Pure Mathematics

Chapter 1

Number Theory

(tags: number theory)

1.1 Natural Numbers

(tags: number theory)

1.1.1 Peano Axioms

(tags: number theory)

Axiom 1. *Closure under addition:*
For all $a, b \in \mathbb{N}$ *we have* $a + b \in \mathbb{N}$.

Axiom 2. *Closure under multiplication:*
For all $a, b \in \mathbb{N}$ *we have* $a \times b \in \mathbb{N}$.

Axiom 3. *Commutative Law for addition:*
For all $a, b \in \mathbb{N}$ *we have* $a + b = b + a$.

Axiom 4. *Associative Law for addition:*
For all $a, b, c \in \mathbb{N}$ *we have* $(a + b) + c = a + (b + c)$.

Axiom 5. *Commutative Law for multiplication:*
For all $a, b \in \mathbb{N}$ *we have* $a \times b = b \times a$.

Axiom 6. *Associative Law for multiplication:*
For all $a, b, c \in \mathbb{N}$ *we have* $(a \times b) \times c = a \times (b \times c)$.

Axiom 7. *Multiplicative Identity:*
There is a special element of $\mathbb{N}$*, denoted by 1, which has the property that for all* $n \in \mathbb{N},\ n \times 1 = n$.

Axiom 8. *Additive cancellation:*
For all $a, b, c \in \mathbb{N}$ *if* $a + c = b + c$ *then* $a = b$.

Axiom 9. *Multiplicative cancellation:*
For all $a, b, c \in \mathbb{N}$ *if* $a \times c = b \times c$ *then* $a = b$.

Axiom 10. *Distributive Law:*
For all $a, b, c \in \mathbb{N},\ a \times (b + c) = (a \times b) + (b \times c)$.

Axiom 11. *Definition of "less than":*
For all $a, b \in \mathbb{N},\ a < b$ *if and only if there is some* $c \in \mathbb{N}$ *s.t.* $a + c = b$.

Axiom 12. *Trichotomous property:*
For all $a, b \in \mathbb{N}$ *exactly one of the following is true:* $a = b,\ a < b,\ b < a$.

Notation. We also write ab for $a \times b$.

Properties following from these axioms

Proposition 1. *If* $a, b \in \mathbb{N}$ *satisfy* $a \times b = a$*, then* $b = 1$.

Proof.

$$\begin{aligned} & a \times b = a = a \times 1 && \text{by Multiplicative Identity axiom} \\ \iff\ & b \times a = 1 \times a && \text{by Commutative Law for multiplication} \\ \iff\ & b = 1 && \text{by Multiplicative cancellation} \end{aligned}$$

□

Proposition 2. *If* $a, b, c \in \mathbb{N}$ *and* $a < b$ *then* $a \times c < b \times c$.

Proof.

$$
\begin{aligned}
& a < b \implies a + d = b \text{ for some } d \in \mathbb{N} && \text{by Definition of "less than"} \\
\therefore \quad & b \times c = (a + d) \times c = (a \times c) + (d \times c) && \text{by Distributive Law} \\
\therefore \quad & a \times c < (a \times c) + (d \times c) = b \times c && \text{by defn. "less than" and closure}
\end{aligned}
$$

□

Proposition 3. 1 *is the least element of* $\mathbb{N}$.

Proof. Assume m is the least element of $\mathbb{N}$. Then, also $m < 1$. So, by Proposition 2,

$$m < 1 \implies m \times m < 1 \times m = m$$

But, closure of multiplication and $m \times m < m$ together contradict the assumption that m is the least element of $\mathbb{N}$.
Therefore m cannot be less than 1. Since we know that $1 \in \mathbb{N}$ and that the minimum element of $\mathbb{N}$, m, cannot be less than 1, it follows that 1 must be the minimum element of $\mathbb{N}$ and $m = 1$. □

1.1.2 Integers

(tags: number theory)

1.1.3 Odd and Even Numbers

(tags: number theory)

Definition. *An* ***even*** *number,* $n \in \mathbb{Z}$*, is one that satisfies,*

$$\exists\, m \in \mathbb{Z} \;\cdot\; n = 2m$$

Definition. *An* ***odd*** *number,* $n \in \mathbb{Z}$*, is one that satisfies,*

$$\exists\, m \in \mathbb{Z} \;\cdot\; n = 2m + 1$$

Consequences

Sum of even numbers, $m + n$:

$$\begin{aligned} m + n &= 2k + 2l \quad \text{where } k, l \in \mathbb{Z} && \text{by defn. of even no.s } m, n \\ &= 2(k + l) \\ &= 2q \quad \text{where } q \in \mathbb{Z} \end{aligned}$$

So $m + n$ is also even. However, if $m + n$ is even:

$$\begin{aligned} m + n &= 2k \quad \text{where } k \in \mathbb{Z} && \text{by defn. of even } m + n \\ k &= \frac{m}{2} + \frac{n}{2} \end{aligned}$$

So m and n are not necessarily even. A counterexample is

$$3 + 5 = 8 \iff \frac{3}{2} + \frac{5}{2} = 4$$

To summarize:

- m, n even $\implies$ $m + n$ even
- $m + n$ even $\implies$ m, n even Wrong!

1.1.4 The Fundamental Theorem of Arithmetic

(tags: number theory)

Definition of prime number: An integer that is only divided cleanly by itself and one. More formally, an integer, p, is prime if it is greater than 1 and,

$$\exists! \, m, n \in \mathbb{Z} \; \cdot \; \frac{p}{m} = n \wedge (m \neq p \wedge m \neq 1)$$

Primality $\implies$ Unique Prime Factorization:

> "Any number either is prime or is measured by some prime number."
> *Euclid, Elements Book VII, Proposition 32*

So, if an integer n is not prime then,

$$\begin{aligned} \exists \, a, b \in \mathbb{Z} \; \cdot \; \frac{n}{a} &= b \\ \iff n &= ab \end{aligned}$$

Then, for a (the same applies to b),

$$\begin{aligned} \exists \, c, d \in \mathbb{Z} \, , \, c, d \notin \{1, a\} \; \cdot \; \frac{n}{a} &= b \\ \iff n &= cd \end{aligned}$$

We can continue to descend like this until we must eventually encounter one or more primes. Furthermore, if a number, n, has a prime factorization, $p_1 p_2$ then,

$$n = p_1 p_2 = p_3 p_4 \iff \frac{p_1}{p_3} = \frac{p_4}{p_2} = n$$

But $\frac{p_1}{p_3} = n$ contradicts the definition of primeness of p_1. Therefore prime factorizations are unique.

Proof of existence

Proof. It must be shown that every integer greater than 1 is either prime or a product of primes. First, 2 is prime. Then, by strong induction, assume this is true for all numbers greater than 1 and less than n. If n is prime, there is nothing more to prove. Otherwise, there are integers a, b where $n = ab$, and $1 < a \leq b < n$. By the induction hypothesis, $a = p_1 p_2 ... p_j$ and $b = q_1 q_2 ... q_k$ are products of primes. But then $n = ab = p_1 p_2 ... p_j q_1 q_2 ... q_k$ is a product of primes. □

Proof of uniqueness

Proof. Suppose, to the contrary, that there is an integer that has two distinct prime factorizations. Let n be the least such integer and write $n = p_1 p_2 ... p_j = q_1 q_2 ... q_k$, where each p_i and q_i is prime. (Note that j and k are both at least 2.) We see that p_1 divides $q_1 q_2 ... q_k$, so p_1 divides some q_i by Euclid's lemma. Without loss of generality, say that p_1 divides q_1. Since p_1 and q_1 are both prime, it follows that $p_1 = q_1$. Returning to our factorizations of n, we may cancel these two terms to conclude that $p_2 ... p_j = q_2 ... q_k$. We now have two distinct prime factorizations of some integer strictly smaller than n, which contradicts the minimality of n. □

1.1.5 Modular Arithmetic

(tags: number theory)

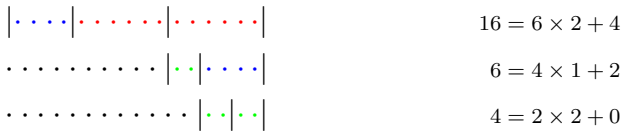
Greatest Common Divisor (also called Highest Common Factor)

Definition. *The* ***Greatest Common Divisor (gcd)*** *of two integers – say* a *and* b *– is an integer* d *that satisfies,*

$$d = z_1 a + z_2 b$$

for some $z_1, z_2 \in \mathbb{Z}$*. This means that, whenever we are adding or subtracting multiples of the two numbers* a *and* b*, the result will always be a multiple of* d *and, therefore also,* d *is the smallest such result obtainable.*

The greatest common divisor of 16 and 6 can be visualized as follows:



This implies the algorithm:

gcd(a, b) :
    if $b == 0$ then
        return a
    else
        return **gcd**$(b, a \bmod b)$
    end if

The greatest common divisor of a and b is the smallest difference of multiples of a and b. This is because – for any difference, d, of multiples of a and b – we have,

$$d = ma + nb \text{ for } m, n \in \mathbb{Z}$$

and, if $g = gcd(a, b)$ then, by definition, g divides both a and b and, therefore, also divides d. So any such sum (or difference) of integer multiples of a and b is a multiple of g.

Proposition 4. *For non-zero integers a and b, if $a = bq + r$ where $q, r \in \mathbb{Z}$, then $gcd(a, b) = gcd(b, r) = gcd(b, a \bmod b)$.*

Proof. $(a \bmod b) = r = a - bq$. For any m s.t. $m \mid a$ and $m \mid b$ we must also have $m \mid (a - bq)$ so the set of divisors of a and b is a subset of the set of divisors of b and $r = (a \bmod b)$. Conversely, for any m s.t. $m \mid b$ and $m \mid r = (a \bmod b)$ we have that $m \mid (bq + r) = a$ so the set of divisors of b and $r = (a \bmod b)$ is a subset of the set of divisors of a and b. So the sets are equal proving that they must have the same maximum element - the greatest common divisor. □

Proposition 5. *If $d = gcd(a, b)$ then there is no integer linear combination of a and b that equals any positive value less than d.*

Proof. Assume $d = gcd(a, b)$ and that $\exists\, e < d \in \mathbb{N}, m, n \in \mathbb{Z}$ s.t. $e = am + bn$. Then,

$$\begin{aligned} & e = am + bn = dz_1 m + dz_2 n = d(z_1 m + z_2 n) \quad \text{for } z_1, z_2 \in \mathbb{Z} \\ \iff \quad & z_1 m + z_2 n = \frac{e}{d} \notin \mathbb{Z} \end{aligned}$$

where we know that $\frac{e}{d} \notin \mathbb{Z}$ because $e < d$. But the field properties of the integers ensures that the integers are closed under integer linear combinations so that $z_1 m + z_2 n \in \mathbb{Z}$. Therefore such an e does not exist. □

Corollary 1. *If $d = gcd(a, b)$ then every integer linear combination of a and b is a multiple of d.*

Proof. Proposition 5 showed that there is no integer linear combination of a and b less than d. Suppose that we have,

$$e > d \in \mathbb{N}, m, n \in \mathbb{Z} \text{ s.t. } e = am + bn$$

then, because d divides both a and b,

$$e = adz_1 + bdz_2 = d(az_1 + bz_2), \quad z_1, z_2 \in \mathbb{Z}.$$

Now the closure of the integer field means that $z_3 = az_1 + bz_2 \in \mathbb{Z}$ so that,

$$e = z_3 d \implies d \mid e.$$

□

Proposition 6. *A number $x \in \mathbb{Z}_m$ has a multiplicative inverse if and only if $gcd(x, m) = 1$.*

Proof. Assume x^{-1} is a multiplicative inverse for $x \in \mathbb{Z}_m$. Then,

$$x^{-1}x = 1 \iff x^{-1}x \equiv 1 \pmod{m} \iff x^{-1}x = am + 1, \quad a \in \mathbb{Z}.$$

This means that we must have $1 = am + bx$ for some $a, b \in \mathbb{Z}$. Now if we have $d = gcd(x, m)$ then by Corollary 1 we must have $d \mid 1$. Therefore $d = 1$.

Clearly, also, if we have $gcd(x, m) = 1$ then we also have $1 = am + bx$ for some $a, b \in \mathbb{Z}$ and by following the previous logic in reverse we obtain that $b = x^{-1}$ is the multiplicative inverse of $x \in \mathbb{Z}_m$. □

Lowest Common Multiple

The lowest common multiple of two numbers is formed by the multiplication of all the prime factors that occur in the two numbers where repititions of prime factors are important. That's to say, the lowest common multiple of 4 and 8 is not 2 (which is the highest common factor/greatest common divisor) but 8 because in 8, the factor 2 occurs three times (as 2^3) and it occurs twice in 4,

$$lcm(4, 8) = lcm(2 \times 2, 2 \times 2 \times 2) = 2 \times 2 \times 2.$$

The general formula for the lowest common multiple may be expressed in terms of the gcd as follows

$$d = gcd(a, b) \implies lcm(a, b) = d \times (a/d) \times (b/d).$$

1.1.6 Some Proofs on the Integers

(tags: number theory)

Proposition 7. *For any integer m, $\sqrt{m}$ is rational iff m is a square, i.e. $m = a^2$ for some integer a.*

To begin with we show the easier direction of implication: $(m = a^2) \implies (\sqrt{m}$ is rational).

Proof. Assume $m, a, b \in \mathbb{Z}$.

$$\begin{aligned} & m = a^2 \\ \iff & \sqrt{m} = |a| \\ & = a/b \text{ for } b = 1 \text{ or } -1. \end{aligned}$$

□

Now the other (harder) direction, $(\sqrt{m}$ is rational$) \implies (m = a^2)$.

Proof. Assume $m, a, b \in \mathbb{Z}$. ($\sqrt{m}$ is rational) can be formalized as:

$$\exists\, m, a, b \in \mathbb{Z} \cdot (\sqrt{m} = \frac{a}{b}) \ \wedge \ (a \text{ and } b \text{ are coprime})$$

$$\begin{aligned} \sqrt{m} &= \frac{a}{b} \\ \implies m &= \frac{a^2}{b^2} \\ \iff mb^2 &= a^2 \end{aligned}$$

But a and b are coprime so they don't share any prime factors. This means that a^2 and b^2 also don't share any prime factors. So, if $|b| > 1$, the prime factorization of mb^2 is necessarily different from that of a^2 meaning that $mb^2 \neq a^2$ contradicting the hypothesis of coprimality. On the other hand, if $|b| = 1$, then b has no prime factors (its prime factorization is empty) and so mb^2 has the same prime factorization as m which may be equal to that of a^2 in the case that $m = a^2$. □

Proposition 8. *For all nonnegative integers $a > b$ the difference of squares $a^2 - b^2$ does not give a remainder of 2 when divided by 4.*

Beginner's attempt - try proof by contradiction:

$$\begin{aligned} a^2 - b^2 &= 4n + 2 \\ 2k &= 4n + 2 && \text{by } a^2 - b^2 \text{ even} \\ k &= 2n + 1 \implies && k \text{ is some odd number.} \end{aligned}$$

So, proof by contradiction is our first instinct but doesn't seem to get us anywhere. Instead, proceed by cases:

Case a, b are even:

$$\begin{aligned} &\exists\, k, l \in \mathbb{Z} \;\cdot\; a = 2k, b = 2l \\ \implies\; & a^2 - b^2 = 4k^2 - 4l^2 \\ & = 4\left(k^2 - l^2\right) \\ & = 4m \;\text{ where }\; m \in \mathbb{Z} \end{aligned}$$

So 4 divides $a^2 - b^2$ with 0 remainder.

Case a, b are odd:

$$\begin{aligned} &\exists\, k, l \in \mathbb{Z} \;\cdot\; a = 2k + 1, b = 2l + 1 \\ \implies\; & a^2 - b^2 = \left(4k^2 + 4k + 1\right) - \left(4l^2 + 4l + 1\right) \\ & = 4\left(k^2 + k - l^2 - l\right) \\ & = 4m \;\text{ where }\; m \in \mathbb{Z} \end{aligned}$$

So, again, 4 divides $a^2 - b^2$ with 0 remainder.

Case a even, b odd:

$$\begin{aligned}
&\exists\, k, l \in \mathbb{Z} \;\cdot\; a = 2k, b = 2l + 1 \\
&\implies a^2 - b^2 = 4k^2 - \left(4l^2 + 4l + 1\right) \\
&\qquad\qquad = 4\left(k^2 - l^2 - l\right) - 1 \\
&\qquad\qquad = 4m + 3 \;\text{ where }\; m = k^2 - l^2 - l - 1 \in \mathbb{Z}
\end{aligned}$$

So, here, 4 divides $a^2 - b^2$ with 3 remainder. So the proposition is proven as we have proven all the possible cases.
There is also another approach given in the Cambridge University Discrete Mathematics lecture notes, TODO

1.1.7 Absolute Value

(tags: number theory, absolute value)

$$|x| \geq x, |y| \geq y \implies |x| + |y| \geq x + y$$

$$|x+y| = \begin{cases} |x| + |y| & x, y \geq 0 \\ |-|x| + |y|| & x < 0, y \geq 0 \\ ||x| - |y|| & x \geq 0, y < 0 \\ |-(|x| + |y|)| & x, y < 0 \end{cases} \iff \begin{cases} ||x| + |y|| & x, y \geq 0 \text{ or } x, y < 0) \\ ||x| - |y|| & x < 0, y \geq 0 \text{ or } x \geq 0, y < 0 \end{cases}$$

Clearly, $||x| + |y|| \geq ||x| - |y||$ so that,

$$|x + y| \leq ||x| + |y|| = |x| + |y|$$

and this is known as the "triangle inequality".

Proposition 9. $|x - y| \leq |x - z| + |y - z|$

Proof.

$$|x - y| = |(x - z) + (z - y)| \leq |x - z| + |z - y| = |x - z| + |y - z|$$

□

Proposition 10. $|x - y| \geq ||x| - |y||$

Proof. Need to show $-|x - y| \leq |x| - |y| \leq |x - y|$. So, prove as two separate inequalities:

$$\begin{aligned} & |y| = |x + (y - x)| \leq |x| + |y - x| \\ \iff \qquad & -|y - x| = -|x - y| \leq |x| - |y| \end{aligned}$$

$$\begin{aligned} & |x| = |(x - y) + y| \leq |x - y| + |y| \\ \iff \qquad & |x| - |y| \leq |x - y| \end{aligned}$$

□

1.1.8 Complex Number

(tags: number theory, complex numbers)

Definition. *The* ***modulus*** *of a complex number, $z = a + bi$, is the quantity defined as,*

$$|z| = \sqrt{a^2 + b^2}.$$

TODO: modulus is also called "absolute value" and can be calculated as product with conjugate The modulus obeys the following properties:

- $|z_1 z_2| = |z_1|\,|z_2|$

Chapter 2

Algebra

(tags: abstract algebra, complex numbers)

2.1 Abstract Algebra

(tags: abstract algebra)

2.1.1 Groups

(tags: abstract algebra)

Definition. *A binary operation is a function,*

$$f : G \times G \longmapsto G$$

which - by the definition of a function - maps a unique tuple from $G \times G$ to a unique value in the codomain G.

Definition. *Let G be a set and $*$ a binary operation on G and denote this $(G, *)$. Then $(G, *)$ is a* ***group*** *if:*

closure $\forall x, y \in G, x * y \in G$;

associativity $\forall x, y, z \in G, (x * y) * z = x * (y * z)$;

identity $\exists e \in G$ *s.t.* $\forall x \in G, e * x = x * e = x$;

inverse $\forall x \in G, \exists x^{-1} \in G$ *s.t.* $x * x^{-1} = x^{-1} * x = e$.

These are known as the ***group axioms****.*

Definition. *The group is an* ***Abelian group*** *if it has the additional property:*

commutativity $\forall x, y \in G, x * y = y * x \in G$.

Notation. from here on we will use juxtaposition notation for the group operation (so $xy = x * y$) and (usually) 1 for the identity element instead of e. This is known as *multiplicative notation.*

Theorem 1. *Suppose an associative law of composition is given on a set S. Then there is a unique way to define a product of n elements $a_1, \ldots, a_n$ for any $n \in \mathbb{N}$.*

Proof. Denote the product of n elements as $[a_1 \ldots a_n]$. We show that a product can be defined with the following properties:

(i) $[a_1] = a_1$;

(ii) $[a_1 a_2] = a_1 * a_2$ is defined by the law of composition;

(iii) for any integer i such that $1 \leq i \leq n$, $[a_1 \ldots a_n] = [a_1 \ldots a_i][a_{i+1} \ldots a_n]$.

Following a proof by induction, firstly note that the product is defined for $n \leq 2$ by (i) and (ii) and that (ii) also satisfies the requirement (iii). Then, assume that the product is defined for $n \leq 2$ and that this product is the unique product satisfying (iii).
Then the induction step is to show that,

$$[a_1 \dots a_n] = [a_1 \dots a_{n-1}][a_n]$$

TODO: complete this from Artin[56] □

Corollaries of the group axioms

The group operation is defined to map a unique tuple in $G \times G$ to a unique value in G so that if we have $x, y \in G$ then $f((x, y)) = f(x, y) = xy \in G$ and for $a, b, c \in G$,

$$a = b \iff (c, a) = (c, b) \implies f((c, a)) = f((c, b)) \iff ca = cb$$

$$\therefore a = b \implies ca = cb$$

Then, using all the group axioms - associativity, inverse and identity,

$$ca = cb \implies c^{-1}(ca) = c^{-1}(cb) \iff (c^{-1}c)a = (c^{-1}c)b \iff 1a = 1b \iff a = b$$

Therefore we have the principle of cancellation,

$$ca = cb \implies a = b$$

Note that, since we have used the axioms of inverse and identity and the definitions of these require these elements to exhibit these properties from both the left and the right, the principle of cancellation can also be shown from both the left and the right. So, also,

$$ac = bc \implies a = b$$

There are (at least) two approaches to finding the other consequences of the group axioms.

First approach. We begin by noticing that the law of cancellation implies that,

unique identity and inverses $\forall a, x, b \in G, ax = b$ has a unique solution because,

$$ax = ax' \iff x = x'$$

That unique solution is $a^{-1}b$. If $b = a$ we have $ax = a$ and x, by identity axiom, is an identity element. Since, the solution to this equation - x - is unique, it follows that there is a unique value that is the identity element. Then, if we let b be this unique identity element we have $ax = 1$ and the unique solution, x, is the inverse of a, i.e. a^{-1}. Therefore, the inverses of group elements are also unique.

Second approach. This approach begins by showing the uniqueness of the identity element solely using the defintion of the identity. Here, for clarity, we revert to using e to denote the identity element.

unique identity Assume there are two identity elements, e, e'. Then, by the definition of the identity $ee' = e'e = e = e'$ so that there is a single value that has the property of the identity element.

Then, using the definition of the inverse we have,

unique inverses Assume there are two distinct inverses of an element a: a^{-1} and a'. Then,

$$\begin{aligned} & aa^{-1} = 1 = aa' && \text{defn. of inverse, uniqueness of identity} \\ \iff \quad & a^{-1} = a' && \text{law of cancellation} \end{aligned}$$

Some Examples of Groups

- $(\mathbb{R} \setminus \{0\}, \times)$ is a group whereas $(\mathbb{R}, \times)$ is not a group because 0 has no multiplicative inverse.

- $(\mathbb{R}, +)$ is a group.
- The set of $n \times n$ invertible matrices is called the General Linear group and denoted GL_n
- Let G denote the set of matrices

$$G = \left\{ \begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix} \;\middle|\; a, b \in \mathbb{Z}_7,\ a \neq 0 \right\}.$$

Then G is a group with respect to matrix multiplication (where all additions and multiplications are carried out in $\mathbb{Z}_7$). This is because closure can be shown by,

$$\begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a' & b' \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} aa' & ab' + b \\ 0 & 1 \end{pmatrix} \in G$$

where the result is in G because $aa' \neq 0$. Next we need to show that we have inverses. So we need to show existence in G of matrices such that,

$$\begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a' & b' \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \iff \begin{pmatrix} aa' & ab' + b \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

which implies that $aa' = 1$ and $ab' + b = 0$. Because we are in $\mathbb{Z}_7$ every non-zero element has a multiplicative inverse so we have,

$$a' = a^{-1} \quad \text{and} \quad b' = -a^{-1}b$$

so that,

$$\begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} a^{-1} & -a^{-1}b \\ 0 & 1 \end{pmatrix}.$$

Permutations and Symmetric Groups

Definition. *A **permutation** is a bijection from a set to itself. Since permutations are bijective, they are invertible and since they are functions, function composition defines an associative law of composition over them. As a result, they form a group.*

Definition. *The* ***symmetric group*** *defined over a set is the group whose elements are the permutations of the objects of the set and whose law of composition is the composition of functions. The name probably comes from the study of symmetries of geometric objects that were eventually realised to be equivalent to permutations of the vertices.*

Notation. The symmetric group over the integers from 1 to n is denoted S_n. The symmetric group over a set G may be denoted $Sym(G)$.

S_2 The symmetric group S_2 consists of the two elements i and τ which are, respectively, the identity map and the transposition which interchanges 1 and 2. The group composition law is described by the fact that the identity map is the identity of the composition and by the relation $\tau\tau = \tau^2 = i$. Which results in the multiplication table:

$$\begin{aligned} i \cdot i &= i \\ i \cdot \tau &= \tau \\ \tau \cdot i &= \tau \\ \tau \cdot \tau &= i \end{aligned}$$

Note that the law of composition is commutative.

S_3 The symmetric group S_3 contains 3! elements. It is the smallest group whose law of composition is not commutative. It can be described using any two permutations of $\{1, 2, 3\}$. For example, if we take,

$$x = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}, \; y = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Then the permutations are,

$$\{1, x, x^2, y, xy, x^2y\} = \{\, x^i y^j \mid 0 \leq i \leq 2, \; 0 \leq j \leq 1 \,\}$$

These are the elements of the group. The composition law over these elements is the function composition of these permutation functions and its

multiplication table is characterized by the rules:

$$x^3 = 1,\ y^2 = 1,\ yx = x^2y$$

These are derived directly from the permutations themselves. Note that this composition law is not associative as $yx \neq xy$.
Any product of the elements x, y and of their inverses can be brought into the form $x^i y^j$ with i, j taking the ranges given above by repeated application of the above rules. To do so, we move all occurrences of y to the right side using the last relation and bring the exponents into the indicated ranges using the first two relations:

$$\begin{aligned} x^{-1}y^3x^2y = x^2yx^2y = x^2(yx)xy = x^2(x^2y)xy = x^4(yx)y \\ = x^4(x^2y)y = x^6y^2 = (x^3)^2y^2 = 1 \cdot 1 = 1 \end{aligned}$$

Rules like these that determine a complete multiplication table are called *defining relations* for the group.

2.1.2 Subgroups

(tags: abstract algebra)

Definition. *A subset H of a group G is called a* ***subgroup*** *if it has the following properties:*

- ***Closure:*** *If $a \in H$ and $b \in H$ then $ab \in H$.*
- ***Identity:*** $1 \in H$.
- ***Inverses:*** *If $a \in H$ then $a^{-1} \in H$.*

These conditions show that the subset H is a group with respect to the induced law of composition created by applying the law of composition of G on the members of H. Note that the associative property is not mentioned because the associativity of the composition of members of G automatically carries over to H.

Notation. If H and G are groups then we may write $H \leq G$ to indicate that H is a subgroup of G.

Note that an alternative, more compact, formulation of the definition of a subgroup is as follows.

Let G be a group and $\emptyset \neq H \subseteq G$. Then H is a subgroup if

$$x, y \in H \implies x^{-1}y \in H.$$

This is because,

$$[(x, y \in H \implies xy \in H) \wedge (x \in H \implies x^{-1} \in H)] \iff (x, y \in H \implies x^{-1}y \in H).$$

The implication,

$$[(x, y \in H \implies xy \in H) \wedge (x \in H \implies x^{-1} \in H)] \implies (x, y \in H \implies x^{-1}y \in H)$$

is obvious. In the other direction,

$$(x, y \in H \implies x^{-1}y \in H) \implies [(x, y \in H \implies xy \in H) \wedge (x \in H \implies x^{-1} \in H)]$$

is because, if we set $x = y$,

$$x^{-1}x = e \in H \implies x^{-1}e = x^{-1} \in H$$

and then, for $x \neq y$,

$$x^{-1}, y \in H \implies xy \in H.$$

Every group, at a minimum, has two trivial subgroups: the maximal subgroup - the group itself; and the minimal subgroup - the set containing just the identity. A subgroup that is neither of these is known as a *proper subgroup*.

Proposition 11. *Suppose H and K are subgroups of G such that neither $H \subseteq K$ nor $K \subseteq H$. Then $H \cup K$ is not a subgroup of G.*

*Note that it might be possible, for this proof, to just show that we have no reason to think that $H \cup K$ is a subgroup (i.e. the definitions don't require it) so, in general, it is not. But, actually, we can show something much stronger: that $H \cup K$ **cannot** be a subgroup. If we can easily show something stronger then in most cases it's going to add clarity.*

Proof. Since neither $H \subseteq K$ nor $K \subseteq H$ we can conclude that $H \setminus K$ and $K \setminus H$ are both non-empty. So, select an element from each,

$$h \in H \setminus K,\ k \in K \setminus H.$$

Then we have $h, k \in H \cup K$ and if $H \cup K$ were a group then the closure property of the group would require that

$$hk \in H \cup K.$$

Assume $hk \in H \cup K$. Then, $hk \in H$ or $hk \in K$. If $hk \in H$ then the group properties of H require that

$$h^{-1}hk = k \in H$$

which contradicts the selection of k. We have a similar situation if $hk \in K$. Therefore, $hk \notin H \cup K$. □

Additive Groups of Integers

Important examples are the subgroups of the additive group of integers $\mathbb{Z}^+$. Denote the subset of $\mathbb{Z}^+$ consisting of all multiples of a given integer b by $b\mathbb{Z}$ such that,

$$b\mathbb{Z} = \{\, n \in \mathbb{Z} \mid n = bk,\ k \in \mathbb{Z} \,\}$$

Proposition 12. *For any integer b, the subset $b\mathbb{Z}$ is a subgroup of $\mathbb{Z}^+$ and every subgroup of $\mathbb{Z}^+$ is of the form $b\mathbb{Z}$ for some integer b.*

Proof. $b\mathbb{Z}$ is a subgroup of $\mathbb{Z}^+$ because,

- $b(0) = 0 \in b\mathbb{Z}$;
- If $a_1, a_2 \in b\mathbb{Z}$ then $a_1 = bk_1, a_2 = bk_2$ for $k_1, k_2 \in \mathbb{Z}$ and so $a_1 + a_2 = bk_1 + bk_2 = b(k_1 + k_2) \in b\mathbb{Z}$
- For any $a = bk \in b\mathbb{Z}$, $-a = b(-k) \in b\mathbb{Z}$

Now we need to prove that any subgroup of $\mathbb{Z}^+$ is $b\mathbb{Z}$ for some b. Let H be an arbitrary subgroup of $\mathbb{Z}^+$. Then by subgroup properties,

- $0 \in H$;

- If $a_1, a_2 \in H$ then $a_1 + a_2 \in H$
- For any $a \in H$, $-a \in H$

We proceed to show that there is always some integer b such that $H = b\mathbb{Z}$. Firstly, if H is the minimal subgroup $\{0\}$ then H trivially conforms to $b\mathbb{Z}$ with $b = 0$.
Otherwise, $\exists a \in H$ s.t. $a \neq 0$ then also $\exists -a \in H$ s.t. $-a \neq 0$. One of these must be a positive non-zero integer so there is at least one such member of H. We take b to be the smallest positive non-zero integer in H. Then,

b*$\mathbb{Z} \in$ *H

- $b \in H$ (by selection) so by subgroup properties $b+b \in H$ and $(b+b)+b \in H$ and $b + \cdots + b \in H$
- By subgroup properties $b \in H \implies -b \in H$

So, $\{\, bk \in \mathbb{Z} \mid k \in \mathbb{Z} \,\}$ is in H.

***H* $\in$ *b*$\mathbb{Z}$** Take any $n \in H$. Using division with remainder and dividing by b we get,

$$n = bq + r \quad q \in \mathbb{Z},\ 0 \leq r < b$$

But, since $b\mathbb{Z} \in H$ this means that $bq \in H$ and so $-bq \in H$. Therefore $n - bq = r \in H$. But $0 \leq r < b$ and, by assumption, b is the smallest positive *non-zero* integer in H and so, $r = 0$. So, every $n \in H$ divides by b. □

Greatest Common Divisor

If we extend this to groups which are generated by two integers a, b, then we have a subgroup of $\mathbb{Z}^+$,

$$a\mathbb{Z} + b\mathbb{Z} = \{\, n \in \mathbb{Z} \mid n = ar + bs \quad r, s \in \mathbb{Z} \,\}$$

This is known as the subgroup *generated* by a, b because it is the smallest subgroup which contains a and b. Proposition 12 tells us that it has the form $d\mathbb{Z}$ for some integer d.

Corollary 2. *If d is the positive integer which generates the subgroup $a\mathbb{Z}+b\mathbb{Z}$ then d is the greatest common divisor of a and b and so,*

- *d can be written in the form $d = ar + bs$ for some integers r and s.*
- *d divides a and b.*
- *If an integer e divides a and b, it also divides d.*

Proof. The first property follows directly from the definition of the subgroup. The second property is a result of the fact that a, b are in the subgroup $a\mathbb{Z} + b\mathbb{Z}$ so that $d\mathbb{Z} = a$ and $d\mathbb{Z} = b$. The third property is evident because $d = ar + bs = ek_1r + ek_2s = e(k_1r + k_2s)$. □

Deductions about Subgroups

Proposition 13. *Suppose $G = \{g_1, g_2, \ldots, g_n\}$ is a finite group of order n and that $x \in G$. Then $\{xg_1, xg_2, \ldots, xg_n\} = G$.*

Proof. Let $X = \{xg_1, xg_2, \ldots, xg_n\}$. By closure in G, every element of X must be in G and by the inverses property in G every element of X is distinct. So, there are n distinct elements of X, each of which are members of G. Since the order of G is n we can conclude that $X = G$. □

Proposition 14. *Suppose that G is a finite group and that H is a non-empty subset of G such that $x, y \in H \implies xy \in H$. Then H is a subgroup.*

Proof. H is non-empty so it contains at least one element, say x. H is closed under the group operation so it must also contain $x^2, x^3, \ldots$. But G is finite so the order of x in G must be finite also and so $\exists n \in \mathbb{N}$ s.t. $x^n = e$. But also $x^n = e \iff x^{n-1} = x^{-1}$. Therefore, for every element in H, the inverse of the element is also in H. □

Proposition 15. *Suppose that p is a prime number and we have integers $1 \leq x, g, h < p$ such that $xg \equiv xh \pmod{p}$. Then $g = h$ and $x \in \mathbb{Z}_p^*$ has an inverse.*

Proof. If $xg \equiv xh \pmod{p}$ then the difference between xg and xh is a multiple of p. But Euclid's Lemma (https://en.wikipedia.org/wiki/Euclid's_lemma) tells us that, because p is prime,

$$p \mid (xg - xh) = x(g - h) \implies (p \mid x) \wedge (p \mid (g - h)).$$

But since we have $x, (g-h) < p$ it is impossible for p to divide either of them unless they are 0. Only $g-h$ can be 0. Therefore,

$$g - h = 0 \iff g = h.$$

So, if we define a function $f : \mathbb{Z}_p^* \longmapsto \mathbb{Z}_p^*$ such that $f(a) = xa$ for some fixed $x \in \mathbb{Z}_p^*$ then f is injective because

$$f(a) = f(b) \iff xa \equiv xb \pmod{p} \iff a \equiv b \pmod{p}.$$

This means that f maps the $p-1$ different values of $\mathbb{Z}_p^*$ to $p-1$ different values in $\mathbb{Z}_p^*$. Therefore f is a bijection and there exists an inverse function f^{-1} such that $f^{-1}(xa) = a$. □

Examples of Subgroups

(1) $(\mathbb{Z}_p^*, \otimes)$ for prime p is a subgroup of the integers. This can be seen as closure of modular multiplication is clear and the existence of inverses has been shown in Proposition 15.
This is not the case however, for non-prime p. For example $(\mathbb{Z}_6^*, \otimes)$ is not a subgroup as it does not have inverses. We can see this by looking at the values generated by selecting a non-identity element and multiplying it by all the elements in $\mathbb{Z}_6^*$:

$$2 \otimes 1 = \mathbf{2},\ 2 \otimes 2 = \mathbf{4},\ 2 \otimes 3 = \mathbf{0},\ 2 \otimes 4 = \mathbf{2},\ 2 \otimes 5 = \mathbf{4}.$$

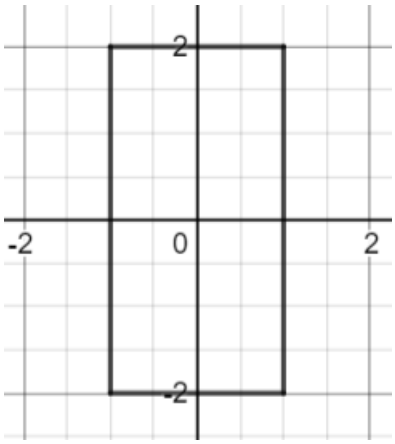
As can be seen, it doesn't generate all the values of $\mathbb{Z}_6^*$ but repeats a subset of them. Compare with the same for $\mathbb{Z}_7^*$:

$$2 \otimes 1 = \mathbf{2},\ 2 \otimes 2 = \mathbf{4},\ 2 \otimes 3 = \mathbf{6},\ 2 \otimes 4 = \mathbf{1},\ 2 \otimes 5 = \mathbf{3},\ 2 \otimes 6 = \mathbf{5}.$$

In the case of prime p all the values are generated so that multplication by other elements is a bijective function with a corresponding inverse.

(2) Let R be a non-square rectangle in $\mathbb{R}^2$ with corners having coordinates $(-1,-1), (-1,2), (1,2), (1,-1)$. Then there are four symmetries i, a, b, c of R, as follows:

- i is the identity
- a is reflection in the x-axis
- b is reflection in the y-axis
- c is a rotation of π radians around the origin.



These symmetry operations form a group whose group table is as follows.

	i	a	b	c
i	i	a	b	c
a	a	i	c	b
b	b	c	i	a
c	c	b	a	i

Cyclic Subgroups

Definition. *If we take a single member of a group (along with its inverse and the identity), the subgroup generated by that element takes the form (using multiplicative notation),*

$$H = \{x^{-(n-1)}, \cdots, x^{-2}, x^{-1}, 1, x, x^2, \cdots, x^{n-1}\}$$

where, either, $x^n = 1$ so that there are n distinct values in the group, or else n is infinite and the values never repeat. This is known as a ***cyclic group*** *and also as the* ***subgroup generated by x*** *and is denoted by $\langle x \rangle$.*

The cyclic subgroup, $\langle x \rangle$, generated by x is the smallest subgroup of G containing x in the sense that, if $H \leq G$ and $x \in H$ then $\langle x \rangle \subseteq H$.

Proposition 16. *Every cyclic group is Abelian.*

Proof. In a cyclic group every element has the form x^i for $i \in \mathbb{Z}$. So we have,

$$x^m x^n = x^{m+n} = x^n x^m$$

for all elements x^i in the group. □

Proposition 17. *The set S of integers n such that $x^n = 1$ is a subgroup of $\mathbb{Z}^+$.*

Proof. If $x^m = 1$ and $x^n = 1$, then $x^{m+n} = x^m x^n = 1$ also so we have closure of addition. Since $x^0 = 1$, 0 is in the subgroup so we have an identity. Finally, for some n in the subgroup, $x^n = 1 \iff x^{-n} = x^n x^{-n} = x^0 = 1$ so n being in the subgroup implies that $-n$ is also in the subgroup and we have inverses. □

Corollary 3. *It follows from S being a subgroup of $\mathbb{Z}^+$ and from Proposition 12 that S has the form $m\mathbb{Z}$ where m is the smallest positive integer such that $x^m = 1$. Therefore, in H, the m elements $1, x, x^2, \cdots, x^{m-1}$ are all different and any element in H will simplify to one of them: for $n \in S$, $n = mq + r$ such that $x^n = (x^m)^q x^r = 1^q x^r = x^r$.*

Order

Definition. *The **order** of a group G is the number of distinct elements it contains. It is typically denoted $|G|$.*
*An element of a group is said to have **order** m (possibly infinity) if the cyclic subgroup it generates has order m. This means that m is the smallest positive integer with the property $x^m = 1$ or, if the order is infinite, that, $x^m \neq 1$ for all $m \neq 0$.*

Theorem 2. *An element and its inverse have the same order.*

Proof. Firstly we need to consider the case that an element x has infinite order. In this case, $\nexists m \in \mathbb{N}$ s.t. $x^m = e$. Now suppose that $\exists n \in \mathbb{N}$ s.t. $(x^{-1})^n = e$. Then we have,

$$(x^{-1})^n = (x^n)^{-1} = e \iff e = x^n$$

which contradicts the hypothesis that x has infinite order. Therefore x^{-1} has infinite order also. Clearly also this argument can be used in reverse to show the reverse implication also holds.

Now consider the case that x has finite order. Let $x \in G$ be an arbitrary member of an arbitrary group such that $x^m = e$. Then $x^{-1} = x^{m-1}$ and if we consider powers $i \in \mathbb{N}$ of the inverse $(x^{-1})^i = (x^{m-1})^i$ then the order is the lowest value of $i(m-1)$ such that $x^{i(m-1)} = e$. But we know that the lowest power of x equal to e is m so we're looking for the lowest multiple of m that has the form $i(m-1)$. So we require,

$$m \mid i(m-1) = im - i \iff (m \mid im) \wedge (m \mid i)$$

which clearly requires that $m \mid i$. Also, clearly, the lowest such i is $i = m$.

Another way to show this is to say that if $x^m = e$ then,

$$(x^{-1})^m = (x^m)^{-1} = e$$

so that the order of x^{-1} is less than or equal to m. Conversely, if x^{-1} has order n then,

$$x^n = ((x^{-1})^{-1})^n = ((x^{-1})^n)^{-1} = e^{-1} = e$$

so that the order of x is less than or equal to n. Thus we have,

$$m \leq n,\ n \leq m \implies m = n.$$

□

Theorem 3. *An element has order 2 iff it is equal to its inverse.*

Proof. Let $x \in G$ be an arbitrary member of an arbitrary group such that $x^2 = e$. Then by Theorem 1 we have,

$$e = x^2 = (x^{-1})^2 = x^{-2} \iff x = x^{-1}.$$

Also,

$$x = x^{-1} \iff x^2 = e.$$

□

Theorem 4. *A group of finite order cannot have any element of infinite order.*

Proof. If G is a group and $x \in G$ has infinite order then,

$$\begin{aligned}
& x^m = x^n \\
\iff \quad & x^{m-n} = 1 = x^{n-m} \\
\iff \quad & x^{|m-n|} = 1 \\
\therefore \quad & |m-n| = 0. \qquad \text{because order of x is infinite}
\end{aligned}$$

So, there are no two distinct powers of x that produce the same object so that $\langle x \rangle \leq G$, the cyclic group generated by x, is infinite. Since $\langle x \rangle \subseteq G$ this requires that G also be infinite. □

Theorem 5. *If a group element x has finite order m then:*

1. *Let $n \in \mathbb{Z}$. If $n = km + r$ where $k, r \in \mathbb{Z}$ and $0 \leq r \leq m-1$, then $x^n = x^r$.*
2. *For $n \in \mathbb{N}$, $x^n = 1 \iff m|n$.*
3. *$1, x, x^2, \ldots, x^{m-1}$ is a complete, repetition-free, list of elements of $\langle x \rangle$.*

4. The subgroup $\langle x \rangle$ generated by x has cardinality m.

Theorem 6. *In an Abelian group the set of all elements of finite order forms a subgroup.*

Proof. Let S be the set of all elements of finite order,

$$S = \{\, g \in G \mid \exists m \in \mathbb{N} \,.\, g^m = e \,\}.$$

- Firstly, S is non-empty because it contains the identity.
- Secondly, if we have $a, b \in S$ then $a^i = b^j = e$ for some $i, j \in \mathbb{N}$. Then, if we let $m = i \times j$,

  $$(ab)^m = a^m b^m = (a^i)^j (b^j)^i = e$$

  where $(ab)^m = a^m b^m$ is valid *only* because the group is Abelian.
- Lastly, S contains all inverses because,

  $$a^i = e \iff a^{i-1} = a^{-1} \iff (a^{-1})^i = (a^i)^{i-1} = e.$$

  Therefore $a^{-1} \in S$.

□

Theorem 7. *If every non-identity element of a group has order 2 then the group is Abelian.*

Proof. Let $x, y \in G$ be two arbitrary elements of order 2 of an arbitrary group. Then by Theorem 1 we have $x = x^{-1},\ y = y^{-1},\ xy = xy^{-1}$ and so,

$$xy = (xy)^{-1} = y^{-1}x^{-1} = yx.$$

Note that $(xy)^{-1} = y^{-1}x^{-1}$ relies only on the associativity of the group operation and is therefore valid for all groups.
We can also show it this way,

$$(xy)^2 = e \iff xyxy = e \iff yxy = xe = x \iff yx = xy.$$

□

Theorem 8. *If a finite group has even-numbered order then it must have at least one element of order 2.*

Proof. By the group properties we know that the group contains the identity element – which has order 1 – and, for every non-identity element, the group also contains its inverse. Also, since the group is finite, every element must have finite order. Now, if every non-identity element is distinct from its inverse then the order of the group will be odd (because of the identity and then every other element is paired with its inverse). For the group's order to be even we must have at least one non-identity element that is not distinct from its inverse which, by Theorem 2, is equivalent to having order 2. □

Corollary 4. *If a finite group has even-numbered order then it must have an odd number of elements of order 2.*

Proof. Let G be a finite group with even-numbered order and M be the number of elements that are distinct from their inverse and N be the number of elements that are not distinct from their inverse (these correspond to elements with order greater than 2 and elements with order 2 respectively). Then the order of G can be expressed as,

$$|G| = 1 + 2M + N.$$

Therefore, $|G|$ is even if N is an odd natural number. □

Theorem 9. *In an infinite cyclic group all elements have infinite order.*

Proof. Let $G = \langle x \rangle$ be an infinite cyclic group. Suppose there is some non-identity element of G, x^n with finite order m. Then,

$$(x^n)^m = e \iff x^{nm} = e$$

which contradicts the hypothesis that $\langle x \rangle$ is infinite. □

Note that the elements of an infinite cyclic group having infinite order does not mean that they generate the group. For example in an infinite cyclic group $\langle x \rangle$, the element x^2 generates the cyclic subgroup,

$$\ldots x^{-4}, x^{-2}, e, x^2, x^4, \ldots$$

which is infinite but clearly doesn't generate the whole group $\langle x \rangle$.

Theorem 10. *An infinite cyclic group has 2 generators.*

Proof. Let $G = \langle x \rangle$ be an infinite cyclic group and suppose there is some non-identity element of G, x^n that generates the group. To show this we only need to show that x^n can generate x because, since x is a member of the group, it is obviously necessary to generate it but, also, if we generate x then we can generate all the other members of the group since they are powers of x.
So let there be an integer a such that $(x^n)^a = x^{an} = x \iff x^n = x^{1/a}$. The cyclic group $\langle x \rangle$ only contains integer powers of x so it therefore follows that $|a| = 1$ which implies that $n = 1$ or -1 and $x^n = x$ or x^{-1}.

We could also say that,

$$x^{an} = x \iff x^{an-1} = e$$

but this implies that the order of x is finite and so contradicts the hypothesis that x generates an infinite cyclic group.

□

Theorem 11. *Let $G = \langle x \rangle$ be a finite cyclic group of order n. If r is a positive integer then $G = \langle x^r \rangle$ if and only if the greatest common divisor of n and r is 1.*

Proof. Members of $\langle x^r \rangle$ have the form $(x^r)^a$ for some $a \in \mathbb{Z}$. For integers b, i,

$$(x^r)^a = x^{ar} = x^{bn+i} = x^{bn}x^i = ex^i = x^i$$

so that the generated elements are x^i where $i = ar - bn$ is the remainder when dividing ar by n. If $d = gcd(n, r)$ then $d \mid i$ and the generated elements are powers of x that are multiples of d. Therefore, to generate every power of x it is necessary to have $d = 1$. Conversely, we can see – by the same argument in reverse – that it is sufficient if $d = 1$ to generate all the powers of x.

Alternatively, we can say if $d > 1$ then n/d is a positive integer less than n and r/d is a positive integer so,

$$(x^r)^{n/d} = (x^n)^{r/d} = e^{r/d} = e$$

which shows that the order of x^r is less than or equal to n/d which is less than n. Therefore the order of the cyclic group it generates is less than n

and so it cannot be equal to G.
Conversely, if $d = 1$ then n and r are coprime and so we have,

$$(x^r)^m = e \iff x^{rm} = e \implies n \mid rm \implies n \mid m \implies m \geq n.$$

This says that the order of x^r in G is greater than or equal to n, the order of G. Well, clearly it cannot be greater than the order of G so it follows therefore, that the order of x^n is n. Since $|\langle x^r \rangle| = |G|$ we can conclude that $\langle x^r \rangle = G$. □

Theorem 12. *A group G is such that G contains at least 2 elements and the only subgroups of G are $\{e\}$ and G itself. Then G is a finite cyclic group of prime order.*

Proof. G contains at least 2 elements so there is at least one non-identity element x. The only subgroups of G are the whole group and $\{e\}$ but $\langle x \rangle$ cannot equal $\{e\}$ so it must equal G. Therefore G is the cyclic group generated by x.
But if G were the infinite cyclic group generated by x then only x and x^{-1} would generate the group and all other non-identity elements – say x^n for $n > 1 \in \mathbb{N}$ – would generate subgroups $\langle x^n \rangle \neq G$. Therefore, G cannot be infinite and is therefore finite.
Now, we have a finite cyclic group where every non-identity element generates the group. Let $|G| = n$. Then, for every m s.t. $0 < m < n$, $\langle x^m \rangle = G$ and, by Theorem 10, m and n are coprime. Therefore, n is prime.
Here we could also use a proof by contradiction: Assume n is not prime and it has factors $r, s > 1 \in \mathbb{N}$. Then,

$$(x^r)^s = x^{rs} = x^n = e$$

so that the order of x^r in G is less than or equal to s which is less than n (because it is a factor of n). It follows then that $\langle x^r \rangle \neq G$, contradicting the definition of G. Therefore n is prime. □

Proposition 18. *Suppose the elements x, y in a group G have orders m, n respectively and that the $gcd(m, n) = 1$. Then $\langle x \rangle \cap \langle y \rangle = \{e\}$ and, if x and y commute, then the order of xy in G is mn.*

Proof. One way to approach this is to say that for $z \in \langle x \rangle \cap \langle y \rangle$ we have some $0 \leq i < m, 0 \leq j < n$ such that,

$$z = x^i = y^j \iff x^{im} = e = y^{jm},\ x^{in} = y^{jn} = e$$

so that $x^{in} = y^{jm} = e \iff (m \mid in \quad \text{and} \quad n \mid jm)$. Note that,

$$(m \mid in \quad \text{and} \quad n \mid jm) \iff m, n \mid in + jm.$$

Now, applying the fact that $gcd(m, n) = 1$ we see that both m and n must divide 1. But both m and n are orders of elements and so, by definition, greater than 1. The only other alternative is that both $i, j = 0$ which results in $z = x^0 = y^0 = e$.

We could also have said,

$$x^{im} = e = x^{in} \iff x^{im-in} = e \iff m \mid im - in.$$

In this case we can apply the fact that $gcd(m, n) = 1$ to the statement that $m \mid i(m - n)$ to deduce that: either $m \mid (m - n) \iff m \mid 1$ which is impossible because m must be greater than 1; or $m \mid i$ which is also impossible because $i < m$. So, again, we are only left with the alternative that $i = 0$ which results in $z = x^0 = y^0 = e$.

Next we prove the order of $xy \in G$ and we begin by noting that, if x and y commute, then $(xy)^r = x^r y^r$. So if we assume that xy has order r then we must have $m, n \mid r$ and the lowest such r is the order. Well, the lowest common multiple of m, n is defined according to the gcd as described in the Number Theory treatment of Modular Arithmetic (1.1.5) as,

$$d = gcd(m, n) \implies lcd(m, n) = d \cdot (m/d) \cdot (n/d).$$

Clearly then, if $d = gcd(m, n) = 1$, then the lowest common multiple is mn and so the order of $xy \in G$ is mn.

Or to describe it a different way: $(xy)^{mn} = x^{mn} y^{mn} = ee = e$ so that the order of xy,

$$|xy| \leq mn.$$

Conversely any r such that $(xy)^r = e$ must have $m, n \mid r$ and the lowest common multiple of m and n is mn so

$$r \geq mn.$$

Therefore, $|xy| = mn$. □

Examples of Cyclic Subgroups

(3) **Cyclic group with order 3**

$$G = \{1, x, x^2\}$$

where $x^3 = 1$ is a cyclic group of order 3 generated by the element x. Note that, since this is a group, it must also contain the inverses, x^{-1}, x^{-2} but $x^3 = 1$ so $x^{-1} = x^2$ and $x^{-2} = x$.

(4) **Symmetries of an equilateral triangle**

Consider an equilateral triangle with vertices labeled A, B, C:

$$\begin{array}{c} A \\ B \;\; C. \end{array}$$

Every permutation of the vertices is a transformation that produces an object that occupies the same space as the original, i.e. a *symmetry*. If we take one of them, say, the clockwise rotation one place that results in,

$$\begin{array}{c} B \\ C \;\; A \end{array}$$

and we name this r, then clearly – since there are 3 vertices – performing this same rotation 3 times leaves us back where we started. So, using function composition as the law of composition and multiplicative notation, $r^3 = i$ where i is the identity transformation. Also the inverse of r is r^2. So, we have a group consisting of $\{i, r, r^2\}$ and function composition. Notice the resemblance of this group to the previous group $\{1, x, x^2\}$; this group is *isomorphic* to the cyclic group of order 3.

(5) **Group $(\mathbb{Z}_5^*, \otimes)$**

Consider the element 2 modulo 5. Using multiplicative notation we have,

$$2^2 = 4, 2^3 = 8 = 3, 2^4 = 16 = 1, 2^5 = 32 = 2.$$

So $2^1 = 2^5 \iff 1 = 2^4$ meaning that the element 2 has order 4 in the group and we see, as expected that the group it generates, $\langle 2 \rangle$ has 4 members. In this case, the members are all the members of the group – that's to say, *the element 2 generates the whole group.* If we consider the element 4 we have,

$$4^2 = 16 = 1, 4^3 = 64 = 4.$$

So this element oscillates between 1 and 4 and so, the cyclic subgroup that it generates $\langle 4 \rangle$ has order 2.
Since the group $(\mathbb{Z}_5^*, \otimes) = \langle 2 \rangle$ it can also be described as a cyclic group. This will be the case for any such group modulo a prime number – i.e. $(\mathbb{Z}_p^*, \otimes)$ where p is prime.

(6) **Cyclic group with infinite order**

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

under matrix multiplication (which is commutative in this case), generates a cyclic group of infinite order because

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}^n = \begin{bmatrix} 1 & n \\ 0 & 1 \end{bmatrix}.$$

(7) **Cyclic groups in a non-Abelian group**

Consider the following two elements in $GL(2, \mathbb{R})$,

$$A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & -1 \\ 1 & -1 \end{pmatrix}.$$

Both of these elements have finite order as,

$$(A^2)^2 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}^2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

$$(B^2)B = \begin{pmatrix} -1 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

But their product AB does not have finite order.

$$AB = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}, \quad (AB)^n = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}^n = \begin{pmatrix} 1 & -n \\ 0 & 1 \end{pmatrix}$$

for any $n \in \mathbb{N}$.

(8) **The Klein Four Group,** V is the simplest group that is not cyclic (it cannot be generated by a single element). It appears in many forms but, as an example, it can be realized as the group consisting of the four matrices,

$$\begin{bmatrix} \pm 1 & 0 \\ 0 & \pm 1 \end{bmatrix}$$

Any two non-identity elements generate V.

2.1.3 Isomorphisms

(tags: abstract algebra)

Definition. *An **isomorphism** is a bijection between two groups that preserves the structure of the groups by being compatible with the law of composition of both groups. More formally, two groups are **isomorphic** if there exists a bijection $\phi : G \longmapsto G'$ such that,*

$$\phi(ab) = \phi(a)\phi(b) \text{ for all } a, b \in G$$

where ab represents composition according to the law of composition of G and $\phi(a)\phi(b)$ represents composition according to the law of composition of G'.

*An **isomorphism** is a **bijection** between two **groups**. That's to say, it is already assumed in the definition of an isomorphism that the codomain G' is a group.*

Proposition 19. *As a consequence of this sole property that, across the bijection, the respective laws of composition are preserved, all other properties of the groups are also preserved.*

Proof. Let e be the identity in G and $e' = \phi(e) \in G'$, and $1'$ be the identity element in G' then,

- Since G' is a group, it has the inverses property that every element has an inverse so,

$$\begin{aligned} & e' = \phi(e) = \phi(ee) = \phi(e)\phi(e) = e'e' && \text{using preservation of law of composition} \\ \iff \quad & (e')^{-1}e' = ((e')^{-1}e')e' && \text{using the inverses property of } G' \\ \iff \quad & 1' = e' \end{aligned}$$

  which implies that e' is the identity in G' so that ϕ maps the identity in G to the identity in G'.

- We can use the fact just shown that $\phi(e) = e' = 1'$ to show,

$$\begin{aligned} & 1' = e' = \phi(e) = \phi(aa^{-1}) = \phi(a)\phi(a^{-1}) && \text{using preservation of law of composition} \\ \iff & \phi(a)^{-1}1' = \phi(a)^{-1}\phi(a)\phi(a^{-1}) && \text{using the inverses property of } G' \\ \iff & \phi(a)^{-1} = \phi(a^{-1}) \end{aligned}$$

  which shows that ϕ maps $a^{-1} \in G$ to $\phi(a)^{-1} \in G'$.

□

For example, if $e \in G$ is the identity of G mapped to an element $e' = \phi(e) \in G'$, then for any $a \in G$ mapped to $a' = \phi(a) \in G'$,

$$a' = \phi(a) = \phi(ea) = \phi(e)\phi(a) = e'a'$$

And $a' = e'a' = a'e'$ means that e' is the identity in G'. Furthermore, the order of elements in G and G' will also be the same as,

$$a^n = e \iff e' = \phi(e) = \phi(a^n) = \phi(a)^n = (a')^n$$

Since two isomorphic groups have the same properties, it is often convenient to identify them with each other when speaking informally. For example, the symmetric group S_n of permutations of $\{1, \cdots, n\}$ is isomorphic to the group of permutation matrices, a subgroup of $GL_n(\mathbb{R})$ and we often blur the distinction between these two groups.

Notation. Sometimes when two groups are isomorphic this is indicated using the notation,

$$G \approx G'$$

Examples

- Let $C = \{\cdots, a^{-2}, a^{-1}, 1, a, a^2, \cdots\}$ be an infinite cyclic group. Then the map,

$$\phi : \mathbb{Z}^+ \longmapsto C \text{ s.t. } \phi(n) = a^n$$

  is an isomorphism where the preservation of the respective laws of composition can be seen as,

$$\phi(m+n) = a^{m+n} = a^m a^n = \phi(m)\phi(n)$$

and also $n + (-n) = 0$ and,

$$\phi(-n) = a^{-n} = (a^n)^{-1}.$$

- Let G be the set of real matrices of the form,

$$\begin{bmatrix} 1 & x \\ 0 & 1 \end{bmatrix}$$

This is a subgroup of $GL_2(\mathbb{R})$ and so, its law of composition is the same as that of $GL_2(\mathbb{R})$, i.e. matrix multiplication.

$$\begin{bmatrix} 1 & x \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & y \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & x+y \\ 0 & 1 \end{bmatrix}$$

So, G is isomorphic to $\mathbb{R}^+$, the additive group of reals.

Definition. *The groups isomorphic to a given group G form what is called the **isomorphism class** of G. Groups are often classified into isomorphism classes, for example, there is one isomorphism class of groups of order 3 and there are two classes of groups of order 4 and five classes of 12.*

Proposition 20. *There is only one isomorphism class for each order of cyclic group.*

Proof. Any two cyclic groups of the same order are isomorphic because, if

$$G = \{1, x, x^2, \cdots, x^{n-1}\},\ G' = \{1, y, y^2, \cdots, y^{n-1}\}$$

are two cyclic groups of order n then the map $\phi(x^i) = y^i$ is an isomorphism. □

Proposition 21. ***Cayley's Theorem*** *states that every group is isomorphic to a group of permutations of the same underlying set, or in other words, to a subgroup of the symmetric group acting on the group.*

Proof. Let G be a group and $x \in G$ and define $f_x : G \longmapsto G$ as $f_x(g) = xg$. Then f_x is a bijection because it has an inverse ${f_x}^{-1}(g) = x^{-1}g = f_{x^{-1}}(g)$. Therefore f_x is a permutation.

Now we define a map, from G to the symmetric group of permutations of G, that maps each element x in G to the permutation defined by f_x. Let $\phi : G \longmapsto Sym(G)$ be defined as $\phi(x) = f_x$ then,

- ϕ is homomorphic because

$$(f_x \circ f_{x'})(g) = f_x(f_{x'}(g)) = x(x'g) = xx'g = f_{xx'}(g)$$

  and so,

$$\phi(xx') = f_{xx'} = f_x \circ f_{x'} = \phi(x) \circ \phi(x').$$

- ϕ is injective because if $x, x' \in G$ such that $x \neq x'$ then $f_x(e) = x \neq x' = f_{x'}(e)$ is sufficient to show that $f_x \neq f_{x'}$. Or alternatively, the kernel of ϕ comprises the elements $k \in G$ such that $f_k(g) = kg = g \iff k = e$ so that the kernel is the trivial subgroup $\{e\}$.

Since ϕ is homomorphic, its image $im\,\phi$ is a subgroup of $Sym(G)$ and since ϕ is injective, it is in bijective correspondence with its image $im\,\phi \leq Sym(G)$. Therefore G is isomorphic to a subgroup of $Sym(G)$. □

Automorphisms

Definition. *The domain and codomain of an isomorphism can be the same set of objects so that $\phi : G \longmapsto G$. This is known as an* ***automorphism****.*

Example Let $G = \{1, x, x^2\}$ be a cyclic group of order 3 so that $x^3 = 1$. The transposition which interchanges x and x^2 is an automorphism of G,

$$\begin{array}{ccc} 1 & \longmapsto & 1 \\ x & \longmapsto & x^2 \\ x^2 & \longmapsto & x \end{array}$$

	1	x	x^2
1	1	x	x^2
x	x	x^2	1
x^2	x^2	1	x

$\longmapsto$

	1	x^2	x
1	1	x^2	x
x^2	x^2	x	1
x	x	1	x^2

This is because the group is cyclic and x and x^2 have the same order ($x^3 = 1$ and also $(x^2)^3 = x^6 = (x^3)^2 = 1^2 = 1$). So the law of composition is preserved.

Conjugation

The most important example of automorphism is conjugation.

Definition. ***Conjugation*** *by $b \in G$ is the map from G to itself defined by,*

$$\phi(a) = bab^{-1}$$

with the result that,

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

so that we can think of conjugation of a by b as the way that we need to change a if we want to move the multiplication by b to the other side.

This is an automorphism (known as an *inner automorphism*) because it

- is compatible with law of composition,

$$\phi(xy) = bxyb^{-1} = bxb^{-1}byb^{-1} = \phi(x)\phi(y).$$

- has an inverse so it is bijective,

$$(\phi^{-1} \circ \phi)(a) = \phi^{-1}(\phi(a)) = b^{-1}(bab^{-1})b = (b^{-1}b)a(b^{-1}b) = a.$$

  Note that this is different from the inverse element of a corresponding under the mapping ϕ,

$$\phi(a)\phi(a^{-1}) = bab^{-1}ba^{-1}b^{-1} = ba(1)a^{-1}b^{-1} = b(1)b^{-1} = 1.$$

A couple more important properties of the conjugate are as follows.

(i) In an abelian group where the composition law is commutative, conjugation becomes the identity map.

$$ba = ab \iff bab^{-1} = a \iff \phi(a) = a.$$

(ii) The inverse of the conjugate $bab^{-1} = b^{-1}ab$.

2.1.4 Homomorphisms

(tags: abstract algebra)

Definition. *A **homomorphism** is a mapping (not necessarily bijective) between two groups, $\phi : G \longmapsto G'$, such that,*

$$\phi(ab) = \phi(a)\phi(b) \text{ for all } a, b \in G$$

where ab represents composition according to the law of composition of G and $\phi(a)\phi(b)$ represents composition according to the law of composition of G'.

So, the difference between a *homomorphism* and a *isomorphism* is that the latter is bijective whereas the former is not. As a result, a *homomorphism* may be one-way only.

*A **homomorphism** is a **mapping** between two **groups**. That's to say, it is already assumed in the definition of a homomorphism that the codomain G' is a group.*

Examples of homomorphisms

(9) Let $C = \{a^{n-1}, \cdots, a^{-2}, a^{-1}, 1, a, a^2, \cdots, a^{n-1}\}$ be a finite cyclic group. Then the map,

$$\phi : \mathbb{Z}^+ \longmapsto C \text{ s.t. } \phi(n) = a^n$$

is a homomorphism. Note that if C were an infinite cyclic group then this would be an isomorphism.

(10) the sign of a permutation $sign : S_n \longmapsto \pm 1$

(11) the determinant function $det : GL_n(\mathbb{R}) \longmapsto \mathbb{R}^\times$

(12) an arguably trivial example is called the *inclusion* map $i : H \longmapsto G$ of a subgroup H into a group G, defined by $i(x) = x$. It functions as the identity for elements in the subgroup H but, since it is not surjective, there is no inverse mapping.

Image of a homomorphism

Since a homomorphism is not bijective it has an image different to the codomain group,

$$im\ \phi = \{\, x \in G' \mid \exists a \in G \text{ s.t. } \phi(a) = x \,\}$$

The image of a homomorphism is a subgroup of the codomain group G' because the homomorphism preserves the group structure as described in Proposition 19.

Notation. The image of the mapping ϕ with domain G is sometimes denoted $\phi(G)$.

Kernel of a homomorphism

Definition. *The **kernel** of a homomorphism is the set of elements in the domain that are mapped to the identity,*

$$ker\ \phi = \{\, a \in G \mid \phi(a) = 1' \,\}$$

Proposition 22. *The kernel of a homomorphism is a subgroup of the domain group G.*

Proof. If $a, b \in ker\ \phi$ then,

- closure: $\phi(ab) = \phi(a)\phi(b) = 1' \cdot 1' = 1'$ which shows that

$$a, b \in ker\ \phi \implies ab \in ker\ \phi.$$

- identity: By Proposition 19, $1' = e' = \phi(e)$ and so $e \in ker\ \phi$.
- inverses: Since $a \in ker\ \phi$, then

$$\begin{aligned} & 1' = e' = \phi(e) = \phi(aa^{-1}) = \phi(a)\phi(a^{-1}) = 1'\phi(a^{-1}) \\ \iff \quad & 1' = \phi(a^{-1}) \end{aligned}$$

so that $a \in ker\ \phi \iff a^{-1} \in ker\ \phi$.

□

Proposition 23. *If $\phi : G \longmapsto G'$ is a group homomorphism with kernel N then, for $a, b \in G$,*

$$\phi(a) = \phi(b) \iff \exists n \in N,\ s.t.\ b = an$$

or, equivalently, $a^{-1}b \in N$.

Proof.

$$\begin{aligned} & b = an \\ \implies \quad & \phi(b) = \phi(an) \\ \implies \quad & \phi(b) = \phi(a)\phi(n) && \text{by homomorphism property} \\ \implies \quad & \phi(b) = \phi(a)1' && \text{n is in the kernel} \\ \implies \quad & \phi(b) = \phi(a) \end{aligned}$$

$$\begin{aligned} & \phi(b) = \phi(a) \\ \implies \quad & \phi(a)^{-1}\phi(b) = 1' && \text{codomain is a group so has inverses} \\ \implies \quad & \phi(a^{-1})\phi(b) = 1' && \text{by Proposition 19} \\ \implies \quad & \phi(a^{-1}b) = 1' && \text{by homomorphism property} \\ \implies \quad & a^{-1}b = n \in N \\ \implies \quad & b = an \end{aligned}$$

□

Theorem 13. *A homomorphism is injective iff its kernel is the trivial subgroup $\{e\}$.*

When asked to prove the proposition that a homomorphism is injective iff its kernel is the trivial subgroup $\{e\}$, it's tempting to begin proving each direction of the bidirectional implication with a proof by contradiction (e.g. "Assuming there is a non-identity element in the kernel...") but the direct positive proof can be made very quick and simple with the above corollary.

Proof. Let $\phi : G \longmapsto G'$ be a homomorphism.

Assume that ϕ is injective and $k \in ker\,\phi$. Remembering that we always at least have $e_G \in ker\,\phi$,

$$\phi(k) = e_{G'} = \phi(e_G) \iff k = e_G$$

where the last implication is by the injectivity of ϕ.

Now assume that $ker\,\phi = \{e_G\}$, $a, b \in G$. Then,

$$\begin{aligned}
& \phi(a) = \phi(b) \\
\iff \quad & \phi(a)\phi(b)^{-1} = e_{G'} \\
\iff \quad & \phi(a)\phi(b^{-1}) = \phi(ab^{-1}) = e_{G'} && \text{using homomorphism properties} \\
\iff \quad & ab^{-1} \in ker\,\phi \\
\iff \quad & ab^{-1} = e_G && \text{by assumption } ker\,\phi = \{e_G\} \\
\iff \quad & a = b. && \square
\end{aligned}$$

Corollary 5. *A homomorphism is an isomorhpism if its kernel contains only the identity and its image is the whole of the codomain (i.e. it's surjective).*

Examples of Kernels of Homomorphisms

(13) The determinant function 11, $det : GL_n(\mathbb{R}) \longmapsto \mathbb{R}^\times$, has a kernel,

$$\{\,\text{real } n \times n \text{ matrices } A \mid det\,A = 1\,\},$$

which is a subgroup of $GL_n(\mathbb{R})$ known as the *special linear group* $SL_n(\mathbb{R})$.

(14) The sign of a permutation 10 has a kernel that is the set of *even* permutations,

$$A_n = \{\text{even permutations}\},$$

which is a subgroup of the symmetric group S_n and is known as the *alternating group*, A_n.

(15) The map from the additive group of integers to a finite cyclic group 9,

$$\phi : \mathbb{Z}^+ \longmapsto C \text{ s.t. } \phi(n) = a^n$$

has the kernel,

$$ker\ \phi = \{\, n \in \mathbb{Z}^+ \mid a^n = 1 \,\}$$

which has been proven to be a subgroup in Proposition 17.

2.1.5 Equivalence Relations and Partitions

(tags: abstract algebra)

Notation. In the following treatment of equivalence relations we will use the notation $a \sim b$ to denote the equivalence of a and b; $\overline{a}$ to indicate the equivalence class of a; and $\overline{S}$ to indicate the partition of S comprised of equivalence classes such as the class $\overline{a} = \overline{b}$ which includes both a and b.

Any map of sets $\phi : S \longmapsto T$ defines an equivalence relation on the domain S such that $a \sim b$ iff $\phi(a) = \phi(b)$. We will refer to this as the *equivalence relation determined by the map*. The corresponding partition is made up of the sets of elements in the domain S that are mapped to the same element in the codomain T.

Definition. *Let $\phi : S \longmapsto T$ be a map, then the* ***inverse image*** *of an element $t \in T$ is defined as,*

$$\phi^{-1}(t) = \{\, s \in S \mid \phi(s) = t \,\}$$

and can also be applied to a set $U \in T$ as,

$$\phi^{-1}(U) = \{\, s \in S \mid \phi(s) \in U \,\}$$

Note that in this notation, ϕ^{-1} ***does not indicate an inverse function*** *as the inverse of the function may not exist but the inverse image is nevertheless defined.*
The inverse images - the sets $\phi^{-1}(t)$ for all $t \in T$ - may also be called the ***fibres*** *of the map ϕ.*

Clearly, the non-empty fibres of the map ϕ form a partition of S. We can express this partition of S as a bijection, that we shall call $\overline{\phi}$, between the fibres of ϕ in S and the element of the image of S to which their members are mapped,

$$\overline{\phi} : \overline{S} \longmapsto im\ \phi$$

so that,

$$\overline{\phi}(\overline{s}) = \phi(s).$$

Congruence

Since a homomorphism maps the identity to the identity and inverses to inverses (Proposition 19), we can deduce that the inverse image of the identity in G' is going to contain at least the identity of G and that the inverse image of an element $(a')^{-1} \in G'$ will contain at least the element $a^{-1} \in G$. So, in terms of equivalence classes we can say that, for a homomorphism ϕ,

$$1 = e \in \phi^{-1}(1') \implies \overline{\phi}(\overline{1}) = 1'$$

$$a^{-1} \in \phi^{-1}((a')^{-1}) \implies \overline{\phi}(\overline{a^{-1}}) = (a')^{-1}$$

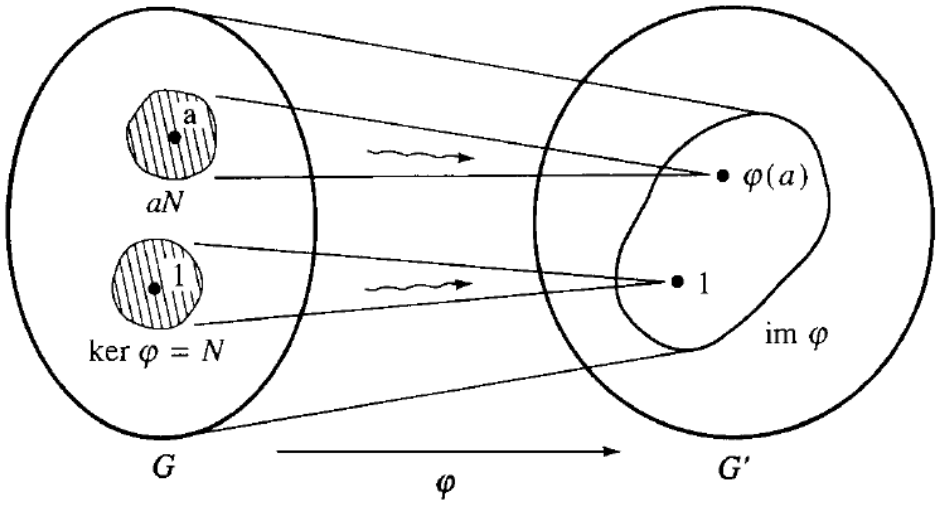


Figure 2.1: A schematic diagram of a group homomorphism

Definition. *The equivalence relation determined by a homomorphism is known as* ***congruence*** *and is commonly denoted using* $\equiv$ *instead of* $\sim$*. For a homomorphism* ϕ*,*

$$a \equiv b \iff \phi(a) = \phi(b).$$

Since ϕ *is a homomorphism we also have,*

$$a \equiv b \iff \phi(ac) = \phi(bc),\ \phi(a^{-1}) = \phi(b^{-1}).$$

More generally, a ***congruence relation*** *is an equivalence relation on an algebraic structure (such as a group, ring, or vector space) that is compatible with the structure in the sense that algebraic operations done with equivalent elements will yield equivalent elements.*

Congruence Examples

(16) The modulus function of complex numbers forms a homomorphism from the multiplicative group of complex numbers to the multiplica-

tive group of reals,

$$\phi : \mathbb{C}^\times \longmapsto \mathbb{R}^\times \text{ s.t. } \phi(a) = |a|$$

and the induced equivalence relation is $a \equiv b \iff |a| = |b|$. The fibres of this map are the concentric circles about 0. They are in bijective correspondence with elements of $im\ \phi$, the set of positive reals.

2.1.6 Cosets

(tags: abstract algebra)

The set of elements of the form an - described in Proposition 23 - is denoted by aN and is called a *coset* of N in G.

Definition. *A coset can be defined for any subgroup H of a group G. A **left coset** is a subset of the form,*

$$aH = \{\, ah \mid h \in H \,\}.$$

Cosets are not, in general, subgroups. This can be easily seen as the left coset aH does not contain the identity as, although H contains the identity, aH contains $a1 = a$.

Note that the arbitrary subgroup H could also be thought of as a coset $1H = H$ and also that the left cosets aH are equivalence classes for the congruence relation,

$$a \equiv b \iff b = ah,\ h \in H.$$

This is a congruence because, for some arbitrary $c \in G$,

$$1 \equiv c \iff \exists h \in H \text{ s.t. } c = 1h = h.$$

That's to say, the elements that are congruent to the identity are precisely the members of the subgroup H so that it plays a similar role to the kernel N in Proposition 23. Furthermore, since the congruence relation is an equivalence relation it forms a partition of the domain G.

Proposition 24. *For a group G with a subgroup H and $x \in G$, the coset xH is equal to H iff $x \in H$.*

Proof. Assume $x \in H$. Then $\forall xh \in xH\ .\ xh \in H$. Therefore $xH \subseteq H$. Conversely, $x^{-1} \in H$ so,

$$\forall h \in H\ .\ x^{-1}h \in H \implies x(x^{-1}h) = h \in xH.$$

Therefore $H \subseteq xH$ and so $H = xH$.

Now assume that $H = xH$. Since $e \in H$ then $xe = x \in xH = H$ and so $x \in H$. □

Proposition 25. *The left cosets of a subgroup partition the group.*

Proof. The left cosets are equivalence classes and, as a result, they partition the group. □

Examples of cosets

(17) The coset of an element with the kernel N,

$$aN = \{\, g \in G \mid g = an,\ n \in N \,\}$$

is the set of all elements that are *congruent* to a. The *congruence classes* are precisely the cosets aN for each $a \in G$. They are also the *fibres* of the homomorphic map.

(18) 2.1.1 Continuing the example of the symmetric group S_3 represented as

$$G = \{1, x, x^2, y, xy, x^2y\}$$

with group multiplication rules,

$$x^3 = 1, y^2 = 1, yx = x^2y.$$

The element xy has order 2 so it generates a cyclic subgroup $H = \{1, xy\}$ of order 2. The left cosets of H in G are the three sets,

$$\{1, xy\} = 1H = xyH,\ \{x, x^2y\} = xH = x^2yH,\ \{x^2, y\} = x^2H = yH.$$

Note that they do partition the group G. Also, notice that the cosets aH for $a \in H$ produce the subgroup H itself as should be expected as the group properties of the subgroup dictate that all products of its elements are already present in the subgroup. For this reason, the cosets aH that are distinct from H are those such that $a \notin H$.

(19) Let $G = (\mathbb{R}^3, +)$ be the group of 3d vectors with vector addition and $\vec{\boldsymbol{w}} \in G$. Then if

$$H = \{\, \vec{\boldsymbol{x}} \in G \mid \vec{\boldsymbol{w}}^T \vec{\boldsymbol{x}} = \vec{\mathbf{0}} \,\}$$

then H is a subgroup, $H \leq G$. H is a vector space representing a plane through the origin in $\mathbb{R}^3$ and it's cosets are

$$\vec{\boldsymbol{v}} + H = \{\, \vec{\boldsymbol{v}} + \vec{\boldsymbol{h}} \mid \vec{\boldsymbol{v}} \in G, \vec{\boldsymbol{h}} \in H \,\}$$

which are the affine spaces representing the translated planes, parallel to H, but not passing through the origin. Once again we see that the cosets partition the space even if there may be an infinite number of them.

The index of a subgroup

Definition. *The **index** of a subgroup is the number of left cosets it forms in the parent group.*

Notation. The **index** of a subgroup H in G is denoted by $[G : H]$.

In the example (18) the index of H is 3. Note that if G were to contain infinitely many elements then the index of a subgroup may also be infinite.

Proposition 26. *Each coset aH has the same number of elements as H.*

Proof. As usual, equal cardinality is demonstrated by showing the existence of a bijection. It is clear that there is a bijective map between the subgroup H and any coset aH because the map $H \longmapsto aH$ is,

- injective because $ah = ah' \implies h = h'$ because by group properties a has an inverse in G;
- surjective because every $c \in aH$ has the form ah and is therefore mapped to by some $h \in H$.

□

Lagrange's Theorem

Since the left cosets of H in G form a partition of G and their order is the same as that of H we see that the order of G is the order of H multiplied by its index in G. This results in a formula known as the *Counting Formula* as follows,

$$|G| = |H| \cdot [G : H].$$

If G is of infinite order and H is finite, then the index of H in G will be infinite.

Theorem 14. ***Lagrange's Theorem:*** *Let G be a finite group, and let H be a subgroup of G. The order of H divides the order of G.*

Corollary 6. *Let G be a finite group, and let a be an element of G. Then the order of a divides the order of G. That's to say, the order of the cyclic group generated by a, $|\langle a \rangle|$, divides $|G|$.*

Corollary 7. *If G is a group of order n, then $g^n = e$ for every element g of G.*

Proof. This is clearly a consequence of the previous corollary. If we let the order of g be m, then by the previous corollary,

$$m \mid n \iff n = km \text{ for } k \in \mathbb{N} \iff g^n = g^{km} = (g^m)^k = e^k = e. \qquad \square$$

Corollary 8. *Suppose that a group G has p elements and that p is a prime integer. Let $a \in G$ be any element, not the identity. Then G is the cyclic group $\{1, a, \ldots, a^{p-1}\}$ generated by a.*

Proof. Since $a \neq 1$ by selection, it has order greater than 1. Since its order must divide the order of G, which is prime, its order is equal to the order of G, p. So, the order of the nonidentity element a is the same as the order of G and so it generates the whole group. $\square$

Corollary 9. *All groups with some prime order, p, are in the same isomorphism class.*

Proof. Any group with prime order p is the cyclic group of order p and by Proposition 20 there is only a single isomorphism class for each cyclic group of a given order. □

Proposition 27. *Suppose the elements x, y in a group G have orders m, n respectively and that the $gcd(m, n) = 1$. Then $\langle x \rangle \cap \langle y \rangle = \{e\}$.*

> *Here we will prove, using Lagrange's Theorem, something that we previously proved here (Proposition 18) using modular arithmetic. Notice how the proofs are similar but the proof with Lagrange's Theorem allows us to remain within Group Theory.*

Proof. Firstly, note that the intersection of the two cyclic groups,

$$H = \langle x \rangle \cap \langle y \rangle$$

is a subgroup both of the parent group G **and** of $\langle x \rangle$ and $\langle y \rangle$. So Lagrange's Theorem tells us that its order must divide into the order of the parent group **and** the orders of the cyclic groups of x and y. Therefore, we have,

$$|H| \mid m \quad \text{and} \quad |H| \mid n.$$

Now, applying the fact that the $gcd(m, n) = 1$ we see that $|H| \mid 1$ and therefore $|H| = 1$. Furthermore, any group of order 1 must be the minimal group $\{e\}$. □

Proposition 28. *Suppose that H is a subgroup of G and $x \in G$. Then there exists some $k \in \mathbb{N}$, $1 \leq k \leq [G : H]$ s.t. $x^k \in H$.*

Proof. Let $n = [G : H]$ be the index of H in G. Then there are precisely n cosets of H in G. But $x \in G \implies x^m \in G$ for any $m \in \mathbb{N}$ (we don't need to consider the negative powers of x because they are inverses of positive powers and are similar for these purposes) and so we have cosets of the form $x^m H$ for each $m \in \mathbb{N}$. Therefore, amongst the $n + 1$ cosets generated by, $x^i H$ for

$i \in \{0, 1, \ldots, n\}$ we must have at least one repetition of the same coset. So, for some fixed x^i, x^j with $0 \leq i, j \leq n$ and $i \neq j$, we have,

$$\begin{aligned} & x^i H = x^j H \\ \iff \quad & \forall h \in H \,.\, x^i h \in x^j H \\ \iff \quad & \forall h \in H \,.\, \exists h' \in H \,.\, x^i h = x^j h' \\ \iff \quad & \forall h \in H \,.\, \exists h' \in H \,.\, x^{i-j} = h^{-1} h' \in H \end{aligned}$$

Since necessarily we have $1 \leq i - j \leq n$ we let $k = i - j$ and then $x^k \in H$ as required. □

Example applications of Lagrange Theorem

(20) **Fermat's Little Theorem**: *If p is a prime number then*

$$a^p \equiv a \bmod p \text{ for all } a \in \mathbb{Z}.$$

We need to be a little careful here. We might assume – given that we are multiplying the integer a in modulo p that the group we want to use is $(\mathbb{Z}_p, \otimes)$. However, this is not a group! The reason is that $\mathbb{Z}_p$ contains 0 which has no inverse under the proposed law of composition, multiplication.
If, however, we take $\mathbb{Z}_p^*$ where the $*$ means $\mathbb{Z}/\{0\}$ then we have a set of $p-1$ distinct elements. Over this set we can form the multiplicative group $G = (\mathbb{Z}_p^*, \otimes)$ because the primality of p means that every element has a multiplicative inverse.
*Note that this is **not** a group of prime order. The primality of p is essential to make sure that every element has a multiplicative inverse but, since we also have to eliminate 0 for the same reason, the order is $p-1$ which is not necessarily prime.*

Proof. Take the set $\mathbb{Z}_p^*$ under multiplication and some arbitrary $a \in \mathbb{Z}$.

(i) Primality of p means that it is possible to find $1 = na + mp$ for $m, n \in \mathbb{Z}$ (see Corollary 2). This implies that there exists a multiplicative inverse of every non-zero element in modulo p. Specifically, n is the inverse of a because $na = (-m)p + 1 \iff na \bmod p \equiv 1$.

(ii) Existence of the multiplicative inverses implies that we have a group $G = (\mathbb{Z}_p^*, \otimes)$.

(iii) G being a group implies that, for any element $a \in G$, by Corollary 7 we have $a^{p-1} = 1$.

(iv) In G, $a^{p-1} = 1 \iff a^p = a$ which translates to $a^p \equiv a \bmod p$.

□

Lagrange's Theorem and Homomorphisms

The Counting Formula can also be applied when a homomorphism is given. Let $\phi : G \longmapsto G'$ be a homomorphism. As we saw in coset example **??**, the left cosets of $ker\ \phi$ are the fibres of the map ϕ. They are in bijective correspondence with the elements of the image. Therefore,

$$[G : ker\ \phi] = |im\ \phi| .$$

Which implies that,

Corollary 10. *If $\phi : G \longmapsto G'$ is a homomorphism of finite groups then,*

$$|G| = |ker\ \phi| \cdot |im\ \phi| .$$

As a result, $|ker\ \phi|$ divides $|G|$, and $|im\ \phi|$ divides both $|G|$ and $|G'|$.

Restriction of a Homomorphism to a Subgroup

A useful way of understanding the structure of a complicated group is to understand its subgroups and then derive an understanding of the parent group from knowledge about the subgroups it contains. This frequently involves the application of Lagrange's Theorem. *Restriction of a Homomorphism to a subgroup* refers to studying the behaviour of a homomorphism on subgroups of the parent group.

Suppose that $\phi : G \longmapsto G'$ is a homomorphism and that H is a subgroup of G. Then we may *restrict ϕ to H* to obtain a homomorphism whose domain is a subset of the original,

$$\phi|_H : H \longmapsto G'.$$

This *restriction* is a homomorphism because ϕ is a homomorphism and the restriction domain is a group. Clearly, the kernel of the restricted homomorphism is the intersection of the domain H with *ker* ϕ.

Examples of using Lagrange's Theorem with a homomorphism restricted to a subgroup

(21) Referring again to the sign of a permutation (10) $S_n \longmapsto \{-1, 1\}$: the order of the codomain of this homomorphism is clearly 2. Suppose we form the restriction of this homomorphism to a subgroup H of S_n. Then, denoting the image by $\phi|_H(H)$, by Corollary 10 we have that $|\phi|_H(H)|$ divides both 2 and $|H|$.
So, if the subgroup H has odd order then $|\phi|_H(H)| = 1$ and – since $\phi|_H(H)$ must be a group because the group structure is preserved across the homomorphism – $\phi|_H(H) = \{1\}$. This means that H is in the kernel of the sign map and that the subgroup of permutations in S_n represented by H consists of only even permutations.
Therefore, every permutation whose order in S_n is odd is an even permutation (since the cyclic group that it generates has odd order). However, we can not make any conclusions about permutations of even order; they may be even or odd permutations.

Right Cosets

Right cosets also exist and are defined as,

$$Ha = \{\, g \in G \mid g = ha,\ h \in H \,\}$$

and these are equivalence classes for the *right congruence* relation,

$$a \equiv b \iff b = ha,\ h \in H.$$

Right cosets are not necessarily the same as left cosets. For instance, continuing the example in 18, the right cosets of the subgroup $\{1, xy\}$ of S_3 are,

$$\{1, xy\} = H1 = Hxy,\ \{x, y\} = Hx = H,\ \{x^2, x^2y\} = Hx^2 = Hx^2y.$$

Note that this generates a different partition of G then was generated by the left cosets.

2.1.7 Normal Subgroups and Centers

(tags: abstract algebra)

Normal Subgroups

Definition. *A subgroup N of a group G is called a* ***normal subgroup*** *if it has the property that,*

$$\forall\, a \in N, b \in G,\ bab^{-1} \in N$$

which is to say, that the conjugate by any element of G of any element in N is also in N.

Proposition 29. *A subset H of a group G is normal if and only if every left coset is also a right coset. If H is normal then,*

$$\forall a \in G,\ aH = Ha.$$

Proof. Suppose that H is normal. For any $h \in H$ and any $a \in G$,

$$ah = (aha^{-1})a.$$

Since H is normal, the conjugate by h of a is also in H, that's to say, $aha^{-1} \in H$ which implies that $(aha^{-1})a \in Ha$. Therefore, any arbitrary member of aH is also a member of Ha. Clearly, the same proof also works in the other

direction so that any member of Ha is also a member of aH and the two cosets are equal. So, we have shown that (H is normal) $\implies$ (left and right cosets of H are equal).
Now we need to show that (left and right cosets of H are equal) $\implies$ (H is normal). Firstly, clearly the above logic doesn't apply if H is not normal; there will be at least one element whose conjugate is not in H so $aH \neq Ha$. However, it could still be the case that each left coset is also a right coset if, for every a in G, there is some b in G such that $aH = Hb$. However, this is not possible because aH and Ha both contain a which means that in a given partition of G they must be the same partition. So $aH \neq Ha$ implies that the partitions are different; Ha creates different equivalence classes. Therefore (left and right cosets of H are equal) $\implies$ (H is normal). □

This is really the point of normal subgroups: That their cosets contain multiplication from both sides. As a result, when the cosets themselves are used as members of groups (see: Quotient Groups 2.1.8), we can define a group composition operation between them such that $aHbH = abH$ and the operation is well defined (i.e. equal arguments give equal results) because,

$$aHbH = abHH = abH \quad \textit{and} \quad aH = a'H \implies abH = aHb = a'Hb = a'bH$$

where $aH = Ha$ because H is normal. .

Examples of Normal Subgroups

(22) The kernel of a homomorphism is a normal subgroup because,

$$\begin{aligned} & a \in ker\ \phi \iff \phi(a) = 1 \\ \implies \quad & \phi(bab^{-1}) = \phi(b) \cdot 1 \cdot \phi(b^{-1}) \\ \iff \quad & \phi(bab^{-1}) = \phi(b)\phi(b)^{-1} \qquad \text{using Proposition 19} \\ \iff \quad & \phi(bab^{-1}) = 1. \end{aligned}$$

For example,

a. $SL_n(\mathbb{R})$ 13 is a normal subgroup of $GL_n(\mathbb{R})$ *even though it is not Abelian*, which can be seen as for $M \in GL_n(\mathbb{R}), A, B \in SL_n(\mathbb{R})$,

$$AB \neq BA \quad \text{and} \quad det\, A = 1,\ det\, M^{-1} = 1/(det\, M)$$

so that,

$$det\, M^{-1}AM = (det\, M^{-1}) \cdot 1 \cdot (det\, M) = (det\, M)/(det\, M) = 1.$$

b. A_n 14 is a normal subgroup of the symmetric group S_n. TODO: explain a bit further (e.g. any group of matrices with the same determinant will be normal)

(23) Any subgroup of an abelian group is normal because when the composition law is commutative, as was mentioned in the section on conjugation,

$$ba = ab \iff bab^{-1} = abb^{-1} = a$$

so that conjugation becomes the identity map and so, trivially, all conjugates of elements in a subgroup are also in the subgroup.

Subgroups of non-abelian groups, however, need not be normal. For example,

(24) Group T of invertible upper triangular matrices is not a normal subgroup of $GL_2(\mathbb{R})$. To show this note,

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, B = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, BAB^{-1} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$$

where $A \in T, B \in GL_2(\mathbb{R})$ but $BAB^{-1} \notin T$.

Proposition 30. *Let $\phi : G \longmapsto G'$ be a homomorphism and let H' be a subgroup of G'. Denote the inverse image $\phi^{-1}(H') = \{\, x \in G \mid \phi(x) \in H' \,\}$ by $\tilde{H}$. Then,*

(i) $\tilde{H}$ is a subgroup of G.

(ii) If H' is a normal subgroup of G' then $\tilde{H}$ is a normal subgroup of G.

(iii) $\tilde{H}$ contains ker ϕ.

(iv) The restriction of phi to $\tilde{H}$ defines a homomorphism $\tilde{H} \longmapsto H'$ whose kernel is ker ϕ.

Proof. Proofs are as follows:

(i) $\tilde{H}$ is a subgroup of G because ϕ is a homomorphism and its image, H', is a group (which is required for a homomorphism).

(ii) If H' is a normal subgroup of G' then $\tilde{H}$ is a normal subgroup of G because for every element in $\tilde{H}$ the mapped element is in H'. Then, since H' is normal, the conjugates of the mapped element are also in H' which means that their inverse images are in $\tilde{H}$. Since the map is homomorphic, the inverse images of the conjugates in G' are the respective conjugates in G.

(iii) $\tilde{H}$ contains *ker* ϕ because it contains every element in G that maps to an element in H' and, since H' is a group, it includes the identity of G'. Therefore $\tilde{H}$ contains every element that maps to the identity of G' which is *ker* ϕ.

(iv) The restriction of *phi* to $\tilde{H}$ is clearly a homomorphism and, since it contains *ker* ϕ, its kernel is equal to the kernel of ϕ.

□

The Center of a Group

Definition. *The* ***center*** *of a group G is the set of elements that commute with every element of G,*

$$Z = \{\, z \in G \mid zx = xz,\ \forall x \in G \,\}.$$

We can also define,

$$C(x) = \{\, g \in G \mid gx = xg \,\}$$

as the set of elements in G that commute with a single fixed element x.

Notation. The **center** of a group G may be denoted by Z or by $Z(G)$.

The center of a group, Z, is a subgroup of G. This can be easily seen as, first of all, Z is non-empty because the identity is in the center of any group.

Then, also, the center Z is closed under the group operation,

$$\forall a, b \in Z, x \in G \,.\, (ab)x = axb = x(ab)$$

and it contains the inverses,

$$\forall a \in Z, x \in G \,.\, ax = xa \iff a^{-1}ax = x = a^{-1}xa \iff xa^{-1} = a^{-1}x.$$

Examples of group centers

(25) Let $G = GL(2, \mathbb{R})$ be the group of invertible 2x2 matrices with real coefficients and take two elements in G,

$$M = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}, \quad N = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

Then we can identify the center of M by observing that an arbitrary matrix in $C(M)$ satisfies,

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 2a & b \\ 2c & d \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 2a & 2b \\ c & d \end{pmatrix}$$

which gives $b = 2b,\ c = 2c$ implying that b and c are 0. So,

$$C(M) = \left\{ \begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix} \,\middle|\, a, d \in \mathbb{R} \setminus \{0\} \right\}.$$

While matrices in the center of N satisfy,

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} a & a+b \\ c & c+d \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a+c & b+d \\ c & d \end{pmatrix}$$

which gives $a = a + c,\ c + d = d,\ a + b = b + d$ implying that $c = 0$ and $a = d$. Therefore,

$$C(N) = \left\{ \begin{pmatrix} a & b \\ 0 & a \end{pmatrix} \,\middle|\, a, b \in \mathbb{R},\ a \neq 0 \right\}.$$

Note that in both cases some coefficients were required to be non-zero because to be members of the general linear group they must be invertible and so their determinant must be non-zero.

(26) The center of the general linear group $GL_n(\mathbb{R})$ is the group of *scalar matrices* of the form cI for $c \in \mathbb{R}$, i.e. matrices of the form,

$$\begin{bmatrix} c & 0 \\ 0 & c \end{bmatrix}$$

in $GL_2(\mathbb{R})$. Note that, for *diagonal* matrices whose elements on the main diagonal are all non-zero but not-necessarily equal as in *scalar* matrices, multiplication is commutative with other diagonal matrices but not generally so with other matrices in the general linear group.

2.1.8 Products Groups and Quotient Groups

(tags: abstract algebra, modular arithmetic)

Definition. *If we take the cartesian product of two sets then:*

- *if the two sets are the underlying sets of two distinct groups then we have no way to combine them (as there is no common group operation) but we can take the pairing and define a component-wise multiplication over the pairs where each component is multiplied using that group's composition operation. In this way we create a new group over the pairs.*
- *if the two sets are subsets of a common group then there is a common group operation between them and so we can multiply them using this group operation. The result is another subset of the common group (not necessarily a subgroup).*

Both of these may at times be referred to as ***Product Groups*** *but the first one is more specifically referred to as a* ***Direct Product*** *and the second one may be referred to as a* ***Product Set****.*

Direct Products

Definition. *Let G, G' be two groups. The* ***direct product*** *is the set $G \times G'$ with component-wise multiplication using the group composition operation for the group corresponding to the component. Its order is the product of the orders of G and G'.*

Notation. The **direct product** of the two groups G, G' may be denoted by $G \times G'$ or GG'. In the case of Abelian groups the direct product may be referred to as the **direct sum** and denoted $G \oplus G'$.

So, if $a, b \in G$ and $a', b' \in G'$ then

- $(a, a'), (a, b'), (b, a'), (b, b') \in G \times G'$
- $(a, a')(b, b') = (ab, a'b')$

- the identity is $(1,1)$ and $(a,a')^{-1} = (a^{-1}, a'^{-1})$.

Definition. *The* ***projections*** *of a direct product $G \times G'$ are the maps p, p' such that,*

$$p(x,x') = x, \quad p'(x,x') = x'.$$

Proposition 31. *The* ***mapping property of direct products****: Let H be any group. The homomorphisms $\Phi : H \longmapsto G \times G'$ are in bijective correspondence to pairs (ϕ, ϕ') of homomorphisms*

$$\phi : H \longmapsto G, \qquad \phi' : H \longmapsto G'.$$

The kernel of Φ is the intersection $(ker\,\phi) \cap (ker\,\phi')$.

Proof. Given a pair of homomorphisms (ϕ, ϕ') we can define $\Phi(x) = (\phi(x), \phi'(x))$. Then this is homomorphic because,

$$\Phi(xy) = (\phi(xy), \phi'(xy)) = (\phi(x), \phi'(x))(\phi(y), \phi'(y)) = \Phi(x)\Phi(y).$$

Conversely, given such a Φ we can recover the pair of homomorphisms with the group projections as such (outer parentheses omitted for clarity),

$$\phi(x), \phi'(x) = p(\Phi(x)), p'(\Phi(x)).$$

Since the correspondence is invertible, it is a bijection.

Clearly, also,

$$\Phi(x) = (\phi(x), \phi'(x)) = (1,1) \iff (\phi(x) = 1) \wedge (\phi'(x) = 1)$$

so that $ker\,\Phi = (ker\,\phi) \cap (ker\,\phi')$. □

Proposition 32. *Let r, s be coprime integers. A cyclic group of order rs is isomorphic to the product of a cyclic group of order r and a cyclic group of order s.*

Proof. Let $C = \{1, x, x^2, \ldots, x^{rs-1}\}$, $C_1 = \{1, y, y^2, \ldots, y^{r-1}\}$, $C_2 = \{1, z, z^2, \ldots, z^{s-1}\}$ and define the map $\phi : C \longmapsto C_1 \times C_2$ as,

$$\phi(x^i) = (y^i, z^i).$$

Then ϕ is homomorphic because it is comprised of two homomorphisms (by the mapping proper Proposition 31),

$$\phi_1(x^i) = y^i \quad \text{and} \quad \phi_2(x^i) = z^i.$$

And ϕ is injective because,

$$\phi(x^i) = (1,1) \iff (y^i = 1) \wedge (z^i = 1) \iff (r \mid i) \wedge (s \mid i)$$

but r and s are coprime so this requires that $i = rs$ which is also the order of $x \in C$. So we have,

$$\phi(x^i) = (1,1) \iff x^i = x^{rs} = 1.$$

Therefore $ker\,\phi = \{1\}$ and, by Theorem 12, ϕ is injective.

Since ϕ is injective, its image has the same order as that of the domain C so we have,

$$|im\,\phi| = |C| = rs = |C \times C|$$

and ϕ is therefore surjective.

Therefore ϕ is a bijection and isomorphic. □

*Note that this is **only** the case for cyclic groups whose order is the product of two coprime numbers. For example, a cyclic group of order 4 is not isomorphic to a product of two cyclic groups of order 2 as every element in a product group $C_2 \times C_2$ has order 1 or 2. Whereas a cyclic group of order 4 has two elements of order 4 (the generating element and its inverse).*

Let $C_4 = \{1, x, x^2, x^3\}$, $C_2 = \{1, y\}$ and define the map $\phi : C_4 \longmapsto C_2 \times C_2$ as $\phi(x^i) = (y^i, y^i)$. Then,

$$\phi(x^i) = (1,1) \iff y^i = 1 \iff 2 \mid i$$

so that we have $ker\,\phi = \{1, x^2\}$ and so ϕ is not injective.

Product Sets

Definition. *Let A and B be subsets of the group G and denote the* ***product set*** *of A and B by*

$$AB = \{\, x \in G \mid x = ab \text{ for some } a \in A \quad \text{and} \quad b \in B \,\}.$$

Note that this notation is the same as one of the alternatives for the notation of the direct product so we need to be clear which is intended when we see this notation.

Relationship Between the Types of Product Groups

Proposition 33. *Let H and K be subgroups of G.*

(i) If $H \cap K = \{1\}$, the product map $p : H \times K \longmapsto G$ defined by $p(h,k) = hk$ is injective. Its image is the subset HK.

(ii) If either H or K is a normal subgroup of G, then the product sets HK and KH are equal and are subgroups of G.

(iii) If H and K are normal, $H \cap K = \{1\}$, and $HK = G$, then G is isomorphic to the direct product $H \times K$.

Proof. Proofs of each property are as follows.

(i) If we assume that $H \cap K = \{1\}$ then, for $h, h' \in H$, $k, k' \in K$,

$$p(h,k) = p(h',k') \iff hk = h'k' \iff (h')^{-1}h = k'k^{-1}$$

so that $(h')^{-1}h = k'k^{-1} \in H \cap K = \{1\}$ therefore,

$$(h')^{-1}h = 1 \iff h = h', \qquad k'k^{-1} = 1 \iff k' = k.$$

Therefore p is injective.

(ii) Assume w.l.o.g. that K is a normal subgroup. Then for all $k \in K, g \in G$, $g^{-1}kg \in K$ and, in particular, for $h \in H, h^{-1}kh \in K$. Therefore,

$$hk \in HK \implies h(h^{-1}kh) = kh \in HK$$

and conversely, using the fact that, $h^{-1} \in H \implies hkh^{-1} \in K$,

$$kh \in KH \implies (hkh^{-1})h = hk \in KH.$$

Therefore $HK = KH$ and this implies that HK is a subgroup because, for $h, h' \in H, k, k' \in K$,

- HK is closed because

  $$kh' \in KH = HK \implies kh' = h''k'' \in HK$$

  for some $h'' \in H$, $k'' \in K$, and so,

  $$hk, h'k' \in HK \implies (hk)(h'k') = h(kh')k' = h(h''k'')k' = (hh'')(k''k') \in HK.$$

- HK has inverses because for $hk \in HK$,

  $$h^{-1}k^{-1} \in HK \implies k^{-1}h^{-1} \in HK.$$

(iii) If $H \cap K = \{1\}$ then the product map p is injective and if $HK = G$ then $im\, p = G$ so p is surjective also and, therefore, is a bijection between $H \times K$ and G.

To show that p is a homomorphism between the direct product and the product set HK we need to show that,

$$p((h,k)(h',k')) = p((hh',kk')) = hh'kk' = hkh'k' = p((h,k))p((h',k'))$$

which will be true if $h'k = kh'$ which, in turn, will be the case if products in HK are commutative.
Now we have $H \cap K = \{1\}$ and so,

$$hk = kh \iff k^{-1}hk = h \iff k^{-1}hkh^{-1} = 1$$

implies that $H \cap K = \{k^{-1}hkh^{-1}\}$. So, if we can show that $k^{-1}hkh^{-1}$ is in both H and K then we have a homomorphism.
Since, in this case, both H and K are normal we have,

$$h, h^{-1} \in H,\ k \in K \implies k^{-1}hk \in H,\ hkh^{-1} \in K$$

which, by the group closure of H and K gives,

$$(k^{-1}hk)h \in H \quad \text{and} \quad k^{-1}(hkh^{-1}) \in K.$$

Therefore, if both H and K are normal subgroups and $H \cap K = \{1\}$, then $hk = kh$ for all $h \in H$, $k \in K$. This, in turn, means that the product map p is a homormorphism between the direct product $H \times K$ and the product set HK. Since p is also bijective, it is an isomorphism.

□

It is important to note that the product map of two subgroups $H \times K \longmapsto HK = G$ will not be a group homomorphism unless the two subgroups commute with each other.

Examples of Product Groups

(27) There is a group with subgroups of orders $1 \ldots 12$. It is a direct product of cyclic groups of orders

$$2 \times 2 \times 2 \times 3 \times 3 \times 5 \times 7 \times 11 = |G| = 27,720$$

so it looks like,

$$G = C_2 \times C_2 \times C_2 \times C_3 \times C_3 \times C_5 \times C_7 \times C_{11}.$$

Products of Cosets

It is possible to define a law of composition on the cosets of normal subgroups. This is because,

$$aH = Ha \implies aHbH = abHH = abH$$

so that we may define a law of composition such that $aH * bH = abH$ which closes over the set of cosets of H. The identity element of this composition is $eH = H$ and the inverse of the element aH is $a^{-1}H$.

*Note that this **only** applies to normal subgroups. The reason is that if H is not normal then there exists $h \in H$ and $a \in G$ such that $aha^{-1} \notin H$ which means that $S = aHa^{-1}H$ is not in any coset.*
This last claim can be proven if we observe – remembering that cosets partition the group – that S contains $a1a^{-1}1 = 1$ which means that it has to be in H. However, S also contains $aha^{-1}1 = aha^{-1} \notin H$ so, since these are equivalence classes, S cannot be in H.

Quotient Groups

Definition. *Suppose H is a normal subgroup of a group G. Then the quotient group G/N is the set of cosets of N in G with the coset product. Its order is the index of N in G, $[G : N]$.*

Notation. Sometimes – when it is not necessary to specify the subgroup against which the cosets are being formed – the set of cosets in G is denoted $\overline{G}$ and a member coset aH is denoted $\overline{a} \in \overline{G}$.

Theorem 15. *Every normal subgroup of a group G is the kernel of a homomorphism.*

Proof. For any normal subgroup $N \leq G$, if we define the map,

$$\pi : G \longmapsto G/N$$

then π is homomorphic because $\pi(ab) = abN = aNbN = \pi(a)\pi(b)$.
Now, Proposition 24 tells us that

$$\pi(x) = 1N = N \iff x \in N$$

which implies that $ker\,\pi = N$. (We could also observe that the cosets are equivalence classes and the kernel is the equivalence class containing the identity. The coset that contains the identity is the original subgroup $N = 1N$.) □

Theorem 16. ***First Isomorphism Theorem****: Let $\phi : G \longmapsto G'$ be a surjective group homomorphism, and let $N = ker\,\phi$. Then G/N is isomorphic to G' by the map $\overline{\phi}$ which sends the coset $\overline{a} = aN$ to $\phi(a)$,*

$$\overline{\phi}(\overline{a}) = \phi(a).$$

Proof. The non-empty fibres of ϕ are the cosets aN as seen in the example **??**. So, G/N can be thought of either as the cosets of the kernel of ϕ or as the non-empty fibres of ϕ. Then, $\overline{\phi}$ bijectively maps the cosets in G/N with the elements of the $im\,\phi$ and, because ϕ is surjective, we have $im\,\phi = G'$ so we have a bijection $\overline{\phi} : G/N \longmapsto G'$.
Also, the map $\overline{\phi}$ is homomorphic because coset multiplication is consistent with multiplication in the group,

$$\overline{\phi}(\overline{ab}) = \phi(ab) = \phi(a)\phi(b) = \overline{\phi}(\overline{a})\overline{\phi}(\overline{b}).$$

□

Examples of Quotient Groups

(28) Let $G = (\mathbb{Z}, +)$ and $H = \{\, 4n \mid n \in \mathbb{Z} \,\}$. Then the cosets of H are $\{\, z + 4n \mid z \in \mathbb{Z} \,\}$ and the product of two cosets,

$$(z_1 + H) + (z_2 + H) = (z_1 + z_2 + H).$$

(29) Let $G = (\mathbb{R}, +)$ and $H = \{\, 2n\pi \mid n \in \mathbb{Z} \,\}$. Then the cosets are the possible angles. This is an example of an infinite quotient group. The affine spaces in example 19 are another example of an infinite quotient group.

(30) In example 16 we saw that the modulus of complex numbers is a homomorphism from complex numbers to the reals. So, its kernel is the unit circle – the set of complex numbers of modulus 1. The cosets of the unit circle are the concentric circles,

$$C_r = \{\, z \mid |z| = r \,\}.$$

Applying the product of cosets gives us $C_r C_s = C_{rs}$ which works out

because,

$$
\begin{aligned}
& |(a+bi)(c+di)| = |(ac-bd)+(ad+bc)i| \\
\iff \quad & |(a+bi)(c+di)| = \sqrt{(ac-bd)^2+(ad+bc)^2} \\
\iff \quad & |(a+bi)(c+di)| = \sqrt{(a^2c^2+b^2d^2-2abcd)+(a^2d^2+b^2c^2+2abcd)} \\
\iff \quad & |(a+bi)(c+di)| = \sqrt{(a^2+b^2)(c^2+d^2)} \\
\iff \quad & |(a+bi)(c+di)| = \sqrt{(a^2+b^2)}\sqrt{(c^2+d^2)} \\
\iff \quad & |(a+bi)(c+di)| = |(a+bi)|\,|(c+di)|\,.
\end{aligned}
$$

TODO: Notes on Quotient Groups: Artin[81]

Modular Arithmetic

TODO: Describe modular arithmetic in terms of cosets: Artin[79] TODO: include in examples Abstract Maths ex. 14.9

2.2 Fields

(tags: abstract algebra)

2.2.1 Complex Numbers

(tags: abstract algebra, complex numbers)

Proposition 34. *For every $\alpha \in \mathbb{C}$, there exists a unique $\beta \in \mathbb{C}$ such that $\alpha + \beta = 0$.*

Proof. By contradiction: Say there are two such elements, β, γ such that,

$$\begin{aligned} \alpha + \beta = 0 = \alpha + \gamma \\ (\alpha + \beta) + \beta = (\alpha + \beta) + \gamma \\ 0 + \beta = \beta = 0 + \gamma = \gamma \end{aligned}$$

□

Proposition 35. *For every $\alpha \in \mathbb{C}$ with $\alpha \neq 0$, there exists a unique $\beta \in \mathbb{C}$ such that $\alpha\beta = 1$.*

Proof. By contradiction: Say there are two such elements, β, γ then,

$$\begin{aligned} \alpha\beta = 1 = \alpha\gamma \\ \beta = \frac{1}{\alpha} = \gamma \end{aligned}$$

□

2.2.2 Complex Numbers Problems

(tags: abstract algebra, complex numbers)

Find all the roots of $x^3 = 1$ for $x \in \mathbb{C}$ (tags: complex numbers)

Since $x^3 - 1 = (x-1)(x^2+x+1)$, we have (via zero-factor theorem) possible roots from,

$$x - 1 = 0 \iff x = 1$$

$$x^2 + x + 1 = 0 \implies x = \frac{-1 \pm \sqrt{-3}}{2} = \frac{-1 \pm \sqrt{3}\, i}{2}$$

More generally,

$$(a + bi) + (a - bi) = 2a$$

and since also,

$$\left[\frac{-1 + \sqrt{3}\, i}{2}\right]^2 = \frac{-1 - \sqrt{3}\, i}{2}$$

as well as the reverse,

$$\left[\frac{-1 - \sqrt{3}\, i}{2}\right]^2 = \frac{-1 + \sqrt{3}\, i}{2}$$

this means that if $x = \frac{-1 \pm \sqrt{3}\, i}{2}$ then $x^2 + x$ is of the form $(a+bi)+(a-bi) = 2a$ and so we have that $x^2 + x = -1 \iff x^2 + x + 1 = 0$.

In addition,

$$(a + bi)(a - bi) = a^2 + b^2$$

which means that if $x = \frac{-1 \pm \sqrt{3}\, i}{2}$ then $x^3 = x^2 x$ is of the form $(a+bi)(a-bi) = a^2 + b^2$ so we have that $x^3 = \frac{-1}{2}^2 + \frac{\sqrt{3}}{2}^2 = \frac{1}{4} + \frac{3}{4} = 1$.

So we see that - allowing for complex x - the cubic polynomial $x^3 - 1$ has 3 roots as we should expect from the Fundamental Theorem of Algebra *(is this the correct interpretation of this?)*.

2.2.3 Vector Spaces

(tags: vector spaces, polynomials, periodic functions)

2.2.4 Vector Space properties

(tags: vector spaces, polynomials)

2.2.5 Definition of a Vector Space

(tags: vector spaces)

A field F is a subfield of $\mathbb{C}$ if the following properties hold:

- If $a, b \in F$, then $a + b \in F$.
- If $a \in F$, then $-a \in F$.
- If $a, b \in F$, then $ab \in F$.
- If $a \in F$ and $a \neq 0$, then $a^{-1} \in F$.
- $1 \in F$.

Note that using the first, second and last of these axioms we can deduce that $1 - 1 = 0$ is an element of F.

Let F denote a field which is a subfield of $\mathbb{C}$ and V denote a vector space over F.

Definition. ***Addition, Scalar Multiplication***

- *An **addition** on a set V is a function that assigns an element $u+v \in V$ to each pair of elements $u, v \in V$.*
- *A **scalar multiplication** on a set V is a function that assigns an element $\lambda v \in V$ to each $\lambda \in F$ and each $v \in V$.*

Note that both functions are closed over V.

Definition. ***A vector space*** *is a set V along with an addition on V and a scalar multiplication on V such that the following properties hold:*

commutativity $\vec{u} + \vec{v} = \vec{v} + \vec{u}$ *for all* $\vec{u}, \vec{v} \in V$*;*

associativity $(\vec{u} + \vec{v}) + \vec{w} = \vec{u} + (\vec{v} + \vec{w})$ *and* $(ab)\vec{v} = a(b\vec{v})$ *for all* $\vec{u}, \vec{v}, \vec{w} \in V$ *and all* $a, b \in F$*;*

additive identity *there exists an element* $\vec{0} \in V$ *such that* $\vec{v} + \vec{0} = \vec{v}$ *for all* $\vec{v} \in V$*;*

additive inverse *for every* $\vec{v} \in V$ *there exists* $\vec{w} \in V$ *such that* $\vec{v} + \vec{w} = \vec{0}$*;*

multiplicative identity $1\vec{v} = \vec{v}$ *for all* $\vec{v} \in V$*;*

distributive properties $a(\vec{u} + \vec{v}) = a\vec{u} + a\vec{v}$ *and* $(a + b)\vec{u} = a\vec{u} + b\vec{u}$ *for all* $a, b \in F$ *and* $\vec{u}, \vec{v} \in V$*;*

2.2.6 Derived properties of a Vector Space

(tags: vector spaces)

Proposition 36. *A vector space contains a unique additive identity element.*

Proof. If $\vec{\mathbf{0}}'$ is also an additive identity then by the additive identity property,

$$\vec{\mathbf{0}} + \vec{\mathbf{0}}' = \vec{\mathbf{0}}$$

but since $\vec{\mathbf{0}}$ is also an additive identity,

$$\vec{\mathbf{0}}' + \vec{\mathbf{0}} = \vec{\mathbf{0}}'$$

Then, by the commutativity of vector addition,

$$\vec{\mathbf{0}} = \vec{\mathbf{0}} + \vec{\mathbf{0}}' = \vec{\mathbf{0}}' + \vec{\mathbf{0}} = \vec{\mathbf{0}}' \qquad \square$$

Proposition 37. *A vector space contains a unique additive inverse for each element.*

Proof. If $\vec{\boldsymbol{v}}$ and $\vec{\boldsymbol{w}}$ are both additive inverses of $\vec{\boldsymbol{u}}$ then, by the additive inverse property we have,

$$\vec{\boldsymbol{u}} + \vec{\boldsymbol{v}} = \vec{\mathbf{0}} \text{ and also } \vec{\boldsymbol{u}} + \vec{\boldsymbol{w}} = \vec{\mathbf{0}}$$

using the uniqueness of the additive identity,

$$\vec{\boldsymbol{u}} + \vec{\boldsymbol{v}} = \vec{\mathbf{0}} = \vec{\boldsymbol{u}} + \vec{\boldsymbol{w}}$$

Then, if we add one of the additive inverses of $\vec{\boldsymbol{u}}$ to both sides,

$$\vec{\boldsymbol{u}} + \vec{\boldsymbol{v}} + \vec{\boldsymbol{v}} = \vec{\boldsymbol{u}} + \vec{\boldsymbol{w}} + \vec{\boldsymbol{v}}$$

and use the associativity of vector addition,

$$\begin{aligned} (\vec{\boldsymbol{u}} + \vec{\boldsymbol{v}}) + \vec{\boldsymbol{v}} &= (\vec{\boldsymbol{u}} + \vec{\boldsymbol{v}}) + \vec{\boldsymbol{w}} \\ \vec{\mathbf{0}} + \vec{\boldsymbol{v}} &= \vec{\mathbf{0}} + \vec{\boldsymbol{w}} \\ \vec{\boldsymbol{v}} &= \vec{\boldsymbol{w}} \end{aligned} \qquad \square$$

Because additive inverses are unique we can use the notation $-\vec{\boldsymbol{v}}$ to denote the additive inverse of $\vec{\boldsymbol{v}}$. Then we define $\vec{\boldsymbol{w}} - \vec{\boldsymbol{v}}$ to mean $\vec{\boldsymbol{w}} + -\vec{\boldsymbol{v}}$.

Definition. *Vector Subtraction*

$$\vec{\boldsymbol{u}} - \vec{\boldsymbol{v}} \coloneqq \vec{\boldsymbol{u}} + -\vec{\boldsymbol{v}}$$

Proposition 38. $0\vec{\boldsymbol{v}} = \vec{\mathbf{0}}$ *for every* $\vec{\boldsymbol{v}} \in V$.

Note that this proposition is asserting something about scalar multiplication and the additive identity of V. The only part of the definition of a vector space that connects scalar multiplication and vector addition is the distributive property. Therefore the distributive property must be used in this proof.

Proof. Firstly take,

$$\vec{\boldsymbol{v}} + 0\vec{\boldsymbol{v}} = 0\vec{\boldsymbol{v}} + 1\vec{\boldsymbol{v}}$$

and then use the properties of the underlying field to say

$$(0+1)\vec{\boldsymbol{v}} = 1\vec{\boldsymbol{v}} = \vec{\boldsymbol{v}}$$

Now we have shown that,

$$\vec{\boldsymbol{v}} + 0\vec{\boldsymbol{v}} = \vec{\boldsymbol{v}}$$

which, by the definition and uniqueness of the additive identity, shows that $0\vec{\boldsymbol{v}} = \vec{\mathbf{0}}$. But if we want to continue algebraically we can now add the additive inverse to both sides,

$$(\vec{\boldsymbol{v}} + -\vec{\boldsymbol{v}}) + 0\vec{\boldsymbol{v}} = (\vec{\boldsymbol{v}} + -\vec{\boldsymbol{v}})$$
$$\vec{\mathbf{0}} + 0\vec{\boldsymbol{v}} = 0\vec{\boldsymbol{v}} = \vec{\mathbf{0}}$$ □

Another, simpler proof exists.

Proof. Using the underlying field properties and the distributivity of scalar vector multiplication,

$$0\vec{\boldsymbol{v}} = (0+0)\vec{\boldsymbol{v}} = 0\vec{\boldsymbol{v}} + 0\vec{\boldsymbol{v}}$$

and then adding the additive inverse to both sides,

$$(0\vec{\boldsymbol{v}} + -(0\vec{\boldsymbol{v}})) = (0\vec{\boldsymbol{v}} + -(0\vec{\boldsymbol{v}})) + 0\vec{\boldsymbol{v}}$$
$$\vec{\mathbf{0}} = \vec{\mathbf{0}} + 0\vec{\boldsymbol{v}} = 0\vec{\boldsymbol{v}}$$ □

Proposition 39. *$a\vec{\mathbf{0}} = \vec{\mathbf{0}}$ for every $a \in F$.*

Proof. Using the distributivity of scalar multiplication of vectors and the additive identity,

$$a\vec{\mathbf{0}} = a(\vec{\mathbf{0}} + \vec{\mathbf{0}}) = a\vec{\mathbf{0}} + a\vec{\mathbf{0}}$$

Then, adding the additive inverse to both sides,

$$(a\vec{\mathbf{0}} + -(a\vec{\mathbf{0}})) = a\vec{\mathbf{0}} + (a\vec{\mathbf{0}} + -(a\vec{\mathbf{0}}))$$
$$\vec{\mathbf{0}} = a\vec{\mathbf{0}} + \vec{\mathbf{0}} = a\vec{\mathbf{0}}$$ □

Proposition 40. $(-1)\vec{\boldsymbol{v}} = -\vec{\boldsymbol{v}}$ *for every* $\vec{\boldsymbol{v}} \in V$.

Proof. Using the distributivity of scalar multiplication of vectors and the underlying field properties we have,

$$(-1)\vec{\boldsymbol{v}} + \vec{\boldsymbol{v}} = (-1)\vec{\boldsymbol{v}} + 1\vec{\boldsymbol{v}} = (-1+1)\vec{\boldsymbol{v}} = 0\vec{\boldsymbol{v}} = \vec{\mathbf{0}}$$

Now we could add the additive inverse to both sides to show that,

$$\begin{aligned}(-1)\vec{\boldsymbol{v}} + (\vec{\boldsymbol{v}} + -\vec{\boldsymbol{v}}) &= \vec{\mathbf{0}} + -\vec{\boldsymbol{v}} \\ (-1)\vec{\boldsymbol{v}} + \vec{\mathbf{0}} &= \vec{\mathbf{0}} + -\vec{\boldsymbol{v}} \\ (-1)\vec{\boldsymbol{v}} &= \vec{\boldsymbol{v}}\end{aligned}$$

□

But we already have,

$$(-1)\vec{\boldsymbol{v}} + \vec{\boldsymbol{v}} = \vec{\mathbf{0}}$$

and this, by the definition of the additive inverse, proves that $(-1)\vec{\boldsymbol{v}}$ is an additive inverse of $\vec{\boldsymbol{v}}$. Since we have previously proven the uniqueness of the additive inverse in Proposition 37 we can conclude, in fact, that $(-1)\vec{\boldsymbol{v}} = -\vec{\boldsymbol{v}}$ the unique additive inverse of v.

2.2.7 The notation F^S

(tags: vector spaces)
If S is a set then F^S denotes the set of functions $S \mapsto F$.

Addition is defined as, for $f, g, (f+g) \in F^S$,

$$(f+g)(x) = f(x) + g(x)$$

for all $x \in S$.

Scalar multiplication is defined as, for $\lambda \in F, \lambda f \in F^S$,

$$(\lambda f)(x) = \lambda f(x)$$

for all $x \in S$.

Example: If S is the interval $[0,1]$ and $F = \mathbb{R}$ then $\mathbb{R}^{[0,1]}$ is the set of real-valued functions on the interval $[0,1]$. $\mathbb{R}^{[0,1]}$ is a vector space with additive identity $0 : [0,1] \mapsto \mathbb{R}$ defined as $0(x) = 0$ and the additive inverse of some function $f \in \mathbb{R}^{[0,1]}$ is the function defined as $(-f)(x) = -f(x)$.
Any *non-empty* set S in conjunction with a subset of $\mathbb{C}$ would similarly produce a vector space. In fact, the vector space F^n can be thought of as the space of functions from the set $\{1, 2, 3, \ldots, n\}$ to F. For example, vectors in 3-dimensional space can be viewed as:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} \equiv f : \{1,2,3\} \mapsto \mathbb{R} \text{ with } f(t) = \begin{cases} x & t = 1 \\ y & t = 2 \\ z & t = 3 \end{cases}$$

2.2.8 Polynomials as a vector space

(tags: vector spaces, polynomials)

A very important example involves treating a polynomial as a vector. A function $p : F \mapsto F$ is called a polynomial with coefficients in F if there exist $a_0, \ldots, a_m \in F$ such that,

$$p(z) = a_0 + a_1 z + a_2 z^2 + \cdots + a_m z^m$$

for all $z \in F$.
Then we can define a vector space, $P(F)$, to be the set of all polynomials with coefficients in F.

Addition on $P(F)$ is defined as,

$$(p+q)(z) = p(z) + q(z) \qquad \text{for } p, q \in P(F), z \in F$$

whose associativity is clear from the definition and the commutativity can be shown by,

$$\begin{aligned} ((p+q)+r)(z) &= (p+q)(z) + r(z) \\ &= p(z) + q(z) + r(z) \\ &= p(z) + (q+r)(z) \\ &= (p + (q+r))(z) \end{aligned}$$

Scalar multiplication on $P(F)$ is defined as,

$$(ap)(z) = ap(z) \qquad \text{for } p \in P(F), a, z \in F$$

whose associativity can be shown by substituting (ab) for a in the definition,

$$[(ab)p](z) = (ab)p(z)$$

Then, by the associativity of the multiplication of the elements of the field F we have,

$$(ab)p(z) = a[b(p(z)]$$

then we use the definition in reverse,

$$a[b(p(z)]) = a[(bp)(z)] = [a(bp)](z)$$

(compare with $(ab)\vec{\boldsymbol{v}} = a(b\vec{\boldsymbol{v}})$)

modeling Concretely, each $p(z) \in P(F)$ is a vector that could be modeled, say, as

$$\vec{\boldsymbol{p}} = \{\, (a_0, a_1, \ldots, a_m) \mid p(z) = a_0 + a_1 z + a_2 z^2 + \cdots + a_m z^m \in P(F) \,\}$$

2.2.9 Subspaces of vector spaces

(tags: vector spaces, polynomials, periodic functions)

2.2.10 Definition of a Subspace

(tags: vector spaces)

Definition. *A set U is a subspace of V if it is a subset of V and if the same addition and multiplication over U forms a vector space.*

Considering the required properties of a vector space, we can see that commutativity and associativity of the addition; associativity of the scalar multiplication; and distributivity of the scalar multiplication over the addition; will all be satisfied as we have the same addition and multiplication over a subset of the elements in V. That's to say, the vector space properties ensure that these properties hold $\forall \vec{\boldsymbol{v}} \in V$ and we have $\forall \vec{\boldsymbol{u}} \in U, \vec{\boldsymbol{u}} \in V$.
Furthermore, the multiplicative identity also holds $\forall \vec{\boldsymbol{v}} \in V$ so will also hold for every element of U.

So what remains to be proven to satisfy the requirements of a subspace?

- Existence of the additive identity
- Existence of an additive inverse for every element of U
- Closure of the addition and scalar multiplication over U

Note, however that - having proved in Proposition 40 that multiplication by -1 gives the additive inverse - closure of the scalar multiplication over U also implies the presence in U of the additive inverse of every element of U. So, actually, what we need to prove for U to be a subspace is only,

- $\vec{\mathbf{0}} \in U$
- Closure of the addition and scalar multiplication over U

2.2.11 A subspace of the polynomials

(tags: vector spaces, polynomials)

An example of a subspace of the polynomials, $P(F)$ is,

$$\{\, p \in P(F) \mid p(3) = 0 \,\}$$

Members of this subspace include:

- $p(z) = 3 - z$
- $p(z) = 9 - z^2$

- $p(z) = 3 - z + 3z^2 - z^3$
- $p(z) = 12z - 4z^2$
- ... etc.

To verify this we need to show that addition and multiplication are closed over this set and that $\vec{\mathbf{0}}$ is a member of the set. It's easy to see that $\vec{\mathbf{0}}$ is a member of the set as,

$$p(3) = 0 + 0(3) + 0(3)^2 + \cdots + 0(3)^m = 0$$

as required. Scalar multiplication is closed as,

$$ap(3) = a(0) = 0$$

whereas addition can be shown to be closed as,

$$(q + r)(3) = q(3) + r(3) = 0 + 0 = 0$$

Note that for values of $z \neq 3$, the closure of these functions is the same as for the general case of $P(F)$.

2.2.12 Sums and Direct Sums

(tags: vector spaces)

Definition. *If $U_1, U_2, \ldots, U_m$ are subspaces of V then their sum is defined as*

$$U_1 + U_2 + \cdots + U_m = \{ \vec{\boldsymbol{u_1}} + \vec{\boldsymbol{u_2}} + \cdots + \vec{\boldsymbol{u_m}} \mid \vec{\boldsymbol{u_1}} \in U_1, \vec{\boldsymbol{u_2}} \in U_2, \ldots, \vec{\boldsymbol{u_m}} \in U_m \}$$

The sum of the subspaces of V is also a subspace of V because,

- Closure of addition

$$\begin{aligned} &(\vec{\boldsymbol{u_1}} + \vec{\boldsymbol{u_2}} + \cdots + \vec{\boldsymbol{u_m}}) + (\vec{\boldsymbol{u_1'}} + \vec{\boldsymbol{u_2'}} + \cdots + \vec{\boldsymbol{u_m'}}) \\ &= (\vec{\boldsymbol{u_1}} + \vec{\boldsymbol{u_1'}}) + (\vec{\boldsymbol{u_2}} + \vec{\boldsymbol{u_2'}}) + \cdots + (\vec{\boldsymbol{u_m}} + \vec{\boldsymbol{u_m'}}) \\ &= \vec{\boldsymbol{v_1}} + \vec{\boldsymbol{v_2}} + \cdots + \vec{\boldsymbol{v_m}} \qquad \text{where } \vec{\boldsymbol{v_1}} \in U_1, \vec{\boldsymbol{v_2}} \in U_2, \ldots, \vec{\boldsymbol{v_m}} \in U_m \end{aligned}$$

- Closure of scalar multiplication

$$\begin{aligned} & a(\vec{\boldsymbol{u_1}} + \vec{\boldsymbol{u_2}} + \cdots + \vec{\boldsymbol{u_m}}) && \text{where } a \in F \\ &= a\vec{\boldsymbol{u_1}} + a\vec{\boldsymbol{u_2}} + \cdots + a\vec{\boldsymbol{u_m}} \\ &= \vec{\boldsymbol{v_1}} + \vec{\boldsymbol{v_2}} + \cdots + \vec{\boldsymbol{v_m}} && \text{where } \vec{\boldsymbol{v_1}} \in U_1, \vec{\boldsymbol{v_2}} \in U_2, \ldots, \vec{\boldsymbol{v_m}} \in U_m \end{aligned}$$

- Existence of $\vec{\mathbf{0}}$

$$\begin{aligned} & U_1, U_2, \ldots, U_m \text{ are subspaces} \\ \implies & \vec{\mathbf{0}} \in U_1, \vec{\mathbf{0}} \in U_2, \ldots, \vec{\mathbf{0}} \in U_m \\ \implies & \vec{\mathbf{0}} + \vec{\mathbf{0}} + \cdots + \vec{\mathbf{0}} \in U_1 + U_2 + \cdots + U_m \end{aligned}$$

  Note though, that this may not be the only way of producing $\vec{\mathbf{0}}$ from the sum of vectors of these subspaces. That's to say, there could be some $(\vec{\boldsymbol{u_1}} + \vec{\boldsymbol{u_2}} + \cdots + \vec{\boldsymbol{u_m}}) = \vec{\mathbf{0}}$ and this is a key difference from direct sums.

Proposition 41. *$U_1 + U_2 + \cdots + U_m$ is the smallest subspace of V containing $U_1, U_2, \ldots, U_m$.*

Proof. $U_1 + U_2 + \cdots + U_m$ is a subspace of V that contains $U_1, U_2, \ldots, U_m$ because we can obtain U_i by setting all the u_j for $j \neq i$ to $\vec{\mathbf{0}}$.

If a subspace of V contains $U_1, U_2, \ldots, U_m$ then, by the closure of addition, it must also contain $U_1 + U_2 + \cdots + U_m$.

Therefore the smallest subspace of V that contains $U_1, U_2, \ldots, U_m$ is $U_1 + U_2 + \cdots + U_m$. $\square$

Definition. *If $U_1, U_2, \ldots, U_m$ are subspaces of V then their **direct sum** is defined as,*

$$U_1 \oplus U_2 \oplus \cdots \oplus U_m = \{\, \vec{\boldsymbol{u_1}} + \vec{\boldsymbol{u_2}} + \cdots + \vec{\boldsymbol{u_m}} \mid \vec{\boldsymbol{u_1}} \in U_1, \vec{\boldsymbol{u_2}} \in U_2, \ldots, \vec{\boldsymbol{u_m}} \in U_m \,\}$$

such that,

$$\vec{\boldsymbol{u_1}} + \vec{\boldsymbol{u_2}} + \cdots + \vec{\boldsymbol{u_m}} = \vec{\mathbf{0}} \implies \vec{\boldsymbol{u_1}} = \vec{\mathbf{0}}, \vec{\boldsymbol{u_2}} = \vec{\mathbf{0}}, \ldots, \vec{\boldsymbol{u_m}} = \vec{\mathbf{0}}.$$

That the unique way of obtaining $\vec{\mathbf{0}}$ is for all of the vectors from each of the subspaces to be $\vec{\mathbf{0}}$ is equivalent to there only being a single unique way of obtaining each resultant vector from an addition of the vectors from the individual subspaces. This can be seen as,

$$\begin{aligned}\vec{\boldsymbol{u_1}} + \vec{\boldsymbol{u_2}} + \cdots + \vec{\boldsymbol{u_m}} &= \vec{\boldsymbol{u'_1}} + \vec{\boldsymbol{u'_2}} + \cdots + \vec{\boldsymbol{u'_m}} \\ (\vec{\boldsymbol{u_1}} + \vec{\boldsymbol{u_2}} + \cdots + \vec{\boldsymbol{u_m}}) - (\vec{\boldsymbol{u'_1}} + \vec{\boldsymbol{u'_2}} + \cdots + \vec{\boldsymbol{u'_m}}) &= \vec{\mathbf{0}} \\ (\vec{\boldsymbol{u_1}} - \vec{\boldsymbol{u'_1}}) + (\vec{\boldsymbol{u_2}} - \vec{\boldsymbol{u'_2}}) + \cdots + (\vec{\boldsymbol{u_m}} - \vec{\boldsymbol{u'_m}}) &= \vec{\mathbf{0}}\end{aligned}$$

Therefore, since vector spaces always contain $\vec{\mathbf{0}}$ and so we will always have the representation,

$$\vec{\mathbf{0}} + \vec{\mathbf{0}} + \cdots + \vec{\mathbf{0}} = \vec{\mathbf{0}}$$

if this is the unique representation of $\vec{\mathbf{0}}$ then it follows that,

$$\begin{aligned}(\vec{\boldsymbol{u_1}} - \vec{\boldsymbol{u'_1}}) = \vec{\mathbf{0}}, (\vec{\boldsymbol{u_2}} - \vec{\boldsymbol{u'_2}}) = \vec{\mathbf{0}}, \ldots, (\vec{\boldsymbol{u_m}} - \vec{\boldsymbol{u'_m}}) &= \vec{\mathbf{0}} \\ \implies \vec{\boldsymbol{u_1}} = \vec{\boldsymbol{u'_1}}, \vec{\boldsymbol{u_2}} = \vec{\boldsymbol{u'_2}}, \ldots, \vec{\boldsymbol{u_m}} &= \vec{\boldsymbol{u'_m}}\end{aligned}$$

which means that these are the same representation. And this clearly holds in reverse also as, if there is a single way of representing each resultant vector then there must be a single way of representing $\vec{\mathbf{0}}$ and due to the definition of a vector space we must always have the representation of all $\vec{\mathbf{0}}$. Therefore, this is the only representation of $\vec{\mathbf{0}}$.

Note that this is a condition on the contents of the subspaces and not on the way that the addition is performed. So, the difference between vector space sum $(U_1 + U_2)$ and vector space direct sum $(U_1 \oplus U_2)$ is not in the operator itself but in the operands they operate over.

For two subspaces, say, U_1, U_2 this condition on the subspaces reduces to the requirement that $U_1 \cap U_2 = \{\vec{\mathbf{0}}\}$ which can be seen as,

$$\begin{aligned}\vec{\boldsymbol{u_1}} + \vec{\boldsymbol{u_2}} &= \vec{\mathbf{0}} \\ \vec{\boldsymbol{u_1}} + -\vec{\boldsymbol{u}}_{\boldsymbol{1}} + \vec{\boldsymbol{u_2}} &= \vec{\mathbf{0}} + -\vec{\boldsymbol{u}}_{\boldsymbol{1}} \\ \vec{\boldsymbol{u_2}} &= -\vec{\boldsymbol{u}}_{\boldsymbol{1}} \\ \implies -\vec{\boldsymbol{u}}_{\boldsymbol{1}} \in U_2 \implies \vec{\boldsymbol{u_1}} &\in U_2\end{aligned}$$

So, for two subspaces, obtaining $\vec{\mathbf{0}}$ as the sum of vectors from the subspaces implies a vector in common between them. So, for $\vec{\mathbf{0}} + \vec{\mathbf{0}}$ to be the only way of obtaining $\vec{\mathbf{0}}$ implies that $\vec{\mathbf{0}}$ is the only vector in common.

However, for more than two subspaces, say U_1, U_2, U_3, the situation is different as we could have,

$$\begin{aligned} & \vec{u_1} + \vec{u_2} + \vec{u_3} = \vec{0} \\ \iff & \vec{u_1} + -\vec{u}_1 + \vec{u_2} + -\vec{u}_2 + \vec{u_3} = \vec{0} + -\vec{u}_1 + -\vec{u}_2 \\ & \iff \vec{u_3} = -\vec{u}_1 + -\vec{u}_2 \end{aligned}$$

which does not imply any vectors held in common.

2.2.13 Vector Space Problems

(tags: vector problems)

Prove that $-(-\vec{v}) = \vec{v}$ for every $\vec{v} \in V$ (tags: vector problems)

$$\begin{aligned} -(-\vec{v}) &= -[(-1)\vec{v}] && \text{using Proposition 40} \\ &= (-1)[(-1)\vec{v}] && \text{using Proposition 40 again} \\ &= [(-1)(-1)]\vec{v} && \text{using associativity of scalar multiplication} \\ &= \vec{v} && \text{using field properties} \end{aligned}$$

Or, a quicker way is,

$$\begin{aligned} -\vec{v} + -(-\vec{v}) &= \vec{0} && \text{using additive identity of } -\vec{v} \\ (-\vec{v} + \vec{v}) + -(-\vec{v}) &= \vec{0} + \vec{v} && \text{adding } \vec{v} \text{ to both sides} \\ -(-\vec{v}) &= \vec{v} \end{aligned}$$

Prove that if $a \in F, \vec{v} \in V$, and $a\vec{v} = \vec{0}$, then $a = 0$ or $\vec{v} = \vec{0}$. (tags: vector problems)

We follow a proof by cases.

Case $a \neq 0$:

$$\begin{aligned} a\vec{\boldsymbol{v}} = \vec{\mathbf{0}}, a \neq 0 &\implies a^{-1}a\vec{\boldsymbol{v}} = a^{-1}\vec{\mathbf{0}} && \text{using field properties} \\ &\iff 1\vec{\boldsymbol{v}} = b\vec{\mathbf{0}} && \text{where } b = a^{-1} \in F \\ &\iff \vec{\boldsymbol{v}} = \vec{\mathbf{0}} && \text{using Proposition 39 and multiplicative identity} \end{aligned}$$

Case $\vec{\boldsymbol{v}} \neq \vec{\mathbf{0}}$:

$$\begin{aligned} a\vec{\boldsymbol{v}} = \vec{\mathbf{0}}, \vec{\boldsymbol{v}} \neq \vec{\mathbf{0}} &\implies a\vec{\boldsymbol{v}} = a\vec{\boldsymbol{v}} + -a\vec{\boldsymbol{v}} \\ &\iff a\vec{\boldsymbol{v}} = (a + -a)\vec{\boldsymbol{v}} = 0\vec{\boldsymbol{v}} && \text{using field properties} \\ \text{Wrong! } a\vec{\boldsymbol{v}} = \vec{\mathbf{0}} &\implies a\vec{\boldsymbol{v}} = a\vec{\boldsymbol{v}} + -a\vec{\boldsymbol{v}} \\ & \quad \text{without need for } \vec{\boldsymbol{v}} \neq \vec{\mathbf{0}} \end{aligned}$$

This indicates that you are proving something that doesn't need proving. In actual fact,

Case $a = 0$: Actually, in this case there is nothing to be proven as we know from Proposition 38 that $0\vec{\boldsymbol{v}} = \vec{\mathbf{0}}$. So we have collectively exhaustive cases by looking at $a = 0$ and $a \neq 0$ and we only need to show that $a \neq 0 \implies \vec{\boldsymbol{v}} = \vec{\mathbf{0}}$ which we have already done.

Give an example of a nonempty subset U of $\mathbb{R}^2$ such that U is closed under scalar multiplication but U is not a subspace of $\mathbb{R}^2$. (tags: vector problems)

For all $\lambda \in \mathbb{R}$ the set $\{\, \lambda\vec{\boldsymbol{v}} \mid \vec{\boldsymbol{v}} \in \{(1,1)(-1,1)\}\,\}$ is closed under scalar multiplication but not addition.

Is $\mathbb{R}^2$ a subspace of the complex vector space $\mathbb{C}^2$? (tags: vector problems)

The definition of a subspace of $\mathbb{C}^2$ is a set of vectors which is a subset of those in $\mathbb{C}^2$ and that forms a vector space under the same addition and scalar multiplication of $\mathbb{C}^2$. The scalar multiplication of the vector space $\mathbb{C}^2$ is multiplication by scalars $\lambda \in \mathbb{C}$.
For a vector, $\vec{\boldsymbol{v}} \in \mathbb{R}^2$, scaling it by a complex number, $\lambda\vec{\boldsymbol{v}}$ will result in a vector that is not necessarily in $\mathbb{R}^2$.

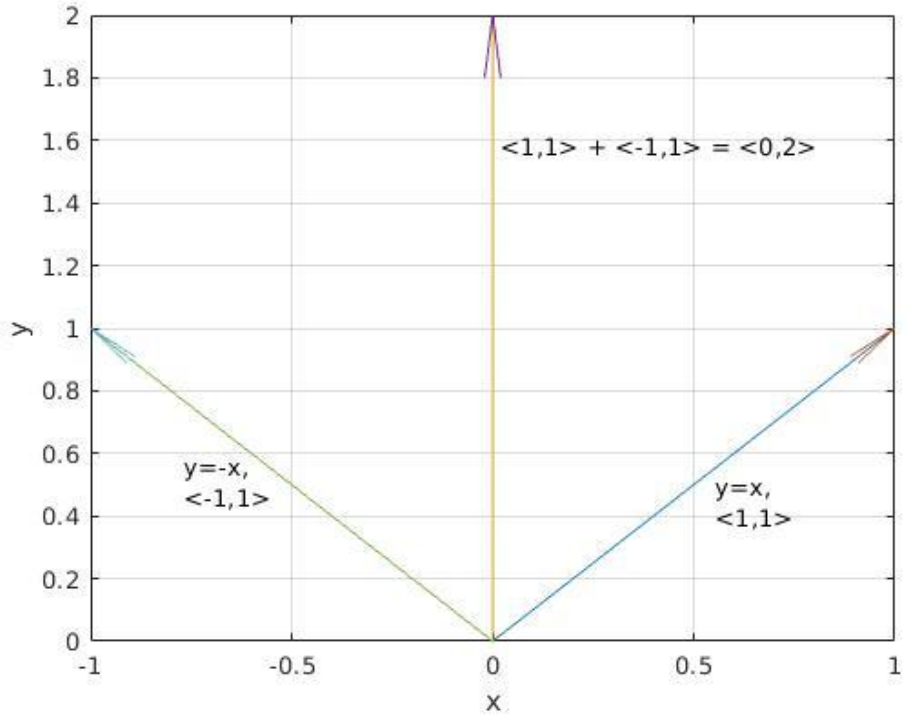


Figure 2.2: The blue arrows are vectors whose scalar multiples will all be in the same line as the blue arrows but the red arrow shows what happens if we add them; the result lies outside of both lines.

Is $\{\,(a,b,c) \in \mathbb{C}^3 \mid a^3 = b^3\,\}$ a subspace of $\mathbb{C}^3$? (tags: vector problems)

For $x \in \mathbb{C}$, x^3 has roots, $1, \frac{-1+\sqrt{3}i}{2}, \frac{-1-\sqrt{3}i}{2}$ so we don't have $a = b$ as we would if we were ranging over the reals.
Concretely, we can have, $(1, \frac{-1+\sqrt{3}i}{2}, 0)$ and $(1, \frac{-1-\sqrt{3}i}{2}, 0)$ but,

$$(1, \frac{-1+\sqrt{3}i}{2}, 0) + (1, \frac{-1-\sqrt{3}i}{2}, 0) = (2, -1, 0)$$

where $(2, -1, 0) \notin \{\,(a,b,c) \in \mathbb{C}^3 \mid a^3 = b^3\,\}$ meaning that addition over this set is not closed. Therefore, this is not a subspace.

Give an example of a non-empty subset U of $\mathbb{R}^2$ such that U is closed under addition and under taking additive inverses (meaning $-\vec{\boldsymbol{u}} \in U$ whenever $\vec{\boldsymbol{u}} \in U$), but U is not a subspace of $\mathbb{R}^2$. (tags: vector problems)

First thought might be $\mathbb{R}^2 - \{\vec{\mathbf{0}}\}$ but this is Wrong!. If the subset is closed under addition and under taking additive inverses then it means that $\vec{\boldsymbol{u}} + -\vec{\boldsymbol{u}} = \vec{\mathbf{0}} \in U$ and so the set $\mathbb{R}^2 - \{\vec{\mathbf{0}}\}$ is not closed under addition and taking additive inverses.
The set $\{\,(x, y) \in \mathbb{R}^2 \mid x, y \in \mathbb{Z}\,\}$ however, is closed under addition because integer addition is closed and under taking additive inverses but scalar multiplication where the scalars range over the reals, will produce non-integer values for x and y.

Is the set of periodic functions over the reals a subspace of $\mathbb{R}^R$? (tags: vector problems, periodic functions)

The definition of two periodic functions over the reals is

$$\exists p > 0 \in \mathbb{R} \cdot f(x) = f(x + p)$$
$$\exists q > 0 \in \mathbb{R} \cdot g(x) = g(x + q)$$

Then for their sum to be periodic we need,

$$\exists \alpha, \beta \in \mathbb{Z}, m \in \mathbb{R} \cdot (m = \alpha p) \wedge (m = \beta q)$$
$$\iff \frac{q}{p} = \frac{\alpha}{\beta} \in \mathbb{Q}$$
$$\therefore (f + g)(x) = (f + g)(x + m) = f(x + m) + g(x + m)$$
$$\iff \frac{q}{p} \in \mathbb{Q}.$$

Prove that the union of two subspaces of V is a subspace of V if and only if one of the subspaces is contained within the other. (tags: vector problems)

Let A, B be subspaces of V and $\vec{\boldsymbol{a}} \in A$, $\vec{\boldsymbol{b}} \in B$ and,

$$C = A \cup B = \{\,\vec{\boldsymbol{c}} \mid \vec{\boldsymbol{c}} \in A \ \vee \ \vec{\boldsymbol{c}} \in B\,\}.$$

Since $\vec{a}, \vec{b} \in C$ we have (C subspace of V) $\iff \forall \alpha, \beta \in F \cdot (\alpha\vec{a} + \beta\vec{b}) \in C$. Then,

$$\begin{aligned} \vec{b} \in A &\implies \forall \alpha, \beta \in F \cdot (\alpha\vec{a} + \beta\vec{b}) \in A \text{ (by subspace properties)} \\ &\implies (\alpha\vec{a} + \beta\vec{b}) \in C. \end{aligned}$$

A similar argument holds for $\vec{a} \in B$. Conversely,

$$\begin{aligned} \forall \alpha, \beta \in F \cdot (\alpha\vec{a} + \beta\vec{b}) \in C &\implies ((\alpha\vec{a} + \beta\vec{b}) \in A) \vee ((\alpha\vec{a} + \beta\vec{b}) \in B) \\ &\implies ((\alpha\vec{a} - \alpha\vec{a} + \beta\vec{b}) = \beta\vec{b} \in A) \vee ((\alpha\vec{a} + \beta\vec{b} - \beta\vec{b}) = \alpha\vec{a} \in B) \\ &\implies (\vec{b} \in A) \vee (\vec{a} \in B) \end{aligned}$$

$$\begin{aligned} \therefore \text{(C subspace of V)} &\iff \forall \alpha, \beta \in F \cdot (\alpha\vec{a} + \beta\vec{b}) \in C \\ &\iff (\vec{b} \in A) \vee (\vec{a} \in B) \\ &\equiv (B \subseteq A) \vee (A \subseteq B). \end{aligned}$$

Can a vector space over an infinite field be a finite union of proper subspaces? (tags: vector problems)

Assume that our vector space V is a finite union of proper subspaces, hence

$$V = \bigcup_{i=1}^{n} U_i.$$

Now, pick a non-zero vector $\vec{x} \in U_1$, and pick another vector $\vec{y} \in V \setminus U_1$.

There are infinitely many vectors $\vec{x} + k\vec{y}$, where $k \in K^*$ (K is our infinite field). Note that $\vec{x} + k\vec{y}$ is not in U_1, hence must be contained in some U_j where $j \neq 1$.

Then since $k \in K^*$, we can have $\vec{x} + k_1\vec{y}$, $\vec{x} + k_2\vec{y} \in U_j$, which implies that it also contains $\vec{y}$ and hence also $\vec{x}$, hence $U_1 \subset U_j$.

Explanation: There are infinitely many vectors $\vec{\boldsymbol{x}}+k\vec{\boldsymbol{y}}$ and only finitely many U_i so they cannot all be in different U_i so we have,

$$\begin{aligned} &\exists\, k_1,\, k_2 \in K^* \,\cdot\, \vec{\boldsymbol{x}} + k_1\vec{\boldsymbol{y}},\, \vec{\boldsymbol{x}} + k_2\vec{\boldsymbol{y}} \in U_j \\ \implies &(\vec{\boldsymbol{x}} + k_1\vec{\boldsymbol{y}}) - (\vec{\boldsymbol{x}} + k_2\vec{\boldsymbol{y}}) = (k_1 - k_2)\vec{\boldsymbol{y}} \in U_j \\ &\implies \vec{\boldsymbol{y}} \in U_j \implies \vec{\boldsymbol{x}} \in U_j \end{aligned}$$

Hence

$$V = \bigcup_{i=2}^{n} U_i.$$

Evidently, this can be continued, hence a contradiction arises.

Prove or give a counterexample: if U_1, U_2, W are subspaces of V such that $V = U_1 \oplus W$ and $V = U_2 \oplus W$ then $U_1 = U_2$. (tags: vector problems)

Counter example: $V = \mathbb{F}^2,\ U_1 = \{\,(x, 0) \in \mathbb{F}^2 \mid x \in F\,\},\ U_2 = \{\,(0, x) \in \mathbb{F}^2 \mid x \in F\,\},\ W = \{\,(x, x) \in \mathbb{F}^2 \mid x \in F\,\}$.

Let U_e denote the set of real-valued even functions on $\mathbb{R}$ and let U_o denote the set of real-valued odd functions on $\mathbb{R}$. Show that $\mathbb{R}^R = U_e \oplus U_o$. (tags: vector problems)

Every function $f \in \mathbb{R}^R$ can be expressed as the sum of an even function and an odd function as,

$$f(x) = \frac{f(x) + f(-x)}{2} + \frac{f(x) - f(-x)}{2} = g(x) + h(x)$$

where $g(x) \in U_e$ and $h(x) \in U_o$. So, $U_e + U_o$ spans $\mathbb{R}^R$.
Furthermore,

$$\begin{aligned} f(x) \in (U_e \cap U_o) &\implies (f(-x) = f(x)) \wedge (f(-x) = -f(x)) \\ &\implies f(x) = -f(x) \\ &\implies f(x) = 0 \end{aligned}$$

Since $f(x) = 0$ is the additive identity of this space, this shows that the intersection is $\vec{\mathbf{0}}$. So, $\mathbb{R}^R = U_e \oplus U_o$.

2.2.14 Span, Dimension and Bases

(tags: vector spaces)

Definition. *The* span *of a list of vectors* $\vec{v_1}, \vec{v_2}, \ldots, \vec{v_k}$ *- written* $\text{span}(\vec{v_1}, \vec{v_2}, \ldots, \vec{v_k})$ *- is defined as*

$$\{ \alpha_1 \vec{v_1} + \alpha_2 \vec{v_2} + \cdots + \alpha_k \vec{v_k} \mid \alpha_1, \alpha_2, \ldots, \alpha_k \in F \}$$

Proposition 42. *The span of a list of vectors is the smallest subspace containing those vectors.*

Note that a vector space over $\mathbb{R}$ or $\mathbb{C}$ is an uncountable set as - while the dimensions of the vector space may be finite - closure under scalar multiplication means that the vectors in the space are continuously valued as the field providing the scalars is continuously valued.
This means that the notion of the *smallest* subspace cannot refer to the cardinality of the set and must refer to ordering based on subset. So, the smallest subspace containing a list of vectors is a subspace that contains the list of vectors and, of which, there is no proper subset which also contains the list of vectors.

Proof.

$$\begin{aligned} \text{Let } S &:= span(\vec{v_1}, \vec{v_2}, \ldots, \vec{v_k}) \\ &:= \{ \alpha_1 \vec{v_1} + \alpha_2 \vec{v_2} + \cdots + \alpha_k \vec{v_k} \mid \alpha_1, \alpha_2, \ldots, \alpha_k \in F \} \\ \text{and let } V &:= \text{ the smallest vector space containing } \vec{v_1}, \vec{v_2}, \ldots, \vec{v_k}. \end{aligned}$$

then S contains every linear combination of $\vec{v_1}, \vec{v_2}, \ldots, \vec{v_k}$ and nothing else and so is a vector space containing $\vec{v_1}, \vec{v_2}, \ldots, \vec{v_k}$,

$$V \subseteq S$$

Additionally, any vector space containing the vectors $\vec{v_1}, \vec{v_2}, \ldots, \vec{v_k}$ must contain all their linear combinations, $span(\vec{v_1}, \vec{v_2}, \ldots, \vec{v_k})$,

$$S \subseteq V$$

Therefore there is no proper subset of $span(\vec{\boldsymbol{v_1}}, \vec{\boldsymbol{v_2}}, \ldots, \vec{\boldsymbol{v_k}})$ that is also a vector space containing $\vec{\boldsymbol{v_1}}, \vec{\boldsymbol{v_2}}, \ldots, \vec{\boldsymbol{v_k}}$, and so $span(\vec{\boldsymbol{v_1}}, \vec{\boldsymbol{v_2}}, \ldots, \vec{\boldsymbol{v_k}})$ is the smallest vector space containing $\vec{\boldsymbol{v_1}}, \vec{\boldsymbol{v_2}}, \ldots, \vec{\boldsymbol{v_k}}$,

$$(V \subseteq S) \wedge (S \subseteq V) \iff V = S \qquad \square$$

Proposition 43. *Length of every linearly independent list in a space is less than or equal to the length of a spanning list in the same space.*

Proof. Let $U = \vec{\boldsymbol{u_1}}, \vec{\boldsymbol{u_2}}, \ldots, \vec{\boldsymbol{u_m}}$ be a linearly independent list of vectors in V and $W = \vec{\boldsymbol{w_1}}, \vec{\boldsymbol{w_2}}, \ldots, \vec{\boldsymbol{w_n}}$ be a spanning list of vectors in V.
If we take $\vec{\boldsymbol{u_1}}$ from U and add it to W then - since the other vectors in W are a spanning list - W must be linearly dependent. That's to say,

$$\begin{aligned}
\exists\, \alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{R} \cdot \alpha_1 \vec{\boldsymbol{w_1}} + \cdots + \alpha_n \vec{\boldsymbol{w_n}} &= \vec{\boldsymbol{u_1}} \\
&\iff \alpha_1 \vec{\boldsymbol{w_1}} + \cdots + \alpha_n \vec{\boldsymbol{w_n}} - \vec{\boldsymbol{u_1}} = -\alpha_i \vec{\boldsymbol{w_i}} \\
&\iff \frac{-\alpha_1}{\alpha_i} \vec{\boldsymbol{w_1}} + \cdots + \frac{-\alpha_n}{\alpha_i} \vec{\boldsymbol{w_n}} + \frac{1}{\alpha_i} \vec{\boldsymbol{u_1}} = \vec{\boldsymbol{w_i}}
\end{aligned}$$

So, $\vec{\boldsymbol{w_i}}$ is in the span of $\vec{\boldsymbol{u_1}}, \vec{\boldsymbol{w_2}}, \ldots, \vec{\boldsymbol{w_n}}$ and we can drop $\vec{\boldsymbol{w_i}}$ from the list, W, and it will still span the vector space.
We can keep doing this with the remaining vectors in U - each time the vector to be removed will be some $\vec{\boldsymbol{w_i}}$ because all the $\vec{\boldsymbol{u_i}}$ are linearly independent - and all the while W remains a spanning list. We continue until we have replaced (potentially) all n vectors in W, which would happen if $m > n$. At this point we would have the spanning list $W = \vec{\boldsymbol{u_1}}, \vec{\boldsymbol{u_2}}, \ldots, \vec{\boldsymbol{u_n}}$ and $(m - n)$ remaining vectors in U.
Now, since W spans the space, the $(m - n)$ vectors that remain in U will be in the span of W. But, all the vectors that originally came from U were linearly independent, so it is impossible for any vectors in U to be in the span of W (which now comprises only vectors that originally came from U). We therefore conclude that there can be no remaining vectors in U and, consequently that m cannot be greater than n, i.e. $m \leq n$. $\square$

2.3 Linear Algebra

(tags: linear algebra)

2.3.1 Basic properties of Matrix Algebra

(tags: linear algebra)

Definition. *Matrix **equality** is defined component-wise so that if $A = B$ then A and B must have the same dimension as well as equal values in each component.*

Definition. *An **identity** element e is defined as $ea = ae = a$.*

The definition of an identity element above is in any context (not just for matrices). For matrices this has certain consequences.

Proposition 44. *Identity matrices must be square*

Proof. For a matrix A and an identity matrix I, $AI = IA = A$ which means that AI, IA and A must all have the same dimensions. If A is of dimension $m \times n$ then I must have dimension $n \times m$ but then AI has dimension $m \times m$ while IA has dimension $n \times n$. We conclude that $m = n$ and both matrices are square. □

If A, B, C are matrices s.t. $AB = AC$, can we, in general, conclude that $B = C$?

The answer is no, as the following example shows:

$$A = \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & -1 \\ 3 & 5 \end{pmatrix}, \qquad C = \begin{pmatrix} 8 & 0 \\ -4 & 4 \end{pmatrix}$$

$$A = B = \begin{pmatrix} 0 & 0 \\ 4 & 4 \end{pmatrix}$$

This is because multiplication by A has no inverse (i.e. it's not a bijection and A^{-1} does not exist) as we can see by the fact that $|A| = 0$.

If A, B, C are matrices s.t. $A + 5B = A + 5C$, can we, in general, conclude that $B = C$?

The answer is yes because the matrix addition and scalar multiplication always have inverses. The inverse of $+A$ is $-A$ and the inverse of scalar multiplication by 5 is scalar multiplication by $\frac{1}{5}$. So we can say,

$$\begin{aligned} & A + 5B = A + 5C \\ \iff \quad & A + 5B - A = A + 5C - A \\ \iff \quad & 5B = 5C \\ \iff \quad & \left(\frac{1}{5}\right) 5B = \left(\frac{1}{5}\right) 5C \\ \iff \quad & B = C \end{aligned}$$

Matrix multiplication

Multiplication of matrices proceeds as a collection of dot-products of individual vectors. As a result, its properties are largely dependent on the properties of the dot-product. These are:

> If $\vec{\boldsymbol{x}} = (a, b)^T$ and $\vec{\boldsymbol{y}} = (e, g)^T$ then the dot-product $\langle \vec{\boldsymbol{x}}, \vec{\boldsymbol{y}} \rangle = ae + bg$ and,

- $\langle \vec{\boldsymbol{x}}, \vec{\boldsymbol{y}} \rangle = \langle \vec{\boldsymbol{y}}, \vec{\boldsymbol{x}} \rangle$
- $\alpha \langle \vec{\boldsymbol{x}}, \vec{\boldsymbol{y}} \rangle = \langle \alpha \vec{\boldsymbol{x}}, \vec{\boldsymbol{y}} \rangle = \langle \vec{\boldsymbol{x}}, \alpha \vec{\boldsymbol{y}} \rangle$
- $\langle \vec{\boldsymbol{x}} + \vec{\boldsymbol{y}}, \vec{\boldsymbol{z}} \rangle = \langle \vec{\boldsymbol{x}}, \vec{\boldsymbol{z}} \rangle + \langle \vec{\boldsymbol{y}}, \vec{\boldsymbol{z}} \rangle$
- $\langle \vec{\boldsymbol{x}}, \vec{\boldsymbol{x}} \rangle \geq 0$ and $\langle \vec{\boldsymbol{x}}, \vec{\boldsymbol{x}} \rangle = 0 \iff \vec{\boldsymbol{x}} = 0$

Matrix multiplication treats the two operand matrices as collections of vectors with the first matrix having the vectors as rows and the second having the vectors as columns.

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} e & f \\ g & h \end{bmatrix} = \begin{bmatrix} ae + bg & af + bh \\ ce + dg & cf + dh \end{bmatrix}$$

This difference in orientation of the vectors in the two operands results in the multiplication not being commutative - the order matters. So, the first property of the dot-product is not preserved but the others are preserved (albeit with a slight modification for the last one).

$$\alpha \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} e & f \\ g & h \end{bmatrix} = \begin{bmatrix} \alpha a & \alpha b \\ \alpha c & \alpha d \end{bmatrix} \begin{bmatrix} e & f \\ g & h \end{bmatrix} = \begin{bmatrix} \alpha(ae+bg) & \alpha(af+bh) \\ \alpha(ce+dg) & \alpha(cf+dh) \end{bmatrix} = \alpha \begin{bmatrix} ae+bg & af+bh \\ ce+dg & cf+dh \end{bmatrix}$$

$$\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix} + \begin{bmatrix} e & f \\ g & h \end{bmatrix} \right) \begin{bmatrix} i & j \\ k & l \end{bmatrix} = \begin{bmatrix} a+e & b+f \\ c+g & d+h \end{bmatrix} \begin{bmatrix} i & j \\ k & l \end{bmatrix} = \begin{bmatrix} i\,(a+e)+k\,(b+f) & j\,(a+e)+l\,(b+f) \\ i\,(c+g)+k\,(d+h) & j\,(c+g)+l\,(d+h) \end{bmatrix}$$
$$= \begin{bmatrix} ia+kb & ja+lb \\ ic+kd & jc+ld \end{bmatrix} + \begin{bmatrix} ie+kf & je+lf \\ ig+kh & jg+lh \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} i & j \\ k & l \end{bmatrix} + \begin{bmatrix} e & f \\ g & h \end{bmatrix} \begin{bmatrix} i & j \\ k & l \end{bmatrix}$$

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix}^T = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} a & c \\ b & d \end{bmatrix} = \begin{bmatrix} a^2+b^2 & ac+bd \\ ac+bd & c^2+d^2 \end{bmatrix}$$

So, to summarize:

If A, B, C are matrices and α is a scalar then,

- $\alpha AB = (\alpha A)B = A(\alpha B) = \alpha(AB)$
- $(A+B)C = C(A+B) = AC + BC$
- AA^T is a symmetric matrix with positive values along the diagonal

Matrix transpose

Denote the ith row of the matrix A as $A[i:]$ and the jth column of the matrix B as $B[:j]$ and a matrix whose components at (i,j) are the dot-products of the ith row of the matrix A with the jth column of the matrix B as $(\langle A[i:], B[:j] \rangle)$. Then,

$$(AB)^T = (\langle A[i:], B[:j] \rangle)^T = (\langle A[j:], B[:i] \rangle)$$
$$B^T A^T = (\langle B^T[i:], A^T[:j] \rangle) = (\langle B[:i], A[j:] \rangle)$$

So, $(AB)^T = B^T A^T$. A consequence of this is that,

$$\begin{aligned} & I = AA^{-1} = (AA^{-1})^T = (A^{-1})^T A^T \\ \iff \quad & I(A^T)^{-1} = (A^{-1})^T A^T (A^T)^{-1} \\ \iff \quad & (A^T)^{-1} = (A^{-1})^T \end{aligned}$$

Matrix inverse

Definition. *Inverse property is if $\exists$ a matrix B s.t. $AB = BA = I$ then B is the **inverse** of A.*

This definition is inherently bound up with the definition of the identity ($\exists$ a matrix I s.t. $AI = IA = A$) and both define the identity and inverse elements as commutatively producing their result under matrix multiplication. Since matrix multiplication is not, in general, commutative there is no guarantee that if $AB = I$ then $BA = I$. An example of this failing is,

$$A = \begin{bmatrix} 1 & 2 \end{bmatrix}, B = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$AB = \begin{bmatrix} 1 \end{bmatrix} = I_1, BA = \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix} \neq I_2$$

But we could have guessed this because Proposition 44 showed that identity matrices must be square and its product with a matrix must be defined from both the left and the right, i.e. $IA = AI = A$ meaning that the matrix A must have the same dimensions as I. So, for non-square matrices, no identity can exist. If there is no identity, then the inverse is not defined either.

Proposition 45. *If the inverses of the matrices A and B both exist then so does the inverse of the product AB and it is equal to $B^{-1}A^{-1}$.*

Proof.

$$\begin{aligned} & (AB)(AB)^{-1} = I \\ \iff \quad & (A^{-1}A)B(AB)^{-1} = A^{-1}I \\ \iff \quad & (B^{-1}B)(AB)^{-1} = B^{-1}A^{-1} \\ \iff \quad & (AB)^{-1} = B^{-1}A^{-1} \end{aligned}$$

and since B^{-1} and A^{-1} both exist then their product exists. Furthermore, this holds for a product of any finite sequence of invertible matrices $A_1 A_2 \cdots A_n$ which can easily be shown by induction on the associative product. $\square$

2.3.2 Matrices as linear transformations

(tags: linear algebra)

Multiplying a vector by a matrix on the left: $A\vec{\boldsymbol{x}} = \vec{\boldsymbol{y}}$

Left multiplication of a matrix A of dimension $m \times n$ on a column vector $\vec{\boldsymbol{x}}$ of dimension $n \times 1$ transforms it to a column vector $\vec{\boldsymbol{y}}$ of dimension $m \times 1$.

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$$

This can be thought of a function from the space of n-dimensional vectors from which $\vec{\boldsymbol{x}}$ is drawn to the space of m-dimensional vectors in which $\vec{\boldsymbol{y}}$ resides. So, for real-valued vectors, the function would be a function $f : \mathbb{R}^n \mapsto \mathbb{R}^m$ such that,

$$f(x_1, \cdots, x_n) = \vec{\boldsymbol{y}} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$$

Or else, this could be thought of as m n-ary functions of the form $f : \mathbb{R}^n \mapsto \mathbb{R}$,

$$\begin{aligned} f_1(x_1, \cdots, x_n) &= a_{11}x_1 + \cdots + a_{1n}x_n = y_1 \\ &\vdots \\ f_m(x_1, \cdots, x_n) &= a_{m1}x_1 + \cdots + a_{mn}x_n = y_m \end{aligned}$$

In this case, each row of the matrix is a real-valued function in n variables. Each of these functions is *homogenous linear* (a function of the form $a_1 x_1 + \cdots + a_k x_k + c$ for scalars $a_1, \cdots, a_k, c$ and $c = 0$) and so the system of functions is called a *linear transformation*.

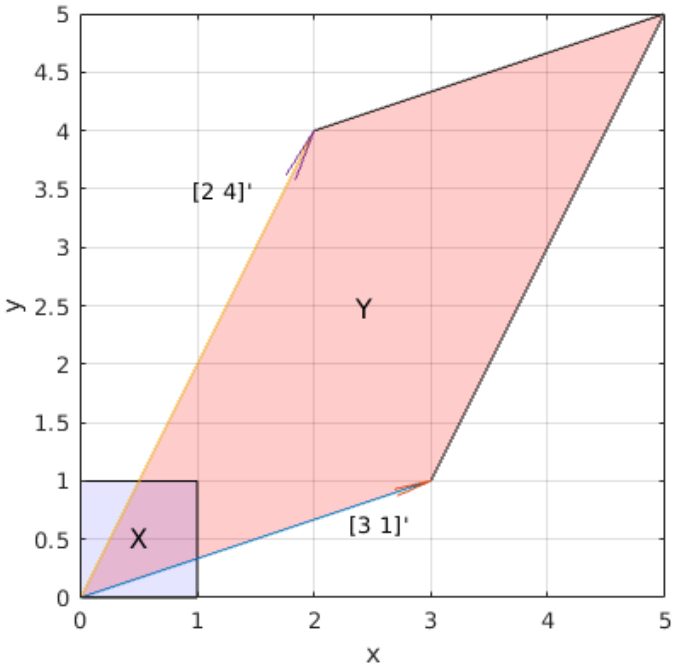
Multiplying a matrix of vectors by a matrix on the left: $AX = Y$

Looking at the matrix as a linear transformation from one co-ordinate space to another, consider $AX = Y$ where X is a matrix - which may be considered a collection of vectors - transformed by the matrix A into the matrix - or collection of vectors - Y.

We transform the unit square in the source space, X in $\mathbb{R}^2$, using the 2D transformation matrix A, into its image in the destination space, Y in $\mathbb{R}^2$.

$$A = \begin{bmatrix} 3 & 2 \\ 1 & 4 \end{bmatrix}, \; X = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

$$AX = Y = \begin{bmatrix} 0 & 3 & 5 & 2 \\ 0 & 1 & 5 & 4 \end{bmatrix}$$



Types of Transformations

There are 3 basic types of transformation:

- **Rigid body** - preserves distances and angles.

  Examples: translation and rotation.

- **Conformal** - preserves angles.

  Examples: translation, rotation and uniform scaling.

- **Affine** - preserves parallelism.

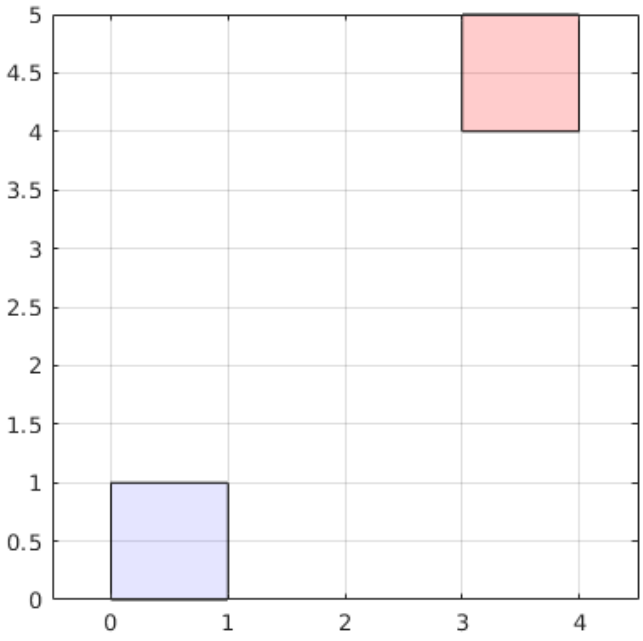
  Examples: translation, rotation, uniform and non-uniform scaling, shearing and reflection.

Rigid Body

Translation So as to perform the translation as multiplication by a transformation matrix we take the approach of homogeneous coordinates (see:https://en.wikipedia.org/wiki/Homogeneous_coordinates) so we form matrix with the identity in the first two columns and then a third column with the translation vector. Then, we add a row of ones to the vectors we will translate and the output vectors also have a 1 in the third row that is ignored.

$$A = \begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \end{bmatrix}, \; X = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

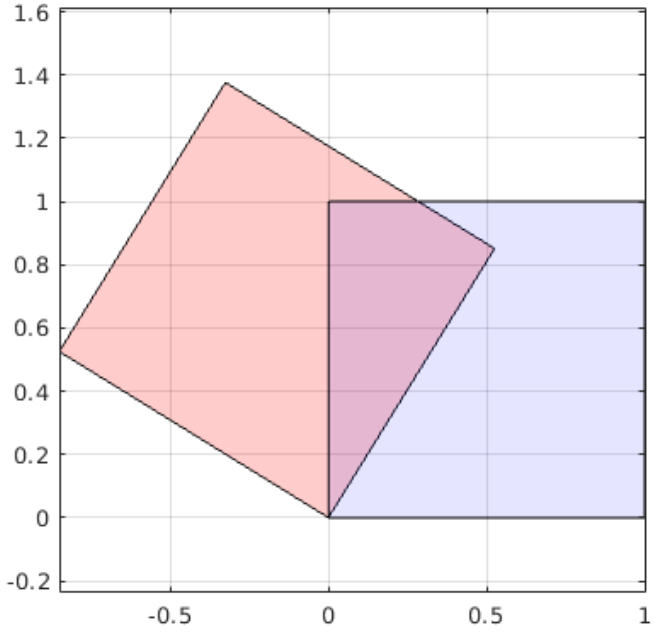
$$AX = Y = \begin{bmatrix} 3 & 4 & 4 & 3 \\ 4 & 4 & 5 & 5 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$



Rotation

$$A = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix},\ X = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

$$AX = Y = \begin{bmatrix} 0 & 0.5253 & -0.3256 & -0.8509 \\ 0 & 0.8509 & 1.3762 & 0.5253 \end{bmatrix}$$

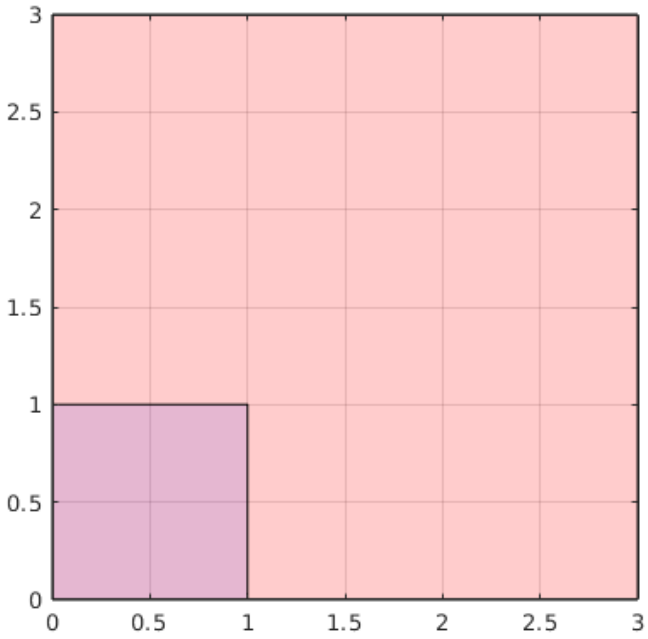


Conformal

Uniform Scaling is scaling by an equal amount in each dimension.

$$A = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}, \; X = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

$$AX = Y = \begin{bmatrix} 0 & 3 & 3 & 0 \\ 0 & 0 & 3 & 3 \end{bmatrix}$$

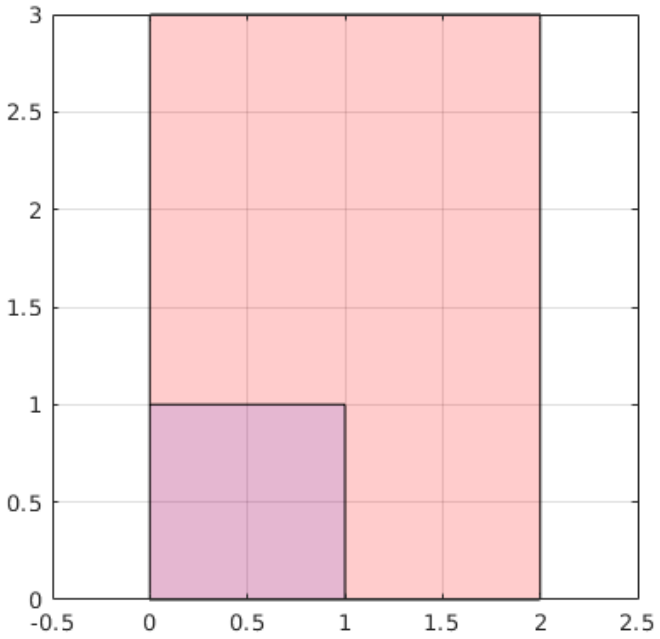


Affine

Non-uniform Scaling is scaling by different amounts in the different dimensions. (The example shown here preserves the angles but for other shapes, a triangle for example, the angles would not be preserved.)

$$A = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}, \; X = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

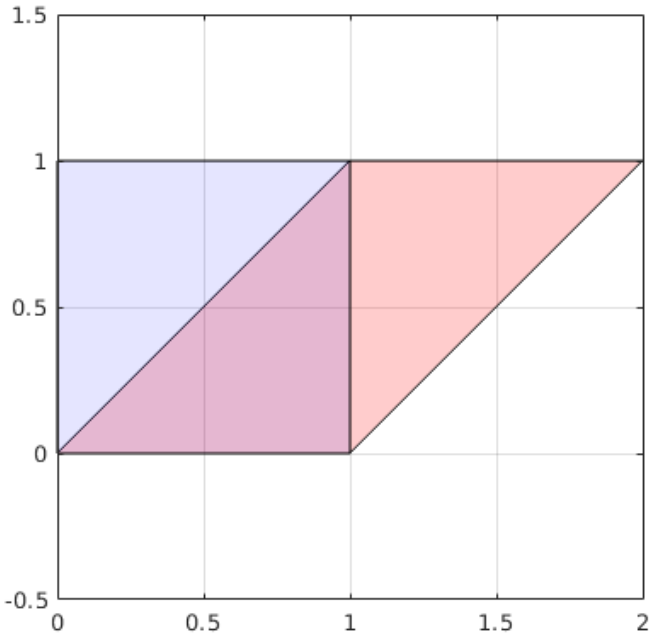
$$AX = Y = \begin{bmatrix} 0 & 2 & 2 & 0 \\ 0 & 0 & 3 & 3 \end{bmatrix}$$



Shearing

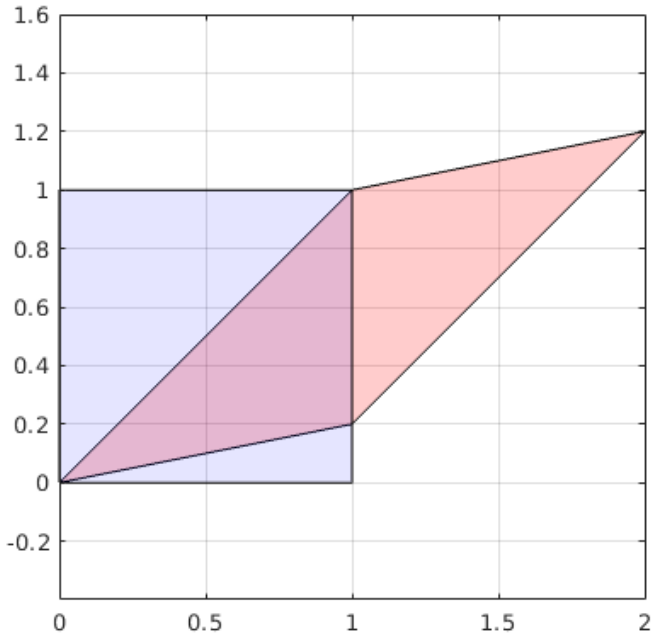
$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix},\ X = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

$$AX = Y = \begin{bmatrix} 0 & 1 & 2 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$



$$A = \begin{bmatrix} 1 & 1 \\ 0.2 & 1 \end{bmatrix},\ X = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

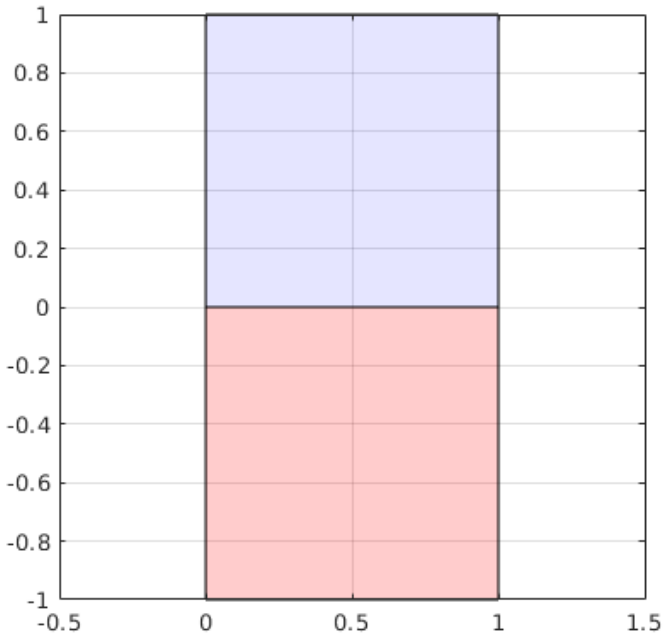
$$AX = Y = \begin{bmatrix} 0 & 1 & 2 & 1 \\ 0 & 0.2 & 1.2 & 1 \end{bmatrix}$$



Reflection

$$A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix},\ X = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

$$AX = Y = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & -1 & -1 \end{bmatrix}$$


1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
-0.5
0
0.5
1
1.5

2.3.3 Elementary Matrices and Row Operations

(tags: linear algebra)

Notation. The **matrix units** - matrices with a single non-zero component whose value is 1 are traditionally named $\boldsymbol{e_{ij}}$ where i, j is the matrix coordinate of the 1.

An arbitrary matrix $A = (a_{ij})$ may be expressed as a sum of such unit matrices as $A = a_{11}e_{11} + \cdots + a_{nn}e_{nn}$.

$$e_{ij} = \begin{bmatrix} \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots \\ \cdots & 1 & \cdots \\ \vdots & \vdots & \vdots \\ \cdots & \cdots & \cdots \end{bmatrix}$$

So matrix units can be used to analyse matrix addition but to analyse matrix multiplication some square matrices called **elementary matrices** are more useful.

Multiplying a matrix from the left (so doing row operations), there are 3 types of elementary matrix:

Adding rows: $\boldsymbol{I + ae_{ij}}$ for $\boldsymbol{i \neq j}$

$$\begin{bmatrix} 1 & & & \\ & \cdot & a & \\ & & \cdot & \\ & & & 1 \end{bmatrix}$$

This adds a times some row to another row.

Swapping rows: $\boldsymbol{I + e_{ij} + e_{ji} - e_{ii} - e_{jj}}$ for $\boldsymbol{i \neq j}$

$$\begin{bmatrix} 1 & & & \\ & 0 & 1 & \\ & 1 & 0 & \\ & & & 1 \end{bmatrix}$$

This swaps the rows i and j.

Scalar-multiplying a row: $I + (c-1)e_{ii}$ for $c \neq 0$

$$\begin{bmatrix} 1 & & & \\ & \cdot & & \\ & & c & \\ & & & 1 \end{bmatrix}$$

This multiplies row i by c.

Proposition 46. *Elementary matrices are invertible and their inverses are also elementary matrices.*

Proof. Proceed by cases on the 3 elementary types of elementary matrices.

Case $I + ae_{ij}$ If R_i is row i and R_j is row j, then this matrix performs $R_i + aR_j$. Clearly this can be "undone" by performing $R_i - aR_j$. So the matrix, $I - ae_{ij}$ is the inverse and clearly this is also an elementary matrix of the same type.

Case $I - e_{ii} - e_{jj} + e_{ij} + e_{ji}$ This matrix swaps 2 rows in a permutation that is its own inverse.

Case $I + (c-1)e_{ii}$ This matrix performs cR_i and so it is "undone" by performing $c^{-1}R_i$ (which for a real-valued matrix would be $\left(\frac{1}{c}\right)R_i$) and this inverse matrix is also an elementary matrix of the same type. □

Proposition 47. *Suppose $AX = B$ and a series of elementary row operations on $[A \mid B]$ produces $[A' \mid B']$, then the solutions of $A'X = B'$ are the same as those of $AX = B$.*

Proof. First note that the series of elementary row operations is described as multiplication on the left by a series of elementary matrices say, $E_1, E_2, \cdots, E_n$ so that,

$$[A' \mid B'] = [(E_n \cdots E_2 E_1)A \mid (E_n \cdots E_2 E_1)B]$$

Now, let $(E_n \cdots E_2 E_1) = E$ and notice that, since each of the individual E_i is invertible the product of them is also invertible by Proposition 45 so,

$$A'X = B' \iff EAX = EB$$

and the existence of the inverse E^{-1} means that the law of cancellation is in effect so,

$$EAX = EB \iff AX = B$$
$$\therefore A'X = B' \iff AX = B$$

□

Proposition 48. *Let A be a square matrix. The following conditions are equivalent:*

- *A can be reduced to the identity by a sequence of elementary row operations.*
- *A is a product of elementary matrices.*
- *A is invertible.*
- *The system of homogeneous equations $AX = 0$ has only the trivial solution $X = 0$.*

Proof. If A can be reduced to the identity by a sequence of elementary row operations then,

$$(E_n \cdots E_2 E_1)A = I$$

and by Proposition 46 and Proposition 45 the matrix $(E_n \cdots E_2 E_1)$ is invertible so,

$$A = (E_n \cdots E_2 E_1)^{-1} I = (E_n \cdots E_2 E_1)^{-1} = {E_1}^{-1} {E_2}^{-1} \cdots {E_n}^{-1}$$

and, also by Proposition 46, A is a product of elementary matrices and is invertible.
Furthermore, if $AX = 0$ then $X = A^{-1}0 = 0$ - i.e. the only solution to $AX = 0$ is $X = 0$. □

2.3.4 Determinants

(tags: linear algebra)

1×1

The determinant of a 1×1 matrix is just its unique component entry,

$$det \begin{bmatrix} a \end{bmatrix} = a$$

2×2

The determinant of a 2×2 matrix is given by the formula,

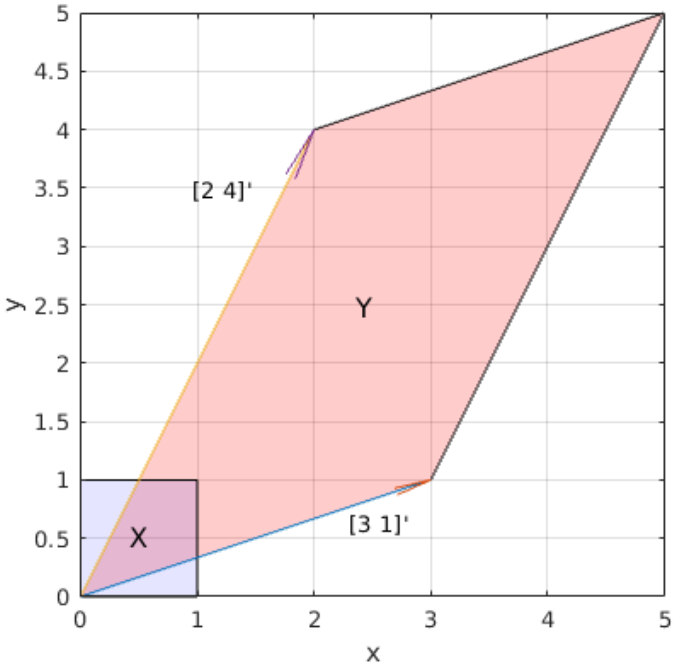
$$det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc$$

Returning to our example of a 2d operator:

We transform the unit square in the source space, X in $\mathbb{R}^2$, using the 2D transformation matrix A, into its image in the destination space, Y in $\mathbb{R}^2$.

$$A = \begin{bmatrix} 3 & 2 \\ 1 & 4 \end{bmatrix}, \; X = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

$$AX = Y = \begin{bmatrix} 0 & 3 & 5 & 2 \\ 0 & 1 & 5 & 4 \end{bmatrix}$$



We see that $det\ A = 10$ and the parallelogram, Y, that is the image of the unit square, X, under the transformation represented by A has area,

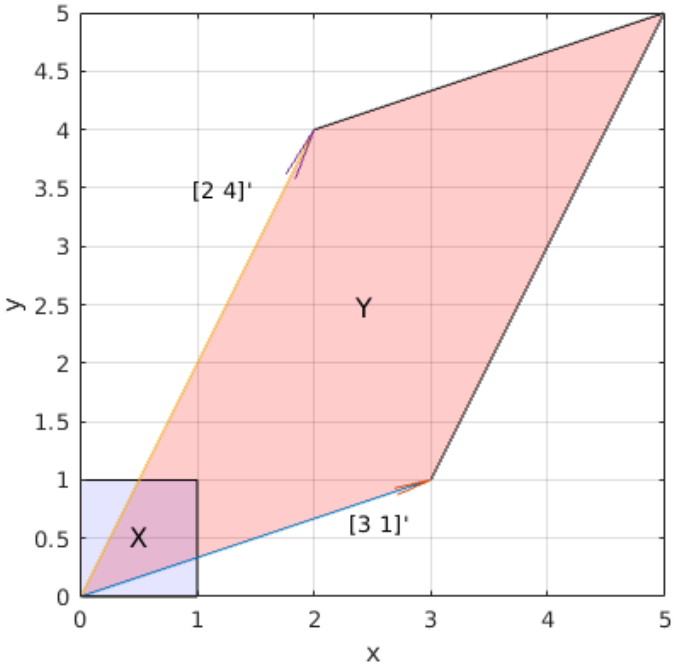
$$\text{area } = b \cdot h = |\langle 3, 1\rangle| \cdot |\langle 2-3, 4-1\rangle| = \sqrt{10} \cdot \sqrt{10} = 10$$

And the determinant would be 0 in the case that the columns were proportional (representing co-linear vectors) and the determinant would be negative if the orientation of the output vectors were reversed w.r.t. the input vectors. So, if we swap either the columns or the rows of the transformation matrix, A, the determinant comes out -10.

Swapping the columns:

$$A_c = \begin{bmatrix} 2 & 3 \\ 4 & 1 \end{bmatrix},\ X = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

$$AX = Y = \begin{bmatrix} 0 & 2 & 5 & 3 \\ 0 & 4 & 5 & 1 \end{bmatrix}$$

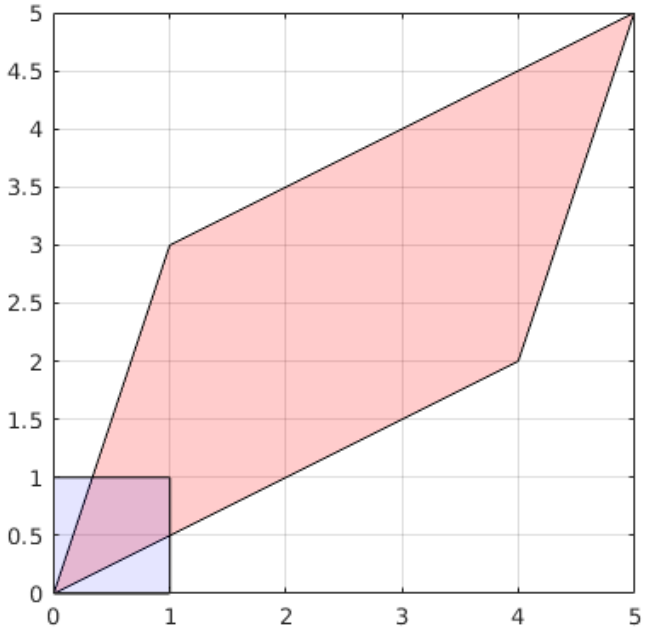


Note that the result looks exactly the same - it's just that now the x-vector $\langle 1, 0\rangle$, produces $\langle 4, 2\rangle$ and the y-vector, $\langle 0, 1\rangle$ produces $\langle 3, 1\rangle$.

Swapping the rows:

$$A_r = \begin{bmatrix} 1 & 4 \\ 3 & 2 \end{bmatrix}, \; X = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

$$AX = Y = \begin{bmatrix} 0 & 1 & 5 & 4 \\ 0 & 3 & 5 & 2 \end{bmatrix}$$



Note that if we swap **both** the columns and the rows then we get back to a transformation with determinant 10.

$$A_{rc} = \begin{bmatrix} 4 & 1 \\ 2 & 3 \end{bmatrix},\ X = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

$$AX = Y = \begin{bmatrix} 0 & 4 & 5 & 1 \\ 0 & 2 & 5 & 3 \end{bmatrix}$$

which produces the same parallelogram as the previous one but with columns reversed.

Summary So we find that,

$$\begin{bmatrix}3 & 2\\1 & 4\end{bmatrix}, \begin{bmatrix}4 & 1\\2 & 3\end{bmatrix} \text{ have determinant } > 0$$

$$\begin{bmatrix}2 & 3\\4 & 1\end{bmatrix}, \begin{bmatrix}1 & 4\\3 & 2\end{bmatrix} \text{ have determinant } < 0$$

If the product of the components on the diagonal of the matrix is greater than the components on the bottom-left to top-right diagonal then the determinant is > 0, if the reverse is true then the determinant is < 0, and if they are equal then the determinant $= 0$.
Note that, for the determinant to be 0 in our example, we need something like $detA = (4\times3)-(4\times3)$ which, due to the commutativity of multiplication can be achieved by both,

$$\begin{bmatrix}4 & 3\\4 & 3\end{bmatrix}, \begin{bmatrix}4 & 4\\3 & 3\end{bmatrix}$$

but in both cases the columns are proportional and therefore co-linear.

$n \times n$

The determinant of an $n \times n$ matrix is defined recursively as:

- if $n = 1$ then $det\, A = a_{11}$, i.e. the determinant is equal to the sole component.
- else if $n > 1$ then, defining A_{ij} as the matrix formed by leaving out the ith row and the jth column,

$$det\, A = a_{11} det\, A_{11} - a_{12} det\, A_{12} + \cdots \pm a_{1n} det\, A_{1n}$$

In the $n = 2$ case, each $det\, A_{ij}$ has $n = 1$ and so is simply equal to the sole component that is neither on the same row or column as the component a_{ij} that is multiplying it. This feature of the determinant calculation continues recursively for higher dimension matrices so that the calculation is always comprised of terms that are a product of components on each of the different

columns and rows. In fact, it comprises the products of all such possible combinations of components.
For example, when $n = 2$ the only combinations are,

$$\{a_{11}, a_{22}\} \text{ and } \{a_{12}, a_{21}\}$$

so there are only 2 terms in the determinant calculation.
When $n = 3$ the possible combinations are,

$$\{a_{11}, a_{22}, a_{33}\},\ \{a_{11}, a_{32}, a_{23}\},$$
$$\{a_{12}, a_{21}, a_{33}\},\ \{a_{12}, a_{31}, a_{23}\},$$
$$\{a_{13}, a_{21}, a_{32}\},\ \{a_{13}, a_{31}, a_{22}\}$$

so there are 6 terms in the determinant calculation. Notice that each term is generated by a different permutation of the columns while holding the rows fixed in ascending order and that the sign of each term is governed by how many permutations the permutation of columns is away from ascending order, $1, 2, \cdots, n$.

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}(a_{22}a_{33} - a_{32}a_{23}) - a_{12}(a_{21}a_{33} - a_{31}a_{23}) + a_{13}(a_{21}a_{32} - a_{31}a_{22})$$
$$= a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31}$$

Consequences

From this feature of the calculation we can see a number of the important properties of the determinant.

Proposition 49. *det* $I = 1$

Proof. Whatever the dimension of the identity matrix there will be only one combination of rows and columns that is the diagonal along which the 1s of the identity matrix reside. So, there will be a single term of the determinant calculation that is a product of 1s and all other terms will contain at 0s. In addition, the term that is along the diagonal has a positive sign in the determinant calculation. Therefore the result is 1. □

Proposition 50. *det A is linear in the rows of the matrix*

Proof. If p and q are row vectors and we have matrices A_p, A_q, A_{pq} in which are present, respectively, the row vector p, the one q, and the row $p+q$, then linearity implies that $det\, A_{pq} = det\, A_p + det\, A_q$. This can be seen since every term of the determinant calculation of A_{pq} will contain one of the components in the row $p+q$. So each term of the calculation will take the form,

$$(p+q)a_{ij}\cdots a_{mn} = p(a_{ij}\cdots a_{mn}) + q(a_{ij}\cdots a_{mn})$$

The other implication of linearity is that - if a row is multiplied by a scalar, c, to produce A_c then $det\, A_c = c\, det\, A$. Using a similar reasoning to the previous argument we have each term of the determinant taking the form,

$$c\, a_{ij}\cdots a_{mn}$$

which obviously results in the determinant being multiplied by c. □

Proposition 51. *If two columns are exchanged in the matrix then the determinant is multiplied by* -1

Proof. If columns p and q are exchanged then the components of p and q appear in terms with signs reversed. Since the components of p and q appear in every term of the determinant calculation, every term has the sign reversed. So the determinant is multiplied by -1. □

Proposition 52. $det\, A = 0$ *if there are two identical columns in the matrix*

Proof. If column p is identical to column q then we can swap columns p and q and we will have the same matrix so the determinant must also remain the same. But Proposition 51 proved that swapping two columns causes the determinant to be multiplied by -1. So, if A_{pq} is the matrix A after swapping the columns,

$$det\, A_{pq} = det\, A = -det\, A \iff det\, A = 0$$

□

Proposition 53. *Adding a multiple of one column to another leaves the determinant unchanged*

Proof. By combining Proposition 50 and Proposition 52 we find that if the columns of A are,

$$\vec{\boldsymbol{x_1}}, \vec{\boldsymbol{x_2}}, \cdots, \vec{\boldsymbol{x_p}}, \vec{\boldsymbol{x_q}}, \cdots, \vec{\boldsymbol{x_n}}$$

and the columns of A_c are,

$$\vec{\boldsymbol{x_1}}, \vec{\boldsymbol{x_2}}, \cdots, \vec{\boldsymbol{x_p}} + c\vec{\boldsymbol{x_q}}, \vec{\boldsymbol{x_q}}, \cdots, \vec{\boldsymbol{x_n}}$$

then,

$$\begin{aligned} det\, A_c &= det\left(\vec{\boldsymbol{x_1}}, \vec{\boldsymbol{x_2}}, \cdots, \vec{\boldsymbol{x_p}}, \vec{\boldsymbol{x_q}}, \cdots, \vec{\boldsymbol{x_n}}\right) + c \cdot det\left(\vec{\boldsymbol{x_1}}, \vec{\boldsymbol{x_2}}, \cdots, \vec{\boldsymbol{x_q}}, \vec{\boldsymbol{x_q}}, \cdots, \vec{\boldsymbol{x_n}}\right) \\ &= det\left(\vec{\boldsymbol{x_1}}, \vec{\boldsymbol{x_2}}, \cdots, \vec{\boldsymbol{x_p}}, \vec{\boldsymbol{x_q}}, \cdots, \vec{\boldsymbol{x_n}}\right) + c \cdot 0 \\ &= det\, A \end{aligned}$$

□

Better formulation (from Rudin's Principles of Mathematical Analysis)

Let $a(i, j)$ be the component in the ith row and jth column of the matrix A and,

$$sign(x) = \begin{cases} -1 & x < 0 \\ 0 & x = 0 \\ 1 & x > 0 \end{cases}$$

$$s(j_1, \cdots, j_n) = \prod_{p<q} sign(j_q - j_p)$$

Then the determinant,

$$det\, A = \sum_i s(j_1, \cdots, j_n) a(i, j_1) \cdots a(i, j_n)$$

defined over all n-tuples of n distinct values, $j_1, \cdots, j_n$ with $1 \leq j_r \leq n$ (i.e, permutations of $[1, n] \subset \mathbb{N}$) with each term being produced by a different permutation.

From this we can see that,

- **The determinant of the identity matrix is 1**
  Every term of the determinant will contain at least one 0 apart from the term that traverses the main diagonal, which is all 1s. We can see that there is only one such term because the main diagonal has $i = j$ and so there is only one such $j_1, \cdots, j_n$ that satisfies this.

- **The determinant is linear in the rows or columns of the matrix, holding the others constant**
  If a column, j_r, is multiplied by a scalar α and another column, j_k, is added to it, then the resulting determinant takes the form,

$$\begin{aligned} & det\, A = \sum_i s(j_1, \cdots, j_n) a(i, j_1) \cdots (\alpha a(i, j_r) + a(i, j_k)) \cdots a(i, j_n) \\ \iff & det\, A = \alpha a(i, j_r) \sum_i s(j_1, \cdots, j_n) a(i, j_1) \cdots a(i, j_n) + \\ & \qquad a(i, j_k) \sum_i s(j_1, \cdots, j_n) a(i, j_1) \cdots a(i, j_n) \end{aligned}$$

- **If two columns are exchanged then the determinant is multiplied by -1**
  If columns p and q are exchanged then this is equivalent to swapping j_p and j_q in the n-tuple so that $s(j_1, j_2, \cdots, j_n)$ changes sign and so the determinant is multiplied by -1.

- **If two columns are equal then the determinant will be 0**
  If two columns are the same then this is equivalent to a repetition of a value in the tuple $j_1, \cdots, j_n$ and so,

$$\exists\, p, q \text{ s.t. } sign(j_q - j_p) = 0 \implies s(j_1, \cdots, j_n) = 0$$

  which results in every term of the determinant being 0.
  This can also be proven by using the previous property that tells us that the determinant is multiplied by -1 when we exchange the identical columns but - since the columns are identical - the resultant matrix is the same - which means that the determinant remains unchanged. Therefore, the determinant must be 0.

Proposition 54. *If A and B are $n \times n$ matrices, then*

$$det\, BA = det\, A\, det\, B$$

Proof. Let the columns of A be the vectors, $\vec{\boldsymbol{x_1}}, \vec{\boldsymbol{x_2}}, \cdots, \vec{\boldsymbol{x_n}}$ so that for each column j,

$$\vec{\boldsymbol{x_j}} = \sum_i a(i, j) \vec{\boldsymbol{e_i}}$$

and define,

$$\Delta_B(\vec{\boldsymbol{x_1}}, \vec{\boldsymbol{x_2}}, \cdots, \vec{\boldsymbol{x_n}}) = \Delta_B(A) = det\, BA$$

so that,

$$det\,(B\vec{\boldsymbol{x_1}}, B\vec{\boldsymbol{x_2}}, \cdots, B\vec{\boldsymbol{x_n}}) = \Delta_B(\vec{\boldsymbol{x_1}}, \vec{\boldsymbol{x_2}}, \cdots, \vec{\boldsymbol{x_n}})$$

Since $B\vec{\boldsymbol{x_j}}$ is linear in $\vec{\boldsymbol{x_j}}$, $\Delta_B(\vec{\boldsymbol{x_1}}, \vec{\boldsymbol{x_2}}, \cdots, \vec{\boldsymbol{x_n}})$ is linear in each $\vec{\boldsymbol{x_j}}$ and so,

$$\begin{aligned}
\Delta_B(\vec{\boldsymbol{x_1}}, \vec{\boldsymbol{x_2}}, \cdots, \vec{\boldsymbol{x_n}}) &= \Delta_B(\sum_i a(i,1)\vec{\boldsymbol{e_i}}, \vec{\boldsymbol{x_2}}, \cdots, \vec{\boldsymbol{x_n}}) \\
&= \sum_i a(i,1)\Delta_B(\vec{\boldsymbol{e_i}}, \vec{\boldsymbol{x_2}}, \cdots, \vec{\boldsymbol{x_n}}) \\
&= \sum_{i_1} a(i_1,1) \sum_{i_2} a(i_2,2) \cdots \sum_{i_n} a(i_n,n)\, \Delta_B(\vec{\boldsymbol{e_{i_1}}}, \vec{\boldsymbol{e_{i_2}}}, \cdots, \vec{\boldsymbol{e_{i_n}}}) \\
&= \sum a(i_1,1)a(i_2,2)\cdots a(i_n,n)\, \Delta_B(\vec{\boldsymbol{e_{i_1}}}, \vec{\boldsymbol{e_{i_2}}}, \cdots, \vec{\boldsymbol{e_{i_n}}})
\end{aligned}$$

the sum being extended over all n-tuples, $(i_1, \cdots, i_n)$ such that $1 \leq i_j \leq n$. Also, by referring again to the properties of the determinant we see that,

$$\Delta_B(\vec{\boldsymbol{e_{i_1}}}, \vec{\boldsymbol{e_{i_2}}}, \cdots, \vec{\boldsymbol{e_{i_n}}}) = t(i_1, i_2, \cdots, i_n)\Delta_B(\vec{\boldsymbol{e_1}}, \vec{\boldsymbol{e_2}}, \cdots, \vec{\boldsymbol{e_n}})$$

where $t(i_1, i_2, \cdots, i_n) = 1, 0, -1$ similar to the function s previously. So, we end up with,

$$det\, BA = \sum a(i_1,1)a(i_2,2)\cdots a(i_n,n)t(i_1, i_2, \cdots, i_n)\, det\, B = det\, A\, det\, B$$

□

Proposition 55. *A linear operator A on $\mathbb{R}^n$ is invertible if and only if $det\, A \neq 0$*

Proof. If A is invertible then, $AA^{-1} = I$ and, using Proposition 49 and Proposition 54, we have,

$$det\, AA^{-1} = det\, A \cdot det\, A^{-1} = 1$$

so $det\, A$ cannot be 0.

Furthermore, if the columns of A are not independent then there is some linear combination of the columns that produces $\vec{\mathbf{0}}$. Since, by Proposition 53 we know that adding multiples of columns to other columns leaves the determinant unchanged, this means that the determinant is equal to the determinant of a matrix with $\vec{\mathbf{0}}$ as a column. Such a matrix has determinant 0, so the determinant of A is also 0. □

Corollary 11. *For invertible matrices,*

$$det\, A^{-1} = \frac{1}{det\, A}$$

Corollary 12. *The determinant is the only function that has the described properties.*

Proof. Every matrix, A, can be transformed by multiplication by elementary matrices to a row-reduced form, R, which is either the identity matrix - in the case that A is invertible - or a matrix with the last row zeroes - in the case where A is not invertible. So, the determinant of the row-reduced matrix, R, is either 1 or 0. Meanwhile, the determinants of the elementary matrices are:

- Add multiple of row to another row - determinant is 1 because this operation maintains the determinant of the identity.
- Swap two rows - determinant is -1 - determinant is -1 because this operation multiplies the determinant of the identity by -1.
- Multiply a row by some scalar c - determinant is c because this operation multiplies the determinant of the identity by c.

So, we have,

$$R = E_1 E_2 \cdots E_n A \implies det\, R = det\, E_1 E_2 \cdots E_n \cdot det\, A$$

where $det\, E_1 E_2 \cdots E_n$ is a known, non-zero quantity - say d_e. Since the determinant of R is either 0 or 1 this leaves the determinant of A being either 0 or $\frac{1}{d_e}$.

So, the value of the determinant of an arbitrary matrix, A, is wholly determined by the properties described. □

2.3.5 Permutation Matrices

(tags: linear algebra)

Definition. *A permutation p is a bijective map from a set S to itself. If a matrix P is the matrix associated with a permutation p then:*

- *the jth column of the matrix is the basis vector $e_{p(j)}$,*
- *P is a sum of the matrix units, $P = e_{p(1)1} + e_{p(2)2} + \cdots + e_{p(n)n} = \sum_j e_{p(j)j}$.*

Proposition 56. *If P, Q are permutation matrices associated with the permutations p, q then the matrix that corresponds to the permutation $p \circ q$ is PQ*

Proof. $pq(i) = p(q(i))$ and $PQX = P(QX)$ □

Proposition 57. *A permutation matrix P is invertible and its inverse is the transpose, $P^{-1} = P^T$*

Proof. A left-multiplying permutation matrix for a permutation, p, maps each row from the input matrix using a column j in the permutation matrix, to the output row, $p(j)$. Since the permutation, by definition, is bijective, we know that this mapping is one-to-one and invertible. If we transpose the matrix P to P^T, swapping rows and columns in the permutation matrix, then the new matrix, P^T maps input rows $p(j)$ into output rows j which is clearly the inverse permutation. □

Since a permutation matrix is a the result of permuting the rows of the identity matrix, clearly, its determinant is ± 1. A permutation is referred to as *odd* or *even* depending on whether its determinant is -1 or 1 respectively. Its determinant is called the *sign of the permutation*,

$$sign\, p = det\, p = \pm 1$$

The determinant of an arbitrary $n \times n$ matrix can be described as,

$$\begin{aligned} det\, A &= \sum_p \left[det \sum_j a_{p(j)j} e_{p(j)j} \right] \\ &= \sum_p \left[(a_{p(1)1} \cdots a_{p(n)n}) \cdot det \sum_j e_{p(j)j} \right] \\ &= \sum_p \Big[(a_{p(1)1} \cdots a_{p(n)n}) \cdot det\, P \Big] \\ &= \sum_p \Big[(sign\, p)(a_{p(1)1} \cdots a_{p(n)n}) \Big] \end{aligned}$$

This is the same formula as earlier and is known as the *complete expansion* of the determinant.

2.3.6 Cramer's Rule

(tags: linear algebra)

Expansion by minors on the jth column:

$$det\, A = (-1)^{j+1} a_{1j}\, det\, A_{1j} + (-1)^{j+2} a_{2j}\, det\, A_{2j} + \cdots + (-1)^{j+n} a_{nj}\, det\, A_{nj}$$

Expansion by minors on the ith row:

$$det\, A = (-1)^{i+1} a_{i1}\, det\, A_{i1} + (-1)^{i+2} a_{i2}\, det\, A_{i2} + \cdots + (-1)^{i+n} a_{in}\, det\, A_{in}$$

Definition. *If we form a matrix with elements* $\alpha_{ij} = (-1)^{i+j}\, det\, A_{ij}$ *and then transpose it we get the* ***adjoint matrix****.*

Notation. The adjoint of A is denoted $adj\, A$.

Following we use $[x]$ to denote a matrix as distinguished from a scalar.

Let $d = det\, A$. Then, if we multiply the adjoint matrix of $[A]$ by $[A]$ we get,

$$[adj\, A][A] = \begin{bmatrix} d & & & & & \\ & d & & & & \\ & & \cdot & & & \\ & & & \cdot & & \\ & & & & \cdot & \\ & & & & & d \end{bmatrix}$$

The off-diagonal elements come out zero because they involve a determinant calculation that involves the same row (or column) repeated and so those determinants are zero.

Theorem 17.

$$[adj\, A][A] = (det\, A)[I] = [A][adj\, A]$$

Corollary 13.

$$\frac{1}{det\, A}[adj\, A][A] = [I] \iff [A^{-1}] = \frac{1}{det\, A}[adj\, A]$$

This formulation of the inverse of a matrix can be used to write the solution to a system of linear equations (reverting to the normal notation) $AX = B$ as, multiplying on the left by A^{-1},

$$X = A^{-1}B = \frac{1}{det\, A}(adj\, A)B$$

so that X is a vector whose components, x_j, are expressed as,

$$\begin{aligned} x_j &= \frac{1}{det\, A}(b_1\alpha_{1j} + \cdots + b_n\alpha_{nj}) \\ &= \frac{1}{det\, A}(b_1(-1)^{1+j}\, det\, A_{1j} + \cdots + b_n(-1)^{n+j}\, det\, A_{nj}) \end{aligned}$$

which is the expansion by minors of A on the jth column but with the components a_{ij} of A replaced with the components of the vector B, divided by the determinant of A.

2.3.7 Linear Transformations

(tags: linear algebra)

The analogue for vector spaces of a homomorphism of groups is a map,

$$T : V \longmapsto W$$

from one vector space over a field $\mathbb{F}$ to another, which is compatible with addition and scalar multiplication:

$$T(\vec{\boldsymbol{v_1}} + \vec{\boldsymbol{v_2}}) = T(\vec{\boldsymbol{v_1}}) + T(\vec{\boldsymbol{v_2}}) \quad \text{and} \quad T(c\vec{\boldsymbol{v_1}}) = cT(\vec{\boldsymbol{v_1}}),$$

for all $\vec{\boldsymbol{v_1}}, \vec{\boldsymbol{v_2}} \in V$, $c \in \mathbb{F}$.

Definition. *A homomorphism between two vector spaces that is also compatible with scalar multiplication is called a* ***linear transformation****.*

The compatibility with addition of vectors implies that a linear transformation is a homomorphism between additive groups of vectors.

The Kernel and the Image of a Linear Transformation

Let $T : V \longmapsto W$ be any linear transformation. Then the kernel (or nullspace) of T is defined as,

$$ker\, T = \{\, \vec{\boldsymbol{v}} \mid T(\vec{\boldsymbol{v}}) = \vec{\boldsymbol{0}} \,\}$$

and the image of T as,

$$im\, T = \{\, \vec{\boldsymbol{w}} \in W \mid \exists \vec{\boldsymbol{v}} \in V \,.\, \vec{\boldsymbol{w}} = T(\vec{\boldsymbol{v}}) \,\}.$$

Proposition 58. *The kernel of $T : V \longmapsto W$ is a subspace of V and the image is a subspace of W.*

Proof. T is a homomorphism between additive groups of vectors and so the proof that the kernel and image are subspaces is the same as for general homomorphisms (see **??**). □

Examples of Linear Transformations

(31) As previously seen in section **??**, matrix multiplication on the left is a linear transformation. Let A be an $m \times n$ matrix with entries in $\mathbb{F}$ and consider A as an operator on column vectors $A : \mathbb{F}^n \longmapsto \mathbb{F}^m$. The kernel of A is the set of vectors that are solutions to $A\vec{\boldsymbol{x}} = \vec{\mathbf{0}}$ while the image (or range) is the set of vectors $\vec{\boldsymbol{b}}$ such that $A\vec{\boldsymbol{x}} = \vec{\boldsymbol{b}}$ has a solution.

(32) Also previously seen in **??** is that polynomials can be modeled as vectors. Let P_n be the vector space of real polynomials of degree $\leq n$. Then the derivative is a linear transformation $P_n \longmapsto P_{n-1}$.

Definition. *The dimension of the image is called the* ***rank*** *while the dimension of the kernel is known as the* ***nullity****.*

The Dimension Formula

Theorem 18. *Let $T : V \longmapsto W$ be a linear transformation, and assume that V is finite dimensional. Then,*

$$dim\, V = dim(ker\, T) + dim(im\, T) = rank + nullity.$$

Proof. □

2.3.8 Relations on matrices

(tags: linear algebra)

Let X be the set of $n \times n$ real matrices. Define a relation $\sim$ on X by:

$$M \sim N \iff \exists \text{ an invertible } P \in X \text{ s.t. } N = P^{-1}MP.$$

Prove that $\sim$ is an equivalence relation. (tags: linear algebra, equivalence relations)

Reflexivity:

$$\begin{aligned} & N = I^{-1}NI \\ \therefore\ & N \sim N \end{aligned}$$

Symmetry:

$$\begin{aligned} & N = P^{-1}MP \\ \iff\ & NP^{-1} = P^{-1}M(PP^{-1}) \\ \iff\ & NP^{-1} = P^{-1}M \\ \iff\ & PNP^{-1} = (PP^{-1})M \\ \iff\ & PNP^{-1} = M \\ \iff\ & R^{-1}NR = M, \ \ R \in X \\ \therefore\ & N \sim M \iff M \sim N \end{aligned}$$

Transitivity:

$$\begin{aligned} & N = P^{-1}MP, \ \ M = Q^{-1}AQ \\ \implies\ & N = P^{-1}(Q^{-1}AQ)P \\ \iff\ & N = (P^{-1}Q^{-1})A(QP) \\ \iff\ & N = R^{-1}AR, \ \ R \in X \\ \therefore\ & (N \sim M) \wedge (M \sim Q) \iff (N \sim Q) \end{aligned}$$

2.3.9 Linear Algebra of Polynomials

(tags: linear algebra, polynomials)

If we look for quadratic polynomials, $p\,(x)$, that pass throught the 3 points (1, 3), (3, 1) and (5, 2):

Then the first has roots at $x = 1, 3$ and passes through the point (5, 2). So, we have:

$$p(1) = p(3) = 0, p(5) = 2$$

meaning that $(x - 1)$and $(x - 3)$are factors. Therefore,

$$\begin{aligned} p(x) &= \alpha(x - 1)(x - 3) \\ &= \alpha(x^2 - 4x + 3) \end{aligned}$$

$$\begin{array}{rcl} & p(5) & = 2 \\ \implies & \alpha(5^2 - 4(5) + 3) & = 2 \\ \iff & 8\alpha & = 2 \\ \iff & \alpha & = \frac{1}{4} \end{array}$$

$$\therefore p(x) = \frac{1}{4}(x^2 - 4x + 3)$$

The second has roots at $x = 1, 5$ and passes throught the point (3, 1):

$$\begin{aligned} p(x) &= \alpha(x - 1)(x - 5) \\ &= \alpha(x^2 - 6x + 5) \end{aligned}$$

$$\begin{array}{rcl} & p(3) & = 1 \\ \implies & \alpha(3^2 - 6(3) + 5) & = 1 \\ \iff & -4\alpha & = 1 \\ \iff & \alpha & = -\frac{1}{4} \end{array}$$

$$\therefore p(x) = -\frac{1}{4}(x^2 - 6x + 5)$$

The third has roots at $x = 3, 5$ and passes through the point (1, 3):

$$\begin{array}{rl} p(x) & = \alpha(x - 3)(x - 5) \\ & = \alpha(x^2 - 8x + 15) \end{array}$$

$$\begin{array}{lcl} & p(1) & = 3 \\ \implies & \alpha(1^2 - 8(1) + 15) & = 3 \\ \iff & 8\alpha & = 3 \\ \iff & \alpha & = \frac{3}{8} \end{array}$$

$$\therefore p(x) = \frac{3}{8}(x^2 - 8x + 15)$$

Adding them together we get,

$$\begin{aligned} & \tfrac{1}{4}(x^2 - 4x + 3) - \tfrac{1}{4}(x^2 - 6x + 5) + \tfrac{3}{8}(x^2 - 8x + 15) \\ = & (\tfrac{1}{4} - \tfrac{1}{4} + \tfrac{3}{8})x^2 + (-1 + \tfrac{6}{4} - 3)x + (\tfrac{3}{4} - \tfrac{5}{4} + \tfrac{45}{8}) \\ = & \ \tfrac{3}{8}x^2 - \tfrac{10}{4}x + \tfrac{41}{8} \end{aligned}$$

Chapter 3

Analysis

(tags: analysis)

3.1 Supremum and Infimum

(tags: analysis)

3.1.1 Definitions

(tags: analysis)

Definition. *An upper bound on a set A is a value x such that,*

$$\forall a \in A,\, a \leq x$$

and a lower bound is similarly defined as a value y such that,

$$\forall a \in A,\, a \geq y.$$

A set is said to be ***upper-bounded*** *if there exists some upper-bound on the set and is said to be* ***lower-bounded*** *if there exists some lower bound on the set. If there exists both upper and lower bounds then the set is said to be* ***bounded****.*

Definition. *The* ***supremum*** *of a upper-bounded set A is a value σ_A such that σ_A is an upper bound on A and,*

$$\sigma'_A < \sigma_A \iff \exists a \in A \ s.t. \ a > \sigma'_A$$

which is to say that if $\sigma'_A < \sigma_A$ then σ'_A is not an upper bound on A and, if σ'_A is not an upper bound on A then it must be less than σ_A since σ_A is an upper bound on A.
An alternative, equivalent definition is,

$$\forall \epsilon > 0, \exists a \in A \ s.t. \ a > \sigma_A - \epsilon.$$

Issue Note that there is an apparent paradox here: This second definition implies that

$$\begin{aligned} & \forall \epsilon > 0 \ . \ \exists a \in A \text{ s.t. } a + \epsilon > \sigma_A \\ \iff \quad & \exists a \in A \text{ s.t. } a \geq \sigma_A \end{aligned}$$

which result, when combined with the upper-bound property, gives

$$\begin{aligned} & \exists a \in A \text{ s.t. } (a \geq \sigma_A) \wedge (a \leq \sigma_A) \\ \iff \quad & \exists a \in A \text{ s.t. } a = \sigma_A \end{aligned}$$

which says that there is always an element in the bounded set that is equal to the supremum. This is not correct - the supremum may be in the set or external to it.
The initial implication is not true, however. We cannot infer that $\exists a \in A$ s.t. $a \geq \sigma_A$. This can be seen with another rearrangement,

$$\begin{aligned} & \forall \epsilon > 0 \ . \ \exists a \in A \text{ s.t. } a + \epsilon > \sigma_A \\ \iff \quad & \forall \epsilon > 0 \ . \ \exists a \in A \text{ s.t. } \epsilon > \sigma_A - a \end{aligned}$$

which shows us that for any positive epsilon there needs to be an a close enough to the value of σ_A that the difference in their values is less than epsilon. Since a can approach arbitrarily close to σ_A this is achievable for any positive epsilon. This property seems to be equivalent to the fact that σ_A is a *limit point* of A but that will be covered properly in Topology.

Definition. *The **infimum** of a lower-bounded set A is defined similarly to the supremum: as a value τ_A such that τ_A is a lower bound on A and,*

$$\tau'_A > \tau_A \iff \exists a \in A \ s.t. \ a < \tau'_A$$

or alternatively,

$$\forall \epsilon > 0, \exists a \in A \ s.t. \ a < \tau_A + \epsilon.$$

Notation. The supremum of A is denoted $sup\, A$ and the infimum is denoted $inf\, A$.

3.1.2 Deductions using the supremum and infimum

(tags: analysis)

Proposition 59. *If a bounded set $A \subset \mathbb{R}$ has the property that,*

$$\forall x, y \in A \ . \ |x - y| < 1$$

then it follows that,

$$(sup\, A - inf\, A) \leq 1.$$

Proof. Let $\sigma_A = sup\, A$ and $\tau_A = inf\, A$ and w.l.o.g. assume that $x > y$. By the definitions of the supremum and infimum we have,

$$\begin{aligned} & \forall \epsilon > 0 \ . \ \exists x, y \in A \ . \ (x > \sigma_A - \epsilon) \wedge (y < \tau_A + \epsilon) \\ \iff \quad & \forall \epsilon > 0 \ . \ \exists x, y \in A \ . \ (x > \sigma_A - \epsilon) \wedge (-y > -\tau_A - \epsilon) \\ \iff \quad & \forall \epsilon > 0 \ . \ \exists x, y \in A \ . \ (x - y) > (\sigma_A - \tau_A) - 2\epsilon \end{aligned}$$

Now suppose, for contradiction, that $(\sigma_A - \tau_A) > 1$ then we can say that,

$$\exists r > 0 \ . \ (\sigma_A - \tau_A) = 1 + r.$$

If we then constrict ϵ such that,

$$\epsilon < \frac{r}{2} \iff 2\epsilon < r \iff r - 2\epsilon > 0$$

then the previous result tells us that, for $0 < \epsilon < \frac{r}{2}$,

$$\begin{aligned} & \exists x, y \in A \,.\, (x - y) > (\sigma_A - \tau_A) - 2\epsilon \\ \iff \quad & \exists x, y \in A \,.\, (x - y) > 1 + r - 2\epsilon > 1 \end{aligned}$$

which contradicts the set property that $\forall x, y \in A,\ |x - y| < 1$. So this shows that $(\sigma_A - \tau_A) \leq 1$. □

Proposition 60. *Let $A \subset \mathbb{R}$ be a bounded set and let B be the set defined by*

$$B = \{\, b \mid b = f(a),\ a \in A \,\}$$

where the function f is some strictly monotonic function. Then it follows that,

$$sup\, B = f(sup\, A).$$

Proof. A is bounded and so $\sigma_A = sup\, A$ exists. So, using the supremum properties we have,

$$\begin{aligned} & \forall a \in A \,.\, a \leq \sigma_A \\ \iff \quad & \forall a \in A \,.\, f(a) \leq f(\sigma_A) && \text{by monotonicity of f} \\ \iff \quad & \forall b \in B \,.\, b \leq f(\sigma_A) \end{aligned}$$

which is to say that $\sigma_B = f(\sigma_A)$ is an upper bound on B.
Furthermore, using the other supremum property, we have that,

$$\begin{aligned} & \sigma'_A < \sigma_A \implies \exists a \in A \text{ s.t. } a > \sigma'_A \\ \iff \quad & f(\sigma'_A) < f(\sigma_A) \implies \exists a \in A \text{ s.t. } f(a) > f(\sigma'_A) && \text{by strict monotonicity of f} \\ \iff \quad & \sigma'_B < \sigma_B \implies \exists b \in B \text{ s.t. } b > \sigma'_B. \end{aligned}$$

Therefore σ_B satisfies both requirements of the supremum and we have shown that,

$$sup\, B = f(sup\, A).$$

□

3.2 Limits

(tags: analysis)

3.2.1 Limits of sequences

(tags: analysis)

Problems with the informal description of a limit

If we say that a sequence tends to some value L when the terms of the sequence *gets closer and closer to* L we have the following problems:

- that the sequence gets closer and closer to many numbers so that this does not specify a single specific limit.
- that the sequence can have a limit but it's not the case that every term is closer than the previous term to the limit. For example,

  $$a_{2k} = 1/k,\ a_{2k-1} = \frac{1}{k+1}$$

  tends to 0 but $a_{2k} > a_{2k-1}$.

Definition. *A sequence a_n is said to* ***tend*** *to L or have the* ***limit*** *L iff,*

$$\forall \epsilon > 0 \in \mathbb{R},\ \exists N \in \mathbb{N}\ s.t.\ \forall n > N,\ |a_n - L| < \epsilon.$$

Definition. *The interval $(L-\epsilon, L+\epsilon)$ is called the ϵ-**neighbourhood of** L.*

Definition. *A sequence a_n is said to* ***tend to infinity*** *iff,*

$$\forall M > 0 \in \mathbb{R},\ \exists N \in \mathbb{N}\ \textit{s.t.}\ \forall n > N,\ a_n > M$$

and ***tend to minus-infinity*** *iff,*

$$\forall M < 0 \in \mathbb{R},\ \exists N \in \mathbb{N}\ \textit{s.t.}\ \forall n > N,\ a_n < M.$$

Definition. *A sequence that has a limit is called* ***convergent*** *and otherwise is called* ***divergent****. Note that* ***divergent*** *sequences include both sequences that remain bounded but oscillate without converging and those that tend to infinity (or minus-infinity).*

Examples of Convergence and Divergence

(33) **Non-convergent Oscillation**

The sequence $a_n = (-1)^n$ is divergent despite always remaining bounded within the interval $[-1, 1]$ as it neither converges to 1 or to -1.

(34) **Limit of an Infinite Recurrence**

Take the sequence given by,

$$a_1 = 1,\ a_{n+1} = \frac{a_n}{2} + \frac{3}{2a_n}\ \ (n \geq 1).$$

Assume there is an equilibrium value, a^*, then

$$\begin{aligned} & a^* = \frac{a^*}{2} + \frac{3}{2a^*} \\ \iff \quad & 2(a^*)^2 = (a^*)^2 + 3 \\ \iff \quad & (a^*)^2 = 3 \\ \iff \quad & a^* = \sqrt{3} \qquad\qquad \forall n, a_n \geq 0 \end{aligned}$$

So $\sqrt{3}$ is the steady-state value that this recurrence converges to as $n \to \infty$. If the recurrence didn't converge then the assumption of an equilibrium value would result in a contradiction. Note, however, that the fact that there is an equilibrium value does *not*, by itself, prove that this sequence converges (although this sequence does).

Proposition 61. *A sequence has at most one limit. In other words, a sequence can only converge, if at all, to a single unique value.*

Proof. Let L and L' both be limits of the sequence a_n, and the constant $\alpha = L - L'$. Then,

$$\forall \epsilon > 0 \in \mathbb{R},\ \exists N \in \mathbb{N} \text{ s.t. } \forall n > N,\ |a_n - L| < \epsilon$$

and

$$\forall \epsilon' > 0 \in \mathbb{R},\ \exists N' \in \mathbb{N} \text{ s.t. } \forall n' > N',\ |a_{n'} - L'| < \epsilon'.$$

But also we have,

$$|a_n - L'| = |(a_n - L) + (L - L')| = |(L - L') + (a_n - L)|$$

and using the triangle inequality,

$$\begin{aligned}
|L - L'| = |(L - L') + (a_n - L) - (a_n - L)| &\leq |(L - L') + (a_n - L)| + |-(a_n - L)| \\
\iff \qquad |L - L'| &\leq |(L - L') + (a_n - L)| + |a_n - L| \\
\iff \qquad |L - L'| - |a_n - L| &\leq |(L - L') + (a_n - L)| \\
\iff \qquad |\alpha| - |a_n - L| &\leq |\alpha + (a_n - L)|
\end{aligned}$$

Since $\alpha = L - L'$ is constant we can consider the situation when $\epsilon = \frac{|\alpha|}{2}$ then we have that,

$$\begin{aligned}
&\exists N \in \mathbb{N} \text{ s.t. } \forall n > N,\ |a_n - L| < \epsilon = \frac{|\alpha|}{2} \\
\iff \qquad &- |a_n - L| > -\frac{|\alpha|}{2} \\
\iff \qquad &|\alpha| - |a_n - L| > |\alpha| - \frac{|\alpha|}{2} \\
\iff \qquad &|\alpha| - |a_n - L| > \frac{|\alpha|}{2}.
\end{aligned}$$

Combining this with the previous result gives, for $\forall n > N$,

$$\frac{|\alpha|}{2} < |\alpha| - |(a_n - L)| \leq |\alpha + (a_n - L)| = |a_n - L'|$$

which, by rearranging a little, is,

$$\forall n > N,\ |a_n - L'| > \frac{|\alpha|}{2}.$$

But this means that if we also choose $\epsilon' = \frac{|\alpha|}{2}$ then there is no N' such that $\forall n' > N'$, $|a_{n'} - L'| < \epsilon'$ which contradicts the hypothesis that L' is also a limit of a_n. □

Proof. Another quicker way of proving the proposition is by letting $\epsilon = \epsilon' = \frac{|\alpha|}{2}$ so that,

$$2\epsilon = |\alpha| = |L - L'| = |L - a_n + a_n - L'| \leq |L - a_n| + |a_n - L'| = |a_n - L| + |a_n - L'|$$

which gives us,

$$2\epsilon \leq |a_n - L| + |a_n - L'|\,.$$

But by the limit definition,

$$|a_n - L| + |a_n - L'| < \epsilon + \epsilon'$$

and since we have set $\epsilon = \epsilon'$ then,

$$|a_n - L| + |a_n - L'| < 2\epsilon$$

which contradicts $2\epsilon \leq |a_n - L| + |a_n - L'|$. □

Definition. *If a_n is a sequence and $S = \{\, a_n \mid n \in \mathbb{N} \,\}$ then a_n is said to be **bounded below** if S has a lower bound and **bounded above** if S has an upper bound, and **bounded** if it is bounded above and below.*

Lemma 1. *Any finite set of elements from an ordered field has a minimum and a maximum.*

Proof. This can be proven quite easily using induction. Taking the base case of a set of cardinality one, clearly there is a maximum and a minimum both of which are the sole element of the set. Then, the induction step is to say, given a set S that has a maximum, s_{max}, and a minimum, s_{min}, if we add a new element e, then if e is greater than s_{max} it is the maximum of the new set and if it is less than s_{min} it is the minimum of the new set. Otherwise, the previous maximum and minimum also pertain to the new set. Therefore, adding a new element to a set that has a maximum and a minimum creates a new set with a maximum and a minimum. □

Proposition 62. *Any convergent sequence is bounded.*

Proof. Firstly, we need to prove that any finite sequence is bounded. We can do this simply by observing that any finite set of elements from an ordered field,

$$S = \{a_1, a_2, \ldots, a_n\}$$

has a minimum and a maximum.
Now, let a_n be an arbitrary convergent sequence so that,

$$\forall \epsilon > 0 \,.\, \exists N \in \mathbb{N} \,.\, \forall n > N \,.\, |a_n - L| < \epsilon$$

for some $L \in \mathbb{R}$.

Then, let S_{max} and S_{min} be the maximum and minimum respectively of the first N terms of a_n, $S = \{a_1, a_2, \ldots, a_N\}$, and,

$$\exists \epsilon > 0 \,.\, \forall n > N \,.\, |a_n - L| < \epsilon$$

so that, for $n > N$, the sequence a_n is bounded in the ϵ-neighbourhood of L. So, if we define $m = min\{S_{min}, L - \epsilon\}$ and $M = max\{S_{max}, L + \epsilon\}$, then the whole sequence a_n for all $n \in \mathbb{N}$ is bounded below by m and bounded above by M.
Therefore a_n is bounded. □

Definition. *An **increasing** sequence is a sequence a_n such that,*

$$\forall n \in \mathbb{N} \,.\, a_{n+1} \geq a_n$$

*and **decreasing** if,*

$$\forall n \in \mathbb{N} \,.\, a_{n+1} \leq a_n$$

*and **monotonic** if either increasing or decreasing.*

Proposition 63. *Any increasing sequence that is bounded above has a limit.*

Proof. Let a_n be an increasing sequence that is bounded above. Then,

$$\forall n \in \mathbb{N} \,.\, a_{n+1} \geq a_n$$

and let $S = \{\, a_n \mid n \in \mathbb{N} \,\}$. Since a_n is bounded above it has a supremum. Let $\sigma = sup\ S$ so that,

$$\forall a_n \in S \,.\, a_n \leq \sigma \quad \text{and} \quad \forall \epsilon > 0 \,.\, \exists a_n \in S \,.\, a_n > \sigma - \epsilon.$$

Therefore, for some arbitrary fixed $\epsilon > 0$,

$$\exists a_n \in S \,.\, a_n > \sigma - \epsilon$$

and setting $N = n$ so that $a_N > \sigma - \epsilon$, we have,

$$\forall n > N \in \mathbb{N} \,.\, a_n \geq a_N > \sigma - \epsilon$$

and so, recalling that σ is an upper bound on S,

$$\begin{aligned}
& \exists N \,.\, \forall n > N \in \mathbb{N} \,.\, (a_n > \sigma - \epsilon) \wedge (a_n \leq \sigma) \\
\iff \quad & \exists N \,.\, \forall n > N \in \mathbb{N} \,.\, a_n \leq \sigma < a_n + \epsilon \\
\iff \quad & \exists N \,.\, \forall n > N \in \mathbb{N} \,.\, 0 \leq \sigma - a_n < \epsilon \\
\implies \quad & \exists N \,.\, \forall n > N \in \mathbb{N} \,.\, |\sigma - a_n| < \epsilon.
\end{aligned}$$

But ϵ was an arbitrary positive value so,

$$\forall \epsilon > 0 \,.\, \exists N \,.\, \forall n > N \in \mathbb{N} \,.\, |\sigma - a_n| < \epsilon$$

and σ is, therefore, the limit of a_n. □

Corollary 14. *Any increasing sequence that is bounded above converges to the supremum of its elements (terms, values, etc.).*

Corollary 15. *A decreasing sequence that is bounded below converges to the infimum of its elements.*

3.2.2 Algebra of limits of sequences

(tags: analysis)

Proposition 64. *Let a_n and b_n be convergent sequences with limits a and b, respectively. Let C be a real number and let k be a positive integer. Then as $n \to \infty$,*

a) $Ca_n \to Ca$

b) $|a_n| \to |a|$

c) $a_n + b_n \to a + b$

d) $a_n b_n \to ab$

e) $a_n^k \to a^k$

f) if, for all n, $b_n \neq 0$ and $b \neq 0$, then $\frac{1}{b_n} \to \frac{1}{b}$.

Proof. We prove each property individually in the given order.

Proof of (a) $Ca_n \to Ca$

If $C = 0$ then $Ca_n = 0 = Ca$ for all n and the proposition holds trivially. If $C \neq 0$ then, since $a_n \to a$,

$$\forall \epsilon' > 0 \,.\, \exists N \,.\, \forall n > N \in \mathbb{N} \,.\, |a_n - a| < \epsilon'.$$

Now let $\epsilon = |C|\,\epsilon'$. Then,

$$\begin{aligned} & \forall \epsilon > 0 \,.\, \exists N \,.\, \forall n > N \in \mathbb{N} \,.\, |C|\,|a_n - a| < |C|\,\epsilon' = \epsilon \\ \iff \quad & \forall \epsilon > 0 \,.\, \exists N \,.\, \forall n > N \in \mathbb{N} \,.\, |Ca_n - Ca| < \epsilon. \qquad |x|\,|y| = |xy| \end{aligned}$$

Proof of (b) $|a_n| \to |a|$

$$\begin{aligned} & |a_n| = |a_n - a + a| \leq |a_n - a| + |a| && \text{the "triangle inequality"} \\ \iff & |a_n| - |a| \leq |a_n - a| \,. && (1) \end{aligned}$$

$$\begin{aligned} & |a| = |a - a_n + a_n| \leq |a - a_n| + |a_n| && \text{the "triangle inequality"} \\ \iff & |a| - |a_n| \leq |a - a_n| = |a_n - a| \\ \iff & |a_n| - |a| \geq -|a_n - a| \,. && (2) \end{aligned}$$

Putting (1) and (2) together we have,

$$\begin{aligned} & -|a_n - a| \leq |a_n| - |a| \leq |a_n - a| && \text{the "triangle inequality"} \\ \iff & ||a_n| - |a|| \leq |a_n - a| \,. \end{aligned}$$

The fact that a_n converges to a implies that $|a_n - a|$ converges to zero. Since it is an upper bound on the value of $||a_n| - |a||$, the value $||a_n| - |a||$ must also converge to zero. Specifically any value of n, ϵ such that $|a_n - a| < \epsilon$ will also satisfy $||a_n| - |a|| \leq |a_n - a| < \epsilon$.

Proof of (c) $a_n + b_n \to a + b$

Using, again, the "triangle inequality",

$$|(a_n + b_n) - (a + b)| = |(a_n - a) + (b_n - b)| \leq |a_n - a| + |b_n - b| \,.$$

So, $|a_n - a| + |b_n - b|$ is an upper bound on the value of $|(a_n + b_n) - (a + b)|$. If we take any arbitrary $\epsilon > 0$ then,

$$\exists N_1 \,.\, \forall n > N_1 \in \mathbb{N} \,.\; |a_n - a| < \frac{\epsilon}{2}$$

and

$$\exists N_2 \,.\, \forall n > N_2 \in \mathbb{N} \,.\; |b_n - b| < \frac{\epsilon}{2}.$$

Then, if we take $N = max\{N_1, N_2\}$, we have,

$$\forall n > N \in \mathbb{N} \,.\; |(a_n + b_n) - (a + b)| \leq |a_n - a| + |b_n - b| < \epsilon.$$

Proof of (d) $a_n b_n \to ab$

$$|a_n b_n - ab| = |a_n b_n - ab_n + ab_n - ab| \leq |b_n(a_n - a)| + |a(b_n - b)| \quad \text{the "triangle inequality"}$$
$$\iff \quad |a_n b_n - ab| \leq |b_n|\,|a_n - a| + |a|\,|b_n - b|$$

Since b_n converges, by Proposition 61, it is bounded. Therefore, $|b_n|$ has some upper bound which we shall call B. Then,

$$\forall \epsilon > 0 \,.\, \exists N_1 \,.\, \forall n > N_1 \in \mathbb{N} \,.\, |a_n - a| < \frac{\epsilon}{2B}$$

and

$$\forall \epsilon > 0 \,.\, \exists N_2 \,.\, \forall n > N_2 \in \mathbb{N} \,.\, |b_n - b| < \frac{\epsilon}{2\,|a|}.$$

Now, let $N = max N_1, N_2$. Then,

$$\forall n > N \in \mathbb{N} \,.\, B\,|a_n - a| + |a|\,|b_n - b| < \epsilon.$$

Proof of (e) $a_n^k \to a^k$

Using (d) - and because k is a positive integer - we can do induction on the power k.
Base cases 0 and 1 are clearly true as $k = 0$ results in a_n being the constant 1 for all n and so trivially converges to $a = 1$; and $k = 1$ results in the same sequence as a_n.
So, we perform the induction step for $k \geq 2$. Then, by the induction hypothesis, $a_n^{k-1} \to a^{k-1}$. But $a_n^k = a_n^{k-1} a_n$ and, by (d) and the induction hypothesis, we have that $a_n^{k-1} a_n \to a^{k-1} a = a^k$. Therefore, $a_n^k \to a^k$.

Proof of (f) $\forall n \,.\, b_n, b \neq 0 \implies \frac{1}{b_n} \to \frac{1}{b}$

Again invoking Proposition 61 and letting the upper bound on the sequence b_n be B,

$$\left|\frac{1}{b_n} - \frac{1}{b}\right| = \left|\frac{b - b_n}{b_n b}\right| = \left|\frac{b_n - b}{b_n b}\right| = \frac{|b_n - b|}{|b_n|\,|b|} \leq \frac{1}{B\,|b|}\,|b_n - b|\,.$$

Now, since $\frac{1}{B|b|}$ is a constant we can define the constant $C = \frac{1}{B|b|}$ and then we see that in (a) we have already proven that $C\,|b_n - b|$ converges to 0. In (a) we used that to prove that $Cb_n \to Cb$ but here it proves that $\frac{1}{b_n} \to \frac{1}{b}$. $\square$

3.2.3 Some theorems on limits of sequences

(tags: analysis)

Theorem 19. *If* $|a| < 1$ *then* $\lim_{n\to\infty} a^n = 0$.

Proof. First of all, note that if $|a| = 0$ then $a = 0 = a^n$ for all n and so the limit holds trivially. For this reason, from here on, we will consider only the case where $a \neq 0$.
There are 3 parts to this proof:

1. $|a| < 1 \implies \lim_{n\to\infty} |a|^n = 0$,
2. $|a|^n = |a^n|$,
3. $\lim_{n\to\infty} |a^n| = 0 \implies \lim_{n\to\infty} a^n = 0$.

1. $\boldsymbol{|a| < 1 \implies \lim_{n\to\infty} |a|^n = 0}$

   It would be natural to prove this by showing that,

   $$\forall \epsilon > 0 \,.\, \exists N \,.\, \forall n > N \in \mathbb{N} \,.\, \left||a|^n - 0\right| = |a|^n < \epsilon.$$

   We can show this by "reverse engineering" the value of N from the requirement that $|a|^n$ be less than ϵ,

   $$|a|^n < \epsilon \iff n \ln |a| < \ln \epsilon \iff n > \frac{\ln \epsilon}{\ln |a|}$$

   with the last step changing the direction of the inequality because we divide by $\ln |a|$ which - remembering that $|a| < 1$ - is a negative value. So, in this way, we have shown that $N = \frac{\ln \epsilon}{\ln |a|}$ is a general formula that relates a value of N with the required property with any arbitrary ϵ.
   However, this proof is not valid because it uses the concept of the logarithm which requires a lot of analysis that has not been proven at this stage. Since we are trying to build the fundamental basis of analysis, at this point we can only use concepts that are pre-requisites (axiomatic) in analysis or have been proven at this stage.

An alternative way to show this, using the properties of limits of sequences just proven, is as follows: Let x_n be the sequence $x_n = |a|^n$. Then, because $0 < |a| < 1$,

$$x_{n+1} = x_n \cdot x = |a|^n |a| < |a|^n = x_n$$

so that x_n is a decreasing sequence. Additionally, $\forall n \in \mathbb{R} \,.\, |a|^n \geq 0$ so 0 is a lower bound on the sequence. Therefore, the sequence converges to a limit (note we haven't yet established that 0 is the limit - only that it is a candidate). Furthermore, $x_{n+1} = |a|^{n+1} = |a|^n |a|$ and, if $x_n \to L$ then $x_{n+1} \to L$ also. But, putting these two facts together, along with property (d) of limits of sequences, means that,

$$L = \lim_{n\to\infty} |a|^{n+1} = \lim_{n\to\infty} |a|^n |a| = (\lim_{n\to\infty} |a|^n)(\lim_{n\to\infty} |a|) = |a| \, (\lim_{n\to\infty} |a|^n) = |a| \, L.$$

So,

$$L = |a| \, L \iff L(1 - |a|) = 0$$

and, since we know that $|a| \neq 1$, therefore L must be 0.

2. $\boldsymbol{|a|^n = |a^n|}$

   For $a \in \mathbb{R}, n \in \mathbb{N}$ it's easy to see that $|a| \, |a| \ldots |a| = |aa \ldots a|$. In actual fact, this appears to hold even for $n \in \mathbb{Q}$, e.g.

$$\left|(-1)^{\frac{1}{2}}\right| = |i| = 1 = \left|1^{\frac{1}{2}}\right| = |1| = 1$$

   but this should be checked when studying complex numbers more thoroughly. Also, the base a, can it also be complex?

3. $\boldsymbol{\lim_{n\to\infty} |a^n| = 0 \implies \lim_{n\to\infty} a^n = 0}$

   This can be proved directly from the definition of the limit.

$$\begin{aligned}
\lim_{n\to\infty} |a^n| = 0 &\iff \forall \epsilon > 0 \,.\, \exists N \,.\, \forall n > N \in \mathbb{N} \,.\, ||a^n| - 0| < \epsilon \\
&\iff \forall \epsilon > 0 \,.\, \exists N \,.\, \forall n > N \in \mathbb{N} \,.\, |a^n| < \epsilon \\
&\iff \forall \epsilon > 0 \,.\, \exists N \,.\, \forall n > N \in \mathbb{N} \,.\, |a^n - 0| < \epsilon \\
&\iff \lim_{n\to\infty} a^n = 0.
\end{aligned}$$

Bear in mind that, in general, $\lim_{n\to\infty} |x_n| = L \not\Longrightarrow \lim_{n\to\infty} x_n = L$. For example, if x_n converges to $-L$ then $|x_n|$ will converge to L.

```
code example
```

Furthermore, the example 32 showed how the fact of a_n and a_{n+1} converging to the same limit produces - when the sequence is expressed as a recurrence - an equilibrium value. □

The Sandwich Theorem

Proposition 65. *Let a_n, b_n, c_n be sequences such that,*

$$\text{for all } n,\ a_n \leq b_n \leq c_n \quad \text{and} \quad \lim_{n\to\infty} a_n = L = \lim_{n\to\infty} c_n.$$

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

Proof. $\lim_{n\to\infty} a_n = L$ means that,

$$\begin{aligned} & \forall \epsilon > 0\ .\ \exists N_1\ .\ \forall n > N_1 \in \mathbb{N}\ .\ |a_n - L| < \epsilon \\ \iff \quad & \forall \epsilon > 0\ .\ \exists N_1\ .\ \forall n > N_1 \in \mathbb{N}\ .\ -\epsilon < a_n - L < \epsilon \\ \iff \quad & \forall \epsilon > 0\ .\ \exists N_1\ .\ \forall n > N_1 \in \mathbb{N}\ .\ L - \epsilon < a_n < L + \epsilon. \end{aligned}$$

By the same reasoning we also have,

$$\forall \epsilon > 0\ .\ \exists N_2\ .\ \forall n > N_2 \in \mathbb{N}\ .\ L - \epsilon < c_n < L + \epsilon.$$

So, if we let $N = max\{N_1, N_2\}$ then we have,

$$\forall \epsilon > 0\ .\ \exists N\ .\ \forall n > N \in \mathbb{N}\ .\ L - \epsilon < a_n, c_n < L + \epsilon$$

and since we also know that $a_n \leq b_n \leq c_n$ it follows that,

$$\forall \epsilon > 0\ .\ \exists N\ .\ \forall n > N \in \mathbb{N}\ .\ L - \epsilon < a_n \leq b_n \leq c_n < L + \epsilon.$$

This shows that,

$$\forall \epsilon > 0\ .\ \exists N\ .\ \forall n > N \in \mathbb{N}\ .\ |b_n - L| < \epsilon.$$

□

An example application of the Sandwich Theorem

(35) **Prove that $|x| < 1 \implies \lim_{n\to\infty} x^n = 0$**

We have already proven this using the properties of limits in Theorem 16 but here we are going to prove it using the Sandwich Theorem.

Proof. Firstly, as we showed in 2, $|x^n - 0| = |x^n| = |x|^n$. So to show that x^n tends to zero we can show that $|x|^n$ tends to zero. So, w.l.o.g. we take $x > 0$ (since $x = 0$ makes the proposition trivially true). Then, notice that $0 < x < 1 \implies x = \frac{1}{1+h}$ for some $h > 0$. Then, we can show inductively that $(1+h)^n \geq 1 + hn$ as follows.

Base cases 0, 1: $(1+h)^0 = 1 = 1 + h(0),\ (1+h)^1 = 1 + h = 1 + h(1).$

Induction step $k > 1$

$$\begin{aligned} & (1+h)^k \geq 1 + hk \\ \iff \quad & (1+h)(1+h)^k \geq (1+h)(1+hk) \\ \iff \quad & (1+h)^{k+1} \geq 1 + hk + h + h^2k = 1 + h(k+1) + h^2k > 1 + h(k+1) \end{aligned}$$

This result implies that,

$$x^n = \frac{1}{(1+h)^n} \leq \frac{1}{1+hn}$$

so that,

$$0 < x^n \leq \frac{1}{1+hn}.$$

Since h is some fixed value, clearly,

$$\lim_{n\to\infty} \frac{1}{1+hn} = 0$$

and, obviously, the limit of the constant 0 is always 0 so, by the Sandwich Theorem,

$$\lim_{n\to\infty} x^n = 0.$$

□

3.2.4 Subsequences

(tags: analysis)

Definition. *Let $(a_n)_{n\in\mathbb{N}}$ be a sequence and consider some strictly increasing natural numbers $(k_1, k_2, k_3, \dots)$ that is, $(k_1 < k_2 < k_3 < k_4 < \dots)$. Then the sequence $(a_{k_n})_{n\in\mathbb{N}}$ is called a **subsequence** of the sequence $(a_n)_{n\in\mathbb{N}}$.*
*Note that a **subsequence** is always infinite (I think).*

Theorem 20. *If a_n is a sequence that tends to a limit, then any subsequence of it tends to the same limit.*

Proof. Firstly, notice that if the nth index of some subsequence is k_n then $k_n \geq n$ (because the subsequence can only skip terms of the original - it can't add in terms). So then, if we have a sequence a_n that tends to a limit a and an arbitrary subsequence a_{k_n} then,

$$\forall \epsilon > 0 \,.\, \exists N \,.\, \forall n > N \in \mathbb{N} \,.\, |a_n - a| < \epsilon \implies |a_{k_n} - a| < \epsilon$$

because $k_n \geq n > N$. □

Theorem 21. *Every sequence has a monotonic subsequence.*

Proof. Either there is an infinite number of terms that are greater than all the following terms or there is not. If there is not, then after the last such term all terms have a term that follows them that is greater than or equal to them.
In the first case we have a strict monotonic decreasing sequence and in the second we have a non-strict monotonic increasing sequence. □

If we put this theorem together with what we learned about bounded sequences in Proposition 62 and its corollaries - that monotonic bounded sequences are convergent - then we get one of the most famous results in analysis, The Bolzano-Weierstrass Theorem.

Theorem 22. ***The Bolzano-Weierstrass Theorem***
Every bounded sequence has a convergent subsequence.

3.2.5 Examples of limits of sequences

(tags: analysis)

(36) **Let $(a_n)_{n\in\mathbb{N}}$ be a sequence, and let $(b_n)_{n\in\mathbb{N}}$ be the sequence defined by $b_n = |a_n|$ for $n \in \mathbb{N}$. Which of the following two statements implies the other?**

Answer: a_n converges $\implies$ b_n converges also but b_n converges $\not\implies$ a_n converges.
The first implication is because,

$$\begin{aligned} & |a_n| = |a_n - a + a| \leq |a_n - a| + |a| \\ \iff \quad & |a_n| - |a| \leq |a_n - a| \end{aligned}$$

$$\begin{aligned} & |a| = |a - a_n + a_n| \leq |a_n - a| + |a_n| \\ \iff \quad & |a| - |a_n| \leq |a_n - a| \\ \iff \quad & |a_n| - |a| \geq -|a_n - a| \end{aligned}$$

which both together imply that $||a_n| - |a|| \leq |a_n - a|$.
The latter non-implication is easy to see if one thinks of a sequence that consists of two subsequences that converge to 2 and -2. Then, their absolute value would converge to 2 but their values do not converge. Remember Theorem 17, for a sequence to converge to a limit, every subsequence of it must converge to the same limit.

(37) **What is the behaviour as $n \to \infty$ of the following:**

(i) $\frac{2n^3+1}{n+1}\left(\frac{3}{4}\right)^n$

$$0 < \frac{2n^3+1}{n+1}\left(\frac{3}{4}\right)^n < \frac{3n^3}{n}\left(\frac{3}{4}\right)^n = 3n^2\left(\frac{3}{4}\right)^n \to 0$$

(ii) $\frac{2^{2n}+n}{n^3 3^n+1}$

$$\frac{2^{2n}+n}{n^3 3^n+1} = \frac{4^n+n}{n^3 3^n+1} > \frac{4^n}{2n^3 3^n} = \frac{(4/3)^n}{2n^3} \to \infty$$

(38) **Let (a_n) be a sequence of non-negative numbers. Prove that if $a_n \to L$ as $n \to \infty$ then $\sqrt{a_n} \to \sqrt{L}$ as $n \to \infty$.**

Proof. We are told that $a_n \to L$ as $n \to \infty$ so we have,

$$\forall \epsilon' > 0 \,.\, \exists N \,.\, \forall n > N \in \mathbb{N} \,.\, |a_n - L| < \epsilon'.$$

We are also told that (a_n) is non-negative so $L \geq 0$.
First, consider the possibility that $L = 0$. Then $|a_n| < \epsilon'$ for $n > N$. But,

$$|a_n| = (\sqrt{a_n})^2 < \epsilon' \iff \sqrt{a_n} < (\epsilon')^2$$

so that taking $\epsilon = (\epsilon')^2$ we obtain,

$$\forall \epsilon > 0 \,.\, \exists N \,.\, \forall n > N \in \mathbb{N} \,.\, |\sqrt{a_n}| < \epsilon.$$

The remaining possibility is that $L > 0$. Notice that the expression we're looking to bound can be rewritten as follows:

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

Furthermore, since $a_n \to L$ and $L > 0$, clearly $0 < L/2 < L$ and there will be some $\epsilon < L/2$ such that,

$$L - L/2 < L - \epsilon < a_n < L + \epsilon < L + L/2 \iff |a_n - L| < \epsilon < L/2.$$

This means that, inside this ϵ-neighbourhood of L, we can find a constant lower bound on $\sqrt{a_n} + \sqrt{L}$ as,

$$\sqrt{a_n} + \sqrt{L} > \sqrt{L/2} + \sqrt{L} = C$$

for constant C. Now we have

$$\left|\sqrt{a_n} - \sqrt{L}\right| = \left|\frac{a_n - L}{\sqrt{a_n}+\sqrt{L}}\right| < \frac{|a_n - L|}{C} < \frac{\epsilon'}{C}$$

so we can take $\epsilon = \epsilon'/C$ to obtain $|a_n - L| < \epsilon$ as required. □

Proof. This is an alternative proof of the $L > 0$ case using the fact that

$$\sqrt{x+y} \leq \sqrt{x} + \sqrt{y}.$$

Choose ϵ such that $0 < \epsilon \leq \sqrt{L}$ and take N so that, for $n > N$, $|a_n - L| < \epsilon^2$, or in other words $L - \epsilon^2 < a_n < L + \epsilon^2$. Then we have

$$\sqrt{L} - \epsilon \leq \sqrt{L - \epsilon^2} < \sqrt{a_n} < \sqrt{L + \epsilon^2} \leq \sqrt{L} + \epsilon$$

which places $\sqrt{a_n}$ in ϵ-neighbourhood of L, so we're done.
Note that we need $\epsilon \leq \sqrt{L}$ for $\sqrt{L} - \epsilon \leq \sqrt{L - \epsilon^2}$ and also – so long as we are doing *real* analysis – this proof is only for the $L > 0$ case because, when $L = 0$, $\sqrt{L - \epsilon^2}$ will be a complex number. □

(39) **Let a_n be a positive decreasing sequence. Show that, if there exist numbers N and α such that**

$$0 < \frac{a_{n+1}}{a_n} < \alpha < 1 \qquad \forall n > N$$

then $a_n \to 0$ as $n \to \infty$. But if, on the other hand, we have

$$0 < \frac{a_{n+1}}{a_n} < 1 \qquad \forall n > N$$

then we cannot conclude that $a_n \to 0$.

The basic principle here is that, if a convergent sequence converges to a non-zero limit, then the ratio of consecutive terms must tend to 1. If the ratio of consecutive terms remains below 1 then the sequence must go to zero.

If $\frac{a_{n+1}}{a_n} < \alpha$ then α is an upper bound on the ratio of consecutive terms so we can deduce that,

$$a_{n+1} < \alpha a_n < a_n \qquad \text{since } \alpha < 1$$

and that, if we let a_N be the value of the sequence at $n = N$ and consider the sequence for $n > N$,

$$a_n < \alpha^{n-N} a_N$$

which is of the form,

$$a_n < \alpha^n C$$

for constant C. So we have bound the values of the sequence below $\alpha^n C$ which clearly goes to 0 as $n \to \infty$. Since we are told that the sequence is positive so that a_n is also bounded below by 0, we can conclude, by Sandwich Theorem, that $a_n \to 0$ as $n \to \infty$. $\square$

An alternative way of showing the same thing is to use the algebra of limits and the fact that, in a convergent sequence, any subsequence must also converge to the same limit (Theorem 17) to deduce that if $a_n \to L$ then we must also have $a_{n+1} \to L$. Then, if $L \neq 0$ we can apply the algebra of limits to obtain,

$$\lim_{n\to\infty} \frac{a_{n+1}}{a_n} = \frac{L}{L} = 1$$

which contradicts $\frac{a_{n+1}}{a_n} < \alpha < 1$. Therefore, $L = 0$. $\square$

On the other hand, if $\frac{a_{n+1}}{a_n} \to 1$ as $n \to \infty$ then the sequence may converge to 0 or to some other value. For example:

(i) $\boldsymbol{a_n = \frac{1}{n} + 1}$

$$\frac{a_{n+1}}{a_n} = \frac{n(n+2)}{(n+1)^2} = \frac{n^2+2n}{n^2+2n+1} \to 1 \text{ as } n \to \infty$$

and clearly,

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

But also,

(ii) $\boldsymbol{a_n = \frac{1}{n}}$

$$\frac{a_{n+1}}{a_n} = \frac{n}{n+1} \to 1 \text{ as } n \to \infty$$

even though

$$\lim_{n\to\infty} \frac{1}{n} = 0.$$

3.2.6 Limits of functions

(tags: analysis)

Definition of the limit of a function

Definition. *Let $f : \mathbb{R} \mapsto \mathbb{R}$ be a function. We say that L is the* ***limit of $f(x)$ as x approaches a*** *if, for each $\epsilon > 0$ there exists $\delta > 0$ such that,*

$$0 < |x - a| < \delta \implies |f(x) - L| < \epsilon.$$

Definition. *Let $f : \mathbb{R} \mapsto \mathbb{R}$ be a function. We say that* ***$f(x)$ tends to infinity as x approaches a*** *if, for each K there exists $\delta > 0$ such that,*

$$0 < |x - a| < \delta \implies f(x) > K.$$

Examples of limits of functions

(40) **Prove that if $f : \mathbb{R} \mapsto \mathbb{R}$ s.t. $f(x) = x^2 + x$ then $\lim_{x \to 2} f(x) = 6$.**

Let $0 < |x - 2| < \delta$ and consider some arbitrary $\epsilon > 0$. Then we have,

$$\left|(x^2 + x) - 6\right| = |(x - 2)(x + 3)| \leq |x - 2|\,|x + 3| < \delta\,|x + 3|\,.$$

It's tempting at this point to say that, since we are examining the behaviour when x approaches 2 so we can assume $x \approx 2 \iff (x + 3) \approx 5$ and then we can set $\delta = \frac{\epsilon}{6}$ so that,

$$0 < |x - 2| < \delta = \frac{\epsilon}{6} \implies |f(x) - 6| < \frac{|x + 3|}{6}\epsilon < \epsilon.$$

However, there is a subtle logical problem here: We have considered any arbitrary $\epsilon > 0$, meaning that ϵ could be large. Then, when we set $\delta = \frac{\epsilon}{6}$ we linked the value of δ to that of ϵ so that δ may also be

arbitrarily large. But $0 < |x-2| < \delta$ so that $|x-2|$ may be arbitrarily large also. This contradicts the assumption that $x \approx 2$ and so the argument we have here is only valid for small ϵ.

So, we need an alternative approach. Consider that,

$$|x+3| = |x-2+2+3| \leq |x-2| + |2+3| = |x-2| + 5 < \delta + 5.$$

This means that,

$$\left|(x^2+x)-6\right| < \delta\,|x+3| < \delta(\delta+5).$$

But note, do **not** start trying to solve a quadratic. There is a much better way.

Now - here comes the clever bit - if we set $\delta = min\{1, \frac{\epsilon}{6}\}$ then *both* 1 and $\frac{\epsilon}{6}$ are an upper bound on the value of δ so that we can say,

$$\left|(x^2+x)-6\right| < \delta(\delta+5) \leq 6\delta \leq 6\frac{\epsilon}{6} = \epsilon.$$

Algebra of limits of functions

Theorem 23. *Let $f, g : \mathbb{R} \mapsto \mathbb{R}$ be two functions and c be any real number. Suppose that $\lim_{x\to a} f(x) = L$ and $\lim_{x\to a} g(x) = M$. Then,*

1. $\lim_{x\to a}(cf)(x) = cL$
2. $\lim_{x\to a}(|f|)(x) = |L|$
3. $\lim_{x\to a}(f+g)(x) = L + M$
4. $\lim_{x\to a}(f-g)(x) = L - M$
5. $\lim_{x\to a} f(x)g(x) = LM$
6. $\lim_{x\to a}(f/g)(x) = L/M$ *provided $g(x) \neq 0$ for any x in the neighbourhood of a.*

Proof. see: Lamar University - Paul's Online Notes - Proofs of limit properties □

One-sided limits

Definition. *Let $f : \mathbb{R} \to \mathbb{R}$ be a function. We say that L is the* ***limit of $f(x)$ as x approaches a from the left (or from below)****, denoted by* $\lim_{x \to a^-} f(x) = L$*, if for each $\epsilon > 0$, there exists $\delta > 0$ such that,*

$$a - \delta < x < a \implies |f(x) - L| < \epsilon.$$

Limits at infinity

Definition. *Let $f : \mathbb{R} \to \mathbb{R}$ be a function. We say that L is the* ***limit of $f(x)$ as x approaches ∞****, denoted by* $\lim_{x \to \infty} f(x) = L$*, if for each $\epsilon > 0$, there exists $M > 0$ such that,*

$$x \geq M \implies |f(x) - L| < \epsilon.$$

3.2.7 Continuity

(tags: analysis)

Definition. *A function f is **continuous at a point a** if*

- *$f(a)$ is defined,*
- *$\lim_{x\to a} f(x) = f(a)$.*

More formally, the definition of $\lim_{x\to a} f(x) = L$ is,

$$\forall \epsilon > 0 \;.\; \exists \delta \text{ s.t. } 0 < |x - a| < \delta \implies |f(x) - L| < \epsilon.$$

But if $f(a)$ is defined, then when $|x - a| = 0$ (i.e. when $x = a$) we have

$$f(x) = f(a) \implies |f(x) - f(a)| = 0 < \epsilon.$$

Therefore, if $f(a)$ is defined and $\lim_{x\to a} f(x) = f(a)$, then

$$\forall \epsilon > 0 \;.\; \exists \delta > 0 \text{ s.t. } |x - a| < \delta \implies |f(x) - f(a)| < \epsilon.$$

Definition. *A function is **continuous** if it is **continuous at every point**.*

Definition. *A function is **continuous on the closed interval $[a, b]$** if it is*

- *continuous at every point in the open interval (a, b),*
- *$\lim_{x\to a^+} f(x) = f(a)$,*
- *$\lim_{x\to b^-} f(x) = f(b)$.*

Definition. *A function is* ***left continuous*** *or* ***continuous on the left*** *at* a *if,*

$$\lim_{x \to a-} f(x) = f(a).$$

Obviously, from the other side, a function can be ***right continuous****.*

Theorem 24. *Let* $f, g : \mathbb{R} \to \mathbb{R}$ *be functions that are continuous at* $a \in R$ *and* c *be any real number. Then* $|f|, (cf), (f-g), (f+g), (f(x)g(x))$ *are all continuous at* a*, and* (f/g) *is continuous provided* $g(x) \neq 0$ *for any* x *in some neighbourhood of* a*.*

Proof. This follows from the algebra of limits of functions given in Theorem 20. □

Corollary 16. *It follows then that every polynomial is continuous. This can be seen as the most simple polynomial is a constant - which is clearly continuous; also* $f(x) = x$ *is clearly continuous. Then powers of* x *are continuous as they are products of continuous functions and when multiplied by coefficients this is a constant multiplying a continuous function so the resultant function is continuous. Then any polynomial is a summation of such terms so the result is continuous as the sum of continuous functions is continuous.*

Theorem 25. *If* g *is a function which is continuous at* a*, and* f *is a function which is continuous at* $g(a)$*. Then* $(f \circ g)$ *is continuous at* a*.*

Proof. Continuity of f at $g(a)$ guarantees that for any $\epsilon > 0$ there exists some $\delta' > 0$ such that $|x' - g(a)| < \delta' \implies |f(x') - f(g(a))| < \epsilon$ and continuity of g at a guarantees that for any $\delta' > 0$ there exists some $\delta > 0$ such that $|x - a| < \delta \implies |g(x) - g(a)| = |x' - g(a)| < \delta'$. □

Corollary 17. $\lim_{x \to a}(f \circ g)(x) = \lim_{x \to a} f(g(x)) = f(\lim_{x \to a} g(x))$

Continuity of functions over sequences

We now give an alternative definition of continuity which ties in the concept of limits for sequences.

Theorem 26. *A function f is continuous at a if and only if for each sequence (x_n) such that $\lim_{n\to\infty} x_n = a$ we have $\lim_{n\to\infty} f(x_n) = f(a)$.*

Before the proof, an important point to note is that this theorem applies to "each sequence" with the described limit. This is important as it is possible to find an individual sequence such that the inference is not valid. For example, the constant sequence $\forall n \in \mathbb{N} \,.\, x_n = a$ clearly tends to a as $n \to \infty$ but this would not imply continuity of f as the limit of $f(x_n)$ for such a sequence would amount to saying that $f(a) = f(a)$. The definition of continuity is assertion of the equality of the limit of f over values in the neighbourhood of a (but not at a itself) with the value of f at a. So, continuity is implied by the fact that the limit of $f(x_n)$ for the constant sequence $x_n = a$ is equal to the limit of $f(x_n)$ for all other sequences x_n whose limit is a.
Another way of looking at it is that f is continuous at a because the limit there equals $f(a)$ however the argument converges to a.

Proof. Breaking it down into two propositions we have,

$$(\forall x_n \text{ s.t. } \lim_{n\to\infty} x_n = a) \ \lim_{n\to\infty} f(x_n) = f(a) \tag{P_1}$$

$$\forall \epsilon > 0 \,.\, \exists \delta > 0 \,.\, |x - a| < \delta \implies |f(x) - f(a)| < \epsilon \tag{P_2}$$

and we need to show that $P_1 \iff P_2$.

So, to begin with we'll assume show that $P_1 \implies P_2$.
Unpacking P_1 we have a function f such that

$$\forall \epsilon > 0 \,.\, \exists N \,.\, \forall n > N \in \mathbb{N} \,.\, |f(x_n) - f(a)| < \epsilon \tag{1}$$

where x_n is a sequence such that

$$\forall \delta > 0 \,.\, \exists N' \,.\, \forall n > N' \in \mathbb{N} \,.\, |x_n - a| < \delta. \tag{2}$$

Then, if we choose a particular $\epsilon > 0$, there exists some N such that for all $n > N$ we have $|f(x_n) - f(a)| < \epsilon$. Now there will also be some $N' \leq N$ so that $\forall n > N$, $n > N'$ and so there is some δ such that $|x_n - a| < \delta$. Since

P_1 says that this is the case *whenever* we have an x_n such as this, P_1 may be rewritten as

$$\forall \epsilon > 0 \,.\, (\exists N \,.\, \forall n > N \in \mathbb{N}) \,.\, \exists \delta > 0 \,.\, |x_n - a| < \delta \implies |f(x_n) - f(a)| < \epsilon.$$

Now, if we remove reference to the number of the term of the sequences - N, n - we have

$$\forall \epsilon > 0 \,.\, \exists \delta > 0 \,.\, |x - a| < \delta \implies |f(x) - f(a)| < \epsilon$$

which is the statement of continuity of f in P_2.

Next we prove $P_2 \implies P_1$.
So now we begin by assuming P_2 which was that,

$$\forall \epsilon > 0 \,.\, \exists \delta > 0 \,.\, |x - a| < \delta \implies |f(x) - f(a)| < \epsilon.$$

Actually, it is quite easy to apply a similar logic as previously but in reverse to make the converse implication. We can choose any ϵ and there exists some δ such that P_2 holds. Then, as we have seen previously in (2), P_1 tells us that, for this value of δ,

$$\exists N' \,.\, \forall n > N' \in \mathbb{N} \,.\, |x_n - a| < \delta$$

and P_2 tells us that,

$$|x_n - a| < \delta \implies |f(x_n) - f(a)| < \epsilon.$$

Putting the two together we get,

$$\forall \epsilon > 0 \,.\, \exists \delta > 0 \,.\, \exists N' \,.\, \forall n > N' \in \mathbb{N} \,.\, |x_n - a| < \delta \implies |f(x_n) - f(a)| < \epsilon$$

which implies that

$$\forall \epsilon > 0 \,.\, \exists N \,.\, \forall n > N \in \mathbb{N} \,.\, |f(x_n) - f(a)| < \epsilon.$$

So we have shown that P_2 implies that (2) implies (1) which is $P_2 \implies P_1$ as required.

For completeness, let's consider another way of proving that $P_1 \implies P_2$ using a proof by contradiction.
So, we are assuming P_1 but also assuming, for contradiction, that P_2 is false. Then we are negating the statement,

$$\forall \epsilon > 0 \,.\, \exists \delta > 0 \,.\, |x - a| < \delta \implies |f(x) - f(a)| < \epsilon$$

so we are asserting that,

$$\forall \epsilon > 0 \,.\, \nexists \delta > 0 \,.\, |x - a| < \delta \implies |f(x) - f(a)| < \epsilon$$

which is equivalent to

$$\forall \epsilon > 0 \,.\, \forall \delta > 0 \,.\, |x - a| < \delta \implies |f(x) - f(a)| < \epsilon$$

or alternatively,

$$\forall \epsilon > 0 \,.\, \forall \delta > 0 \,.\, \exists x \,.\, (|x - a| < \delta) \wedge (|f(x) - f(a)| \geq \epsilon).$$

So, this says that we can choose any arbitrary $\epsilon > 0$ and for all $\delta > 0$ there will be some x in the δ-neighbourhood of a such that $|f(x) - f(a)| \geq \epsilon$.

Now, if we choose a value of δ that depends on a natural number n in such a way that $\delta \to 0$ as $n \to \infty$ – for example, if $\delta = 1/n$ – then the δ-neighbourhoods around a will get smaller as $n \to \infty$. Then we can select an x from the δ-neighbourhood that corresponds to a particular value of n and call it x_n and, in this way, we create a sequence x_n that converges to a. So we have $\lim_{n\to\infty} x_n = a$.
But now, for every δ there is an x_n in the δ-neighbourhood of a with $|f(x_n) - f(a)| \geq \epsilon$ and, if we choose this value for x_n, we have a constructed a sequence that converges to a but

$$\lim_{n\to\infty} f(x_n) \neq f(a)$$

which contradicts hypothesis P_1. □

Corollary 18. $\lim_{n\to\infty} f(x_n) = f(\lim_{n\to\infty} x_n)$

Continuous Functions on Closed Intervals

Definition. *For a subset X of the domain of a function f, we say that f is bounded on X if there exists M such that $|f(x)| \leq M$ for each $x \in X$.*

Definition. *We define the supremum (or maximum) of f on X as $sup\,\{\, f(x) \mid x \in X \,\}$ (or $max\,\{\, f(x) \mid x \in X \,\}$ if it exists).*

Examples of bounded and unbounded functions

(41) The indicator function for rational numbers within the reals, known as the **Dirichlet Function**,

$$f(x) = \begin{cases} 0 & x \text{ is irrational} \\ 1 & x \text{ is rational} \end{cases}.$$

This function is nowhere continuous because between every two irrational numbers there is a rational number (and vice-versa) so $f(x)$ is flipping between 0 and 1 in every neighbourhood of every point however small a neighbourhood we consider. So this function never converges anywhere but *is* bounded because it only ever takes values of 1 or 0 so, clearly, its maximum is 1 and minimum is 0.

(42) The function $f(x) = 1/x$ over the interval $(0, 1]$ is continuous at every point but unbounded as it goes to infinity as $x \to 0$. If we consider the same function over the closed interval $[0, 1]$ then we no longer have continuity over this interval as there is a singularity at $x = 0$.

Extreme Value Theorem

Theorem 27. *Let f be continous on $[a, b]$. Then f is bounded on $[a, b]$ and it achieves its maximum; that's to say, the supremum is equal to the maximum.*

Note that, even if f is defined at every point in $[a, b]$, if it is not continuous then it may not be bounded. There do exist functions that are not continuous but bounded (for example the Dirichlet Function 39) but there also exist functions that are not continuous and unbounded such as the reciprocal function 40. So, functions that are not continuous on a closed interval may be bounded or not; and functions that are continuous on an open interval also may or may not be bounded (again, the reciprocal function is an example of a function that is continuous on an open interval but not bounded); but here we will show that functions that are continuous on a closed interval must be bounded on that interval.

Proof. It may be tempting to begin trying to prove this by reasoning as follows.

> That f is continuous on $[a, b]$ means that, for any $c \in (a, b)$,
>
> $$\forall \epsilon > 0 \,.\, \exists \delta > 0 \,.\, |x - c| < \delta \implies |f(x) - f(c)| < \epsilon.$$
>
> This means that the value of $f(x)$ must be finite everywhere in (a, b) as, choosing any fixed point c in the open interval, $|x - c|$ is finite and, therefore, less than some δ thus implying that $|f(x) - f(c)| < \epsilon$ for some finite $\epsilon \ldots$

However this is **dead wrong!** If we take the example of the reciprocal function (40) over the interval $(0, 1]$: If we take x-values approaching 0, the value of $f(x)$ grows unbounded. For any given x-value it will be less than some finite ϵ but we can always find another x-value with a greater value of $f(x)$. So, there is no maximum ϵ and so, also no maximum value of $f(x)$ on the interval.

Another tempting way to prove this is as follows.

> A closed interval is an interval such that every sequence of values in the interval converges to a point in the interval.
> That f is continuous on $[a, b]$ implies that for every sequence, x_n, of values in $[a, b]$ that converges to some point in $c \in [a, b]$, $\lim_{n\to\infty} f(x_n) = f(c)$. Furthermore, that the interval is closed implies that every sequence of values in the interval converges to a point in the interval. Therefore,

> we can conclude that $f(x)$ at every point in $[a, b]$ exists and is finite so that f is bounded and obtains a maximum on the interval.

There are two issues with this:

1. The statement about the nature of a closed interval that the proof relies upon has not been proven.
2. That $f(x)$ is defined and finite for all $x \in [a, b]$ is taken as proof that the function is bounded and obtains a maximum in the interval.

To deal with issue (1) – if we're not going to prove the proposition about closed intervals as a pre-requisite of the proof – we need to develop a proof that doesn't rely on this characteristic of closed intervals. To deal with (2) meanwhile, we need to explicitly prove boundedness and that f obtains a maximum.
The proof given in LSE Abstract Mathematics course material follows.

> Suppose first that f is unbounded above. For each $n \in N$, let x_n be a point in $[a, b]$ such that $f(x_n) > n$. The sequence (x_n) is bounded, so has a convergent subsequence (x_{k_n}), tending to some limit c (by Theorem 10.11). Necessarily $c \in [a, b]$. Since f is continuous at c, $f(x_{k_n}) \to f(c)$ as $n \to \infty$. But this contradicts the construction of the sequence (x_n), since $f(x_{k_n}) > n \to \infty$. So f is bounded above. Let $M = \sup\{\, f(x) \mid x \in [a, b] \,\}$. For each $n \in N$, let x_n be a point in $[a, b]$ such that $f(x_n) > M - \frac{1}{n}$. Again take a convergent subsequence (x_{k_n}) of (x_n), tending to some limit $c \in [a, b]$. Arguing as before, we see $f(c) = M$.

This proof says: Assume that f is unbounded on the interval. Then we can construct a sequence of x-values such that, for the nth value x_n, $f(x_n) > n$. This is possible because, if f is unbounded, for any value of n, there is some subinterval of x-values such that for each of them $f(x) > n$ and any interval of the real numbers contains an infinite number of real numbers and so, a sequence x_n with $n \to \infty$. Note that, at this point, x_n is an arbitrary sequence which is not necessarily convergent (it could bounce around the interval).
Then, we notice that this sequence x_n is necessarily bounded (as it is a subinterval of $[a, b]$) and so we invoke the Bolzano-Weierstrass Theorem, Theorem 19 (called Theorem 10.11 in the quoted proof), to deduce that it has a

convergent subsequence which we call x_{k_n} and call its limit c.
At this point we can use the continuity of f on the interval to deduce that $f(x_{k_n}) \to f(c)$ as $n \to \infty$ by which we obtain a contradiction to the condition we set on the values of x_{k_n} when we constructed the sequence – namely that $f(x_{k_n}) > n$. Notice that this is constructing a sequence x_{k_n} such that $f(x_{k_n})$ grows without bound as $n \to \infty$ and then saying, "but the limit of x_{k_n} as $n \to \infty$ is c which is inside the interval and so (by continuity) the limit of $f(x_{k_n})$ as $n \to \infty$ is $f(c)$ – a fixed finite value". This is the point where the fact that the interval $[a, b]$ is closed comes into play – if the interval were not closed it would be possible that c was not inside the interval and then we would not be able to invoke continuity to assert that the limit of $f(x_{k_n})$ over this sequence was finite.
So, now we have shown that if a function is continuous over a closed interval then assuming that the function is unbounded produces a contradiction and, therefore, we can conclude that it is, in fact, bounded.
The last thing that needs to be proven is that f obtains its maximum in the interval. Having shown that the function is bounded on the interval we now know that there exists

$$M = \sup\{\, f(x) \mid x \in [a, b] \,\}$$

and we need to show that there is such an x-value in $[a, b]$ that $f(x) = M$. In an open interval this might not be the case as the supremum of the function on the interval might occur as the limit of f over a sequence of x-values converging to a point that lies outside the interval. So, in this case, we construct a convergent sequence such that $f(x_{k_n}) \to M$ as $n \to \infty$ by selecting x_n such that $f(x_n) = M - \frac{1}{n}$. This is possible because the definition of the supremum says that, because it is the *lowest* upper bound,

$$\forall \epsilon > 0 \;.\; \exists f(x_n) \;.\; f(x_n) > M - \epsilon$$

and so we are letting $\epsilon = \frac{1}{n}$. Then, as previously, we take a convergent subsequence of this sequence and name it x_{k_n}. So, as before, we have a sequence converging on some point, we'll call it $c \in [a, b]$, at which f is continuous so that $f(x_{k_n}) \to M$ as $n \to \infty$ implies that $M = f(c)$. This, in turn, means that f obtains a maximum in the interval. □

Intermediate Value Theorem

Theorem 28. *Let f be continuous on $[a, b]$ with $f(a) < f(b)$. Then, for all K s.t. $f(a) < K < f(b)$, there exists some $c \in (a, b)$ with $f(c) = K$.*

> *Note that this theorem is* ***not*** *written for an interval such that $f(a) \leq f(b)$ because if $f(a) = f(b)$ then the only K s.t. $f(a) \leq K \leq f(b)$ is $f(a) = K = f(b)$. But now it is* ***not*** *true to say that there exists some $c \in (a, b)$ with $f(c) = K$ as there is no reason why the value of $f(x)$ at the bounds of the interval should be repeated somewhere in the interior of the interval.*

We begin by proving a special case from which the general proof will follow.

Lemma 2. *Let f be continuous on $[a, b]$ with $f(a) < 0 < f(b)$. Then there exists some $c \in (a, b)$ with $f(c) = 0$.*

Proof. A first attempt at this proof is given below.

Continuity at the interval bounds a and b means that, for some $\epsilon_1, \epsilon_2 > 0$,

$$\exists \delta_1 \,.\, 0 \leq x - a < \delta_1 \implies |f(x) - f(a)| < \epsilon_1,$$

$$\exists \delta_2 \,.\, 0 \leq b - x < \delta_2 \implies |f(x) - f(b)| < \epsilon_2.$$

So, we have a lower neighbourhood around $f(a)$ and an upper neighbourhood around $f(b)$ as follows,

$$f(a) - \epsilon_1 < f(x) < f(a) + \epsilon_1,$$

$$f(b) - \epsilon_2 < f(x) < f(b) + \epsilon_2$$

If $\epsilon_1 > |f(a)|$ and $\epsilon_2 > |f(b)|$ then both neighbourhoods contain $f(x) = 0$. Therefore, there is some interval of x such that $f(x)$ lies inside both the lower and upper neighbourhood – in the overlap of the two. In the overlap we have,

$$f(b) - \epsilon_2 < f(x) < f(a) + \epsilon_1.$$

Now if we let ϵ_1 vary freely but make ϵ_2 a function of ϵ_1 like so,

$$\epsilon_2 = f(a) + f(b) + \epsilon_1$$

then we still have $\epsilon_1 > |f(a)|$ and $\epsilon_2 > |f(b)|$ as,

$$\epsilon_1 > |f(a)| \iff -\epsilon_1 < f(a) < \epsilon_1 \iff 0 < f(a) + \epsilon_1 < 2\epsilon_1$$

so $\epsilon_2 > f(b)$ and, conversely, if we assume that $\epsilon_2 > |f(b)|$ then,

$$\epsilon_2 > |f(b)| \iff -\epsilon_2 < f(b) < \epsilon_2 \iff -\epsilon_2 - f(b) < 0 < \epsilon_2 - f(b)$$

$$\epsilon_1 = \epsilon_2 - f(a) - f(b) > -f(a) = |f(a)| \qquad \text{since } f(a) < 0.$$

Therefore,

$$\epsilon_2 = f(a) + f(b) + \epsilon_1 \implies [\epsilon_1 > |f(a)| \iff \epsilon_2 > |f(b)|].$$

Now we have,

$$f(b) - \epsilon_2 = -f(a) - \epsilon_1 < f(x) < f(a) + \epsilon_1$$

which is equivalent to

$$|f(x)| < f(a) + \epsilon_1 = \epsilon_3.$$

Now, since $f(a) + \epsilon_1 > 0$ we also have $\epsilon_3 > 0$ and it can become arbitrarily small by making $|f(a)| - \epsilon_1$ arbitrarily small. So we have shown that continuity over the closed interval and $f(a) < f(b)$ imply that we can find subintervals of $[a, b]$ such that $f(x)$ is constricted to ever-decreasing neighbourhoods of 0. In other words, for some c s.t. $a < c < b$, $f(x) \to 0$ as $x \to c$ and since f is continuous on the interval this implies that $f(c) = 0$ also.

This is not bad but suffers from vagueness in a couple of areas: the overlap of the lower and upper neighbourhoods probably needs to be more precisely defined and, certainly, the final part of the proof stating that confining $f(x)$ to ever-decreasing neighbourhoods of 0 implies that there is some c such that $f(c) = 0$ needs to be drawn much more explicitly.

Here is the proof given in LSE Abstract Mathematics.

We construct a sequence of intervals $[a_n, b_n]$ such that

1. $f(a_n) < 0,\ f(b_n) > 0$ for each n
2. $[a_{n+1}, b_{n+1}] \subseteq [a_n, b_n]$ for each n.

We start by letting $[a_1, b_1] = [a, b]$. Then for each $n \geq 1$, we define $[a_{n+1}, b_{n+1}]$ as follows.

Let $c_n = (a_n + b_n)/2$, be the midpoint of the previous interval. If $f(c_n) = 0$, then the theorem is proved and so we need not continue constructing intervals!

Otherwise, if $f(c_n) < 0$, we define $a_{n+1} = c_n$ and $b_{n+1} = b_n$. And if $f(c_n) > 0$, we define $b_{n+1} = c_n$ and $a_{n+1} = a_n$. Note that the condition 1. is satisfied by choosing our intervals in this manner. Moreover, note that the $(n+1)$st interval is half the size of the nth interval and so $b_{n+1} - a_{n+1} \leq (b_1 - a_1)/2^n$. It follows that

$$\lim_{n\to\infty} (b_n - a_n) = 0. \tag{3}$$

Finally, note that (a_n) is increasing and bounded above (by b_1) and so it has a limit; similarly (b_n) is decreasing and bounded below and so has a limit. Thus by (3) (and algebra of limits) these limits are equal to, say, c. Thus by continuity (using Theorem 23),

$$f(c) = \lim_{n\to\infty} f(b_n) \geq 0,$$

where the last inequality follows from the fact that each $f(b_n) \geq 0$ (in fact > 0). Similarly,

$$f(c) = \lim_{n\to\infty} f(a_n) \leq 0.$$

Thus $f(c)$ must be equal to zero, and the proof is complete.

□

Clearly, the general proof of the Intermediate Value Theorem follows naturally from this because,

- If $f(a) > f(b)$ then we can consider the function $g(x) = -f(x)$ so that we have $g(a) < g(b)$,
- If we have K s.t. $f(a) < K < f(b)$ with $K \neq 0$ we can consider $g(x) = f(x) - K$ so that we have $g(a) < 0 < g(b)$.

So, the general problem posed in the Intermediate Value Theorem is reducible to the case we have proved.

Corollary 19. *Suppose that the real function f is continuous on the closed interval $[a, b]$ and that f maps $[a, b]$ into $[a, b]$. Then there is $c \in [a, b]$ with $f(c) = c$.*

Proof. Let $h(x) = f(x) - x$ so that $f(c) = c$ if and only if $h(c) = 0$. Then also we have,

$$a \leq f(x) \leq b \implies h(a) \geq 0,\ h(b) \leq 0.$$

But this means that, either one of $h(a)$ or $h(b)$ is equal to 0 or neither are. In the case that one of them is equal to 0 then we have found our c such that $f(c) = c$. Otherwise, if neither is equal to 0, then we must have $h(a) > 0$ and $h(b) < 0$. So, we may apply the Intermediate Value Theorem to conclude that there exists $c \in (a, b)$ such that $h(c) = 0$ which is to say $f(c) = c$. □

Examples of reasoning with the Intermediate Value Theorem

(43) *Suppose the real function f is continuous, positive and unbounded on $\mathbb{R}$ and that $\inf\{\, f(x) \mid x \in \mathbb{R} \,\} = 0$. Use the Intermediate Value Theorem to prove that the range of f is $(0, \infty)$.*

Let $y \in (0, 1)$. We show that there is some $c \in \mathbb{R}$ such that $f(c) = y$. This shows that the range is the whole of $(0, 1)$. (The fact that it is no larger follows from the given fact that f is positive.)
Now, $\inf f(\mathbb{R}) = \inf\{\, f(x) \mid x \in \mathbb{R} \,\} = 0$, so, since $y > 0$, there must be some $y_1 \in f(\mathbb{R})$ with $y_1 < y$. This means there is some $x_1 \in \mathbb{R}$ such that $y_1 = f(x_1) < y$.
Similarly, because f is unbounded, which means $f(\mathbb{R})$ is unbounded, there must be some $y_2 \in f(\mathbb{R})$ with $y_2 > y$ and there will be some $x_2 \in \mathbb{R}$ such that $y_2 = f(x_2) > y$.
Then y lies between $f(x_1)$ and $f(x_2)$ and, since f is continuous, the Intermediate Value Theorem shows that there is some c between x_1 and x_2 with $f(c) = y$.

(44) *Suppose the real function g is continuous on $\mathbb{R}$ and that g maps $[a, b]$ into $[d, e]$ and maps $[d, e]$ into $[a, b]$ where $a < b,\ d < e$. By considering the function*

$$k(x) = g(g(x)),$$

prove that there are $p, q \in \mathbb{R}$ *such that*

$$g(p) = q,\ g(q) = p.$$

Hence show that there is $c \in \mathbb{R}$ *such that* $g(c) = c$.

The function k, being a composition of continuous functions, is continuous and also maps $[a, b]$ into $[a, b]$ so that we can use Corollary 19 to deduce that there exists $c \in [a, b]$ such that $k(c) = c$. If we let $p = c$ and $q = g(p)$ then we have,

$$k(p) = p \iff g(g(p)) = p \iff g(q) = p.$$

Now we can employ the same trick again by defining

$$h(x) = g(x) - x$$

and then we have

$$h(p) = g(p) - p = q - p,\ h(q) = g(q) - q = p - q$$

so that $h(p) = -h(q)$ and, therefore, $h(x)$ changes sign between p and q. (Note that we don't know which of p and q is the lower end and upper end of the interval but we know that between the two values the function changes sign.) Then, applying the Intermediate Value Theorem we have some c between p and q such that,

$$h(c) = 0 \iff g(c) - c = 0 \iff g(c) = c.$$

3.2.8 Relationship Between Sequences and Functions

(tags: analysis)

Let $f : \mathbb{R} \mapsto \mathbb{R}$, $f(x) = y$. Now imagine that we take regular intervals on the domain, say, interval 1. We name the x-values at the upper bound of these intervals $x_1, x_2, \ldots$ for $x = 1, 2, \ldots$. Then, the corresponding y-values, $y = f(x_1), f(x_2), \ldots$ can be named $y_1, y_2, \ldots$. In this way, we have defined a sequence, $y_n = f(x_n)$ where $x_n \in \mathbb{N}$ (note that this is a different notation from that used before where a sequence $x_n = f(n)$ for $n \in \mathbb{N}$).
Looking at the derivative of the function,

$$\frac{dy}{dx} \approxeq \frac{\Delta y}{\Delta x} = \frac{f(x_{n+1}) - f(x_n)}{1} = f(x_{n+1}) - f(x_n).$$

While the ratio of consecutive terms is,

$$\frac{y_{n+1}}{y_n} = \frac{y_n + \Delta y}{y_n} = \frac{f(x_{n+1})}{f(x_n)} = \frac{f(x_n) + (f(x_{n+1}) - f(x_n))}{f(x_n)}.$$

Note, also, that

$$\frac{y_n + \Delta y}{y_n} = 1 + \frac{\Delta y}{y_n} \approxeq 1 + \frac{dy/dx}{y} = 1 + \frac{d}{dx} \ln y$$

which is to say that $\frac{\Delta y}{y_n}$ is the discrete form of the log derivative. Clearly, when $\Delta y > 0$, we can put this in the form

$$\frac{y_{n+1}}{y_n} = 1 + \frac{\Delta y}{y_n} = 1 + \frac{\Delta y / y_n}{1} = \frac{1 + h}{1}.$$

If $\Delta y < 0$ however,

$$\frac{y_n + \Delta y}{y_n} = \frac{1}{y_n/(y_n + \Delta y)} = \frac{1}{\frac{y_n + \Delta y - \Delta y}{y_n + \Delta y}} = \frac{1}{1 + \frac{-\Delta y}{y_n + \Delta y}} = \frac{1}{1 + h}.$$

If $\frac{\Delta y}{y_n}$ does not go to 0 then the ratio of consecutive terms stays below 1 and the sequence converges to 0. If it does go to 0 then the ratio of consecutive terms goes to 1 and the sequence may converge to 0 or to some non-zero value. These cases appear (proof?) to be distinguishable by looking at how fast $\frac{\Delta y}{y_n}$ goes to 0. For example,

(i) $\boldsymbol{a_n = \frac{1}{n} + 1}$

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

$$\frac{\Delta a}{a_n} = \frac{-1}{(n+1)^2}$$

(ii) $\boldsymbol{a_n = \frac{1}{n}}$

$$\lim_{n\to\infty} \frac{1}{n} = 0$$

$$\frac{\Delta a}{a_n} = \frac{-1}{n+1}$$

Maybe the fact that $\frac{\Delta y}{y_n}$ goes to 0 faster as $n \to \infty$ in the first case indicates that it converges before it gets to 0?