$\}$ (Not well defined)
   $\{x \in \mathbb{Z}/10 \ : \ x$ can be represented by a prime.$\}$ (1,2,3,5,7,9)
   $\{x \in \mathbb{Z}/10 : x$ can be represented by infinitely many primes.$\}$ (1,3,7,9)

   The last is an important statement in number theory!

Induction. Principle of induction: if $S \subset \mathbb{N}$ satisfies $0 \in S$ and $n \in S \implies (n+1) \in S$, then $S = \mathbb{N}$. Alternatively, if we can prove $n+1 \in S$ once we know $\{0, 1, \ldots, n\} \subset S$, then $S = \mathbb{N}$. Here are some examples of its use.

For all $n \in \mathbb{N}$, $S(n) = 1 + 2 + \cdots + n = P(n) = n(n+1)/2$. Proof: True for $n = 0, 1$ and $S(n+1) = S(n) + 1$ which implifies to $P(n+1)$.

Every $n > 1$ *is a product of primes.* Proof: True for $n = 1$ (the empty product) and $n = 2$ (which is prime). Suppose true up to n. If $n+1$ is prime, then we are done; otherwise, $n + 1 = rs$ with $r, s \leq n$, and each of r and s is a product of primes, so $n + 1$ is as well.

Proof by induction that people can live arbitrarily long: let $P(n)$ be the assertion: it is possible to live n microseconds. Then $P(n) \implies P(n+1)$. (?)

The (Google) job interview. Each candidate holds a playing card to his forehead, so the others can see it but he cannot. Each time the second hand on the clock crosses 12, each candidates must call out if he or she can deduce that the card they hold is the ace of spaces.

But in fact all the cards are aces of spades!

The examiner tells them all, before they start, "At least one of you holds an ace of spades."

(1) What happens? (2) What did the examiner tell them that they did not already know?

(There are many variations on this theme.)

Proof by contradiction. One frequent use of the contrapositive is called *proof by contradiction.*

Example: *there are infinitely many prime numbers.* Proof (Euclid). Suppose not. Let $p_1, \ldots, p_n$ be all the primes, and consider $N = p_1 \cdots p_n + 1$. Then N must be divisible by some prime, but it leaves a remainder of 1 when divided by any p_i. This contradicts the fact that any $N > 1$ is a product of primes.

Logically, what we are doing to prove A is we show that $\overline{A} \implies B$ where B is false. The only way this formula can hold is if A is true; equivalently, $\overline{B} \implies A$ and thus A is true.

This method of proof is similar to tracing down a variation in a game of chess. I.e. we suppose our opponent 'plays $\overline{A}$', and show we can win in that case.

Coda: Writing proofs. When you hit a home run, you just have to step once on the center of each base as you round the field. You don't have to circle first base and raise a cloud of dust so the umpire can't quite see if you touched the base but will probably give you the benefit of the doubt.

2 Set theory

We are now ready to investigate the first of our three topics, the theory of sets. Set theory forms the *foundation* of all mathematics, hence its central importance. It is also provides the basic *language* in which all rigorous mathematical disciplines can be expressed.

2.1 The axioms of set theory.

Here is a quick summary of all the axioms we will need.

- Axiom I (Extension). A set is determined by its elements. That is, if $x \in A \implies x \in B$ and vice-versa, then $A = B$.
- Axiom II (Specification). If A is a set then $\{x \in A : P(x)\}$ is also a set.
- Axiom III (Pairs). If A and B are sets then so is $\{A, B\}$.
- Axiom IV (Unions). If A is a set, then $\bigcup A = \{x : \exists B, B \in A \,\&\, x \in B\}$ is also a set.

- Axiom V (Powers). If A is a set, then $\mathcal{P}(A) = \{B \,:\, B \subset A\}$ is also a set.

- Axiom VI (Infinity). There exists a set A such that $0 \in A$ and $x+1 \in A$ whenever $x \in A$.

- Axiom VII (The Axiom of Choice). For any set A there is a function $c : \mathcal{P}(A) - \{\emptyset\} \to A$, such that $c(B) \in B$ for all $B \subset A$.

The two axioms we have omitted are more technical: the Axiom of Foundation (or Regularity) and the Axiom of Replacement. We will not need or use them, but for the record we briefly explain them.

The Axiom of Replacement says that if we can define, for every set a, a new set $f(a)$, then for any set A, $f(A) = \{f(a) \,:\, a \in A\}$ is a set. (As an example, we might have $f(a) = \mathcal{P}(A)$.)

The Axiom of Foundation says that there is no sequence of sets x_i such that $x_{i+1} \in x_i$ for all i. Thus every set is 'founded' on the empty set.

Suppose you think of a set A as defining a game. Player 1 must choose $x_1 \in x_0 = A$, then player 2 must choose $x_2 \in x_1$, and so on. The game ends when a player has no move, i.e. when $x_i = \emptyset$. Then Foundation insures that the game always ends after finitely many moves, no matter what set we start with.

This means that one can think of a set as a branching network of 1-way streets. Every street leads to a dead end, but the number of streets can still be huge, because infinitely many streets can leave a given intersection. (One might even consider a set as a metaphor for life, choice and mortality.)

2.2 Discussion of the Axioms

To have a well–defined domain of discourse, the elements of sets are *also sets*. The only primitive relation that can hold between sets is *membership*; that is, for any two sets A and B we can ask if $A \in B$. More elaborate concepts must be defined in terms of these. For example, the statement

$$A \subset B$$

means $\forall x(x \in A) \implies (x \in B)$. Here the quantified variable x ranges over all sets.

Axiom I (Extension). A set is determined by its elements. That is, if $x \in A \implies x \in B$ and vice-versa, for all sets x, then $A = B$.

There is a subtle point in Axiom I: what does the conclusion, $A = B$, mean anyway? In fact the idea of equality is a notion in logic rather than set theory. It means that for any logical sentence $P(x)$, $P(A)$ has the same answer as $P(B)$. For example, if $A = B$, and $A \in Y$, then $B \in Y$.

Axiom II (Specification). If A is a set then $\{x \in A \,:\, P(x)\}$ is also a set.

More precisely, this axiom asserts that given A and $P(x)$, there exists a unique set B such that $x \in B$ iff $x \in A$ and $P(x)$ is true. We will use the informal language 'is a set' in the sequel for this more precise notion.

Examples. The intersection of two sets is defined by

$$A \cap B = \{x \in A \,:\, x \in B\}.$$

Similarly, the difference is defined by

$$A - B = \{x \in A \,:\, x \notin B\}.$$

Thus $\mathbb{R} - \mathbb{Q} =$ the irrationals; while $\mathbb{Q} - \mathbb{R} = \emptyset$.

$$\{x \in \mathbb{Z} \,:\, \exists y \in \mathbb{Z}, y + y = x\} = \text{the even numbers.}$$
$$\{x \in \mathbb{Z} \,:\, x/n \in \mathbb{Z}\ \forall n > 0\} = \{0\}.$$
$$\{x \in \mathbb{Z} \,:\, x^2 < 0\} = \emptyset.$$

The empty set exists. We provisionally assume that at least one set A exists (later we will assume much more). Given that, we can now form

$$0 = \emptyset = \{x \in A \,:\, x \neq x\},$$

but nothing else for sure. (E.g. A might be $\emptyset$.)

Odd perfect numbers as a set. A number $n > 0$ is *perfect* if it is the sum of its divisors, other than n itself. Thus $6 = 1+2+3$ and $28 = 1+2+4+7+14$ are perfect. It is unknown if there is an *odd perfect number*. Nevertheless, we can form the set

$$A = \{n \in \mathbb{N} \,:\, n \text{ is an odd perfect number}\}.$$

We just cannot decide whether this set is empty or not!

The Barber of Seville; Russell's paradox. If $X = \{A : A \notin A\}$, is $X \in X$? Either answer leads to a contradiction.

In our axiom system, the set X cannot be formed, since A ranges in the 'universe' of all sets. Indeed, in our system there is no universe, i.e. there is no set U that contains all sets. In fact, this paradox *proves* there is no universe, for otherwise X *would* be a set.

One can think of the axioms in general as ways of reining in the size of sets, to keep them from getting to large that paradoxes result.

One solution to the classic paradox — who shaves the barber of Seville? — is of course that the barber is a woman. In the Gödel-Bernays theory, you are allowed to form X, but X is not a set; it is called a class.

Historical aside: Frege and Russell. In a famous episode, Bertrand Russell wrote to Frege, just as Vol. 2 of his *Grundgesetze* was about to go to press in 1903, showing that Russell's paradox could be derived from Frege's Basic Law V. The system of the Grundgesetze is thus inconsistent. Frege wrote a hasty, last-minute Appendix to Vol. 2, deriving the contradiction and proposing to eliminate it by modifying Basic Law V. Frege opened the Appendix with the exceptionally honest comment:

> *Hardly anything more unfortunate can befall a scientific writer than to have one of the foundations of his edifice shaken after the work is finished. This was the position I was placed in by a letter of Mr. Bertrand Russell, just when the printing of this volume was nearing its completion.*

(This letter and Frege's reply are translated in Jean van Heijenoort 1967.)

Axiom III (Pairs). If A and B are sets then so is $\{A, B\}$.

This is our first axiom that allows us to make sets bigger. From this axiom and $\emptyset = 0$, we can now form $\{0, 0\} = \{0\}$, which we call 1; and we can form $\{0, 1\}$, which we call 2; but we cannot yet form $\{0, 1, 2\}$.

Axiom IV (Unions). If A is a set, then

$$\bigcup A = \{x : \exists B, B \in A \,\&\, x \in B\}$$

is also a set.

From this axiom and that of pairs we can form $\bigcup\{A, B\} = A \cup B$. Thus we can define $x^+ = x+1 = x \cup \{x\}$, and form, for example, $7 = \{0, 1, 2, 3, 4, 5, 6\}$.

Intersections. If $A \neq \emptyset$, we can define $\bigcap A = \{x \,:\, \forall B \in A, x \in B\}$. Since A has at least one element B_0, we have $\bigcap A \subset B_0$ and thus the intersection is a set. Note: $\bigcap \emptyset$ is undefined!

Examples: $\bigcap\{A\} = A$, $\bigcap\{A, B\} = A \cap B$.

Axiom V (Powers). If A is a set, then

$$\mathcal{P}(A) = \{B \,:\, B \subset A\}$$

is also a set.

We can also defined $\mathcal{P}_k(A) \subset \mathcal{P}(A)$ to be the subsets $B \subset A$ with exactly k elements.

Examples: $|\mathcal{P}(52)| = 2^{52}$, while $|\mathcal{P}_5(52)| = \binom{52}{5} = 2,598,960$. The latter set can be thought of as the set of possible poker hands. Exactly 4 of these are royal flushes.

Pascal's triangle. This well–known figure is a convenient way of organizing the coefficients of $(1+x)^n$, or equivalently the values of $\binom{n}{k}$. Each new row is determined from the preceding one by adding together adjacent entries.

To explain this, note that $\mathcal{P}_k(n+1)$ can be partitioned into the subsets that contain $n+1$, and those that do not. The first type of set has $k-1$ elements from n, while the second type has k of them.

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

Axiom VI (Infinity) . There exists a set A such that $0 \in A$ and $x+1 \in A$ whenever $x \in A$.

For precision we emphasize that $0 = \emptyset$ and $x + 1 = x \cup \{x\}$.

Definition of the natural numbers. Let us call a set A as above 'inductive'. The smallest inductive set is unique, and we refer to it as the set of natural numbers:

$$\mathbb{N} = \{0, 1, 2, 3, \ldots\}.$$

Here is precise proof that the smallest inductive set is unique. First, given A as provided by the Axiom, we can define

$$\mathbb{N} = \bigcap\{B \in \mathcal{P}(A) \,:\, 0 \in B \quad \text{and} \quad x \in B \implies x + 1 \in B\}.$$

Now consider any other inductive set C. Then $B = C \cap A$ is also inductive, so $\mathbb{N} \subset C$ by the definition above. This is what is means to say $\mathbb{N}$ is the smallest inductive set.

Justifying induction. The principle of induction now follows from the *definition* of $\mathbb{N}$. Namely, we have seen that the only inductive subset of $\mathbb{N}$ is

$\mathbb{N}$ itself. On the other hand, if we know $P(0)$ and we know that $P(n) \implies P(n+1)$, then

$$A = \{n \in \mathbb{N} \,:\, P(n)\}$$

is an inductive subsets of $\mathbb{N}$, so it is equal to $\mathbb{N}$.

Other inductive sets. In Axiom VI, there are in fact other possibilities for A besides $\mathbb{N}$. For example, if we let $\omega = \mathbb{N}$ (this is the standard notation for $\mathbb{N}$ as an *ordinal*), then we can form the inductive set

$$2\omega = \{0, 1, 2, \ldots, \omega, \omega + 1, \omega + 2, \ldots\}.$$

We can similarly form $n\omega$, ω^2, ω^ω, and an infinite tower of ω's. These *countable ordinals* have a rich structure worth a separate investigation.

Negative numbers? Note that there is no set x such that $x + 1 = 0$, since $x + 1$ is always nonempty. While it is tempting to thinking of $\mathbb{Z}$ as an inductive set, this is not quite correct, since the notion of addition in $\mathbb{Z}$ is necessarily different (at least if 0 is represented by the empty set set).

Arithmetic. We can proceed to define, by induction, the usual arithmetic operations on $\mathbb{N}$. For example, we have already define $x + 1$. Having defined $x + n$, we let $x + (n + 1) = (x + n) + 1$. Similarly, we define $2x = x + x$ and $(n + 1)x = nx + x$.

Ordering numbers. We can now also define, for $i, j \in \mathbb{N}$, $i < j$ iff $i \in j$. What is $i \cap j$? $i \cup j$?

2.3 Functions and relations

We are now in a position to develop most mathematical notions. To proceed systematically, we start with the definition of $A \times B$.

Ordered pairs. The *ordered pair* of two sets a and b is defined by

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

Then $(a, b) = (a', b')$ iff $a = a'$ and $b = b'$. The *product* of two sets is defined by

$$A \times B = \{(a, b) \,:\, a \in A, b \in B\}.$$

Note that $A \times B \subset \mathcal{P}(\mathcal{P}(A \cup B))$, so it is a set.

Set theory as a programming language. The point of the definitions of $\mathbb{N}$ and (a, b) is not so much that they are natural or canonical, but that they

work. In other words set theory provides a very simple language in which the rest of mathematics can be *implemented*.

Relations and graphs. A *relation* R between A and B is a subset $R \subset A \times B$.

A relation can be visualized as a directed graph with vertices $A \cup B$ and with an edge from a to b exactly when $(a, b) \in \mathbb{R}$.

Examples: The relation $i < j$ on $\{0, 1, 2\}$ is a directed triangle. The relation $b|a$ on $\{1, 2, \ldots, 10\}$.

A relation is *reflexive* if there is a loop at each vertex; it is *symmetric* if all edges go in both directions; it is an *equivalence relation* if each component is a complete graph. This is a good way to visualize the fact that an equivalence relation is the same as a partition.

Functions. A *function* $f : A \to B$ is a relation between A and B such that for each $a \in A$, there is a unique b such that $(a, b) \in f$. We write this as $b = f(a)$. Functions are also called *maps*.

The set of all $f : A \to B$ is denoted B^A. Why? How many elements does 3^5 have? (Answer: 243.)

A function is *surjective* (or onto) if $f(A) = B$; it is *injective* (or one–to–one) if $f(a) = f(a') \implies a = a'$.

A function is *bijective* if it is both injective and surjective.

Graphs. Traditionally a function $f : [0, 1] \to [0, 1]$ was given by a formula, such as $f(x) = x^2$, and then one can draw its graph as a subset of the square. From a modern perspective, f *is the same* as its graph, and any graph defines a function. One can also picture relations directly by their 'graphs'. See Figure 2 for some examples.

Composition. We define $(f \circ g)(x) = f(g(x))$.

If $f : A \to B$ is bijective, then there is a unique map $g : B \to A$ such that $g \circ f(x) = x \, \forall x \in A$. This map is called f^{-1}. (We will later see that there is a map with the same name that sends B into $\mathcal{P}(A)$ and is defined for all f.)

Examples. Consider the map $f : A \to A$ given by $f(x) = x^2$, for $A = \mathbb{N}$, $\mathbb{Z}$, $\mathbb{R}_+$ and $\mathbb{C}$. It is injective for $\mathbb{N}$, bijective for $\mathbb{R}_+$, and surjective for $\mathbb{C}$. It is neither injective nor surjective on $\mathbb{Z}$.

The function $\sin : \mathbb{R} \to [-1, 1]$ is surjective, but not injective. Its restriction, $\sin : [-\pi/2, \pi/2] \to [0, 1]$, is bijective. Its restriction, $\sin : [0, 1] \to [-1, 1]$, is injective but not surjective.

Multivalued functions. It is sometimes useful to think of a number such as 9 as having 2 square-roots, 3 and -3. This notion is nicely capture by

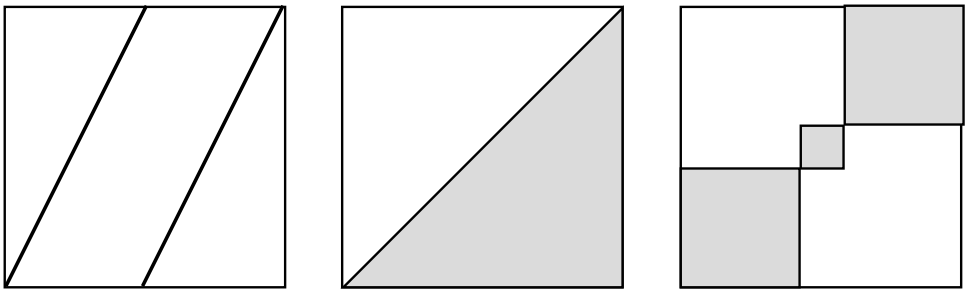

Figure 2. Relations on $[0,1] \times [0,1]$: the graph of $y = 2x \bmod 1$; $y \leq x$; and an equivalence relation.

making the square–root into a *relation*:

$$R = \{(x, y) \ : \ y^2 = x\} \subset \mathbb{R} \times \mathbb{R}.$$

One can treat a relation as a set–valued function, defined by

$$R(a) = \{b \ : \ (a, b) \in \mathbb{R}\}.$$

Then for $x \geq 0$, $R(x) = \{+\sqrt{x}, -\sqrt{x}\}$, where as usual $\sqrt{x} \geq 0$; while $R(x) = \emptyset$ for $x < 0$.

From this perspective, the square–root is like a subroutine that returns a list.

Different sets that are the same. There is a natural bijection between $A \times A$ and A^2.

There is a natural bijection between $\mathcal{P}(A)$ and 2^A.

It is common in mathematics (outside of set theory and logic) to use these (and many other) identifications without mention!

Equivalence relations. Given an equivalence relation R on a set A, we can now give a precise meaning to the set A/R where any two points $x, y \in A$ with xRy are identified. Namely, for any $x \in A$ we define its *equivalence class* by

$$[x] = \{y \in A \ : \ (x, y) \in R\} \in \mathcal{P}(A).$$

The fact that R is an equivalence relation shows that $x \in [x]$ and that for any x, y, either $[x] = [y]$ or $[x]$ and $[y]$ are disjoint.

We can now define:

$$A/R = \{[x] \,:\, x \in A\} \subset \mathcal{P}(\mathcal{P}(A)).$$

We have a natural surjective map $\pi : A \to A/R$ sending x to $[x]$, such that $\pi(x) = \pi(y)$ iff x is equivalent to y.

The collection $P = A/R$ forms a *partition* of A. Conversely, any partition P of A (into nonempty, disjoint sets) determines an equivalence relation, namely:

$$R = \bigcup \{B \times B \,:\, B \in P\}.$$

Example: $\mathbb{Z}/10$. This set consists of 10 subsets of $\mathbb{Z}$, namely $10\mathbb{Z}$, $1 + 10\mathbb{Z}$, $\ldots 9 + 10\mathbb{Z}$.

Is it true that the map $\pi : \mathbb{Z} \to \mathbb{Z}/10$ sends x and y to the same point iff they have the same last digit? Not quite: $\pi(-3) = \pi(7)$.

Constructing the integers. We have now built up the natural numbers via set theory, and can proceed to the real numbers, functions on them, etc., with everything resting on the empty set.

For example, we can define $\mathbb{Z}$ to be a quotient of $\mathbb{N}$. The idea is that every $n \in \mathbb{Z}$ can be thought of as a difference of two positive numbers, $n = b - a$. The only issue is that these two numbers are not unique; but if $n = b' - a'$, then $a + b' = a' + b$. Using this condition to define an equivalence relation on pairs (a, b), we then let:

$$\mathbb{Z} = \mathbb{N} \times \mathbb{N}/\sim .$$

Similarly, rational numbers are equivalence classes of pairs of integers, and real numbers can be described in terms of raitonal numbers by Dedekind cuts.

As an exercise, let us see how to define multiplication in $\mathbb{Z}$: we have

$$(a - b)(c - d) = ac + bd - bc - ad,$$

so we define

$$(a, b) \cdot (c, d) = (ac + bd, bc + ad).$$

One should verify that the equivalence class on the right only depends on the equivalence class on the left.

$\mathcal{P}(X)$ as an algebra. For a more complete discussion of functions, it is useful to observe that the power set comes with lots of additional structure.

First, we remark that $\mathbb{Z}/2$ has natural laws of addition and multiplication. The multiplicative laws are obvious, and for addition the main thing to keep in mind is that $1 + 1 = 0$.

Mimicking this idea, we can give $\mathcal{P}(X)$ laws of addition and multiplication. The law of multiplication is easy:

$$AB = A \cap B.$$

Note that $1 = X$ serves as the multiplicative identity; also, observe that $A^2 = A$ for all $A \in \mathcal{P}(A)$.

Addition is less obvious. It is defined using the *symmeteric difference*: that is,

$$A + B = (A \cup B) - (A \cap B).$$

In different notation, one also writes:

$$A + B = A \triangle B = (A - B) \cup (B - A).$$

Note that $0 = \emptyset$ satisfies $0 + A = A$, and that $A + A = 0$, that is, $-A = A$. Also, $X + A =$ the complement of A, i.e.

$$X + A = X - A$$

where the term $X - A$ means the difference of sets.

(*Why not set* $A + B = B \cup A$*?* Then we would have $A + A = A$; which gives $A = 0$, if we have cancellation. Our definition permits cancellation, while $A \cup B$ does not.)

It can now be checked that these operations make $\mathcal{P}(X)$ into a ring; for example, $A(B + C) = AC + AC$. In fact $\mathcal{P}(X)$ is an algebra over the field with 2 elements, $\mathbb{F}_2 = \mathbb{Z}/2$.

The source of this algebraic structure is clear if we consider the isomorphism

$$\chi : \mathcal{P}(X) \to 2^X,$$

denote the map that sends A to its *indicator function* $\chi_A : X \to \mathbb{F}_2$. These functions can be added and multiplied since they take values in a field, and we get the laws of addition and multiplication just formulated from the laws of addition and multiplication on $\mathbb{F}_2$.

Functions, unions and intersections. Now let $f : X \to Y$ be a function. We set $f(A) = \{f(a) \ : \ a \in A\}$. In this way we obtain a map $f : \mathcal{P}(X) \to \mathcal{P}(Y)$.

Is this map a ring homomorphism?

In general, $f(A \cap B) \neq f(A) \cap f(B)$. We only have $f(A \cap B) \subset f(A) \cap f(B)$. However, if f is *injective*, then equality holds. We always have, however, $f(A \cup B) = f(A) \cup f(B)$.

If $f : A \to B$ is a function, for any subset $X \subset B$ we define $f^{-1}(X) = \{a \in A : f(a) \in X\}$. Thus we have $f^{-1} : \mathcal{P}(B) \to \mathcal{P}(A)$. This map preserves intersection and unions: e.g. $f^{-1}(X \cap Y) = f^{-1}(X) \cap f^{-1}(Y)$. Thus we obtain a ring homomorphism:

$$f^{-1} : \mathcal{P}(Y) \to \mathcal{P}(X).$$

Abusing notation, one also writes $f^{-1}(b)$ for $f^{-1}(\{b\})$.

The ubiquity of f^{-1}. The fact that f^{-1} preserves set-theoretic operations means that a good theory of maps often turns on properties of f^{-1} rather than f.

For example, in topology a continuous function can be defined as one such that $f^{-1}(U)$ is open for every open set U. Similarly, in real analysis a measurable function is one such that $f^{-1}(U)$ is measurable for every open set U. In differential topology, we find that forms pullback as well, that $d(f^*\omega) = f^*(d\omega)$, etc.

Another reason for this ubiquity is that functions naturally pull back. In particular, the indicator functions satisfy

$$\chi_A \circ f = f^*(\chi_A) = \chi_{f^{-1}(A)}.$$

This pullback operation preserves the algebra structure of the space of functions. That is, it is obvious that:

$$(\chi_A \circ f) \cdot (\chi_B \circ f) = (\chi_A \cdot \chi_B) \circ f \quad \text{and} \quad (\chi_A \circ f) + (\chi_B \circ f) = (\chi_A + \chi_B) \circ f.$$

Aside: Category theory. A pervasive modern perspective in mathematics is that *morphisms* between mathematical objects are just as important as the object themselves. This perspective is made precise by *category theory*. In the case at hand, the objects are *sets*, and the morphisms are *maps*.

In particular, two objects in set theory are *isomorphic* if there is a bijection between them. This means that they are essential the *same set*, but with (possibly) different names for its elements. We now make this discussion more precise.

2.4 Cardinality and the Axiom of Choice

We say sets A and B have the same *cardinality* if there is a bijection between A and B. We will write this relation as $|A| = |B|$. It is an equivalence relation.

Finite sets. A set A is *finite* iff there is a bijection $f : A \to n$ for some $n \in \mathbb{N}$. That is, A is finite iff $|A| = |n|$ for some $n \in \mathbb{N}$.

Theorem 2.1 (Pigeon-hole principle) *If A is finite, then any injective map $f : A \to A$ is surjective.*

Proof. By induction on $|A| = n$. It suffices to treat the case where $A = n$; and the case $A = 0$ is obvious, because any map to the empty set is surjective. Suppose we know the pigeon–hole principle for $A = n$, and we wish to prove it for $A = n + 1$. Let $f : (n + 1) \to (n + 1)$ be an injective map, and let $x = f(n)$. We can the find a *permutation* σ of $n + 1$ — that is, a bijection — such that $\sigma(x) = n$. Then the map $\sigma \circ f$ sends $n \subset n + 1$ into n. Clearly $\sigma \circ f | n$ is injective, and hence $f(n+1)$ contains n by our induction hypothesis. On ther other hand, $n = \sigma(f(x))$, so f is surjective as well. ∎

Here is an 'obvious' but important result that follows.

Corollary 2.2 *Given $n, m \in \mathbb{N}$, we have $|n| = |m|$ if and only if $n = m$.*

Proof. If $n > m$ and we have a bijective map $f : n \to m$, then we can compose it with the proper inclusion $m \subset n$ to get a violation of the pigeon–hole principle. ∎

Application: inversion mod p. Here is a typical use of the pigeon–hole principle to prove a result that is not at all obvious.

Theorem 2.3 *For any prime p and $a > 0$ not divisible by p, there is an integer b such that $ab = 1 \bmod p$.*

(I.e. $b = 1/a$).

Proof. The map $b \mapsto ab$ is $1 - 1$ on $(\mathbb{Z}/p)^*$, so it is onto. (If $ab = a'b \bmod p$, then $p(a - a')b$, so $p|(a - a')$ since $\gcd(b, p) = 1$.) ∎

Example: $1/10 = 12 \bmod 17$; in fact $10 * 12 = 120 = 7 * 17 + 1$.

Infinite sets. A set is *infinite* iff it is not finite. The next result is very similar to Corollary 2.2.

Theorem 2.4 $\mathbb{N}$ *is infinite.*

Proof. Otherwise, there would for some n be an injective map $\mathbb{N} \hookrightarrow n$, and hence an injective map $n + 1 \hookrightarrow n$. This contradicts the pigeon–hole principle. ∎

The next theorem was at one time considered shocking.

Theorem 2.5 $|\mathbb{N}| \neq |\mathbb{R}|$.

Proof. Suppose $x(n)$ is a list of all real numbers, and write their fractional parts as

$$\{x_n\} = 0.x_1(n)x_2(n)\dots$$

in base 10. Now choose any sequence of digits y_i with $y_n(n) \neq x_n(n)$. We can also arrange that y_i keep changing, e.g. they are not all equal to 0 or 9 from some point on. Then

$$y = 0.y_1y_2y_3\dots$$

disagrees with x_n in its nth digit, so it is not on the list. ∎

Note that we have finessed the objection that $0.9999\dots = 1.0000$.

Theorem 2.6 (Cantor) *A and $\mathcal{P}(A)$ do not have the same cardinality.*

Proof. Given $f : A \to \mathcal{P}(A)$, let $B = \{a \,:\, a \notin f(a)\}$. Suppose $B = f(a)$. Then $a \in B$ iff $a \notin B$. This is a contradiction so f does not exist. ∎

What is a sequence? We remark that it is common and natural to think of $A^{\mathbb{N}}$ as the set of sequences $(a_0, a_1, \dots)$ with $a_i \in A$. This is just another way of presenting a the function $f : \mathbb{N} \to A$ with $f(i) = a_i$.

If we think of $\mathcal{P}(\mathbb{N}) \cong 2^{\mathbb{N}}$ as sequences (a_i) of binary digits, then the proof that $|\mathbb{N}| \neq |\mathcal{P}(\mathbb{N})|$ is almost the same as digit diagonalization.

Corollary 2.7 *There are many different sizes of infinity.*

The real numbers. We will see below that $|\mathcal{P}(\mathbb{N})| = |\mathbb{R}|$. Thus the real numbers are an example of the 'second kind' of infinity, the continuum. In any case, it is easy to see, using decimal sequences of 0's and 1s, that $|[0,1]| \geq |\mathcal{P}(\mathbb{N})|$.

The Axiom of Choice. Next we show that every infinite set contains a copy of $\mathbb{N}$. The proof uses (and requires) a new axiom.

Axiom VII (The Axiom of Choice). For any set A there is a function $c : \mathcal{P}(A) - \{\emptyset\} \to A$, such that $c(B) \in B$ for all $B \subset A$.

In concrete cases it is possible to find explicit choice functions. For the natural numbers, we can let $c(A) = \min(A)$. For the rational numbers, we can take this 'simplest' rational number $x = p/q$ in A, say with q minimal, $|p|$ minimal given q, and with $x \geq 0$ to break ties.

Theorem 2.8 *A is infinite iff there is an injective map $f : \mathbb{N} \to A$.*

Proof. If A is finite then any subset of A is finite, so there is no injection of $\mathbb{N}$ into A.

Now assume A is infinite; we will construct f. Pick some $a \in A$. Then define, by induction, $f(0) = a$ and $f(n+1) = c(A - \{f(0), \ldots, f(n)\})$. The resulting map is injective by construction. ∎

Corollary 2.9 (Cantor's definition of infinity) *A set A is infinite iff there exists a map $f : A \to A$ which is injective but not surjective.*

Proof. Use the map $f(n) = n + 1$ on a copy of $\mathbb{N}$ inside A, and set f equal to the identity elsewhere. Then f is injective but not surjective. ∎

Hilbert's hotel. The set $\mathbb{N}$ (or any infinite set) serves to illustrate *Hilbert's hotel*. The hotel is full, and yet but just shuffling the residents around we can create an empty room.

Note that there is a bijection between $\mathbb{N}$ and *any* infinite subset of $\mathbb{N}$, such that the odd numbers or the squares.

There was once a University with a long line of offices had become a little top–heavy: the professors could only occupy the offices with numbers $1, 4, 9, 16, \ldots, n^2 \ldots$ because the rest were taken up by the Deans. Outraged, the president required that each professor be assigned his own Dean, and the

rest fired. The Dean in office n was assigned to the professor in room n^2, and now the Dean's were fully employed as personal assistants to professors.

No one had to be fired. In fact the professors in offices 16, 81, 256, ... were still left without assistants, so more Deans were hired.

Other applications of AC. Every vector space has a basis. The Hahn–Banach theorem. Every set can be well–ordered. Choice of coset representatives for G/H. Existence of non-measurable sets.

The Banach-Tarski paradox. As a consequence of AC, you can cut a grapefruit into 5 pieces and reassemble them by rigid motions to form 2 grapefruits. (Now you've gone too far.)

Explicit choice for $\mathbb{Q}$ and $\mathbb{R}$. One can construct an explicit choice function for subsets of $\mathbb{Q}$, by choosing the rational number of 'least complexity'. No one has ever found a concrete choice function for subsets of $\mathbb{R}$.

Relative size. It is natural to say $|A| \geq |B|$ if there is a surjective map from A to B. But it is equally natural to require that there is an injective map from B to A. The result above is used to show these two definitions are equivalent.

Small point: if $|B| = 0$ then the surjective definition does not work.

Let us say $|A| \leq |B|$ if

(1) there is an injection $f : A \hookrightarrow B$; or
(2) there is a surjection $g : B \twoheadrightarrow A$, or $A = \emptyset$.

Theorem 2.10 *(1) and (2) are equivalent.*

Proof. Given the inclusion f we obtain from f^{-1} a surjection from $f(A)$ back to A, which we can extend to the rest of B as a constant map so long as $A \neq \emptyset$. Conversely, using the Axiom of Choice, we take f to be a section of g, i.e. set $f(a) = c(g^{-1}(\{a\}))$. ∎

Theorem 2.11 *If $|A|$ is finite and $|B| \leq |A|$ then B is finite.*

Proof 1. If B is infinite, then there exists an injective map $\mathbb{N} \to B$, so there is an injective map $\mathbb{N} \to A$, so A is infinite.

Proof 2. We can assume B is nonempty. If A is finite then we have an $n \in \mathbb{N}$ and a bijective map $f : n \to A$ and hence a surjective map $F : n \to B$ (send the points of $A - B$ to a single point of B). Then we can consider the *least* n for which such a surjection exists. For the least n, F must be bijective. ∎

Theorem 2.12 (Schröder-Bernstein) *If* $|A| \leq |B|$ *and* $|B| \leq |A|$ *then* $|A| = |B|$.

Proof. We will assume A and B are disjoint — this can always be achieved, if necessary, by replacing A, B with $A \times \{0\}, B \times \{1\}$.

Suppose we have injections $f : A \to B$ and $g : B \to A$. Then we obtain an injection

$$F = f \cup g : A \cup B \to A \cup B.$$

To clarify the proof, say $F(x)$ is the *child* of x, and x is the *parent* of $F(x)$. Since F is injective, a child can have only one parent, and every element of $A \cup B$ is a parent. However some parents are no-one's child; let us call them *godparents.*

For any $x \in A \cup B$, either x is descended from a unique godparent (possibly x itself), or x has no godparent; it has an infinite line of ancestors (or x is descended from itself.)

Now partition A into 3 pieces, A_0, A_A and A_B. A_0 is the elements $x \in A$ with no godparent; A_A consists of those x whose godparent is in A; and A_B is those whose godparent is in B. Similarly define B_0, B_A, B_B.

There is a bijection $A_0 \leftrightarrow B_0$ defined by sending a to its child $F(a)$. It is injective because F is, and it is surjective because every $x \in B_0$ has a parent, which must lie in A_0.

There is a bijection $A_A \leftrightarrow B_A$ defined by sending each $a \in A_A$ to its child $F(a)$. The inverse map sends children to their parents. There are no godparent in B_A, so the inverse is well-defined.

Similarly there is a bijection $A_B \leftrightarrow B_B$, sending $a \in A$ to its parent $F^{-1}(a)$ in B_B. Putting these three bijections together shows $|A| = |B|$. ∎

Application: $|\mathbb{Z}| = |\mathbb{N}$. Clearly $\mathbb{N} \subset \mathbb{Z}$, while we can inject $\mathbb{Z}$ into $\mathbb{N}$ by $f(x) = 10x + \text{sign}(x)$, where $\text{sign}(0) = 0$ and $\text{sign}(x) = x/|x|$ otherwise. By the Schröder–Bernstein theorem, we have a bijection between $\mathbb{Z}$ and $\mathbb{N}$.

Historical aiside: Brouwer's skepticism. It is interesting that even this simple and useful result, whose proof does not use the axiom of choice, was one controversial.

> *Brouwer did not use the Cantor–Bernstein theorem, which according to him 'must be regarded as an open problem ... but the proof of this theorem is considered inconclusive by many mathematicians', hence he defined 'as large as' by 'A can be injected*

into B and B can be injected into A'.

—Dirk van Danlen, *LEJ Brouwer*, p. 237.

Countable sets. We say A is countable if $|A| \leq |\mathbb{N}|$. Finite sets are countable.

Theorem 2.13 *If A is countable and infinite, then $|A| = |\mathbb{N}|$.*

Proof. Infinite implies $|\mathbb{N}| \leq |A|$, and countable implies $|A| \leq |\mathbb{N}|$; apply SB. ∎

Theorem 2.14 *The product $\mathbb{N}^2$ is countable; that is, $|\mathbb{N}^2| = |\mathbb{N}|$.*

Proof. Define a bijection $f : \mathbb{N}^2 \to \mathbb{N}$ by $f(n) = (a, b)$ where $2^a(2b + 1) - 1$. ∎

Alternatively: it is obvious that $|\mathbb{N}| \leq |\mathbb{N}^2|$, and there are many injections $f : \mathbb{N}^2 \to \mathbb{N}$, e.g. $f(a, b) = 10^a(10b + 1)$. The number a is the number of trailing zeros in the decimal expansion of $n = f(a, b)$. We can then apply Schröder–Bernstein to conclude $|\mathbb{N}| = |\mathbb{N}^2|$.

Corollary 2.15 *A countable union of countable sets is countable.*

Proof. We will treat the case of a countably infinite union of nonempty, countable sets; the other cases are easier. Let $A = \bigcup_{i \in \mathbb{N}} B_i$, and let $f_i : \mathbb{N} \to B_i$ be a surjection, certifying the countability of B_i. We can then If $X = \bigcup A_i$ we can send the jth element of A_i to $(i, j) \in \mathbb{N} \times \mathbb{N}$. ∎

Examples.

1. The set of things that can be described in words is countable. Thus most real numbers have no names.
2. The integers $\mathbb{Z}$ can be constructed from $2 \times \mathbb{N}$; they satisfy $|\mathbb{Z}| = |\mathbb{N}|$.
3. The rationals $\mathbb{Q}$ are $\mathbb{Q} = \mathbb{Z} \times \mathbb{Z}^* / \sim$, where $(a, b) \sim (c, d)$ if $ad - bc = 0$. The $\mathbb{Q}$ is countable.
4. The ring of polynomials $\mathbb{Z}[x]$ is countable.

Corollary 2.16 *Most real numbers are transcendental.*

Warning. While a countable sum of countable sets is countable, the same is not true for products. Indeed, $2^{\mathbb{N}}$ is a countable product of *finite* sets, but it naturally isomorphic to $\mathcal{P}(\mathbb{N})$.

Uncountable sets. Now let us examine some larger infinite sets.

Theorem 2.17 $|\mathbb{R}| = |\mathcal{P}(\mathbb{N})|$.

Proof. We can associate to each subset $A \subset \mathbb{N}$ a unique real number defined in base 10 by

$$x_A = 0.x_1x_2x_3\ldots = \sum_{n\in A} 10^{-(n+1)},$$

where $x_n = 1$ if $n \in A$ and zero otherwise. Conversely, a real number x is uniquely determined by the set

$$A_x = \{y \in \mathbb{Q} \,:\, y < x\} \subset \mathbb{Q}.$$

Thus $|\mathcal{P}(\mathbb{N})| \leq |\mathbb{R}| \leq |\mathcal{P}(\mathbb{Q})| = |\mathcal{P}(\mathbb{N})|$. ∎

Peano curves. The next result was, again, considered shocking at one time.

Theorem 2.18 *We have* $|\mathbb{R}^2| = |\mathbb{R}|$.

Proof. Using the preceding results, we have

$$|\mathbb{R}^2| = |\mathcal{P}(\mathbb{N})^2| = |\mathcal{P}(2 \times \mathbb{N})| = |\mathcal{P}(\mathbb{N})| = |\mathbb{R}|.$$

∎

In fact one can find a *continuous* surjective map

$$f : [0, 1] \to [0, 1]^2.$$

That is, a sufficiently erratic ant can walk through *every point* in San Marco square in the course of a day.

Still larger infinite sets. A third level of infinity is still within reach of our imaginations.

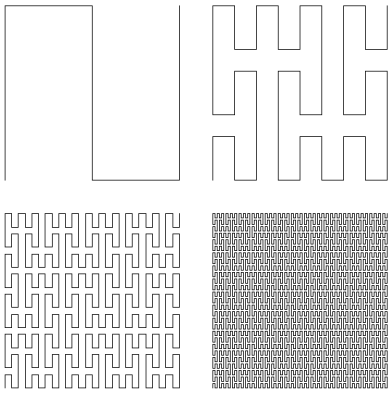

Figure 3. Approximations to a Peano curve.

Theorem 2.19 *We have* $|\mathbb{R}^{\mathbb{R}}| = |\mathcal{P}(\mathcal{P}(\mathbb{N}))|$. *Thus the functions on* $\mathbb{R}$ *represent a third kind of infinity.*

Proof. We use the fact that $|\mathbb{R}^2| = |\mathbb{R}|$ and that a function $f : \mathbb{R} \to \mathbb{R}$ can be identified with its graph. Thus we have

$$|\mathbb{R}^{\mathbb{R}}| \leq |\mathcal{P}(\mathbb{R}^2)| = |\mathcal{P}(\mathbb{R})| = |\mathcal{P}(\mathcal{P}(\mathbb{N}))| = |2^{\mathbb{R}}| \leq |\mathbb{R}^{\mathbb{R}}|.$$

By the Schröder–Bernstein theorem, equality must hold throughout. ∎

The continuum hypothesis. Using the Axiom of Choice, one can prove that for any two sets A and B, $|A| \leq |B|$ or $|B| \leq |A|$.

Is there a set A such that $|\mathbb{N}| < |A| < |\mathbb{R}|$? It is now known that this question *cannot be answered* using the axioms of set theory (assuming these axioms are themselves consistent). Some logicians have argued that CH is obviously false (Cohen, Woodin), while others have argued that it must be true (Woodin).

Finite sets, reprise. As an important exercise, to bring us back to earth, we remark that:

A finite union of finite sets is finite.

By induction, one can reduce to showing: if A and B are finite, so is $A \cup B$.

3 Group Theory

We now turn to group theory. Just as numbers are an abstract way of capturing the notion of cardinality, groups are an abstract way of capturing the notion of symmetry.

3.1 Definitions and examples.

Binary operations. Let A be a set.

A *binary operation* on A is just a map $* : A \times A \to A$, sending each *ordered* pair of elements (a, b) in A to its *product* $a * b \in A$.

An *isomorphism* between $(A, *)$ and $(A', *')$ is a *bijection* $f : A \to A'$ such that

$$f(a * b) = a *' b$$

for all $a, b \in A$. If we drop the requirement that f is a bijection, then f is called a *homomorphism*.

Example. The map $f : \mathbb{R} \to \mathbb{R}_+$ given by $f(x) = \exp(x)$ is an isomorphism between $(\mathbb{R}, +)$ and $(\mathbb{R}_+, *)$.

Define $(a, b) * (c, d) = (a + c, b + d)$ on $\mathbb{N} \times \mathbb{N}$. The map $f : \mathbb{N} \times \mathbb{N} \to \mathbb{Z}$ given by $f(a, b) = a - b$ is a homomorphism between $(\mathbb{N} \times \mathbb{N}, *)$ and $(\mathbb{Z}, +)$, but not an isomorphism.

Groups. A *group* $\langle G, * \rangle$ is a set G with a binary operation such that the following axioms hold.

1. There exists an *identity element* $e \in G$ such that $e * a = a * e = a$ for all $a \in G$.

2. For every $a \in G$ there exists an *inverse* $a' \in G$ such that $a * a' = a' * a = e$.

3. The product is associative: for all $a, b, c \in G$, we have $(a * b) * c = a * (b * c)$.

Theorem 3.1 *The identity element in G, and the inverse of a given element in G, are both unique.*

Proof. If e and e' are both identities, then $e = e * e' = e'$ by the first axiom.

If a' and a'' are two inverses for a, then by axioms (1) and (3) we can simplify $a' * a * a''$ in 2 different ways, to obtain

$$a' = a' * (a * a'') = (a' * a) * a'' = a''.$$

■

Theorem 3.2 *For any $a, b \in G$, the equation $a * x = b$ has a unique solution $x \in G$.*

Proof. Clearly $x = a' * b$ is a solution, and conversely any solution must satisfy $x = a' * a * x = a' * b$. ■

Theorem 3.3 *In any group G, we have $(a * b)' = b' * a'$.*

Proof. $(a * b) * (b' * a') = a * (b * b') * a' = a * e * e' = a * a' = e$. ■

Warning: it is not always true that $(a * b)' = a' * b'$!

Order of an element. For any element $a \in G$, we let $a^i = a * a * \cdots a$, the product with i terms, if $i > 0$; $a^0 = e$; $a^{-i} = (a^i)'$.

Then $a^i a^j = a^{i+j}$. The *order* of a is the least $n > 0$ such that $a^n = e$, or infinity if no such n exists.

Theorem 3.4 *If a has finite order n, the $a' = a^{n-1}$.*

Examples of groups and non–groups.

1. Consider $G = \mathbb{Z}$ with the following operations.
   (a) $a * b = a + b$. This is a group.
   (b) $a * b = ab$. Not a group (most elements have no inverse).
   (c) $a * b = ab/2$. Not a group ($ab/2 \notin \mathbb{Z}$). One says G is not *closed* under the group operation.
   (d) $a * b = a + b - 2$. This is a group! We have $e = 2$, $a' = 4 - a$.
   (e) $a*b = ab+1$. This is not associative. We have $(a*b)*c = abc+c+1$, while $a * (b * c) = abc + a + 1$.

2. The group $\mathbb{Z}/n = \{0, 1, 2, \ldots, n-1\}$. We define $a * b = a + b \bmod n$. That is, we form the sum and take the remainder after division by n.

   In this group, $a' = n - a$.

   The map $f : \mathbb{Z} \to \mathbb{Z}/n$, the sends a number to its remainder after division by n, is an example of a group homomorphism. It is defined so that

   $$n | x - f(x)$$

   for all $x \in \mathbb{Z}$. There is a unique value of $f(x)$ with $0 \leq f(x) < n$ satisfying this condition. To check this is a homomorphism, observe that if $n|a$ and $n|b$ then $n|a + b$. Thus

   $$n|(x + y) - (f(x) + f(y)),$$

   so $f(x + y) = f(x) + f(y)$.

   Remark: We also have $f(xy) = f(x)f(y)$. This is the basis of modular arithmetic.

3. *Casting out nines.* You can take any whole number and reduce it $\bmod n$. This gives a map $\mathbb{Z} \to \mathbb{Z}/n$ compatible with addition.

   Now notice that $1 = 10 = 100 = \ldots = 1 \bmod 9$. Thus reduction $\bmod 9$ is the same as adding up the digits. This is the famous trick of 'casting out nines' to check arithmetic.

   Example: We are all familiar with the fact that the 2-digit multiples of 9 add up 9: 18, 27, ... 81. Similarly, 522 is divisible by 9; 741 is divisible by 3, as is 174 and 147.

4. Finite examples. Consider $G = \{0, 1, 2, 3, 4\}$.

   (a) $a * b = a + b \bmod 5$. This is a group.

   (b) $a * b = ab$. This is not a group; the identity is 1, but $0 * a = 0$ for all a.

   (c) $\langle (\mathbb{Z}/5) - \{0\}, a * b = ab \rangle$. This is a group; $2 * 3 = 1$, $4 * 4 = 1$, $1 * 1 = 1$.

   (d) $\langle (\mathbb{Z}/10) - \{0\}, a * b = ab \rangle$. This is not a group; $2 * 5 = 10 = 0$ is not in G.

(e) Theorem. If we let $G = (\mathbb{Z}/n)^*$ consists of those residue classes a such that $(a, n) = 1$, then G forms a group under multiplication. In particular $(\mathbb{Z}/p) - \{0\}$ is a multiplicative group for any prime p.

5. *Continuous examples.* The groups $\mathbb{R}$, $\mathbb{R}^n$, $\mathbb{R}^*$.

6. *Multiplication of complex numbers.* Multiplication of complex numbers is based on the fact that $i^2 = -1$; it is given, for $a, b, c, d \in \mathbb{R}$, by

$$(a + ib)(c + id) = (ac - bd) + i(bc + ad).$$

The absolute value of a complex number $z = x + iy$ is defined by $|z|^2 = x^2 + y^2$. The complex conjugate of z is $\overline{z} = x - iy$. The polar coordinates of a complex number are given by

$$z = (r, \theta) = (|z|, \arg z),$$

where $\theta \in \mathbb{R}/2\pi\mathbb{Z}$.

Note that $|z|^2 = z\overline{z}$. From this it follows a key property:

$$|zw| = |z| \cdot |w|.$$

From this we find that multiplication by a complex number sends triangles to similar triangles, which in turn shows that

$$\arg(zw) = \arg(z) + \arg(w).$$

In other words, multiplication in polar coordinates is given by

$$(r, \theta) \cdot (r', \theta') = (rr', \theta + \theta').$$

Euler's famous formula states that

$$e^{i\theta} = \cos\theta + i \sin\theta.$$

In particular, $z = r \exp(i\theta)$, and $\exp(\pi i) = -1$.

The complex numbers of unit length, denoted S^1, form a group under multiplication, and we have

$$S^1 \cong \mathbb{R}/2\pi\mathbb{Z},$$

the isomorphism given by $z \mapsto \arg(z)$.

7. *Finite groups of complex numbers.* Recall that $i^2 = -1$. Thus $G = \{1, i, -1, -i\}$ forms a subgroup of $\mathbb{C}^*$ isomorphic to $\mathbb{Z}/4$.

   We also have $e^{i\theta} = \cos\theta + i\sin\theta$, the unit vector at angle θ. Since $e^{i\alpha}e^{i\beta} = e^{i(\alpha+\beta)}$, we see (a) all nth roots of unity are of the form $e^{2\pi ik/n}$; and (b) altogether these form a cyclic group $U_n \cong \mathbb{Z}/n$.

8. *Commutative groups.* A group is *commutative* if $a * b = b * a$ for all $a, b \in G$. All the groups we have seen so far are commutative. But not all groups are commutative.

   One also says such groups are *abelian*. (What's purple and commutes?)

9. *Products of groups.* If G and H are groups then so is $G \times H$, with the obvious group law, $(g_1, h_1) * (g_2, h_2) = (g_1g_2, h_1h_2)$.

10. The Klein 4-group, $V_4 \cong \mathbb{Z}/2 \times \mathbb{Z}/2$, is an example of an abelian group that is not cyclc. (In the redundant notation V_4, V stands for *vier*.)

11. *Symmetric groups.* For any set A, let $\mathrm{Sym}(A)$ be the set of all bijections $f : A \to A$, with the group law $f * g = f \circ g$. Then $\mathrm{Sym}(A)$ is a group.

12. Now consider $\mathrm{Sym}(\mathbb{R})$. Let $f(x) = x + 1$, $g(x) = 2x$. Then $f \circ g(x) = 2x + 1$, while $g \circ f(x) = 2x + 2$. Thus $f * g \neq g * f$. (On the other hand, $f * h = h * f$ if $h(x) = x + t$, and $g * h = h * g$ if $h(x) = ax$, $a \neq 0$.)

13. *The symmetric groups.* We let $S_n = \mathrm{Sym}(\{1, 2, \ldots, n\})$, $n \in \mathbb{N}$.

14. *Symmetries of a triangle.* The group S_3 is not commutative. It can be thought of as the symmetries of a triangle with vertices labeled $\{1, 2, 3\}$. It contains, besides the identity element, two rotations and three reflections. A rotation r and a reflection f do not commute: in fact, we have $frf = r^{-1}$.

    As a Corollary, S_n is non-commutative for $n \geq 3$. In fact S_3 is the smallest nonabelian group.

15. What about S_2? This group just has two elements, $\langle e, a\rangle$, and $a * a = e$. What about S_1? This is the *trivial group*, with just the identity map. What about $S_0 = \mathrm{Sym}(\emptyset)$?! This is also trivial — and nonempty!

Matrices. If you have matrices $A = (a_{ij})$ and $B = (b_{ij})$, of dimension $I \times J$ and $I' \times J'$ respectively; then if $J = I'$ you can form the product

$$(AB)_{ik} = \sum_{j=1}^{J} a_{ij} b_{jk}.$$

The result is an $I \times K$ matrix.

We always have $(AB)C = A(BC)$ when the products are all defined, because $(ABC)_{il} = \sum_{j,k} a_{ij} b_{jk} c_{kl}$.

Determinant and $\mathbf{GL_n(\mathbb{R})}$ Square matrices of rank n. The identity matrix. The powers of a diagonal matrix. Not every square matrix is invertible; those that are form the group $GL_n(\mathbb{R})$. They are characterized by $\det(A) \neq 0$.

It is a famous fact that $\det(AB) = \det(A)\det(B)$. In other words, we have a homomorphism

$$\det : \mathrm{GL}_n(\mathbb{R}) \to \mathbb{R}^*.$$

The group $\mathbf{SL_2(\mathbb{Z})}$. This is one of the most important groups in mathematics; it is essential to the study of elliptic curves and modular forms; it links analysis, geometry and number theory.

The group $\mathrm{SL}_2(\mathbb{Z})$ consists of all matrices $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ with $a, b, c, d \in \mathbb{Z}$ and $\det(A) = ad - bc$. The inverse is given by $A' = \left(\begin{smallmatrix} d & -b \\ -c & a \end{smallmatrix}\right)$.

The diagonal matrices in $\mathrm{SL}_2(\mathbb{Z})$ are $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} -1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$. The matrix $A = \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$ has order 2. The matrix $A = \left(\begin{smallmatrix} 1 & -1 \\ 1 & 0 \end{smallmatrix}\right)$ has order 6. These give finite subgroups $H \subset SL_2(\mathbb{Z})$. The matrix $A = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ has infinite order. The matrix $A = \left(\begin{smallmatrix} 2 & 1 \\ 1 & 1 \end{smallmatrix}\right)$ also has infinite order, and is related to the Fibonacci numbers. This is a hint of the relationship with number theory.

Classification of groups. Recall that $\langle G, * \rangle$ and $\langle H, \# \rangle$ are *isomorphic* if there is a bijection $f : G \to H$ such that $f(a * b) = f(a) \# f(b)$.

One of the *main problems* in group theory is to classify groups up to isomorphism. Let us address this problem for groups of small order. (The *order* of G is its number of elements, $|G|$.)

The groups $\mathbb{Z}/6$ and S_3 both have order 6, but they are *not* isomorphic. Because $\mathbb{Z}/6$ is commutative, but S_3 is not! Or, because $\mathbb{Z}/6$ has an element of order 6, but S_3 does not.

The table of a group. One can completely describe a finite group by given its *multiplication table*. That is, we write $G = (g_1, g_2, \ldots, g_n)$, with $g_1 = e$, and then make a table whose (i, j) entry is $g_i * g_j$.

A couple of useful principles. (i) The first row and column are copies of the edges of the table. (ii) Every row and every column lists a permutation of the group.

Note that commutativity corresponds to symmetry of the table. Note that e's on the diagonal tell you the elements that satisfy $a = a'$. The examples of $G = \mathbb{Z}/3$ and $G = S_3$ are shown below.

$*$	e	a	b
e	e	a	b
a	a	b	e
b	b	e	a

Table 4. Multiplication table for a group of order 3.

$*$	e	r	r^2	f	rf	r^2f
e	e	r	r^2	f	rf	r^2f
r	r	r^2	e	rf	r^2f	f
r^2	r^2	e	r	r^2f	f	rf
f	f	r^2f	rf	e	r^2	r
rf	rf	f	r^2f	r	e	r^2
r^2f	r^2f	rf	f	r^2	r	e

Table 5. Multiplication table for $S_3 = \langle r, f \ : \ r^3 = f^2 = e, frf = r^{-1} \rangle$.

Theorem 3.5 *Let G be a group with $n = |G|$.*

1. *If $n = 1$ then $G \cong \mathbb{Z}/1$, i.e. G is the trivial group.*

2. *If* $n = 2$ *then* $G \cong \mathbb{Z}/2$.

3. *If* $n = 3$ *then* $G \cong \mathbb{Z}/3$.

4. *If* $n = 4$ *then* $G \cong \mathbb{Z}/2 \times \mathbb{Z}/2$ *or* $\mathbb{Z}/4$.

Proofs. The first two statements are almost obvious. For example, when$G = \langle e, a \rangle$, the only question is, what is $a * a$? But it must be e, since a needs an inverse.

Suppose $|G| = 3$. Write $G = \langle e, a, b \rangle$, and start filling in the group table. What goes in position (a, b)? It can't be a, because there's already an a in that row, and it can't be b, because there's a b in that column! So it must be e. Similarly $b * a = e$ and we quickly see $a^3 = e$ and $G \cong \mathbb{Z}/3$.

Now suppose $|G| = 4$. Every element has order 2, 3 or 4. Suppose there is an element of order 4; then we have $\mathbb{Z}/4$. Suppose 3 is the maximal order. Then the group is $\langle e, a, a^2, b \rangle$, but what is ab? It must be a power of a, contradiction. So finally we can assume every element has order two. Then the group table is easy to complete, and we find the group is V_4. ∎

Subgroups. Let $H \subset G$ be a subset of a group $\langle G, * \rangle$. Then H is a *subgroup* of G if (a) $*(H \times H) \subset H$ and (b) $\langle H, *|H \times H \rangle$ is a group. Part (a) say H is *closed* under the product $*$.

We always have $H = \{e\}$ and $H = G$ as subgroups of G. The first is called the *trivial* group.

Theorem 3.6 *Let* H *be a subset of* G. *Then* H *is a subgroup iff*

1. $e \in H$;

2. $a, b \in H \implies a * b \in H$; *and*

3. $a \in H \implies a' \in H$.

Proof. Just check: (0) we have $*(H \times H) \subset H$; and (1) identity exists, (2) inverses exist; and (3) associativity is inherited. ∎

In brief, H is a subgroup if it is closed under multiplication and inverse, and it contains the identity element.

Examples. We have $\mathbb{Q} \subset \mathbb{R}$ as a subgroup; $\mathbb{R}^2 \subset \mathbb{R}^3$; $\{1, i, -1, -i\} \subset \mathbb{C}^*$; $\mathbb{Q}^+ \subset \mathbb{R}^*$; $\{a + b\sqrt{2} \ : \ a, b \in \mathbb{Z}\} \subset \mathbb{R}$; $\{3^n \ : \ n \in \mathbb{Z}\} \subset \mathbb{R}^*$; $\{f \ : \ f|[0,1] = 0\} \subset \mathbb{R}^{\mathbb{R}}$, under addition.

Theorem 3.7 *For any nonempty set of subgroups $H_i \subset G$, the intersection $H = \bigcap H_i$ is also a subgroup.*

Proof. Since $e \in H_i$ for all i, we have $e \in \bigcap H_i$. If $a, b \in \bigcap \mathcal{H}_i$, then $a, b \in H_i$ for every H_i, and hence $a * b \in H_i$ and $a' \in H_i$, so $a * b$ and a' also belong to $\bigcap H_i$. ∎

The lattice of subgroups. Thus the subgroups of G form a lattice, closed under intersection. Drawing this lattice is one way to start to visualize G

Example: In V_4, there are 3 nontrivial subgroups; in $\mathbb{Z}/4$, there is just one.

In S_3, there are 4 nontrivial subgroups: the cyclic group $\langle r \rangle$, and the three reflection subgroups of order 2: $\langle f \rangle$, $\langle rf \rangle$ and $\langle r^2 f \rangle$. (It is not hard to show that any reflection and rotation together generate the whole group.)

Generators of a group. For any set $S \subset G$, we can consider the collection $\mathcal{H}$ of all subgroups H with $S \subset H \subset G$. Note that $\mathcal{H}$ has at least one element, namely G itself.

We define *subgroup generated by S* to be $\langle S \rangle = \bigcap \mathcal{H}$.

The group $\langle S \rangle$ can also be built up from the inside.

Theorem 3.8 *The group generated by S consists of all elements of the form $g = \prod_1^n s_i^{n_i}$, with $s_i \in S$ and $n_i \in \mathbb{Z}$.*

In the expression above, repetitions in the list (s_i) are allowed. For an abelian group we do not need these, and we can write

$$\langle S \rangle = \left\{ \sum_S n_s \cdot s \ : \ n_s \in \mathbb{Z} \text{ and } n_s = 0 \text{ for all but finitely many } s \right\}.$$

Proof. Clearly $\langle S \rangle$ must contain these elements, and it is easy to see that the collection of elements on the right forms a subgroup of G. ∎

The Cayley graph. Given generators a_i for G, we draw a directed graph with a vertex for each $g \in G$ and an edge from g to $a_i g$, colored by a_i. If a_i has order two, the arrow is dropped.

Examples: $\langle \mathbb{Z}, 1\rangle$; $\langle \mathbb{Z}/n, 1\rangle$; $\langle V_4, a, b\rangle$; generators i, j; the star, i.e. $\mathbb{Z}/5$ with generator 2.

Examples: (S_3, f, r) vs. $(\mathbb{Z}/6, 2, 3)$. Two triangles running opposite directions in one case, the same direction in the other. Visualizing commutativity.

3.2 Cyclic groups and greatest common divisor

A group is *cyclic* if it is generated by a single element. In this section classify these groups and explore their structure.

If $G = \langle a\rangle$ then G is a cyclic group and a is a *generator* for G.

A cyclic group can have more than one generator. In $\mathbb{Z}$, the only generators are ± 1. The generators of $\mathbb{Z}/10$ are $\{1, 3, 7, 9\}$.

Classification. Recall that the order of an element $a \in G$ is the least $n \geq 1$ such that $a^n = e$. If no such n exists, we say a has *infinite order*.

Theorem 3.9 *Every cyclic group is isomorphic to $\mathbb{Z}/n$ or $\mathbb{Z}$.*

Proof. Suppose $G = \langle a\rangle$, and a has order n. Then the elements of G are simply $(e, a, a^2, \ldots, a^{n-1})$, and $a^n = e$. It is then readily verified that the map $f : \mathbb{Z}/n \to G$ given by $f(i) = a^i$ is an isomorphism.

Similarly, if a has infinite order, then the map $f : \mathbb{Z} \to G$ given by $f(i) = a^i$ is an isomorphism. ∎

Divisibility, primes etc. The additive group of the integers has a rich structure, connected with *multiplicative number theory*. To present this connection, we first recall some elementary notions related to factorization in the integers.

The first and most important fact is that given any integers p, q, with $q \neq 0$, we can write

$$\frac{p}{q} = a + \frac{r}{q},$$

where $a \in \mathbb{Z}$ and $0 \leq r < q$. The number $r = p \bmod q$ is the *remainder* of division of p by q, and the *residue* of p modulo q.

If $r = 0$ we say q *divides* p and write $q|p$. This just means that $p = aq$ for some $a \in \mathbb{Z}$. The set of all divisors $q \geq 1$ of a will be denoted by $\mathrm{Div}(a)$. It always contains 1.

GCD. Given $a, b \geq 1$, the greatest common divisor is given by

$$\gcd(a, b) = \max \mathrm{Div}(a) \cap \mathrm{Div}(b).$$

Note that the set at the right is nonempty, since it always contains 1.

Similarly, a and b have at least one common multiple — namely ab — and hence they have a *least* common multiple, denote by $\mathrm{lcm}(a, b)$.

Example: $\gcd(2, 3) = 6$, while $\mathrm{lcm}(12, 15) = 60$. We have $\gcd(3, 5) = 1$; $\gcd(10, 7) = 1$; $\gcd(21, 15) = 3$; $\gcd(84, 120) = 12$.

One way to find these numbers is to use the *prime factorization* of a and b, say $a = \prod p^{e_p}$ and $b = \prod p^{f_p}$, where the product is over all primes, and all but finitely many exponents are zero. Then

$$\gcd(a, b) = \prod p^{\min(e_p, f_p)} \quad \text{and} \quad \mathrm{lcm}(a, b) = \prod p^{\max(e_p, f_p)}.$$

For example, $12 = 2^2 \cdot 3$, while $15 = 3 \cdot 5$, so $\mathrm{lcm}(12, 15) = 2^4 \cdot 3 \cdot 5 = 60$.

The group $\mathbb{Z}$. Next we turn to the additive group $(\mathbb{Z}, +)$. We will see that the lattice of subgroups of $\mathbb{Z}$ knows about the multicative structure of $\mathbb{N}$.

Theorem 3.10 *Every subgroup $H \subset \mathbb{Z}$ is cyclic; that is, $H = a\mathbb{Z}$ for a unique $a \in \mathbb{N}$.*

Proof. This is clear when H is trivial, so suppose $H \neq \{0\}$. Let $q > 0$ be the smallest positive element of H. Given $p \in H$, write $p = aq + r$ with $0 \leq r < q$. Then $r = p - aq \in H$, but $r < q$ so $r = 0$ and thus $p = aq$. This shows $H = q\mathbb{Z}$. ∎

Theorem 3.11 *We have $a\mathbb{Z} \subset b\mathbb{Z}$ iff $b|a$.*

Proof. Immediate. ∎

Group theory of the lcm. Given that subgroups of $\mathbb{Z}$ are cyclic, it is easy to show:

Theorem 3.12 *The group $a\mathbb{Z} \cap b\mathbb{Z}$ is generated by $n \operatorname{lcm}(a, b)$.*

Proof. Let us write $G = a\mathbb{Z} \cap b\mathbb{Z} = c\mathbb{Z}$ with $c > 0$. Note that the elements of G are exactly the common divisors of a and b. Since c is the least positive element of G, it is also the least common divisor of a and b. ∎

The lattice of subgroups. The lattice of subgroups of $\mathbb{Z}$ has $\mathbb{Z}$ as the largest subgroup, then $p\mathbb{Z}$ for p prime as the next largest subgroups, and so on. The smallest subgroup is 0.

Group theory of the gcd. We can now give an alternative characterization of the greatest common divisor.

Theorem 3.13 *The subgroup of $\mathbb{Z}$ generated by $a, b \geq 0$ is given by $\langle a, b \rangle = \gcd(a, b)\mathbb{Z}$.*

Proof. Let $H = \langle a, b \rangle = c\mathbb{Z}$. Then $c|a$ and $c|b$, so c is a common divisor of a, b, and thus $c \leq \gcd(a, b)$. On the other hand, clearly $a, b \in \gcd(a, b)\mathbb{Z}$, so $c\mathbb{Z} \subset \gcd(a, b)\mathbb{Z}$, and thus $\gcd(a, b)|c$. Thus $c = \gcd(a, b)$. ∎

Corollary 3.14 *There exist $r, s \in \mathbb{Z}$ with $\gcd(r, s) = 1$ such that $ar + bs = \gcd(a, b)$.*

Proof. Since $\gcd(a, b) \in \langle a, b \rangle$, there exist r, s as above. Sincegcd(a, b) is the *least* positive element of $\langle a, b \rangle$, the numbers r and s must be relatively prime. ∎

Remark: gcd$(0, a) = a$. Consistent with the result above, it is natural to set $\operatorname{Div}(0) = \mathbb{N}$ and then $\gcd(0, a) = a$ for any $a \geq 0$.

The Euclidean algorithm. It is clear that $\langle a, b \rangle = \langle a, b - na \rangle$ for any n. Thus $\gcd(a, b) = \gcd(a, b - na)$ as well.

Thus we can recursive compute $\gcd(a, b)$ for $a > b > 0$: namely, we define $\gcd(a, 0) = a$, and otherwise:

$$\gcd(a, b) = \gcd(b, a \bmod b).$$

Examples:

$$\gcd(120, 84) = \gcd(84, 36) = \gcd(36, 12) = \gcd(12, 0) = 12.$$
$$\gcd(84, 35) = \gcd(35, 14) = \gcd(14, 7) = \gcd(7, 0) = 7.$$
$$\gcd(112, 45) = \gcd(45, 22) = \gcd(22, 1) = \gcd(1, 0) = 1.$$

The Euclidean algorithm is much faster than factoring!

Since $\operatorname{lcm}(a, b) = ab/\gcd(a, b)$, computing the GCD also allows computation of the LCM.

The golden ratio: the first irrational number. The golden rectangle has aspect ratio $x > 1$ satisfying $1 : (x - 1) = x : 1$, i.e. $x^2 - x - 1 = 0$. By geometry, the Euclidean algorithm for x never terminates, so x is irrational.

Given this infinite divisibility, it is interesting that Democritis should have advocated the atomic theory.

The finite cyclic groups.

We can now find the generators of $\mathbb{Z}/n$.

Theorem 3.15 *An element $a \in \mathbb{Z}/b$ generates $\mathbb{Z}/b$ iff $\gcd(a, b) = 1$.*

Proof. Proof. The element a generates $\mathbb{Z}/b$ iff $ar = 1 \bmod b$ for some $r \in \mathbb{Z}$, iff $ar + bs = 1$ for some $r, s \in \mathbb{Z}$, iff $\gcd(a, b) = 1$. ∎

Example: the generators of $\mathbb{Z}/9$ are $\{1, 2, 4, 5, 7, 8\}$.

Theorem 3.16 *More generally, for any $a \in \mathbb{Z}/b$ we have $\langle a \rangle = \langle \gcd(a, b) \rangle$ and the order of a in $\mathbb{Z}/b$ is $b/\gcd(b, a)$.*

Proof. Let $c = \gcd(a, b)$. Notice that, since $c|b$, the group $\langle c \rangle$ just consists of the multiples of c up to b, so it has order b/c.

Now we will show $\langle a \rangle = \langle c \rangle$. We have $ar + bs = c$ for some $r, s \in \mathbb{Z}$. Since $ar = c \bmod b$, we have $\langle c \rangle \subset \langle a \rangle$. On the other hand, c divides a, so $a = nc$ and thus $\langle a \rangle \subset \langle c \rangle$. Thus $\langle a \rangle = \langle c \rangle$. ∎

Corollary 3.17 *Every subgroup of $\mathbb{Z}/b$ is cyclic.*

Proof. Let $H \subset \mathbb{Z}/b$ be a nontrivial subgroup, with least element c. Since $\langle c \rangle$ contains $\gcd(b, c) \leq c$, we must have $c = \gcd(b, c)$ and hence $c|b$. Now for any other $x \in H$ we can write $x = nc + r$ with $0 \leq r < c$; since c was the last element of H, $r = 0$ and therefore $H = \langle c \rangle$. ∎

Corollary 3.18 *The subgroups of $\mathbb{Z}/b$ correspond bijectivitly to the divisors of b.*

Under this bijection, $b = 0 \bmod b$ gives the trivial group.

Corollary 3.19 *Every subgroup of a cyclic group is cyclic.*

Examples. In the group $G = \mathbb{Z}/18$ the possible subgroups are $\langle 0\rangle$, $\langle 1\rangle$, $\langle 2\rangle$, $\langle 3\rangle$, $\langle 6\rangle$ and $\langle 9\rangle$, ordered by divisibility.

What is the order of $a = 1024$ in $\mathbb{Z}/9999$? Every divisor of $a = 2^{10}$ is a power of 2, and 9999 is odd, so $\gcd(1024, 9999) = 1$ — and thus a generates, $\mathrm{ord}(a) = 9999$.

Relative primality. We say $a, b > 0$ are *relatively prime* if there is no prime p that divides both a and b. Putting together the preceding results, we have the following equivalent characterizations of this condition:

The numbers $a, b > 0$ are relatively prime iff:

1. No prime p that divides both a and b.
2. We have $\gcd(a, b) = 1$.
3. There exist $r, s \in \mathbb{Z}$ such that $ar + bs = 1$.
4. We have $\langle a, b\rangle = \mathbb{Z}$.
5. The element a is a generator of $\mathbb{Z}/b$.
6. There exists a matrix in $\mathrm{SL}_2(\mathbb{Z})$ with (a, b) as one of its rows or columns.

The high road. For a more pristine development of the theory of cyclic groups, it is useful to note that all the results for $\mathbb{Z}/n$ follow from the corresponding results for $\mathbb{Z}$. The reason is that we have a surjective homomorphism

$$f : \mathbb{Z} \to \mathbb{Z}/n,$$

given by $f(i) i \bmod n$. For any subgroup $H \subset \mathbb{Z}/n$, its preimage $H_0 = f^{-1}(H)$ is a subgroup of $\mathbb{Z}$ containing $n\mathbb{Z}$.

For example, to show H is cyclic, just use the fact that $H_0 = \langle a_0\rangle$ is cyclic; then $a = f(a_0)$ generates H.

To classify the subgroups of $\mathbb{Z}/n$, we just need to classifying the subgroups $H = a\mathbb{Z}$ of $\mathbb{Z}$ that contains $n\mathbb{Z}$. But these just correspond to integers $a \geq 1$ such that $a|n$, i.e. they correspond to the divisors of n

Automorphisms. An *automorphism* of a group G is a bijective homomorphism $\phi : G \to G$. The set $\mathrm{Aut}(G)$ of all automorphisms itself forms a group, under composition.

Theorem 3.20 $\mathrm{Aut}(\mathbb{Z}/n) \cong (\mathbb{Z}/n)^*$.

Proof. An automorphism ϕ is determined by $\phi(1) = k$. It has to send 1 to an element k that also generates $\mathbb{Z}/n$. If k and n are both divisible by some number $d > 1$, then all multiples of k would be divisible by d, so we'd get a proper subgroup. Thus we must have $\gcd(k, n) = 1$. Then k does generate.

What is the group law? $\phi_k(\phi_\ell(1)) = k\ell$, so the group law is indeed multiplication. ∎

Examples. In $\mathbb{Z}/100$, is multiplication by 3 an automorphism? (Yes). What is its order? (This is tricky! $3^{20} = 3486784401 = 1 \bmod 100$.)

In how many ways is $\mathbb{Z}/10$ isomorphic to $\mathbb{Z}/2 \times \mathbb{Z}/5$? Answer: there are 4 generators for $\mathbb{Z}/10$, so there are 4 isomorphisms.

3.3 Symmetric groups

In this section we study the finite symmetric groups S_n, which can be thought of as permutations of the set $A_n = \{1, 2, \ldots, n\}$.

The symmetric groups. We have seen that for any set A, the set of bijective maps $f : A \to A$ forms a group.

Note also that any bijection $\phi : A \to B$ gives rise to an isomorphism

$$\mathrm{Sym}(A) \cong \mathrm{Sym}(B),$$

sending f to the permutation $\phi \circ f \circ \phi^{-1}$.

The group of permutations of the numbers $\{1, 2, \ldots, n\}$ is denoted by S_n and referred to as the *symmetric group* on n elements. If $|A| = n$ then $\mathrm{Sym}(A) \cong S_n$.

Every finite group is a permutation group. We can now easily prove:

Theorem 3.21 (Cayley) *Every finite group G is isomorphic to a subgroup of S_n for some n.*

Proof. To each element $g \in G$ we associate the permutation $\sigma_g \in \mathrm{Sym}(G)$ with $\sigma_g(a) = ga$. To see σ_g is a permutation, note that it is invertible, in fact $\sigma_g^{-1} = \sigma_{g^{-1}}$. Also $\sigma_g(e) = g$ so the map $g \mapsto \sigma_g$ is $1-1$.

Let $H = \{\sigma_g \ : \ g \in G\} \subset \mathrm{Sym}(G)$. Then it is easy to verify that H is a subgroup of $S(G)$. Clearly $\sigma_{gh} = \sigma_g \circ \sigma_h$, so the map $g \mapsto \sigma_g$ is an isomorphism to its image.

Finally we note that $\mathrm{Sym}(G)$ itself is isomorphic to S_n where $n = |G|$. Any bijection between G and $\{1, 2, \ldots, n\}$ gives such an isomorphism. ∎

Example. The group $G = \mathbb{Z}/3$ is isomorphic to the subgroup of rotations inside $S_3 \cong S(G)$.

Remark. This theorem shows if $|G| = n$ then we can find G inside S_n, a group of size $n!$ If we allow permutation groups of infinite sets, then the same theorem works for infinite groups.

Notation for permutations. We can express any element $\sigma \in S_n$ as a $2 \times n$ matrix with the first row listing $1, 2, \ldots, n$ and the second row listing $f(1), f(2), \ldots, f(n)$. This is just like ordered pairs only you have to read vertically!

Example. Let $\alpha = \left(\begin{smallmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 3 & 4 & 5 & 1 \end{smallmatrix}\right)$, $\beta = \left(\begin{smallmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 1 & 5 & 3 & 4 \end{smallmatrix}\right)$. Then α^{-1} is obtained by writing α upside-down and reordering: $\alpha^{-1} = \left(\begin{smallmatrix} 1 & 2 & 3 & 4 & 5 \\ 5 & 1 & 2 & 3 & 4 \end{smallmatrix}\right)$. Similarly, $\beta^{-1} = \left(\begin{smallmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 1 & 4 & 5 & 3 \end{smallmatrix}\right)$. And the product is given by $\alpha\beta = \left(\begin{smallmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 2 & 1 & 4 & 5 \end{smallmatrix}\right)$.

The standard generators for S_3 are $r = \left(\begin{smallmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{smallmatrix}\right)$ and $f = \left(\begin{smallmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{smallmatrix}\right)$.

Cycles. Suppose $\sigma \in \mathrm{Sym}(A)$. We define an equivalence relation by $a \sim b$ if $\sigma^i(a) = b$ for some $i \in \mathbb{Z}$. The equivalence classes are the *orbits* of σ.

Example: for $\sigma = \left(\begin{smallmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 4 & 8 & 5 & 1 & 3 & 9 & 6 & 7 & 2 \end{smallmatrix}\right)$, the orbits are $\{1, 4\}$, $\{2, 6, 7, 8, 9\}$, $\{3, 5\}$.

A *cycle* is a permutation with at most one interesting orbit. We use the shorthand $(a_1 \ldots a_n)$ for the permutation that sends a_i to a_{i+1} and fixes everything else. The *length* of the cycle is n.

Properties of cycles. If $a = (a_1 \ldots a_n)$ then $a' = (a_n \ldots a_1)$. If $b = (b_1 \ldots b_m)$ and $\{a_i\}$ and $\{b_j\}$ are disjoint, then $ab = ba$.

Cycle notation. If A is finite, then every permutation $\sigma \in \mathrm{Sym}(A)$ is a product of disjoint cycles, $\sigma = \mu_1 \cdots \mu_\ell$. This expression for σ is unique up to permuting the cycles.

Proof. On each orbit A_i of σ, define $\mu_i(x) = \sigma(x)$ for $x \in A_i$ and $\mu_i(x) = x$ elsewhere. Then μ_i is a cycle, and clearly σ is their product. ∎

Example. The permutation σ above is $(14)(28769)(35)$. We can write this is any order and cyclically permute the elements of any cycle. So we also have $\sigma = (53)(76928)(14)$.

Note: We *leave out* the trivial cycles in this product.

Powers of permutations. For *disjoint* cycles, we have

$$(\mu_1 \cdots \mu_n)^k = \mu_1^k \cdots \mu_n^k.$$

Since a cycle of length L has order L, this shows:

Proposition 3.22 *The order in S_n of a product of disjoint cycles is the least common multiple of their lengths.*

Products of cycles. How to take the product of cycles that might not be disjoint. Examples: $(12)(23) = (123)$, $(23)(12) = (213)$. More examples: $(12345)(35)(13) = (1)(23)(45)$.

The dihedral group $D_n \subset S_n$. The symmetries D_n of a regular n-gon P_n are generated by a *rotation* r and a *flip* f, satisfying $frf = r^{-1}$ and $r^n = f^2 = e$. We can regard these elements as permutations of the n vertices P_n, which we label $1, 2, \ldots, n$. Then $r = (123 \ldots n)$, and the flip fixing 1 is given by $f = (2n)(3, n-1) \cdots$. (It has a slightly different shape depending on whether n is odd or even).

Any element of D_n can be written in the form r^i or $r^i f$, for $0 \leq i < n$. The elements with f in them all have order two! In fact $r^i f r^i f = f r^{-i} r^i f = f^2 = e$.

Similarly we can multiply any two elements and write the result in the standard form: e.g.

$$r^a f r^b f = r^a r^{-b} f^2 = r^{a-b}.$$

Theorem 3.23 *The dihedral group is generated by r and any flip $r^i f$.*

Extended example: D_4 as a subgroup of S_4. What are the symmetries of the square? Numbering the vertices counter-clockwise, we have $r = (1234)$, the counter-clockwise rotation. We also have flip on the ascending and descending diagonals, $a = (24)$ and $d = (13)$; and the vertical and horizontal flips, $v = (14)(23)$ and $h = (12)(34)$.

It is then not hard to work out the group table.

The lattice of subgroups of D_4. To practice some cycle computations, lets look at the subgroup $\langle v, h\rangle$ generated by two orthogonal flips in D_4. Their product vh should be a rotation! Indeed, $vh = (14)(23)(12)(34) = (13)(24) = r^2$. Now r^2 commutes with any flip, so $\langle v, h\rangle$ is a Klein 4-group. Similarly $\langle a, d\rangle$ is a Klein 4-group. On the other hand, $\langle r\rangle$ is isomorphic to $\mathbb{Z}/4$.

What about the group $\langle v, a\rangle$ — generated by two flips that are *not* orthogonal? This contains $av = (24)(14)(23) = (1234) = r$, and r and any flip generate D_4. In fact we have found all the subgroups of order 4.

The lattice of groups is:

$$\begin{array}{ccccccc} & & & D_4 & & & \\ & & \langle a, d\rangle & \langle r\rangle & \langle v, h\rangle & & \\ \langle a\rangle & & \langle d\rangle & \langle r^2\rangle & \langle v\rangle & & \langle h\rangle \\ & & & \langle e\rangle & & & \end{array}$$

Here r^2 forms the intersection of any pair of subgroups of order 4.

General theory of the symmetric group.

Theorem 3.24 *The transpositions generate S_n.*

Proof 1. Any cycle is a product of transpositions and any permutation is a product of cycles. Example: $(12345) = (12)(23)(34)(45)$.

Proof 2. Draw the picture of the map as a braid! ∎

Theorem 3.25 *In fact S_n is generated by $\sigma = (123\ldots n)$ and $\tau = (12)$.*

Idea of proof. By conjugating τ by σ we get all adjacent transpositions, and these suffice. ∎

Parity. Let us say an element of S_n is *even* if it can be expressed as a product of an even number of transpositions; otherwise it is odd. The parity of an element defines a homomorphism

$$\pi : S_n \to \mathbb{Z}/2.$$

But first we must show it is well–defined!

Theorem 3.26 *Every element of S_n is either even or odd but not both.*

Lemma 3.27 *Let $N(\sigma)$ be the number of orbits of σ. Then for any transposition τ, $N(\tau\sigma) = 1 + N(\sigma) \bmod 2$.*

Proof. Multiplication by τ either joins two orbits together, or breaks one orbit into two. So in either case, the number of orbits changes by one. ∎

Proof of the Theorem. If σ can be expressed as both an even and odd product of permutations, then $N(\sigma) = N(e) = N(e) + 1 \bmod 2$, which is obviously impossible. (Thus N is even iff $N(\sigma) = N(e) \bmod 2$.) ∎

The alternating group. The even permutations form the *alternating group* $A_n \subset S_n$. It satisfies $|A_n| = |S_n|/2$, provided $n \geq 2$ so there is at least one odd element.

The alternating group A_4 turns out to be the symmetry group of a tetrahedron (see below). For $n \geq 5$, the group A_n is a *simple* group. This means that for any other group G, any homomorphism $f : A_n \to G$ is either trivial or injective. The simple groups are the basic building blocks of all groups. The groups $\mathbb{Z}/p$ are also simple, but the A_n are much more interesting because they are nonabelian.

Other perspectives on parity. There are many different ways of looking at the idea of the parity of a permutation, some of which we briefly recount.

Parity and orientation. One can also view parity in terms of orientation of an $(n-1)$–simplex. The permutations which preserve orientation are even, the rest are odd. Alternatively, one can regard permutations as *matrices*, and then $\det(A_\sigma) = +1$ when σ is an even permutation, and -1 otherwise. For this, we just have to notice that transpositions have negative determinant.

Examples: The group S_2 acts on a directed segment; the subgroup A_2 preserves the arrow. The group S_3 acts on a triangle; the subgroup A_3 keeps the front face forward. The group S_4 acts on a tetrahedron; the subgroup A_4 doesn't turn the tetrahedron inside-out.

Parity and determinants. Essentially the same distinction can be drawn by associating to $g \in S_n$ a linear transformation $A(g) : \mathbb{R}^n \to \mathbb{R}^n$ that sends (x_i) to $(x_{\sigma(i)})$. Then $\det A(g) = \pm 1$, and A_n is the subgroup where the determinant is $+1$.

Parity and braids. Draw $g \in S_n$ as a braid. Then g can be written as a product of transpositions with one transposition for each crossing in the

picture! As one moves the strands around, the parity never changes. Thus A_n is a proper subgroup of S_n.

Example: D_4. The rotation $r = (1234) = (12)(23)(34)$ is *odd*, as are the flips a and d. Thus $D_4 \cap A_4 = \{e, r^2, v, h\}$. (Note that the square of any element in S_n is an even permutation.)

Why symmetric? A function $f(x_1, \ldots, x_n)$ is said to be *symmetric* if $f(x_i) = f(x_{\sigma(i)})$ for any $\sigma \in S_n$. Thus the symmetric functions are those invariant under the symmetric group. That seems to be where the terminology comes from.

Fermions. The Pauli exclusion principle states that two electrons (or more generally, fermions) cannot occupy the same state. More technically, electrons are symmetric under A_n but anti-symmetric under odd permutations. (Thus $f(x_1)f(x_2)$ can be the wave function for a pair of bosons but not for a pair of fermions; the latter might typically have instead the form $f(x_1)g(x_2) - f(x_2)g(x_1)$, where $f \neq g$.)

This principle is what forces electrons to occupy higher and higher shells around an atom, and thus gives rise to chemistry.

Sliding puzzles. The famous Sam Lloyd puzzle with small sliding squares, sometimes cannot be done!

The original puzzle consisted of 15 numbered squares in a box that holds 16. Starting with 14 and 15 reversed, Lloyd offered a $1000 reward for sliding the numbers around until they are all in order. 'Puzzle fever' reached its height around 1880, in both America and Europe. Then mathematicians proved the problem has no solution!

The reason has to do with the alternating group. Imagine the squares of the puzzle sit on a checkerboard. Each move of the puzzle is a transposition. But each move also moves the blank square from a white to a black square, vice-versa. So if we start and end with the blank on the same square, we have made an even number of moves, hence an even number of transpositions. But half the elements of S_n cannot be expressed in this way! So if the puzzle is assembled randomly, there is a 50% chance that it cannot be solved.

A particularly clear example is the 2×2 puzzle, where the pieces can only be moved cyclically.

It is remarkable that Lloyd was unable to patent his puzzle. To earn a patent, one has to submit a working model. The patent officer asked if, in fact, the model could be manipulated so the numbers come out in order. Lloyd confessed that it could not. The officer then declared that the model

did not to work, so no patent was granted.

3.4 Cosets and group actions

In this section we prove one of the simplest but most important results about finite group, *Lagrange's theorem:*

Theorem 3.28 *If $H \subset G$ is a subgroup of a finite group, then the order of H divides the order of G.*

Example: in studying the subgroups of D_4, we only needed to find subgroups of orders 8, 4, 2 and 1.

Corollary 3.29 *The order of any element of g dividies $|G|$.*

Corollary 3.30 *If $|G| = p$ is prime, then G is isomorphic to $\mathbb{Z}/p$.*

Proof. Take any $a \in G$ other than the identity; then $n = |\langle a \rangle| > 1$ must divide p, so $n = p$. ∎

In particular, $|G| = 5$ implies $G \cong \mathbb{Z}/5$. Thus we have completed the classification of groups of order ≤ 5.

Multiplicative notation for sets. Given $x \in G$ and $A \subset G$, we write

$$xA = \{xa \,:\, a \in A\}.$$

We also write $AB = \{ab \,:\, a \in A, b \in B\}$. For example, if H is a subgroup, then $HH = H$, $HG = G$. If $xH = H$ then $xh = e$ for some h and hence $x = h' \in H$.

Cosets The proof of Lagrange's theorem relies on the idea of *cosets*, which is important in its own right.

Let G be a group, and fix a subgroup $H \subset G$. A *left coset* of G is a subset of the form

$$xH \subset G.$$

Note that xH is just the image of H under the bijection given by multiplication by x. This shows:

All cosets have the same size, i.e. $|xH| = |H|$. .

Note that

$$hH = H$$

for any $h \in H$. Consequently:

The cosets form a partition *of* G.

Indeed, if xH meets yH, then $xh_1 = yh_2$ for some $h_1, h_2 \in H$; but then:

$$xH = xh_1H = yh_2H = yH.$$

We denote this partition by:

$$G/H = \{xH \ : \ x \in G\}.$$

This is also the quotient of G by the equivalence relation:

$$x \sim y \iff xH = yH \iff y^{-1}x \in H.$$

Proof of Theorem 3.28. The left cosets of G give a partition of G into $|G/H|$ subsets, each of cardinality H; thus $|H| \cdot |G/H| = |G|$. In particular, $|H|$ divides $|G|$. ∎

Examples of cosets. In $G = \mathbb{Z}/6$, there are 3 cosets of $\langle 3 \rangle$. The right and left cosets agree, as they would in any abelian group.

n $G = S_3$, let $H = \langle r \rangle$. Then the cosets are H and fH. (In fact you can tell which coset x is in by whether or not it reverses the orientation of a triangle.) Note that the right and left cosets agree.

Let $H = \langle f \rangle$. Then the left cosets are H, rH, r^2H. (The coset is determined by where the element maps the vertex fixed by f.)

Now consider the right cosets. These are different! $rH = \{r, rf\}$ while $Hr = \{r, fr\} = \{r, r^2f\}$.

These cosets are nicely pictured on the Cayley graph.

The index of a subgroup. The *index* of $H \subset G$ is given by

$$[G : H] = |G/H|.$$

For finite groups, we have $[G : H] = |G|/|H|$, but the index make sense of infinite groups as well. For example, if $G = \mathbb{Z}$ and $H = 3\mathbb{Z}$, then $[G : H] = 3$.

Group actions. An *action* of G on a set A is a map $G \times A \to A$, usually written as left multiplication, such that

$(gh)a = g(ha)$, and
$ea = a$.

To give a group action is the same as to give a homomorphism $G \to \mathrm{Sym}(A)$.

Examples. The group $\mathrm{Sym}(A)$ actions on A by $\sigma a = \sigma(a)$. Any subgroup $G \subset \mathrm{Sym}(A)$ acts on A. The dihedral group acts on the vertices of a square, and on its edges, and on its diagonals.

Transitivity. A group action is *transitive* if for any $x, y \in A$, there is a $g \in G$ such that $g(x) = y$. This g need not be unique.

Examples. The group S_n acts transitively on $\{1, \ldots, n\}$. But so does the cyclic group $\mathbb{Z}/n$ generated by $\sigma = (123 \ldots n)$.

Let A be the set of all 6 edges obtained from a square by adding in the diagonals. Then D_4 acts on A, but *not* transitively; it has two orbits.

Stabilizer. The *stabilizer* or *isotropy subgroup* of a given point $x \in A$ is just defined by

$$H_x = \{h \in G \ : \ h \cdot x = x\}.$$

Coset space. The group G acts naturally on the set G/H, by $g(xH) = gxH$. The stabilizer of the identity coset is H itself. We will see that this abstract picture models any transitive group action.

Theorem 3.31 *Let G act transitively on a set A, and let $H \subset G$ be the stabilizer of $x \in A$. Then $|A| = |G|/|H|$.*

Proof. Define a map $f : G/H \to A$ by $f(gH) = gHx$. Since $Hx = x$, this map is well–defined. By transivity, f is surjective. Finally, f is one-to-one. For if $f(gH) = f(kH)$, then $gx = kx$, so $k^{-1}gx = x$ and thus $k^{-1}g \in H$, which implies that $gH = kH$.

Thus A and G/H are in bijection, and since $|G/H| = |G|/|H|$, we have the theorem. ∎

Examples in the plane: The group D_3 acts on a triangle with the stabilizer of a vertex given by $H_v = \{e, f\}$. The number of vertices is $|D_3|/|H_v| = 6/2 = 3$. Similarly for D_4.

The group D_4 also acts on the *diagonals* of a square. The stabilizer of the ascending diagonal is $\langle a, d \rangle$; it has order 4.

The Burnside counting theorem. When G acts on A, the *orbits* of G are the sets of the form Ga, $a \in A$. They form a partition of A, which is often denote by $G\backslash A$.

Assuming G and A are finite, we have

Theorem 3.32 (The Burnside counting theorem I) *The number of orbits of G is given by*

$$|G\backslash A| = \frac{1}{|G|}\sum_{a\in A}|G_a|.$$

Proof. First, we observe that if a and b are in the same orbit of G then $|G_b| = |G_a|$. Therefore the theorem is true when G acts transitively on A; it reduces to the theorem we have just proved above. For the general case, decompose the sum over A into a sum over orbits; the action of G on each orbit is transitive, and the sum over a given orbit is 1. ∎

Let $A^g = \mathrm{Fix}(A)$ be the set of points of A fixed by $g \in G$. The sum can be re-arranged to prove:

Theorem 3.33 (The Burnside counting theorem II) *The number of orbits of G is given by*

$$|G\backslash A| = \frac{1}{|G|}\sum_{g\in G}|A^g|,$$

Proof. First, we observe that if a and b are in the same orbit of G then $|G_b| = |G_a|$. Therefore the theorem is true when G acts transitively on A; it reduces to the theorem we have just proved above. For the general case, decompose the sum over A into a sum over orbits; the action of G on each orbit is transitive, and the sum over a given orbit is 1. ∎

This is the most useful form. It can be used, for example, to show that there are 57 ways to color a cube using 3 colors (up to rotation). (Here $|A| = 3^6$ and $G \cong S_4$.)

3.5 Geometric examples of groups

The Platonic solids. A *Platonic solid* S is a polyhedron in space such that all faces, edges and vertices are equivalent. In other words, the symmetry group of S must act transitively on the vertices, edges and faces.

As a consequence, the number of vertices, faces and edges must divide the order of the symmetry group.

There are exactly 5 Platonic solids: the tetrahedron, the cube, the octahedron, the dodecahedron and the icosahedron.

The tetrahedron. The symmetry group of a tetrahedron is A_4; it can be described as the orientation-preserving permutations of the vertices.

The cube. The symmetry group of a cube has 24 elements, since there are 6 faces each with stabilizer of order 4.

In fact G is isomorphic to S_4, acting on the long diagonals! To see this, note that a rotation fixing a face gives the permutation $\sigma = (1234)$, and a rotation fixing an edge gives the permutation (12). These two elements together generate S_4.

The cube is dual to the octahedron.

Relating the tetrahedron and the cube. There are 2 ways that a tetrahedron can be inscribed in a cube. An element of S_4 is even if it preserves these two sub-tetrahedra, and odd if it interchanges them.

The dodecahedron. How large is the symmetry group of a dodecahedron? A face has stabilizer of order 5, and there are 12 faces, so $|G| = t \times 12 = 60$. Similarly there are 30 edges (since each has stabilizer 2) and 20 vertices (since 5 faces come together at each).

It turns out we have $G \cong A_5$. To see this, one can find *5 cubes* whose vertices lie on the vertices of a dodecahedron. There are 20 vertices all together, and each belongs to two cubes — which works out, since 5 cubes have $5 \cdot 8 = 40$ vertices all together.

It is important to note that not every symmetry of an inscribed cube extends to a symmetry of the dodecahedron. In fact we have $S_4 \cap A_5 = A_4$ under the embedding.

The dodecahedron is dual to the icosahedron.

A non-Platonic solid. The rhombic dodecahedron is *not* a Platonic solid. All its 12 faces are equivalent, and their stabilizer is of order 2, so $|G| = 24$. There are 14 vertices, but they are *not* all equivalent! In fact they fall into two classes of sizes $6 + 8 = 14$, and each of those divides 24.

Higher Platonic solids. (Daniel Allcock.) There are 6 4D Platonic solids, described in Coxeter's book, *Regular Polytopes*. (A $(d+1)$-dimensional solid is Platonic if its symmetries act transitively on faces of the same dimension, and each d-dimensional face is also Platonic, and its symmetries arise from those of the full polyhedron.)

They are: the simplex, the cube, the dual cube, a solid with 120 dodecahedral faces, its dual (with 600 tetrahedral faces), and a solid with 24 octahedral faces (self-dual). Their symmetry groups in the last two cases are $(2A_5 \times 2A_5)/2$ (of order $60 \cdot 120$) and a degree two extension of $(2A_4 \times 2A_4)/2$ (of order $24 \cdot 24$).

	G	$\|G\|$	V	E	F	$V-E+F$
Tetrahedron	A_4	12	4	6	4	2
Cube	S_4	24	8	12	6	2
Octahedron	S_4	24	6	12	8	2
Dodecahedron	A_5	60	20	30	12	2
Icosahedron	A_5	60	12	30	20	2
Rhombic Dodecahedron	S_4	24	6+8=14	12	12	2

Table 6. The Platonic solids and one of their cousins.

In dimensions 5 or more, there are only 3 Platonic solids: the simple, the cube and the dual to the cube.

Kepler's Cosmology. Kepler believed that the orbits of the planets were determined by the Platonic solids. Each eccentric orbit determines a thickened sphere or orb, centered at the Sun, that just encloses it. The 5 Platonic solids thus correspond exactly to the gaps between the 6 planets known at that time. Between the orbits of Saturn and Jupiter you can just fit a cube; between Jupiter and Mars, a tetrahedron; between Mars and Earth, a dodecahedron; between Earth and Venus an icosahedron, and between Venus and Mercury, an octahedron.

This theory is discussed in the *Mysterium Cosmographicum*, 1596. Using the astronomical data of Copernicus, Kepler found a reasonable agreement between his theory and observations.

	Predicted	Actual
Jupiter/Saturn	577	635
Mars/Jupiter	333	333
Earth/Mars	795	757
Venus/Earth	795	794
Mercury/Venus	577	723

Table 7. Kepler's data.

See 'Kepler's Geometrical Cosmology', J. V. Field, University of Chicago Press, 1988.

Quaternions. Hamilton made the amazing discovery that you get a reasonable algebra by adjoining not one but 3 square-roots of unity to $\mathbb{R}$! But you have to give up commutativity.

In this 'quaternion algebra', every number is of the form $a+bi+cj+dk$, where $i^2=j^2=k^2=-1$, $ij=-ji=k$, $jk=-kj=i$ and $ki=-ik=j$.

Then $G=\{\pm 1,\pm i,\pm j,\pm k\}$ gives a non-commutative group of order 8, called *the quaternion group.*

More Cayley graphs: groups of order 8. The dihedral group (D_4,f,r). Th group $\mathbb{Z}/2\times\mathbb{Z}/4$.

The quaternion group Q with generators i,j; 8 points on S^3! (Put $\pm 1,\pm i,\pm j$ on the vertices of an octahedron. Then put k in the center of the octahedron and $-k$ at infinity!)

Plane isometries. Finally we mention the important group $\mathrm{Isom}(\mathbb{E}^2)$ of isometries of the Euclidean plane. This group consists of all maps $f:\mathbb{E}^2\to\mathbb{E}^2$ such that $d(f(P),f(Q))=d(P,Q)$.

Our groups D_n — the symmetries of a polygon – can be thought of as special cases of plane isometries. Some other subgroups of $\mathrm{Isom}(\mathbb{E}^2)$: $\mathbb{Z}^2$, the symmetries of a checkerboard with a rook (written R) on each square.

Types of isometries. Other than the identity, there are four types of plane isometry. An isometry *preserves orientation* if handwriting stays the same way.

Orientation preserving, with a fixed-point P: rotation.
Orientation preserving, with no fixed-point: translations.
Orientation reversing, with a fixed-point: reflection.
Orientation reversing, with no fixed-point: glide-reflection.

3.6 Abelian groups

Going beyond the case of cyclic groups, which we know are always isomorphic to $\mathbb{Z}$ or $\mathbb{Z}/n$, in this section we describe the classification of finitely–generated abelian groups. In particular we classify all *finite* abelian groups.

Orders of elements in products. The main result will describe every finitely–generated abelian group as a product of cyclic groups. Before stating this result, we note the following easy fact:

Proposition 3.34 *Let (g, h) be an element of the product group $G \times H$. Then*

$$\operatorname{ord}(g, h) = \operatorname{lcm}(\operatorname{ord}(g), \operatorname{ord}(h)).$$

Proof. We have $(g, h)^n = (g^n, h^n) = (e, e)$ iff $\operatorname{ord}(g)|n$ and $\operatorname{ord}(h)|n$, and the order of (g, h) is the least $n \geq 1$ for which this equality holds. ∎

Corollary 3.35 *If $\gcd(a, b) = 1$, then $G = \mathbb{Z}/a \times \mathbb{Z}/b \cong \mathbb{Z}/ab$.*

Proof. Consider the element $g = (1, 1)$. Then $\operatorname{ord}(g) = \operatorname{lcm}(a, b) = ab/\gcd(a, b) = ab$, so g generates G and thus G is cyclic. ∎

Thus even if a group is a product, it may be cyclic!

Examples. We have $\mathbb{Z}/2 \times \mathbb{Z}/3 \cong \mathbb{Z}/6$. However, $G = \mathbb{Z}/2 \times \mathbb{Z}/4$ is *not* isomorphic to $\mathbb{Z}/8$; indeed every element in $\mathbb{Z}/a \times \mathbb{Z}/b$ has order *at most* $\operatorname{lcm}(a, b)$, which is less than ab if $\gcd(a, b) > 1$.

We may now state the main result.

Theorem 3.36 (Fundamental Theorem of Abelian Groups) *Let G be a finitely-generated abelian group. Then there are prime numbers p_i and exponents e_i and an integer b such that*

$$G \cong \mathbb{Z}/p_1^{e_1} \times \cdots \times \mathbb{Z}/p_n^{e_n} \times \mathbb{Z}^b.$$

This expression is unique up to the ordering of the factors. *The primes need not be distinct!*

The proof is best given after one has the notion of a factor group, so it is omitted for now. The statement by itself is nevertheless very useful and easy to apply.

Corollary 3.37 *If G is an abelian group and $|G|$ is square–free, then G is cyclic.*

Proof. If $n = |G| = p_1 \dots p_m$ is a product of distinct primes, then the only possibility for G is $\prod \mathbb{Z}/p_i$, which is cyclic because $\operatorname{lcm}(p_1, \dots, p_m) = n$. ∎

Prime factorization. We can think of the preceding theorem as saying that the groups of the form $G = \mathbb{Z}/p^e$ are 'prime': they cannot be 'factored' as $G = G_1 \times G_2$ except in trivial ways. So the fundamental theorem of abelian groups is like the prime factorization of integers.

The Classification Theorem can be proved in a similar way too: given a finite abelian group, we factor it as much as possible.

Examples.

1. If n has the prime factorization $n = p_1^{e_1} \dots p_n^{e_n}$, then

$$\mathbb{Z}/n \cong \mathbb{Z}/p_1^{e_1} \times \cdots \times \mathbb{Z}/p_n^{e_n}.$$

   Notice that $\operatorname{lcm}(p_1^{e_1}, \dots, p_n^{e_n}) = n$.

2. Any two generator group $G \subset \mathbb{R}^3$ is isomorphic to $\mathbb{Z}^n$, $n = 0$, $n = 1$ or $n = 2$. Proof: G has no torsion (other than the identity).

3. How does an abelian group like $\mathbb{Q}$ fit into this scheme? *The group $\mathbb{Q}$ is not finitely-generated!* However, it does have finitely-generated subgroups. E.g. what is the subgroup $\langle 1/3, 1/7 \rangle \subset \mathbb{Q}$? It's a *cyclic group*, namely $\langle 1/21 \rangle$. (That's because $\gcd(3, 7) = 1$, and thus $3 \times 5 - 7 \times 2 = 1$, so $5/7 - 2/3 = 1/21 \in G$.)

4. How does the trivial group fit into this scheme? It is the 'empty product', or $\mathbb{Z}^0$.

Abelian p-groups. A finite group G is a *p-group* if $|G| = p^e$ for some *prime* p some $e > 0$. By the classification theorem, every abelian p-group has the form

$$G \cong \mathbb{Z}/p^{e_1} \times \cdots \times \mathbb{Z}/p^{e_k},$$

where $\sum e_i = e$. Thus the number of abelian groups of order p^e, up to isomorphism, is given by $p(e) =$ the number of *partitions* of e. For example, the partitions of 3 are $3, 1+2$ and $1+1+1$, so there are 3 abelian groups of order 8:

$$\mathbb{Z}/8, \quad \mathbb{Z}/2 \times \mathbb{Z}/4, \quad \text{and} \quad \mathbb{Z}/2 \times \mathbb{Z}/2 \times \mathbb{Z}/2.$$

Finding all abelian groups of a given order n. In general, the isomorphism classes of abelian groups of order n correspond to the ways of factoring n into a product of powers of primes.

The only abelian group of order 6 is $\mathbb{Z}/2 \times \mathbb{Z}/3$. For order 12, there are 2 groups, 4×3 and $2 \times 2 \times 3$. For order 360, there are 6 groups; $360 = 2^3 \times 3^2 \times 5$, and the factorizations are

$$\begin{array}{ll} 2 \times 2 \times 2 \times 9 \times 5, & 2 \times 2 \times 2 \times 3 \times 3 \times 5 \\ 2 \times 4 \times 9 \times 5, & 2 \times 4 \times 3 \times 3 \times 5 \\ 8 \times 9 \times 5, & 8 \times 3 \times 3 \times 5. \end{array}$$

In general, to find the abelian groups of G order n, one first writes n as a product of powers of *distinct* primes, $n = p_1^{e_1} \cdots p_k^{e_k}$. Then, any G of order n can be written as

$$G \cong G_1 \times \cdots \times G_k,$$

where G_i is a group of order $p_i^{e_i}$. In other words, each factor is a p–group (for some prime p). Then the number of possibilities for G can be expressed in terms of the partition function: it is $p(e_1) \cdots p(e_k)$.

When are 2 finite abelian groups isomorphic? The factorization theorem gives a canonical form for any finite abelian group. Moreover, any cyclic group can be put into this form by replacing $\mathbb{Z}/n$ with $\prod_p \mathbb{Z}/p^{e_p}$. Using this fact, it is easy to check if two products of cyclic groups are isomorphic. The procedure is to replace each factor with a product of p–groups, and then see if one gets the same result.

Example. Consider the groups of order 24 given by:

$$\begin{array}{rcl} G_1 & = & \mathbb{Z}/2 \times \mathbb{Z}/12, \\ G_2 & = & \mathbb{Z}/4 \times \mathbb{Z}/6, \quad \text{and} \\ G_3 & = & \mathbb{Z}/24. \end{array}$$

Notice that 12, 6 and 24 are *not* powers of primes, so we need to factor these cyclic groups further to put the groups into canonical form. For this first 2 groups we obtain the factorization $24 = 2 \cdot 4 \cdot 6$ and $24 = 4 \cdot 2 \cdot 3$, so $G_1 \cong G_2$; while for the last group, we obtain $24 = 8 \cdot 3$, so G_3 is *not* isomorphic to either of the first two groups.

Decomposable groups. Let us say a group is *decomposable* if it is isomorphic to a nontrivial product $A \times B$; otherwise it is indecomposable.

Theorem 3.38 *A finite abelian group G is indecomposable iff $G \cong \mathbb{Z}/p^e$ for some prime p and $e \geq 0$.*

Proof. If G is indecomposable then in the classification, only one term can occur, so $G \cong \mathbb{Z}/p^e$. Conversely, if $G = \mathbb{Z}/p^e$ and $G = A \times B$, then $|A| = p^a$ and $|B| = p^b$, where $a + b = e$, and thus the order of every element in G is at most $\text{lcm}(p^a, p^b) = p^{\max(a,b)}$. But G has an element of order p^e, so $a = 1$ or $b = 1$ and hence the product is trivial. ∎

3.7 Homomorphisms and factor groups

It is often said that in mathematics, the maps between objects are as important (or more important) than the objects themselves. In this section we will study maps between groups in more detail.

Group homomorphisms. Let $\phi : G \to H$ be a map between groups. We say ϕ is a group *homomorphism* if

$$\phi(ab) = \phi(a)\phi(b)$$

for all $a, b \in G$. The following important properties of homomorphisms are easily verified:

$\phi(e) = e$;
$\phi(a') = \phi(a)'$; and
the image $\phi(G)$ is a subgroup of H.

Examples of homomorphisms.

1. Let $\phi : G \to H$ be defined by $\phi(a) = e$ for all $a \in G$. This is the *trivial* homomorphism.

2. Let $\phi : \mathbb{Z} \to \mathbb{Z}/n$ be reduction mod n.

3. Define $\phi : D_n \to \mathbb{Z}/2$ by $\phi(r^i) = 0$ and $\phi(fr) = 1$.

4. Let $\phi : S_n \to \mathbb{Z}/2$ by parity.

5. For any element $a \in G$, we obtain a homomorphism $\phi : \mathbb{Z} \to G$ by $\phi(n) = a^n$.

   If a has order n, we also get a homomorphism $\phi : \mathbb{Z}/n \to G$. This map is $1 - 1$.

6. Let $\phi : GL_n(\mathbb{R}) \to \mathbb{R}^*$ by $\phi(A) = \det(A)$.

7. For any real numbers a, b, let $\phi : \mathbb{R}^2 \to \mathbb{R}$ be given by $\phi(x, y) = ax + by$.

8. Let $\phi : \mathbb{R} \to \mathbb{C}^*$ be given by $\phi(\theta) = \exp(i\theta)$.

9. Let $\phi : \mathbb{R}^* \to \mathbb{R}^*$ be given by $\phi(x) = |x|$.

10. Let $\phi : \mathbb{R}^* \to \mathbb{R}^*$ be given by $\phi(x) = x^2$.

11. Let $\phi : C^\infty[0, 1] \to \mathbb{R}$ be given by $\phi(f) = \int_0^1 f(x)\, dx$.

12. Let $\phi : C^\infty[0, 1] \to C^\infty[0, 1]$ be given by $\phi(f) = f'(x)$.

13. Let $\phi : S_4 \to \mathrm{Sym}(\{x, y, z\}) \cong S_3$ map the symmetries of the cube into the space of permutations of the coordinate axes.

    (Equivalently, ϕ gives the action of S_4 on pairs of opposite faces of a cube, or on pairs of elements of $\{1, 2, 3, 4\}$.)

Kernels and normal subgroups. A subgroup $K \subset G$ is *normal*, usually written $K \lhd G$, if $gK = Kg$ for all $g \in G$. Equivalently, K is normal iff

$$gKg^{-1} = K$$

for all $g \in G$. It suffices to show $gKg^{-1} \subset K$ for all g, since this implies the reverse inclusion:

$$K = g(g^{-1}Kg^{-1})g \subset gKg^{-1}.$$

Examples. (1) $K = \langle r \rangle$ is normal in $G = S_3$, but $H = \langle f \rangle$ is not, since rHr^{-1} contains $rfr^{-1} = r^2 f \notin H$. (2) Any subgroup of an abelian group is normal.

The *kernel* of a homomorphism $\phi : G \to H$ is defined by

$$\mathrm{Ker}(\phi) = \{g \in G \ : \ \phi(g) = e\}.$$

The importance of normal subgroups comes from:

Theorem 3.39 *The kernel of ϕ is a normal subgroup of G.*

Proof. Recall that to show $K = \mathrm{Ker}(\phi)$ is a subgroup, we must show it contains the identity and is closed under inversion and multiplication. All three properties follow from the basic facts about homomorphisms. As for normality, just note that if $a \in K$ then

$$\phi(gag') = \phi(g)\phi(a)\phi(g)' = \phi(g)\phi(g)' = e,$$

so $gKg^{-1} \subset K$. ∎

Examples: the kernels in our various examples of homomorphisms are:

1. For the trivial homomorphism $\phi : G \to H$, we have $\mathrm{Ker}(\phi) = G$.
2. For reduction mod n, we have $\mathrm{Ker}(\phi) = n\mathbb{Z} \subset \mathbb{Z}$.
3. For the parity map on D_n, we have $\mathrm{Ker}(\phi) = \langle r \rangle$.
4. For the parity map on S_n, we have $\mathrm{Ker}(\phi) = A_n$.
5. The map $n \mapsto a^n$ has kernel $k\mathbb{Z} \subset \mathbb{Z}$, where $k = \mathrm{ord}(a)$ if this is finite, and otherwise $k = 0$.
6. The kernel of the determinant map is $\mathrm{SL}_n(\mathbb{R})$.
7. The line $\langle (bt, -at) \ : \ t \in \mathbb{R} \rangle$ is the kernel of the map $(x, y) \mapsto ax + by$.
8. The kernel of the exponential map is $2\pi\mathbb{Z} \subset \mathbb{R}$.
9. The kernel of $x \mapsto |x|$ is (± 1).
10. The kernel of $x \mapsto x^2$ is also (± 1). (This shows different maps can have the same kernel.)
11. The functions of mean zero are the kernel of the integration map.
12. The constant functions are the kernel of the derivative map.
13. The Klein four subgroup, generated by 180 degree rotations about the coordinate axes, is the kernel of the map $S_4 \to S_3$.

Solving equations. The kernel often intervenes when the solution to an equation is not unique.

Theorem 3.40 *Let $\phi : G \to H$ be a homomorphism, and suppose $\phi(x_0) = y$. Then the set of all solutions to $\phi(x) = y$ in G is given by the coset*

$$S = x_0 \, \mathrm{Ker}(\phi).$$

Examples.

1. The kernel of the squaring map on $\mathbb{R}^*$ is ± 1. Thus, if $x_0^2 = y$, the set of all solutions to $x^2 = y$ is given by $\pm x_0$.

The plus or minus sign in square–roots comes from the kernel of the squaring map.

2. Define $\phi : C^\infty[0,1] \to C^\infty[0,1]$ by $\phi(f) = f'$. Then $\mathrm{Ker}(\phi)$ consists of the constant functions, $f(x) = C$.

   Now let us try to find all solutions to the equation $\phi(f) = f'(x) = x$. One solution is given by $f_0(x) = x^2/2$. Thus all solutions are given by $f(x) = x^2/2 + C$.

   The ubiquitous $+C$ in indefinite integrals comes from the kernel of the derivative map.

Theorem 3.41 *A homomorphism ϕ is injective iff* $\mathrm{Ker}(\phi) = \{e\}$.

Corollary 3.42 *If $\phi : G \to H$ has trivial kernel, then G is isomorphic to a subgroup of H.*

Quotient groups. We have seen that $\mathrm{Ker}(\phi)$ is a normal subgroup. Now given a normal subgroup $K \subset G$, can we somehow find a homomorphism $\phi : G \to H$ such that $\mathrm{Ker}(\phi) = K$? The answer is yes!

To see this, note that when K is normal, *the product of two cosets* of K is another coset. Namely,

$$(aK)(bK) = a(Kb)K = a(bK)K = abK.$$

Thus we can define a binary operation on G/K by $(aK) * (bK) = (aK)(bK)$, and we have:

Theorem 3.43 *If K is normal, then $(G/K, *)$ is a group.*

Proof. (1) The identity element is K; (2) we have $(aK)' = a'K$; and associative follows from:

$$((aK)(bK))(cK) = abcK = (aK)((bK)(cK)).$$

■

Note that we also have a natural *quotient homomorphism*

$$\pi : G \to G/K,$$

given by $\phi(a) = (aK)$, and that

$$\mathrm{Ker}(\pi) = K.$$

Thus we have explicitly realized every normal subgroup as a kernel.

Theorem 3.44 *Let $\phi : G \to H$ be a homomorphism. Then $\phi(G)$ is naturally isomorphic to G/K.*

Proof. We can reduce to the case where $\phi(G) = H$. The map $\psi : G/K \to H$ given by $\psi(aK) = \phi(a)$ is well–defined, surjective and $1-1$, so it is an isomorphism. ∎

Example: The integers mod n. Since the integers form an abelian group, $K = n\mathbb{Z}$ is a normal subgroup of $\mathbb{Z}$. The quotient

$$G/K = \mathbb{Z}/n\mathbb{Z}$$

is natural isomorphic to $\mathbb{Z}/n$. (In fact this description of $\mathbb{Z}/n$ motivates the notation.) This may change the way you think about $\mathbb{Z}/n$! For example, in $\mathbb{Z}/10$, the number 3 is really the coset $3 + 10\mathbb{Z} = \{\ldots, -7, 3, 13, \ldots\}$.

Lagrange's theorem for quotients. Since $|G/K| = |G|/|K|$, the preceding discussion yields:

Corollary 3.45 *If $\phi : G \to H$ is a homomorphism, then $|\phi(H)|$ divides both $|H|$ and $|G|$.*

Thus Lagrange's theorem holds for both *subgroups* and *quotient groups.*

Examples.

1. If $\phi : \mathbb{Z}/22 \to S_{10}$ is a homomorphism, then its image must have order 1 or 2 (11 does not divide $10! = |S_{10}|$).

2. There is no nontrivial homomorphism $\phi : \mathbb{Z}/12 \to \mathbb{Z}/5$. In fact:

3. There is a nontrivial map $\phi : \mathbb{Z}/a \to \mathbb{Z}/b$ iff $\gcd(a, b) > 1$.

   (Proof. Suppose ϕ is nontrivial, and let $G = \phi(\mathbb{Z}/a)$, $d = |\phi(\mathbb{Z}/a)| > 1$. Then $d|a$ since G is a *quotient group* of $\mathbb{Z}/a$, and $d|b$ since G is a *subgroup* of $\mathbb{Z}/b$. Thus $1 < d \leq gcd(a, b)$.

   Conversely, if $d = \gcd(a, b) > 1$, then we can then we can define $\phi(x) = (b/d)x$. If x is changed by a multiple of a, then $\phi(x)$ is changed by a multiple of $(b/d)a = b(a/d) = 0 \bmod b$, so ϕ is well-defined; and $\phi(1) = b/d \neq 0 \bmod b$, so ϕ is nontrivial.)

Exercise. (Fraleigh, 3.2(21)). If G is abelian then G/H is abelian. What is wrong with a proof that starts "let a and b be elements of G/H"?

Simple groups. A group G is *simple* if its only normal subgroups are G and $\{e\}$. The group $\mathbb{Z}/p$ is simple whenever p is prime; $\mathbb{Z}/n$ is *not* simple if n is composite. The smallest nonabelian simple group is A_5.

A famous puzzle, appropriate for anniversary conferences: what is the next term in this sequence:

$$2, 3, 5, 7, 11, 13, 17, 19, 21, 23, 29, 31, 37, 39, 41, 43, 47, 53, 59?$$

Answer: 60. These are the orders of the finite simple groups!

3.8 Generators and relations

In this section we give a glimpse of *combinatorial group theory* by explaining how to define a group by generators and relations.

This perspective is important in topology and fits well with the Cayley graph. For example, we will later use generators and relations to associate a group to every knot; this is an example of *algebraic topology*.

Introduction to group presentations. One way to specify a group G is to list a finite set of *generators* $g_1, \ldots, g_n$ for G, together with a finite set of *relations* $r_1 = s_1, \ldots, r_m = s_m$ that we would like to hold between these generators. The notation for such a finitely–generated group is:

$$G = \langle g_1, \ldots, g_n \ : \ r_1 = s_1, \ldots, r_m = s_m \rangle.$$

The *relations* are between elements of the group specified as products of generators and their inverses. The group G is defined as the *largest possible group* where the relations hold.

It is customary to use the shorthand $\overline{g}_i$ for g_i^{-1}.

Examples.

1. Consider the group

$$G = \langle a, b \ : \ a^5 = b^2 = (ab)^2 = e \rangle.$$

This specifies a group with 3 properties: (i) G is generated by a, b; (ii) the elements b and ab in G have order 2, while a has order 5; and (iii) G is otherwise as large as possible.

To make (iii) precise we require that for *any other* group H, with elements $\alpha, \beta \in H$ such that

$$\alpha^2 = \beta^5 = (\alpha\beta)^2 = e,$$

there is a (unique) homomorphism $\phi : G \to H$ such that $\phi(a) = \alpha$ and $\phi(b) = \beta$.

Using (iii) we can show that $G \cong D_5$. First we note that, since $ba = a^{-1}b$, every element of G can be written in the form $a^i b^j$. Then, since a and b have orders 5 and 2 respectively, we know $|G| \leq 10$. On the other hand, by (iii) there is a homomorphism $\phi : G \to D_5$ defined by $\phi(a) = r$ and $\phi(b) = f$. Since r and f generate D_5, this map is surjective. Therefore $|G| = 10$ and ϕ is a bijection; hence an isomorphism.

2. Consider the group

$$G = \langle a, b \ : \ ab = ba \rangle.$$

Because of the relation above, G is abelian; every element can be written in the form $g = a^i b^j$. The relation $\alpha\beta = \beta\alpha$ also holds for $\alpha = (1, 0)$, $\beta = (0, 1)$ in $\mathbb{Z}^2$. Thus we have a homomorphism

$$\phi : G \to \mathbb{Z}^2$$

defined by $\phi(a) = \alpha$ and $\phi(b) = \beta$. Clearly $\phi(a^i b^j) = (i, j)$, so the only element of G sent to zero is the identity; thus ϕ is an isomorphism and we have $G \cong \mathbb{Z}^2$.

3. Consider the group

$$G = \langle a \ : \ a^5 = e \rangle.$$

Arguing as above, we conclude that $G \cong \mathbb{Z}/5$.

4. Let us return to our first example, the dihedral group. Write

$$D_{2n} = \langle r, f \ : \ r^n = f^2 = (rf)^2 = e \rangle.$$

Note that the elements $\alpha = 0$ and $\beta = 1$ in $\mathbb{Z}/2$ satisfy $\alpha^n = \beta^2 = (\alpha\beta)^2 = e$. Thus there is a unique homomorphism

$$\pi : D_{2n} \to \mathbb{Z}/2$$

given by $\pi(r) = 0$ and $\pi(f) = 1$. It records whether or not a symmetry of the regular n-gon reverses its orientation.

Free groups. So far we have implicitly assumed that a largest group with given relations exists. We now turn to the proof, first in the case where there are *no relations*. That is, we will construct the *free group*

$$F_n = \langle g_1, \ldots, g_n \rangle.$$

The number n is called the *rank* of the free group.

It is characterized by two properties: (i) the elements $(g_1, \ldots, g_n)$ generate F_n; and (ii) for any group G and any $\gamma_1, \ldots, \gamma_n \in G$, there is a (unique) homomorphism

$$\phi : F_n \to G$$

such that $\phi(g_i) = \gamma_i$.

Example: the free group on one generator. The elements of $F_1 \cong \langle a \rangle$ are given by a^n, $n \in \mathbb{Z}$. Thus $F_1 \cong \mathbb{Z}$ with $a = 1$. Given any group G and $\alpha \in G$, we have a unique map $\phi : F_1 \to G$ sending a to α, namely $\phi(a^n) = \alpha^n$.

The free group on two generators. The first interesting example is $F_2 \cong \langle a, b \rangle$.

Given a finite *alphabet* $\mathcal{A}$, the set of *words* $W(\mathcal{A})$ consists of all finite sequences of the form $w = a_1 \cdots a_n$, with $a_i \in \mathcal{A}$. There is a unique word of length zero which we denote by e. Words are multiplied by concatenation:

$$ww' = a_1 \cdots a_n a'_1 \cdots a'_{n'}.$$

To describe $F_2 = \langle a, b \rangle$ we use the alphabet $\mathcal{A} = \langle a, b, \overline{a}, \overline{b} \rangle$. Here $\overline{a}$ and $\overline{b}$ play the roles of the inverse of a and b. We have a map $g \mapsto \overline{g}$ on $\mathcal{A}$ that interchanges a with $\overline{a}$ and b with $\overline{b}$.

Reduction. Whenever a product of the form $g\overline{g}$ occurs in a word $w \in W(\mathcal{A})$, we can 'cancel' these two letters (eliminate them) to achieve a shorter word. Doing this repeatedly until no such products appear yields a unique *reduced word* $\mathrm{red}(w)$. It has the property that

$$\mathrm{red}(w_1 g\overline{g} w_2) = \mathrm{red}(w_1 w_2)$$

for any $g \in \mathcal{A}$. For example, we have

$$\mathrm{red}(ab\overline{a}\overline{b}ab) = \mathrm{red}(ab\overline{a}ab) = \mathrm{red}(abb) = abb = ab^2.$$

Multiplication in F_2 is defined by $w * v = \mathrm{red}(wv)$.

Theorem 3.46 *With this binary operation, F_2 is a group.*

Proof. (i) The empty word e serves as the identity; (ii) the inverse of $w = g_1 \cdots g_n$ is $\overline{g}_n \cdots \overline{g}_1$; and (iii) associativity follows from properties of reduction that can be easily proved by induction. ∎

The Cayley graphs of F_2 and $\mathbb{Z}^2$; the tree and the checkerboard. It is instructive now to draw the Cayley graph of F_2; it is an infinite tree with degree 4 at every vertex. It is also useful to compare the Cayley graphs of the free group $\langle a, b\rangle$ and of the free abelian group $\langle a, b \ : \ ab = ba\rangle$. The relation gives loops in the second graph.

The universal property for free groups. To show F_2 is as large as possible, we need to show that for any group G and any $\alpha, \beta \in G$, there is a unique homomorphism $\phi : F_2 \to G$ satisfying $\phi(a) = \alpha$ and $\phi(b) = \beta$. To this end, we first define $\Phi : \mathcal{A} \to G$ by sending $(a, \overline{a}, b, \overline{b})$ to $(\alpha, \alpha^{-1}, \beta, \beta^{-1})$. Then we extend Φ to arbitrary words so that

$$\Phi(g_1 \cdots g_n) = \Phi(g_1) \cdots \Phi(g_n),$$

where the product on the right takes places in G. Then clearly

$$\Phi(wv) = \Phi(w)\Phi(v).$$

Finally we observe that

$$\Phi(\mathrm{red}(w)) = \Phi(w).$$

Thus, if we define $\phi : F_2 \to G$ by simplying restrict Φ to the set of reduced words, we have:

$$\phi(w * v) = \Phi(\text{red}(wv)) = \Phi(wv) = \Phi(w)\Phi(v) = \phi(w)\phi(v).$$

Thus ϕ is a homomorphism.

Example. For every n, we have a surjective map $\phi : F_2 \to S_n$ given by $\phi(a) = (12)$ and $\phi(b) = (12\cdots n)$. Surjectivity holds because these two cycles generator S_n.

This map is certainly not injective, e.g. $\phi(a^2) = e$.

Construction of a group from a presentation. We can now define the finitely–presented group

$$G = \langle g_1, \ldots, g_n \ : \ r_1 = s_1, \ldots, r_m = s_m \rangle.$$

First, we observe that the relation $r = s$ is the same as the relation $rs^{-1} = e$. Thus it suffices to treat the case where all the $s_i = e$. To construct G, we take the free group F_n on $g_1, \ldots, g_n$ — using an alphabet with $2n$ letters this time — and proceed to 'kill' the words $r_1, \ldots, r_m$. More precisely, we let N be the *smallest normal subgroup* of F_n containing $r_1, \ldots, r_m$, and define

$$G = F_n/N.$$

(To verify that N exists, use the fact that the intersection of any collection of normal subgroups is normal.)

The universal property, with relations. The main property of a finitely–presented group

$$G = \langle g_1, \ldots, g_n \ : \ r_1 = \cdots = r_n = e \rangle$$

is the following.

Theorem 3.47 *For any group H, the homomorphisms $\phi : G \to H$ are in bijection with the homomorphisms*

$$\psi : F_n \cong \langle g_1, \ldots, g_n \rangle \to H$$

such that $\psi(r_i) = e$ for each relation r_i.

Proof. The condition on relations says that $\text{Ker}(\psi)$ contains R, which is exactly what is needed for ψ to factor through $G = F_n/R$. ∎

In particular, if H is any other group with the same generators where the same relations hold, we get a surjective homomorphism $G \to H$, showing that G is indeed the largest group with the given relations.

Example: The infinite dihedral group. Let $D_\infty = \langle f, r \,:\, rf = fr^{-1} \rangle$. We can think of D_∞ as acting on $\mathbb{R}$ as the boundary of an 'infinite polygon', with vertices at the integers, by $f(x) = -x$, $r(x) = x + 1$. Then $frf(x) = -((-x) + 1) = x - 1 = r^{-1}(x)$ as required.

Now let $G = \langle a, b \,:\, a^2 = b^2 = e \rangle = \mathbb{Z}/2 * \mathbb{Z}/2$. It is easy to draw the Cayley graph of G; it's a straight line, just like the boundary of an infinite polygon.

Theorem 3.48 *D_∞ and G are isomorphic.*

Proof. Define a map $\phi : G \to D_\infty$ by $\phi(a) = f$, $\phi(b) = rf$. Then clearly $\phi(a^2) = e$ and $\phi(b^2) = rfrf = rr^{-1} = e$, so ϕ is a homomorphism.

Now define a map $\psi : D_\infty \to G$ by $\psi(f) = a$ and $\psi(r) = ba$. Then $\psi(f^2) = a^2 = e$ and

$$\psi(fr^{-1}) = a(ba)' = aa'b' = b' = b = (ba)a = \psi(rf),$$

so ψ is a homomorphism. We then compute $\psi \circ \phi(a) = a$,

$$\psi \circ \phi(b) = \psi(rf) = baa = b,$$

so $\psi \circ \phi$ is the identity. Similarly $\phi \circ \psi$ is the identity, so these two groups are isomorphic. ∎

Generators and relations for S_n.

Theorem 3.49 *The symmetric group S_n has generators $\tau_i = (i, i+1)$, $i = 1, \ldots n-1$, with relations*

$\tau_i^2 = e$;
$\tau_i \tau_{i+1} \tau_i = \tau_{i+1} \tau_i \tau_{i+1}$; *and*
$\tau_i \tau_j = \tau_j \tau_i$ *if* $|i - j| > 1$.

Sketch of the proof. To check the main relation, let $(i, i+1, i+2) = (i, j, k)$; then we have: $(ij)(jk)(ij) = (ik) = (jk)(ij)(jk)$. So there is a map of the group above to S_n, and since adjacent permutations generate, it is *onto*.

Now by the picture of changing crossings, it is clear that any two diagrams of the same permutation differ by these relations. ∎

Corollary 3.50 *The parity of an element in S_n is well-defined.*

Proof. The relations preserve parity. Alternatively, define a map from S_n to $\mathbb{Z}/2$ by sending each τ_i to one, and observe that the relations are satisfied. ∎

Trivial groups. It is not always easy to tell whether or not a presentation is just giving the trivial group. For example, $\langle a \,:\, a^{12} = e, a^{25} = e\rangle$ is trivial. For a more interesting example, consider

$$G = \langle a, b \,:\, b^2 a = b, ba^2 b = a\rangle.$$

From the first relation we get $b = a^{-1}$, so the second relation gives $e = a$ and hence G is trivial.

Remarkably: *there is no algorithm to check if a given finitely–presented group is trivial.*

Classification of groups of order 10. We can use the idea of relations to classify groups. Here is an example, in the case of groups of order 10.

Theorem 3.51 *Any group of order 10 is isomorphic to $\mathbb{Z}/10$ or D_5.*

Proof. To begin we note that the only abelian group of order 10 is $\mathbb{Z}/10$. So we can assume G is nonabelian.

Since $|G| = 10$, the elements of G have orders $1, 2, 5$ or 10. If every element has order 2, G is abelian, so we can exclude this case. If there is an element of order 10, then G is cyclic, so we can exclude this case as well. Thus G contains an element a of order 5.

Let $H = \langle a\rangle \subset G$. Since $|G/H| = 2$, H is normal. Take any element $b \in G - H$. Then we have

$$bab^{-1} = a^k$$

for $k = 2, 3$ or 4. Moreover, the order n of b is 2 or 5. Since

$$b^n a b^{-n} = a^{k^n} = a,$$

we have $k^n = 1 \bmod 5$. But $k^5 = k \bmod 5$, so the order of b must be 2 and k must be 4. In other words, $ba = ab$ and $b^2 = e$.

This shows that G satisfies the defining relations for D_5, so there is a surjective map $D_5 \to G$. Since $|G| = 10$, this map is an isomorphism.

$G \to \mathrm{Aut}(H) \cong (\mathbb{Z}/5)^*$, which is a group of order 4. Since G is nonabelian, the image of this map is nontrivial; and since $|G| = 10$, the image has order 2. Thus we can find an element $b \in G$ such that $bab^{-1} = a^4 b = a^{-1} b$. Then b has even order, so b has order 2, and hence $G \cong D_{10}$. ∎

Proof variant. As above we have a normal subgroup $H \subset G$ of order 5. Hence we have a homomorphism $\phi : G \to \mathrm{Aut}(H)$ given by conjugation. Now $\mathrm{Aut}(H) \cong (\mathbb{Z}/5)^* \cong \mathbb{Z}/4$, so $|\phi(G)| = 1$ or 2. If $|\phi(G)| = 1$ then G is abelian, so $G \cong \mathbb{Z}/10$; otherwise, $\phi(G) = \{1, 4\}$, and hence $G \cong D_5$. ∎

4 Knot Theory

We now turn to the final topic of this course, *knot theory*. The ideas we will discuss in this section belong to *topology*. Unlike algebra and set theory, topology is concerned with continuous objects like circles and surfaces. However we will see that in the case of knot theory, the study of loops in space can be reduced to a combinatorial theory, and that algebra and combinatorics play a useful role in attacking this subject.

4.1 Knots and links

Introduction. What is a knot? It is a smooth closed curve in 3-space. A knots is not allowed to cross itself. A knot can be moved a little so it is a polygon. We do not allow wild knots.

Two knots K_0 and K_1 are *equivalent* if you can make a smoothly moving family of knots K_t that connects them. You can imagine this motion taking place in discrete steps, $K_0, \ldots, K_n$, where K_i and K_{i+1} differ by a triangle move.

A *link* is defined similarly as a finite number of disjoint closed loops.

Knot projections. A useful way to discuss knots is by projections: you put the knot almost in a plane, with pairs of strands meeting at crossings.

Any knot (or link) can be given a knot projection; in fact a generic projection will work. You just have to avoid the directions tangent to the knot, and the directions of lines passing through the knot in 3 points (taken with multiplicities). Each locus we must avoid forms a one-dimensional set in the 2-sphere of projections $\pi : \mathbb{R}^3 \to \mathbb{R}^2$.

Examples of knots and links.

1. The unknot. There are several projections. Any knot projection with 0, 1 or 2 crossings is the unknot. Any knot projection you can draw without anticipating crossings is the unknot.

2. The trefoil knot, 3_1.

3. The figure-eight knot, 4_1.

4. The (p, q)-torus knot/link. Start with q strands and form a braid of the form β^p, where β is a cyclic permutation; then close. If p and q are relatively prime, you get a knot. The $(1, 3)$ torus knot is the unknot; $(2, 3)$ is the trefoil.

   In general the (p, q) torus link as $q/\gcd(p, q)$ components. For example, the $(1, 2)$ torus link is the Hopf link.

5. The unlink on two components.

6. The Hopf link.

7. The Borromean rings, after the Renaissance family crest of the Borromeas. (Exercise: these rings are not round!)

8. The Whitehead link.

History. Lord Kelvin conjectured that atoms are knots in ether. Tait and Little undertook the tabulation of knots up to ten crossings, a 10 year project completed around 1899. It was not until 1974 that the lawyer Ken Perko found a mistake in their tables.

In recent times biologists have discovered that DNA is often knotted. The classification of 3-dimensional spaces is intimately tied up with knot theory.

Showing two knots are the same. Suppose K_1 and K_2 are projections that happen to correspond to the same knot. Then you can transform K_1 to K_2 be a sequence of *Reidemeister moves* (or their inverses). These moves are:

I Transform one strand into a loop with one crossing. The singularity is a cusp.

II Transform two parallel strands by adding two crossings. The singularity is a tangency.

III Transform three strands preserving the number of crossings. The singularity is a triple-point.

The Reidemeister moves can be remembered by the number of strands they involve. Planar isotopy is also allowed. The Reidemeister moves also work for links.

Proof. One way to approach the proof is to consider what kinds of singularities can arise as you view a generic knot projection during isotopy. The generic singularities are cusps, tangencies and triple points, accounting for the 3 moves. ∎

Examples.

1. Draw a trefoil with one crossing wrong. This can be undone by the sequence III, I, II.

2. Tangle up the trefoil.

3. For each crossing of 6_3, change it and simplify the result.

4.2 Linking number and tricoloring

Oriented knots and links. A knot or link is **oriented** if we have chosen a direction (usually indicated by an arrow) to traverse each component. A link with n components has 2^n possible orientations.

Factorization and prime knots. Given two oriented knots, there is a natural way to join them together to form their sum $K = K_1 \# K_2$. If K_1 is the unlink, then $K = K_2$. It turns out that any oriented knot can be unique factored into **prime knots** . Knot tables list only prime knots.

Showing two links are different. Let $L = K_1 \cup K_2$ be a two component oriented link. The **linking number** $\ell(K_1, K_2)$ is defined as follows: at each crossing between K_1 and K_2, count $+1$ if it is a right-hand turn to get onto the overpass, otherwise -1. Add up and divide by two; this is $\ell(K_1, K_2)$.

Theorem 4.1 *The linking number is an invariant of an oriented pair of knots* K_1, K_2.

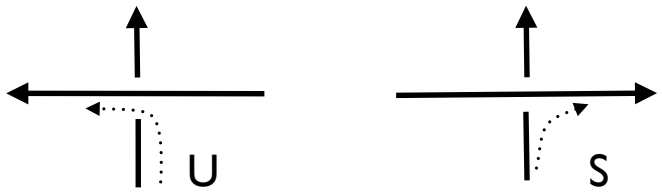

Figure 8. Safe and unsafe crossings of oriented strands of a knot or link.

Even though it is defined using a projection, the answer for two different projections is the same.

Proof. Type I moves don't involve both components. A type two moves creates a pair of crossings of opposite sign. And type III doesn't really change any pair of strands, only the configuration of all three. ∎

Examples: the unlink, Hopf link, the Whitehead link, and the Borromean rings.

Unoriented links. The *absolute value* of the linking number is independent of orientation.

Tricoloring. We still haven't shown there are any real knots. To do this, let's say a **tricoloring** of a knot or link projection is an assignment of colors R, G, B to the arcs such that:

1) at least 2 colors are used; and at each crossing either
2a) all three strands have the same color or
2b) all three strands have different colors.

Theorem 4.2 *If one projection of a knot or link can be tricolored, then they all can.*

Proof. Proof. We must check the Reidemeister moves.

(I) The entire loop must be a single color.
(II) If the parallel strands are colored R and G, then we color the underloop B.
(III) If 3 colors appear at the central crossing, then a crossing where 3 colors are the same either appears or disappears.

In all 3 cases, if 2 colors were used before, then 2 are used after (see especially 2, which can increase the number of colors from 2 to 3). ∎

Examples of tricolorings.

1. The unknot cannot be tricolored.

2. The trefoil **can** be tricolored. Thus the trefoil is really knotted!

3. The figure 8 knot cannot be tricolored. To check this, focus on one crossing. Either the two understrands are the same, or they are different. In the first case one finds all the strands have the same color, and in the second case one runs into a contradiction.

   Thus tricoloring is not powerful enough to detect all knots.

4. The unlink on 2 comonents L can be tricolored with 2 colors — if drawn with no crossings — and with 3 colors — if drawn with 2 crossings.

5. The Hopf link cannot be tricolored.

6. The Whitehead link cannot be tricolored. This shows it is not the unlink.

The *number* of tricolorings is also an invariant of the knot or link.

For example, the trefoil can be tricolored in 6 different ways. The connect sum $T_1 \# T_2$ of *two* trefoils can be tricolored in 24 different ways! (Each of the 6 tricolorings of T_1 can be extended to a tricoloing on $T_1 \# T_2$ in 3 different ways — including the monochromatic colors on T_2. Then, the 3 monochromatic colrings of T_1 can each be extended to 2 different tricolorings on $T_1 \# T_2$. This gives $6 \cdot 3 + 3 \cdot 2 = 24$ tricolorings.

4.3 The fundamental group

In this section we introduce a very powerful invariant of a knot or link L, namely the *fundamental group* G_L.

We begin by describing how to compute this group and what its basic properties are. Then we explain its topological meaning, and finally check that it really is invariant under Reidemeister moves. Using this group, we can clarify the idea of 3-coloring and find other ways to tell knots and links apart.

The group G_K. Let K be an *oriented* knot or link *projection*. We define a finitely presented group G_K as follows:

1. There is one *generator* for each strand $a, b, c \ldots$ of K; and

2. There is one relation for each *crossing* in the diagram. Assuming that running along the orientation of the knot, strand a passes under strand c to become strand b. Then we add the relation:

   $ac = cb$ if the crossing was *safe*, and
   $ca = bc$ if the crossing was *unsafe*.

Using Reidemeister moves, we will later prove:

Theorem 4.3 *If K and K' are equivalent knots or links, then G_K is isomorphic to $G_{K'}$.*

But first some examples.

The trivial knot. The unknot can be presented with one strand and no crossings, so have

$$G_K = \langle a \rangle \cong \mathbb{Z}.$$

The trefoil knot. In practice it is useful to label the strands $a, b, c, \ldots$ so they occur as consecutive letters in the alphabet as one runs around the knot in the direction dictated by its orientation. Then first relation has the form $ax = xb$ or $xa = bx$, the second has the form $yb = cy$ or $by = yc$, etc.

Now in the standard projection of the trefoil knot, we have just three strands a, b, c, all distinct at each crossing, and all crossings are of the same type. Assuming all crossings are safe, this gives:

$$G_K = \langle a, b, c \ : \ ac = cb, ba = ac, cb = ba \rangle.$$

Is this group nontrivial? Yes, in fact there is a surjective homomorphism from G_K to $\mathbb{Z}$ defined by $\phi(a) = \phi(b) = \phi(c)$. This map exists because it sends each relation in G_K to the relation $1 + 1 = 1 + 1$, which is true in $\mathbb{Z}$. The same reasoning applies to any knot.

Theorem 4.4 *For any knot K, there is a surjective homomorphism $\phi : G_K \to \mathbb{Z}$.*

Braid relation and the trefoil knot. Here is another presentation for G_K. First, we can drop the final relation $cb = ba$, because it follows from the other two. Then, we can use the equation $c = a'ba$ to eliminate c. What is left is the presentation

$$G_K = \langle a, b \,:\, aba = bab \rangle.$$

This is very close to the presentation we gave earlier for S_3. In fact, to get S_3 one just needs to add the relations $a^2 = b^2 = \mathrm{id}$. This shows:

Theorem 4.5 *There is a surjective homomorphism from the trefoil knot group G_K to S_3.*

Corollary 4.6 *The trefoil knot group G_K is nonabelian, and hence K is not equivalent to the unknot.*

Tricolorings revisited. Let us now consider for a moment relations like $ac = cb$ where a, b, c are elements of D_n. The following fact is easily verified:

> *We have $(r^i f)(r^k f) = (r^k f)(r^j)f$ in D_n if and only if $i + j = 2k \bmod n$.*

In particular, in S_3 this relation holds iff $i + j + k = 0 \bmod 3$. And this in turn holds iff all the integers $i, j, k \in \mathbb{Z}/3$ are the same, or all are different.

Theorem 4.7 *There is a surjective homomorphism $\phi : G_K \to S_3$ if and only if K can be tricolored.*

Proof. Suppose K can be tricolored, and let us label the colorings $0, 1, 2 \in \mathbb{Z}/3$. Define a map $\phi : G_K \to S_3$ on the generators of G_K by $\phi(x) = r^i f$ if the strand x has color i. Now at a typical crossing, we have a relation like $ac = cb$ where (a, b, c) are colored i, j, k. This relation is satisfied in S_3 if and only if $i + j + k = 0 \bmod 3$, which happens iff all 3 colors are the same or all 3 are different. But the definition of a tricoloring is exactly that this coloring condition holds at each crossing. Thus ϕ gives a homomorphism, and it is surjective because at least two colors are used.

The converse is similar. Suppose we have a surjective homomorphism $\phi : G_K \to S_3$. First, we observe that any two generators of G_K are *conjugate* in G_K. This implies that their images in S_3 are all conjugate. Thus if one of them is a power of r, they all are; but then the map is not surjective. Consequently, on every generator x we have $\phi(x) = r^i f$ for some $i \in \mathbb{Z}/3$. Using i to color the strand labeled x, we obtain a 3-coloring of K by the same reasoning as above. ∎

Links. The same method associates a group to a link. In particular, if H is the Hopf link and L is the unlink on two components, then the group

$$G_H = \langle a, b \,:\, ab = ba \rangle \cong \mathbb{Z} \oplus \mathbb{Z}$$

is abelian for the Hopf link, while

$$G_L = \langle a, b \rangle = \mathbb{Z} * \mathbb{Z}$$

is free for the unlink.

Tricoloring links. One must be careful with links because not all generators of G_L are conjugate. We find that L admits a tricoloring iff there is a surjective homomorphism

$$\phi : G_L \to S_3$$

that sends every generator to a flip in S_3. (The last condition is automatic for a knot.)

The topological fundamental group. We now pause to explain the more general mathematical theory underlying the definition of G_K.

Let A be a reasonable connected space, like a curve or surface in $\mathbb{R}^3$. Picking a basepoint $a \in A$, we wish to make the set of *loops* $\gamma \subset A$ that begin and end at a into a group. (Formally, a loop is a continuous function $\gamma : [0, 1] \to A$ such that $\gamma(0) = \gamma(1) = a$.)

Composition is defined in an obvious way: $\gamma_1 * \gamma_2$ means you first traverse γ_1, then traverse γ_2. This makes sense because γ_1 ends at a and γ_2 starts there.

The 'trivial loop', given by a constant function γ, serves as the identity. As usual, the main problem is *inverses*. To create them, we declare that $\gamma_0 \sim \gamma_1$ if they can be interpolated by a family of continuous loops

$$\gamma_t \subset A,$$

all based at a. The space of equivalence classes of such loops is denoted by

$$\pi_1(A, a)$$

and called the *fundamental group* of A.

Example: the trivial group. In $\mathbb{R}^3$ with basepoint $a = (0, 0, 0)$, every loop γ_0 can be shrunk to the constant loop γ_1 by setting $\gamma_t = (1-t)\gamma_0$. Thus

the fundamental group of $\mathbb{R}^3$ is *trivial*. When $\pi_1(A, a)$ is trivial, we say A is *simply–connected*.

Example: the circle and $\mathbb{Z}$. Let $A = S^1 \subset \mathbb{C}$ with the basepoint $a = 1$. Let $w(\gamma) \in \mathbb{Z}$ count the net number of times that γ winds counter–clockwise around the origin. It turns out that the *winding number*

$$w : \pi_1(S^1, a) \to \mathbb{Z}$$

is an isomorphism.

Quotient spaces. Here is another useful way to look at $\pi_1(A, a)$. Suppose X is a simply–connected space, and a group G acts freely on X in a reasonable way such that $A = X/G$. Then, it turns out we have

$$\pi_1(X/G) \cong G.$$

For example, if we describe S^1 as $\mathbb{R}/\mathbb{Z}$, then we get $\pi_1(S^1) = \mathbb{Z}$.

The torus. For a more interesting example, observate that the torus can be described as $T = \mathbb{R}^2/\mathbb{Z}^2$, and thus

$$\pi_1(T) \cong \mathbb{Z}^2.$$

It is useful to take two standard generators a, b of $\pi_1(T)$ and draw a picture of $[a, b] = aba^{-1}b^{-1}$ and verify that it is trivial, showing that $\pi_1(T)$ is abelian.

The bouquet of two circles. Let Y be two circles A and B joined at a single point p. Every loop based at p first winds some number of times around A, then some number of times around B, etc. Writing down the corresponding product of a's and b's, we obtain an isomorphism to the free group

$$\pi_1(Y, p) \cong \mathbb{Z} * \mathbb{Z} = \langle a, b \rangle.$$

We can also take the Cayley group T of $G = \langle a, b \rangle$ and regard Y as the quotient T/G. Since a tree is simply–connected, this also shows $\pi_1(Y) \cong G$.

Torus with a hole. If we remove a disk from the torus, to obtain a surface with boundary $T^0 = T - D$, then we can no longer argue that $[a, b] = \mathrm{id}$. In fact, $[a, b]$ represents a loop around ∂D, and it turns out that

$$\pi_1(T^0) \cong \langle a, b \rangle.$$

A surface of genus two. Finally let S be a surface of genus two. We can build S by gluing together two copies of T^0 along their boundaries. This only introduce one relation into the fundamental group, and we have

$$\pi_1(S) \cong \langle a, b, c, d \ : \ [a, b] = [c, d] \rangle.$$

The knot group. We can now explain where G_K comes from: for a suitable basepoint p, we have

$$G_K \cong \pi_1(\mathbb{R}^3 - K, p).$$

In other words, G_K is the group of flight plans for airplanes leaving from a base located outside the knot and flying around it.

For each strand a of K, we have a path γ_a that starts at p, descends to the plane of the knot projection, makes a right–hand turn around the directed strand, and returns to p. The relations at a crossing are explained in Figure 9.

This algorithm we have given computes the *Wiritinger presentation* of $G_K = \pi_1(S^3 - K)$.

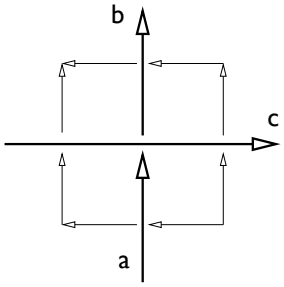

Figure 9. Proof that $ac = cb$ at safe underpass.

The Carabiner Trick. The Hopf link L is *less knotted* than the unlink, since $G(L) \cong \mathbb{Z} \oplus \mathbb{Z}$. As a trick, one can weave the commutator through two unlinked carabiner, in such a way that the loop comes free when the carabiner are linked! (Cf. homework on computing $G(L)$.)

Changing presentations. To prove G_K is a knot invariant, not just an invariant of the knot projection, it is important to understand elementary (or Tietze) moves on a presentation. There are just two:

(1) $\langle g_i \ : \ r_i \rangle \iff \langle g_i \ : \ r_i, s \rangle$, where s is a consequence of the given relations. That means s is a product of conjugates of the r_i.

(2) $\langle g_i \ : \ r_i \rangle \iff \langle g_i, h \ : \ r_i, h = w(g_1, \ldots, g_n) \rangle$, where $w(\cdot)$ is a word in the generators g_i. This means we can add a new generator so long as we add a relation putting it in the group generated by the (g_i).

Invariance of G_K. We can now prove Theorem 4.3.

Proof. We must check the Reidemeister moves.

(I) A loop gives $\langle a, b \ : \ aa = ba \rangle$, so we can use Tietze move (2) to eliminate b.

(II) Suppose the arc a is underpassed by (b, c, d). Then we get from the 2 crossings the relations $ba = ac, ac = da$. From this we derive $b = d$ (Tietze 1), then eliminate c, d (Tietze 2). We are left with a, b and no relations, which is the contribution of two parallel arcs.

(III) Let the topmost arc be c, NW to SE, over (a, b), which runs SW to NE, with (d, x, e) passing W to E under a and c. The big diagonal safe crossing gives the relation

$$ac = cb,$$

while the other two crossings give

$$R = \langle ad = xa, xc = ce \rangle.$$

See Figure 10. We can solve for $x = cec'$ and substitute this in $ad = xa$ to get $ad = cec'a$ or equivalently:

$$R' = \langle ada' = cec' \rangle.$$

After the Reidemeister move we get a new arc y and relations

$$S = \langle dc = cy, by = eb \rangle.$$

We can again solve for $y = b'eb$ and get $dc = cb'eb$ or equivalent:

$$S' = \langle c'dc = b'eb \rangle.$$

To show the group hasn't changed, we need to show S' is a consequence of R' and vice–versa. To do these we use the relation $ac = cb$, which holds

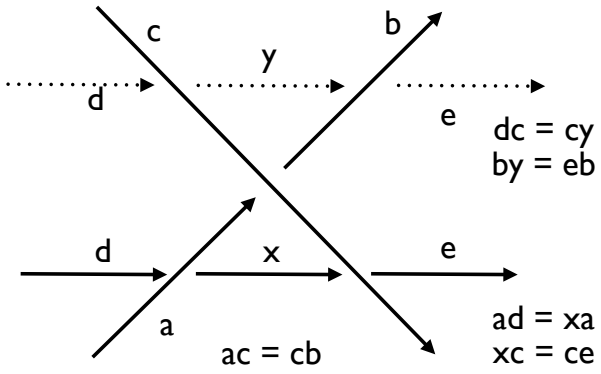

Figure 10. Reidemeister move III.

both before and after the Reidemeister move. This relation is equivalent to $c'a = bc'$, so R' is equivalent to

$$R'' = \langle c'ada'c = e \rangle$$

from which we get

$$R''' = \langle bc'dcb' = e \rangle$$

which is equivalent to S'. The equivalence can be reversed so the groups are isomorphic.

The figure eight knot group. The relations for the figure 8 knot are computed in Figure 11. Note that it is helpful to label the consecutive strands by consecutive letters, (a, b, c, d); and then record the overcrossings as shown at the right, as an intermediate step in the calculation. We have the presentation

$$G = \langle a, b, c, d \ : \ ad = db, ab = ca, cb = bd, cd = ac \rangle.$$

This presentation can be simplified to:

$$G = \langle a, b \ : \ aba'ba = bab'ab \rangle.$$

To show the figure–eight knot is really knotted, one can check:

Theorem 4.8 *There exists a surjective homomorphism $\phi : G_K \to A_4$.*

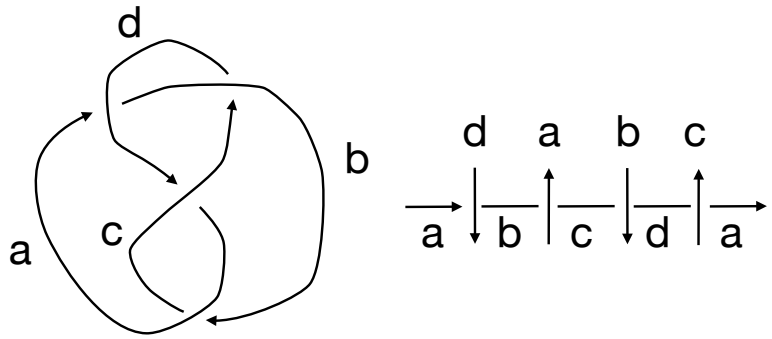

Figure 11. Computing relations for the figure eight knot.

Proof 1. Define $\phi(a) = (123)$ and $\phi(b) = (134)$. We have $ab \mapsto (123)(134) = (234)$, $ba \mapsto (134)(123) = (124)$, so $aba'ba \mapsto (234)(321)(124) = \mathrm{id}$, and similarly $bab'ab \mapsto (124)(431)(234) = \mathrm{id}$, so ϕ gives a homomorphism on G_K which is clearly surjective. ∎

Proof 2. A better way to think of this is to draw a tetrahedron flattened out in the plane, with A an equilateral triangle and B, C, D arrayed counter–clockwise around it. Let us identify each fact with its counter-clockwise twist of 120°. Then conjugating by A cyclically permutes B, C, D; that is, $ABA' = C$, $ACA' = D$ and $ADA' = B$. In particular the desired relation $AB = CA$ holds. One can similar check that the other 3 relations defining G_K hold as well. ∎

There is also a surjective map to D_5, as a '5-coloring' argument shows.

Topology of the linking number. Now that we understand the meaning of G_K more geometrically, we can give a group–theoretic perspective on the linking number of a pair of knots.

Theorem 4.9 *The linking number for $L = K_1 \cup K_2$ corresponds to the abelianization of the element of $\pi_1(\mathbb{R}^3 - K_1)$ represented by K_2.*

Compare Theorem 4.7, which gives a natural map $G_K \to \mathbb{Z}$.

Proof. The proof is a little tricky. Working with a link projection, one can first change crossings of K_2 with itself so K_2 becomes the unknot. This change does not change our projection-based calculation of the linking number (obviously), nor does it change the image of K_2 in $\pi_1(\mathbb{R}-K_1)$ (obviously).

Now re-arrange the projection so K_2 is a counter-clockwise round circle. It is then clear that the number of outward-heading crossings of K_1 with K_2 is the same as the number of inward-heading crossings.

Count the crossings in 4 groups, TO, BO, TI, BI, where T/B means K_1 is on top/bottom, and O/I means it is headed out/in. Letting P be the linking number in π_1, and L the linking number from the diagram, we have

$$\begin{aligned} TO + BO &= TI + BI \\ L &= (TO + BI - TI - BO)/2 \quad \text{and} \\ P &= TO - TI. \end{aligned}$$

Using the first equation we have $P = BI - BO$; averaging these two expressions for P, we obtain $P = L$. ∎

How good an invariant is the knot group? Since G_K can be pretty complicated, one might wonder if G_K determines K. The fact that this is essentially true is a contemporary theorem.

Theorem 4.10 (Gordon–Luecke) *Let $(G_K, H(K))$ be a knot group and a subgroup $H(K) \cong \mathbb{Z}$ generated by a meridian. Then K is equivalent to K', or its mirror image, iff there is an isomorphism $G_K \cong G(K')$ sending $H(K)$ to $H(K')$.*

The $H(K)$ is needed for the square and granny knots, which have isomorphic groups. Often the $H(K)$ is unnecessary.

4.4 Knot polynomials

Our final knot/link invariant will be a *Laurent polynomial* $X(L)$ in one variable A, discovered (in a different form) by Vaughn Jones in 1984.

While the knot group G_K is a powerful invariant (in fact a complete invariant of the knot), it is hard to work with; for example, it is hard to

check in general if two different knot groups are isomorphic or not. (Even proving that G_K is nontrivial requires work.) The coloring invariants we can define from it are easy to work with, but not powerful enough to tell many knots apart.

It is trivial, however, to check if two polynomial agree. The Jones polynomial also seems to be very good at telling knots apart; for example, there only known example with trivial Jones polynomial is the unknot.

The Kaufmann bracket. To define the Jones polynomial, we begin with the *Kaufmann bracket* $\langle L \rangle$. This is also a polynomial, defined for a knot projection by:

(i) $\langle O \rangle = 1$;
(ii) $\langle L \rangle = A\langle L_s \rangle + A^{-1}\langle L_u \rangle$; and
(iii) $\langle L \cup O \rangle = -(A^2 + A^{-2})\langle L \rangle$.

Here L_u and L_u are obtain from L by focusing on one undercrossing and replacing it with safe, right on-ramps (L_s) or unsafe, left onramps (L_u). The circle O denotes the unknot.

Example. For the unknot with one twist, $\langle L \rangle = -A^3$ (or $-A^{-3}$, depending on the direction of the twist).

Example. For the Hopf link, we have

$$\langle H \rangle = -A^4 - A^{-4}.$$

Exercise: why is $\langle L \rangle$ well-defined? Why doesn't it depend, for example, on the order in which we resolve crossings?

Reidemeister II. This move leaves $\langle L \rangle$ unchanged, as an easy computation shows.

Reidemeister III. By resolving the middle crossing in two different ways, then applying move II, we see that Reidemeister move III also leaves $\langle L \rangle$ unchanged.

In fact, equations (ii) and (iii) are chosen just to insure that I and II leave the bracket invariant, and (i) is just a normalizing factor.

Reidemeister I. Now let L^+ denote L with a safe loop added by Reidemeister move I. Then we have

$$\langle L^+ \rangle = (-A^3)\langle L \rangle.$$

The writhe. To account for these changes, we use another invariant of a projection that is affected by move I but not moves II or III. This is the

writhe $w(L)$ or 'self-linking number' of a knot or oriented link, obtained by adding up the signs of *all* self-crossings.

Note that the writhe of a *knot* does not depend on an orientation — while the writhe of a link *does*. It is *not* an invariant of the knot; for example, a single twist gives the unknot writhe ± 1.

The writhe is unaffected by moves II and III for the same reason that the linking number is an invariant. But Reidemeister move I changes it: we have

$$w(L^+) = w(L) + 1.$$

The polynomial $X(L)$. We can combine the writhe with the bracket polynomial to get an honest invariant of an oriented link, or unoriented knot, namely:

$$X(L) = (-A^3)^{-w(L)}\langle L\rangle.$$

The reason this works is that:

$$\begin{aligned} X(L^+) &= (-A^3)^{-w(L^+)}\langle L^+\rangle \\ &= (-A^3)^{-w(L^+)-1}(-A^3)\langle L\rangle \\ &= X(L). \end{aligned}$$

The Hopf link revisited. Give both components the same orientation; then both crossings are positive, so we must multiply the bracket by $(-A^3)^{-2} = A^{-6}$, and we then obtain

$$X(H^+) = -A^{-2} - A^{-10}.$$

Knot without orientations. Reversing the orientation does *not* change the writhe of a knot, and does not affect the bracket polynomial. Thus $X(K)$ is an invariant of *unoriented* knots.

Mirror images. If we replace K by its mirror image $-K$, then we $\langle -K\rangle$ is just $\langle K\rangle$ with A replaced by A^{-1}. Similarly, $w(-K) = -w(K)$ — safe and unsafe crossings are interchanged. Therefore $X(-K)(A)$ is $X(K)(A^{-1})$.

The positive trefoil. Consider the trefoil 3_1^+, with all 3 crossings safe (positive). Then $w(3_1^+) = 3$. The bracket polynomial is given by

$$\langle 3_1^+\rangle = A^{-7} - A^{-3} - A^5.$$

Multiplying by $(-A^3)^{-3}$, we obtain

$$X(3_1^+) = -A^{-16} + A^{-12} + A^{-4}.$$

The fact that the polynomial is *not* symmetric proves that trefoils come in two types, right and left handed! That is,

$$X(3_1^-) = A^4 + A^{12} - A^{16}.$$

(*Note:* Adams' table shows 3_1^-, the mirror image of 3_1^+.)

The figure eight knot. Start with the projection of 4_1 in Adams' tables, then change the crossings at the top. The two resulting knots are the Hopf link with a safe twist and the positive trefoil. Using the rule for Reidemeister move I, we find

$$\langle 4_1 \rangle = A\langle H_s \rangle + A^{-1}\langle 3_1^+ \rangle = A(-A^3)\langle H \rangle + A^{-1}\langle 3_1^+ \rangle.$$

We already computed $\langle H^+ \rangle = -A^{-4} - A^4$ and $\langle 3_1^+ \rangle = A^{-7} - A^{-3} - A^5$, from which we find:

$$\langle 4_1 \rangle = A^{-8} - A^{-4} + 1 - A^4 + A^8.$$

In addition, $w(4_1) = 0$, so we get $X(4_1) = \langle 4_1 \rangle$.

Jones' polynomial. In all our examples of $X(K)$ we see only 4th powers of A. To get a simpler expression, the **Jones polynomial** $V(L)$ is defined by replacing A with $t^{-1/4}$. Thus

$$V(3_1^+) = t + t^3 - t^4,$$

and

$$V(4_1) = t^{-2} - t^{-1} + 1 - t + t^2.$$

Skein relations. The Jones polynomial of an oriented knot or link can be computed directly using the following properties: $V(U) = 1$ for the unknot, and

$$(t^{1/2} - t^{-1/2})V(L_0) = t^{-1}V(L_s) - tV(L_u),$$

where L_s and L_u are unsafe crossings, and L_0 is the oriented link obtained when the crossing is removed.

Why does this definition suffice for computations? Because, by changing enough crossing, any diagram becomes a diagram for the unknot. To find

the changes to make draw the knot diagram so that as you run along the knot, all crossings are undercrossings.

Computations with the Jones polynomial. By definition, for the unknot we have

$$V(U) = 1.$$

If we put one twist in the unknot, and then apply skein relations, we obtain $(L_0, L_s, L_u) = (2U, U, U)$ which gives

$$V(2U) = -(t^{1/2} + t^{-1/2}).$$

If we start with the positive Hopf link H^+ as L_s, then $(L_0, L_s, L_u) = (U, H^+, 2U)$, which gives

$$V(H^+) = -t^{1/2}(1 + t^2).$$

If we start with the trefoil $3_1^+ = L_s$, we get $(L_0, L_s, L_u) = (H^+, 3_1^+, U)$, which gives

$$V(3_1^+) = t + t^3 - t^4.$$

If we start with the figure eight knot as $4_1 = L_u$, we get $(L_0, L_s, L_u) = (H, U, 4_1)$, giving

$$V(4_1) = t^{-2} - t^{-1} + 1 - t + t^2.$$

Continuing in this fashion, one can compute the Jones polynomial for a table of knots and links without too much difficulty. It is necessary however to do links at the same time. The next case to be covered in the pattern above is the closure 4_1^{2+} of the $(2, 4)$ torus knot with all crossings safe. This is a link with two components with linking number two. We have $(L_0, L_s, L_u) = (H^+, 4_1^{2+}, 3_1^+)$, which gives

$$V(4_1^{2+}) = -t^{3/2} - t^{7/2} + t^{9/2} - t^{11/2}$$

Computational complexity. The most straightforward computation of $X(K)$ takes 2^c steps where c is the number of crossings of K.

Quantum field theory. One can think of the crossings of a knot as undergoing fluctuations, to different states with different energies. Weighting the states by their energies we get the **partition function** which is the knot polynomial. The variable is then related to the temperature of the system (statistical mechanics).

Representations of the braid group. Jones' discovery of $V(K)$ emerged from his study of operator algebras, which led to some interesting representztions of the braid group,

$$\phi_n : B_n \to \mathrm{GL}_N(\mathbb{Z}[t, t^{-1}]).$$

Then $\mathrm{Tr}(\phi_n(\beta))$ is an invariant of the conjugacy class of the braid β. By studying the behavior of these invariants under the *Markov moves*, he was able to extract a knot invariant. The complexity of computing this trace is polynomial if the number of strands n is fixed.

To apply this method one can appeal to:

Theorem 4.11 (Alexander) *Every knot or link arises as the closure of a braid.*

Sketch of the proof. ∎

An unsolved problem. The Jones polynomial, or its variant $X(K)$, gives different answers for every knot with 9 or fewer crossings.

The *only known knot* with $X(K) = 1$ is the unknot.

Alternating links. A link *projection* L is *alternating* if the crossings are alternate over and under as one traverses any component of the link. It *reduced* if there is no obvious way to simplify the projection by Reidemeister move I or variants thereof; more precisely, if each component of $\mathbb{R}^2 - L$ is bounded by an embedded loop in the plane (it does not touch itself).

The *width* of a Laurent polynomial $P(A)$ is the difference between the largest and smallest degrees of A which occur in the polynomial.

Theorem 4.12 *The number of crossings of L is given by 4 times the width of $\langle L \rangle$. In particular, any two alternating diagrams of the same link have the same number of crossings.*

(The final statement was a long–standing conjecture.)

Proof. Let c be the total number of crossings of L, and consider the state expansion of $\langle L \rangle$. It is a sum of terms of the form

$$A^{s-u}(-A^2 - A^{-2})^{m-1},$$

where s is the number of crossings made safe, u is the number made unsafe, and m is the resulting number of circles or unknots that appear after the crossings are resolved. Of courser $s + u = c$.

Color the regions on the plane complementary to L (including the unbounded component) alternating colors. The total number of regions is $R_s + R_u = R = c + 2$ (exercise). When we make all the crossings safe, the $m = R_s$ regions with one color become unknots, and when we make them all unsafe, the remaining R_u regions become unknots. It can be shown that these terms determine the width, and hence

$$W = c + 2(R_s - 1) + c + 2(R_u - 1) = 2c + 2(R - 2) = 4c.$$

■

4.5 Immersed spheres

1. **The time traveler.** Suppose at noon we begin to travel in time, making the dial move on the clock in side our time machine. It goes around and around, maybe forward, maybe back — but at the end the clock says 12:00 again. Now we have traveled a definite integral numbers of half-days n – this is the *winding number* of the hand around the clock. The actual return is to the present plus $n/2$ days. (Even though n may be negative.)

2. **Circle eversion.** Can you turn the circle inside out? Consider changing closed immersed loops in the plane by the three Reidemeister moves. Then it turns out to be possible to evert the circle — just apply II followed by two I's.

   Now let's rule out I's, since they can't be done continuously without pinching. Then the circle cannot be everted!

3. **Degree.** To prove that C and $-C$ are not equivalent, we must put an arrows on our loops. Now given an arrow, we can walk around the loop, with a clock whose hour hand points in the direction of travel. If we start at an upward heading point, then we begin and end at 12:00. The *degree* of the loop is the (signed) number of 12 hour periods our clock has turned by the time we come back.

The degree is always an integer. So if you move the loop gradually, this integer can never change! But $d(-C) = -d(C)$ so the circle cannot be everted.

4. **Smiles and frowns.** One way to compute the degree is to count the number of frowns, minus the number of smiles, where the time is 3pm.

5. **Turning the sphere inside out.** Incredibly, a 2-sphere can be turned inside out through immersions.

5 Summary

1. *Sets.*
   (a) Axioms. $A \in B$.
   (b) Paradoxes. Let $S = \{A : A \notin A\}$. Then is $S \in S$? The barber of Seville.
   (c) $0 = \{\}$, $n + 1 = n \cup \{n\}$.
   (d) Sizes of infinity (Cantor): $|\mathcal{P}(A)| > |A|$. A line ($\mathbb{R}$) is bigger than $\mathbb{N}$.
   (e) Schröder-Bernstein. $|A| \leq |B|$ and $|B| \leq |A|$ implies $|A| = |B|$.

2. **Groups.**
   (a) Axioms.
   (b) Isomorphism. Classification of groups of order up to 7.
   (c) Cyclic groups. Every subgroup of $\mathbb{Z}$ is cyclic. $\langle a, b\rangle = \gcd(a, b)\mathbb{Z}$. $(a\mathbb{Z}) \cap (b\mathbb{Z}) = \mathrm{lcm}(a, b)\mathbb{Z}$. a generates $\mathbb{Z}/n$ iff $\gcd(a, n) = 1$.
   (d) The Cayley graph.
   (e) Lagrange: $H \leq G \implies |H|$ divides $|G|$. The order of any quotient of G also divides $|G|$.
   (f) Group actions. $|A| = |G|/|\operatorname{Stab}(a)|$.
   (g) Examples: $\mathbb{Z}/n$, S_3, S_n, A_n (sliding puzzle), D_n, V_4, the quaternion group Q_8, $SL_2\mathbb{Z}$, the free group on 2 generators.
   (h) 5 Platonic solids, 3 groups: A_4, S_4, A_5. Kepler.

(i) Homomorphism and quotient groups.

(j) Normal subgroups and group presentations.

3. **Knots.**

(a) Isomorphism (equivalence).

(b) Knot and link projections and Reidemeister moves.

(c) The unknot, unlink, Hopf link, trefoil, figure eight and Borromean rings.

(d) Tricoloring — an invariant.

(e) Linking number.

(f) The knot and link group.

(g) The fundamental group of space.

(h) The Wirtinger presentation for G_K.

(i) The trefoil group maps to S_3, the figure eight group maps to A_4, the Hopf link has $G(L) = \mathbb{Z}^2$, the unlink has $G(L) = \mathbb{Z} * \mathbb{Z}$.

(j) The carabiner trick.

(k) Knot polynomials. The bracket and the writhe.

(l) Unsolved problem: does $X(K) = 1$ imply K is unknotted?

Turning the sphere inside out.