Chapter 0

Preliminaries

0.1 Sets and logic

0.1.1 First-order logic

Definition 0.1. A *set* $\mathcal{S}$ is a collection of *distinct* objects x's, often denoted with the following notation

$$\mathcal{S} = \{x \mid \text{ the conditions that } x \text{ satisfies. } \}. \tag{0.1}$$

Notation 1. $\mathbb{R}, \mathbb{Z}, \mathbb{N}, \mathbb{Q}, \mathbb{C}$ denote the sets of real numbers, integers, natural numbers, rational numbers and complex numbers, respectively. $\mathbb{R}^+, \mathbb{Z}^+, \mathbb{N}^+, \mathbb{Q}^+$ the sets of positive such numbers. In particular, $\mathbb{N}$ contains the number zero while $\mathbb{N}^+$ does not.

Definition 0.2. $\mathcal{S}$ is a *subset* of $\mathcal{U}$, written $\mathcal{S} \subseteq \mathcal{U}$, if and only if (iff) $x \in \mathcal{S} \Rightarrow x \in \mathcal{U}$. $\mathcal{S}$ is a *proper subset* of $\mathcal{U}$, written $\mathcal{S} \subset \mathcal{U}$, if $\mathcal{S} \subseteq \mathcal{U}$ and $\exists x \in \mathcal{U}$ s.t. $x \notin \mathcal{S}$.

Definition 0.3 (Statements of first-order logic)**.** A *universal statement* is a logical statement of the form

$$\mathsf{U} = (\forall x \in \mathcal{S},\ \mathsf{A}(x)). \tag{0.2}$$

An *existential statement* has the form

$$\mathsf{E} = (\exists x \in \mathcal{S},\ \text{s.t. } \mathsf{A}(x)), \tag{0.3}$$

where $\forall$ ("for each") and $\exists$ ("there exists") are the *quantifiers*, $\mathcal{S}$ is a set, "s.t." means "such that," and $\mathsf{A}(x)$ is the *formula*.
A statement of *implication/conditional* has the form

$$\mathsf{A} \Rightarrow \mathsf{B}. \tag{0.4}$$

Example 0.1. Universal and existential statements:
$\forall x \in [2, +\infty),\ x > 1$;
$\forall x \in \mathbb{R}^+,\ x > 1$;
$\exists p, q \in \mathbb{Z},\ \text{s.t. } p/q = \sqrt{2}$;
$\exists p, q \in \mathbb{Z},\ \text{s.t. } \sqrt{p} = \sqrt{q} + 1$.

Definition 0.4. *Uniqueness quantification* or *unique existential quantification*, written $\exists!$ or $\exists_{=1}$, indicates that exactly one object with a certain property exists.

Exercise 0.2. Express the logical statement $\exists! x,\ \text{s.t. } \mathsf{A}(x)$ with $\exists$, $\forall$, and $\Leftrightarrow$.

Definition 0.5. A *universal-existential statement* is a logical statement of the form

$$\mathsf{U}_E = (\forall x \in \mathcal{S},\ \exists y \in \mathcal{T} \text{ s.t. } \mathsf{A}(x, y)). \tag{0.5}$$

An *existential-universal statement* has the form

$$\mathsf{E}_U = (\exists y \in \mathcal{T},\ \text{s.t. } \forall x \in \mathcal{S},\ \mathsf{A}(x, y)). \tag{0.6}$$

Example 0.3. True or false:
$\forall x \in [2, +\infty),\ \exists y \in \mathbb{Z}^+ \text{ s.t. } x^y < 10^5$;
$\exists y \in \mathbb{R} \text{ s.t. } \forall x \in [2, +\infty),\ x > y$;
$\exists y \in \mathbb{R} \text{ s.t. } \forall x \in [2, +\infty),\ x < y$.

Example 0.4 (Translating an English statement into a logical statement)**.** Goldbach's conjecture states *every even natural number greater than 2 is the sum of two primes.* Let $\mathbb{P} \subset \mathbb{N}^+$ denote the set of prime numbers. Then Goldbach's conjecture is $\forall a \in 2\mathbb{N}^+ + 2, \exists p, q \in \mathbb{P}, \text{ s.t. } a = p + q$.

Theorem 0.6. The existential-universal statement implies the corresponding universal-existential statement, but not vice versa.

Example 0.5 (Translating a logical statement to an English statement)**.** Let $\mathcal{S}$ be the set of all human beings.
U_E =($\forall p \in \mathcal{S}, \exists q \in \mathcal{S}$ s.t. q is p's mom.)
E_U=($\exists q \in \mathcal{S}$ s.t. $\forall p \in \mathcal{S}$, q is p's mom.)
U_E is probably true, but E_U is certainly false.
If E_U were true, then U_E would be true. Why?

Axiom 0.7 (First-order negation of logical statements)**.** The negations of the statements in Definition 0.3 are

$$\neg\mathsf{U} = (\exists x \in \mathcal{S},\ \text{s.t. } \neg\mathsf{A}(x)). \tag{0.7}$$
$$\neg\mathsf{E} = (\forall x \in \mathcal{S},\ \neg\mathsf{A}(x)). \tag{0.8}$$

Rule 0.8. The negation of a more complicated logical statement abides by the following rules:

- switch the type of each quantifier until you reach the last formula without quantifiers;
- negate the last formula.

One might need to group quantifiers of like type.

Example 0.6 (The negation of Goldbach's conjecture)**.** $\exists a \in 2\mathbb{N}^+ + 2 \text{ s.t. } \forall p, q \in \mathbb{P},\ a \neq p + q$.

Exercise 0.7. Negate the logical statement in Definition 0.51.

Axiom 0.9 (Contraposition)**.** A conditional statement is logically equivalent to its contrapositive.

$$(\mathsf{A} \Rightarrow \mathsf{B}) \Leftrightarrow (\neg\mathsf{B} \Rightarrow \neg\mathsf{A}) \tag{0.9}$$

Example 0.8. "If Jack is a man, then Jack is a human being." is equivalent to "If Jack is not a human being, then Jack is not a man."

Exercise 0.9. Draw an Euler diagram of subsets to illustrate Example 0.8.

Exercise 0.10. Rewrite each of the following statements and its *negation* into *logical statements* using symbols, quantifiers, and formulas.

(a) The only even prime is 2.

(b) Multiplication of integers is associative.

(c) Goldbach's conjecture has at most a finite number of counterexamples.

0.1.2 Ordered sets

Definition 0.10. The *Cartesian product* $\mathcal{X} \times \mathcal{Y}$ between two sets $\mathcal{X}$ and $\mathcal{Y}$ is the set of all possible ordered pairs with first element from $\mathcal{X}$ and second element from $\mathcal{Y}$:

$$\mathcal{X} \times \mathcal{Y} = \{(x, y) \mid x \in \mathcal{X},\ y \in \mathcal{Y}\}. \tag{0.10}$$

Axiom 0.11 (Fundamental principle of counting)**.** A task consists of a sequence of k independent steps. Let n_i denote the number of different choices for the i-th step, the total number of distinct ways to complete the task is then

$$\prod_{i=1}^{k} n_i = n_1 n_2 \cdots n_k. \tag{0.11}$$

Example 0.11. Let A, E, D be the set of appetizers, main entrees, desserts in a restaurant. $A \times E \times D$ is the set of possible dinner combos. If $\#A = 10$, $\#E = 5$, $\#D = 6$, $\#(A \times E \times D) = 300$.

Definition 0.12 (Maximum and minimum)**.** Consider $\mathcal{S} \subseteq \mathbb{R}$, $\mathcal{S} \neq \emptyset$. If $\exists s_m \in \mathcal{S}$ s.t. $\forall x \in \mathcal{S}$, $x \leq s_m$, then s_m is the *maximum* of $\mathcal{S}$ and denoted by $\max \mathcal{S}$. If $\exists s_m \in \mathcal{S}$ s.t. $\forall x \in \mathcal{S}$, $x \geq s_m$, then s_m is the *minimum* of $\mathcal{S}$ and denoted by $\min \mathcal{S}$.

Definition 0.13 (Upper and lower bounds)**.** Consider $\mathcal{S} \subseteq \mathbb{R}$, $\mathcal{S} \neq \emptyset$. a is an *upper bound* of $\mathcal{S} \subseteq \mathbb{R}$ if $\forall x \in \mathcal{S}$, $x \leq a$; then the set $\mathcal{S}$ is said to be *bounded above*. a is a *lower bound* of $\mathcal{S}$ if $\forall x \in \mathcal{S}$, $x \geq a$; then the set $\mathcal{S}$ is said to be *bounded below*. $\mathcal{S}$ is *bounded* if it is bounded above and bounded below.

Definition 0.14 (Supremum and infimum)**.** Consider a nonempty set $\mathcal{S} \subseteq \mathbb{R}$. If $\mathcal{S}$ is bounded above and $\mathcal{S}$ has a least upper bound then we call it the *supremum* of $\mathcal{S}$ and denote it by $\sup \mathcal{S}$. If $\mathcal{S}$ is bounded below and $\mathcal{S}$ has a greatest lower bound, then we call it the *infimum* of $\mathcal{S}$ and denote it by $\inf \mathcal{S}$.

Example 0.12. If a set $\mathcal{S} \subset \mathbb{R}$ has a maximum, we have $\max \mathcal{S} = \sup \mathcal{S}$.

Example 0.13. $\sup[a, b] = \sup[a, b) = \sup(a, b] = \sup(a, b)$.

Axiom 0.15 (Completeness of $\mathbb{R}$)**.** Every nonempty subset of $\mathbb{R}$ that is bounded above has a least upper bound.

Corollary 0.16. Every nonempty subset of $\mathbb{R}$ that is bounded below has a greatest lower bound.

Definition 0.17. A *binary relation between two sets* $\mathcal{X}$ and $\mathcal{Y}$ is an ordered triple $(\mathcal{X}, \mathcal{Y}, \mathcal{G})$ where $\mathcal{G} \subseteq \mathcal{X} \times \mathcal{Y}$.
A *binary relation on* $\mathcal{X}$ is the relation between $\mathcal{X}$ and $\mathcal{X}$.
The statement $(x, y) \in R$ is read "x is R-related to y," and denoted by xRy or $R(x, y)$.

Definition 0.18. A binary relation "$\leq$" on some set $\mathcal{S}$ is a *total order* or *linear order* on $\mathcal{S}$ iff, $\forall a, b, c \in \mathcal{S}$,

- $a \leq b$ and $b \leq a$ imply $a = b$ (antisymmetry);
- $a \leq b$ and $b \leq c$ imply $a \leq c$ (transitivity);
- $a \leq b$ or $b \leq a$ (totality).

A set equipped with a total order is a *chain* or *totally ordered set*.

Example 0.14. The real numbers with less or equal.

Example 0.15. The English letters of the alphabet with dictionary order.

Example 0.16. The Cartesian product of a set of totally ordered sets with the *lexicographical order*.

Example 0.17. Sort your book in lexicographical order and save a lot of time. $\log_{26} N \ll N!$

Definition 0.19. A binary relation "$\leq$" on some set $\mathcal{S}$ is a *partial order* on $\mathcal{S}$ iff, $\forall a, b, c \in \mathcal{S}$, antisymmetry, transitivity, and reflexivity ($a \leq a$) hold.
A set equipped with a partial order is a called a *poset*.

Example 0.18. The set of subsets of a set $\mathcal{S}$ ordered by inclusion "$\subseteq$."

Example 0.19. The natural numbers equipped with the relation of divisibility.

Example 0.20. The set of stuff you will put on your body every morning with the time ordered: undershorts, pants, belt, shirt, tie, jacket, socks, shoes, watch.

Example 0.21. Inheritance ("is-a" relation) is a partial order. $A \to B$ reads "B is a special type of A".

Example 0.22. Composition ("has-a" relation) is also a partial order. $A \rightsquigarrow B$ reads "B *has an* instance/object of A."

Example 0.23. Implication "$\Rightarrow$" is a partial order on the set of logical statements.

Example 0.24. The set of definitions, axioms, propositions, theorems, lemmas, etc., is a poset with inheritance, composition, and implication. It is helpful to relate them with these partial orderings.

0.2 Basic analysis

0.2.1 Limits and continuity

Definition 0.20. A *function/map/mapping* f from $\mathcal{X}$ to $\mathcal{Y}$, written $f : \mathcal{X} \to \mathcal{Y}$ or $\mathcal{X} \mapsto \mathcal{Y}$, is a subset of the Cartesian product $\mathcal{X} \times \mathcal{Y}$ satisfying that $\forall x \in \mathcal{X}$, there is exactly one $y \in \mathcal{Y}$ s.t. $(x, y) \in \mathcal{X} \times \mathcal{Y}$. $\mathcal{X}$ and $\mathcal{Y}$ are the *domain* and *range* of f, respectively.

Definition 0.21. A function $f : \mathcal{X} \to \mathcal{Y}$ is said to be *injective* or *one-to-one* iff

$$\forall x_1 \in \mathcal{X}, \forall x_2 \in \mathcal{X}, \quad x_1 \neq x_2 \;\Rightarrow\; f(x_1) \neq f(x_2). \tag{0.12}$$

It is *surjective* or *onto* iff

$$\forall y \in \mathcal{Y},\ \exists x \in \mathcal{X},\ \text{s.t. } y = f(x). \tag{0.13}$$

It is *bijective* iff it is both injective and surjective.

Definition 0.22. A set $\mathcal{S}$ is *countably infinite* iff there exists a bijective function $f : \mathcal{S} \to \mathbb{N}^+$ that maps $\mathcal{S}$ to $\mathbb{N}^+$. A set is *countable* if it is either finite or countably infinite.

Example 0.25. Are the integers countable? Are the rationals countable? Are the real numbers countable?

Definition 0.23. A *scalar function* is a function whose range is a subset of $\mathbb{R}$.

Definition 0.24 (Limit of a scalar function with one variable)**.** Consider a function $f : I \to \mathbb{R}$ with $I(c, r) = (c-r, c) \cup (c, c+r)$. The *limit* of $f(x)$ exists as x approaches c, written $\lim_{x\to c} f(x) = L$, iff

$$\forall \epsilon > 0, \exists \delta > 0,\ \text{s.t. } \forall x \in I(c, \delta),\ |f(x) - L| < \epsilon. \tag{0.14}$$

Example 0.26. Show that $\lim_{x\to 2} \frac{1}{x} = \frac{1}{2}$.

Proof. If $\epsilon \geq \frac{1}{2}$, choose $\delta = 1$. Then $x \in (1, 3)$ implies $\left|\frac{1}{x} - \frac{1}{2}\right| < \frac{1}{2}$ since $\frac{1}{x} - \frac{1}{2}$ is a monotonically decreasing function with its supremum at $x = 1$.

If $\epsilon \in (0, \frac{1}{2})$, choose $\delta = \epsilon$. Then $x \in (2-\epsilon, 2+\epsilon) \subset (\frac{3}{2}, \frac{5}{2})$. Hence $\left|\frac{1}{x} - \frac{1}{2}\right| = \frac{|2-x|}{|2x|} < |2-x| < \epsilon$. The proof is completed by Definition 0.24. □

Definition 0.25. $f : \mathbb{R} \to \mathbb{R}$ is *continuous* at c iff

$$\lim_{x\to c} f(x) = f(c). \tag{0.15}$$

f is *continuous on* (a, b), written $f \in \mathcal{C}(a, b)$ if (0.15) holds $\forall x \in (a, b)$.

Definition 0.26. Let $I = (a, b)$. A function $f : I \to \mathbb{R}$ is *uniformly continuous* on I iff

$$\begin{aligned} &\forall \epsilon > 0, \exists \delta > 0,\ \text{s.t.} \\ &\forall x, y \in I,\ |x - y| < \delta \Rightarrow |f(x) - f(y)| < \epsilon. \end{aligned} \tag{0.16}$$

Example 0.27. Show that, on (a, ∞), $f(x) = \frac{1}{x}$ is uniformly continuous if $a > 0$ and is not so if $a = 0$.

Proof. If $a > 0$, then $|f(x) - f(y)| = \frac{|x-y|}{xy} < \frac{|x-y|}{a^2}$.
Hence $\forall \epsilon > 0, \exists \delta = a^2\epsilon$, s.t.
$|x - y| < \delta \Rightarrow |f(x) - f(y)| < \frac{|x-y|}{a^2} < \frac{a^2\epsilon}{a^2} = \epsilon$.

If $a = 0$, negating the condition of uniform continuity, i.e. eq. (0.16), yields $\exists \epsilon > 0$ s.t. $\forall \delta > 0\ \exists x, y > 0$ s.t. $|x - y| < \delta \Rightarrow |f(x) - f(y)| \geq \epsilon$.

We prove a stronger version: $\forall \epsilon > 0$, $\forall \delta > 0\ \exists x, y > 0$ s.t. $|x - y| < \delta \Rightarrow |\frac{1}{x} - \frac{1}{y}| \geq \epsilon$.

If $\delta \geq \frac{1}{2\epsilon}$, choose $x = \frac{1}{2\epsilon}$, $y = \frac{1}{4\epsilon}$. This choice satisfies $|x - y| < \delta$ since $x - y = \frac{1}{4\epsilon} < \frac{1}{2\epsilon} \leq \delta$. However, $|f(x) - f(y)| = \frac{|x-y|}{xy} = 2\epsilon > \epsilon$.

If $\delta < \frac{1}{2\epsilon}$, then $2\epsilon\delta < 1$. Choose $x \in (0, \epsilon\delta^2)$ and $y \in (2\epsilon\delta^2, \delta)$. This choice satisfies $|x - y| < \delta$ and $|x - y| > \epsilon\delta^2$. However, $|f(x) - f(y)| = \frac{|x-y|}{xy} > \frac{\epsilon\delta^2}{xy} > \frac{1}{y} > \frac{1}{\delta} > 2\epsilon > \epsilon$. □

Exercise 0.28. On (a, ∞), $f(x) = \frac{1}{x^2}$ is uniformly continuous if $a > 0$ and is not so if $a = 0$.

Theorem 0.27. Uniform continuity implies continuity but the converse is not true.

Proof. exercise. □

Theorem 0.28. $f : \mathbb{R} \to \mathbb{R}$ is *uniformly continuous* on (a, b) iff it can be extended to a continuous function $\tilde{f}$ on $[a, b]$.

Definition 0.29. The *derivative* of a function $f : \mathbb{R} \to \mathbb{R}$ at a is the limit

$$f'(a) = \lim_{h\to 0} \frac{f(a + h) - f(a)}{h}. \tag{0.17}$$

If the limit exists, f is *differentiable* at a.

Example 0.29. For the power function $f(x) = x^\alpha$, we have $f' = \alpha x^{\alpha-1}$ due to Newton's generalized binomial theorem,

$$(a + h)^\alpha = \sum_{n=0}^{\infty} \binom{\alpha}{n} a^{\alpha-n} h^n.$$

Definition 0.30. A function $f(x)$ is k times *continuously differentiable* on (a, b) iff $f^{(k)}(x)$ exists on (a, b) and is itself continuous. The set or space of all such functions on (a, b) is denoted by $\mathcal{C}^k(a, b)$. In comparison, $\mathcal{C}^k[a, b]$ is the space of functions f for which $f^{(k)}(x)$ is bounded and uniformly continuous on (a, b).

Theorem 0.31. A scalar function f is bounded on $[a, b]$ if $f \in \mathcal{C}[a, b]$.

Theorem 0.32 (Intermediate value)**.** A scalar function $f \in \mathcal{C}[a, b]$ satisfies

$$\forall y \in [m, M]\,,\ \exists \xi \in [a, b],\ \text{s.t. } y = f(\xi) \tag{0.18}$$

where $m = \inf_{x\in[a,b]} f(x)$ and $M = \sup_{x\in[a,b]} f(x)$.

Theorem 0.33. If $f : (a, b) \to \mathbb{R}$ assumes its maximum or minimum at $x_0 \in (a, b)$ and f is differentiable at x_0, then $f'(x_0) = 0$.

Proof. Suppose $f'(x_0) > 0$. Then we have

$$f'(x_0) = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} > 0.$$

The definition of a limit implies

$$\exists \delta > 0 \text{ s.t. } a < x_0 - \delta < x_0 + \delta < b,$$

which, together with $|x - x_0| < \delta$, implies $\frac{f(x)-f(x_0)}{x-x_0} > 0$. This is a contradiction to $f(x_0)$ being a maximum when we choose $x \in (x_0, x_0 + \delta)$. □

Theorem 0.34 (Rolle's)**.** If a function $f : \mathbb{R} \to \mathbb{R}$ satisfies

(i) $f \in \mathcal{C}[a, b]$ and f' exists on (a, b),

(ii) $f(a) = f(b)$,

then $\exists x \in (a, b)$ s.t. $f'(x) = 0$.

Proof. By Theorem 0.32, all values between $\sup f$ and $\inf f$ will be assumed. If $f(a) = f(b) = \sup f = \inf f$, then f is a constant on $[a, b]$ and thus the conclusion holds. Otherwise, Theorem 0.33 completes the proof. □

Theorem 0.35 (Mean value)**.** If $f \in \mathcal{C}[a, b]$ and if f' exists on (a, b), then $\exists \xi \in (a, b)$ s.t. $f(b) - f(a) = f'(\xi)(b - a)$.

Proof. Construct a linear function $L : [a, b] \to \mathbb{R}$ such that $L(a) = f(a)$, $L(b) = f(b)$, then $\forall x \in (a, b)$, we have $L'(x) = \frac{f(b)-f(a)}{b-a}$. Consider $g(x) = f(x) - L(x)$ on $[a, b]$. $g(a) = 0$, $g(b) = 0$. By Theorem 0.34, $\exists \xi \in [a, b]$ such that $g'(\xi) = 0$, which completes the proof. □

0.2.2 Taylor series

Definition 0.36. A *sequence* is a map on $\mathbb{N}^+$ or $\mathbb{N}$.

Example 0.30. Whether the sequence starts from 0 or 1 is a matter of convention and convenience according to the context.

Definition 0.37 (Limit of a sequence)**.** A sequence $\{a_n\}$ has the *limit* L, written $\lim_{n\to\infty} a_n = L$, or $a_n \to L$ as $n \to \infty$, iff

$$\forall \epsilon > 0, \ \exists N, \text{ s.t. } \ \forall n > N, \ |a_n - L| < \epsilon. \tag{0.19}$$

If such a limit L exists, we say that $\{a_n\}$ *converges* to L.

Example 0.31 (A story of π)**.** A famous estimation of π in ancient China is given by Zu, ChongZhi 1500 years ago,

$$\pi \approx \frac{355}{113} \approx 3.14159292.$$

In modern mathematics, we approximate π with a sequence for increasing accuracy, e.g.

$$\pi \approx 3.141592653589793\ldots \tag{0.20}$$

As of March 2019, we human beings have more than 31 trillion digits of π. However, real world applications never use even a small fraction of the 31 trillion digits:

- If you want to build a fence over your backyard swimming pool, several digits of π is probably enough;
- in NASA, calculations involving π use 15 digits for Guidance Navigation and Control;
- if you want to compute the circumference of the entire universe to the accuracy of less than the diameter of a hydrogen atom, you need only 39 decimal places of π.

On one hand, computational mathematics is judged by a metric that is different from that of pure mathematics; this may cause a huge gap between what needs to be done and what has been done. On the other hand, a computational mathematician cannot assume that a fixed accuracy is good enough for all applications. In the approximation a number or a function, she must develop theory and algorithms to provide the user the choice of an ever-increasing amount of accuracy, so long as the user is willing to invest an increasing amount of computational resources. This is one of the main motivations of infinite sequence and series.

Theorem 0.38 (Bolzano-Weierstrass)**.** Every bounded sequence has a convergent subsequence.

Definition 0.39. A *series* associated with an infinite sequence $\{a_n\}$ is defined as $\sum_{i=0}^{\infty} a_n$, the sum of all terms of the sequence.

Definition 0.40. The *sequence of partial sums* S_n associated to a series $\sum_{i=0}^{\infty} a_i$ is defined for each n as the sum of the sequence $\{a_i\}$ from a_0 to a_n

$$S_n = \sum_{i=0}^{n} a_i. \tag{0.21}$$

Lemma 0.41. A series converges to L iff the associated sequence of partial sums converges to L.

Definition 0.42. A *power series* centered at c is a series of the form

$$p(x) = \sum_{n=0}^{\infty} a_n (x - c)^n, \tag{0.22}$$

where a_n's are the *coefficients*. The *interval of convergence* is the set of values of x for which the series converges:

$$I_c(p) = \{x \mid p(x) \text{ converges}\}. \tag{0.23}$$

Definition 0.43. If the derivatives $f^{(i)}(x)$ with $i = 1, 2, \ldots, n$ exist for a function $f : \mathbb{R} \to \mathbb{R}$ at $x = c$, then

$$T_n(x) = \sum_{k=0}^{n} \frac{f^{(k)}(c)}{k!}(x - c)^k \tag{0.24}$$

is called the nth *Taylor polynomial* for $f(x)$ at c.
In particular, the *linear approximation* for $f(x)$ at c is

$$T_1(x) = f(c) + f'(c)(x - c). \tag{0.25}$$

Example 0.32. If $f \in \mathcal{C}^\infty$, then $\forall n \in \mathbb{N}$, we have

$$T_n^{(m)}(x) = \begin{cases} \sum_{k=m}^{n} \frac{f^{(k)}(c)}{(k-m)!}(x - c)^{k-m}, & m \in \mathbb{N}, m \le n; \\ 0, & m \in \mathbb{N}, m > n. \end{cases}$$

This can be proved by induction. In the inductive step, we regroup the summation into a constant term and another shifted summation.

Definition 0.44. The *Taylor series* (or Taylor expansion) for $f(x)$ at c is

$$\sum_{k=0}^{\infty} \frac{f^{(k)}(c)}{k!}(x-c)^k. \tag{0.26}$$

Definition 0.45. The *remainder* of the nth *Taylor polynomial* in approximating $f(x)$ is

$$E_n(x) = f(x) - T_n(x). \tag{0.27}$$

Theorem 0.46. Let T_n be the nth Taylor polynomial for $f(x)$ at c.

$$\lim_{n\to\infty} E_n(x) = 0 \;\Leftrightarrow\; \lim_{n\to\infty} T_n(x) = f(x). \tag{0.28}$$

Lemma 0.47. $\forall m = 0, 1, 2, \ldots, n$, $E_n^{(m)}(c) = 0$.

Proof. This follows from Definition 0.43 and Example 0.32. □

Theorem 0.48 (Taylor's theorem with Lagrangian form)**.** Consider a function $f : \mathbb{R} \to \mathbb{R}$. If $f \in \mathcal{C}^n[c-d, c+d]$ and $f^{(n+1)}(x)$ exists on $(c-d, c+d)$, then $\forall x \in [c-d, c+d]$, there exists some ξ between c and x such that

$$E_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!}(x-c)^{n+1}. \tag{0.29}$$

Proof. Fix $x \neq c$, let M be the unique solution of

$$E_n(x) = f(x) - T_n(x) = \frac{M(x-c)^{n+1}}{(n+1)!}.$$

Consider the function

$$g(t) := E_n(t) - \frac{M(t-c)^{n+1}}{(n+1)!}. \tag{0.30}$$

Clearly $g(x) = 0$. By Lemma 0.47, $g^{(k)}(c) = 0$ for each $k = 0, 1, \ldots, n$. Then Rolle's theorem implies that

$$\exists x_1 \in (c, x) \text{ s.t. } g'(x_1) = 0.$$

If $x < c$, change (c, x) above to (x, c). Apply Rolle's theorem to $g'(t)$ on (c, x_1) and we have

$$\exists x_2 \in (c, x_1) \text{ s.t. } g^{(2)}(x_2) = 0.$$

Repeatedly using Rolle's theorem,

$$\exists x_{n+1} \in (c, x_n) \text{ s.t. } g^{(n+1)}(x_{n+1}) = 0. \tag{0.31}$$

Since T_n is a polynomial of degree n, we have $T_n^{(n+1)}(t) = 0$, which, together with (0.31) and (0.30), yields

$$f^{(n+1)}(x_{n+1}) - M = 0.$$

The proof is completed by identifying ξ with x_{n+1}. □

Example 0.33. How many terms are needed to compute e^2 correctly to four decimal places?

The requirement of four decimal places means an accuracy of at least $\epsilon = 10^{-5}$. By Definition 0.44, the Taylor series of e^x at $c = 0$ is

$$e^x = \sum_{n=0}^{+\infty} \frac{x^n}{n!}.$$

By Theorem 0.48, we have

$$\exists \xi \in [0, 2] \text{ s.t. } E_n(2) = e^{\xi} 2^{n+1}/(n+1)! < e^2 2^{n+1}/(n+1)!$$

Then $e^2 2^{n+1}/(n+1)! \leq \epsilon$ yields $n \geq 12$, i.e., 13 terms.

0.2.3 Riemann integral

Definition 0.49. A *partition of an interval* $I = [a, b]$ is a finite ordered subset $T_n \subseteq I$ of the form

$$T_n(a, b) = \{a = x_0 < x_1 < \cdots < x_n = b\}. \tag{0.32}$$

The interval $I_i = [x_{i-1}, x_i]$ is the ith *subinterval* of the partition. The *norm* of the partition is the length of the longest subinterval,

$$h_n = h(T_n) = \max(x_i - x_{i-1}), \qquad i = 1, 2, \ldots, n. \tag{0.33}$$

Definition 0.50. The *Riemann sum* of $f : \mathbb{R} \to \mathbb{R}$ over a partition T_n is

$$S_n(f) = \sum_{i=1}^{n} f(x_i^*)(x_i - x_{i-1}), \tag{0.34}$$

where $x_i^* \in I_i$ is a *sample point* of the ith subinterval.

Definition 0.51. $f : \mathbb{R} \to \mathbb{R}$ is *integrable* (or more precisely *Riemann integrable*) on $[a, b]$ iff

$$\begin{aligned} &\exists L \in \mathbb{R}, \text{ s.t. } \forall \epsilon > 0, \ \exists \delta > 0 \text{ s.t.} \\ &\forall T_n(a, b) \text{ with } h(T_n) < \delta, \ |S_n(f) - L| < \epsilon. \end{aligned} \tag{0.35}$$

Example 0.34. The following function $f : [a, b] \to \mathbb{R}$ is not Riemann integrable.

$$f(x) = \begin{cases} 1 & x \text{ is rational;} \\ 0 & x \text{ is irrational.} \end{cases}$$

To see this, we first negate the logical statement in (0.35) to get

$$\begin{aligned} &\forall L \in \mathbb{R}, \exists \epsilon > 0, \text{ s.t. } \forall \delta > 0 \\ &\exists T_n(a, b) \text{ with } h(T_n) < \delta, \ \text{ s.t. } |S_n(f) - L| \geq \epsilon. \end{aligned}$$

If $|L| < \frac{b-a}{2}$, we choose all x_i^*'s to be rational so that $f(x_i^*) \equiv 1$; then (0.34) yields $S_n(f) = b - a$. For $\epsilon = \frac{b-a}{4}$, the formula $|S_n(f) - L| \geq \epsilon$ clearly holds.

If $|L| \geq \frac{b-a}{2}$, we choose all x_i^*'s to be irrational so that $f(x_i^*) \equiv 0$; then (0.34) yields $S_n(f) = 0$. For $\epsilon = \frac{b-a}{4}$, the formula $|S_n(f) - L| \geq \epsilon$ clearly holds.

Definition 0.52. If $f : \mathbb{R} \to \mathbb{R}$ is integrable on $[a, b]$, then the limit of the Riemann sum of f is called the *definite integral* of f on $[a, b]$:

$$\int_a^b f(x)\mathrm{d}x = \lim_{h_n \to 0} S_n(f). \tag{0.36}$$

Theorem 0.53. A scalar function f is integrable on $[a, b]$ if $f \in \mathcal{C}[a, b]$.

Definition 0.54. A *monotonic* function is a function between ordered sets that either preserves or reverses the given order. In particular, $f : \mathbb{R} \to \mathbb{R}$ is *monotonically increasing* if $\forall x, y,\ x \le y \Rightarrow f(x) \le f(y)$; $f : \mathbb{R} \to \mathbb{R}$ is *monotonically decreasing* if $\forall x, y,\ x \le y \Rightarrow f(x) \ge f(y)$.

Theorem 0.55. A scalar function is integrable on $[a, b]$ if it is monotonic on $[a, b]$.

Exercise 0.35. True or false: a bijective function is either order-preserving or order-reversing?

Theorem 0.56 (Integral mean value)**.** Let $w : [a, b] \to \mathbb{R}^+$ be integrable on $[a, b]$. For $f \in \mathcal{C}[a, b]$, $\exists \xi \in [a, b]$ s.t.

$$\int_a^b w(x)f(x)\mathrm{d}x = f(\xi)\int_a^b w(x)\mathrm{d}x. \tag{0.37}$$

Proof. Denote $m = \inf_{x\in[a,b]} f(x)$, $M = \sup_{x\in[a,b]} f(x)$, and $I = \int_a^b w(x)\mathrm{d}x$. Then $mw(x) \le f(x)w(x) \le Mw(x)$ and

$$mI \le \int_a^b w(x)f(x)\mathrm{d}x \le MI.$$

$w > 0$ implies $I \neq 0$, hence

$$m \le \frac{1}{I}\int_a^b w(x)f(x)\mathrm{d}x \le M.$$

Applying Theorem 0.32 completes the proof. □

0.2.4 Uniform convergence in metric spaces

Definition 0.57. A *metric* is a function $d : \mathcal{X} \times \mathcal{X} \to [0, +\infty)$ that satisfies, for all $x, y, z \in \mathcal{X}$,

(1) non-negativity: $d(x, y) \ge 0$;

(2) identity of indiscernibles: $x = y \Leftrightarrow d(x, y) = 0$;

(3) symmetry: $d(x, y) = d(y, x)$;

(4) triangle inequality: $d(x, z) \le d(x, y) + d(y, z)$.

A *metric space* is an ordered pair $(\mathcal{X}, d)$ where $\mathcal{X}$ is a set and d is a metric on $\mathcal{X}$.

Example 0.36. Set $\mathcal{X}$ to be $C[a, b]$, the set of continuous functions $[a, b] \to \mathbb{R}$. Then the following is a metric on $\mathcal{X}$,

$$d(x, y) = \max_{t\in[a,b]} |x(t) - y(t)|. \tag{0.38}$$

Definition 0.58. The *sequence space* ℓ^∞ is a metric space $(\mathcal{X}, d)$, where $\mathcal{X}$ is the set of all bounded sequences of complex numbers,

$$\forall x = (\xi_1, \xi_2, \ldots) \in \mathcal{X}, \exists c_x \in \mathbb{R},\ \text{s.t. } \forall i = 1, 2, \ldots,\ |\xi_i| \le c_x,$$

and the metric is given by

$$d(x, y) = \sup_{i\in\mathbb{N}^+} |\xi_i - \eta_i|$$

where $y = (\eta_1, \eta_2, \ldots) \in \mathcal{X}$.

Exercise 0.37. Let $\mathcal{X}$ be the set of all bounded and unbounded sequences of complex numbers. Show that the following is a metric on $\mathcal{X}$,

$$d(x, y) = \sum_{j=1}^{\infty} \frac{1}{2^j}\frac{|\xi_j - \eta_j|}{1 + |\xi_j - \eta_j|}, \tag{0.39}$$

where $x = (\xi_j)$ and $y = (\eta_j)$.

Definition 0.59. For a real number $p \ge 1$, the ℓ^p *space* is the metric space $(\mathcal{X}, d)$ with

$$\mathcal{X} = \left\{(\xi_j)_{j=1}^{\infty} : \xi_j \in \mathbb{C}; \sum_{j=1}^{\infty} |\xi_j|^p < \infty\right\}; \tag{0.40}$$

$$d(x, y) = \left(\sum_{j=1}^{\infty} |\xi_j - \eta_j|^p\right)^{1/p}, \tag{0.41}$$

where $x = (\xi_j)$ and $y = (\eta_j)$ are both in $\mathcal{X}$. In particular, the *Hilbert sequence space* ℓ^2 is the ℓ^p space with $p = 2$.

Definition 0.60. Two positive real numbers p, q are called *conjugate exponents* iff they satisfy

$$p > 1, \qquad \frac{1}{p} + \frac{1}{q} = 1. \tag{0.42}$$

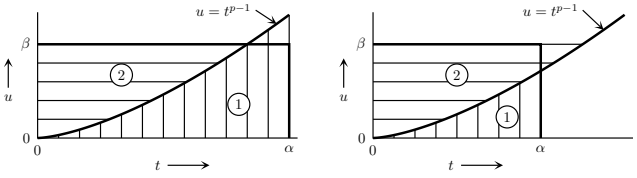
Lemma 0.61. Any two positive real numbers α, β satisfy

$$\alpha\beta \le \frac{\alpha^p}{p} + \frac{\beta^q}{q}, \tag{0.43}$$

where p and q are conjugate exponents.

Proof. By (0.42), we have

$$u = t^{p-1} \ \Rightarrow\ t = u^{q-1}.$$



It follows that

$$\alpha\beta \le \int_0^\alpha t^{p-1}\mathrm{d}t + \int_0^\beta u^{q-1}\mathrm{d}u = \frac{\alpha^p}{p} + \frac{\beta^q}{q},$$

where the equality holds if $\alpha = 0$ and $\beta = 0$. □

Exercise 0.38. Prove that (0.41) is indeed a metric. In particular, prove that (0.41) satisfies the triangular inequality by showing

(a) Lemma 0.61 implies the *Hölder inequality*, i.e., for conjugate exponents p, q and for any $(\xi_j) \in \ell^p$, $(\eta_j) \in \ell^q$,

$$\sum_{j=1}^{\infty} |\xi_j\eta_j| \le \left(\sum_{k=1}^{\infty} |\xi_k|^p\right)^{1/p}\left(\sum_{m=1}^{\infty} |\eta_m|^q\right)^{1/q}. \tag{0.44}$$

(b) The Hölder inequality implies the *Minkowski inequality*, i.e. for any $p \ge 1$, $(\xi_j) \in \ell^p$, and $(\eta_j) \in \ell^p$,

$$\left(\sum_{j=1}^{\infty} |\xi_j + \eta_j|^p\right)^{1/p} \le \left(\sum_{k=1}^{\infty} |\xi_k|^p\right)^{1/p} + \left(\sum_{m=1}^{\infty} |\eta_m|^p\right)^{1/p}. \tag{0.45}$$

(c) The Minkowski inequality implies that the triangular inequality holds for (0.41).

Definition 0.62. In a metric space $(\mathcal{X}, d)$, an *open ball* $B_r(x)$ centered at $x \in \mathcal{X}$ with radius r is the subset

$$B_r(x) := \{y \in \mathcal{X} : d(x, y) < r\}. \tag{0.46}$$

Definition 0.63. Let $(\mathcal{X}, d)$ be a metric space. A point $x_0 \in \mathcal{X}$ is an *adherent point* or a *closure point* of $E \subset \mathcal{X}$ or a *point of closure* or a *contact point* iff

$$\forall r > 0,\ E \cap B_r(x_0) \neq \emptyset. \tag{0.47}$$

Definition 0.64 (Limiting value of a function)**.** Let $(\mathcal{X}, d_{\mathcal{X}})$ and $(\mathcal{Y}, d_{\mathcal{Y}})$ be metric spaces. Let E be a subset of $\mathcal{X}$ and $x_0 \in \mathcal{X}$ be an adherent point of E. A function $f : \mathcal{X} \to \mathcal{Y}$ is said to *converge* to $L \in \mathcal{Y}$ as x converges to $x_0 \in E$, written

$$\lim_{x \to x_0; x \in E} f(x) = L, \tag{0.48}$$

iff

$$\begin{aligned} &\forall \epsilon > 0, \exists \delta > 0 \text{ s.t. } \forall x \in E, \\ &\qquad |x - x_0|_{\mathcal{X}} < \delta \ \Rightarrow\ |f(x) - L|_{\mathcal{Y}} < \epsilon. \end{aligned} \tag{0.49}$$

Notation 2. In Definition 0.64 we used the synonym notation

$$|u - v|_{\mathcal{X}} := d_{\mathcal{X}}(u, v). \tag{0.50}$$

Definition 0.65 (Pointwise convergence)**.** Let $(f_n)_{n=1}^{\infty}$ be a sequence of functions from one metric space $(\mathcal{X}, d_{\mathcal{X}})$ to another $(\mathcal{Y}, d_{\mathcal{Y}})$, and let $f : \mathcal{X} \to \mathcal{Y}$ be another function. We say that $(f_n)_{n=1}^{\infty}$ *converges pointwise* to f on $\mathcal{X}$ iff

$$\forall x \in \mathcal{X}, \qquad \lim_{n \to \infty} f_n(x) = f(x), \tag{0.51}$$

or, equivalently,

$$\forall \epsilon > 0, \forall x \in \mathcal{X},\ \exists N \in \mathbb{N}^+ \text{ s.t. } \forall n > N,\ |f_n(x) - f(x)|_{\mathcal{Y}} < \epsilon. \tag{0.52}$$

Example 0.39. Consider $f_n : [0, 1] \to \mathbb{R}$ defined by $f_n(x) := x^n$ and $f : [0, 1] \to \mathbb{R}$ defined by

$$f(x) := \begin{cases} 1 & \text{if } x = 1; \\ 0 & \text{if } x \in [0, 1). \end{cases}$$

The functions f_n are continuous and converge pointwise to f, which is discontinuous. Hence pointwise convergence does not preserve continuity.

Example 0.40. For the functions in Example 0.39, we have $\lim_{x \to 1; x \in [0,1)} x^n = 1$ for all n and $\lim_{x \to 1; x \in [0,1)} f(x) = 0$; it follows that

$$\lim_{n \to \infty} \lim_{x \to x_0; x \in \mathcal{X}} f_n(x) \neq \lim_{x \to x_0; x \in \mathcal{X}} \lim_{n \to \infty} f_n(x).$$

Hence pointwise convergence does not preserve limits.

Example 0.41. Consider the interval $[a, b] = [0, 1]$, and the function sequence $f_n : [a, b] \to \mathbb{R}$ given by

$$f_n(x) := \begin{cases} 2n & \text{if } x \in \left[\frac{1}{2n}, \frac{1}{n}\right]; \\ 0 & \text{otherwise.} \end{cases}$$

Then (f_n) converges pointwise to $f(x) = 0$. However, $\int_a^b f_n = 1$ for every n while $\int_a^b f = 0$. Hence

$$\lim_{n \to \infty} \int_a^b f_n \neq \int_a^b \lim_{n \to \infty} f_n.$$

Hence pointwise convergence does not preserve integral.

Example 0.42. Pointwise convergence does not preserve boundedness. For example, the function sequence

$$f_n(x) = \begin{cases} \exp(x) & \text{if } \exp(x) \leq n; \\ n & \text{if } \exp(x) > n \end{cases} \tag{0.53}$$

converges pointwise to $f(x) = \exp(x)$. Similarly, the function sequence

$$f_n(x) = \begin{cases} \frac{1}{x} & \text{if } x \geq \frac{1}{n}; \\ 0 & \text{if } x \in (0, \frac{1}{n}) \end{cases} \tag{0.54}$$

converges pointwise to $f(x) = \frac{1}{x}$. As another example, the function sequence

$$f_n(x) = n \sin \frac{x}{n} \tag{0.55}$$

converges pointwise to $f(x) = x$.

Definition 0.66 (Uniform convergence)**.** Let $(f_n)_{n=1}^{\infty}$ be a sequence of functions from one metric space $(\mathcal{X}, d_{\mathcal{X}})$ to another $(\mathcal{Y}, d_{\mathcal{Y}})$, and let $f : \mathcal{X} \to \mathcal{Y}$ be another function. We say that $(f_n)_{n=1}^{\infty}$ *converges uniformly* to f on $\mathcal{X}$ iff

$$\forall \epsilon > 0,\ \exists N \in \mathbb{N}^+ \text{ s.t. } \forall x \in \mathcal{X}, \forall n > N,\ |f_n(x) - f(x)|_{\mathcal{Y}} < \epsilon. \tag{0.56}$$

The sequence (f_n) is *locally uniformly convergent* to f iff for every point $x \in \mathcal{X}$ there is an $r > 0$ such that $(f_n|_{B_r(x) \cap \mathcal{X}})$ is uniformly convergent to f on $B_r(x) \cap \mathcal{X}$.

Theorem 0.67. Uniform convergence implies pointwise convergence.

Proof. This follows directly from (0.52), (0.56), and Theorem 0.6. □

Example 0.43 (Uniform convergence of Taylor series)**.** Consider $f : \mathbb{R} \to \mathbb{R}$ and the sequence of its Taylor polynomial $(T_n)_{n=1}^{\infty}$ in Definition 0.43. For any interval $I_r := (a - r, a + r)$, $(T_n)_{n=1}^{\infty}$ converges locally uniformly to $f|_{I_r}$ if r is less or equal to the radius of convergence of f at a. In particular, $(T_n)_{n=1}^{\infty}$ converges locally uniformly to f if the radius of convergence of f is $+\infty$.

0.3 Linear algebra

0.3.1 Vector spaces and the basis

Definition 0.68. A *field* is a commutative division ring. More commonly, a *field* $\mathbb{F}$ is a set together with two binary operations, usually called "addition" and "multiplication" and denoted by "+" and "∗", such that $\forall a, b, c \in \mathbb{F}$, the following axioms hold,

- commutativity: $a + b = b + a$, $ab = ba$;
- associativity: $a + (b + c) = (a + b) + c$, $a(bc) = (ab)c$;
- identity: $a + 0 = a$, $a1 = a$;
- invertibility: $a + (-a) = 0$, $aa^{-1} = 1$ $(a \neq 0)$;
- distributivity: $a(b + c) = ab + ac$.

Definition 0.69. A *vector space* or *linear space* over a field $\mathbb{F}$ is a set $\mathcal{V}$ together with two binary operations "+" and "×" respectively called vector addition and scalar multiplication that satisfy the following axioms:

(VSA-1) commutativity
$\forall \mathbf{u}, \mathbf{v} \in \mathcal{V}$, $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$;

(VSA-2) associativity
$\forall \mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathcal{V}$, $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$;

(VSA-3) compatibility
$\forall \mathbf{u} \in \mathcal{V}$, $\forall a, b \in \mathbb{F}$, $(ab)\mathbf{u} = a(b\mathbf{u})$;

(VSA-4) additive identity
$\forall \mathbf{u} \in \mathcal{V}$, $\exists \mathbf{0} \in \mathcal{V}$, s.t. $\mathbf{u} + \mathbf{0} = \mathbf{u}$;

(VSA-5) additive inverse
$\forall \mathbf{u} \in \mathcal{V}$, $\exists \mathbf{v} \in \mathcal{V}$, s.t. $\mathbf{u} + \mathbf{v} = \mathbf{0}$;

(VSA-6) multiplicative identity
$\forall \mathbf{u} \in \mathcal{V}$, $\exists 1 \in \mathbb{F}$, s.t. $1\mathbf{u} = \mathbf{u}$;

(VSA-7) distributive laws

$$\forall \mathbf{u}, \mathbf{v} \in \mathcal{V},\ \forall a, b \in \mathbb{F},\ \begin{cases} (a+b)\mathbf{u} = a\mathbf{u} + b\mathbf{u}, \\ a(\mathbf{u}+\mathbf{v}) = a\mathbf{u} + a\mathbf{v}. \end{cases}$$

The elements of $\mathcal{V}$ are called *vectors* and the elements of $\mathbb{F}$ are called *scalars*.

Definition 0.70. A vector space with $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$ is called a *real vector space* or a *complex vector space*, respectively.

Example 0.44. The simplest vector space is $\{\mathbf{0}\}$. Another simple example of a vector space over a field $\mathbb{F}$ is $\mathbb{F}$ itself, equipped with its standard addition and multiplication.

Definition 0.71. A *list of length* n or *n-tuple* is an ordered collection of n elements (which might be numbers, other lists, or more abstract entities) separated by commas and surrounded by parentheses: $\mathbf{x} = (x_1, x_2, \ldots, x_n)$.

Definition 0.72. A vector space composed of all the n-tuples of a field $\mathbb{F}$ is known as a *coordinate space*, denoted by $\mathbb{F}^n$ $(n \in \mathbb{N}^+)$.

Example 0.45. The properties of forces or velocities in the real world can be captured by a coordinate space $\mathbb{R}^2$ or $\mathbb{R}^3$.

Example 0.46. The set of continuous real-valued functions on the interval $[a, b]$ forms a real vector space.

Notation 3. For a set $\mathcal{S}$, define a vector space

$$\mathbb{F}^{\mathcal{S}} := \{f : \mathcal{S} \to \mathbb{F}\}.$$

$\mathbb{F}^n$ is a special case of $\mathbb{F}^{\mathcal{S}}$ because n can be regarded as the set $\{1, 2, \ldots, n\}$ and each element in $\mathbb{F}^n$ can be considered as a constant function.

Definition 0.73. A *linear combination* of a list of vectors $\{\mathbf{v}_i\}$ is a vector of the form $\sum_i a_i \mathbf{v}_i$ where $a_i \in \mathbb{F}$.

Example 0.47. $(17, -4, 2)$ is a linear combination of $(2, 1, -3), (1, -2, 4)$ because

$$(17, -4, 2) = 6(2, 1, -3) + 5(1, -2, 4).$$

Example 0.48. $(17, -4, 5)$ is not a linear combination of $(2, 1, -3), (1, -2, 4)$ because there do not exist numbers a_1, a_2 such that

$$(17, -4, 5) = a_1(2, 1, -3) + a_2(1, -2, 4).$$

Solving from the first two equations yields $a_1 = 6$, $a_2 = 5$, but $5 \neq -3 \times 6 + 4 \times 5$.

Definition 0.74. The *span* of a list of vectors $(\mathbf{v}_i)$ is the set of all linear combinations of $(\mathbf{v}_i)$,

$$\operatorname{span}(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m) = \left\{ \sum_{i=1}^{m} a_i \mathbf{v}_i :\ a_i \in \mathbb{F} \right\}. \tag{0.57}$$

In particular, the span of the empty set is $\{\mathbf{0}\}$. We say that $(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m)$ *spans* $\mathcal{V}$ if $\mathcal{V} = \operatorname{span}(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m)$.

Example 0.49.

$$(17, -4, 2) \in \operatorname{span}((2, 1, -3), (1, -2, 4))$$
$$(17, -4, 5) \notin \operatorname{span}((2, 1, -3), (1, -2, 4))$$

Definition 0.75. A vector space $\mathcal{V}$ is called *finite dimensional* if some list of vectors span $\mathcal{V}$; otherwise it is *infinite dimensional*.

Example 0.50. Let $\mathbb{P}_m(\mathbb{F})$ denote the set of all polynomials with coefficients in $\mathbb{F}$ and degree at most m,

$$\mathbb{P}_m(\mathbb{F}) = \left\{ p : \mathbb{F} \to \mathbb{F};\ p(z) = \sum_{i=0}^{m} a_i z^i, a_i \in \mathbb{F} \right\}. \tag{0.58}$$

Then $\mathbb{P}_m(\mathbb{F})$ is a finite-dimensional vector space for each non-negative integer m. The set of all polynomials with coefficients in $\mathbb{F}$, denoted by $\mathbb{P}(\mathbb{F}) := \mathbb{P}_{+\infty}(\mathbb{F})$, is infinite-dimensional. Both are subspaces of $\mathbb{F}^{\mathbb{F}}$ for $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$.

Definition 0.76. A list of vectors $(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m)$ in $\mathcal{V}$ is called *linearly independent* iff

$$a_1\mathbf{v}_1 + \ldots + a_m\mathbf{v}_m = \mathbf{0} \ \Rightarrow\ a_1 = \cdots = a_m = 0. \tag{0.59}$$

Otherwise the list of vectors is called *linearly dependent*.

Example 0.51. The empty list is declared to be linearly independent. A list of one vector $(\mathbf{v})$ is linearly independent iff $\mathbf{v} \neq \mathbf{0}$. A list of two vectors is linearly independent iff neither vector is a scalar multiple of the other.

Example 0.52. The list $(1, z, \ldots, z^m)$ is linearly independent in $\mathbb{P}_m(\mathbb{F})$ for each $m \in \mathbb{N}$.

Example 0.53. $(2,3,1)$, $(1,-1,2)$, and $(7,3,8)$ is linearly dependent in $\mathbb{R}^3$ because

$$2(2,3,1) + 3(1,-1,2) + (-1)(7,3,8) = (0,0,0).$$

Example 0.54. Every list of vectors containing the $\mathbf{0}$ vector is linearly dependent.

Lemma 0.77 (Linear dependence lemma)**.** Suppose $V = (\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_m)$ is a linearly dependent list in $\mathcal{V}$. Then there exists $j \in \{1, 2, \ldots, m\}$ such that

- $\mathbf{v}_j \in \operatorname{span}(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{j-1})$;
- if the jth term is removed from V, the span of the remaining list equals $\operatorname{span}(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m)$.

Lemma 0.78. In a finite-dimensional vector space, the length of every linearly independent list of vectors is less than or equal to the length of every spanning list of vectors.

Definition 0.79. A *basis* of a vector space $\mathcal{V}$ is a list of vectors in $\mathcal{V}$ that is linearly independent and spans $\mathcal{V}$.

Definition 0.80. The *standard basis* of $\mathbb{F}^n$ is the list of vectors

$$(1,0,\cdots,0)^T,\ (0,1,0,\cdots,0)^T,\ \ldots,\ (0,\cdots,0,1)^T. \tag{0.60}$$

Example 0.55. $(z^0, z^1, \ldots, z^m)$ is a basis of $\mathbb{P}_m(\mathbb{F})$ in (0.58).

Lemma 0.81. A list of vectors $(\mathbf{v}_1, \ldots, \mathbf{v}_n)$ is a basis of $\mathcal{V}$ iff every vector $\mathbf{u} \in \mathcal{V}$ can be written uniquely as

$$\mathbf{u} = \sum_{i=1}^{n} a_i \mathbf{v}_i, \tag{0.61}$$

where $a_i \in \mathbb{F}$.

Lemma 0.82. Every spanning list in a vector space $\mathcal{V}$ can be reduced to a basis of $\mathcal{V}$.

Lemma 0.83. Every linearly independent list of vectors in a finite-dimensional vector space can be extended to a basis of that vector space.

Definition 0.84. The *dimension* of a finite-dimensional vector space $\mathcal{V}$, denoted $\dim \mathcal{V}$, is the length of any basis of the vector space.

Lemma 0.85. If $\mathcal{V}$ is finite-dimensional, then every spanning list of vectors in $\mathcal{V}$ with length $\dim \mathcal{V}$ is a basis of $\mathcal{V}$.

Lemma 0.86. If $\mathcal{V}$ is finite-dimensional, then every linearly independent list of vectors in $\mathcal{V}$ with length $\dim \mathcal{V}$ is a basis of $\mathcal{V}$.

0.3.2 Inner product spaces

Definition 0.87. Let $\mathbb{F}$ be the underlying field of a vector space $\mathcal{V}$. The *inner product* $\langle \mathbf{u}, \mathbf{v} \rangle$ on $\mathcal{V}$ is a function $\mathcal{V} \times \mathcal{V} \to \mathbb{F}$ that satisfies

(IP-1) real positivity: $\forall \mathbf{v} \in \mathcal{V}$, $\langle \mathbf{v}, \mathbf{v} \rangle \geq 0$;

(IP-2) definiteness: $\langle \mathbf{v}, \mathbf{v} \rangle = 0$ iff $\mathbf{v} = \mathbf{0}$;

(IP-3) additivity in the first slot:
$\forall \mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathcal{V}$, $\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$;

(IP-4) homogeneity in the first slot:
$\forall a \in \mathbb{F}$, $\forall \mathbf{v}, \mathbf{w} \in \mathcal{V}$, $\langle a\mathbf{v}, \mathbf{w} \rangle = a \langle \mathbf{v}, \mathbf{w} \rangle$;

(IP-5) conjugate symmetry: $\forall \mathbf{v}, \mathbf{w} \in \mathcal{V}$, $\langle \mathbf{v}, \mathbf{w} \rangle = \overline{\langle \mathbf{w}, \mathbf{v} \rangle}$.

An *inner product space* is a vector space $\mathcal{V}$ equipped with an inner product on $\mathcal{V}$.

Exercise 0.56. Deduce *additivity in the second slot* and *conjugate homogeneity in the second slot* from Definition 0.87.

Definition 0.88. The *Euclidean inner product* on $\mathbb{F}^n$ is

$$\langle \mathbf{v}, \mathbf{w} \rangle = \sum_{i=1}^{n} v_i \overline{w_i}. \tag{0.62}$$

Definition 0.89. Let $\mathbb{F}$ be the underlying field of an inner product space $\mathcal{V}$. The *norm induced by an inner product* on $\mathcal{V}$ is a function $\mathcal{V} \to \mathbb{F}$:

$$\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}. \tag{0.63}$$

Definition 0.90. The *Euclidean ℓ_p norm* of a vector $\mathbf{v} \in \mathbb{F}^n$ is

$$\|\mathbf{v}\|_p = \left(\sum_{i=1}^{n} |v_i|^p \right)^{\frac{1}{p}} \tag{0.64}$$

and the *Euclidean ℓ_∞ norm* is

$$\|\mathbf{v}\|_\infty = \max_i |v_i|. \tag{0.65}$$

Theorem 0.91 (Equivalence of norms)**.** Any two norms $\|\cdot\|_N$ and $\|\cdot\|_M$ on a finite dimensional vector space $\mathcal{V} = \mathbb{C}^n$ satisfy

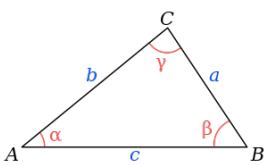
$$\exists c_1, c_2 \in \mathbb{R}^+,\ \text{s.t. } \forall \mathbf{x} \in \mathcal{V},\ c_1 \|\mathbf{x}\|_M \leq \|\mathbf{x}\|_N \leq c_2 \|\mathbf{x}\|_M. \tag{0.66}$$

Definition 0.92. The angle between two vectors $\mathbf{v}, \mathbf{w}$ in an inner product space with $\mathbb{F} = \mathbb{R}$ is the number $\theta \in [0, \pi]$,

$$\theta = \arccos \frac{\langle \mathbf{v}, \mathbf{w} \rangle}{\|\mathbf{v}\| \|\mathbf{w}\|}. \tag{0.67}$$

Theorem 0.93 (The law of cosines)**.** Any triangle satisfies

$$c^2 = a^2 + b^2 - 2ab\cos\gamma. \tag{0.68}$$



Proof. The dot product of AB to $AB = CB - CA$ yields

$$c^2 = \langle AB, CB\rangle - \langle AB, CA\rangle .$$

The dot products of CB and CA to $AB = CB - CA$ yield

$$\begin{aligned}\langle CB, AB\rangle &= a^2 - \langle CB, CA\rangle ;\\ -\langle CA, AB\rangle &= -\langle CA, CB\rangle + b^2.\end{aligned}$$

The proof is completed by adding up all three equations and applying (0.67). □

Theorem 0.94 (The law of cosines: abstract version)**.** Any induced norm on a real vector space satisfies

$$\|\mathbf{u} - \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - 2\langle \mathbf{u}, \mathbf{v}\rangle . \tag{0.69}$$

Proof. Definitions 0.89 and 0.87 and $\mathbb{F} = \mathbb{R}$ yield

$$\begin{aligned}\|\mathbf{u} - \mathbf{v}\|^2 &= \langle \mathbf{u} - \mathbf{v}, \mathbf{u} - \mathbf{v}\rangle \\ &= \langle \mathbf{u}, \mathbf{u}\rangle + \langle \mathbf{v}, \mathbf{v}\rangle - \langle \mathbf{u}, \mathbf{v}\rangle - \langle \mathbf{v}, \mathbf{u}\rangle \\ &= \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - 2\langle \mathbf{u}, \mathbf{v}\rangle .\end{aligned}$$

□

Definition 0.95. Two vectors $\mathbf{u}, \mathbf{v}$ are called *orthogonal* if $\langle \mathbf{u}, \mathbf{v}\rangle = 0$, i.e., their inner product is the additive identity of the underlying field.

Example 0.57. An inner product on the vector space of continuous real-valued functions on the interval $[-1, 1]$ is

$$\langle f, g\rangle = \int_{-1}^{+1} f(x)g(x)\mathrm{d}x.$$

f and g are said to be orthogonal if the integral is zero.

Theorem 0.96 (Pythagorean)**.** If $\mathbf{u}, \mathbf{v}$ are orthogonal, then $\|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$.

Proof. This follows from (0.69) and Definition 0.95. □

Theorem 0.97 (Cauchy-Schwarz inequality)**.**

$$|\langle \mathbf{u}, \mathbf{v}\rangle| \le \|\mathbf{u}\|\|\mathbf{v}\|, \tag{0.70}$$

where the equality holds iff one of $\mathbf{u}, \mathbf{v}$ is a scalar multiple of the other.

Proof. For any complex number λ, (IP-1) implies

$$\begin{aligned}&\langle \mathbf{u} + \lambda\mathbf{v}, \mathbf{u} + \lambda\mathbf{v}\rangle \ge 0\\ \Rightarrow\ &\langle \mathbf{u}, \mathbf{u}\rangle + \lambda\langle \mathbf{v}, \mathbf{u}\rangle + \bar{\lambda}\langle \mathbf{u}, \mathbf{v}\rangle + \lambda\bar{\lambda}\langle \mathbf{v}, \mathbf{v}\rangle \ge 0.\end{aligned}$$

If $\mathbf{v} = 0$, (0.70) clearly holds. Otherwise (0.70) follows from substituting $\lambda = -\frac{\langle \mathbf{u}, \mathbf{v}\rangle}{\langle \mathbf{v}, \mathbf{v}\rangle}$ into the above equation. □

Exercise 0.58. To explain the choice of λ in the proof of Theorem 0.97, what is the geometric meaning of (0.70) in the plane? When will the equality hold?

Example 0.59. If $x_i, y_i \in \mathbb{R}$, then for any $n \in \mathbb{N}^+$

$$\left|\sum_{i=1}^{n} x_i y_i\right|^2 \le \sum_{j=1}^{n} x_j^2 \sum_{k=1}^{n} y_k^2.$$

Example 0.60. If $f, g : [a, b] \to \mathbb{R}$ are continuous, then

$$\left|\int_a^b f(x)g(x)\mathrm{d}x\right|^2 \le \left(\int_a^b f^2(x)\mathrm{d}x\right)\left(\int_a^b g^2(x)\mathrm{d}x\right)$$

0.3.3 Normed vector spaces

Definition 0.98. A function $\|\cdot\| : \mathcal{V} \to \mathbb{F}$ is a *norm* for a vector space $\mathcal{V}$ iff it satisfies

(NRM-1) real positivity: $\forall \mathbf{v} \in \mathcal{V}$, $\|\mathbf{v}\| \ge 0$;

(NRM-2) point separation: $\|\mathbf{v}\| = 0 \Rightarrow \mathbf{v} = \mathbf{0}$.

(NRM-3) absolute homogeneity: $\forall a \in \mathbb{F}$, $\forall \mathbf{v} \in \mathcal{V}$, $\|a\mathbf{v}\| = |a|\|\mathbf{v}\|$;

(NRM-4) triangle inequality: $\forall \mathbf{u}, \mathbf{v} \in \mathcal{V}$, $\|\mathbf{u} + \mathbf{v}\| \le \|\mathbf{u}\| + \|\mathbf{v}\|$.

A *normed vector space* or simply a *normed space* is a vector space $\mathcal{V}$ equipped with a norm on $\mathcal{V}$.

Exercise 0.61. Explain how (NRM-1,2,3,4) relate to the geometric meaning of the norm of vectors in $\mathbb{R}^3$.

Lemma 0.99. The norm induced by an inner product is a norm as in Definition 0.98.

Proof. The induced norm as in (0.63) satisfies (NRM-1,2) trivially. For (NRM-3),

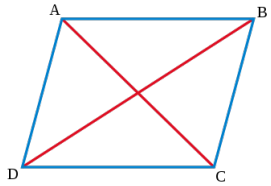
$$\|a\mathbf{v}\|^2 = \langle a\mathbf{v}, a\mathbf{v}\rangle = a\langle \mathbf{v}, a\mathbf{v}\rangle = a\bar{a}\langle \mathbf{v}, \mathbf{v}\rangle = |a|^2\|\mathbf{v}\|^2.$$

To prove (NRM-4), we have

$$\begin{aligned}\|\mathbf{u} + \mathbf{v}\|^2 &= \langle \mathbf{u} + \mathbf{v}, \mathbf{u} + \mathbf{v}\rangle \\ &= \langle \mathbf{u}, \mathbf{u}\rangle + \langle \mathbf{v}, \mathbf{v}\rangle + \langle \mathbf{u}, \mathbf{v}\rangle + \overline{\langle \mathbf{u}, \mathbf{v}\rangle} \\ &= \langle \mathbf{u}, \mathbf{u}\rangle + \langle \mathbf{v}, \mathbf{v}\rangle + 2|\langle \mathbf{u}, \mathbf{v}\rangle| \\ &\le \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 + 2\|\mathbf{u}\|\|\mathbf{v}\| \\ &= (\|\mathbf{u}\| + \|\mathbf{v}\|)^2 ,\end{aligned}$$

where the second step follows from (IP-5) and the fourth step from Cauchy-Schwarz inequality. □

Theorem 0.100 (The parallelogram law)**.** The sum of squares of the lengths of the four sides of a parallelogram equals the sum of squares of the two diagonals.



More precisely, we have in the above plot

$$(AB)^2 + (BC)^2 + (CD)^2 + (DA)^2 = (AC)^2 + (BD)^2. \tag{0.71}$$

Proof. Apply the law of cosines to the two diagonals, add the two equations, and we obtain (0.71). □

Theorem 0.101 (The parallelogram law: abstract version)**.** Any induced norm (0.63) satisfies

$$2\|\mathbf{u}\|^2 + 2\|\mathbf{v}\|^2 = \|\mathbf{u} + \mathbf{v}\|^2 + \|\mathbf{u} - \mathbf{v}\|^2. \tag{0.72}$$

Proof. Replace $\mathbf{v}$ in (0.69) with $-\mathbf{v}$ and we have

$$\|\mathbf{u}+\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 + 2\langle \mathbf{u}, \mathbf{v}\rangle .$$

(0.72) follows from adding the above equation to (0.69). □

Exercise 0.62. In the case of Euclidean ℓ_p norms, show that the parallelogram law (0.72) holds if and only if $p = 2$.

Theorem 0.102. The induced norm (0.63) holds for some inner product $\langle \cdot, \cdot \rangle$ if and only if the parallelogram law (0.72) holds for every pair of $\mathbf{u}, \mathbf{v} \in \mathcal{V}$.

Exercise 0.63. Prove Theorem 0.102.

Example 0.64. By Theorem 0.102 and Exercise 0.62, the ℓ^1 and ℓ^∞ spaces do not have a corresponding inner product for the ℓ_1 and ℓ_∞ norms.

0.4 Abstract algebra

0.4.1 Binary algebraic structures

Definition 0.103. A *binary operation* on a set $\mathcal{S}$ is a map $\mathcal{S} \times \mathcal{S} \to \mathcal{S}$. A *binary algebraic structure* or a *magma* is an ordered pair $(\mathcal{S}, *)$ where $\mathcal{S}$ is a set and $*$ a binary operation on $\mathcal{S}$.

Definition 0.104. Let $(\mathcal{S}, *)$ and $(\mathcal{S}', *')$ be two binary algebraic structures. A *homomorphism* between $\mathcal{S}$ and $\mathcal{S}'$ is a map $\phi : \mathcal{S} \to \mathcal{S}'$ satisfying

$$\forall a, b \in \mathcal{S}, \qquad \phi(a * b) = \phi(a) *' \phi(b). \tag{0.73}$$

Definition 0.105 (Type of homomorphisms)**.** A *monomorphism* is an injective homomorphism, an *epimorphism* is a surjective homomorphism, An *endomorphism* is a homomorphism $\phi : \mathcal{S} \to \mathcal{S}$. An *isomorphism* is a bijective homomorphism. If such an isomorphism exists between $\mathcal{S}$ and $\mathcal{S}'$, they are said to be *isomorphic*, written $\mathcal{S} \simeq \mathcal{S}'$. An *automorphism* is an isomorphism $\phi : \mathcal{S} \to \mathcal{S}$.

Exercise 0.65. $(\mathbb{R}, +)$ is isomorphic to $(\mathbb{R}^+, \times)$.

Definition 0.106. An element e of a binary structure $(\mathcal{S}, *)$ is an *identity element for* $*$ iff

$$\forall s \in \mathcal{S},\ e * s = s * e = s. \tag{0.74}$$

Theorem 0.107 (Uniqueness of the identity element)**.** A binary structure has at most one identity element.

Proof. Suppose there are two identity elements e and e'. Then (0.74) implies that they are equal. □

0.4.2 Groups

Definition 0.108. A *group* is an ordered pair $\langle \mathcal{G}, * \rangle$ where $\mathcal{G}$ is a set and '$*$' is a binary operation on $\mathcal{G}$ satisfying the following axioms:

(GRP-1) associativity
$\forall a, b, c \in \mathcal{G},\ \ (a * b) * c = a * (b * c)$;

(GRP-2) identity element
$\exists e \in \mathcal{G},$ s.t. $\forall x \in \mathcal{G}, e * x = x * e = x$;

(GRP-3) inverse element
$\forall a \in \mathcal{G}, \exists a' \in \mathcal{G}.$ s.t. $a' * a = a * a' = e$;

$\mathcal{G}$ is *abelian* if $*$ is commutative.

Example 0.66. An example of the "is-a" relation is: Abelian group → group → binary algebraic structure.

Exercise 0.67. Find out the definitions of *semigroup*, *monoid*, and *groupoid*; what are their relations to the concepts of magma and group?

Definition 0.109. The *order of a group* G, written $|G|$, is the number of elements in G.

Example 0.68. Each of $\mathbb{Q}, \mathbb{R}, \mathbb{C}$ is an abelian group under addition, and is not under multiplication. With 0 deleted, each resulting set is a group under multiplication.

Example 0.69. $(\mathbb{R}^{m\times n}, +)$ is a group while $(\mathbb{R}^{m\times n}, \times)$ is not. The set of invertible matrices $\mathbb{R}^{n\times n}$ is a group.

Theorem 0.110 (Cancellation laws)**.** For a group $(\mathcal{G}, *)$, the *left and right cancellation laws* hold in $\mathcal{G}$,

$$\forall a, b, c \in \mathcal{G}, \qquad \begin{cases} a * b = a * c \Rightarrow b = c, \\ b * a = c * a \Rightarrow b = c. \end{cases} \tag{0.75}$$

Proof. This follows from multiplying the first equation by a' and applying the associative law and identity element. □

Theorem 0.111. For a group $(\mathcal{G}, *)$, the linear equations $a * x = b$ and $y * a = b$ have unique solutions if $a, b \in \mathcal{G}$.

Proof. The existence follows from multiplying the first equation by a', the associative law, and the identity element. The uniqueness follows from Theorem 0.110. □

Example 0.70. In the case of solving a linear system $A\mathbf{x} = \mathbf{b}$, we know it has a unique solution so long as A belongs to the group of invertible matrices.

Corollary 0.112. The identity element of a group $(\mathcal{G}, *)$ is unique.

Proof. A group *is a* binary algebraic structure and the rest of the proof follows from Theorem 0.107. □

Corollary 0.113. For each element a in a group $\mathcal{G}$, there is only one element $a' \in \mathcal{G}$ such that $a' * a = a * a' = e$.

Proof. This follows from Theorem 0.110. □

Corollary 0.114. Let a' denote the inverse of a in a group $\mathcal{G}$. Then

$$\forall a, b \in \mathcal{G}, \qquad (a * b)' = b' * a'. \tag{0.76}$$

Proof. $(b' * a') * (a * b) = e$. The uniqueness is guaranteed by Corollary 0.113. □

0.4.3 Subgroups and generating sets

Definition 0.115. If a subset $H \subseteq G$ of a group G is closed under the binary operation and H itself forms a group, then H is a *subgroup* of G, written $H \leq G$.

Definition 0.116. The *improper subgroup* of a group G is the subgroup G; all other subgroups of G are *proper subgroups* of G. The subgroup $\{e\}$ is the *trivial subgroup* of G; all other subgroups are *nontrivial subgroups*.

Definition 0.117. Let G be a group and $a \in G$. The *cyclic subgroup of G generated by a* is the subgroup

$$\langle a \rangle = \{a^n \in G : n \in \mathbb{Z}\}. \tag{0.77}$$

Theorem 0.118. The cyclic subgroup of G generated by a is the smallest subgroup of G that contains a.

Definition 0.119. An element a of a group G *generates* G and is a *generator for* G if $\langle a \rangle = G$.

Definition 0.120. The *intersection of the sets* S_i is the set of all elements that are all in the sets S_i,

$$\cap_{i \in I} S_i = \{x : \forall i \in I, \quad x \in S_i\}. \tag{0.78}$$

Theorem 0.121. The intersection of some subgroups H_i of a group G for $i \in I$ is again a subgroup of G.

Definition 0.122. Let G be a group and let $a_i \in G$ for $i \in I$. The smallest subgroup of G containing $\{a_i : i \in I\}$ is the *subgroup generated by* $\{a_i : i \in I\}$. If this subgroup is all of G, then $\{a_i : i \in I\}$ *generates* G and the a_i's are *generators of* G. If there is a finite set $\{a_i : i \in I\}$ that generates G, then G is *finitely generated*.

0.4.4 Permutations and symmetric groups

Definition 0.123. A *permutation of a set* A is a bijective function $\sigma : A \to A$.

Notation 4. A permutation on $\{1, 2, \ldots, n\}$ can be denoted by *Cauchy's two-line notation*,

$$\begin{pmatrix} 1 & 2 & \cdots & n \\ \downarrow & \downarrow & \cdots & \downarrow \\ \sigma(1) & \sigma(2) & \cdots & \sigma(n) \end{pmatrix} \tag{0.79}$$

with arrows often omitted, or *Cauchy's one-line notation*,

$$(\sigma(1)\ \sigma(2)\ \cdots\ \sigma(n)), \tag{0.80}$$

or *cyclic notation*, e.g.,

$$(2, 1, 3)(4, 5) := 2 \mapsto 1 \mapsto 3 \mapsto 2; 4 \mapsto 5 \mapsto 4.$$

The cyclic notation does not display any element that is fixed under the permutation.

Theorem 0.124. Let A be a non-empty set, and let S_A be the collection of all permutations of A. Then $(S_A, \circ, ^{-1}, e)$ is a group where the identity e is the identity function.

Definition 0.125. Let A be the finite set $\{1, 2, \ldots, n\}$. The group of all permutations of A is called the *symmetric group on n letters*, and is denoted by S_n.

Exercise 0.71. Show that S_n has $n!$ elements.

Definition 0.126. The nth *dihedral group* D_n is the group of symmetries of the regular n-gon.

Example 0.72. S_3 is also called the *group D_3 of symmetries of an equilateral triangle*. Label the vertices of an equilateral triangle by $X = \{1, 2, 3\}$. Then the six elements in S_X can be interpreted as three reflections and three rotations around the centroid of $\frac{2i}{3}\pi$ with $i = 1, 2, 3$.

Example 0.73. S_4 is also called the *octic group* or the *group D_4 of symmetries of the square*. Label the vertices of an equilateral triangle by $X = \{1, 2, 3, 4\}$. Then the eight elements in S_X can be interpreted as four rotations, two reflections along the two diagonal directions, and two reflections along the horizontal and vertical axis.

Lemma 0.127. Let G and G' be groups and let $\phi : G \to G'$ be an injective homomorphism. then $\phi(G)$, the image of G, is a subgroup of G' and ϕ is an isomorphism.

Exercise 0.74. Prove Lemma 0.127.

Theorem 0.128 (Cayley)**.** Every group G is isomorphic to a subgroup of S_G.

Proof. For $x \in G$, define $\lambda_x : G \to G$ as $\lambda_x(g) = xg$ for all $g \in G$. λ_x is surjective because

$$\forall c \in G, \exists x^{-1}c \in G, \text{ s.t. } \lambda_x(x^{-1}c) = c.$$

λ_x is injective because

$$\lambda_x(a) = \lambda_x(b) \ \Rightarrow\ xa = xb \ \Rightarrow\ a = b.$$

Therefore λ_x is a permutation of G.

Define $\phi : G \to S_G$ by $\phi(x) = \lambda_x$ for all $x \in G$. ϕ is injective because

$$\phi(x) = \phi(y) \ \Rightarrow\ \lambda_x(e) = \lambda_y(e) \ \Rightarrow\ x = y.$$

ϕ is a homomorphism because

$$\begin{aligned} \forall g \in G, \phi(xy)(g) &= \lambda_{xy}(g) = (xy)g = \lambda_x(\lambda_y(g)) \\ &= (\lambda_x\lambda_y)(g). \end{aligned}$$

The proof is completed by Lemma 0.127. □

0.4.5 Group action on a set

Definition 0.129. An *action of a group G on a set X* is a map $* : G \times X \to X$ such that

(1) $\forall x \in X$, $ex = x$,

(2) $\forall x \in X$, $\forall g_1, g_2 \in G$, $(g_1g_2)(x) = g_1(g_2(x))$.

X is called a *G-set* if G has an action on X.

Example 0.75. The set $X = \{1, 2, \ldots, n\}$ in Examples 0.72 and 0.73 is a S_X-set since the action of S_X on X can be defined as $\cdot(\sigma, x) = \sigma x$.

Theorem 0.130. Let X be a G-set. For each $g \in G$, the function $\sigma_g : X \to X$ defined by $\sigma_g(x) = gx$ is a permutation of X. Also, the map $\phi : G \to S_G$ defined by $\phi(g) = \sigma_g$ is a homomorphism.

Proof. The proof is similar to that of Theorem 0.128. □

Exercise 0.76. Let H be a subgroup of G. Show that G is an H-set under conjugation with

$$\forall g \in G, \forall h \in H, \ \cdot(h, g) = hgh^{-1}.$$

0.4.6 Orbits and alternating groups

Definition 0.131. For a permutation $\sigma : A \to A$, define an equivalence relation by

$$\forall a, b \in A, \quad a \sim b \ \Leftrightarrow \ \exists n \in \mathbb{Z} \text{ s.t. } b = \sigma^n(a). \tag{0.81}$$

The equivalence classes in A determined by (0.81) are called the *orbits of the permutation* σ.

Exercise 0.77. Show that (0.81) is indeed an equivalence relation.

Example 0.78. The orbits of the permutation

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 3 & 8 & 6 & 7 & 4 & 1 & 5 & 2 \end{pmatrix}$$

are $\{1, 3, 6\}$, $\{2, 8\}$, and $\{4, 5, 7\}$.

Definition 0.132. A permutation $\sigma \in S_n$ is a *cycle* if at most one of its orbits contains more than one element. Two cycles are *disjoint* if any integer is moved by at most one of these cycles.

Example 0.79. Cyclic notations now make perfect sense.

$$(1, 3, 5, 4) = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 2 & 5 & 1 & 4 \end{pmatrix}$$

Clearly $(1, 3, 5, 4) = (3, 5, 4, 1) = (5, 4, 1, 3) = (4, 1, 3, 5)$.

Theorem 0.133. Every permutation σ of a finite set is a product of disjoint cycles.

Proof. Let $B_1, B_2, \ldots, B_r$ be the orbits of σ, and define corresponding cycles as

$$\mu_i(x) = \begin{cases} \sigma(x) & \text{if } x \in B_i; \\ x & \text{otherwise.} \end{cases} \tag{0.82}$$

Clearly $\sigma = \mu_1\mu_2 \cdots \mu_r$. Because the orbits are pairwise disjoint, so are the cycles. □

Example 0.80. The multiplication of disjoint cycles is commutative.

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 6 & 5 & 2 & 4 & 3 & 1 \end{pmatrix} = (1, 6)(2, 5, 3) = (2, 5, 3)(1, 6).$$

Example 0.81. If two cycles are not disjoint, their multiplication is not commutative; in fact, their multiplication might not even be a cycle:

$$(1, 4, 5, 6)(2, 1, 5) = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 6 & 4 & 3 & 5 & 2 & 1 \end{pmatrix},$$
$$(2, 1, 5)(1, 4, 5, 6) = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 4 & 1 & 3 & 2 & 6 & 5 \end{pmatrix}.$$

Definition 0.134. A *transposition* is a cycle of length 2.

Example 0.82. The permutation (2 1 3) in Cauchy's one-line notation is a transposition $(1, 2)$.

Exercise 0.83. Consider the set of standard basis vectors for $\mathbb{R}^n$ as in Definition 0.80,

$$E := \{e_1, e_2, \ldots, e_n\}. \tag{0.83}$$

Show that

(1) Every set X of n elements is isomorphic to E.
(2) Every permutation $\sigma : E \to E$ is a matrix $\mathbb{Z}_2^{n \times n}$ where there is exactly one 1 in each column and each row.
(3) In particular, a transposition $\tau_{i,j}$ is an elementary matrix A of type II, i.e., A is the identity matrix except $a_{i,i} = a_{j,j} = 0$ and $a_{i,j} = a_{j,i} = 1$.

Lemma 0.135. Any permutation of a finite set X of at least two elements is a product of transpositions. In other words, the set of transpositions generates the symmetric group.

Proof. It is easily verified that

$$(a_1, a_2, \ldots, a_n) = (a_1, a_n)(a_1, a_{n-1}) \cdots (a_1, a_2)$$

and the rest of the proof follows from Theorem 0.133. □

Example 0.84. In S_n for $n \geq 2$, the identity permutation is the product $(1, 2)(1, 2)$ of transpositions.

Lemma 0.136. For a permutation $\sigma \in S_n$ and a transposition $\tau = (i, j)$ in S_n, the numbers of orbits of σ and of $\tau\sigma$ differ by 1.

Proof. Suppose i and j are in the same orbit of σ. By Theorem 0.133 we can write σ as a product of disjoint cycles with the first cycle of the form

$$(a, i, c, \ldots, b, j, d, \ldots).$$

Then we have

$$\tau\sigma = (i, j)(a, i, c, \ldots, b, j, d, \ldots) = (a, j, d, \ldots)(b, i, c, \ldots).$$

Suppose i and j are in different orbits of σ. WLOG, we write the first two cycles as

$$(b, j, d, \cdots)(a, i, c, \ldots).$$

Then we have

$$\tau\sigma = (i, j)(b, j, d, \cdots)(a, i, c, \ldots) = (a, j, d, \ldots, b, i, c, \ldots).$$

The statement is proved for other cases similarly. □

Theorem 0.137. No permutations in S_n can be expressed both as a product of an even number of transpositions and as a product of an odd number of transpositions.

Proof. By Lemma 0.135, any $\sigma \in S_n$ can be expressed as

$$\sigma = \tau_1 \tau_2 \cdots \tau_m I,$$

where the identity I has n orbits. Then Lemma 0.136 completes the proof. □

Definition 0.138. The *signature of a permutation* σ, denoted by $\operatorname{sgn}(\sigma)$, is $+1$ or -1 if σ can be expressed as an even or odd number of transpositions, respectively; we also say that the permutation is an *even permutation* or an *odd permutation*, respectively.

Example 0.85. The identity in S_n is an even permutation.

$$\sigma = (1,4,5,6)(2,1,5) = (1,6)(1,5)(1,4)(2,5)(2,1)$$

is an odd permutation.

Theorem 0.139. If $n \geq 2$, then the collection of all even permutations of $\{1, 2, \ldots, n\}$ forms a subgroup of order $n!/2$ of the symmetric group S_n.

Exercise 0.86. Prove Theorem 0.139.

Definition 0.140. The *alternating group* A_n *on* n *letters* is the subgroup of S_n consisting of all even permutations of n letters.

0.4.7 Determinants

Definition 0.141. The *signed volume of a parallelotope* spanned by n vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n \in \mathbb{R}^n$ is a function $\delta : \mathbb{R}^{n\times n} \to \mathbb{R}$ that satisfies

(SVP-1) $\delta(I) = 1$;

(SVP-2) $\delta(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n) = 0$ if $\mathbf{v}_i = \mathbf{v}_j$ for some $i \neq j$;

(SVP-3) δ is linear, i.e., $\forall j = 1, \ldots, n,\ \forall c \in \mathbb{R}$,

$$\begin{aligned} &\delta(\mathbf{v}_1, \ldots, \mathbf{v}_{j-1}, \mathbf{v} + c\mathbf{w}, \mathbf{v}_{j+1}, \ldots, \mathbf{v}_n) \\ &= \delta(\mathbf{v}_1, \ldots, \mathbf{v}_{j-1}, \mathbf{v}, \mathbf{v}_{j+1}, \ldots, \mathbf{v}_n) \\ &\quad + c\delta(\mathbf{v}_1, \ldots, \mathbf{v}_{j-1}, \mathbf{w}, \mathbf{v}_{j+1}, \ldots, \mathbf{v}_n). \end{aligned} \tag{0.84}$$

Lemma 0.142. Adding a multiple of one vector to another does not change the determinant.

Proof. This follows directly from (SVP-2,3). □

Lemma 0.143. If the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ are linearly dependent, then $\delta(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n) = 0$.

Proof. WLOG, we assume $\mathbf{v}_1 = \sum_{i=2}^n c_i \mathbf{v}_i$. Then the result follows from (SVP-2,3). □

Lemma 0.144. The signed volume δ is alternating, i.e.,

$$\delta(\mathbf{v}_1, \ldots, \mathbf{v}_i, \ldots, \mathbf{v}_j, \ldots, \mathbf{v}_n) = -\delta(\mathbf{v}_1, \ldots, \mathbf{v}_j, \ldots, \mathbf{v}_i, \ldots, \mathbf{v}_n). \tag{0.85}$$

Exercise 0.87. Prove Lemma 0.144 using (SVP-2,3).

Lemma 0.145. Let M_σ denote the matrix of a permutation $\sigma : E \to E$ where E is in (0.83). Then we have $\delta(M_\sigma) = \operatorname{sgn}(\sigma)$.

Proof. There is a one-to-one correspondence between the vectors in the matrix

$$M_\sigma = [e_{\sigma(1)}, e_{\sigma(2)}, \ldots, e_{\sigma(n)}]$$

and the scalars in the one-line notation

$$(\sigma(1)\ \sigma(2)\ \ldots\ \sigma(n)).$$

A sequence of transpositions taking σ to the identity map also takes M_σ to the identity matrix. By Lemma 0.144, each transposition yields a multiplication factor -1. Definition 0.138 and (SVP-1) give $\delta(M_\sigma) = \operatorname{sgn}(\sigma)\delta(I) = \operatorname{sgn}(\sigma)$. □

Definition 0.146 (Leibniz formula of determinants)**.** The *determinant of a square matrix* $A \in \mathbb{R}^{n\times n}$ is

$$\det A = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^n a_{\sigma(i),i}, \tag{0.86}$$

where the sum is over all permutations σ in the symmetric group and $a_{\sigma(i),i}$ is the element of A at the $\sigma(i)$th row and the ith column.

Exercise 0.88. Show that the determinant formula in (0.86) reduces to

$$\det \begin{bmatrix} a & c \\ b & d \end{bmatrix} = ad - bc \tag{0.87}$$

for $n = 2$. Give a geometric proof that $ad - bc$ is the signed volume of the parallelogram determined by the vectors $(a, b)^T$ and $(c, d)^T$ on the plane.

Theorem 0.147. The signed volume function satisfying (SVP-1,2,3) in Definition 0.141 is unique and is the same as the determinant in (0.86).

Proof. Let the parallelotope be spanned by the column vec-

tors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$. We have

$$
\begin{aligned}
&\delta \begin{bmatrix} v_{11} & v_{12} & \ldots & v_{1n} \\ v_{21} & v_{22} & \ldots & v_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ v_{n1} & v_{n2} & \ldots & v_{nn} \end{bmatrix} \\
&= \sum_{i_1=1}^{n} v_{i_1 1} \delta \begin{bmatrix} | & v_{12} & \ldots & v_{1n} \\ e_{i_1} & v_{22} & \ldots & v_{2n} \\ | & \vdots & \ddots & \vdots \\ | & v_{n2} & \ldots & v_{nn} \end{bmatrix} \\
&= \sum_{i_1, i_2=1}^{n} v_{i_1 1} v_{i_2 2} \delta \begin{bmatrix} | & | & v_{13} & \ldots & v_{1n} \\ e_{i_1} & e_{i_2} & v_{23} & \ldots & v_{2n} \\ | & | & \vdots & \ddots & \vdots \\ | & | & v_{n2} & \ldots & v_{nn} \end{bmatrix} \\
&= \cdots \\
&= \sum_{i_1, i_2, \ldots, i_n=1}^{n} v_{i_1 1} v_{i_2 2} \cdots v_{i_n n} \delta \begin{bmatrix} | & | & \ldots & | \\ e_{i_1} & e_{i_2} & \ldots & e_{i_n} \\ | & | & \ldots & | \end{bmatrix} \\
&= \sum_{\sigma \in S_n} v_{\sigma(1),1} v_{\sigma(2),2} \cdots v_{\sigma(n),n} \delta \begin{bmatrix} | & | & \ldots & | \\ e_{\sigma(1)} & e_{\sigma(2)} & \ldots & e_{\sigma(n)} \\ | & | & \ldots & | \end{bmatrix} \\
&= \sum_{\sigma \in S_n} v_{\sigma(1),1} v_{\sigma(2),2} \cdots v_{\sigma(n),n} \operatorname{sgn}(\sigma) \\
&= \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} v_{\sigma(i),i},
\end{aligned}
$$

where the first four steps follow from (SVP-3), the sixth step from Lemma 0.145, and the fifth step from (SVP-2). In other words, the signed volume $\delta(\cdot)$ is zero for any $i_j = i_k$ and hence the only nonzero terms are those of which $(i_1, i_2, \ldots, i_n)$ is a permutation of $(1, 2, \ldots, n)$. □

Exercise 0.89. Use the formula in (0.86) to show that $\det A = \det A^T$.

Definition 0.148. The i, j *cofactor* of $A \in \mathbb{R}^{n \times n}$ is

$$C_{ij} = (-1)^{i+j} M_{ij}, \tag{0.88}$$

where M_{ij} is the i, j *minor matrix* of A, i.e. the determinant of the $(n-1) \times (n-1)$ matrix that results from deleting the i-th row and the j-th column of A.

Theorem 0.149 (Laplace formula of determinants)**.** Given fixed indices $i, j \in 1, 2, \ldots, n$, the determinant of an n-by-n matrix $A = [a_{ij}]$ is given by

$$\det A = \sum_{j'=1}^{n} a_{ij'} C_{ij'} = \sum_{i'=1}^{n} a_{i'j} C_{i'j}. \tag{0.89}$$

Exercise 0.90. Prove Theorem 0.149 by induction.