SOCR MATH

KEN RICHARDSON

Contents

1. Notation

Below is some notation I will use.

Notation	**meaning**
$g : A \to B$	g is a function with domain A and codomain B
$\mathbb{R}$	the set of all real numbers
$\mathbb{Z}$	the set of all integers
$\mathbb{C}$	the set of all complex numbers
$x \in A$	x is an element of the set A.
$C \subseteq D$	The set C is a subset of the set D.
$\exp(x)$	e^x
$\log(x)$	$\log_e(x) = \ln(x)$
$\mathbb{R}^2$	the set of ordered pairs (x, y) such that $x, y \in \mathbb{R}$
$\mathbb{N}$	the set of natural numbers, i.e. $\{1, 2, 3, 4, ...\}$
$\lfloor x \rfloor$	floor(x), the greatest integer that is $\leq x$
$g : A \to B$	g is a function with domain A and codomain B
$\mathbb{R}$	the set of all real numbers
$\mathbb{Z}$	the set of all integers
$\mathbb{C}$	the set of all complex numbers
$x \in A$	x is an element of the set A.
$C \subseteq D$	The set C is a subset of the set D.
$\exp(x)$	e^x
$\log(x)$	$\log_e(x) = \ln(x)$
$\mathbb{R}^2$	the set of ordered pairs (x, y) such that $x, y \in \mathbb{R}$
$\mathbb{N}$	the set of natural numbers, i.e. $\{1, 2, 3, 4, ...\}$
$\lfloor x \rfloor$	floor(x), the greatest integer that is $\leq x$
$\lceil x \rceil$	ceiling(x), the least integer that is $\geq x$.

2. What is rigorous math?

In my view, there are a couple of reasons why mathematicians learn to use rigorous mathematics to state problems and theorems, to give solutions to problems and proofs of theorems.

Reasons for doing mathematics rigorously:

(1) **Mathematicians want to be sure that their results are correct, so that mathematics is exact and precise.** One feature of mathematics that distinguishes the subject among all other disciplines is that if a result of mathematics is proved, then it is correct for all time. Results of mathematics proved 2000 years ago are still and will always be correct. This is not like other fields, where new information or better models can sometimes show that what was previously believed to be true is really not true. This is not to say that mathematicians don't make mistakes; there are in fact some interesting historical examples of this, but in those cases, it was discovered that assumptions made in the logic were not correct (i.e. there was a mistake in the proof). And another thing: opinion or prejudice do not change the truth of mathematical statements.

(2) **Mathematicians want to communicate to others to show why results are true.** Part of the function of a good proof is to allow mathematicians and students of mathematics to understand a complicated result in terms of much more rudimentary and believable results. The idea is to check a theorem in a precise way so that conceivably a computer could check it. So rigorous mathematics has an educational value in that it cuts through the mystery and allows us to connect more difficult concepts with simpler ones.

Keeping these things in mind, we need to make sure that when we state things and prove things, we do it exactly, and we don't allow any loopholes. And also, we want to make sure that the ways that we state things and prove things are done in the simplest and cleanest ways possible.

Things to think about when doing rigorous math:

(1) **When making statements:** Is there any way this could be misinterpreted? Are all my variables and symbols defined? Is the mathematical meaning of every word precise? Make sure that everything is written in complete sentences.
(2) **When proving and calculating:** In the back of your head, there is an evil person saying "I don't believe you!" after every sentence. Make sure that there is absolutely no doubt that everything you have written is true, and that the logical flow leaves no room for error. Make sure that everything is written in complete sentences.
(3) **After proof or calculation is complete:** Read over it again. Can anything be written more clearly? Is it easy to understand? Are there any loopholes?

To illustrate, let's do a couple of examples of correct and incorrect statements, proofs, calculations.

3. The quadratic formula

Example 3.1. ***(The Quadratic Formula) Derive the quadratic formula, solving the equation*** $ax^2+bx+c=0$.

If you were doing this in a high school algebra class, you would just go through some steps, which could look like this:

$$\begin{aligned}
ax^2+bx+c &= 0\\
x^2+\frac{b}{a}x+\frac{c}{a} &= 0\\
x^2+\frac{b}{a}x+\left(\frac{b}{2a}\right)^2+\frac{c}{a} &= \left(\frac{b}{2a}\right)^2\\
\left(x+\frac{b}{2a}\right)^2 &= \frac{b^2}{4a^2}-\frac{c}{a}=\frac{b^2-4ac}{4a^2}\\
x+\frac{b}{2a} &= \pm\frac{\sqrt{b^2-4ac}}{2a}\\
x &= -\frac{b}{2a}\pm\frac{\sqrt{b^2-4ac}}{2a}=\frac{-b\pm\sqrt{b^2-4ac}}{2a}.
\end{aligned}$$

However, to do a rigorous proof, we would need to state things clearly and to justify every step, all with complete sentences. Here is a rigorous version of the work above:

Statement: *If x is a complex number that satisfies the equation $ax^2+bx+c=0$, where $a,b,c\in\mathbb{R}$ and $a\neq 0$, then*

$$x=\frac{-b+\sqrt{b^2-4ac}}{2a} \text{ or } x=\frac{-b-\sqrt{b^2-4ac}}{2a}.$$

Proof. With the notation above, suppose that

$$ax^2 + bx + c = 0.$$

Then since $a \neq 0$, we may multiply both sides of the equation by $\frac{1}{a}$ to get another true equation. Using the distributive property,

$$x^2 + \frac{b}{a}x + \frac{c}{a} = 0.$$

Then we may add $\frac{b^2}{4a^2}$ to both sides to get another true equation, and by factoring we get

$$\begin{aligned} x^2 + \frac{b}{a}x + \frac{c}{a} + \frac{b^2}{4a^2} &= \frac{b^2}{4a^2}, \\ \left(x + \frac{b}{2a}\right)^2 + \frac{c}{a} &= \frac{b^2}{4a^2}. \end{aligned}$$

Next, we add $-\frac{c}{a}$ to both sides, simplify, and then take the square root:

$$\begin{aligned} \left(x + \frac{b}{2a}\right)^2 &= \frac{b^2}{4a^2} - \frac{c}{a} = \frac{b^2 - 4ac}{4a^2} \text{ implies} \\ x + \frac{b}{2a} &= \sqrt{\frac{b^2 - 4ac}{4a^2}} \text{ or } -\sqrt{\frac{b^2 - 4ac}{4a^2}}. \end{aligned}$$

Observe that

$$\sqrt{\frac{b^2 - 4ac}{4a^2}} = \frac{\sqrt{b^2 - 4ac}}{\sqrt{4a^2}} = \begin{cases} \frac{\sqrt{b^2-4ac}}{2a} & \text{if } a > 0 \\ -\frac{\sqrt{b^2-4ac}}{2a} & \text{if } a < 0 \end{cases},$$

so that

$$\begin{aligned} x + \frac{b}{2a} &= \sqrt{\frac{b^2 - 4ac}{4a^2}} \text{ or } -\sqrt{\frac{b^2 - 4ac}{4a^2}} \\ &= \begin{cases} \frac{\sqrt{b^2-4ac}}{2a} \text{ or } -\frac{\sqrt{b^2-4ac}}{2a} & \text{if } a > 0 \\ -\frac{\sqrt{b^2-4ac}}{2a} \text{ or } \frac{\sqrt{b^2-4ac}}{2a} & \text{if } a < 0 \end{cases} \\ &= \frac{\sqrt{b^2 - 4ac}}{2a} \text{ or } -\frac{\sqrt{b^2 - 4ac}}{2a}. \end{aligned}$$

Then, by adding $-\frac{b}{2a}$ to both sides of the equation in both cases, we get the final equations

$$\begin{aligned} x &= -\frac{b}{2a} + \frac{\sqrt{b^2 - 4ac}}{2a} = \frac{-b + \sqrt{b^2 - 4ac}}{2a} \text{ or} \\ x &= -\frac{b}{2a} - \frac{\sqrt{b^2 - 4ac}}{2a} = \frac{-b - \sqrt{b^2 - 4ac}}{2a}. \end{aligned}$$

□

After giving an exact proof of a result like this, it opens up many more questions to explore. For instance, is the statement actually "if and only if"? In other words, is the converse true? Another question: what do the numbers a, b, c tell us about the types of solutions we can have? Can we generalize this fact so that it applies to polynomials with complex number coefficients? This stream of thought is common when doing rigorous mathematics. When we really prove things, we really deeply understand why things are true — and this allows us to understand other problems that have not yet been addressed.

Let's first consider whether the statement has a true converse. Recall that the **converse** of a statement of the form "A implies B" ($A \Rightarrow B$) is the statement "B implies A" ($B \Rightarrow A$). If both $A \Rightarrow B$ and $B \Rightarrow A$ are true, we say that "A iff B" or "A if and only if B" or "$A \Leftrightarrow B$".

The converse of our statement above is

Statement: If $a, b, c \in \mathbb{R}$ and $a \neq 0$ and

$$x = \frac{-b + \sqrt{b^2 - 4ac}}{2a} \text{ or } x = \frac{-b - \sqrt{b^2 - 4ac}}{2a},$$

then $ax^2 + bx + c = 0$.

Note that "$a, b, c \in \mathbb{R}$ and $a \neq 0$" is background information on the context of the problem, so we did not consider that to really be part of the "if" part of the original statement. Also, note that we did not have to say that x is a complex number, because the formulas imply that x is a complex number.

The converse statement is actually true!

Proof. (of the converse statement) Suppose first that

$$x = \frac{-b + \sqrt{b^2 - 4ac}}{2a}.$$

Then, using algebra,

$$\begin{aligned}
ax^2 + bx + c &= a\left(\frac{-b + \sqrt{b^2 - 4ac}}{2a}\right)^2 + b\left(\frac{-b + \sqrt{b^2 - 4ac}}{2a}\right) + c \\
&= a\frac{\left(-b + \sqrt{b^2 - 4ac}\right)^2}{4a^2} + \frac{-b^2 + b\sqrt{b^2 - 4ac}}{2a} + c \\
&= \frac{b^2 - 2b\sqrt{b^2 - 4ac} + b^2 - 4ac}{4a} + \frac{-b^2 + b\sqrt{b^2 - 4ac}}{2a} + c \\
&= \frac{b^2 - 2b\sqrt{b^2 - 4ac} + b^2 - 4ac - 2b^2 + 2b\sqrt{b^2 - 4ac}}{4a} + c \\
&= \frac{-4ac}{4a} + c = -c + c = 0.
\end{aligned}$$

On the other hand, if

$$x = \frac{-b - \sqrt{b^2 - 4ac}}{2a},$$

we have similarly that

$$\begin{aligned}
ax^2 + bx + c &= a\left(\frac{-b - \sqrt{b^2 - 4ac}}{2a}\right)^2 + b\left(\frac{-b - \sqrt{b^2 - 4ac}}{2a}\right) + c \\
&= a\frac{\left(-b - \sqrt{b^2 - 4ac}\right)^2}{4a^2} + \frac{-b^2 - b\sqrt{b^2 - 4ac}}{2a} + c \\
&= \frac{b^2 + 2b\sqrt{b^2 - 4ac} + b^2 - 4ac}{4a} + \frac{-b^2 - b\sqrt{b^2 - 4ac}}{2a} + c \\
&= \frac{b^2 + 2b\sqrt{b^2 - 4ac} + b^2 - 4ac - 2b^2 - 2b\sqrt{b^2 - 4ac}}{4a} + c \\
&= \frac{-4ac}{4a} + c = -c + c = 0.
\end{aligned}$$

□

Some more interesting questions that come up from the proof of the quadratic formula:

(1) Does the quadratic formula generalize to formulas for the roots of higher degree polynomials?
Answer: Yes and no. There are formulas for the roots of a third degree polynomial and for the roots of a fourth degree polynomial involving rational expressions, powers, roots, etc. However, using Galois theory [in Abstract Algebra II], one can prove that no general formula like those can exist for polynomials of degree 5 and above. In fact, there are specific polynomials of degree 5 with integer coefficients whose complex roots cannot be written in terms of algebraic expressions involving powers, roots, fractions of rational numbers.

(2) What do the coefficients a, b, c tell you about the graph of $y = ax^2 + bx + c$, and does the quadratic formula give any insight into what the graph looks like?
Answer: Well a lot of things. If $a > 0$, then we have a parabola pointing upward (since

$$\lim_{x\to\pm\infty} \left(ax^2 + bx + c\right) = \lim_{x\to\pm\infty} ax^2 \left(1 + \frac{b}{x} + \frac{c}{x^2}\right) = \infty$$

in that case), and similarly the parabola points down if $a < 0$. The quadratic formula says that the roots (i.e. x-intercepts of $y = ax^2 + bx + c$) of $ax^2 + bx + c$ are

$$x = \frac{-b + \sqrt{b^2 - 4ac}}{2a}, \frac{-b - \sqrt{b^2 - 4ac}}{2a}.$$

If $b^2 - 4ac > 0$, then the two roots are real and distinct, and the parabola intersects the x-axis in those two places. The average of the two roots gives the line of symmetry of the parabola as the vertical line $x = \frac{-b}{2a}$. Note that $\frac{\sqrt{b^2-4ac}}{2|a|}$ is the distance from the line of symmetry to each of the roots $-\frac{b}{2a} + \frac{\sqrt{b^2-4ac}}{2|a|}$, $-\frac{b}{2a} - \frac{\sqrt{b^2-4ac}}{2|a|}$. If $b^2 - 4ac = 0$, both roots are equal, so the parabola hits the x-axis in one spot, at $x = -\frac{b}{2a}$. If $b^2 - 4ac < 0$, then the roots are not real numbers, so the parabola does not intersect the x-axis at all.

What about that last case? Is it still true that $x = \frac{-b}{2a}$ is the line of symmetry? Yes, because as we can see from calculus,

$$\frac{d}{dx}\left(ax^2 + bx + c\right) = 0$$

implies

$$2ax + b = 0, \text{ or } x = -\frac{b}{2a}.$$

Thus the critical point of the parabola occurs when $x = -\frac{b}{2a}$ in all cases, so in all cases this is the point of symmetry of the parabola. Of course, to verify symmetry rigorously, we need to show that if $f(x) = ax^2 + bx + c$ then for all real t, $f\left(-\frac{b}{2a} + t\right) = f\left(-\frac{b}{2a} - t\right)$. Let's

check:

$$\begin{aligned} f\left(-\frac{b}{2a}+t\right) &= a\left(-\frac{b}{2a}+t\right)^2+b\left(-\frac{b}{2a}+t\right)+c \\ &= a\left(\frac{b^2}{4a^2}-\frac{bt}{a}+t^2\right)-\frac{b^2}{2a}+tb+c \\ &= \frac{b^2}{4a}-bt+at^2-\frac{b^2}{2a}+tb+c=at^2-\frac{b^2}{4a}+c, \text{ and} \\ f\left(-\frac{b}{2a}-t\right) &= a\left(-\frac{b}{2a}-t\right)^2+b\left(-\frac{b}{2a}-t\right)+c=...= \\ &= at^2-\frac{b^2}{4a}+c=f\left(-\frac{b}{2a}+t\right).\checkmark \end{aligned}$$

In the case when $b^2-4ac<0$, can we say anything about the positioning of the parabola?

Yes; observe that from the calculation above with $t=0$,

$$f\left(-\frac{b}{2a}\right)=-\frac{b^2}{4a}+c=\frac{4ac-b^2}{4a},$$

so when the parabola does not intersect the x-axis, this number $\frac{4ac-b^2}{4a}=-\frac{b^2-4ac}{4a}$ gives the vertical position of the parabola's vertex.

(3) Is it possible to allow a,b,c to be any complex numbers? That is, do the formulas still work?

Answer: Very good question. We need to talk a little about complex numbers first and then will get back to this question.

4. Complex numbers

The set of complex numbers

$$\mathbb{C}=\{x+yi : x,y\in\mathbb{R}\}$$

is really a different way to describe the plane $\mathbb{R}^2$, but with the additional feature that you can multiply points together in the special way: for $a,b,c,d\in\mathbb{R}$,

$$\begin{aligned} i^2 &= -1,\ ia=ai, \\ (a+bi)(c+di) &= (ac-bd)+i(bc+ad). \end{aligned}$$

Notice that addition of complex numbers

$$(a+bi)+(c+di)=(a+c)+(b+d)i$$

can be visualized as ordinary vector addition in $\mathbb{R}^2$. Subtraction is defined in a similar way to addition.

Division is a little trickier. Is it really true that we can divide two complex numbers and then get another complex number? Observe that if $a,b,x,y\in\mathbb{R}$ with x,y not both 0, then

$$\begin{aligned} \frac{a+bi}{x+yi} &= \frac{a+bi}{x+yi}\left(\frac{x-yi}{x-yi}\right) \\ &= \frac{(a+bi)(x-yi)}{(x+yi)(x-yi)}=\frac{(ax+by)+(bx-ay)i}{x^2+y^2} \\ &= \frac{ax+by}{x^2+y^2}+\frac{bx-ay}{x^2+y^2}i, \end{aligned}$$

which by definition is in $\mathbb{C}$. We often use single variables refer to complex numbers, such as $z = 3+4i$; you should be able to tell from context or from what is given.

The set of complex numbers satisfies the same properties that the set of real numbers satisfies, namely these properties. These can be taken as a given set of axioms. Later in the course, if time allows, we will show how to derive these properties from the axioms of set theory and logic.

Properties of complex numbers

(1) (Closure for addition) For all $z, w \in \mathbb{C}$, $z + w \in \mathbb{C}$.
(2) (Associative property of addition) For all $z, w, v \in \mathbb{C}$, $(z + w) + v = z + (w + v)$.
(3) (Identity property for addition) For all $z \in \mathbb{C}$, $0 + z = z + 0 = z$.
(4) (Inverse property for addition) For all $z \in \mathbb{C}$, there exists a $w \in \mathbb{C}$ such that $z+w = w+z = 0$. [In fact, $w = -z$.]

Properties (1)-(4) imply that $(\mathbb{C}, +)$ is a **group**.

5. (Commutative property of addition) For all $z, w \in \mathbb{C}$, $z + w = w + z$.

Propertoes (1)-(5) imply that $(\mathbb{C}, +)$ is an **abelian group**.

6. (Closure for multiplication) For all $z, w \in \mathbb{C}$, $zw \in \mathbb{C}$.
7. (Associative property for multiplication) For all $z, w, v \in \mathbb{C}$, $(zw)\, v = z\,(wv)$.
8. (Distributive property #1) For all $z, w, v \in \mathbb{C}$, $z\,(w + v) = zw + zv$.
9. (Distributive property #2) For all $z, w, v \in \mathbb{C}$, $(z + w)\, v = zv + wv$.

Properties (1)-(9) imply that $(\mathbb{C}, +, \cdot)$ is a **ring**.

10. (Property of 1) There is an element $1 \in \mathbb{C}$ such that for all $z \in \mathbb{C}$, $1z = z$.

Properties (1)-(10) imply that $(\mathbb{C}, +, \cdot)$ is a **ring with unity** or **ring with** 1.

11. (Commutative property of multiplication) For all $z, w \in \mathbb{C}$, $zw = wz$.

Properties (1)-(9) and (11) imply that $(\mathbb{C}, +, \cdot)$ is a **commutative ring**.
Properties (1)-(11) imply that $(\mathbb{C}, +, \cdot)$ is a **commutative ring with** 1.

12. (No zero divisors) For any $z, w \in \mathbb{C}$, if $zw = 0$ then at least one of z and w must be 0.

Properties (1)-(9) and (11)-(12) imply that $(\mathbb{C}, +, \cdot)$ is an **integral domain**.

13. (Inverse property of multiplication) For any $z \in \mathbb{C} \setminus \{0\}$, there exists a $w \in \mathbb{C}$ such that $zw = wz = 1$. [In fact, $w = \frac{1}{z}$.]

Properties (1) through (13) imply that $(\mathbb{C}, +, \cdot)$ is a **field**.

Note that only properties of addition and multiplication are given. By definition, for all $z, w \in \mathbb{C}$

$$z - w = z + (-w)\,,$$

and also if $a, b \in \mathbb{C}$ and $b \neq 0$ then

$$\frac{a}{b} = a\left(\frac{1}{b}\right).$$

Thus, properties of subtraction and division reduce to properties of addition and multiplication, respectively.

Also, we emphasize that these properties of complex numbers certainly also hold for subsets of complex numbers, such as $\mathbb{R}$, $\mathbb{Z}$, etc., because these smaller sets consist of particular examples of complex numbers.

The conjugation operation has some nice properties. For $x, y \in \mathbb{C}$, the **conjugate of** $x + yi$ is defined as

$$\overline{x + yi} = x - yi.$$

For instance,

$$\overline{3-4.76i} = \overline{3+(-4.76)\,i} = 3-(-4.76)\,i = 3+4.76i.$$

Proposition 4.1. *Complex conjugation satisfies the following properties.*

(1) For all $z \in \mathbb{C}$, $\overline{(\overline{z})} = z$.
(2) For all $z, w \in \mathbb{C}$, $\overline{z+w} = \overline{z} + \overline{w}$.
(3) For all $z, w \in \mathbb{C}$, $\overline{z-w} = \overline{z} - \overline{w}$.
(4) For all $z, w \in \mathbb{C}$, $\overline{zw} = \overline{z}\,\overline{w}$.
(5) For all $z, w \in \mathbb{C}$, $\overline{\left(\frac{z}{w}\right)} = \frac{\overline{z}}{\overline{w}}$.
(6) For all $z \in \mathbb{C}$, $z\overline{z} \geq 0$ and $z\overline{z} = 0$ only if $z = 0$.

The proofs of these facts are exercises for the students.

For any $x, y \in \mathbb{R}$, the **real part** $\mathrm{Re}\,(z)$ of a complex number $z = x + iy$ is defined as

$$\mathrm{Re}\,(z) = x = \frac{z+\overline{z}}{2}.$$

The **imaginary part** $\mathrm{Im}\,(z)$ is defined as

$$\mathrm{Im}\,(z) = y = \frac{z-\overline{z}}{2i}.$$

The **absolute value** $|z|$ or **magnitude** of a complex number $z = x+iy$ (with $x, y \in \mathbb{R}$) is defined as

$$|z| = \sqrt{z\overline{z}} = \sqrt{x^2+y^2} = \sqrt{(\mathrm{Re}z)^2 + (\mathrm{Im}z)^2}.$$

Proposition 4.2. *The complex absolute value satisfies the following properties.*

(1) For all $z \in \mathbb{C}$, $|z| \geq 0$ and $|z| = 0$ only if $z = 0$.
(2) For all $z \in \mathbb{C}$, $|\overline{z}| = |z|$.
(3) For all $z, w \in \mathbb{C}$, $|zw| = |z|\,|w|$.
(4) (Triangle inequality) For all $z, w \in \mathbb{C}$, $|z+w| \leq |z| + |w|$, and equality occurs only when z and w lie on the same ray through the origin in $\mathbb{C}$.

A very useful description of complex numbers comes from the polar form of complex numbers. First of all, we will show how to define e^z for complex numbers z. From calculus, we know that the Taylor series for certain functions converge in certain intervals to the function itself. For example, for $x \in \mathbb{R}$,

$$\begin{aligned}
e^x &= 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + ... = \sum_{k=0}^{\infty} \frac{x^k}{k!}, \\
\cos(x) &= 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + ... = \sum_{k=0}^{\infty} \frac{(-1)^k\, x^{2k}}{(2k)!}, \\
\sin(x) &= x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + ... = \sum_{k=0}^{\infty} \frac{(-1)^k\, x^{2k+1}}{(2k+1)!},
\end{aligned}$$

and I emphasize that these series converge to the respective functions for **every** $x \in \mathbb{R}$. We now can actually extend the domains of these functions to allow any complex number inputs. In other words, a complex number can be substituted for x in the right hand sides of each of these functions, and it turns out that one can prove that the series converges. The answer we get is taken to be the meaning of the left hand side.

Now, it turns out that there are simpler expressions for e^{x+iy} for $x, y \in \mathbb{R}$. First, we note that

$$\begin{aligned} e^{iy} &= \sum_{k=0}^{\infty} \frac{(iy)^k}{k!} = \sum_{k=0}^{\infty} \frac{i^k y^k}{k!} \\ &= \sum_{m=0}^{\infty} \frac{i^{2m} y^{2m}}{(2m)!} + \sum_{r=0}^{\infty} \frac{i^{2r+1} y^{2r+1}}{(2r+1)!} \\ &= \sum_{m=0}^{\infty} \frac{\left(i^2\right)^m y^{2m}}{(2m)!} + \sum_{r=0}^{\infty} \frac{\left(i^2\right)^r i y^{2r+1}}{(2r+1)!} \\ &= \sum_{m=0}^{\infty} \frac{(-1)^m y^{2m}}{(2m)!} + \sum_{r=0}^{\infty} \frac{i\,(-1)^r y^{2r+1}}{(2r+1)!} \\ &= \cos(y) + i \sin(y). \end{aligned}$$

Thus, using the exponential rule $e^z e^w = e^{z+w}$, we have that

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

As a result of this calculation, we notice that $\cos(y) + i \sin(y)$ is the point on the unit circle in $\mathbb{C}$ corresponding to an angle y (arclength measured counterclockwise from the positive x-axis).

Now that we have this information about the exponential function, we can describe polar coordinates in $\mathbb{C} = \mathbb{R}^2$ in a very nice way. Recall that r is the distance from (x, y) to the origin and θ is the angle between the positive x-axis and the ray from $(0, 0)$ to (x, y). Thus, since $z = x + iy$ corresponds to (x, y), we have

$$x + iy = re^{i\theta}.$$

We have $|z| = r$ is the distance from 0 to z, and $\theta = \arg(z)$ is called the argument of z. Using this interpretation, we have a nice geometric picture of the multiplication of complex numbers. If $z = re^{i\theta}$ and $w = se^{i\varphi}$, then

$$\begin{aligned} zw &= \left(re^{i\theta}\right)\left(se^{i\varphi}\right) \\ &= rse^{i(\theta+\varphi)}. \end{aligned}$$

So that when we multiply z and w, the distances from the origin get multiplied as positive real numbers, and the arguments get added.

Using this information, it allows us to solve some equations. First, consider the equation

$$z^n = 1,$$

where n is a positive integer at least 2. Writing in terms of polar coordinates $z = re^{i\theta}$, we have that

$$r^n e^{in\theta} = 1,$$

so $r = 1$, and $n\theta$ is a multiple of 2π. Thus, the n numbers

$$z = e^{2k\pi i/n},\ k = 0, 1, 2, ..., n-1$$

are the distinct n complex numbers that satisfy $z^n = 1$. Note that $k = 0$ gives us the simple real solution $z = 1$. Thus, we can factor

$$z^n - 1 = (z-1)\left(z - e^{2\pi i/n}\right)\left(z - e^{4\pi i/n}\right) ... \left(z - e^{2(n-1)\pi i/n}\right).$$

Note that for a complex number z, expressions like $\sqrt{z}$ and $z^{1/3}$ are not well-defined, because there is not necessarily a "positive" or "real" square root or cube root. But we can say something like

this. If w is a single solution to $w^2 = z$ (in other words, **a** square root of z, then the two square roots of z are $\pm w$. Similarly, if p is a fixed n^{th} root of z, then

$$\left\{p, pe^{2\pi i/n}, pe^{4\pi i/n}, ..., pe^{2(n-1)\pi i/n}\right\}$$

are all the n^{th} roots of z. Why is this?

$$\left(pe^{2k\pi i/n}\right)^n = p^n e^{2k\pi i} = p^n = z.$$

Observe that these n roots are all on the same circle of radius $|p| = |z|^{1/n}$, and they are the vertices of a regular n-gon. And we can exactly calculate the positions of the vertices using the $e^{i\theta} = \cos\theta + i\sin\theta$ formula.

Now back to the quadratic formula. Using the information we now know about complex numbers, we can now solve the quadratic equation $ax^2+bx+c = 0$ when a, b, c are arbitrary complex numbers when $a \neq 0$. The same algebra works, except we show replace $\sqrt{b^2 - 4ac}$ with w, one of the two square roots of $b^2 - 4ac$. So the solution is

$$x = \frac{-b + w}{2a} \text{ or } x = \frac{-b - w}{2a},$$

where w is one fixed solution to $w^2 = b^2 - 4ac$.

An important fact about polynomial equations and the complex numbers is the fundamental theorem of algebra.

Theorem 4.3. *(**Fundamental Theorem of Algebra**) Let $p(x) = a_n z^n + a_{n-1}z^{n-1} + ... + a_0$ be a polynomial such that $a_n \neq 0$ and $a_j \in \mathbb{C}$ for all j. Then*

$$p(z) = a_n (z - z_1)(z - z_2)...(z - z_n)$$

for some complex numbers $z_1, ..., z_n$. That is, $p(x)$ has exactly n roots in $\mathbb{C}$, counting multiplicities.

5. Rational and irrational numbers

Let's start with some facts about integers.

Here we will often use a standard proving trick called "Proof by Contradiction." The way this works is this:

To prove $A \Rightarrow B$:

- Assume A is true.
- Then assume that (not B) is true.
- Proceed through logical steps to show that these assumptions reach a false conclusion (called the "contradiction")
- Therefore, the hypothesis that (not B) is true must be false, so in fact B is true. Q.E.D.

Proof by contradiction is a great technique, but you should not overuse it. A rule of thumb is that it should only be used if it makes the proof shorter. And sometimes it is the only way to prove something.

Notation: if $a, b \in \mathbb{Z}$, we write $a|b$ to mean that " a divides b", i.e. there exists another integer k such that $ak = b$.

An important theorem concerning the integers is the Fundamental Theorem of Arithmetic.

A **prime number** is a positive integer greater than 1 such that its only divisors are 1 and itself.

Theorem 5.1. *(**Fundamental Theorem of Arithmetic**) Every positive integer $x > 1$ can be written as*

$$x = p_1^{r_1} p_2^{r_2} ... p_k^{r_k},$$

where $p_1 < p_2 < ... < p_k$ are prime numbers and $r_1, ..., r_k$ are positive integers. The expression above is unique.

Proof. Uses induction. Postponed..... □

Definition 5.2. *Let $x, y \in \mathbb{Z}$ such that not both of x and y are 0. Then the* ***greatest common divisor*** $\gcd(x, y)$ *is the greatest positive integer g such that $g|x$ and $g|y$.*

For examples, $\gcd(-24, 36) = 12$, $\gcd(0, -34) = 34$, $\gcd(-13, 7) = 1$. We say that integers c and d are **relatively prime** if $\gcd(c, d) = 1$.

Definition 5.3. *Let x, y be positive integers. The* ***least common multiple*** $\operatorname{lcm}(x, y)$ *is the least positive integer L such that $x|L$ and $y|L$.*

Proposition 5.4. *If $x = p_1^{a_1} p_2^{a_2} ... p_k^{a_k}$ and $y = p_1^{b_1} p_2^{b_2} ... p_k^{b_k}$, where $p_1 < ... < p_k$ are primes and $a_j \geq 0$, $b_j \geq 0$ for all j. Then*

$$\begin{aligned} \gcd(x, y) &= p_1^{\min\{a_1, b_1\}} p_2^{\min\{a_2, b_2\}} ... p_k^{\min\{a_k, b_k\}}, \\ \operatorname{lcm}(x, y) &= p_1^{\max\{a_1, b_1\}} p_2^{\max\{a_2, b_2\}} ... p_k^{\max\{a_k, b_k\}}. \end{aligned}$$

Remark 5.5. *This proposition can be used to find the* gcd *and* lcm *of any two positive numbers, since $\{p_1, ..., p_k\}$ can be chosen to be the union of all prime numbers dividing either x or y.*

Definition 5.6.

Proposition 5.7. *If x, y are positive integers, then $xy = \operatorname{lcm}(x, y) \gcd(x, y)$.*

Proof. Follows from the previous proposition... □

Lemma 5.8. *If x, y are both integers and are not both 0, and if $a \in \mathbb{Z}$ satisfies*

$$a|x \text{ and } a|y,$$

then $a| \gcd(x, y)$.

Proof. Follows from previous two propositions... □

Proposition 5.9. *(**Division algorithm**) Let $x, y \in \mathbb{Z}$ and let $y > 0$. Then there exists a unique integer q and a unique integer r such that $0 \leq r < y$ such that*

$$x = qy + r.$$

Proof. Let r be the smallest nonnegative integer in the set $A = \{x - by : b \in \mathbb{Z}\}$, and let q be the specific b such that $r = x - by$. Then we have $x = qy + r$ and just have to show that $r < y$. Suppose instead that $r \geq y$. Then

$$r' = r - y = x - (q + 1) y \geq 0$$

is smaller than r and is nonnegative and is in A, so this is a contradiction to our choice of r. Thus $r < y$ and we are done. □

Proposition 5.10. *(**Bezout's Identity**) Let $x, y \in \mathbb{Z}$, with x and y are not both zero. Let $S = \{rx + sy : r, s \in \mathbb{Z}\}$. Then the smallest positive integer in S is $\gcd(x, y)$. In particular, that means that there exist integers a, b such that $ax + by = \gcd(x, y)$.*

Proof. Let $p = ax + by$ be the smallest positive element of S, with $a, b \in \mathbb{Z}$. Then it is clear that $d = \gcd(x, y)$ is a factor of p (i.e. $d|p$), since it is a factor of both x and y. Next, we use the division algorithm to divide x by p:

$$\begin{aligned} x &= pq + r \\ &= (ax + by)\, q + r, \end{aligned}$$

where $0 \leq r < p$. But then

$$\begin{aligned} r &= x - (ax + by)\, q \\ &= (1 - aq)\, x + (-bq)\, y \end{aligned}$$

is an integer linear combination of x and y, so since p is the smallest postive integer like that, $r = 0$. Thus, $p|x$. Similarly, $p|y$, so that p is a common factor of x and y. But then $p|d$. So, since $p|d$ and $d|p$ and d, p are positive integers, $d = p$. □

A **rational number** x is a real number that can be expressed as $x = \frac{p}{q}$, where $p, q \in \mathbb{Z}$ and $q \neq 0$. An **irrational number** is a real number that is not a rational number.

Lemma 5.11. *Every rational number x can be expressed as $x = \frac{p}{q}$, where $p, q \in \mathbb{Z}$ and $q \neq 0$, and where p and q are relatively prime.*

Proof. Suppose that p and q have a greatest common factor k, so that $p = km$ and $q = ks$ for some positive integer $k > 1$ and integers m and s. Then

$$\frac{p}{q} = \frac{km}{ks} = \frac{m}{s}.$$

If m and s have a common factor $a > 1$, then $m = au$ and $s = av$ for some integers u and v. But then $p = kau$ and $q = kav$, so that ka is a positive common factor of p and q, and that is a contradiction to the fact that $k = \gcd(p, q)$. □

Proposition 5.12. *The number $\sqrt{2}$ is irrational.*

Proof. Suppose instead that $\sqrt{2}$ is rational, so that $\sqrt{2} = \frac{p}{q}$, where $p, q \in \mathbb{Z}$ and $q \neq 0$, and where p and q have no common factors other than 1 or -1 (by the previous lemma). Then $2 = \frac{p^2}{q^2}$, so $2q^2 = p^2$. Since 2 is a prime number and p^2 is a multiple of 2, since p^2 has the same individual prime factors as p does, it must be true that p is a multiple of 2, say $p = 2k$ for some $k \in \mathbb{Z}$. Then our equation reads $2q^2 = (2k)^2 = 4k^2$, so in fact $q^2 = 2k^2$. But then, using similar reasoning, we conclude that q is a multiple of 2, so that p and q have a common factor of 2. But this is a contradiction to the fact that p and q have no common factors other than 1 or -1. Therefore, the assumption that $\sqrt{2}$ is rational is wrong, so that $\sqrt{2}$ must be an irrational number. □

Some questions about the rational and irrational numbers come up. How common are they? Are there the same amount of each? What are their decimal expansions?

Let's talk about their decimal expansions first. A **positive decimal expansion** is a series of the form

$$\begin{aligned} x &= \sum_{j=-\infty}^{k} a_j 10^j \\ &= a_k 10^k + a_{k-1} 10^{k-1} + ... + a_0 10^0 + a_{-1} 10^{-1} + ... \\ &= a_k a_{k-1} ... a_0 . a_{-1} a_{-2} ..., \end{aligned}$$

where a_j are integers such that $0 \leq a_j \leq 9$ for $j < k$ and $1 \leq a_k \leq 9$. Every positive real number x has such an expansion. For example,

$$\begin{aligned} & x = 34.5269489034120... \\ = \ & 3 \cdot 10^1 + 4 \cdot 10^0 + 5 \cdot 10^{-1} + 2 \cdot 10^{-2} + 6 \cdot 10^{-3} + ... \end{aligned}$$

Another way to conceive of a real number is that it is a limit of an increasing sequence of rational numbers. For example, x above is the limit of the sequence

$$(30, 34, 34.5, 34.52, 34.526, 34.5269, 34.52694, ...) .$$

We may ask: can we tell from the decimal expansion of a real number if it is a rational number?

Someone who plays with calculators a lot might be able to guess the answer. We say that a decimal expansion of a number x is **eventually repeating** if x can be written as a decimal expansion of the form

$$x = x_1 x_2 x_3 ... x_N \overline{x_{N+1} x_{N+2} ... x_{N+p}},$$

where first of all we have not described where the decimal point is (but it is somewhere), p is a natural number, and $\overline{x_{N+1} x_{N+2} ... x_{N+p}}$ means

$$\begin{aligned} & \overline{x_{N+1} x_{N+2} ... x_{N+p}} \\ = \ & x_{N+1} x_{N+2} ... x_{N+p} \ x_{N+1} x_{N+2} ... x_{N+p} \ x_{N+1} x_{N+2} ... x_{N+p} \ ... \end{aligned}$$

— that is, a sequence of p numbers repeated indefinitely. In all cases, x_j is a positive integer between 0 and 9.

Theorem 5.13. *If $x \in \mathbb{R}$ has a decimal expansion that is eventually repeating, then x is a rational number.*

The converse is also true!

Theorem 5.14. *If $x \in \mathbb{R}$ is a rational number, then x has a decimal expansion that is eventually repeating.*

We will now prove the first theorem, and we postpone the proof of the second theorem until after we have discussed modular arithmetic later in the course.

Proof. (Proof of Theorem 5.13)Suppose that $x \in \mathbb{R}$ has an eventually repeating decimal expansion. Then there exists $k \in \mathbb{Z}$ and a positive integer p such that

$$x = 10^k \times x_1 x_2 x_3 ... x_N \ . \ \overline{x_{N+1} x_{N+2} ... x_{N+p}};$$

in other words,

$$k = \begin{cases} 0 & \text{if the decimal point occurs just before} \\ & \text{the beginning of the repeating part, as above} \\ m & \text{if the decimal point occurs } m \text{ positions to the} \\ & \text{left of the beginning of the repeating part} \\ -h & \text{if the decimal point occurs } h \text{ positions to the} \\ & \text{right of the beginning of the repeating part} \end{cases}$$

Then

$$\begin{aligned} 10^p x &= 10^k \times x_1 x_2 x_3 ... x_N x_{N+1} x_{N+2} ... x_{N+p} \ . \ \overline{x_{N+1} x_{N+2} ... x_{N+p}}, \\ x &= 10^k \times x_1 x_2 x_3 ... x_N \ . \ \overline{x_{N+1} x_{N+2} ... x_{N+p}}, \end{aligned}$$

so if we subtract, we get

$$10^p x - x = 10^k \times x_1 x_2 x_3 ... x_N x_{N+1} x_{N+2} ... x_{N+p} - 10^k \times x_1 x_2 x_3 ... x_N.$$

Then

$$x\left(10^p - 1\right) = 10^k \times x_1 x_2 x_3 ... x_N x_{N+1} x_{N+2} ... x_{N+p} - 10^k \times x_1 x_2 x_3 ... x_N,$$

so

$$x = \frac{10^k \times x_1 x_2 x_3 ... x_N x_{N+1} x_{N+2} ... x_{N+p} - 10^k \times x_1 x_2 x_3 ... x_N}{10^p - 1},$$

which is a fraction of integers (if k is negative, we simply put 10^k in the denominator), and $10^p - 1 \neq 0$ since $p \geq 1$. Thus, x is rational. □

Remark 5.15. *We have mentioned that the converse is true, but note that this means that any fraction of integers can be written as a fraction with* 99...900...0 *in the denominator. This implies the result that every positive integer is a factor of some integer of the type*

$$99...900...0.$$

This is a bit surprising! In fact, if we trace through the argument above backwards, we realize that if we prove this result, we can make any rational number into an eventually repeating decimal.

We have now a better handle on what rational and irrational numbers are. A couple of questions come up:

(1) Are there more rational numbers than irrational numbers, or vice versa?
Answer: we will examine this question when we explore measures of sizes of infinite sets, in the next section.
(2) Is there a rational number between any two real numbers (including possibly two rationals, one rational and the other irrational, or the case of two irrationals)?
Answer: Yes!
(3) Is there an irrational number between any two real numbers (including the possibilities mentioned above)?
Answer: Yes!

We will now explain why the answers to (2) and (3) above are "Yes". Related to this question is the **Archimedean Property** of the real numbers.

Axiom 5.16. *(**Archimedean Property**) Given any two positive real numbers x and y, there is a natural number n such that $nx > y$. [In particular, this is true if $x = 1$; there exists a natural number n such that $n > y$.]*

This seems really obvious! Thus, it is a very believable axiom.
We now can prove (2) and (3) above.

Proposition 5.17. *Suppose that $x, y \in \mathbb{R}$ such that $x < y$. Then there exists a rational number r such that $x < r < y$.*

Proof. With $x, y \in \mathbb{R}$ such that $x < y$, we see that $\frac{2}{y-x} > 0$. Then, by the Archimedean Property, there exists a (positive) natural number q such that $q > \frac{2}{y-x}$. Then, taking reciprocals, $\frac{1}{q} < \frac{y-x}{2}$.

Now, let p be the least integer so that $x \leq \frac{p-1}{q}$; in other words, let $p-1 = \lceil qx \rceil$. Then

$$\begin{aligned} x &\leq \frac{p-1}{q} < \frac{p}{q} = x + \left(\frac{p}{q} - x\right) \\ &< x + \frac{2}{q} = x + 2\left(\frac{1}{q}\right) \\ &< x + 2\left(\frac{y-x}{2}\right) = y. \end{aligned}$$

Then $\frac{p}{q}$ is a rational number such that $x < \frac{p}{q} < y$. □

To prove the next part, observe that we already proved that $\sqrt{2}$ is irrational, and also we know that $\sqrt{2} > 1$.

Lemma 5.18. *For any nonzero integers p, q, $\frac{p}{q\sqrt{2}}$ is irrational.*

Proof. Suppose instead that $\frac{p}{q\sqrt{2}}$ is rational, so that $\frac{p}{q\sqrt{2}} = \frac{c}{d}$ for some integers c, d such that $d \neq 0$. Also, $c \neq 0$ since $p \neq 0$. Then, multiplying by $qd\sqrt{2}$, $pd = qc\sqrt{2}$, so that $\sqrt{2} = \frac{pd}{qc}$. But $pd, qc \in \mathbb{Z}$ and $qc \neq 0$, so this shows $\sqrt{2}$ is rational, a contradiction. Thus $\frac{p}{q\sqrt{2}}$ is irrational. □

And now, just a little tweak of the previous proposition.

Proposition 5.19. *Suppose that $x, y \in \mathbb{R}$ such that $x < y$. Then there exists a irrational number u such that $x < u < y$.*

Proof. With $x, y \in \mathbb{R}$ such that $x < y$, let $\frac{\sqrt{2}}{y-x} > 0$. Then, by the Archimedean Property, there exists a (positive) natural number q such that $q > \frac{\sqrt{2}}{y-x}$. Then, taking reciprocals, $\frac{1}{q} < \frac{y-x}{\sqrt{2}}$, so $\frac{1}{q\sqrt{2}} < \frac{y-x}{2}$ Now, let p be the least integer so that $x \leq \frac{p-1}{q\sqrt{2}}$; in other words, let $p-1 = \lceil qx\sqrt{2} \rceil$. Then

$$\begin{aligned} x &\leq \frac{p-1}{q\sqrt{2}} < \frac{p}{q\sqrt{2}} = x + \left(\frac{p}{q\sqrt{2}} - x\right) \\ &< x + \frac{2}{q\sqrt{2}} = x + 2\left(\frac{1}{q\sqrt{2}}\right) \\ &< x + 2\left(\frac{y-x}{2}\right) = y. \end{aligned}$$

Then $u = \frac{p}{q\sqrt{2}}$ is a rational number such that $x < u < y$. □

Given these two propositions above, one might conclude that if one were to measure the size of infinities, the infinite number of rational numbers should be the same as the infinite number of irrational numbers. As we will soon find out, um, not really.

6. Set Theory and Cardinality

We now wish to have a way of measuring the size of sets, even if they are infinite. Before doing this, we will review some definitions from set theory.

A **set** is a collection of things, called **elements** of the set. Sets are sometimes described using set notation:

$$A = \{n \in \mathbb{Z} : n = 3k \text{ for some } k \in \mathbb{Z}\}$$

means

$$A \text{ is the set of all integers } n \text{ such that } n = 3k \text{ for some integer } k.$$

We write $6 \in A$ to mean "6 is an element of A."

If A and B are sets, we say A is a **subset** of B and write $A \subseteq B$ if the following condition is met:

$$\text{If } x \in A, \text{ then } x \in B.$$

We say $A = B$ if $A \subseteq B$ and $B \subseteq A$; in other words, A and B are equal if they have the same set of elements. Note that

$$\{1, 2, 5\} = \{5, 2, 1\},$$

because the order of listing elements is not important; sets are unordered by definition.

A **function** $f : A \to B$ between two sets A and B is the assignment to each element $x \in A$ a **single** element $f(x) \in B$. The set A is called the **domain** of f, and the set B is called the **codomain** of f. Functions are sometimes called **maps**. The set

$$\text{Im}(f) = Range(f) = f(A) = \{f(x) : x \in A\}$$

is known as the **image of** f or as the **range of** f. Notice that $f(A) \subseteq B$.

A function $f : A \to B$ is called **1-1** (one-to-one) [other terms: **injective**, **monic**] if

$$f(x) = f(y) \text{ implies } x = y.$$

That is, if $x \neq y$, then $f(x) \neq f(y)$. Another way to say it: a function is $1-1$ if each element of the range comes from exactly one element of the domain.

A function $f : A \to B$ is called **onto** [other terms: **surjective**, **epic**] if for every $y \in B$, there exists an $x \in A$ such that $f(x) = y$. In other words, f is onto if $f(A) = B$.

A function $f : A \to B$ is called a **one-to-one correspondence** [other term: **bijective**] if f is both 1-1 and onto. Note that if $g : A \to B$ is bijective, there exists an **inverse function** $g^{-1} : B \to A$, defined by

$$g^{-1}(y) = \text{the unique } x \text{ such that } g(x) = y.$$

Lemma 6.1. *If $g : A \to B$ is bijective, then $g^{-1} : B \to A$ is bijective.*

Proof. (left as an exercise) □

Note that the inverse notation is also used to denote the **inverse image of a set**. Given a function $\alpha : C \to D$ and a subset $S \subseteq D$, the **inverse image of** S is

$$\alpha^{-1}(S) = \{c \in C : \alpha(c) \in S\}.$$

We now give the definition of the **cardinality** of a set, a measure of the "number of elements" in a set, which can be used even if the set is infinite.

Definition 6.2. *The **cardinality** of a set S is denoted $|S|$ or $card(S)$ or $\#(S)$. We have the following definitions:*

1. *If S and T are two sets, we say that $|S| = |T|$ if there exists a bijection $F : S \to T$.*
2. *If S and T are two sets, we say that $|S| \leq |T|$ if there exists a 1-1 function $G : S \to T$.*
3. *If S and T are two sets, we say that $|S| < |T|$ if there exists a 1-1 function $G : S \to T$ and there does not exist a bijection from S to T.*
4. *If S and T are two sets, we say that $|S| \geq |T|$ if there exists an onto function $H : S \to T$.*
5. *If S and T are two sets, we say that $|S| > |T|$ if there exists an onto function $H : S \to T$ and there does not exist a bijection from S to T.*

There are many issues that come up in the definition. Lots of things to investigate.

Proposition 6.3. *Equal cardinality is an equivalence relation. That is,*

(1) (Reflexivity) If A is a set, then $|A| = |A|$.
(2) (Symmetry) If A and B are two sets such that $|A| = |B|$, then $|B| = |A|$.
(3) (Transitivity) If A, B, and C are sets and $|A| = |B|$ and $|B| = |C|$, then $|A| = |C|$.

Proof. (1) If A is any set, let $I : A \to A$ be the function defined by $I(x) = x$ for all $x \in A$. Then I is onto, since for all $y \in A$, we can see that $I(y) = y$. Also, I is 1-1, since if $I(y) = I(x)$, then $y = x$. Thus I is bijective, so $|A| = |A|$.
(2) If A and B are two sets such that $|A| = |B|$, then there exists a bijective function $F : A \to B$. Then also $F^{-1} : B \to A$ is bijective, so $|B| = |A|$.
(3) If A, B, and C are sets and $|A| = |B|$ and $|B| = |C|$, then there exist bijective functions $f : A \to B$ and $g : B \to C$. Then consider $g \circ f : A \to C$: If y is any element of C, then there exists $z \in B$ such that $g(z) = y$ since g is onto. Then also there exists $x \in A$ such that $f(x) = z$ since f is onto. But then $(g \circ f)(x) = g(f(x)) = g(z) = y$, so $g \circ f$ is onto. Next, suppose $(g \circ f)(s) = (g \circ f)(t)$ for some $s, t \in A$. Then $g(f(s)) = g(f(t))$, so $f(s) = f(t)$ since g is 1-1. Then $s = t$ since f is 1-1. Thus, $(g \circ f)$ is a bijection, so $|A| = |C|$. □

Proposition 6.4. *If A and B are sets such that $A \subseteq B$, then $|A| \leq |B|$ and $|B| \geq |A|$.*

Proof. (Left as an exercise) □

Proposition 6.5. *If A is a set, then $|A| \leq |A|$ and $|A| \geq |A|$.*

Proof. (Left as an exercise) □

Some of the next few results need to use the **axiom of choice**. This is a fundamental set theory property that must be assumed in order to prove certain things.

Axiom 6.6. ***(Axiom of Choice)*** *Let $\{S_\alpha\}_{\alpha \in A}$ be a collection of nonempty sets, where A is the index set. Then there exists a function $f : \{S_\alpha\} \to \bigcup_\alpha S_\alpha$ such that $f(S_\alpha) =$ a single element of S_α. [That is, $f(S_\alpha) \in S_\alpha$.] In other words, we can **choose** one element from each set S_α.*

This axiom has an interesting history, and some quite shocking theorems, can be derived from it, like the "Banach-Tarski Paradox".

Proposition 6.7. *If A and B are two sets, then $|A| \leq |B|$ if and only if $|B| \geq |A|$.*

Proof. ($\Rightarrow$) Suppose that A and B are two sets, and $|A| \leq |B|$. Then, there exists a 1-1 function $f : A \to B$. Then let a_0 be one particular element of A, and we define the function $g : B \to A$ by

$$g(y) = \begin{cases} x & \text{if } y \in \text{Im}(f) \text{ and } f(x) = y \\ a_0 & \text{otherwise} \end{cases}.$$

Then we see that g is onto, because in particular g maps $\text{Im}(f)$ to the domain of f, i.e. all of A. Thus, $|B| \geq |A|$.
($\Leftarrow$) Suppose that A and B are two sets, and $|A| \geq |B|$. Then, there exists an onto function $p : A \to B$. We now define a function $q : B \to A$ by choosing for each $b \in B$ a single element $q(b)$ of the set $p^{-1}(\{b\})$ (here we use the axiom of choice). Then if $q(b_1) = q(b_2)$ for some $b_1, b_2 \in B$, then $p(q(b_1)) = p(q(b_2))$, so by the way q is defined, $p(q(x)) = x$ for all $x \in B$, so $b_1 = b_2$. Thus, q is 1-1. Then, $|B| \leq |A|$. □

Proposition 6.8. *If A, B, C are sets such that $|A| \leq |B|$ and $|B| \leq |C|$, then $|A| \leq |C|$.*

Proof. (Left as an exercise) □

Proposition 6.9. *If A, B, C are sets such that $|A| \geq |B|$ and $|B| \geq |C|$, then $|A| \geq |C|$.*

Proof. (Left as an exercise) □

Theorem 6.10. ***(Schröder-Bernstein Theorem)*** *If A and B are twos sets, and if $|A| \leq |B|$ and $|B| \leq |A|$, then $|A| = |B|$.*

Proof. See Wikipedia or another source. Some proofs use the axiom of choice, but others do not. □

Corollary 6.11. *If A and B are two sets, and if $|A| \geq |B|$ and $|B| \geq |A|$, then $|A| = |B|$.*

Proof. (Left as exercise; just use one of the previous propositions and the Schröder-Bernstein Theorem) □

We now examine some examples of computing cardinality. Recall that the set of natural numbers is

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

Definition 6.12. *A set S is called **countable** if $|S| = |\mathbb{N}|$. A set T is called **uncountable** if $|T| = |\mathbb{N}|$.*

Example 6.13. *Let $U = \{20p + 1 : p \in \mathbb{N}\} = \{21, 41, 61, ...\}$.*

Observe that $f : U \to \mathbb{N}$ defined by $f(x)$ is 1-1, so $|U| \leq |\mathbb{N}|$. but also, observe that $g : \mathbb{N} \to U$ defined by

$$g(x) = 20x + 1$$

is 1-1, because if $g(x) = g(y)$, then $20x + 1 = 20y + 1$, so $20x = 20y$, so $x = y$. Also, g is onto, because every element of U can be written as $g(p)$ for some $p \in \mathbb{N}$. Thus, $|U| = |\mathbb{N}|$, so U is countable.

Example 6.14. *Consider the set $\mathbb{Z} = \{..., -3, -2, -1, 0, 1, 2, 3, 4, ...\}$ of all integers.*

Observe that there exists a function $I : \mathbb{N} \to \mathbb{Z}$ given by $I(x) = x$ that is 1-1. Thus $|\mathbb{N}| \leq |\mathbb{Z}|$. On the other hand,we can order the set of integers like this:

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

so we could make a bijection $A : \mathbb{N} \to \mathbb{Z}$ by

$$A(1) = 0, A(2) = 1, A(3) = -1, A(4) = 2, A(5) = -2, A(6) = 3, A(7) = -3, ...$$

i.e. $A(n) = n^{th}$ item on the list above. Therefore, $|\mathbb{Z}| = |\mathbb{N}|$, so $\mathbb{Z}$ is also countable.

The previous example demonstrates that a set S is countable if and only if it can be arranged in a list

$$\{s_1, s_2, s_3, ...\}.$$

The bijection $f : \mathbb{N} \to S$ is $f(j) = s_j$ for all j, can be obtained from the list. Conversely, if there does exist such a bijection, the set can be arranged in the list $\{f(1), f(2), f(3), ...\}$.

Consider the set $\mathbb{Q}$ of all rational numbers. Since $\mathbb{N} \subseteq \mathbb{Q}$, the $f(x) = x$ gives an injective map so that we know $|\mathbb{N}| \leq |\mathbb{Q}|$. Is it true that $|\mathbb{N}| < |\mathbb{Q}|$ or that $|\mathbb{N}| = |\mathbb{Q}|$? It turns out that we can make a list of all rational numbers. Here is how. First, we make a two-dimensional grid that contains all the rational numbers, with numerators (N) fixed in each column and denominators (D) fixed in each row:

D \ N	0	1	−1	2	−2	3	−3	...
1	$\frac{0}{1}$	$\frac{1}{1}$	$\frac{-1}{1}$	$\frac{2}{1}$	$\frac{-2}{1}$	$\frac{3}{1}$	$\frac{-3}{1}$	...
2	$\frac{0}{2}$	$\frac{1}{2}$	$\frac{-1}{2}$	$\frac{2}{2}$	$\frac{-2}{2}$	$\frac{3}{2}$	$\frac{-3}{2}$	...
3	$\frac{0}{3}$	$\frac{1}{3}$	$\frac{-1}{3}$	$\frac{2}{3}$	$\frac{-2}{3}$	$\frac{3}{3}$	$\frac{-3}{3}$	...
4	$\frac{0}{4}$	$\frac{1}{4}$	$\frac{-1}{4}$	$\frac{2}{4}$	$\frac{-2}{4}$	$\frac{3}{4}$	$\frac{-3}{4}$	...
5	$\frac{0}{5}$	$\frac{1}{5}$	$\frac{-1}{5}$	$\frac{2}{5}$	$\frac{-2}{5}$	$\frac{3}{5}$	$\frac{-3}{5}$	...
6	$\frac{0}{6}$	$\frac{1}{6}$	$\frac{-1}{6}$	$\frac{2}{6}$	$\frac{-2}{6}$	$\frac{3}{6}$	$\frac{-3}{6}$	...
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

We see that every possible rational number is on this infinite grid, although there are many repeats. Also, we can make an infinite list of all the items on the grid by setting up the following sequence (followed alphabetically):

D \ N	0	1	−1	2	−2	3	−3	...
1	A	C	F	J	O	$\cdots$	$\frac{-3}{1}$	...
2	B	E	I	N	T	$\frac{3}{2}$	$\frac{-3}{2}$	...
3	D	H	M	S	$\frac{-2}{3}$	$\frac{3}{3}$	$\frac{-3}{3}$	...
4	G	L	R	$\frac{2}{4}$	$\frac{-2}{4}$	$\frac{3}{4}$	$\frac{-3}{4}$	...
5	K	Q	$\frac{-1}{5}$	$\frac{2}{5}$	$\frac{-2}{5}$	$\frac{3}{5}$	$\frac{-3}{5}$	...
6	P	$\frac{1}{6}$	$\frac{-1}{6}$	$\frac{2}{6}$	$\frac{-2}{6}$	$\frac{3}{6}$	$\frac{-3}{6}$	...
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

We may then list off all the rational numbers by using the sequence above but then at each position deleting the entry if it has already appeared in the list. Thus, we have

$$\mathbb{Q} = \left\{ \frac{0}{1}, \frac{0}{\not 2}, \frac{1}{1}, \frac{0}{\not 3}, \frac{1}{2}, \frac{-1}{1}, \frac{0}{\not 4}, \frac{1}{3}, \frac{-1}{2}, \frac{2}{1}, \frac{0}{\not 5}, \frac{1}{4}, \frac{-1}{3}, \frac{2}{\not 2}, \ldots \right\}.$$

Therefore, $|\mathbb{Q}| = |\mathbb{N}|$, and $\mathbb{Q}$ is countable!

An interesting consequence of this is the following.

Proposition 6.15. *For any positive number $\varepsilon > 0$, there exists a collections of intervals $\{I_j\}_{j=1}^{\infty}$ such that their total length is $\leq \varepsilon$, and such that*

$$\mathbb{Q} \subseteq \bigcup_{j=1}^{\infty} I_j.$$

Proof. Let

$$\mathbb{Q} = \{q_1, q_2, q_3, ...\}$$

be the list of all rational numbers. Let $I_1 = \left(q_1 - \frac{\varepsilon}{2^2}, q_1 + \frac{\varepsilon}{2^2}\right)$, which has total length $\frac{\varepsilon}{2}$. Similarly, let

$$I_n = \left(q_n - \frac{\varepsilon}{2^{n+1}}, q_n + \frac{\varepsilon}{2^{n+1}}\right),$$

of length $\frac{\varepsilon}{2^n}$. Then the total length of the intervals is at most the geometric series

$$\sum_{n=1}^{\infty} \frac{\varepsilon}{2^n} = \frac{\varepsilon/2}{1 - \frac{1}{2}} = \varepsilon.$$

□

This is a bit shocking!

At this point, you may ask yourself, is **every** infinite set countable? G. Cantor proved that $\mathbb{R}$ is actually uncountable, using the famous diagonalization argument. But first, we must think about what we consider to be a real number. We will consider a real number to be a number that has a decimal expansion. (Recall that rational numbers are exactly those real numbers whose decimal expansions are eventually repeating.) Really, a decimal expansion is an infinite series. For example, the number

$$x = 7890.437289041970...$$

means

$$x = 7 \cdot 10^3 + 8 \cdot 10^2 + 9 \cdot 10^1 + 0 \cdot 10^0 + 4 \cdot 10^{-1} + 3 \cdot 10^{-2} + ...$$

So really, real numbers are convergent sequences (partial sums of series) of rational numbers.

Also, observe that

$$y = 1.00\overline{0} \text{ and } z = 0.99\overline{9}$$

are different decimal expansions, but observe that

$$\begin{aligned} 10z - z &= 9.99\overline{9} - 0.99\overline{9} = 9, \\ z &= 1 = y. \end{aligned}$$

So they refer to the same number. Another way to see this is that

$$\begin{aligned} z &= 0.99\overline{9} = \frac{9}{10} + \frac{9}{10^2} + \frac{9}{10^3} + ... \\ &= \sum_{n=1}^{\infty} 9 \left(\frac{1}{10} \right)^n, \end{aligned}$$

a geometric series that converges to

$$z = \frac{a}{1-r} = \frac{9/10}{1 - \frac{1}{10}} = 1.$$

Similarly,

$$678.4799\overline{9} = 678.480\overline{0}.$$

Thus, there is not exactly a 1-1 relationship between decimal expansions and real numbers, unless we outlaw repeated nines, for example.

Theorem 6.16. *(Cantor) There does not exist a bijection $F : \mathbb{N} \to \mathbb{R}$. [That is, $|\mathbb{N}| < |\mathbb{R}|$.]*

Proof. Suppose instead that F exists. Then

$$\mathbb{R} = \{F(1), F(2), F(3), ...\}.$$

Then we write in terms of decimal expansions:

$$\begin{aligned} F(1) &= \pm * * * .d_1^1 d_2^1 d_3^1 d_4^1 ... \\ F(2) &= \pm * * * .d_1^2 d_2^2 d_3^2 d_4^2 ... \\ F(3) &= \pm * * * .d_1^3 d_2^3 d_3^3 d_4^3 ... \\ F(4) &= \pm * * * .d_1^4 d_2^4 d_3^4 d_4^4 ... \\ &... \end{aligned}$$

(where $\pm * **$ is a wild card - it stands for anything in front of the decimal point). Now, for $j \in \mathbb{N}$, let

$$a_j = \begin{cases} 2 & \text{if } d_j^j \neq 2 \\ 3 & \text{if } d_j^j = 2 \end{cases}.$$

Let

$$x = 0.a_1a_2a_3...$$

By construction, x is not in the list, but $x \in \mathbb{R}$. This is a contradiction. Therefore, no such F exists. □

Remark 6.17. *Thus, $\mathbb{R}$ is uncountable!*

Remark 6.18. *This type of argument is called a **diagonalization** argument.*

We see that we can construct old sets from new sets in many different ways.

Definition 6.19. *If A and B are two sets, the **Cartesian Product** $A \times B$ is the set defined by*

$$A \times B = \{(x, y) : x \in A, y \in B\}.$$

Remark 6.20. *Note that (x, y) is an **ordered pair**, so for instance $(1, 2)$ and $(2, 1)$ are different points in the Cartesian product of $\mathbb{N} \times \mathbb{N}$. By contrast, sets are unordered, so that $\{1, 2\} = \{2, 1\} = \{2, 1, 2\}$ as subset of $\mathbb{N}$.*

Proposition 6.21. *The Cartesian product of two countable sets is a countable set.*

Theorem 6.22. *Proof.* (Left as an exercise. Similar to the proof that the set of rational numbers is countable.) □

Proposition 6.23. *The union of a countable number of countable sets is countable.*

Proof. (Left as an exercise. Similar to the proof that the set of rational numbers is countable.) □

Proposition 6.24. *We have $|\mathbb{R}^2| = |\mathbb{R}|$.*

Proof. (Sketch) Let the map $G : \mathbb{R} \to \mathbb{R}^2$ be given by $G(x) = (x, 0)$. This is 1-1 since $G(x) = G(y)$ $\Rightarrow (x, 0) = (y, 0) \Rightarrow x = y$. Thus $|\mathbb{R}| \leq |\mathbb{R}^2|$. Next, let the map $H : \mathbb{R}^2 \to \mathbb{R}$ be defined as follows. For $(x, y) \in \mathbb{R}^2$, let the decimal expansions (with no repeated 9's) be

$$\begin{aligned} x &= \pm...x_2x_1x_0.x_{-1}x_{-2}... \\ y &= \pm...y_2y_1y_0.y_{-1}y_{-2}... \end{aligned}$$

and let

$$s = \begin{cases} 1 & \text{if } x \geq 0, y \geq 0 \\ 2 & \text{if } x < 0, y \geq 0 \\ 3 & \text{if } x \geq 0, y < 0 \\ 4 & \text{if } x < 0, y < 0 \end{cases}$$

determine the signs of x, y. Then we let

$$H(x, y) = ...x_2y_2x_1y_1x_0y_0s.x_{-1}y_{-1}x_{-2}y_{-2}...$$

It can be shown that by construction H is $1-1$. Thus $|\mathbb{R}^2| \leq |\mathbb{R}|$. Then, by the Schröder-Bernstein Theorem, $|\mathbb{R}^2| = |\mathbb{R}|$. □

A result of Cantor shows that for any given set, we can construct a set of **greater** cardinality.

Definition 6.25. *(The **power set**) Given a set S, the **power set** $\mathcal{P}(S)$ is defined as the set of all subsets of S. Thus,*

$$\mathcal{P}(S) = \{T : T \subseteq S\}.$$

For example,

$$\mathcal{P}(\{2,3,8\}) = \{\varnothing, \{2\}, \{3\}, \{8\}, \{2,3\}, \{3,8\}, \{2,8\}, \{2,3,8\}\}.$$

Proposition 6.26. *If S is a finite set with n elements, then the number of elements in $\mathcal{P}(S)$ is 2^n.*

Proof. (Left as an exercise) □

Theorem 6.27. *(Cantor) For any set S, $|S| < |\mathcal{P}(S)|$.*

Proof. There is an injection $f : S \to \mathcal{P}(S)$ given by $f(x) = \{x\}$, so $|S| \leq |\mathcal{P}(S)|$. We will show that there is no bijection from S to $\mathcal{P}(S)$.
Suppose instead that there exists a bijection $g : S \to \mathcal{P}(S)$. Then we define the set

$$A := \{x \in S : x \notin g(x)\},$$

which is itself a subset of S. Now, since g is onto, there exists $y \in A$ such that $g(y) = A$. Then we ask the question whether $y \in g(y)$. If $y \in g(y) = A$, then $y \notin g(y)$, a contradiction. If $y \notin g(y)$, then $y \in A = g(y)$, also a contradiction. Therefore, the assumption that there exists a bijection is false. Therefore, $|S| < |\mathcal{P}(S)|$. □

As a result, there is a infinite sequence of cardinalities of infinite sets, each strictly larger than the previous:

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

There are special symbols for these different **cardinal numbers**, using the Hebrew letter aleph ($\aleph$):

$$\aleph_0 = |\mathbb{N}| < \aleph_1 = |\mathcal{P}(\mathbb{N})| < ... < \aleph_n = |\mathcal{P}^n(\mathbb{N})| < ...$$

7. Numbers with different bases and the Cantor set

Given a real number x, we are accustomed to expressing this number is base 10 (decimal) notation. For example, the number 34.7480320243... is

$$3 \times 10^1 + 4 \times 10^0 + 7 \times 10^{-1} + 4 \times 10^{-2} + ...,$$

and in fact any real number can be expressed as

$$x = \pm \sum_{\substack{j \leq k \\ j \in \mathbb{Z}}} a_j 10^j,$$

where $a_j \in \{0, 1, 2, 3, ..., 9\}$. But what is special about 10? We certainly can use other bases to describe numbers. For examples:

base 10	base 2	base 3	base 5
1	1	1	1
2	10	2	2
3	11	10	3
4	100	11	4
5	101	12	10
6	110	20	11
7	111	21	12
8	1000	22	13
9	1001	100	14
10	1010	101	20

For example,

$$\begin{aligned} 10_{10} &= 1010_2 = 1\times 2^3 + 0\times 2^2 + 1\times 2^1 + 0\times 2^0 \\ &= 101_3 = 1\times 3^2 + 0\times 3^1 + 1\times 3^0. \end{aligned}$$

Any operations normally done with decimal expansions can be done in other bases. Here are some examples.

$$101_2\,(1101_2) = ?$$

$$\begin{array}{r} 1\ 1\ 0\ 1 \\ \times\quad 1\ 0\ 1 \\ \hline 1\ 1\ 0\ 1 \\ 1\ 1\ 0\ 1 \\ \hline 1\ 0\ 0\ 0\ 0\ 0\ 1 \end{array}$$

$$\frac{2021_3}{12_3} = ?$$

$$\begin{array}{r} 1\ 1\ 0.\ \overline{0\ 1\ 2\ 1} \\ 1\ 2\ \overline{)\ 2\ 0\ 2\ 1} \\ \underline{1\ 2} \\ 1\ 2 \\ \underline{1\ 2} \\ 1.\ 0\ 0 \\ \underline{1\ 2} \\ 1\ 1\ 0 \\ \underline{1\ 0\ 1} \\ 2\ 0 \\ \underline{1\ 2} \\ 1\ 0\ 0 \end{array}$$

We will use these different number bases when we consider the following example of an interesting set. This set is generated from a sequence of operations. Let

set	# of intervals	length of each interval	total length
$C_0 = [0,1] \subseteq \mathbb{R}$	1	1	1
$C_1 = \left[0, \frac{1}{3}\right] \cup \left[\frac{2}{3}, 1\right]$	2	$\frac{1}{3}$	$\frac{2}{3}$
$C_2 = \left[0, \frac{1}{9}\right] \cup \left[\frac{2}{9}, \frac{1}{3}\right] \cup \left[\frac{2}{3}, \frac{7}{9}\right] \cup \left[\frac{8}{9}, 1\right]$	4	$\frac{1}{9}$	$\frac{4}{9}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$C_n = C_{n-1}$ with middle thirds removed	2^n	$\frac{1}{3^n}$	$\frac{2^n}{3^n}$

Then we let the **Cantor set** C be

$$C = \bigcap_{n=1}^{\infty} C_n.$$

That is $x \in C$ if and only if $x \in C_n$ for $n = 1, 2, ...$

Observe that if we express the numbers in $[0,1]$ as ternary numbers (base 3),

$$\begin{aligned} C_1 &= \text{ base 3 numbers of the form } 0.a_1 * * * ... \\ \text{where } a_1 &= 0 \text{ or } 2 \\ C_2 &= \text{base 3 numbers of the form } 0.a_1 a_2 * * * .. \\ \text{where } a_1, a_2 &\in \{0, 2\} \\ C_3 &= \text{base 3 numbers of the form } 0.a_1 a_2 a_3 * * * .. \\ \text{where } a_1, a_2, a_3 &\in \{0, 2\} ... \end{aligned}$$

So

$$C = \{x : \text{in base 3, } x = 0.a_1 a_2 a_3 ... \text{ where each } a_j \in \{0, 2\}\}.$$

Note that this includes numbers like

$$0.0\overline{2}_3 = 0.1_3 = \frac{1}{3}.$$

Lemma 7.1. *We have $|(0,1)| = |\mathbb{R}|$, and in fact if I is any closed or open interval, $|I| = \mathbb{R}$.*

Proof. (Sketch) Use $f : \mathbb{R} \to (0,1)$ defined by $f(x) = \frac{1}{\pi} \arctan(x) + \frac{1}{2}$. You can show $f'(x) > 0$, so f is $1-1$, and you can also show it is onto. A similar function works for any open interval. And then it is not hard to see that closed intervals (since they are larger than open ones) also work. □

Proposition 7.2. *The Cantor set C satisfies $|C| = |[0,1]| = |\mathbb{R}|$, and yet it can be covered by a set of intervals of arbitrarily small size (i.e. it has measure 0).*

Proof. First, $C \subseteq [0,1]$, so using the inclusion map $f : C \to [0,1]$, we see that $|C| \leq |[0,1]|$. Next, we construct a 1-1 function $h : [0,1] \to C$. For any $x \in [0,1]$, we expand it as a binary expansion with only 0's and 1's. In order to make it well-defined, we do not use repeated 1's in these expansions (ie we use $.0101\overline{0}_2$ instead of $.0100\overline{1}_2$), with the exception of $1 = .\overline{1}_2$. Then we map to C by "multiplying by 2" to get a **ternary** expansion with only 0's and 2's. This is now a well-defined function that is 1-1. Thus, $|[0,1]| \leq |C|$, so by the Schröder-Berstein Theorem, $|C| = |[0,1]|$, which is the same as $|\mathbb{R}|$ by the lemma. The last part follows from the fact that C_n covers C for all n. □

8. Mathematical Induction

If we wish prove a statement $P(n)$ which depends on an integer n for all $n \geq a$ (where $a \in \mathbb{Z}$), then we may use the following procedure.

Induction Proof:

(1) Prove $P(a)$ is true (the **basis step**).

(2) Then, for some $k \geq a$, assume $P(j)$ is true for all $j \in \mathbb{Z}$ such that $a \leq j \leq k$ (the **induction hypothesis**).
(3) Then prove that $P(k+1)$ must also be true.

This completes the proof, by the principal of mathematical induction.

Example 8.1. *Prove that $2^m \geq m^2 + 3m$ for all $m \geq 7$.*

Proof. (by induction). Let $P(m)$ be the statement "$2^m \geq m^2 + 3m$". Observe that $2^7 = 128 \geq 70 = 49 + 21 = 7^2 + 3(7)$, so that the case $P(7)$ is true. Now, assume that $P(k) =$ "$2^k \geq k^2 + 3k$" for some $k \geq 7$. Then

$$\begin{aligned} 2^{k+1} &= 2 \cdot 2^k \\ &\geq 2\left(k^2 + 3k\right) \text{ by the induction hypothesis} \\ &= 2k^2 + 6k. \end{aligned}$$

On the other hand,

$$\begin{aligned} (k+1)^2 + 3(k+1) &= k^2 + 2k + 1 + 3k + 3 \\ &= k^2 + 5k + 4 \\ &< k^2 + 6k + 4 \\ &< k^2 + 6k + k^2 = 2k^2 + 6k, \end{aligned}$$

since $k \geq 7$. Thus we have

$$2^{k+1} \geq 2k^2 + 6k > (k+1)^2 + 3(k+1),$$

so that $P(k+1)$ is true. By the principle of mathematical induction, $P(m)$ is true for all $m \geq 7$. □

Example 8.2. *Prove that for any $n \geq 2$ and any $a_1, a_2, ..., a_n \in \mathbb{C}$, $a_1 + a_2 + ... + a_n = a_2 + ... + a_n + a_1$.*

Proof. We prove by induction on n. When $n = 2$, $a_1 + a_2 = a_2 + a_1$ by the commutative property of addition. Next, we assume that the statement has been shown for $2 \leq n \leq k$. Then

$$\begin{aligned} a_1 + a_2 + ... + a_k + a_{k+1} &= (a_1 + a_2 + ... + a_k) + a_{k+1} \text{ by the order of operations} \\ &= (a_2 + ... + a_k + a_1) + a_{k+1} \text{ by the induction hypothesis} \\ &= ((a_2 + ... + a_k) + a_1) + a_{k+1} \text{ by the order of operations} \\ &= (a_2 + ... + a_k) + (a_1 + a_{k+1}) \text{ by the assoc. prop. of addition} \\ &= (a_2 + ... + a_k) + (a_{k+1} + a_1) \text{ by the comm. prop. of addition} \\ &= ((a_2 + ... + a_k) + a_{k+1}) + a_1 \text{ by the assoc. prop. of addition} \\ &= a_2 + ... + a_k + a_{k+1} + a_1 \text{ by the order of operations.} \end{aligned}$$

Thus, the statement is true for $n = k + 1$. Therefore, by induction, the statement is true for all positive integers $n \geq 2$. □

We now prove the celebrated fundamental theorem of arithmetic. First we need a lemma.

Lemma 8.3. ***(Euclid's Lemma)*** *If $a, b \in \mathbb{Z}$ and p is a prime number, and if $p|ab$, then $p|a$ or $p|b$.*

Proof. With the given $p|ab$, if $p|a$, we are done. Otherwise, $\gcd(a, p) = 1$, so by Proposition 5.10, there exist integers x and y such that $ax + py = 1$. Multiplying by b, we get

$$abx + pyb = b.$$

We see that p is a factor of the left hand side (since $p|ab$), so $p|b$. □

Corollary 8.4. *If for $m \geq 1$, $a_1, a_2, ..., a_m \in \mathbb{Z}$ and p is a prime number, and if $p|a_1a_2...a_m$, then p divides one of the a_j.*

Proof. We proceed by induction on m. The statement is clearly true for $m = 1$. Next, we assume that we have proved the statement for $1 \leq m \leq k$. Then we consider the case $m = k + 1$. Then if

$$\begin{aligned} &p|a_1a_2...a_ka_{k+1}, \text{ then} \\ &p|\left((a_1a_2...a_k)\, a_{k+1}\right). \end{aligned}$$

By Euclid's lemma, $p|a_1a_2...a_k$ or $p|a_{k+1}$. In the second case, we are done. In the first case, the induction hypothesis implies that $p|a_j$ for some j such that $1 \leq j \leq k$. Thus, the case $m = k + 1$ has been proved. By induction, the statement is true for every integer $m \geq 1$. □

Theorem 8.5. *(**Fundamental Theorem of Arithmetic**) Every positive integer $x > 1$ can be written as*

$$x = p_1^{r_1}p_2^{r_2}...p_k^{r_k},$$

where $p_1 < p_2 < ... < p_k$ are prime numbers and $r_1, ..., r_k$ are positive integers. The expression above is unique.

Proof. (Existence part) We prove by induction on x. If $x = 2$, then it is already in the correct form, with $p_1 = 2$, $r_1 = 1$, and $k = 1$. Now, we assume that the statement has be proven for all integers x such that $2 \leq x \leq k$. Now, consider the integer $x = k + 1$. If this integer is prime, then we are done. Otherwise, it can be written as $x = ab$, with

$$\begin{aligned} 2 \;&\leq\; a \leq \frac{k+1}{2} < \frac{k+k}{2} = k \\ 2 \;&\leq\; b \leq \frac{k+1}{2} < \frac{k+k}{2} = k \end{aligned}$$

By the induction hypothesis, both a and b can be written as products of primes, so therefore $x = k + 1$ is also a product of prime numbers. By induction, any integer $x \geq 2$ can be written as a product of primes.
(Uniqueness part) Suppose that x can be written in two different ways as a product of primes, say

$$x = p_1...p_\ell = q_1...q_m,$$

where $p_1 \leq p_2 \leq ... \leq p_\ell$ and $q_1 \leq q_2 \leq ... \leq q_m$ are all primes (we allow equality now so that we don't have to deal with the exponents). Next, we divide out by all the common prime factors on both sides. If we end up with $1 = 1$, that would show that both sides are identical, so that the uniqueness proof is complete. Otherwise, we will be left with an equation similar to the one above with the p_j and q_j distinct primes. But then p_1 divides the left side, so it divides the right side. By the corollary above, it must divide one of the q_j, and since they are prime $p_1 = q_j$. But this contradicts the fact that we have divided out all the prime factors, so the assumption that the reduced equation has primes in it is false. Thus, the two expressions are the same, so that the prime factorization of x is unique. □

9. Properties of Groups and addition in modular arithmetic

Given any positive integer n, and any $k \in \mathbb{Z}$, by the division algorithm (Proposition 5.9), there exists a unique $q \in \mathbb{Z}$ and a unique $r \in \{0, 1, ..., n-1\}$ such that

$$k = qn + r,$$

and r is called the **remainder** when k is divided by n. We define the symbol " k modulo n "

$$k \mod n = r.$$

For examples,

$$\begin{aligned} -17 \mod 5 &= 3 \text{ because } -17 = (-4)\,5 + 3 \\ 467 \mod 6 &= 5 \text{ because } 467 = (77)\,6 + 5 \end{aligned}$$

For $a, b \in \mathbb{Z}$, we also write "a is congruent to b modulo n"

$$a \equiv b \mod n$$

whenever $a \mod n = b \mod n$.

Lemma 9.1. *For any positive integer n and any integers a, b, $a \equiv b \mod n$ if and only if $(a-b)$ is a multiple of n.*

Proof. ($\Rightarrow$) Suppose $a \equiv b \mod n$ as above. Then $a \mod n = b \mod n = r$, so

$$\begin{aligned} a &= q_1 n + r \\ b &= q_2 n + r \end{aligned}$$

for some $q_1, q_2 \in \mathbb{Z}$. Subtracting, we see that

$$a - b = (q_1 - q_2)\, n.$$

($\Leftarrow$) On the other hand, suppose $a - b = kn$ for some $k \in \mathbb{Z}$. Then by the division algorithm, there exist q_1, q_2 in $\mathbb{Z}$ and $r_1, r_2 \in \{0, 1, ..., n-1\}$ such that

$$\begin{aligned} a &= q_1 n + r_1 \\ b &= q_2 n + r_2, \end{aligned}$$

so then

$$a - b = (q_1 - q_2)\, n + (r_1 - r_2) = kn,$$

so that

$$r_1 - r_2 = (q_1 - q_2 + k)\, n.$$

By construction,

$$-n + 1 \leq r_1 - r_2 \leq n - 1,$$

so the only way this can be a multiple of n is that $r_1 - r_2 = 0$, so that $r_1 = a \mod n = r_2 = b \mod n$. Thus $a \equiv b \mod n$. □

One important fact about congruence modulo n is that it is an **equivalence relation**. Note that a relation R on a set S is a function that takes any two elements $x, y \in S$ and produces the statement xRy, which is either true or false.

Definition 9.2. *If A is a set and R is a relation on A, then R is called an **equivalence relation** if*

(1) (Reflexive Property) For every $a \in A$, aRa.
(2) (Symmetry) For every $a, b \in A$, if aRb, then also bRa.

(3) (Transitivity) For every $a, b, c \in A$, if aRb and bRc, then also aRc.

Examples of equivalence relations include the set of complex numbers, with the relation $=$, the set of triangles in the plane with the relation being congruence, the set of triangles in the plane with the relation being similarity,

Lemma 9.3. *Congruence modulo n is an equivalence relation on $\mathbb{Z}$.*

Proof. (Reflexive Property) For all $x \in \mathbb{Z}$, $x - x = 0\,(n)$ so that $x \equiv x \mod n$.
(Symmetry) If $a, b \in \mathbb{Z}$ and $a \equiv b \mod n$, then $a - b = kn$ for some $k \in \mathbb{Z}$, so $b - a = (-k)\,n$, which implies $b \equiv a \mod n$.
(Transitivity) If $a, b, c \in \mathbb{Z}$ and $a \equiv b \mod n$ and $b \equiv c \mod n$, then $a - b = kn$ and $b - c = \ell n$ for some $k, \ell \in \mathbb{Z}$, so by adding the equations, we get $a - c = (k + \ell)\,n$, so $a \equiv c \mod n$. □

Given an equivalence relation R on a set S, an **equivalence class** $[s]$ **of** $s \in S$ is the set of all $t \in S$ such that tRs. That is,

$$[s] = \{t \in T : tRs\}\,.$$

Recall that a **group** is a set along with an operation that satisfies closure, the associative property, the existence of the identity, and the existence of inverses for that operation. An **abelian group** is a group that in addition satisfies the commutative property. For example, the set $\mathbb{C}$ with operation addition is a group, and the set $(0, \infty)$ is a group under the operation multiplication.

We define the set $\mathbb{Z}_n$ to be $\{0, 1, ..., n-1\}$ where the operation is addition mod n. In other words, if $x, y \in \mathbb{Z}_n$, then the sum $x + y$ is $(x + y) \mod n$. Sometimes, $\mathbb{Z}_n$ is defined to be the set of equivalence classes of integers $\mod n$. That is, for example

$$\begin{aligned} 1 \mod 5 &= 6 \mod 5 = -9 \mod 5 = ... \\ [1] &= [6] = [-9] = \{..., -14, -9, -4, 1, 6, 11, ...\} \\ 1 + 7 &\equiv 8 \equiv 3 \mod 5 \text{ so } [1] + [7] = [3]\,. \end{aligned}$$

So, we see that addition modulo n is **well-defined** in $\mathbb{Z}_n$. This means that if I add x and y in $\mathbb{Z}_n$, I get the same result (mod n) if I were to choose different representatives. A proof:

Statement: Addition modulo n is well-defined. If $x_1 \equiv x_2 \mod n$ and $y_1 \equiv y_2 \mod n$, then

$$x_1 + x_2 \equiv y_1 + y_2 \mod n$$

Proof: Given x_1, x_2, y_1, y_2 as above,

$$x_2 = x_1 + kn,\ y_2 = y_1 + \ell n$$

for some $k, \ell \in \mathbb{Z}$. Then

$$x_2 + y_2 = x_1 + kn + y_1 + \ell n,$$

so

$$(x_2 + y_2) - (x_1 + y_1) = (k + \ell)\,n,$$

so

$$x_2 + y_2 \equiv (x_1 + y_1) \mod n.$$

QED

Lemma 9.4. *The set $\mathbb{Z}_n = \{0, 1, ..., n-1\}$ with operation addition $\mod n$ is an abelian group.*

Proof. (Closure) For any $x, y \in \mathbb{Z}_n$, $x + y$ is congruent to one element of $\mathbb{Z}_n \mod n$, since every integer has a remainder in $\mathbb{Z}_n$ when divided by n.
(Associative Property) For any $x, y, z \in \mathbb{Z}_n$, by the associative property of addition in $\mathbb{Z}$,

$$(x + y) + z = x + (y + z),$$

so

$$[(x + y) + z] \mod n = [x + (y + z)] \mod n,$$

so

$$(x + y) + z \equiv x + (y + z) \mod n.$$

(Existence of identity) Since for all $x \in \mathbb{Z}_n$,

$$\begin{aligned} 0 + x &\equiv x \mod n \\ x + 0 &\equiv x \mod n, \end{aligned}$$

so 0 is the identity.
(Existence of inverses) The identity 0 is its own inverse. For all nonzero $x \in \mathbb{Z}_n$, $n - x \in \mathbb{Z}_n$, and

$$\begin{aligned} x + (n - x) &\equiv n \equiv 0 \mod n, \\ (n - x) + x &\equiv n \equiv 0 \mod n, \end{aligned}$$

so each element has an inverse.
(Commutative property) For any $x, y \in \mathbb{Z}_n$,

$$x + y = y + x$$

by the commutative property in $\mathbb{Z}$, so

$$x + y \equiv y + x \mod n.$$

Thus, $\mathbb{Z}_n$ is an abelian group under addition modulo n. □

Many of the standard properties of equations carry over to the addition modulo n setting.

Lemma 9.5. *Suppose that* $a \equiv b \mod n$ *and* $c \equiv d \mod n$. *Then*

$$\begin{aligned} a + c &\equiv b + d \mod n \\ a - c &\equiv b - d \mod n \end{aligned}$$

Proof. (Left as an exercise) Note that this Lemma is just showing that addition and subtraction mod n are well-defined. We proved the addition part above. □

In a group, a **subgroup** is a subset that is itself a group under the same group operation. For example, the set $S = \{0, 4\}$ is a subgroup of $\mathbb{Z}_8 = \{0, 1, 2, 3, 4, 5, 6, 7\}$ because

$$\begin{aligned} 0 + 0 &\equiv 0 \mod 8 \\ 0 + 4 &\equiv 4 + 0 \equiv 4 \mod 8 \\ 4 + 4 &\equiv 8 \equiv 0 \mod 8 \end{aligned}$$

implies that it satisfies closure, existence of identity, existence of inverses (note that 4 is its own inverse), and automatically the associative property holds since it holds in $\mathbb{Z}_8$.

Here are some important facts about groups and subgroups.

Proposition 9.6. *The identity of a group is unique.*

Proof. Let e_1 and e_2 be two identity elements of a group G under the operation $*$. Observe that

$$\begin{aligned} e_1 * e_2 &= e_1 \text{ since } e_2 \text{ is an identity, and} \\ e_1 * e_2 &= e_2 \text{ since } e_1 \text{ is an identity.} \end{aligned}$$

Then $e_1 = e_2$. Thus, the identity in G is unique. □

Proposition 9.7. *Given an element x of a group G, its inverse is unique.*

Proof. (Left as an exercise. Assume y and z are both inverses, and prove $y = z$.) □

Proposition 9.8. *If e_G is the identity of a group G and H is a subgroup of G, then e_G is also in H and is the identity in H.*

Proof. With the given e_G in the group G with operation $*$, suppose that e_H is the identity in the subgroup H. Then

$$e_H * e_H = e_H.$$

Multiplying on the left by the inverse e_H^{-1} of e_H in G, we get

$$e_H^{-1} * (e_H * e_H) = e_H^{-1} * e_H$$

Then,

$$\left(e_H^{-1} * e_H\right) * e_H = e_G,$$

so

$$e_G * e_H = e_G,$$

so that

$$e_H = e_G.$$

□

Proposition 9.9. ***(Two-step Subgroup Test)*** *If S is a nonempty subset of a group G that is closed under the group operation and such that every element in S has its inverse in S, then S is a subgroup of G.*

Proof. Let the group operation for G be denoted $*$. The elements of S automatically satisfy the associative property, because $S \subseteq G$. Also, by the given properties, for any $x \in S$, also $x^{-1} \in S$, and so $x * x^{-1} = e \in S$, so it also satisfies the identity property as well. The inverse and closure properties follow automatically from the given, so S is a subgroup of G. □

Given an element a of a finite group G with operation $*$ and identity e, the **subgroup** $\langle a \rangle$ **generated by** a is defined as

$$\langle a \rangle = \{e, a, a * a, (a * a) * a, ...\}.$$

This can even be applied to infinite groups, but then the inverses need to be added in:

$$\langle a \rangle = \left\{e, a, a^{-1}, a * a, a^{-1} * a^{-1}, (a * a) * a, \left(a^{-1} * a^{-1}\right) * a^{-1}, ...\right\}$$

Lemma 9.10. *If G is a group with operation $*$ and $a \in G$, then $\langle a \rangle$ is a subgroup of G.*

Proof. (left as an exercise) □

Let's look at a specific example. Let $G = \mathbb{Z}_{12}$. Then

$$\begin{aligned}
\langle 0\rangle &= \{0\} \\
\langle 1\rangle &= \{0, 1, 1+1=2, 2+1=3, ..., 11\} = \mathbb{Z}_{12} \\
\langle 2\rangle &= \{0, 2, 4, 6, 8, 10\} \\
\langle 3\rangle &= \{0, 3, 6, 9\} \\
\langle 4\rangle &= \{0, 4, 8\} \\
\langle 5\rangle &= \{0, 5, 10, 3, 8, 1, 6, 11, 4, 9, 2, 7\} = \mathbb{Z}_{12} \\
\langle 6\rangle &= \{0, 6\} \\
&... \\
\end{aligned}$$

The number of elements in a group G is called the **order of** G and is denoted $|G|$. For example, $|\mathbb{Z}_{12}| = 12$. If a is an element of a group G with identity e, the smallest number k such that

$$\underbrace{a * a * a * ... * a}_{k \text{ factors}} = e$$

is called the **order of** a and is denoted $|a|$. Note that $|a|$ is also the same number as the number of elements in the subgroup $\langle a\rangle$, so

$$|a| = |\langle a\rangle| .$$

We call an element a of a group G with operation $*$ a **generator** if $\langle a\rangle = G$; in other words, if *every* element in G is in $\langle a\rangle$. A group with a generator is called a **cyclic** group. So a cyclic group has an element a such that $|a| = |G|$.

For example, consider the subgroups $\langle a\rangle$ above of $\mathbb{Z}_{12}$ for different values of a. Note that $\langle 1\rangle = \langle 5\rangle = \mathbb{Z}_{12}$ so that 1 and 5 are generators, but for example $|3| = |\langle 3\rangle| = 4$, so 3 is not a generator.

You may notice that the order of every element in $\mathbb{Z}_{12}$ is some factor of 12. This is due to this more general fact.

Theorem 9.11. *(**Lagrange's Theorem**) If G is a finite group and if H is a subgroup of G, then $|H|$ is a factor of $|G|$.*

Proof. (Sketch) If H is a subgroup of G with operation $*$, for every element $a \in G$, we can find the right **coset** $a * H$ of H, defined as

$$a * H = \{a * h : h \in H\} .$$

One can prove that
(a) every element of G is in some coset of H
(b) two cosets of H are either exactly the same or have no elements in common.
(c) every coset has the same number of elements, $|H|$.
Then this means that there are a finite number of disjoint cosets that make a subdivision of G into pieces

$$G = \bigcup_{j=1}^{p} a_j * H,$$

and since the number of elements in each $a_j * H$ is $|H|$,

$$|G| = p\,|H| .$$

□

Corollary 9.12. *If G is a finite group with identity e and $a \in G$, then*

$$a^{|G|} = e.$$

Proof. Since the subgroup $\langle a \rangle$ generated by a has order k equal to the order of a, $a^k = e$. But since k is a factor of $|G|$ by Lagrange's Theorem, $a^{|G|} = e$. □

So, for example, in $\mathbb{Z}_{12}$, the only possible orders of elements are $1, 2, 3, 4, 6, 12$ (since they all generate subgroups of that order). To see the cosets in action, take for example

$$\begin{aligned} H &= \langle 3 \rangle = \{0, 3, 6, 9\} \subseteq \mathbb{Z}_{12} \\ 0 + H &= H \\ 1 + H &= \{1, 4, 7, 10\} \\ 2 + H &= \{2, 5, 8, 11\} \\ G &= \mathbb{Z}_{12} = (0 + H) \cup (1 + H) \cup (2 + H). \end{aligned}$$

Here is a summary of numerical facts we know about groups. Let G be a finite group, and let H be a subgroup of G.

- $|H| \cdot [G : H] = G$, where $[G : H]$ is the **index** of H in G, the number of distinct cosets of H in G.
- For any $a \in G$, $a^{|G|} = e$.
- For any $a \in G$, the order $|a|$ is a factor of $|G|$.

10. Multiplication in Modular Arithmetic

In $\mathbb{Z}_n$, conveniently we have another operation we can play with — multiplication $\mod n$. First of all, we will check that multiplication $\mod n$ is **well-defined**. This means that we need the Lemma below.

Lemma 10.1. ***(Multiplication $\mod n$ is well-defined)*** *If $a \equiv b \mod n$ and $c \equiv d \mod n$, then*

$$ac \equiv bd \mod n.$$

Proof. Suppose $a \equiv b \mod n$ and $c \equiv d \mod n$. Then

$$a - b = kn,\ c - d = \ell n$$

for some $k, \ell \in \mathbb{Z}$. Thus,

$$\begin{aligned} ac &\equiv (b + kn)(d + \ell n) \mod n \\ &\equiv bd + kdn + b\ell n + k\ell n^2 \mod n \\ &\equiv bd \mod n. \end{aligned}$$

□

Note that the number 1 acts as the identity for multiplication $\mod n$.

Lemma 10.2. *The integer k has a multiplicative inverse $\mod n$ if and only if $\gcd(k, n) = 1$.*

Proof. ($\Rightarrow$) If k has a multiplicative inverse $x \mod n$, then

$$\begin{aligned} kx &\equiv 1 \mod n, \text{ or} \\ kx - 1 &= pn \end{aligned}$$

for some integer p. Then

$$kx + (-p)\, n = 1,$$

so a positive integer divides both k and n implies by that equation that it divides 1, so the integer is 1. Thus, $\gcd(k, n) = 1$.
($\Leftarrow$) If $\gcd(k, n) = 1$, then by Bezout's Identity (Proposition 5.10), there exist integers r and s such that

$$ks + ns = 1,$$

so that

$$ks \equiv 1 \mod n.$$

Thus, s is the multiplicative inverse of $k \mod n$. □

With this information, we can now define a new group. Let

$$U(\mathbb{Z}_n) = \{k \in \mathbb{Z}_n : \gcd(k, n) = 1\}$$

denote the **group of units of** $\mathbb{Z}_n$. By simple modular arithmetic properties and the Lemma above, all of the properties of groups are satisfied. For example,

$$\begin{aligned} U(\mathbb{Z}_7) &= \{1, 2, 3, 4, 5, 6\} \\ U(\mathbb{Z}_8) &= \{1, 3, 5, 7\} \\ U(\mathbb{Z}_9) &= \{1, 2, 4, 5, 7, 8\} \\ U(\mathbb{Z}_{10}) &= \{1, 3, 7, 9\}. \end{aligned}$$

Notice that $H = \{1, 4, 7\} = \langle 4 \rangle$ is a subgroup of $U(\mathbb{Z}_9)$. The cosets are H and $2H = \{2, 8, 5\}$.

The Euler totient function ϕ is defined on $n \in \mathbb{N}$ as

$$\begin{aligned} \phi(n) = &\text{ the number of positive integers } \leq n \text{ that} \\ &\text{are relatively prime to } n. \end{aligned}$$

For example $\phi(7) = 6$, $\phi(8) = 4$, $\phi(10) = 4$. Observe that

$$|U(\mathbb{Z}_n)| = \phi(n).$$

Thanks to Lagrange's Theorem (Theorem 9.11), we have the following result concerning $U(\mathbb{Z}_n)$.

Theorem 10.3. *(Euler's Theorem)If a and n are relatively prime integers, then*

$$a^{\phi(n)} \equiv 1 \mod n.$$

Proof. Considering a as an element of $U(\mathbb{Z}_n)$, we have that

$$a^{\phi(n)} = a^{|U(\mathbb{Z}_n)|} \equiv 1 \mod n.$$

□

For example, this allows us to find out interesting things such as:

Example 10.4. *Find the units digit of the integer*

$$67905287^{78905238}.$$

Solution: *We just need to compute this number* $\mod 10$. *So,*

$$67905287^{78905238} \equiv 7^{78905238} \mod 10.$$

Since $\phi(10) = 4$*, we see that* $7^k \equiv 1 \mod 10$ *if* k *is a multiple of* 4*. Then*

$$\begin{aligned} 7^{78905238} \mod 10 &= 7^{78905238 \mod 4} \mod 10 \\ &= 7^2 \mod 10 \\ &= 49 \mod 10 = 9. \end{aligned}$$

So 9 *is the last digit.*

Here is an important application.

Theorem 10.5. *A number* $x \in \mathbb{R}$ *is rational if and only if* x *can be written in a decimal expansion that is eventually repeating.*

Proof. Suppose first that x can be written as a decimal that is eventually repeating. Then we write

$$x = \pm x_1 ... x_k . y_1 ... y_m \overline{z_1 ... z_n},$$

where $k \geq 1$, $m \geq 0$, $n \geq 1$, and all the x_j, y_ℓ, z_p are digits in $\{0, 1, 2, ..., 9\}$. (We will allow even the leading digit x_j to be 0 any other digits to be 0) Then we see that

$$\begin{aligned} 10^{m+n} x &= \pm x_1 ... x_k . y_1 ... y_m z_1 ... z_n . \overline{z_1 ... z_n} \\ 10^m x &= \pm x_1 ... x_k . y_1 ... y_m . \overline{z_1 ... z_n}, \end{aligned}$$

We subtract these equations. If we let

$$s = \pm x_1 ... x_k . y_1 ... y_m z_1 ... z_n - \pm x_1 ... x_k . y_1 ... y_m,$$

then $s \in \mathbb{Z}$ and

$$x = \frac{s}{10^{m+n} - 10^m}.$$

Since $n > 0$, the denominator is a positive integer, so that we see that x is a rational number.✓
Next, suppose x is a rational number. Then we have that $x = \frac{p}{q}$ with $p, q \in \mathbb{Z}$ and $q \neq 0$. Changing the sign of p if necessary, we can assume q is positive. We now write q as $q = 2^r 5^s \widetilde{q}$, where $\widetilde{q}$ has no factors of 2 or 5. Since $\widetilde{q}$ is relatively prime to 10, by Euler's Theorem (Theorem 10.3),

$$10^{\phi(\widetilde{q})} \equiv 1 \mod \widetilde{q},$$

so there exists an integer ℓ such that

$$\ell \widetilde{q} = 10^{\phi(\widetilde{q})} - 1.$$

Then, observe that

$$\begin{aligned} (2^s 5^r \ell)\, q &= (2^s 5^r \ell)\, 2^r 5^s \widetilde{q} \\ &= 10^{r+s} \left(10^{\phi(\widetilde{q})} - 1\right) \\ &= 10^{m+n} - 10^m, \end{aligned}$$

where $m = r + s \geq 0$ and $n = \phi(\widetilde{q}) \geq 1$. Thus,

$$x = \frac{p}{q} = \frac{(2^s 5^r \ell)\, p}{(2^s 5^r \ell)\, q} = \frac{\text{integer}}{10^{m+n} - 10^m}$$

for some $m \geq 0$, $n > 0$. By the Lemma below, this implies that x can be expressed as a decimal number that is eventually repeating. □

Lemma 10.6. *If a rational number* x *can be written as* $\frac{a}{10^{m+n} - 10^m}$*, where* $a \in \mathbb{Z}$ *and* $m \geq 0$*,* $n > 0$*,* $m, n \in \mathbb{Z}$*, then there exists a decimal expansion of* x *that is eventually repeating.*

Proof. With the hypothesis, we first observe that

$$\frac{a}{10^{m+n}-10^m} = \frac{1}{10^m}\frac{a}{10^n-1}.$$

It is sufficient to show that a decimal expansion for $\frac{a}{10^n-1}$ is eventually repeating, because the factor $\frac{1}{10^m}$ simply shifts that expansion by m places. Using the division algorithm, there exists an integer q and a nonnegative integer $r < 10^n - 1$ (at most n decimal places) such that

$$a = q\left(10^n - 1\right) + r.$$

Dividing by $10^n - 1$,

$$\frac{a}{10^n-1} = q + \frac{r}{10^n-1}.$$

We now write the integers q and r in decimal form as

$$q = \pm q_1...q_t,\ r = r_1...r_n,$$

where each q_i, r_i are in $\{0, 1, 2, ..., 9\}$. Then we have

$$\frac{a}{10^n-1} = \pm q_1...q_t\ .\ \overline{r_1...r_n}$$

as a decimal number. □

11. Other interesting groups

In this section we examine some other interesting groups.

The **dihedral group** D_n is the set of symmetries of a regular n-gon (regular polygon with n sides). It can be shown that D_n consists of n rotations (including the identity) and n reflections, a total of $2n$ elements. The operaton is composition of functions.

For example, D_4 is the set of symmetries of the square. The set D_4 is

$$D_4 = \{e, R_{90}, R_{180}, R_{270}, h, v, f_+, f_-\},$$

where symmetries are labels as functions in this way: e is the identity, mapping the square to itself without moving. Th map R_{90} rotates the square by 90° counterclockwise, and R_{180} and R_{270} are defined similarly. So this means, for example, that $R_{90} \circ R_{90} = R_{90}^2 = R_{180}$, $R_{90}^3 = R_{270} = R_{90}^{-1}$. The reflection across horizontal line of symmetry of the square is labeled h, and v is the reflection across the vertical line of symmetry. The function f_+ reflects the square across the diagonal line of symmetry with positive slope, and f_- reflects across the diagonal line of symmetry with negative slope. The group multiplication is function composition, so for example $R_{90} \circ h$ means that we first reflect across the horizontal line of symmetry and then rotate 90° counterclockwise. Note that we do h first and then R_{90}, as with function composition: $(F \circ G)(x) = F(G(x))$.

We can check that indeed this set is closed under the group operation. Here are some examples. We let $\begin{bmatrix} A & B \\ C & D \end{bmatrix}$ denote the square

$$\begin{aligned} R_{90} \circ h \begin{bmatrix} A & B \\ C & D \end{bmatrix} &= R_{90} \begin{bmatrix} C & D \\ A & B \end{bmatrix} = \begin{bmatrix} D & B \\ C & A \end{bmatrix} \\ &= f_+ \begin{bmatrix} A & B \\ C & D \end{bmatrix}, \end{aligned}$$

so $R_{90} \circ h = f_+$. On the other hand,

$$\begin{aligned} h \circ R_{90} \begin{bmatrix} A & B \\ C & D \end{bmatrix} &= h \begin{bmatrix} B & D \\ A & C \end{bmatrix} = \begin{bmatrix} A & C \\ B & D \end{bmatrix} \\ &= f_- \begin{bmatrix} A & C \\ B & D \end{bmatrix}, \end{aligned}$$

so $h \circ R_{90} = f_-$. So this is an example of a group that is not abelian, because the commutative property does not always hold. You can check that

$$S = \{e, R_{90}, R_{180}, R_{270}\}$$

is a subgroup of D_4 of index 2. The left cosets of S are

$$e \circ S = R_{90} \circ S = R_{180} \circ S = R_{270} \circ S = S$$

and you can verify that

$$\begin{aligned} h \circ S &= \{h \circ e, h \circ R_{90}, h \circ R_{180}, h \circ R_{270}\} \\ &= \{h, f_-, v, f_+\} \end{aligned}$$

Note that since the cosets partition groups, this was the only choice for the other coset, so really without calculation we can also conclude that

$$h \circ S = f_- \circ S = v \circ S = f_+ \circ S.$$

The **symmetric group** S_n is the set of permutations of $\{1, 2, ..., n\}$. A **permutation** is simply a bijection, which may be thought of as a rearrangement of the numbers. The group operation is the composition of functions. For example, one element α of S_4 satisfies

$$\alpha(1) = 3, \alpha(3) = 2, \alpha(2) = 1, \alpha(4) = 4.$$

Or, we could write it like this:

$$\alpha \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \\ 1 \\ 4 \end{pmatrix}.$$

There is a shorthand notation for permutations called cycle notation. The permutation α can be written as

$$\alpha = (123).$$

This means $1 \mapsto 2$, $2 \mapsto 3$, $3 \mapsto 1$ (the cycle) and $4 \mapsto 4$ since it is not changed. Since the cycle has length 3, it is called a "3-cycle". Let β be the permutation defined by

$$\beta \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} = \begin{pmatrix} 4 \\ 3 \\ 2 \\ 1 \end{pmatrix}.$$

Note that $\beta = (14)(23)$ because it switches 2 and 3 and also switches 1 and 4. Then $\alpha \circ \beta = \alpha\beta$ (people usually drop the $\circ$) satisfies

$$\begin{aligned} \alpha\beta \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} &= \alpha \left(\beta \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} \right) = \alpha \begin{pmatrix} 4 \\ 3 \\ 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 4 \\ 1 \\ 3 \\ 2 \end{pmatrix} \\ &= (142). \end{aligned}$$

On the other hand,

$$\begin{aligned} \beta\alpha \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} &= \beta \left(\alpha \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} \right) = \beta \begin{pmatrix} 2 \\ 3 \\ 1 \\ 4 \end{pmatrix} = \begin{pmatrix} 3 \\ 2 \\ 4 \\ 1 \end{pmatrix} \\ &= (134). \end{aligned}$$

So, again we have an example of a group that is not commutative. Note that we could have done these calculations entirely with cycle notation. For example,

$$\begin{aligned} \beta\alpha &= (14)(23)(123) \\ &= (134). \end{aligned}$$

The thought process: First remember that the permutation on the right goes first. Now we find out what happens to each number through the combination $n \mapsto (123)\, n \mapsto (23)(123)\, n \mapsto (14)(23)(123)$. First look at the number 1. The map (123) changes it to 2. Then (23) changes it to 3. Then (14) does nothing to 3. So 1 maps to 3. So where does 3 go? The map (123) maps 3 to 1, then (23) does nothing to the 1, then (14) maps 1 to 4. Thus 3 maps to 4. Where does 4 go? (123) does not change 4, and then (23) also does nothing to 4, and then (14) maps 4 to 1. So we see we have the cycle (134). Now we already know that 2 must map to itself (nowhere else to go), but we can check: (123) maps 2 to 3, then (23) maps 3 to 2, and then (14) does not change 2, so 2 maps to 2.

The procedure used above can be used to write any permutation as a **product of disjoint cycles**. The word **disjoint** means that the cycles do not have any numbers in common. For example,

$$(125)(34)(78)$$

is a product of disjoint cycles in S_8, but

$$(123)(345)$$

is not a product of disjoint cycles. But we can see that

$$(123)(345) = (12345),$$

which is a "product" of disjoint cycles (only one cycle, but still called a "product"). Here are some facts about disjoint cycles. Note that

$$\begin{aligned} (12)^2 &= e \\ (134)^3 &= (134)(134)(134) = e. \end{aligned}$$

In general, the order of an m-cycle is m, so for example, $|(1276)| = 4$.

Proposition 11.1. *We have the following facts about elements of S_n.*

(1) Every element of S_n can be written as a product $\alpha_1\alpha_2...\alpha_k$ of disjoint cycles.
(2) Disjoint cycles commute.
(3) The order of an m-cycle is m.
(4) The order of a product of disjoint cycles $\alpha_1\alpha_2...\alpha_k$ is the least common multiple

$$\operatorname{lcm}\left\{\left|\alpha_1\right|,\left|\alpha_2\right|,...,\left|\alpha_k\right|\right\}.$$

Proof. (Left as an exercise) □

Here is an example of a calculation. Observe that (1346) is a 4-cycle in S^6. Let's construct the subgroup $\langle(1346)\rangle$ generated by this 4-cycle:

$$\begin{aligned}\langle(1346)\rangle &= \left\{e,(1346),(1346)^2,(1346)^3\right\}\\ &= \{e,(1346),(14)(36),(1643)\}.\end{aligned}$$

12. Introduction to sequences

A **sequence** is a function $a:\mathbb{N}\to X$, where X is a set. In most applications of sequences, the values of sequences are real or complex numbers, so that $X=\mathbb{R}$, or $X=\mathbb{C}$. In other cases we specialize to integer sequences, when $X=\mathbb{Z}$. In the notation of sequences, $a(n)$ is often denoted a_n. Also, since such a function is just an ordered list of elements of X, a sequence can be written as an ordered list $(a_1,a_2,...)=(a_j)_{j\geq 1}$. An examples of the notation:

$$a:\mathbb{N}\to\mathbb{Z}$$

is defined by

$$a(n)=3n^2.$$

Alternately,

$$\begin{aligned}a_j &= 3j^2\\ (a_j)_{j\geq 1} &= \left(3j^2\right)_{j\geq 1}\\ &= \left(3\cdot 1^2,3\cdot 2^2,3\cdot 3^2,...\right).\end{aligned}$$

The formula

$$a_n=3n^2$$

is called a **closed form formula**, because for any n, the value of a_n can be computed independent of knowledge of any other values of the sequence. On the other hand, we sometimes describe sequences using recursive formulas, where a_{n+1} is given in terms of a_n, a_{n-1},.... When describing a sequence using a recursive formula, it is important to give a small set of initial values (such as a_0 and/or a_1) so that the rest of the terms of the sequence can be determined from those initial values alone. For example, the sequence $a_n=3n^2$ can alternatively be described by the recursive formula

$$\begin{aligned}a_1 &= 3 \text{ (initial condition)}\\ a_{n+1} &= a_n+6n+3 \text{ for } n\geq 1\end{aligned}$$

From this we see

$$\begin{aligned} a_2 &= a_1 + 6(1) + 3 = 3 + 6(1) + 3 = 12 \\ a_3 &= 12 + 6(2) + 3 = 27 \\ a_4 &= 27 + 6(3) + 3 = 48 \\ &\dots \end{aligned}$$

Other examples of sequences are as follows:

(1) $\left(0, 1, 0, \frac{1}{2}, 0, \frac{1}{3}, 0, \frac{1}{4}, ...\right)$
Closed form formula:

$$b_n = \begin{cases} 0, & \text{if } n \text{ is odd} \\ \frac{2}{n}, & \text{if } n \text{ is even} \end{cases}$$

Recursive formula:

$$\begin{aligned} b_1 &= 0;\ b_2 = 1; \\ b_{n+2} &= \frac{n}{(n+2)} b_n \text{ for } n \geq 0 \end{aligned}$$

(2) $(1, 2, 6, 24, 120, ...)$
Closed form formula:

$$c_j = j!$$

Recursive formula:

$$\begin{aligned} c_1 &= 1 \\ c_k &= k c_{k-1} \text{ for } k \geq 2 \end{aligned}$$

(3) $(1, -1, 1, -1, ...)$
Closed form formula:

$$d_m = (-1)^{m+1}$$

Recursive formula:

$$\begin{aligned} d_1 &= 1 \\ d_n &= -d_{n-1} \text{ for } n \geq 2. \end{aligned}$$

Here is a famous example of a sequence, called the **Fibonacci sequence**:

$$\begin{aligned} F_1 &= F_2 = 1; \\ F_{n+1} &= F_n + F_{n-1} \text{ for } n \geq 2 \end{aligned}$$

We see that

$$(F_n)_{n \geq 1} = (1, 1, 2, 3, 5, 8, 13, 21, 34, 55, ...).$$

There is a closed form formula for the Fibonacci sequence, but at this point we don't have the technology to prove it is correct. For your entertainment, here it is:

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

Another sequence that comes out of the Fibonacci sequence is the sequence of ratios of successive terms. That is,

$$R_n = \frac{F_{n+1}}{F_n} \text{ for } n \geq 1$$

So

$$\begin{aligned}(R_n)_{n\geq 1} &= \left(\frac{1}{1}, \frac{2}{1}, \frac{3}{2}, \frac{5}{3}, \frac{8}{5}, \frac{13}{8}, \frac{21}{13}, \frac{34}{21}...\right) \\ &= \left(1, 2, 1.5, 1.\overline{6}, 1.6, 1.625, 1.61538, 1.619047, ...\right)\end{aligned}$$

We will later show that the ratios **converge** to the golden ratio

$$\frac{1+\sqrt{5}}{2} = 1.\,618\,033\,988\,7....$$

Another famous sequence is the following, known as the **3x + 1 sequence**. The sequence has an initial value that starts at any positive integer x_1. Then

$$x_{n+1} = \begin{cases} 3x_n + 1, & \text{if } x_n \text{ is odd} \\ \frac{x_n}{2}, & \text{if } x_n \text{ is even} \end{cases}$$

For example, for the initial values $x_1 = 7$ or $x_1 = 15$, we have

$$\begin{aligned}(x_n)_{n\geq 1} &= (7, 22, 11, 34, 17, 52, 26, 13, 40, 20, 10, \\ &\qquad 5, 16, 8, 4, 2, 1, 4, 2, 1, \overline{4, 2, 1}) \\ (x_n)_{n\geq 1} &= (15, 46, 23, 70, 35, 106, 53, 160, 80, 40, 20, 10, \\ &\qquad 5, 16, 8, 4, 2, 1, 4, 2, 1, \overline{4, 2, 1}).\end{aligned}$$

The following is an unsolved problem in mathematics.

Conjecture 12.1. *(**The 3x+1 Conjecture**) If x_1 is chosen to be any positive integer, the resulting $3x + 1$ sequence eventually becomes the cycle $\overline{4, 2, 1}$.*

The following are two well-known types of sequences.

A **arithmetic sequence** is a sequence where the next number is a constant (usually an integer) added to the previous number. That is, for given $k, m \in \mathbb{N}$, the associated arithmetic sequence

$$\begin{aligned}z_1 &= k \\ z_n &= z_{n-1} + m \text{ for } n \geq 1\end{aligned}$$

may also be written in closed form as

$$z_n = k + (n-1)\, m \text{ for } n \geq 1.$$

A **geometric sequence** is a sequence where the next number is a constant times the previous number. For example, if $y_0 = a$ and

$$y_k = r y_{k-1} \text{ for } k \geq 1,$$

and this sequence can be written in closed form as

$$y_k = a r^k \text{ for } k \geq 0.$$

13. Limits of Sequences

When we say that a sequence $(a_n)_{n\geq 1}$ converges, intuitively it means that the numbers a_n get closer and closer to a limit L as n increases, and we write

$$\lim_{n\to\infty} a_n = L.$$

Turning this idea into a precise statement in mathematics is not an easy task. Here is the formal definition of limit.

Definition 13.1. *Given a sequence $(a_n)_{n\geq 1}$ of real numbers and a real number L, we say that*

$$\lim_{n\to\infty} a_n = L$$

if and only if

For every $\varepsilon > 0$, there exists an $N > 0$ such that whenever $n \geq N$,

$$|a_n - L| < \varepsilon.$$

*When this occurs, L is called the **limit** of the sequence a_n.*

Remark 13.2. *Often we will denote $\lim_{n\to\infty} a_n$ by simply $\lim a_n$. When a limit of a sequence (a_n) exists, we say that (a_n) **converges**.*

The picture of this definition is that the limit L is on the number line, and the numbers $a_1, a_2, a_3, ...$ are also on that line, and ε is a small (and arbitrarily chosen) distance from L, as seen in the interval $(L-\varepsilon, L+\varepsilon)$. What must be true is that after some point in the sequence, say the N^{th} term, **all** the terms a_n with $n \geq N$ must be within that interval $(L-\varepsilon, L+\varepsilon)$. So, no matter how small $\varepsilon > 0$ is chosen, there is some N so that the terms a_n for $n \geq N$ are **all** within ε of the limit L.

The general strategy for proving that a limit exists is:

(1) Compute the limit using calculus skills.
(2) Start off with some scratchwork: Assume we are given $\varepsilon > 0$, and set up the inequality to be proved:
$$|a_n - L| < \varepsilon.$$
(3) Then, continuing the scratchwork, use algebra and inequality skills to figure out how big the n has to be to make that inequality above work. In other words, you shoot for
$$n > \text{some function of } \varepsilon = N.$$
(4) Finally, work backwords to write a rigorous proof. That is, assume we are given $\varepsilon > 0$, then let N be that expression involving ε above. Then retrace your steps to finally get to
$$|a_n - L| < \varepsilon.$$
If you do get to that point, you have completed the proof.

In working with inequalities, remember the important concepts:

(1) If
$$A < B$$
is true and $C \in R$, then also
$$\begin{aligned} A + C &< B + C \text{ and} \\ A - C &< B - C \end{aligned}$$
are true, and in fact because these changes are reversible, both of these new inequalities are equivalent to the original one.
(2) If
$$A < B$$
is true and $C > 0$, then
$$AC < BC$$
is also true (and the new inequality is equivalent to the old one). If $D < 0$, then
$$AD > BD$$

is also true (and the new inequality is equivalent to the old one).

(3) Same as (2) above, but when dividing by either positive or negative numbers.

(4) If

$$0 < A < B,$$

then

$$\frac{1}{A} > \frac{1}{B}.$$

We now work out some examples.

Example 13.3. *Prove that* $\lim_{n\to\infty} \frac{n}{n+2} = 1$.

Scratchwork: *We want to say that for any* $\varepsilon > 0$ *given, we can choose the* n *large enough so that*

$$\left| \frac{n}{n+2} - 1 \right| < \varepsilon.$$

So we need to figure out what that means. The inequality can be rewritten

$$-\varepsilon < \frac{n}{n+2} - 1 < \varepsilon,$$

or

$$\begin{aligned} -\varepsilon &< \frac{n}{n+2} - 1 \qquad (1) \\ &\text{and} \\ \frac{n}{n+2} - 1 &< \varepsilon \qquad (2) \end{aligned}$$

Working with (1) first, we add 1 *to both sides and then multiply by* $n+2$ *(a positive real number!):*

$$\begin{aligned} 1 - \varepsilon &< \frac{n}{n+2} \\ (1-\varepsilon)(n+2) &< n \\ (1-\varepsilon)n + 2(1-\varepsilon) &< n \end{aligned}$$

Subtracting $(1-\varepsilon)n$ *from both sides:*

$$\begin{aligned} 2(1-\varepsilon) &< n - (1-\varepsilon)n \\ &= n - n + \varepsilon n = \varepsilon n \end{aligned}$$

And so

$$\frac{2(1-\varepsilon)}{\varepsilon} < n \qquad (3)$$

We'll remember this later. Now let's work with the other inequality (2), using similar techniques:

$$\begin{aligned} \frac{n}{n+2} - 1 &< \varepsilon \\ \frac{n}{n+2} &< \varepsilon + 1 \\ n &< (\varepsilon+1)(n+2) \\ &= (\varepsilon+1)n + 2(\varepsilon+1) \\ -2(\varepsilon+1) &< (\varepsilon+1)n - n = \varepsilon n \\ -\frac{2(\varepsilon+1)}{\varepsilon} &< n. \end{aligned}$$

Note that this inequality is automatically true for all $n \geq 1$. Actually, we could have seen that from the top of the above calculation because $\frac{n}{n+2} - 1$ must be negative! So (3) tells us that the N we will need to use in our proof is some N such that

$$N > \frac{2(1-\varepsilon)}{\varepsilon}.$$

Now, time for the ***actual proof****.*

Proof. For any $\varepsilon > 0$, choose any positive number N so that

$$N > \frac{2(1-\varepsilon)}{\varepsilon}.$$

Then, if $n \geq N$,

$$n > \frac{2(1-\varepsilon)}{\varepsilon},$$

so

$$\varepsilon n = n - (1-\varepsilon) n > 2(1-\varepsilon).$$

Then

$$n > (1-\varepsilon) n + 2(1-\varepsilon) = (1-\varepsilon)(n+2),$$

so

$$\frac{n}{n+2} > 1 - \varepsilon,$$

or

$$\frac{n}{n+2} - 1 > -\varepsilon.$$

Also, since $n < n+2$, $\frac{n}{n+2} < 1$, so

$$\frac{n}{n+2} - 1 < 0 < \varepsilon.$$

Thus,

$$-\varepsilon < \frac{n}{n+2} - 1 < \varepsilon,$$

or

$$\left| \frac{n}{n+2} - 1 \right| < \varepsilon.$$

Therefore, by the definition of limit,

$$\lim_{n\to\infty} \frac{n}{n+2} = 1.$$

□

Example 13.4. *Find* $\lim_{n\to\infty} \frac{4n^2 - 2n\cos(2n) + 1}{n^2}$ *and prove it exists.*

*[**Scratchwork:** We see that $-1 \leq \cos(2n) \leq 1$, so the numerator is between $4n^2 - 2n + 1$ and $4n^2 + 2n + 1$, which is approximately $4n^2$ for large n. So the limit should be 4. So we need to prove*

$$\left| \frac{4n^2 - 2n\cos(2n) + 1}{n^2} - 4 \right| < \varepsilon,$$

or

$$-\varepsilon < \frac{4n^2 - 2n\cos(2n) + 1}{n^2} - 4 < \varepsilon$$

Since $|\cos(2n)| \leq 1$, *it is sufficient to show*

$$-\varepsilon < \frac{4n^2 - 2n + 1}{n^2} - 4 \text{ and } \frac{4n^2 + 2n + 1}{n^2} - 4 < \varepsilon,$$

or

$$-\varepsilon < \frac{-2n + 1}{n^2} \text{ and } \frac{2n + 1}{n^2} < \varepsilon.$$

For the ***LHS inequality it is sufficient that we show:***

$$-\varepsilon < \frac{-2n}{n^2} = -\frac{2}{n} < \frac{-2n + 1}{n^2},$$

so

$$-n < -\frac{2}{\varepsilon}, \text{ i.e. } n > \frac{2}{\varepsilon}.$$

For the ***RHS inequality it is sufficient to show:***

$$\frac{2n + 1}{n^2} < \frac{2n + n}{n^2} = \frac{3}{n} < \varepsilon,$$

or

$$n > \frac{3}{\varepsilon}.$$

So now we are ready for the real proof.

Proof. For any $\varepsilon > 0$, let N be any real number such that $N > \frac{3}{\varepsilon}$. Then, if $n \geq N$

$$\begin{aligned} n &> \frac{3}{\varepsilon}, \text{ so since } \frac{\varepsilon}{n} > 0 \\ \varepsilon &> \frac{3}{n} = \frac{2n + n}{n^2} > \frac{-2n\cos(2n) + 1}{n^2} = \frac{4n^2 - 2n\cos(2n) + 1}{n^2} - 4. \end{aligned}$$

Also,

$$\begin{aligned} n &> \frac{3}{\varepsilon} > \frac{2}{\varepsilon}, \text{ so since } -\frac{\varepsilon}{n} < 0, \\ -\varepsilon &< -\frac{2}{n} = -\frac{2n}{n^2} < \frac{4n^2 - 2n + 1}{n^2} - 4 \\ &< \frac{4n^2 - 2n\cos(2n) + 1}{n^2} - 4. \end{aligned}$$

Thus,

$$-\varepsilon < \frac{4n^2 - 2n\cos(2n) + 1}{n^2} - 4 < \varepsilon,$$

so

$$\left| \frac{4n^2 - 2n\cos(2n) + 1}{n^2} - 4 \right| < \varepsilon.$$

Therefore,

$$\lim_{n \to \infty} \frac{4n^2 - 2n\cos(2n) + 1}{n^2} = 4.$$

□

The following squeeze theorem is often useful in computing limits.

Theorem 13.5. ***(Squeeze Theorem)*** *Suppose that* $(a_n), (b_n), (c_n)$ *are three sequences such that for some* $N > 0$, $a_n \leq b_n \leq c_n$ *for all* $n \geq N$, *and suppose that*

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty} c_n = L.$$

Then $\lim_{n\to\infty} b_n$ *exists and*

$$\lim_{n\to\infty} b_n = L.$$

Proof. By the given information , for any given $\varepsilon > 0$, there exists N_1 such that for all $n \geq N$

$$|a_n - L| < \varepsilon, \text{ or } -\varepsilon < a_n - L < \varepsilon,$$

and there exists N_2 such that for all $m \geq N_2$,

$$|c_m - L| < \varepsilon, \text{ or } -\varepsilon < c_m - L < \varepsilon.$$

Then let

$$N_3 = \max\{N_1, N_2, N\}.$$

Then for $n \geq N_3$,

$$\begin{aligned} -\varepsilon &< a_n - L \leq b_n - L \\ &\leq c_n - L < \varepsilon, \end{aligned}$$

so

$$|b_n - L| < \varepsilon.$$

Thus, $\lim_{n\to\infty} b_n$ exists and

$$\lim_{n\to\infty} b_n = L.$$

□

It is interesting to think about how to **show a limit does not exist**. This is an interesting logical exercise. To understand this, it is first important to understand how to **negate statements with quanitfiers**. That is, if a statement is of the form similar to:

$$\begin{aligned} \text{statement } P &\equiv \forall A \exists B \text{ s.t.} \forall C \forall D \exists E \text{ s.t. } F \\ \text{statement } P &\equiv \text{“For all } A \text{, there exists } B \text{ such that for all } C \text{ and for all } D, \\ &\quad \text{there exists } E \text{ such that } F \text{ is true.”} \end{aligned}$$

Then the **negation is:**

$$\begin{aligned} \text{negation of statement } P &\equiv \text{not } P \equiv \neg P \equiv \text{ “}P \text{ is false.”} \\ &\equiv \exists A \text{ s.t. } \forall B \exists C \text{ s.t. } \exists D \text{ s.t. } \forall E.\ \neg F \\ &\equiv \text{“There exists } A \text{ such that for all } B \text{, there exists } C \text{ such that} \\ &\quad \text{there exists } D \text{ such that for all } E, F \text{ is false.”} \end{aligned}$$

In other words when we move the “NOT” symbol $\neg$ past a quantifier (like “$\forall$” or “$\exists$___ s.t.”), the quantifier switches to the other type.

Now, let's apply this to the definition of limit. Here is the logical version of the definition of limit:

$$\lim_{n\to\infty} x_n = L \Leftrightarrow \forall \varepsilon > 0, \exists N > 0 \text{ s.t. } \forall n \geq N,\ |a_n - L| < \varepsilon.$$

Then the negation would be

$$\lim_{n\to\infty} x_n \neq L \Leftrightarrow \exists \varepsilon > 0 \text{ s.t.} \forall N > 0,\ \exists n \geq N \text{ s.t. } |a_n - L| \geq \varepsilon.$$

If we are just trying to show that a limit of a sequence does not exist, then we must prove

$$\begin{aligned}\lim_{n\to\infty} a_n \text{ does not exist } &\Leftrightarrow \forall L \in \mathbb{R}, \lim_{n\to\infty} x_n \neq L \\ &\Leftrightarrow \forall L \in \mathbb{R}, \exists \varepsilon > 0 \text{ s.t.} \forall N > 0, \exists n \geq N \text{ s.t. } |a_n - L| \geq \varepsilon.\end{aligned}$$

Let's now apply this logical discussion in the following example. Another useful tool we will recall are the various versions of the triangle inequality: For all $x, y \in \mathbb{C}$ (or $\mathbb{R}$ in our case)

$$\begin{aligned}|x+y| &\leq |x|+|y| \\ |x-y| &\leq |x|+|y| \\ |x+y| &\geq |x|-|y| \\ |x-y| &\geq |x|-|y|\end{aligned}$$

Example 13.6. *Prove that* $\lim_{n\to\infty} \frac{(-1)^n}{5}$ *does not exist.*

Proof. For all $L \in \mathbb{R}$, let $\varepsilon = \frac{1}{5}$. Then for all $N > 0$, for any odd $n \geq N$,

$$\left|\frac{(-1)^n}{5} - L\right| = \left|\frac{-1}{5} - L\right|,$$

and for any even $n \geq N$,

$$\left|\frac{(-1)^n}{5} - L\right| = \left|\frac{1}{5} - L\right|.$$

Then

$$\frac{2}{5} = \left|\frac{1}{5} - L - \left(\frac{-1}{5} - L\right)\right| \leq \left|\frac{1}{5} - L\right| + \left|\frac{-1}{5} - L\right|,$$

by the triangle inequality. At least one of these $\left|\frac{1}{5} - L\right| + \left|\frac{-1}{5} - L\right|$ must be $\geq \frac{1}{5}$ in order that the equation is true, so there does exist $n \geq N$ such that

$$\left|\frac{(-1)^n}{5} - L\right| \geq \varepsilon = \frac{1}{5}.$$

Therefore, for all $L \in \mathbb{R}$, $\lim_{n\to\infty} \frac{(-1)^n}{5} \neq L$. Thus, $\lim_{n\to\infty} \frac{(-1)^n}{5}$ does not exist. □

14. Properties of Limits and the Monotone Convergence Theorem

The following proposition states many properties of limits that we often use in calculus.

Proposition 14.1. *Let* (a_n) *and* (b_n) *be two sequences in* $\mathbb{R}$ *(or* $\mathbb{C}$*). Assume that* $\lim_{n\to\infty} a_n$ *and* $\lim_{n\to\infty} b_n$ *both exist. Let* $c \in \mathbb{R}$ *(or* $\mathbb{C}$*) be a nonzero constant.*

(1) $\lim_{n\to\infty} (a_n + b_n) = \lim_{n\to\infty} a_n + \lim_{n\to\infty} b_n$.
(2) $\lim_{n\to\infty} (a_n - b_n) = \lim_{n\to\infty} a_n - \lim_{n\to\infty} b_n$.
(3) $\lim_{n\to\infty} (a_n b_n) = \lim_{n\to\infty} a_n \lim_{n\to\infty} b_n$.
(4) $\lim_{n\to\infty} \left(\frac{a_n}{b_n}\right) = \frac{\lim_{n\to\infty} a_n}{\lim_{n\to\infty} b_n}$, *if* $\lim_{n\to\infty} b_n \neq 0$.
(5) $\lim_{n\to\infty} (ca_n) = c \lim_{n\to\infty} a_n$

[**Scratchwork for #4: if** $\lim_{n\to\infty} a_n = L$ and $\lim_{n\to\infty} b_n = P \neq 0$, assume $P > 0$: we want $\lim_{n\to\infty} \frac{a_n}{b_n} = \frac{L}{P}$:

$$\begin{aligned} -\varepsilon &< \frac{a_n}{b_n} - \frac{L}{P} < \varepsilon \\ -\varepsilon &< \frac{Pa_n - Lb_n}{Pb_n} < \varepsilon \\ -\varepsilon &< \frac{P(a_n - L) - L(b_n - P)}{Pb_n} < \varepsilon \end{aligned}$$

So we can force this to happen if we force $|a_n - L| < \frac{P(P/2)\varepsilon}{2P}$, $|b_n - P| < \frac{P(P/2)\varepsilon}{2|L|}$ and $|b_n - P| < \frac{P}{2}$.]

Proof. (Most of this left as exercises. Some of these proofs can be quite tricky! We prove one case as an example.) Suppose that $\lim_{n\to\infty} a_n = L$ and $\lim_{n\to\infty} b_n = P \neq 0$. For the purposes of this proof, we assume $P > 0$; simple modifications would take care of the case where $P < 0$. Then for any $\varepsilon > 0$, there exists $N_1 > 0$ such that for all $n \geq N_1$,

$$|a_n - L| < \frac{P(P/2)\varepsilon}{2P},$$

and there exists $N_2 > 0$ such that for all $m \geq N_2$,

$$|b_m - P| < \min\left\{\frac{P(P/2)\varepsilon}{2|L|}, \frac{P}{2}\right\}$$

Since $|b_m - P| < \frac{P}{2}$, note that for $m \geq N_2$,

$$b_m - P < -\frac{P}{2} \text{ so that } b_m > \frac{P}{2} > 0.$$

Then, letting $N = \max\{N_1, N_2\}$, for all $n \geq N$,

$$\begin{aligned} \left|\frac{a_n}{b_n} - \frac{L}{P}\right| &= \left|\frac{P(a_n - L) - L(b_n - P)}{Pb_n}\right| \leq \left|\frac{P(a_n - L)}{Pb_n}\right| + \left|\frac{L(b_n - P)}{Pb_n}\right| \\ &\leq \frac{P}{Pb_n}|a_n - L| + \frac{|L|}{Pb_n}|b_n - P| \leq \frac{P}{P(P/2)}|a_n - L| + \frac{|L|}{P(P/2)}|b_n - P| \\ &< \frac{P}{P(P/2)}\frac{P(P/2)\varepsilon}{2P} + \frac{|L|}{P(P/2)}\frac{P(P/2)\varepsilon}{2|L|} = \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \end{aligned}$$

Therefore,

$$\lim_{n\to\infty}\left(\frac{a_n}{b_n}\right) = \frac{L}{P}.$$

□

Remark 14.2. *The requirement that* $\lim_{n\to\infty} a_n$ *and* $\lim_{n\to\infty} b_n$ *both exist is important.*

Here is another example of algebraic properties of limits.

Lemma 14.3. *(1) Suppose that* (a_n) *is a sequence of real numbers, and suppose that* $\lim a_n$ *exists. Then the sequence* (a_n^2) *converges as well, and*

$$\lim a_n^2 = (\lim a_n)^2.$$

(2) *Suppose that* (b_n) *is a sequence of positive real numbers, suppose that* $\lim b_n$ *exists. Then the sequence* $\left(\sqrt{b_n}\right)$ *converges as well, and*

$$\lim \sqrt{b_n} = \sqrt{\lim b_n}.$$

Proof. (1) This is just a special case of the previous proposition.
(2) Suppose that $b_k > 0$ for all k, and $\lim b_n = L \geq 0$. First consider the case where $L = 0$. Then, given any $\varepsilon > 0$, let $\varepsilon_2 = \min\{\varepsilon^2, 1\} \leq \varepsilon^2$. Then $\exists N > 0$ such that for all $n \geq N$,

$$|b_n| < \varepsilon_2 \leq 1.$$

Then $\left|\sqrt{b_n}\right| < \sqrt{\varepsilon_2} \leq \varepsilon$, so $\lim \sqrt{b_n} = 0$ also.
On the other hand, if $L > 0$, then given any $\varepsilon > 0$, $\exists N > 0$ such that for all $n \geq N$,

$$|b_n - L| < \varepsilon' = \varepsilon\sqrt{L}.$$

Then

$$\left|\left(\sqrt{b_n} - \sqrt{L}\right)\right|\left(\sqrt{b_n} + \sqrt{L}\right) = \left|\left(\sqrt{b_n} - \sqrt{L}\right)\left(\sqrt{b_n} + \sqrt{L}\right)\right| = |b_n - L| < \varepsilon\sqrt{L},$$

so

$$\left|\left(\sqrt{b_n} - \sqrt{L}\right)\right| < \frac{\varepsilon\sqrt{L}}{\sqrt{b_n} + \sqrt{L}} < \frac{\varepsilon\sqrt{L}}{\sqrt{L}} = \varepsilon.$$

Thus,

$$\lim \sqrt{b_n} = \sqrt{L}.$$

□

We now discuss monotone sequences.

Definition 14.4. *A sequence* $(a_n)_{n\geq 1}$ *is called*

- ***increasing*** *if* $a_{n+1} \geq a_n$ *for all* $n \in \mathbb{N}$
- ***strictly increasing*** *if* $a_{n+1} > a_n$ *for all* $n \in \mathbb{N}$
- ***decreasing*** *if* $a_{n+1} \leq a_n$ *for all* $n \in \mathbb{N}$
- ***strictly decreasing*** *if* $a_{n+1} < a_n$ *for all* $n \in \mathbb{N}$
- ***monotone*** *if* (a_n) *is either increasing or decreasing*
- ***strictly monotone*** *if* (a_n) *is either strictly increasing or strictly decreasing*

Definition 14.5. *An* ***upper bound*** *for a set* $S \subseteq \mathbb{R}$ *is a real number* M *such that* $x \leq M$ *for all* $x \in S$. *A* ***lower bound*** *for a set* $S \subseteq \mathbb{R}$ *is a real number* D *such that* $x \geq D$ *for all* $x \in S$.

Definition 14.6. *A* ***least upper bound (or supremum)*** *of a set* $S \subseteq \mathbb{R}$ *is an upper bound* U *such that* $M \geq U$ *for every upper bound* M *of* S. *[Note that a least upper bound does not necessarily exist.] A* ***greatest lower bound (or infimum)*** *of a set* $S \subseteq \mathbb{R}$ *is an lower bound* L *such that* $L \geq T$ *for every lower bound* T *of* S. *[Again, a greatest lower bound need not exist.]*

One of the definitions of real numbers uses the following axiom as the starting point.

Axiom 14.7. *(****Least Upper Bound Axiom****, or* ***Completeness Axiom****) Any subset* $S \subseteq \mathbb{R}$ *that has an upper bound has a least upper bound.*

Remark 14.8. *If the LUB Axiom is given, automatically a similar fact is true for sets that have lower bounds. That is, if a set* $A \subseteq \mathbb{R}$ *has a lower bound* T, *then the set* $B = \{-x : x \in A\}$ *satisfies* $y \leq T$ *for all* $y \in B$. *By the LUB Axiom,* B *has a supremum* U, *and then you can show that* $-U$ *is an infimum for* A.

Example 14.9. *The sets* $\mathbb{Z}$, $\mathbb{Q}$. $\mathbb{R}$ *have no upper or lower bounds, but* $\mathbb{N}$ *has lower bounds such as* -5, -1.3, 0.2. *The greatest lower bound of* $\mathbb{N}$ *is* 1.

Example 14.10. *The set* $(0,1)$ *has upper bounds* 3.8, 67849682. *Its supremum is* 1.

Now, we are ready for the monotone convergence theorem.

Theorem 14.11. *(**Monotone Convergence Theorem**) Every bounded monotone sequence converges.*

Proof. We will prove this in the case where the sequence is increasing. The proof can be easily modified for the case of decreasing sequences. Let $(a_n)_{n\geq 1}$ be an increasing sequence such that $a_n \leq M$ for some $M \in \mathbb{R}$. Let L be the least upper bound of $\{a_n : n \in \mathbb{N}\}$. We will now show $\lim_{n\to\infty} a_n = L$. Given any $\varepsilon > 0$, I claim that there exists $N \in \mathbb{N}$ such that $a_N < L - \varepsilon$. [If not, then $L - \varepsilon$ would also be an upper bound for the sequence but smaller than L, a contradiction.] Then, for any $n \geq N$,

$$L - \varepsilon < a_N \leq a_n \leq L < L + \varepsilon,$$

so

$$|a_n - L| < \varepsilon.$$

Therefore, $\lim_{n\to\infty} a_n = L$. □

As an application, we now find the limit of a sequences.

Example 14.12. *Prove that the sequence* (b_n) *defined by* $b_1 = 1$; $b_{n+1} = \sqrt{6 + b_n}$ *converges, and find the limit.*

Proof. First of all, observe that $0 < b_n < 100$ for all n: we use induction to prove it. Clearly, this is true for $b_1 = 1$. Assume we have proved this for b_k for some $k \in \mathbb{N}$. Then $b_{k+1} = \sqrt{6 + b_k} > \sqrt{6} > 0$ and $b_{k+1} < \sqrt{6 + 100} = \sqrt{106} < 11 < 100$. By induction, $0 < b_n < 100$ for all $n \in \mathbb{N}$. Next, we prove by induction that (b_n) is a strictly increasing sequence. Observe that $b_2 = \sqrt{6+1} > 1 = b_1$. Assume that $b_{j+1} > b_j$ for some $j \geq 1$. Then

$$\begin{aligned} b_{j+2}^2 - b_{j+1}^2 &= \left(\sqrt{6 + b_{j+1}}\right)^2 - \left(\sqrt{6 + b_j}\right)^2 \\ &= 6 + b_{j+1} - 6 + b_j = b_{j+1} - b_j > 0, \end{aligned}$$

so since b_{j+1} and b_{j+2} are positive, $b_{j+2}^2 > b_{j+1}^2$ implies $b_{j+2} > b_{j+1}$. By induction (b_n) is strictly increasing. Thus, we have shown that (b_n) is bounded and monotone, so it converges to a limit L. From the recursion formula and the algebraic limit properties,

$$\begin{aligned} L &= \lim b_{n+1} = \lim \sqrt{6 + b_n} \\ &= \sqrt{\lim (6 + b_n)} = \sqrt{6 + \lim b_n} \\ &= \sqrt{6 + L}, \end{aligned}$$

so $L^2 = 6 + L$, or $L^2 - L - 6 = 0 = (L - 3)(L + 2)$. Thus, the limit is either 3 or -2. Since $b_n > 0$ for all n, -2 is not possible, so

$$\lim b_n = 3.$$

□