Mathematical Atlas

Eldar Akhmetov

Mathematics is the language of spaces

The room in which we sit is a space.
The set of all possible temperatures in this room is also a space.
The set of all functions describing temperature is also a space.

Mathematics studies spaces of different types and connections between them.

Each branch of mathematics is a way of looking at space:

BRANCH	WHAT IT SEES IN SPACE
Set theory	Only points, no structure
Topology	Which points are "close" (but without distances)
Metric	Distances between points
Linear algebra	Addition and multiplication by numbers
Groups	Symmetries — transformations preserving structure
Manifolds	Locally like ℝⁿ, globally curved
Funct. analysis	Infinite-dimensional spaces of functions

One and the same physical space can be studied by all methods.
Different problems require different views.

This is a map of mathematical territory. The atlas shows connections between branches.

The branch order in the table above answers the question "what does each branch see". The table below ("what is added to space") uses a different order — the sequence of enrichment, from the poorest structure to the richest, following the atlas's own scaffolding.

Philosophy of this atlas

emptiness → boundaries → space → structure → measurement

BRANCH	WHAT IS ADDED TO SPACE
Set theory	Nothing. Dust — points without connections.
Topology	Fabric — notion of "nearby", connectedness, continuity.
Groups	Mobility — a catalog of allowed motions.
Metric	Rigidity — numerical distances between points.
Linear algebra	Flatness — can add and scale.
Manifolds	Curvature — locally flat, globally curved.
Analysis	Measurement — functions as sensors on space.

In all of mathematics there is a fundamental division:

Object — that which exists independently of the method of description.
(vector, tensor, manifold, operator)

Observer — one who chooses a coordinate system and writes down numbers.
(basis, chart, reference frame)

Invariant — that on which all observers will agree.
(length, angle, rank, spectrum, topological type)

Notations

∈ ∉ ⊂ ⊆	belongs, inclusion	∀ ∃ ⇒ ⇔	quantifiers, implication
∩ ∪ \	intersection, union	¬ ∧ ∨	not, and, or
ℕ ℤ ℚ ℝ ℂ	number sets	ℍ 𝕆	quaternions, octonions
≅	isomorphism	⊗ ⊕	tensor, direct sum
⟨·,·⟩ ‖·‖	inner product, norm	V* Aᵀ A⁻¹	duality, transpose

Hierarchy of spaces — main diagram

```
                         Set
                (points without structure)
                          |
        +-----------------+-----------------+
        |                 |                 |
        ▼                 ▼                 ▼
   Topological        Algebraic        Partial order
     space             (group,           (lattice)
   (closeness)          ring)
        |                 |
        ▼                 ▼
    Metric            Vector space
   (distance)       (addition, scaling)
        |                 |
        +---------+-------+
                  |    ← here both branches converge
                  ▼
               Normed
        (norm defines metric:
         d(x,y) = ‖x − y‖)
                  |
               Banach
             (+ completeness)
                  |
             Hilbert
       (+ inner product,
        angles between vectors)
```

Second branch (from Topological space):

```
   Topological
```

```
        space
           |
           ▼
        Manifold
      (locally ≅ ℝⁿ)
           |
      Smooth manifold
    (+ diff. structure)
           |
     Riemannian manifold
    (+ metric tensor:
     inner product on
     tangent space)
```

Connection of branches:
Normed = Metric ∩ Vector: norm uniquely defines metric, and both structures are compatible.
Riemannian manifold: at each point p has inner product on T_pM — locally this is a Hilbert structure.

Why Is Calculus Not at the Beginning?

Traditional education: school → calculus → everything else.
This atlas is arranged differently.

Calculus is analysis on a concrete space $\mathbb{R}^n$.
We first answer: what is a space in general?

- Topology: what does "close" and "continuous" mean
- Linear algebra: what does "add" and "multiply by a number" mean
- Groups: what does "symmetry" mean
- Manifolds: what does "locally like $\mathbb{R}^n$" mean

and then we show: here is how all this works on $\mathbb{R}^n$ (calculus).

One problem — many languages

Heat conduction in a rod. One physics, but:

LANGUAGE	HOW IT LOOKS

Physics	Heat flows from hot to cold
Calculus	$\partial T/\partial t = \alpha\cdot\partial^2 T/\partial x^2$ (partial differential eq.)
Fourier	$T(x,t) = \Sigma c_n e^{-\alpha n^2 t}\sin(n\pi x/L)$
Funct.Anal.	$dT/dt = AT$, where $A = \alpha\cdot d^2/dx^2$ — operator in L^2
Semigroups	$T(t) = e^{At}T_0$ — one-parameter semigroup
Probability	Brownian motion, particle diffusion

All these languages describe the same thing. The atlas shows how to move between them. Sometimes a problem is simpler in one language, sometimes — in another.

Second problem — even more languages (flow in a pipe)

Water flows through a pipe. What is the flow rate? What are the pressure losses?

ATLAS SECTION	WHAT IT GIVES FOR THIS PROBLEM
Vectors	Velocity $v = (v_x, v_y, v_z)$ — vector field
Tensors	Stresses τ_{ij} — rank-2 tensor Relation τ and velocity: $\tau_{ij} = \mu(\partial v_i/\partial x_j + \partial v_j/\partial x_i)$
Forms	Flow rate $= \iint_S v\cdot dS$ — integral of 2-form $*v$
Calculus	Navier–Stokes equation: $\rho(\partial v/\partial t + (v\cdot\nabla)v) = -\nabla p + \mu\nabla^2 v + \rho g$
Funct.Anal.	Weak solutions, Sobolev spaces $W^{1,2}$
Groups	Symmetry of problem: axial → Poiseuille profile $v(r) = v_{max}(1 - r^2/R^2)$ — parabolic profile
Dimensions	Reynolds number $Re = \rho vL/\mu$ — dimensionless. $Re < 2300$: laminar, $Re > 4000$: turbulence

PART I: FOUNDATION

Philosophical foundation

Hierarchy of thinking: from emptiness to physics

All of mathematics and our cognition of the world arise from a sequence of acts of thinking:

```
        ∅ Emptiness
         |
         | Act 1: Choice
         ↓
   +============================================================+
   |  A subject is endowed with the ability to act upon emptiness  |
   +============================================================+
         |
         | Act 2: Drawing boundaries
         ↓
   +============================================================+
   |  Emptiness is cut by boundaries into images                |
   +============================================================+
         |
```

```
      | Act 3: Manipulations of images
      ↓
+=============================================================+
|  Universal to all living things                             |
|  Do not require symbols                                     |
|  Direct operation with patterns                             |
+=============================================================+
      |
      | Act 4: Categorization
      ↓
+=============================================================+
|  Set theory                                                 |
|  Minimal language for describing collections of objects    |
|  Bridge between images and communication                    |
+=============================================================+
      |
      | Act 5: Communication
      ↓
+=============================================================+
|  Natural languages                                          |
|  Symbolic representation of images                         |
|  Loss of precision in transmission                          |
+=============================================================+
      |
      | Act 6: Verification of communication
      ↓
+=============================================================+
|  Logic / Proofs                                             |
|  Needed due to unreliability of language                    |
|  Attempt to restore original clarity of images              |
+=============================================================+
      |
      | Act 7: Application to the world
      ↓
+=============================================================+
|  Physics                                                    |
|  Experimental science with high reproducibility             |
|  Mathematics = experimental physics                         |
+=============================================================+
```

Key philosophical propositions

1	The universe is emptiness, perpetually redrawing its own boundaries
2	Existence of an object = the possibility for someone to point to it
3	Thinking = indicating which sets objects belong to
4	Proof = an explicit path along the map of set embeddings
5	Logic and mathematics = experimental physics with high reproducib.
6	To understand = to be able to represent visually
7	"Object" and "space" are not properties of a thing but roles assigned by the observer's act of drawing a boundary

Object or space? — a question of point of view

One and the same entity can be both object and space — depending on where we draw the boundary of observation.

Example: donut (torus)

Donut as object (view from outside):

We look at the torus as a whole. It is one element in the space of all surfaces, alongside the sphere, double torus, etc. We are interested in its global properties: genus, area, embedding in $\mathbb{R}^3$.

Donut as space (view from inside):

We "live" on the torus. Now we are interested in points on it, paths between them, functions on it. For an ant crawling on the donut, the donut is the whole world, the space in which it moves.

Both views are correct. The donut is the same object, but the act of drawing the boundary deter

---

Recursion: space becomes object

Level 0: Point p on torus T ← p is object
Level 1: Torus T ← T is a space for p
Level 2: Moduli space of tori ← T becomes object.
Level 3: Space of all moduli ← previous becomes object
...

At each level, what was space becomes object in a space of higher level. The boundary rises.

Example from physics: liquid and gas

This example is especially important because it shows how the same physical object requires different mathematical descriptions.

APPROACH	LIQUID/GAS IS:
Thermodynamics	Object with parameters (P, V, T, S) "What is the pressure of gas in the cylinder?" Internal structure is unimportant — only state.
Hydrodynamics (Navier–Stokes equations)	Space with fields v(x,t), P(x,t), ρ(x,t) "How does liquid flow around a wing?" Each point is a place where velocity, pressure, density are defined. Flow itself = trajectories.
Kinetic theory	Space of molecules (phase space) Each molecule is object with coordinates (x, v). Gas = cloud of points in 6N-dimensional space.

Key observation:

- In thermodynamics: gas = point in state space (P,V,T)
- In hydrodynamics: gas = space itself, where fields live
- In kinetics: gas = set of particles, each of which is an object

Three different levels of description — three different answers to the question "what here is object, and what is space".

Navier–Stokes equations, Boltzmann equation, equation of state — these are not competitors, but descriptions at different levels of hierarchy.

Practical criterion

Object — when we ask "what is it like?" (properties of the whole)
Space — when we ask "what is in it?" (structure inside)

This is not a property of the thing, but a property of the question we ask.

Philosophical conclusion

"Object" and "space" are not absolute properties, but roles. The role is determined by the act of drawing a boundary by the observer.

Mathematics studies structures at all levels simultaneously — and provides language for transition between them.

Set theory — basic concepts

Set theory is a view of space as dust: there are points, and nothing more. No structure, no connections. We do not yet know which points are "nearby", we cannot add them, we cannot measure distances. Only the bare fact: this point belongs to this set, or does not belong.

This is the poorest view — but this is where everything begins. All other structures (topology, algebra, metric) will be superstructures over sets.

In terms of "object–observer": at the level of sets there is no observer yet. No coordinate system, no way to "write" an element with numbers. There are only the objects themselves and the question: belongs or not?

What is a set

A set is a collection of objects considered as a single whole.

Objects contained in a set are called its elements.

Ways of defining:
- Enumeration: A = {1, 2, 3}
- Property description: B = {x : x > 0} = "all positive x"

Special sets:
∅ = {} — empty set (contains no elements)
U — universe (set of all considered objects)

Two main relations

```
 +-----------------------------------------------------------------------+
 | x ∈ A     "x — element of set A"                                      |
 |            x is one object lying inside A                           |
 |                                                                       |
 |           +===============+                                           |
 |           |       A       |                                           |
 |           |    ●          |  ← point x inside A                       |
 |           |    x          |                                           |
 |           +===============+                                           |
```

```
+-----------------------------------------------------------------------+

+-----------------------------------------------------------------------+
| B ⊆ A     "B – subset of A"                                           |
|           Every element of B is an element of A                       |
|           (B lies entirely inside A)                                  |
|                                                                       |
|           +=======================+                                   |
|           |         A             |                                   |
|           |    +---------+        |                                   |
|           |    |    B    |        |  ← B entirely inside A            |
|           |    +---------+        |                                   |
|           +=======================+                                   |
+-----------------------------------------------------------------------+
```

Important not to confuse:
- x ∈ A – x is an object inside A
- B ⊆ A – B is a set, all elements of which are in A

Example: A = {1, 2, 3}
- 2 ∈ A – yes (2 – element of A)
- {2} ⊆ A – yes ({2} – subset of A)
- {2} ∈ A – no ({2} is not an element of A, elements are numbers)
- 2 ⊆ A – makes no sense (2 is not a set)

Empty set:
- ∅ ⊆ A – always true for any A
- ∅ ∈ A – true only if ∅ is explicitly specified as an element

```
 Operations on sets
+--------------+--------------------------+-----------------------------------+
| OPERATION    | DEFINITION               | VENN DIAGRAM                      |
+--------------+--------------------------+-----------------------------------+
|              |                          |                                   |
| A ∪ B        | {x : x ∈ A or x ∈ B}     |     .-"""-.   .-"""-.             |
| UNION        |                          |    /███████\ /███████\            |
|              | All elements that are    |   |█████████████████████|         |
|              | in at least one          |   |█████████████████████|         |
|              |                          |    \███████/ \███████/            |
|              |                          |     `-._.-"   `-._.-"             |
|              |                          |  Everything shaded                |
|              |                          |                                   |
+--------------+--------------------------+-----------------------------------+
|              |                          |                                   |
| A ∩ B        | {x : x ∈ A and x ∈ B}    |     .-"""-.   .-"""-.             |
| INTERSECTION |                          |    /       \ /       \            |
|              | All elements that are    |   |         █         |           |
|              | in both sets             |   |       █████       |           |
|              |                          |    \       █ /       /            |
|              |                          |     `-._.-" `-._.-"               |
|              |                          |  Only intersection                |
|              |                          |                                   |
+--------------+--------------------------+-----------------------------------+
|              |                          |                                   |
| A \ B        | {x : x ∈ A and x ∉ B}    |     .-"""-.   .-"""-.             |
| DIFFERENCE   |                          |    /███████\ /       \            |
|              | Elements of A that are   |   |█████████         |            |
|              | not in B                 |   |███████           |            |
|              |                          |    \███████/ \       /            |
|              |                          |     `-._.-"   `-._.-"             |
|              |                          |  A without intersection           |
|              |                          |                                   |
+--------------+--------------------------+-----------------------------------+
|              |                          |  +===========================+    |
| Aᶜ or A'     | {x : x ∉ A}              |  |█████████████████████████  |    |
| COMPLEMENT   |                          |  |██████ .-"""-. ██████████  |    |
|              | All elements of universe,|  |█████/       \ █████████   |    |
|              | not contained in A       |  |████|    A    |█████████   |    |
|              |                          |  |█████\       /██████████   |    |
|              | Aᶜ = U \ A               |  |██████ `-._.-" ██████████  |    |
|              |                          |  +===========================+    |
|              |                          |  Everything except A              |
|              |                          |                                   |
+--------------+--------------------------+-----------------------------------+
```

Laws of set theory

Commutativity:	$A \cup B = B \cup A$	$A \cap B = B \cap A$
Associativity:	$(A\cup B)\cup C = A\cup(B\cup C)$	$(A\cap B)\cap C = A\cap(B\cap C)$
Distributivity:	$A\cap(B\cup C) = (A\cap B)\cup(A\cap C)$	$A\cup(B\cap C) = (A\cup B)\cap(A\cup C)$

De Morgan's laws:

$(A \cup B)^c = A^c \cap B^c$	"not(A or B)" = "not A and not B"
$(A \cap B)^c = A^c \cup B^c$	"not(A and B)" = "not A or not B"

Properties of empty set and universe:

$A \cup \varnothing = A$	$A \cap \varnothing = \varnothing$
$A \cup U = U$	$A \cap U = A$
$A \cup A^c = U$	$A \cap A^c = \varnothing$

Cardinality of a set

$|A|$ – number of elements in finite set A.

Examples:

$|\varnothing| = 0$

$|\{a, b, c\}| = 3$

$|\{1, 2, \{1,2\}\}| = 3$ (three elements: 1, 2, and the set $\{1,2\}$)

Inclusion-exclusion formula:

$|A \cup B| = |A| + |B| - |A \cap B|$

Cardinality of power set:

If $|A| = n$, then A has 2^n subsets (including $\varnothing$ and A itself)

Cartesian product

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

The set of all ordered pairs where the first element is from A, the second from B.

Example: $\{1, 2\} \times \{a, b\} = \{(1,a), (1,b), (2,a), (2,b)\}$

Geometrically: If A and B are segments on axes, then A × B is a rectangle.

```
b +    +---------+
  |    |         |
  +    |  A × B  |
  |    |         |
a +----+---------+----
       1         2
```

$|A \times B| = |A| \cdot |B|$

$\mathbb{R} \times \mathbb{R} = \mathbb{R}^2$ – plane
$\mathbb{R} \times \mathbb{R} \times \mathbb{R} = \mathbb{R}^3$ – three-dimensional space

Logic and proofs

Logical laws are a formalization of self-evident results observed when manipulating sets – results grounded in the behavior of physical objects in reality. However, these results need not always hold.

Statements and sets

Philosophical foundation: thinking = classification by sets

> Ultimately, all our thoughts reduce to indicating which sets the conceivable objects belong to.

When we think "this table is wooden", we place the object (table) in the set (wooden things). When we think "5 is a prime number", we place 5 in the set of prime numbers.

Structure of any statement

Any statement can be decomposed into two parts:

- x – the object being discussed
- P – the set (property) to which the object belongs or does not

```
+-------------------------------------------------------------+
| Statement  =  "object x has property P"  =  x ∈ P |
+-------------------------------------------------------------+
```

Examples of decomposition:

"Socrates is mortal"
x = Socrates
P = {all mortal beings}
Statement: Socrates $\in P$

"7 is an odd number"
$x = 7$
$P = \{\text{odd numbers}\} = \{1, 3, 5, 7, 9, \ldots\}$
Statement: $7 \in P$ (true)

"This liquid is an acid"
x = this liquid
P = {acids}
Statement: $x \in P$ (requires verification)

Why this is important:

All logic is rules for working with membership of objects in sets.
"P and Q" = x belongs to both set P and set Q = $x \in (P \cap Q)$
"P or Q" = x belongs to at least one = $x \in (P \cup Q)$
"not P" = x does not belong to P = $x \in P^c$ (complement)

What is logic and why is it needed

Logic is a normative discipline: it prescribes how a person should think, but does not describe how a person actually thinks.

Goals of logic:

- To be able to prove new statements based on already known ones
- To provide means for evaluating arguments for correctness

Basic concepts

Statement – an assertion that can be verified for truth/falsity.

Examples:

✓ "2 + 2 = 4" – statement (true)
✓ "Moscow is the capital of France" – statement (false)
✗ "What time is it?" – not a statement (question)
✗ "x > 5" – not a statement while x is not defined (predicate)

Formula – a statement composed of other statements through logical connectives (∧, ∨, ¬, →, ↔) or quantifiers (∀, ∃).

Tautology – a formula that is true for any values of its constituents.

Example: $P \lor \neg P$ (law of excluded middle) – always true.
Tautologies are used to construct proof methods.

Two Levels of Logic

Propositional (logic of propositions, zero-order logic):

- Works with ready-made propositions P, Q, R, ...
- Connectives: ∧, ∨, ¬, →, ↔
- Can be verified by truth table (mechanically)

Predicate (first-order logic):

- Adds variables, predicates and quantifiers ∀, ∃
- Allows talking about properties of objects and relations between them
- Needed for mathematics (statements like "for all n")

Variables – represent elements of some set.
Predicates – functions returning true/false for objects.
  Example: P(x) = "x is even", Q(x,y) = "x < y"
Quantifiers – allow making statements about sets of objects.

```
--------------------------------------------------------------------------------
 Two Ways to Verify Logical Consequence
--------------------------------------------------------------------------------
```

Method 1: Truth table (mechanical, for propositional logic)

- Write out all combinations of variable values
- Calculate values of premises and conclusion
- Check: if premises = 1, then conclusion = 1?

Method 2: Ready-made methods (tautologies with known structure)

- Modus Ponens, Modus Tollens, proof by contradiction, etc.
- Don't require tables — use already proven tautologies

Fundamental principle:

Set theory is primary with respect to logic.

- Set theory = rules of operations on objects of any nature
- Logic = rules of operations on propositions of any language

Logic is a superstructure over set theory.
First we think in terms of membership in sets,
and only then give definitions to logical operations.

```
+--------------+-------------------------+----------------------------------+
| LOGICAL      |  DEFINITION THROUGH     |  VENN DIAGRAM                    |
| OPERATION    |  SET THEORY             |  (shaded = result)               |
+--------------+-------------------------+----------------------------------+
|              |                         |                                  |
| P ∧ Q        | x ∈ (P ∩ Q)             |       .-"""-.   .-"""-.          |
| "P and Q"    | x belongs to            |      /   P   \ /   Q   \         |
|              | intersection            |     |         █         |        |
|              |                         |     |       █████       |        |
|              |                         |      \       █ /       /         |
|              |                         |       `-._.-" `-._.-"           |
|              |                         |               ↑                  |
|              |                         |     Only this (intersection)     |
|              |                         |                                  |
+--------------+-------------------------+----------------------------------+
|              |                         |                                  |
| P ∨ Q        | x ∈ (P ∪ Q)             |       .-"""-.   .-"""-.          |
| "P or Q"     | x belongs to            |      /███████\ /███████\         |
|              | union                   |     |█████████████████████|      |
|              |                         |     |█████████████████████|      |
```

```
|              |                         |      \███████/ \███████/         |
|              |                         |       `-._.-"   `-._.-"          |
|              |                         |            ↑                     |
|              |                         |    All shaded (both circles)     |
|              |                         |                                  |
+--------------+-------------------------+----------------------------------+
|              |                         |  +===========================+   |
| ¬P           | x ∈ Pᶜ                  |  |█████████████████████████  |   |
| "not P"      | x belongs to            |  |██████ .-"""-. ██████████  |   |
|              | complement              |  |█████/       \█████████    |   |
|              |                         |  |████|    P    |█████████   |   |
|              |                         |  |█████\       /█████████    |   |
|              |                         |  |██████ `-._.-" ██████████  |   |
|              |                         |  |█████████████████████████  |   |
|              |                         |  +===========================+   |
|              |                         |     ↑ Everything except P        |
|              |                         |                                  |
+--------------+-------------------------+----------------------------------+
|              |                         |                                  |
| P ⇒ Q        | P ⊆ Q                   |       .------""""""------.       |
| "if P,       | Set P is contained      |      /          Q         \      |
|  then Q"     | in set Q                |     |    .---.             |     |
|              |                         |     |   |  P  |            |     |
|              | (every x from P         |     |    `---"             |     |
|              |  is necessarily in Q)   |      \                    /      |
|              |                         |       `------.-----------"       |
|              | Equivalent: ¬P ∨ Q      |     P entirely inside Q          |
|              |                         |                                  |
+--------------+-------------------------+----------------------------------+
|              |                         |                                  |
| P ⇔ Q        | P = Q                   |         .-------.                |
| "P if and    | Sets coincide           |        /         \               |
| only if      |                         |       |   P = Q   |              |
| Q"           | (P ⊆ Q) ∧ (Q ⊆ P)       |        \         /               |
|              |                         |         `-------"                |
|              |                         |     One and the same circle.     |
|              |                         |                                  |
+--------------+-------------------------+----------------------------------+
|              |                         |                                  |
| ⊤            | U (universe)            |  +===========================+   |
| "true"       |                         |  |█████████████████████████  |   |
|              | x ∈ U  for any x        |  |█████████████████████████  |   |
|              |                         |  |█████████ U █████████████  |   |
|              |                         |  |█████████████████████████  |   |
|              |                         |  +===========================+   |
|              |                         |     Entire space                 |
|              |                         |                                  |
+--------------+-------------------------+----------------------------------+
|              |                         |                                  |
| ⊥            | ∅ (empty set)           |      { }                         |
| "false"      |                         |      (nothing there)             |
|              | x ∈ ∅ never             |                                  |
|              |                         |                                  |
+--------------+-------------------------+----------------------------------+
```

Quantifiers (for predicate logic):

```
+--------------+-------------------------+----------------------------------+
| QUANTIFIER   |  DEFINITION             |  VISUALIZATION                   |
+--------------+-------------------------+----------------------------------+
|              |                         |                                  |
| ∀x ∈ A: P(x) | P(x) is true for all x  |  +===========+                   |
|              | from set A              |  | · · · · · | ← every point     |
| "for all x   |                         |  | · · · · · |   in A has        |
|  from A      | Equivalent: A ⊆ P       |  | · · · · · |   property P      |
|  P(x) holds" | (A is contained in set  |  +===========+                   |
|              |  of points with         |       A ⊆ P                      |
|              |  property P)            |                                  |
|              |                         |                                  |
+--------------+-------------------------+----------------------------------+
|              |                         |                                  |
| ∃x ∈ A: P(x) | P(x) is true for at     |  +===========+                   |
|              | least one x from A      |  |           |                   |
| "there exists|                         |  |     ●     | ← at least one    |
|  x from A    | Equivalent: A ∩ P ≠ ∅   |  |           |   point with P    |
|  with        | (intersection of A and  |  +===========+                   |
|  property P" |  set of points with P   |      A ∩ P ≠ ∅                   |
|              |  is nonempty)           |                                  |
|              |                         |                                  |
+--------------+-------------------------+----------------------------------+
```

Correspondence: logic ↔ set theory

LOGIC	SET THEORY
Proposition P	Set {x : P(x) is true}
Logical consequence P ⊢ Q	Set inclusion: P ⊆ Q
Tautology	Universal set
Contradiction	Empty set ∅

Tabular method for propositional logic (zeroth order)

Allows mechanical verification of any formula without creativity
Works only for finite number of variables, without quantifiers ∀, ∃

Truth table of logical operations

P	Q	P ∧ Q	P ∨ Q	¬P	P → Q	P ↔ Q
0	0	0	0	1	1	1
0	1	0	1	1	1	0
1	0	0	1	0	0	0
1	1	1	1	0	1	1

Algorithm for classifying formulas

Problem 1: determining the type of formula

Problem: Determine the type of formula F

Input: formula F with variables $P_1, P_2, ..., P_n$
Output: tautology / contradiction / satisfiable

Algorithm for tabular verification

STEP	ACTION
1	Write out variables: $P_1, P_2, ..., P_n$
2	Build a table with 2^n rows
3	Compute F for each row
4	Classify the result

Classification

RESULT	TYPE OF FORM
All rows = 1	Tautology
All rows = 0	Contradiction
Mixed	Satisfiable

Example 1.1: Verify $(P \to Q) \leftrightarrow (\neg P \lor Q)$

P	Q	P → Q	¬P	¬P ∨ Q	(P→Q) ↔ (¬P ∨ Q)
0	0	1	1	1	1
0	1	1	1	1	1
1	0	0	0	0	1
1	1	1	0	1	1

All rows = 1 → tautology

Example 1.2: Verify $P \land \neg P$

P	¬P	P ∧ ¬P
0	1	0
1	0	0

```
+---+----+--------+
```

All rows = 0 → contradiction

Example 1.3: Verify de Morgan's law $\neg(P \wedge Q) \leftrightarrow (\neg P \vee \neg Q)$

P	Q	P ∧ Q	¬(P ∧ Q)	¬P	¬Q	¬P ∨ ¬Q	↔
0	0	0	1	1	1	1	1
0	1	0	1	1	0	1	1
1	0	0	1	0	1	1	1
1	1	1	0	0	0	0	1

All rows = 1 → tautology

Problem 2: checking entailment

Problem: determine whether B follows from premises $A_1, A_2, \ldots, A_k$

Notation: $A_1, A_2, \ldots, A_k \vdash B$

Key point: comma between premises means ∧ (logical AND).

$A_1, A_2, \ldots, A_k \vdash B$ means $(A_1 \wedge A_2 \wedge \ldots \wedge A_k) \rightarrow B$

Algorithm:

Step 1. Write out all variables from all formulas $A_1, \ldots, A_k, B$
Step 2. Build a table with 2^n rows
Step 3. For each row compute $A_1, A_2, \ldots, A_k$, their conjunction, and B
Step 4. Search for counterexample: a row where $(A_1 \wedge \ldots \wedge A_k) = 1$, but $B = 0$

- Counterexample found → entailment is false
- No counterexample → entailment is true

Example 2.1: Verify modus ponens $(P \rightarrow Q) \wedge P \vdash Q$

Denote: $A_1 = (P \rightarrow Q)$, $A_2 = P$, $B = Q$

P	Q	A_1	A_2	$A_1 \wedge A_2$	B	Verification
0	0	1	0	0	0	$A_1 \wedge A_2=0$
0	1	1	0	0	1	$A_1 \wedge A_2=0$
1	0	0	1	0	0	$A_1 \wedge A_2=0$
1	1	1	1	1	1	$A_1 \wedge A_2=1$, B=1 ✓

The only row where $A_1 \wedge A_2=1$: P=1, Q=1.
In it B=1. No counterexample → entailment is true.

Example 2.2: Verify modus tollens $(P \to Q) \land \neg Q \vdash \neg P$

Denote: $A_1 = (P \to Q)$, $A_2 = \neg Q$, $B = \neg P$

P	Q	A_1	A_2	$A_1 \land A_2$	B	Verification
0	0	1	1	1	1	$A_1 \land A_2=1$, $B=1$ ✓
0	1	1	0	0	1	$A_1 \land A_2=0$
1	0	0	1	0	0	$A_1 \land A_2=0$
1	1	1	0	0	0	$A_1 \land A_2=0$

The only row where $A_1 \land A_2=1$: $P=0$, $Q=0$.
In it $B=1$. No counterexample → entailment is true.

Example 2.3: Verify invalid entailment $(P \to Q) \land Q \vdash P$

Denote: $A_1 = (P \to Q)$, $A_2 = Q$, $B = P$

P	Q	A_1	A_2	$A_1 \land A_2$	B	Verification
0	0	1	0	0	0	$A_1 \land A_2=0$
0	1	1	1	1	0	$A_1 \land A_2=1$, $B=0$ ✗
1	0	0	0	0	1	$A_1 \land A_2=0$
1	1	1	1	1	1	$A_1 \land A_2=1$, $B=1$ ✓

Row $P=0$, $Q=1$: $A_1 \land A_2=1$, but $B=0$.
Counterexample found → entailment is false.
(This is the fallacy "affirming the consequent")

Limitations of the method

- Works only for propositional logic (zeroth-order)
- Does not work for predicate logic (with ∀, ∃)
- With n variables requires 2^n rows (exponential growth)
- For $n = 10$ already 1024 rows, for $n = 20$ – over a million

Connection with set theory:

Truth table – enumeration of all points in the space $\{0,1\}^n$.
Formula defines a subset (where it is true).

Tautology = entire space.
Contradiction = empty set.
Entailment $(A_1 \land \ldots \land A_k) \vdash B$ holds $\iff$ set $(A_1 \land \ldots \land A_k) \subseteq$ set B.

Proof methods — with concrete problem examples

What does "prove" mean in mathematics?

Intuitively:
To prove = to convince any reasonable person, following the rules that they accepted in advance.

Formally:
A proof is a finite sequence of statements, where each:

- is either an axiom (accepted without proof)
- or follows from previous ones by rules of logic

The last statement = what we are proving.

What a proof does not do:

✗ does not explain "why this is true" (intuition does this)
✗ does not show how it was discovered (this is history)
✗ does not guarantee understanding (one can verify a proof without understanding)

A proof is verification, not explanation.
A good proof explains, a bad one — only convinces.

Types of statements and what is needed for proof

STATEMENT	HOW TO PROVE	HOW TO REFUTE
∀x: P(x) "for all"	Prove for arbitrary x	one counterexample
∃x: P(x) "exists"	one example	Prove for all that P(x) is false
∀x∃y: P(x,y) "for all exists"	For arbitrary x exhibit such a y	Find x for which there is no such y
∃x∀y: P(x,y) "exists for all"	Exhibit such an x holding for all y	For any x find y violating P

Important: Examples never prove ∀-statements.
Even a million examples do not prove that "for all".
But one counterexample refutes.

Constructive vs non-constructive proofs

Constructive: Presents the object explicitly.
  Example: "There exists an irrational number" – here is √2, and here is the proof.

Non-constructive: Proves existence without showing the object.
  Example: "There exist irrational a, b such that a^b is rational."
    Consider √2^√2.
    If rational – done: take a = b = √2.
    If irrational – take a = √2^√2, b = √2;
    then a^b = (√2^√2)^√2 = $\sqrt{2}^2$ = 2 – rational.
    We don't know which of the two cases holds, but one of them definitely does.

How to read proofs (practical advice)

1. Write separately: what is given, what we are proving
2. At each step ask: "where does this follow from?"
3. Look for the key idea – usually one trick, the rest is technique
4. Try to break the proof – where does it use the conditions?
5. After reading – retell in your own words

If you cannot explain in simple words – you have not understood.

Proof = explicit description of a path along the map of inclusions and intersections of sets, leading to a statement about membership of an object in a set.

```
+------------------+---------------------+-----------------------------------+
|  METHOD          |  STRUCTURE          |  GEOMETRIC INTERPRETATION         |
+------------------+---------------------+-----------------------------------+
|                  |                     |                                   |
| Modus ponens     |  1. P → Q (true)    |  If object in P, and P contained  |
| (inference rule) |  2. P (true)        |  in Q, then object in Q           |
|                  |  ---------------    |                                   |
|                  |  ∴ Q (true)         |     +-------------+               |
|                  |                     |     |      Q      |               |
|                  | "If P, then Q.      |     |  +-----+    |               |
|                  |  P is true.         |     |  |  P ●|    |               |
|                  |  Therefore, Q"      |     |  +-----+    |               |
|                  |                     |     +-------------+               |
|                  |                     |     object ● ∈ P ⊂ Q              |
|                  |                     |                                   |
| Example problem: | Prove: if n         | Solution:                         |
|                  | is divisible by 6,  | 1. 6 | n ⇒ n = 6k                 |
|                  | then n is divisible | 2. n = 6k = 3·(2k)                |
|                  | by 3                | 3. ⇒ 3 | n  ✓                     |
|                  |                     | {n : 6|n} ⊂ {n : 3|n}             |
|                  |                     |                                   |
+------------------+---------------------+-----------------------------------+
|                  |                     |                                   |
| Modus tollens    |  1. P → Q           |  If object not in Q, and P ⊂ Q,   |
| (denying the     |  2. ¬Q              |  then object not in P             |
|  consequent)     |  ---------------    |                                   |
```

```
|                  |  ∴ ¬P                |     +-------------+            |
|                  |                      |     |      Q      |            |
|                  | "If P, then Q.       |     |  +-----+    |            |
|                  |  Q is false.         |     |  |  P  |    |   ● not in Q |
|                  |  Therefore, P        |     |  +-----+    |            |
|                  |  is false"           |     +-------------+            |
|                  |                      |     ⇒ ● cannot be in P         |
|                  |                      |                                |
+------------------+----------------------+--------------------------------+
|                  |                      |                                |
| REDUCTIO AD      |  Want to prove Q     |  Assume ¬Q, arrive at          |
| ABSURDUM         |                      |  contradiction                 |
| (proof by        |  1. Assume ¬Q        |                                |
| contradiction)   |  2. Derive           |  +----------+                  |
|                  |     contradiction P∧¬P|  |Universe  |                  |
|                  |  ---------------     |  |  +--+     |                 |
|                  |  ∴ Q is true         |  |  |Q | ¬Q | ← empty.          |
|                  |                      |  |  +--+     |                 |
|                  | "Suppose the         |  +----------+                  |
|                  |  opposite.           |  If ¬Q empty, then all in Q    |
|                  |  Get absurdity.      |                                |
|                  |  Therefore, original |                                |
|                  |  is true"            |                                |
|                  |                      |                                |
| Example problem: | Prove √2 ∉ ℚ         | Solution:                      |
|                  |                      | 1. Assume √2 ∈ ℚ               |
|                  |                      | 2. √2 = p/q (in lowest terms)  |
|                  |                      | 3. 2q² = p² ⇒ p even (=2k)     |
|                  |                      | 4. 2q² = 4k² ⇒ q² = 2k²        |
|                  |                      | 5. ⇒ q even                    |
|                  |                      | 6. But p,q both even —         |
|                  |                      |    contradiction               |
|                  |                      | 7. ⇒ √2 ∉ ℚ  ✓                 |
|                  |                      |                                |
+------------------+----------------------+--------------------------------+
|                  |                      |                                |
| direct           |  Given: premises     |  Direct path from premises to  |
| proof            |  P₁, P₂, ..., Pₙ     |  conclusion through inclusions |
|                  |                      |                                |
|                  |  Goal: Q             |  P₁ → P₂ → ... → Pₙ → Q        |
|                  |                      |                                |
|                  |  Build chain:        |  Each arrow = inclusion        |
|                  |  P₁ ⇒ ... ⇒ Pₙ ⇒ Q   |  of sets or logical            |
|                  |                      |  entailment                    |
|                  |                      |                                |
| Example problem: | Prove: n even        | Solution:                      |
|                  | ⇒ n² even            | 1. n even ⇒ n = 2k             |
|                  |                      | 2. n² = (2k)² = 4k²            |
|                  |                      | 3. = 2·(2k²) = 2m, m=2k²       |
|                  |                      | 4. ⇒ n² even  ✓                |
|                  |                      |                                |
```

```
+------------------+----------------------+----------------------------------+
|                  |                      |                                  |
| mathematical     |  Base: P(0)          |  Dominoes fall:                  |
| INDUCTION        |  Step: P(n) ⇒ P(n+1) |                                  |
|                  |  -------------       |  ●→●→●→●→●→...                   |
|                  |  ∴ ∀n: P(n)          |  0  1  2  3  4                   |
|                  |                      |                                  |
|                  | "True for 0.         |  If pushed 0, and each           |
|                  |  If true for n,      |  pushes the next, then all       |
|                  |  then for n+1.       |  will fall                       |
|                  |  Therefore, true for |                                  |
|                  |  all natural numbers"|                                  |
|                  |                      |                                  |
| Example problem: | Prove                | Solution:                        |
|                  | 1+2+...+n=n(n+1)/2   | Base: n=1: 1=1·2/2=1  ✓          |
|                  |                      | Step: Assume true for n.         |
|                  |                      | For n+1:                         |
|                  |                      | 1+...+n+(n+1) =                  |
|                  |                      | = n(n+1)/2 + (n+1)               |
|                  |                      | = (n+1)(n/2+1)                   |
|                  |                      | = (n+1)(n+2)/2  ✓                |
|                  |                      |                                  |
+------------------+----------------------+----------------------------------+
|                  |                      |                                  |
| case analysis    |  P ∨ Q               |  Universe = P ∪ Q                |
|                  |  P ⇒ R               |                                  |
|                  |  Q ⇒ R               |  +--------------+                |
|                  |  -------------       |  |   R          |                |
|                  |  ∴ R                 |  | +---+  +---+ |                |
|                  |                      |  | | P |  | Q | |                |
|                  | "Either P or Q.      |  | +---+  +---+ |                |
|                  |  in both cases       |  +--------------+                |
|                  |  R follows"          |  Both cases lead to R            |
|                  |                      |                                  |
+------------------+----------------------+----------------------------------+
|                  |                      |                                  |
| contraposition   |  P → Q               |  P ⊂ Q equivalent to             |
|                  |  =========           |  Qᶜ ⊂ Pᶜ                         |
|                  |  ¬Q → ¬P             |                                  |
|                  |                      |  +------------+                  |
|                  | "If Q follows from P,|  |  Q    Qᶜ   |                  |
|                  |  then NOT-P follows  |  |+--+  +--+  |                  |
|                  |  from NOT-Q"         |  ||P |  |Pᶜ|  |                  |
|                  |                      |  |+--+  +--+  |                  |
|                  | Equivalent           |  +------------+                  |
|                  | formulations.        |                                  |
|                  |                      |                                  |
| Example problem: | Prove: n² odd        | Solution (contraposition):       |
|                  | ⇒ n odd              | Prove: n even ⇒ n² even          |
|                  |                      | (we already know this)           |
|                  |                      | By contraposition:               |
|                  |                      | n² odd ⇒ n odd  ✓                |
|                  |                      |                                  |
+------------------+----------------------+----------------------------------+
```

Key idea:

All proof methods reduce to demonstrating that some object is located in sets nested within each other.

To prove Q = to show a path of inclusions leading to Q.

Geometric interpretation of proof methods

Each proof method has a simple visualization through nestedness of sets on a Venn diagram. This makes logic intuitively understandable.

Modus ponens: "If P, then Q. P is true. Therefore, Q is true."

Formula: $P \to Q,\ P \vdash Q$

Geometrically:
If $P \subset Q$ (circle P completely inside circle Q)
and object $x \in P$
then $x \in Q$ (obviously)

```
+================================+
|            Q                   |
|      +------------+            |
|      |     P      |            |
|      |   ● x      |            |  ← x in P, but P inside Q
|      +------------+            |    ⇒ x automatically in Q
+================================+
```

Example: "All students have a student ID."
"Masha is a student."
⇒ "Masha has a student ID."

Reductio ad absurdum: "Assume the opposite. We get absurdity."

Formula: Assume $\neg Q$. Derive contradiction. Therefore, Q.

Geometrically:
Assume $P \subset Q$ (P inside Q).
Suppose: object x belongs to P and simultaneously does not belong to Q.
But this is a contradiction. (x cannot be in P and outside Q, if $P \subset Q$)
Therefore, if $x \in P$, then $x \in Q$.

```
+================================+
|            Q         ████████  | ← Qᶜ (complement)
|      +------------+  █ ¬Q  ██  |
|      |     P      |  ████████  |
|      |            |  ████████  |
|      +------------+  ████████  |
+================================+
```

Attempt to place a point simultaneously in P and in Q^c – contradiction. These regions do not intersect.

Example: "Prove $\sqrt{2} \notin \mathbb{Q}$"
Assume $\sqrt{2} \in \mathbb{Q} \Rightarrow$ we get contradiction $\Rightarrow \sqrt{2} \notin \mathbb{Q}$

Philosophical remark

Proofs are not the pinnacle of mathematics, but a crutch to compensate for losses during linguistic transmission. In an ideal world of telepaths, where one can directly "show" the geometry of inclusions, formal proofs would not be needed.

Formalization of mathematics is not rigor in itself, but a forced measure during the transition to linguistic description. Mathematics itself exists at the pre-linguistic level of manipulations with images and inclusions of sets.

Unity: logic = sets = order

Preliminarily: What is order?

Partial order $\leq$ on set X is a relation that:

- Reflexive: $x \leq x$ for any x
- Antisymmetric: $x \leq y$ and $y \leq x \Rightarrow x = y$
- Transitive: $x \leq y$ and $y \leq z \Rightarrow x \leq z$

Examples: $\subseteq$ on sets, $\leq$ on numbers, "divides" on naturals.

Logic, sets and order are three languages for describing one structure.

This is not an analogy. This is an identity. Stone's theorem (1936) proves:
Boolean algebra ≅ algebra of sets ≅ propositional logic

One object – three languages

LOGIC	SETS	ORDER/LATTICE
Statement P	Set A	Element a
$P \land Q$ (AND)	$A \cap B$ (intersection)	$a \land b$ (meet, inf)
$P \lor Q$ (or)	$A \cup B$ (union)	$a \lor b$ (join, sup)
$\neg P$ (NOT)	A^c (complement)	$\neg a$ (complement)
$P \Rightarrow Q$ (implies)	$A \subseteq B$ (inclusion)	$a \leq b$ (order)
truth $\top$	Universe U	Maximum 1
false $\bot$	Empty $\varnothing$	Minimum 0

Laws – the same in three languages

De Morgan:	$\neg(P\land Q) = \neg P\lor\neg Q$	$(A\cap B)^c = A^c\cup B^c$	analogously
	$\neg(P\lor Q) = \neg P\land\neg Q$	$(A\cup B)^c = A^c\cap B^c$	
Distributivity:	$P\land(Q\lor R)=(P\land Q)\lor(P\land R)$	$A\cap(B\cup C)=(A\cap B)\cup(A\cap C)$	
	$P\lor(Q\land R)=(P\lor Q)\land(P\lor R)$	$A\cup(B\cap C)=(A\cup B)\cap(A\cup C)$	
Excluded middle:	$P \lor \neg P = \top$	$A \cup A^c = U$	$a \lor \neg a = 1$
Contradiction:	$P \land \neg P = \bot$	$A \cap A^c = \varnothing$	$a \land \neg a = 0$

These are not "similar laws" – this is one law, written in three ways.

Lattice and Boolean algebra

Lattice = partial order, where any two elements have:

$a \lor b = \sup\{a, b\}$ (least upper bound, join)

$a \land b = \inf\{a, b\}$ (greatest lower bound, meet)

Boolean algebra = lattice with complement:
For each a there exists $\neg a$: $a \vee \neg a = 1$ and $a \wedge \neg a = 0$

Example	∨ (join)	∧ (meet)
Subsets 2^X	$A \cup B$	$A \cap B$
Divisors of n	LCM(a, b)	GCD(a, b)
Statements	$P \vee Q$ (or)	$P \wedge Q$ (AND)
Open sets	$U_1 \cup U_2$	$U_1 \cap U_2$

Important: Divisors of 12 — lattice, but not Boolean. (no complement for 2, 3, 4, 6)
Subsets — Boolean algebra (complement A^c always exists)

Intuitionistic logic and topology

The table above is true for classical logic (Boolean algebra).
But there is intuitionistic logic, where $\neg\neg P \neq P$ and $P \vee \neg P$ is not always true.

CLASSICAL	INTUITIONISTIC
Boolean algebra	Heyting algebra
Subsets	Open sets of topology
$P \vee \neg P = \top$ (always)	$P \vee \neg P \neq \top$ (not always)
$\neg\neg P = P$	$\neg\neg P \neq P$
$A \cup A^c = U$	$U \cup \mathrm{Int}(U^c) \neq X$ in general

Why open sets:

- $\neg U = \mathrm{Int}(U^c)$ = interior of complement (not the complement itself)
- $U \cup \neg U = U \cup \mathrm{Int}(U^c)$ may not equal X
- Example: $U = (0,1)$ on $\mathbb{R}$. Then $\neg U = \mathrm{Int}((-\infty,0]\cup[1,\infty)) = (-\infty,0)\cup(1,\infty)$ and $U \cup \neg U = \mathbb{R} \setminus \{0,1\} \neq \mathbb{R}$.

Connection with constructivism:
In intuitionistic logic "exists" = "can be constructed".
The law of excluded middle ($P \vee \neg P$) is not valid, because
from impossibility to construct a counterexample does not follow the existence of an example.

Axiomatic Method — The Rules of the Game in Mathematics

What is an Axiom

An axiom is a statement that is accepted without proof.

Why without proof? Because any proof relies on some preceding statements. If we require proving everything, we get an infinite chain or a closed circle. A point of support is needed.

Analogy: Rules of a Board Game

- You don't "prove" that in chess a knight moves in an L-shape
- This is a rule of the game – we accept it in order to play
- Axioms = rules of the mathematical game
- Theorems = everything that can be derived from the rules

Important: Axioms are not "true" or "false" in an absolute sense.
They are agreements. Different sets of axioms give different mathematics.

Why the Axiomatic Method is Needed

History: Until the end of the 19th century mathematicians worked "intuitively".
Then paradoxes were discovered – logical contradictions.

Russell's Paradox (1901):

Let $R = \{x : x \notin x\}$ – "the set of all sets not containing themselves"

Question: $R \in R$ or $R \notin R$?

- If $R \in R$, then by definition of R we must have $R \notin R$ - contradiction
- If $R \notin R$, then by definition of R we must have $R \in R$ - contradiction

Conclusion: One cannot simply create "the set of everything satisfying a condition". Restrictions are needed – axioms defining which sets are "legal".

Solution: The ZFC axiom system (Zermelo–Fraenkel with axiom of choice)

ZFC Axioms – Foundation of Modern Mathematics

ZFC = Zermelo–Fraenkel + Choice (Zermelo–Fraenkel + Choice)

This is a system of ~9 axioms on which almost all of modern mathematics is built.

Notation

$\forall x$ – "for all x" (universal quantifier)
$\exists x$ – "there exists x" (existential quantifier)
$\rightarrow$ – "implies", "entails"
$\leftrightarrow$ – "equivalent", "if and only if"
$\wedge$ – "and" (conjunction)
$\vee$ – "or" (disjunction)

ZFC Axioms — Summary Table

AXIOM	FORMULA	WHAT IT GIVES
1. Extensional Extensionality	$(\forall x: x \in A \leftrightarrow x \in B) \to A=B$	Set = its elements. $\{1,2,3\} = \{3,1,2\}$
2. Empty Set Empty Set	$\exists \varnothing: \forall x: x \notin \varnothing$	"Zero" of set theory. Starting point for constructs
3. Pairing Pairing	$\forall a \forall b\ \exists P: x \in P \leftrightarrow (x=a \lor x=b)$	Can create $\{a,b\}$. Consequence: $\{a\}=\{a,a\}$
4. Union Union	$\forall A\ \exists U: x \in U \leftrightarrow \exists B(B \in A \land x \in B)$	Can merge sets. $\cup\{\{1,2\},\{3\}\} = \{1,2,3\}$
5. Power Set Power Set	$\forall A\ \exists P: B \in P \leftrightarrow B \subseteq A$	Set of all subsets. $\mathcal{P}(\{1,2\}) = \{\varnothing,\{1\},\{2\},\{1,2\}\}$ $\lvert\mathcal{P}(A)\rvert = 2^{\lvert A\rvert}$
6. Infinity Infinity	$\exists I: \varnothing \in I \land (x \in I \to x \cup \{x\} \in I)$	Infinite sets exist. Minimal such $I = \mathbb{N}$
7. Separation Separation	$\forall A\ \exists B: x \in B \leftrightarrow (x \in A \land \varphi(x))$	$\{x \in A: \varphi(x)\}$ — legal. $\{x: \varphi(x)\}$ — illegal. (avoiding Russell's paradox)
8. Replacement Replacement	Image of set under function — a set	$\{f(a): a \in A\}$ — a set. Needed for large cardinals
9. Foundation Foundation	$A \neq \varnothing \to \exists x \in A: x \cap A = \varnothing$	No $x \in x$, no cycles $a \in b \in a$ Everything built "from below" from $\varnothing$

Important subtlety: axiom schemas

Axioms 7 (Separation) and 8 (Replacement) are not single axioms, but schemas: for each formula $\varphi(x)$ we get its own axiom.

- "Separation with $\varphi(x)$ = (x is even)" — one axiom
- "Separation with $\varphi(x)$ = (x is prime)" — another axiom
- ...and so on for each possible formula

Therefore ZFC formally contains infinitely many axioms.
This is not a problem — we can still verify any concrete proof in finite time.

Russell's Paradox and how ZFC resolves it

Paradox: Let $R = \{x : x \notin x\}$. Then $R \in R \iff R \notin R$. Contradiction.

Solution in ZFC:
The Axiom of Separation forbids constructing $\{x : \varphi(x)\}$.
Only $\{x \in A : \varphi(x)\}$ is allowed – a subset of an already existing A.

To construct $R = \{x : x \notin x\}$, one would first need to have a set A containing R. But such A does not exist (by Regularity $x \notin x$ always).

Conclusion: ZFC is "freedom with responsibility". One cannot create sets "out of thin air" – only from already constructed ones.

Philosophical meaning of axioms

AXIOM	PHILOSOPHICAL INTERPRETATION
Empty set	Formalization of "emptiness" – we postulate its existence
Infinity	Infinity is not obvious, but a choice to play this kind of mathematics. One can build mathematics without it
Separation	Restriction on "naive" set creation – lesson from Russell's paradox. Freedom with responsibility
Regularity	Everything is built hierarchically from ∅. No "hanging in the air" or self-referential constructions

Visualization: cumulative hierarchy – "tower of sets"

The ZFC axioms generate all of mathematics from nothing (the empty set).
This happens level by level – the cumulative hierarchy V_α.

```
V_ω+1  | 𝓟(V_ω) – power set of the infinite set
       | Includes ℝ, all functions ℕ → ℕ, ...
       | This is already uncountable.
-------+---------------------------------------------------------
       |
V_ω    | V₀ ∪ V₁ ∪ V₂ ∪ ... – first infinite level
       | Includes ℕ, all finite sets, all finite structures
-------+---------------------------------------------------------
 ⋮      | ⋮
V₃     | 𝓟(V₂) = all subsets of V₂
       | |V₃| = 2² = 4 elements
```

```
-------+-----------------------------------------------------------
V₂     | 𝒫(V₁) = {∅, {∅}}
       | 2 elements — can be identified with {0, 1}
-------+-----------------------------------------------------------
V₁     | 𝒫(V₀) = 𝒫(∅) = {∅}
       | 1 element — this is "1" in the construction of natural numbers
-------+-----------------------------------------------------------
V₀     | ∅ — empty set, beginning of everything
       | 0 elements — this is "0"
=======╧===========================================================
```

Size growth:

$|V_0| = 0$
$|V_1| = 2^0 = 1$
$|V_2| = 2^1 = 2$
$|V_3| = 2^2 = 4$
$|V_4| = 2^4 = 16$
$|V_5| = 2^{16} = 65536$
$|V_6| = 2^{65536} \approx 10^{19728}$ — already inconceivable.
...
$|V_\omega| = \aleph_0$ (countable infinity)
$|V_{\omega+1}| = 2^{\aleph_0} = |\mathbb{R}|$ (continuum)

Conclusion: All of mathematics "grows" from ∅ by applying one operation 𝒫.

Where familiar objects live

OBJECT	FIRST LEVEL WHERE IT APPEARS
$\varnothing = 0$	V_0
$1 = \{\varnothing\}$	V_1
$2 = \{\varnothing, \{\varnothing\}\}$	V_2
n (any finite)	V_n
ℕ (as a set)	V_ω
Function f: ℕ → ℕ	$V_{\omega+1}$ (as subset of ℕ×ℕ)
ℝ (as Dedek. set)	$V_{\omega+1}$
Function f: ℝ → ℝ	$V_{\omega+2}$
Space C[0,1]	$V_{\omega+2}$

Almost all "working" mathematics lives in $V_{\omega+\omega}$ — relatively low.
Large cardinals (inaccessible, measurable, ...) require very high V_α.

Axiom of choice — special status

Axiom of choice (AC)

Formula: $\forall\mathcal{A}: (\forall A\in\mathcal{A}: A\neq\varnothing) \to \exists f: \forall A\in\mathcal{A}: f(A)\in A$

Meaning: For any family of nonempty sets there exists a function
choosing one element from each.

The word "simultaneously" is psychological, not mathematical.
Mathematically: a choice function f exists.
There is no "process of choosing" — the function simply exists (or doesn't).

```
        +---+  +---+  +---+  +---+
        | ● |  | ▲ |  | ■ |  | ◆ |  ...  infinitely many boxes
        | ○ |  | △ |  | □ |  | ◇ |
        +---+  +---+  +---+  +---+
          ↓      ↓      ↓      ↓
          ●      ▲      ■      ◆    ← choice function f exists
```

Without AC in some models of ZF: the product of nonempty sets is empty.
This is not "we cannot choose", but "a choice literally does not exist".

Why the axiom of choice is non-obvious

CASE	SITUATION
Finite number of boxes	Obvious — just enumerate (AC not needed)
Countable number	Axiom of countable choice (AC_ω) suffices — weaker than full AC
Uncountable number	How to choose? No algorithm. AC asserts that the function exists, even without explicit construction

What follows from the axiom of choice

Good consequences:

- Any vector space has a basis
- Any ideal is contained in a maximal ideal
- Tychonoff's theorem: product of compact spaces is compact
- Hahn–Banach theorem: extension of linear functionals
- Any two cardinals are comparable: $|A| \le |B|$ or $|B| \le |A|$

Paradoxical consequences:

- Banach–Tarski paradox:
  A ball can be cut into 5 pieces and assembled into two identical balls.
  (pieces are not "ordinary" pieces, but pathological sets of points)
- There exist non-measurable sets (Lebesgue)

Status:
The axiom of choice is independent of ZF – it can neither be proved nor refuted. One can work with it (ZFC) or without it (ZF). Most mathematicians accept AC.

"Choice" in philosophy vs "choice" in the axiom

In the philosophical foundation "choice" = the subject's ability to distinguish, draw boundaries, extract objects from emptiness. This is ontological choice.

In the axiom of choice "choice" = existence of a function assigning an element to each set. This is a technical possibility.

Connection:
Philosophical choice is primary (without it there is no mathematics at all).
Axiom of choice is a specific rule of the game, which can be accepted or not.

Difference:
Philosophical choice: "I can extract an object"
Axiom of choice: "For any (even uncountable) family there exists a method of simultaneous choice"

Why all this is needed

ZFC axioms are not "truths about the world".
They are rules of the game which:

1. Are strong enough to construct all known mathematics
2. Are restricted enough to avoid paradoxes
3. Are (as far as is known) consistent
4. Are incomplete (Gödel's theorem, 1931): there exist true statements about natural numbers that are unprovable in ZFC

Incompleteness is not a bug, but a fundamental property of any sufficiently rich formal system. ZFC cannot prove its own consistency from within.

When further on we write "there exists a set X" or "let us choose an element from Y", we rely on these axioms. Mathematics is the consequences of these rules.

Now, having the foundation, we can construct numbers.

Gödel's theorems — boundaries of formal systems

Context:

Gödel's theorems (1931) are not abstract logic. They are fundamental results about what mathematics can and cannot do:

- Why some problems are fundamentally unsolvable
- Why a machine cannot completely replace a mathematician
- Why "truth" and "provability" are different things
- Why we cannot be certain of the consistency of mathematics

For an engineer: if a problem is unsolvable — it's not our fault, but a boundary of theory.

First Gödel incompleteness theorem

Formulation (simplified):

In any consistent formal system rich enough to describe arithmetic of natural numbers, there exists a statement G which:
- is true (in the standard interpretation)
- is unprovable in this system

Intuition — "Liar paradox":

Consider the sentence: "This sentence is false."
If it is true — then it is false (contradiction).
If it is false — then it is true (contradiction).
Paradox. This sentence has no truth value.

Gödel constructed a mathematical analogue:
"This statement is unprovable in system S."

If it is provable — we proved a falsehood (system is inconsistent).
If the system is consistent — it is unprovable.
But then it is true (since it says it is unprovable).

Proof idea:

1. Gödel encoded formulas as numbers (Gödel numbering)
   Each symbol — a number, formula — product of powers of primes

2. Proofs also became numbers
   "x is a proof of formula y" is an arithmetic relation.

3. Constructed a formula G saying about itself:
   G = "There does not exist a number encoding a proof of formula G"

Corollary: ZFC (and any "reasonable" system) is incomplete.
There exist true statements about numbers which ZFC will not prove.

Second Gödel incompleteness theorem

Formulation:

A consistent system rich enough for arithmetic
cannot prove its own consistency.

Intuition:

The statement Con(S) = "System S is consistent" can be written
as an arithmetic formula (via Gödel numbering).

Gödel showed: $\mathrm{Con}(S) \to G$
(If S is consistent, then G is true)

If S proved Con(S), it would prove G.
But G is unprovable (by the first theorem).
Therefore, Con(S) is also unprovable in S.

Philosophical meaning:

Mathematics cannot guarantee its own reliability.
We cannot prove that ZFC will not lead to a contradiction.
We simply believe this (and no contradictions have been found in 100 years).

This is like bootstrapping: one cannot verify a compiler with itself.

What this means in practice

Incompleteness does not mean	Incompleteness means
"Mathematics is useless"	There are fundamental boundaries
"Nothing can be proved"	Some questions are undecidable
"Computers cannot prove"	Complete automation is impossible
"Formalism must be abandoned"	Intuition remains important

Examples of undecidable statements:

- Continuum hypothesis: $|\mathbb{R}| = \aleph_1$? – independent of ZFC
- Axiom of choice – independent of ZF
- Some combinatorial statements (Paris–Harrington)

For an engineer:

99.9% of practical problems are not affected by incompleteness.
Gödel's theorems are important for understanding boundaries, not for daily work.
If an algorithm does not find a solution – possibly the problem is undecidable,
but usually one just needs to search better.

Unifying idea: There exist questions which a formal system cannot answer. This is not a weakness – it is a theorem.

Relations – how structure emerges from sets

"All mathematics = study of relations between objects"

Relation = a way to say "these objects are somehow connected"

In everyday life we constantly use relations:

- "Ivan is the father of Peter" (relation "to be a father")
- "Moscow is larger than Tver" (relation "larger by population")
- "5 is less than 7" (relation "less than")
- "These two triangles are similar" (relation "similarity")
- "A and B are friends" (relation "friendship")

Key idea: A relation is simply a set of pairs (or triples, etc.), which we consider "connected".

"less than" on $\mathbb{N} = \{(1,2), (1,3), (2,3), (1,4), (2,4), (3,4), ...\}$

The notation "$3 < 5$" simply means "$(3, 5) \in R$", where R is the relation "less than".

Why formalization?
- We can prove properties of relations
- We can combine relations (composition)
- We can classify relations by properties
- Foundation for order, equivalence, functions

Formal definition

Cartesian product

$A \times B = \{(a, b) \mid a \in A, b \in B\}$ – set of all ordered pairs

Example: $\{1, 2\} \times \{a, b\} = \{(1,a), (1,b), (2,a), (2,b)\}$

Formal definition of a pair:
$(a, b) := \{\{a\}, \{a, b\}\}$

This allows us to distinguish (a,b) from (b,a): $\{\{a\},\{a,b\}\} \neq \{\{b\},\{a,b\}\}$ when $a \neq b$
Important: an ordered pair is constructed from unordered sets.

Generalization to n sets:
$A_1 \times A_2 \times ... \times A_n = \{(a_1, a_2, ..., a_n) \mid a_i \in A_i\}$ – n-tuples

Example: $\mathbb{R} \times \mathbb{R} \times \mathbb{R} = \mathbb{R}^3$ – three-dimensional space

What is a relation

n-ary relation on sets $A_1, ..., A_n$ is a subset:

$$R \subseteq A_1 \times A_2 \times ... \times A_n$$

A tuple $(a_1, ..., a_n) \in R$ means "the elements are connected by relation R".

ARITY	EXAMPLES
Unary (n = 1)	$R \subseteq A$ – simply a subset (property) "even" $\subseteq \mathbb{Z} = \{x \in \mathbb{Z} \mid x \text{ is divisible by } 2\}$
Binary (n = 2)	$R \subseteq A \times B$ – connection between two elements "less than" $\subseteq \mathbb{R} \times \mathbb{R} = \{(x,y) \mid x < y\}$ "divides" $\subseteq \mathbb{N} \times \mathbb{N} = \{(a,b) \mid a \mid b\}$
Ternary (n = 3)	$R \subseteq A \times B \times C$ – connection between three elements "between" $\subseteq \mathbb{R} \times \mathbb{R} \times \mathbb{R} = \{(x,y,z) \mid x < y < z\}$ "sum" $= \{(a,b,c) \mid a + b = c\}$
n-ary	A table in a database with n columns. Each row is a tuple, the table is a relation

Most important binary relations

Below – a table of binary relations from which all of mathematics is built.

RELATION	PROPERTIES	EXAMPLES / VISUALIZATION
1. membership $x \in A$ "element of a set"	Basic relation Not reflexive ($x \notin x$ usually) Foundation of set theory	• $3 \in \mathbb{N}$ • $\pi \in \mathbb{R}$ • $\{1\} \in \mathcal{P}(\mathbb{N})$ A +-----+ | ● | ← x ∈ A +-----+
2. inclusion $A \subseteq B$ "subset"	• Reflexive: $A \subseteq A$ • Transitive: $A \subseteq B,\ B \subseteq C \Rightarrow A \subseteq C$ • Antisymmetric:	• $\mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R} \subseteq \mathbb{C}$ • $\{1,2\} \subseteq \{1,2,3\}$ • $\varnothing \subseteq A$ for any A +---------+

```
|                  |  A⊆B, B⊆A ⇒ A=B     |   |    B    |                     |
|                  |                      |   |  +---+  |                     |
|                  | Partial order        |   |  | A |  |                     |
|                  | on 𝒫(X)              |   |  +---+  |                     |
|                  |                      |   +---------+                     |
|                  |                      |                                   |
+------------------+----------------------+-----------------------------------+
|                  |                      |                                   |
| 3.equivalence    | • Reflexive:         | • Equality of numbers (=)         |
|    x ~ y         |   x ~ x              | • Similarity of triangles         |
|                  | • Symmetric:         | • Congruence modulo:              |
| "indistinguishable|   x~y ⇒ y~x         |   a ≡ b (mod n)                   |
|  with respect to | • Transitive:        | • Isomorphism of structures       |
|  a property"     |   x~y, y~z ⇒ x~z     |                                   |
|                  |                      | Partitions X into classes:        |
|                  | Gives factorization: |  +---+---+---+                    |
|                  | X/~ (quotient set)   |  |[x]|[y]|[z]|                    |
|                  |                      |  +---+---+---+                    |
|                  |                      |  equivalence classes              |
|                  |                      |                                   |
+------------------+----------------------+-----------------------------------+
|                  |                      |                                   |
| 4. (total)       | • Reflexive:         | • ≤ on ℝ                          |
|    ORDER         |                      |                                   |
|     x ≤ y        |   x ≤ x              | • Lexicographic on strings        |
|                  | • Antisymmetric:     | • Chronological (time)            |
| "not greater"    |   x≤y, y≤x ⇒ x=y     |                                   |
|                  | • Transitive:        |     a ≤ b ≤ c ≤ d                 |
|                  |   x≤y, y≤z ⇒ x≤z     |     ●---●---●---●                 |
|                  | • TOTALITY:          |                                   |
|                  |   ∀x,y: x≤y ∨ y≤x    |  Any two elements are comparable |
|                  |                      |                                   |
+------------------+----------------------+-----------------------------------+
|                  |                      |                                   |
| 5. partial       | Like total order,    | • ⊆ on 𝒫(X)                       |
|    order         | but without totality | • Divisibility | on ℕ             |
|    x ⪯ y          |                      | • Implication ⇒ on propositions  |
|                  | ∃x,y: incomparable   |                                   |
| "not greater, but| Hasse diagram:       |     +--d--+                       |
|  not always      |                      |     |     |                       |
|  comparable"     |                      |     b     c  ← incomparable       |
|                  |                      |     |     |                       |
|                  |                      |     +--a--+                       |
|                  |                      |                                   |
|                  |                      | Example: {1,2} and {1,3}          |
|                  |                      | incomparable by ⊆                 |
|                  |                      |                                   |
+------------------+----------------------+-----------------------------------+
|                  |                      |                                   |
| 6. function      | • Total:             | • f(x) = x²                       |
|    f: A → B      |   ∀a ∈ A: ∃b         | • sin: ℝ → [-1,1]                 |
|                  | • Single-valued:     | • det: M_n → ℝ                    |
| "rule of         |   f(a) is uniquely   |                                   |
|  correspondence" |   defined            |   A = {1, 2}    B = {a, b, c}     |
```

```
|                  |                      |                                  |
|                  | Subset of A×B with   |    1 --→ a                       |
|                  | conditions above     |    2 --→ b                       |
|                  |                      |                                  |
|                  |                      |  To each element of A            |
|                  |                      |  corresponds exactly one         |
|                  |                      |  element of B                    |
|                  |                      |                                  |
+------------------+----------------------+----------------------------------+
|                  |                      |                                  |
| 7. injection     | Function with        | • f(x) = 2x                      |
|    f: A ↣ B      | additional property: | • exp: ℝ → ℝ⁺                    |
|                  |                      | • embedding ℕ in ℤ               |
| "one-to-one"     | f(x₁) = f(x₂)        |                                  |
|                  |   ⇒ x₁ = x₂          |  A = {1, 2}    B = {a, b, c}     |
|                  |                      |                                  |
|                  | Different inputs →   |    1 --→ a                       |
|                  | different outputs    |    2 --→ b                       |
|                  |                      |          c  (not used)           |
|                  | Preserves distinctions|                                 |
|                  |                      |  |A| ≤ |B|, in B there may be    |
|                  |                      |  "extra" elements                |
|                  |                      |                                  |
+------------------+----------------------+----------------------------------+
|                  |                      |                                  |
| 8. surjection    | Function with        | • f: ℝ → ℝ, f(x) = x³            |
|    f: A ↠ B      | additional property: | • projection ℝ² → ℝ: (x,y) ↦ x   |
|                  |                      |                                  |
| "onto"           | ∀b ∈ B: ∃a ∈ A       |  A = {1, 2, 3}   B = {a, b}      |
|                  | such that f(a) = b   |                                  |
|                  |                      |    1 --→ a                       |
|                  | Covers all of B      |    2 --→ a  (both to one point)  |
|                  |                      |    3 --→ b                       |
|                  | Each element of B    |                                  |
|                  | has a preimage       |  |A| ≥ |B|, each element of B    |
|                  |                      |  is reached by at least one      |
|                  |                      |                                  |
+------------------+----------------------+----------------------------------+
|                  |                      |                                  |
| 9. bijection     | Function that is     | • f: ℝ → ℝ, f(x) = x             |
|    f: A ↔ B      | simultaneously:      | • exp: ℝ → ℝ⁺                    |
|                  | • Injection          | • any isomorphism                |
| "one-to-one      | • Surjection         |                                  |
|  correspondence" |                      |  A = {1, 2, 3}   B = {a, b, c}   |
|                  | Inverse exists       |                                  |
|                  | f⁻¹: B → A           |    1 ←-→ a                       |
|                  |                      |    2 ←-→ b                       |
|                  | |A| = |B|            |    3 ←-→ c                       |
|                  | (equinumerous)       |                                  |
|                  |                      |  Perfect correspondence:         |
|                  |                      |  to each a — exactly one b,      |
|                  |                      |  and vice versa                  |
|                  |                      |                                  |
+------------------+----------------------+----------------------------------+
```

Morphism — structure-preserving function

Now that we know what a function is, we can define a morphism.

Key concept: Morphism

Function (= Mapping) — a rule $f: A \to B$, assigning to each element $a \in A$ a unique element $f(a) \in B$.

Morphism — a function between structures that preserves structure.

Isomorphism — an invertible morphism (bijection preserving structure).
Means "structures are the same up to renaming of elements".

STRUCTURE	MORPHISM IS CALLED	WHAT IT PRESERVES
Sets	Function	(nothing — base case)
Groups	Homomorphism	Operation: $\varphi(ab) = \varphi(a)\varphi(b)$
Rings	Ring homomorphism	Both operations: +, ×
Vector spaces	Linear map	Linearity: $T(\alpha u+\beta v)=\alpha Tu+\beta Tv$
Topologies	Continuous map	Open sets
Manifolds	Smooth map	Smoothness
Orders	Monotone map	Order: $a\leq b \Rightarrow f(a)\leq f(b)$

Formula ≠ Function (common mistake)

Formula: a method of computation, for example $f(x) = x^2$
Function: a triple (domain A, codomain B, correspondence rule)

Why these are different things:

1. One function — many formulas:
   $f(x) = x^2 = (x+1)^2 - 2x - 1 = x\cdot x = \ldots$ (different notations for the same thing)

2. One formula — different functions (depends on domain):

FORMULA	DOMAIN	CODOMAIN	PROPERTIES
$f(x) = x^2$	$\mathbb{R}$	$\mathbb{R}$	not injective: $f(2) = f(-2)$
$g(x) = x^2$	$\mathbb{R}^+$	$\mathbb{R}^+$	Injective and surjective.
$h(z) = z^2$	$\mathbb{C}$	$\mathbb{C}$	Completely different object

3. Functions without formulas:
   $D(x) = 1$ if $x \in \mathbb{Q}$, otherwise 0 (Dirichlet function — no "formula".)

Conclusion: Function = triple $(A, B, f: A \to B)$.
Formula — just one way to specify the rule.

Image and preimage of a set

For a function $f: A \to B$ and subsets $S \subseteq A$, $T \subseteq B$:

CONCEPT	DEFINITION
Image of a set $f(S)$	$f(S) = \{f(x) : x \in S\}$ = all values of f on elements of S
Preimage of set $f^{-1}(T)$	$f^{-1}(T) = \{x \in A : f(x) \in T\}$ = all points that f sends to T

Important: $f^{-1}(T)$ — this is not an inverse function.
Preimage is always defined, even if f is not invertible.
Inverse function $f^{-1}: B \to A$ exists only for bijections.

Examples:

$f(x) = x^2$, $f: \mathbb{R} \to \mathbb{R}$

$f(\{1, 2, 3\}) = \{1, 4, 9\}$ — image of a set

$f^{-1}(\{4\}) = \{-2, 2\}$ — preimage of a point

$f^{-1}(\{y : y > 0\}) = \mathbb{R} \setminus \{0\}$ — preimage of an interval

$f^{-1}(\{-1\}) = \varnothing$ — preimage can be empty

Properties:

- $f(S_1 \cup S_2) = f(S_1) \cup f(S_2)$ — image of union
- $f^{-1}(T_1 \cup T_2) = f^{-1}(T_1) \cup f^{-1}(T_2)$ — preimage of union
- $f^{-1}(T_1 \cap T_2) = f^{-1}(T_1) \cap f^{-1}(T_2)$ — preimage of intersection
- $f(S_1 \cap S_2) \subseteq f(S_1) \cap f(S_2)$ — for image only $\subseteq$!

Application: Topology uses preimages to define continuity:
f continuous $\Longleftrightarrow$ preimage of an open set is open

Key patterns:

Checking properties of a relation R:

1. Reflexivity: xRx for all x?
2. Symmetry: $xRy \Rightarrow yRx$?
3. Transitivity: xRy, $yRz \Rightarrow xRz$?
4. Antisymmetry: xRy, $yRx \Rightarrow x=y$?

Combinations define the type:

- Reflexivity + Symmetry + Transitivity = Equivalence (partition into classes)
- Reflexivity + Antisymmetry + Transitivity = Partial order
- + Totality = Linear order

Equivalence relations and equivalence classes

Intuition: Equivalence = "Same from a certain point of view"

Equality (=) — too strict a requirement. Often we care not about complete identity, but "sameness" by some criterion:

- Two people are "equivalent" by age (both are 25 years old)
- Two triangles are "equivalent" by shape (similar)
- Two fractions are "equivalent" by value ($1/2 = 2/4 = 3/6$)
- Two numbers are "equivalent" by remainder when divided by 3

Formally: a relation $\sim$ is called an equivalence if it is:

1. Reflexive: $a \sim a$ (each is equivalent to itself)
2. Symmetric: $a \sim b \Rightarrow b \sim a$ (order doesn't matter)
3. Transitive: $a \sim b, b \sim c \Rightarrow a \sim c$ (equivalence is "inherited")

Equivalence class = all those "identical" to a given one

If $\sim$ is an equivalence on set X, then the equivalence class of element a is the set of all elements equivalent to a:

$$[a] = \{x \in X \mid x \sim a\}$$

Key fact: Equivalence classes partition the set X:

- Each element $x \in X$ lies in exactly one class
- Different classes do not intersect
- Union of all classes = X

Visualization:

```
  Set X:                        After partition X/~:
  +-----------------------+     +-----------------------+
  | ● ●   ○ ○ ○    ▪ ▪    |     |  [●]   [○]   [▪]      |
  |   ●   ○        ▪ ▪ ▪  |  →  |                       |
  |       ○ ○          ▪  |     | 3 classes             |
  +-----------------------+     +-----------------------+
  (elements of the same "type"  (each class is one
   marked identically)           element of quotient set)
```

Concrete examples of equivalence classes

Example 1: Remainders from division (modular arithmetic)

Set: $\mathbb{Z}$ (all integers)
Relation: $a \sim b \iff a - b$ is divisible by 3 (written: $a \equiv b \pmod{3}$)

Equivalence classes:
$[0] = \{..., -6, -3, 0, 3, 6, 9, ...\}$ – numbers with remainder 0
$[1] = \{..., -5, -2, 1, 4, 7, 10, ...\}$ – numbers with remainder 1
$[2] = \{..., -4, -1, 2, 5, 8, 11, ...\}$ – numbers with remainder 2

Quotient set: $\mathbb{Z}/3\mathbb{Z} = \{[0], [1], [2]\}$ – total of 3 classes.
(This is a ring of residues – foundation of cryptography)

Example 2: Fractions (rational numbers)

Set: $\mathbb{Z} \times \mathbb{Z}^* = \{(a, b) \mid a,b \in \mathbb{Z}, b \neq 0\}$ – pairs "numerator/denominator"
Relation: $(a,b) \sim (c,d) \iff a \cdot d = b \cdot c$ (fractions are equal)

Equivalence classes:
$[(1,2)] = \{(1,2), (2,4), (3,6), (-1,-2), ...\}$ – all forms of "one
$[(1,3)] = \{(1,3), (2,6), (-1,-3), ...\}$ half"

A rational number is an equivalence class, not a specific fraction.
$\mathbb{Q} = (\mathbb{Z} \times \mathbb{Z}^*)/\sim$ – set of equivalence classes of pairs.

Example 3: Similarity of triangles

Set: all triangles on the plane
Relation: $T_1 \sim T_2 \iff$ angles of T_1 equal angles of T_2

Equivalence class: all triangles of one "shape" (of different sizes)

```
      +-+              +---+
     /   \            /     \
    / 60° \          /  60°  \
   /       \        /         \        All three are in one
  / 60° 60° \      /  60° 60°  \       equivalence class
 +-----------+    +-------------+      (equilateral)
```

Example 4: Real numbers (construction via Cauchy)

Set: all Cauchy sequences of rational numbers
Relation: $\{a_n\} \sim \{b_n\} \iff |a_n - b_n| \to 0$

A real number is an equivalence class of sequences.
Example: $\sqrt{2} = [\{1, 1.4, 1.41, 1.414, 1.4142, ...\}]$

(Infinitely many sequences converge to $\sqrt{2}$ – they are all equivalent and form one class, which we call the number $\sqrt{2}$)

Example 5: Homotopy classes of paths (topology)

Set: all paths from A to B in space X
Relation: $\gamma_1 \sim \gamma_2 \iff \gamma_1$ can be continuously deformed into γ_2

Fundamental group $\pi_1(X)$ = equivalence classes of closed paths.
This construction is the foundation of algebraic topology.

Quotient set X/~ – "gluing" of equivalent elements

The quotient set $X/\sim$ is the set of equivalence classes:

$$X/\sim \; = \{[a] \mid a \in X\}$$

Intuition: We "glue" all equivalent elements into one point.

Projection: The canonical map $\pi: X \to X/\sim$, where $\pi(a) = [a]$.
It "forgets" distinctions between equivalent elements.

Main theorem (Universal property):

Any function $f: X \to Y$, constant on equivalence classes (i.e. $a \sim b \Rightarrow f(a) = f(b)$), factors uniquely through the quotient set:

$$\begin{tikzcd} X \arrow[r, "f"] \arrow[d, "\pi"'] & Y \\ X/\sim \arrow[ru, "\bar{f}\ \text{(exists unique)}"'] & \end{tikzcd}$$

Where this is used:
- $\mathbb{Z}_n = \mathbb{Z}/n\mathbb{Z}$ (modular arithmetic, cryptography)
- $\mathbb{Q} = (\mathbb{Z}\times\mathbb{Z}^*)/\sim$ (construction of rational numbers)
- $\mathbb{R} = \text{Cauchy}/\sim$ (construction of real numbers)
- $\mathbb{C} = \mathbb{R}[x]/(x^2+1)$ (construction of complex numbers)
- Projective space $\mathbb{P}^n = (\mathbb{R}^{n+1} \setminus 0)/\sim$
- Tori, Klein bottles and other topological objects

Intuition: from clocks to complex numbers

$\mathbb{Z}/12\mathbb{Z}$ – clock arithmetic. We "wrap" integers onto a clock face:
13 o'clock = 1 o'clock, because $13 \equiv 1 \pmod{12}$. We declared $12 = 0$.

$\mathbb{R}[x]/(x^2+1)$ — the same thing, but with polynomials. We "wrap" polynomials, declaring $x^2+1 = 0$, that is $x^2 = -1$.
We get: x is a "new number" whose square equals −1.
This is the imaginary unit i. Complex numbers $a + bi$ are polynomials of degree ≤ 1 with arithmetic modulo (x^2+1).

Why does an engineer need equivalence classes?

1. Modular arithmetic ($\mathbb{Z}_n$):
   Hash functions, checksums, cryptography — all this is work with residue classes.

2. Finite automata:
   Minimization of an automaton = merging "equivalent" states (those that cannot be distinguished by input data).

3. Dimensional analysis:
   Physical quantities with the same dimensionality — one class.
   "Meters per second" is the class of all ways to express velocity.

4. Coordinates:
   A tensor is an equivalence class of sets of numbers (coordinates), where two sets are equivalent if they are related by a transformation law.

5. Clustering:
   Partitioning data into clusters is essentially constructing equivalence classes by the criterion of "similarity".

Philosophy:

Mathematics studies not the objects themselves, but the relations between them.
A function $f: A \to B$ is more important than A and B separately.
Modern mathematics (categorical) is "mathematics of arrows".

==

Mappings — how structures are connected

==

Notes: This section is an overview of types of mappings. The terms "group", "topology", "manifold" are defined in Part II.
Here they are used to illustrate a general pattern.

In the "Relations" section we defined function and morphism. Here are detailed examples.

"A good morphism preserves structure"

Preliminary remark on categories

All examples below are special cases of one idea: morphism.

- Objects (sets, groups, spaces, ...)
- Arrows between them (functions, homomorphisms, continuous maps, ...)
- Composition of arrows (apply one, then another)

This is a category — a language that unifies all of mathematics.
For now it suffices to know: different types of mappings in the table below are morphisms in different categories.

Category of sets: morphisms = functions
Category of groups: morphisms = homomorphisms
Category of topol. sp.: morphisms = continuous mappings
Category of vect. sp.: morphisms = linear operators

TRANSFORMATION	STRUCTURE	EXAMPLES
1. function $f: A \to B$ Basic mapping	Set → Set Preserves nothing (no additional structure)	• $f(x) = x + 1$ • projection $(x,y) \mapsto x$ • any mapping
2. homomorphism $\varphi: G \to H$ Preserves operation Automorphism if G = H	Group → Group $\varphi(g_1 * g_2) =$ $= \varphi(g_1) \odot \varphi(g_2)$ Isomorphism if bijection (structures "identical")	• exp: $(\mathbb{R},+) \to (\mathbb{R}^+,\times)$ $\exp(a+b) = \exp(a)\exp(b)$ • det: $(GL(n),\times) \to (\mathbb{R}^*,\times)$ $\det(AB) = \det(A)\det(B)$ ($GL(n)$ = invertible $n \times n$ matrices) • complex conjugation $z \mapsto \bar{z}$ on $\mathbb{C}$
3. linear operator $T: V \to W$ Preserves linear structure	Vect.sp. → Vect.sp. $T(\alpha v + \beta w) =$ $= \alpha T(v) + \beta T(w)$ Matrix in basis	• Rotation in $\mathbb{R}^2$ • Projection onto subsp. • Differentiation $D: C^1 \to C^0$ $D(af + bg) =$ $= aD(f) + bD(g)$
4. continuous mapping $f: X \to Y$ Preserves "closeness"	Topol.sp. → Topol.sp. Preimage of open is open: $U \in \tau\_Y \Rightarrow f^{-1}(U) \in \tau\_X$	• $f: \mathbb{R} \to \mathbb{R}$ continuous • Deformation of rubber • Any $f: M \to N$ (smooth) Homeomorphism if bijection + f^{-1} continuous

```
|                      |                           | (topologically "the same")|
|                      |                           |                          |
+----------------------+---------------------------+--------------------------+
|                      |                           |                          |
| 5. diffeomorphism    | Manifold →                | • Any smooth invertible  |
|    f: M → N          | Manifold                  |   mapping                |
|                      |                           | • ℝ ≅ (0,1) via          |
| Smooth + invertible  | f and f⁻¹ smooth          |   f(x) = x/(1+|x|)       |
|                      |                           |                          |
```

	Preserves smooth structure	Shows geometric equivalence
6. derivative d/dx Instantaneous rate of change	Function → Function Linear operator: $D(af + bg) = aD(f)+bD(g)$ Local approximation	• $D(x^2) = 2x$ • D(sin) = cos • $D(e^x) = e^x$ Geometry: tangent to graph
7. integral $\int_a^b f(x)dx$ Accumulation (area, mass)	Function → Number Linear functional: $\int(af+bg) = a\int f + b\int g$ Inverse to derivative	• $\int_0^1 x^2 dx = 1/3$ • $\int \sin x dx = -\cos x+C$ Geometry: area under graph
8. tangent space T_pM "Plane at a point"	Manifold → Vector space Curved → Flat Functor: $M \mapsto T_pM$	• T_pS^2 – plane tangent to sphere at p • T_pM – where "live" velocity vectors Linearization of geometry
9. fundamental group π_1: Top → Grp functor Geometry → Algebra	Topology → Algebra Space ↦ Group Loops with homotopies "Holes" transforms into algebraic structure	• $\pi_1(S^1) = \mathbb{Z}$ (circle → integers) • $\pi_1(S^2) = \{e\}$ (sphere trivial) • $\pi_1(T^2) = \mathbb{Z} \times \mathbb{Z}$ (torus → two holes) Computation of topology via groups
10. Fourier transform $\mathcal{F}: L^2(\mathbb{R}) \to L^2(\mathbb{R})$ Functor Time ↔ Frequency	Function → Function $\mathcal{F}(f)(\omega) = \int f(t)e^{-i\omega t}dt$ Unitary operator Transforms diff. eq. into algebraic	• $\cos(\omega t) \leftrightarrow \delta(\omega - \omega_0)$ • Convolution ↔ Multiply Applications: • Signal processing • Quantum mechanics • Diff. equations

General pattern:

A good transformation preserves what is important:

- Homomorphism preserves operation
- Continuous mapping preserves closeness
- Linear operator preserves linearity
- Isometry preserves distance

Functor — transformation between categories (in detail):
Translates objects into objects and arrows into arrows, preserving composition.
Example: π_1 translates topological spaces into groups,
and continuous mappings — into homomorphisms.

Why this is needed:

1. Connection between areas: π_1 transforms topology into algebra
2. Simplification: Fourier transforms diff. eq. into algebraic
3. Understanding: tangent space makes geometry flat
4. Computations: isomorphism allows working in a convenient structure

Numbers — how all numbers arise from emptiness

Construction of natural numbers

Construction of natural numbers from the empty set (von Neumann)

All mathematics arises from a single primitive: the empty set $\varnothing$

```
+-------+----------------------+-----------------------------------+
```

```
| NUMBER| DEFINITION            | VISUALIZATION                      |
+-------+-----------------------+------------------------------------+
|       |                       |                                    |
| 0     | ∅                     | { }                                |
|       | Empty set             | Nothing there                      |
|       |                       |                                    |
+-------+-----------------------+------------------------------------+
|       |                       |                                    |
| 1     | {∅}                   | { { } }                            |
|       | Set containing        | Box with empty box                 |
|       | empty set             |                                    |
|       |                       |                                    |
+-------+-----------------------+------------------------------------+
|       |                       |                                    |
| 2     | {∅, {∅}}             | { { }, { { } } }                   |
|       | = {0, 1}              | Box with: (empty) and (box         |
|       |                       | with empty)                        |
|       |                       |                                    |
+-------+-----------------------+------------------------------------+
|       |                       |                                    |
| 3     | {∅, {∅}, {∅, {∅}}}   | { { }, { { } }, { { }, { { } } } } |
|       | = {0, 1, 2}           | Box with 0, 1 and 2                |
|       |                       |                                    |
+-------+-----------------------+------------------------------------+
|       |                       |                                    |
| n     | {0, 1, 2, ..., n-1}   | Set of all previous numbers        |
|       |                       |                                    |
|       | n = {k : k < n}    | n contains within itself all       |
|       |                       | numbers less than itself           |
|       |                       |                                    |
+-------+-----------------------+------------------------------------+

Recursive definition:

  0 := ∅
  n+1 := n ∪ {n}   (next number = previous + set with previous)

Deep meaning:

  • From nothing (∅) arises everything
  • Each number contains all previous: 3 ⊃ 2 ⊃ 1 ⊃ 0
  • Number = "memory" of all previous steps
  • Construction of numbers = successive drawing of boundaries in emptiness

From natural to the rest:
```

```
 ℕ = {0, 1, 2, 3, ...}     (defined above)
       ↓
 ℤ = ℕ × ℕ / ~             (pairs (a,b) as "a - b", with equivalence)
       ↓
 ℚ = ℤ × ℤ* / ~            (pairs (p,q) as "p/q", q≠0)
       ↓
 ℝ = {Dedekind cuts}       (or Cauchy sequences in ℚ)
       ↓
 ℂ = ℝ × ℝ                 (pairs (a,b) as "a + bi")
```

Everything is built from ∅ by successive acts of drawing boundaries.

Dedekind cut — how to create irrational numbers from rational

Problem: In $\mathbb{Q}$ there are "holes"
Equation $x^2 = 2$ has no solution in $\mathbb{Q}$ (proof — by contradiction)
But on the number line there must be a point $\sqrt{2}$.

Dedekind's idea (1858): define a number through a cut of the line

Definition:
Cut (A, B) of the set $\mathbb{Q}$ is a partition of $\mathbb{Q}$ into two nonempty classes:

- A — "lower class": all rational numbers to the left of the cut
- B — "upper class": all rational numbers to the right of the cut
- $A \cup B = \mathbb{Q}$, $A \cap B = \varnothing$
- Any element of A is less than any element of B
- In class A there is no greatest element (key condition)

Visualization:

```
  Number line ℚ:
  ←--●--●--●--●--●--|--●--●--●--●--●--→
        A          ✂         B
                 (cut)
```

Cut for $\sqrt{2}$:
$A = \{q \in \mathbb{Q} : q < 0 \text{ or } q^2 < 2\}$ (all "to the left of $\sqrt{2}$")
$B = \{q \in \mathbb{Q} : q > 0 \text{ and } q^2 \geq 2\}$ (all "to the right of $\sqrt{2}$")

In A there is no greatest. For any $q \in A$ with $q^2 < 2$ there exists rational $r > q$ with $r^2 < 2$.

Definition of real number:
$\mathbb{R} := \{\text{all cuts } (A, B) \text{ of the set } \mathbb{Q}\}$

Two types of cuts:
1. B has smallest element $q \in \mathbb{Q}$ → cut defines rational q
2. B has no smallest element → cut defines irrational number

Arithmetic of cuts:
$(A_1, B_1) + (A_2, B_2) := (A_1 + A_2, \ldots)$ where $A_1 + A_2 = \{a_1 + a_2\}$
Similarly defined multiplication, order etc.

Deep meaning:

- Irrational number is not an object, but a boundary between objects
- $\sqrt{2}$ is not a concrete fraction, but a place of cut of the number line
- Again the act of drawing a boundary creates a new entity

Alternative – Cauchy sequences:

$\mathbb{R}$ can also be defined as equivalence classes
of fundamental sequences in $\mathbb{Q}$
(sequences where $|a_n - a_m| \to 0$ as $n, m \to \infty$)

Example: 1, 1.4, 1.41, 1.414, 1.4142, ... $\to \sqrt{2}$

Both approaches give the same set $\mathbb{R}$.

Cauchy Sequence – Convergence Without Knowing the Limit

Problem: How to understand that a sequence converges
if we don't know what it converges to?

Cauchy's Idea: A sequence converges if its terms
become arbitrarily close to each other.

Definition:

A sequence $\{a_n\}$ is called a Cauchy sequence
(or fundamental), if:

$$\forall \varepsilon > 0 \;\; \exists N \in \mathbb{N}: \;\; \forall n,m > N \;\Rightarrow\; |a_n - a_m| < \varepsilon$$

In words: starting from some number N, the distance between
any two terms is less than any predetermined ε.

Visualization:

Ordinary sequence:
"All terms are close to limit a"

```
     a₁    a₂   a₃ a₄a₅
--●---●----●--●●●●------------→
               ↑
               a (limit)
```

Cauchy sequence:
"All terms are close to each other"

```
     a₁      a₂    a₃  a₄ a₅ a₆ a₇.
--●-------●-----●---●--●●●●●●--------→
                   +--+
                    <ε
                "stick together"
```

Examples:

✓ Cauchy: $a_n = 1/n$
$|a_n - a_m| = |1/n - 1/m| \le 1/n + 1/m \to 0$
Terms: 1, 0.5, 0.33, 0.25, 0.2, ... (stick together around 0)

✓ Cauchy: $a_n = 1 + 1/2 + 1/4 + ... + 1/2^n$
Partial sums of geom. progression: 1, 1.5, 1.75, 1.875, ...
Converges to 2, terms closer and closer to each other

✗ Not Cauchy: $a_n = n$
$|a_{n+1} - a_n| = 1$ — does not tend to 0, terms diverge

✗ Not Cauchy: $a_n = (-1)^n$
$|a_{n+1} - a_n| = 2$ — terms jump between −1 and 1

Theorem (Cauchy criterion):

```
+--------------------------------------------------------+
|                                                        |
| In ℝ (real numbers):                                   |
|                                                        |
| Sequence converges ⟺ Sequence is Cauchy                 |
|                                                        |
+--------------------------------------------------------+
```

Important: In ℚ this is not the case.

The sequence 1, 1.4, 1.41, 1.414, 1.4142, ... is Cauchy in ℚ,
but does not converge in ℚ, because $\sqrt{2} \notin \mathbb{Q}$.

This is called incompleteness of ℚ.

Completeness of ℝ:
The set ℝ is called complete, because in it
every Cauchy sequence has a limit.
This is a key property distinguishing ℝ from ℚ.

Cauchy construction for ℝ:
ℝ = {equivalence classes of Cauchy sequences in ℚ}

Two sequences $\{a_n\} \sim \{b_n\}$ are equivalent if $|a_n - b_n| \to 0$

Example: (1, 1.4, 1.41, ...) ~ (1.5, 1.42, 1.415, ...) — both "represent" $\sqrt{2}$

Deep meaning:
The Cauchy criterion allows one to speak of convergence
in an internal way — through the terms of the sequence itself,
without invoking the concept of "limit" from outside.

Cardinalities of Sets — Different Sizes of Infinity

Two sets have the same cardinality if there is a bijection between them.
For finite sets this is obvious: $|\{a,b,c\}| = |\{1,2,3\}| = 3$.

And for infinite ones? Cantor showed: there are different infinities.

Countable sets (cardinality $\aleph_0$ — "aleph-null"):
A set is called countable if there exists a bijection with ℕ.
That is, elements can be "enumerated": $a_1, a_2, a_3, \ldots$

- ℕ — countable (trivially)
- ℤ — countable. Bijection: 0, 1, −1, 2, −2, 3, −3, ...
- ℚ — countable. (surprisingly — "more" numbers, but same cardinality)
- ℕ × ℕ — countable (diagonal traversal)

Uncountable sets (cardinality $> \aleph_0$):

- ℝ — uncountable.
- (0,1) — uncountable (bijection with ℝ via tan)
- $\mathcal{P}(\mathbb{N})$ — uncountable (set of all subsets of ℕ)

Cantor's Diagonal Argument — Why ℝ is Uncountable

Theorem: The set ℝ (or even the interval [0,1]) is uncountable.

Proof (by contradiction):

Suppose [0,1] is countable. Then all numbers can be enumerated:

$x_1 = 0.\ a_{11}\ a_{12}\ a_{13}\ a_{14}\ \ldots$
$x_2 = 0.\ a_{21}\ a_{22}\ a_{23}\ a_{24}\ \ldots$
$x_3 = 0.\ a_{31}\ a_{32}\ a_{33}\ a_{34}\ \ldots$
$x_4 = 0.\ a_{41}\ a_{42}\ a_{43}\ a_{44}\ \ldots$
... ↘ ↘ ↘ ↘
diagonal

Construct a number $y = 0.\ b_1\ b_2\ b_3\ b_4\ \ldots$, where:

$b_n \neq a_{nn}$ (differs from n-th number in n-th digit)

Then $y \neq x_1$ (differs in 1st digit)
$y \neq x_2$ (differs in 2nd digit)
$y \neq x_n$ for any n!

But $y \in [0,1]$, so y must be in the list. Contradiction.

Conclusion: [0,1] cannot be enumerated. The set ℝ is uncountable.

Deep meaning:

The diagonal argument is a universal method. It shows:
$\mathcal{P}(A)$ always has greater cardinality than A (Cantor's theorem).
Infinities form a hierarchy: $\aleph_0 < 2^{\aleph_0} < 2^{(2^{\aleph_0})} < \ldots$

Continuum Hypothesis — A Mystery Without an Answer

The cardinality of $\mathbb{R}$ is called continuum and is denoted c or $2^{\aleph_0}$.

Cantor's question (1878):

Does there exist a set with cardinality between $\aleph_0$ and c?
That is: is there an infinity "in between" $\mathbb{N}$ and $\mathbb{R}$?

Continuum hypothesis (CH): No, there does not exist. That is $c = \aleph_1$.

Shocking result of the 20th century:

- Gödel (1940): CH is consistent with ZFC (cannot be refuted)
- Cohen (1963): ¬CH is also consistent with ZFC (cannot be proven)

⇒ CH is independent of the axioms of ZFC.

This is not a question of "we don't know yet". This is fundamental: within ZFC the question has no answer. One can work in mathematics with CH or without it.

Philosophical lesson:

There are meaningful mathematical questions to which the axioms do not answer.
Mathematics is not a single "true" system, but a family of possible ones.

Extension of numbers — what is missing?

Main idea:

Each number system arises as an answer to the question:
"What equation has no solution?"

We extend numbers → obtain solution → lose some property

Terminology (strict definitions):

- Field = set with +, ×, where everything except 0 is invertible ($\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$)
- Ring = field without mandatory division ($\mathbb{Z}$)
- Division algebra = "almost field", but possibly noncommutative ($\mathbb{H}$, $\mathbb{O}$)

Hierarchy of number systems

NUMBER	EQUATION WITHOUT SOLUTION	SOLUTION	WHAT WE LOSE	WHAT WE GAIN
ℕ	$x + 3 = 1$	not in ℕ	–	Counting
ℤ	$2x = 3$	not in ℤ	"always ≥ 0"	Subtraction
ℚ	$x^2 = 2$	not in ℚ	Discreteness	Division
ℝ	$x^2 = -1$	not in ℝ	Countability	Continuity
ℂ	all exist.	–	Order (<, >)	Alg. closedness
ℍ	(special)	–	Commutativity $ab \neq ba$	4D rotations
𝕆	(special)	–	Associativity $(ab)c \neq a(bc)$	8D, exceptional structures

Visualization

```
ℕ ⊂ ℤ ⊂ ℚ ⊂ ℝ ⊂ ℂ
|   |   |   |   |
|   |   |   |   +-- Plane (2D): a + bi
|   |   |   +------ Line (1D): continuous
|   |   +---------- Line (1D): with holes (no √2)
|   +-------------- Line (1D): integer points
+------------------ Ray: 0, 1, 2, 3, ...

ℍ (quaternions): 4D, basis {1, i, j, k},  i² = j² = k² = ijk = −1
𝕆 (octonions):   8D, last division algebra over ℝ
```

Frobenius theorem (1877) and Hurwitz theorem (1898)

Frobenius: Associative division algebras over ℝ: ℝ, ℂ, ℍ — and that's all.
Hurwitz: Normed division algebras over ℝ: ℝ, ℂ, ℍ, 𝕆 — and that's all.

Dimensions: 1, 2, 4, 8 — no more.

With each extension we lose a fundamental property:

ℝ → ℂ: we lose order
ℂ → ℍ: we lose commutativity
ℍ → 𝕆: we lose associativity
𝕆 → 𝕊: we lose the ability to divide (zero divisors appear)

Sedenions 𝕊 (16D) and beyond — mathematically exist, but are useless for physics: in them $ab = 0$ is possible with $a \neq 0$ and $b \neq 0$. Therefore the chain ℝ → ℂ → ℍ → 𝕆 — this is all there is.

Fundamental Theorem of Algebra

Any polynomial with complex coefficients has a root in ℂ.

This means: ℂ is closed under solving equations.
There is no need to extend further "for the sake of equations".

Therefore ℂ is called an "algebraically closed field".

Numbers as points in spaces — topological differences

Hierarchy of embeddings:

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}$$

Each subsequent set "fills the holes" of the previous one:

ℕ → ℤ : added negatives (solution of $a + x = 0$)
ℤ → ℚ : added fractions (solution of $ax = b$)
ℚ → ℝ : filled "holes" (limits of sequences)
ℝ → ℂ : added √(-1) (roots of all polynomials)

Geometric picture:

ℤ — lattice on the line ℂ — plane

```
  (separate points)            (ℝ — horizontal axis)

  --●--●--●--●--●--●--          Im
    -2 -1  0  1  2  3             |
                                  |        ● 2+i
                             ----+----●--------► Re
                                  |     1
                                  ● -i
                                  |

  ℚ — "dust" on the line         ℝ — continuous line
  (everywhere dense, but holey)    (without holes, complete)

  -·-·-·-·-·-·-·-·-·-·-         ━━━━━━━━━━━━━━━━━━━
      √2 — hole.                   √2 exists.
```

Topological properties

SPACE	TOPOLOGICAL PROPERTIES
ℕ, ℤ	Discrete topology: each point — open set Not compact, not connected (each point — component)
ℚ	Everywhere dense in ℝ, but "holey" Not complete (sequence → √2 has no limit in ℚ) Not connected (ℚ = (−∞,√2)∩ℚ ∪ (√2,+∞)∩ℚ — partition)
ℝ	Complete, connected, locally compact The unique complete ordered field.
ℂ	Complete, connected, locally compact Algebraically closed (any polynomial has a root) No order compatible with operations.

Key idea:
ℕ ⊂ ℤ ⊂ ℚ ⊂ ℝ ⊂ ℂ — this is not just "adding numbers".
At each step the topology and algebraic structure of the space changes.

Dictionary of mathematical jargon

Mathematicians use words that are rarely explained in textbooks.
Here is a dictionary of the most common "invisible" terms.

Canonical

= "natural", "unique", "not requiring an arbitrary choice"

Example:

- $V^{**} \cong V$ — canonical isomorphism (there is a natural correspondence)
- $V^{*} \cong V$ — not canonical (need to choose a basis, result depends on choice)

When they say "canonical" — it means there is one correct way, and no additional choices need to be made.

Correctly defined / well defined

= "result does not depend on the method of representation"

Problem: Sometimes the same object can be written in different ways. A function is well-defined if it gives the same answer for all representations.

Example 1 (incorrect):

Let $f: \mathbb{Z}/3\mathbb{Z} \to \mathbb{Z}$, $f([x])$ = remainder of x when divided by 2.
Check: $f([0]) = 0$, $f([3]) = 1$.
But $[0] = [3]$ in $\mathbb{Z}/3\mathbb{Z}$. Therefore $f([0])$ must $= f([3])$.
Contradiction: $0 \neq 1$. Function f is not well-defined.

Example 2 (correct):

Let $g: \mathbb{Z}/6\mathbb{Z} \to \mathbb{Z}/2\mathbb{Z}$, $g([x]) = [x \bmod 2]$.
Check: if $[x] = [y]$ in $\mathbb{Z}/6\mathbb{Z}$, then $x \equiv y \pmod 6$.
Therefore $x \equiv y \pmod 2$, and $[x \bmod 2] = [y \bmod 2]$.
Function g is well-defined. ✓

Conclusion: When working with equivalence classes one must always check correctness of definition.

Without loss of generality

= "it suffices to consider the special case, the rest reduce to it"

Example 1:

Theorem: $|a + b| \leq |a| + |b|$
Proof: "W.l.o.g. let $a \geq 0$."
Why can we? If $a < 0$, replace a with $-a$ (the absolute value won't change), and return to the case $a \geq 0$.

Example 2:
Prove something for two points A and B.
"W.l.o.g. let A be to the left of B."
Why can we? We can rename: $A \leftrightarrow B$.

When it can be used:

- There is symmetry in the condition (a and b are interchangeable)
- Can be reduced to the needed case by substitution of variables
- The operation (renaming, change of sign) doesn't change the essence of the problem

When it cannot:

- Conditions for a and b are different
- The transformation changes the structure of the problem

Trivial / nontrivial

Trivial = "obvious", "empty", "simplest possible"
Nontrivial = "substantial", "not reducing to the obvious"

Examples:

- Trivial solution of $Ax = 0$: $x = 0$
- Trivial subgroup: $\{e\}$ or all of G
- Nontrivial root: $x \neq 0$

Irony: What is "trivial" for a professor may be difficult for a student.
"Trivially follows." often means "I don't want to explain this".

Why the symbol ∘ for composition

The notation $(f \circ g)(x) = f(g(x))$ means: "apply g first, then f".

Why right to left:
From the notation $f(g(x))$ – first we compute $g(x)$, then f of the result.
The inner function is applied first.

Origin of the symbol:
∘ – small circle, "linking element" between functions.
Introduced in the 19th century for brevity.

Alternative notation (rare):
$f \,;\, g = g \circ f$ – "first f, then g" (reads left to right)
Used in some programming languages and category theory.

Remember: $f \circ g$ read as "f after g" or "f circle g".

Unique up to

= "unique if we don't distinguish objects of a certain type"

Examples:

- "A basis is unique up to order" — vectors can be permuted
- "Solution is unique up to sign" — there are exactly two: x and $-x$
- "Group is unique up to isomorphism" — all such groups are isomorphic to each other

Meaning: There are several objects, but they are "the same" in a certain sense.

Typical errors in reasoning

Error 1: Confusing implication $A \Rightarrow B$ with equivalence $A \Longleftrightarrow B$

$A \Rightarrow B$ does not mean $B \Rightarrow A$

True: "If n is divisible by 4, then n is divisible by 2"
False: "If n is divisible by 2, then n is divisible by 4"
Counterexample: $n = 6$ is divisible by 2, but 6 is not divisible by 4.

Correct relations:
$A \Rightarrow B$ (direct implication)
$B \Rightarrow A$ (converse implication) — different statement
$\neg B \Rightarrow \neg A$ (contrapositive) — equivalent to direct

Error 2: Circular proof

Wrong: "Prove A. Assume A. Then. Therefore A. ∎"

This is not a proof. You used A to prove A.

More subtle version:
"From A follows B. From B follows C. From C follows A."
This proves $A \Longleftrightarrow B \Longleftrightarrow C$, but doesn't prove that A is true.

Error 3: Confusing the order of quantifiers $\forall x \exists y$ and $\exists y \forall x$

$\forall x\ \exists y: P(x,y)$ — for each x there exists its own y (y depends on x)
$\exists y\ \forall x: P(x,y)$ — there exists one y working for all x

Example: Continuity vs uniform continuity

$\forall\varepsilon>0\ \forall x\ \exists\delta>0: \ldots$ – δ can depend on x and ε (simple continuity)
$\forall\varepsilon>0\ \exists\delta>0\ \forall x: \ldots$ – δ is one for all x (uniform continuity)

These are different properties. $f(x) = 1/x$ is continuous on $(0,1)$, but not uniformly.

Error 4: Dividing by an expression that may be zero

Wrong: "From $ax = ay$ divide by a, we get $x = y$"

What if $a = 0$? Then $0 = 0$ is true for any x, y.

Correct: "If $a \neq 0$, then from $ax = ay$ follows $x = y$."
"If $a = 0$, then the equality holds for any x, y."

Error 5: One example is not a proof

Wrong: "41 is prime, $41+2=43$ is also prime.
Therefore, if p is prime, then $p+2$ is also prime."

Counterexample: 7 is prime, but $7+2 = 9 = 3^2$ is not prime.

Asymmetry:
- To prove $\forall x\ P(x)$ need to check all x
- To refute $\forall x\ P(x)$ one x with $\neg P(x)$ is enough

Error 6: Identifying necessary and sufficient

A – sufficient condition for B: $A \Rightarrow B$ (if A, then certainly B)
A – necessary condition for B: $B \Rightarrow A$ (without A there is no B)

Example: "Being a square" for "divisible by 4"

Sufficient? $n = k^2 \Rightarrow n$ is divisible by 4? No. ($9 = 3^2$, but $4 \nmid 9$)
Necessary? n is divisible by $4 \Rightarrow n$ is a square? No. (8 is divisible by 4, $8 \neq k^2$)

Why Symbols Instead of Words

Question: Why do mathematicians write $\forall\varepsilon>0\ \exists\delta>0$ instead of words?

Reason 1: Compactness

In words:
"For any positive number epsilon there exists such a positive number delta that for all x, if the absolute value of the difference x and a is less than de
then the absolute value of the difference f of x and L is less than epsilon."

In symbols:
$\forall\varepsilon>0\ \exists\delta>0\ \forall x: |x-a|<\delta \Rightarrow |f(x)-L|<\varepsilon$

One line instead of four.

Reason 2: Precision

The order of quantifiers is visible immediately:

$\forall\varepsilon\ \exists\delta$ – for each ε its own δ (ordinary continuity)
$\exists\delta\ \forall\varepsilon$ – one δ for all ε (different property)

In words this is easy to confuse.

Reason 3: Manipulation

Logical rules are applied mechanically:

$\neg(\forall x\ P(x)) = \exists x\ \neg P(x)$ – simply flip the quantifier and negate P
$\neg(\exists x\ P(x)) = \forall x\ \neg P(x)$

In words such transformations are easy to confuse.

Advice for reading:
On first reading translate symbols into words aloud.
"For all epsilon greater than zero there exists delta greater than zero."

Advice for writing:
Use symbols for precision of structure.
Add verbal explanations for intuition.

Why Terms Are Called This Way (etymology)

Understanding the origin of terms helps remember their meaning.

Algebraic terms

Kernel (from Ger. Kern = grain, core):
Historical metaphor: "core" of what is lost under the mapping.
The kernel is what "collapses" into the neutral element.
$\ker(\varphi) = \{x : \varphi(x) = 0\}$ – the central part of "losses".

Image (from Lat. imago = reflection):
Literally: "picture", "reflection" of the original set.
$\mathrm{Im}(f) = f(A)$ – where set A "maps to".

Homomorphism (ὁμός = same, μορφή = form):
"Preserving form" — transfers operations without changing structure.
$\varphi(a \cdot b) = \varphi(a) \cdot \varphi(b)$ — the operation "looks the same" before and after.

Isomorphism (ἴσος = equal):
"Equal-form" — structures are completely identical algebraically.
Bijective homomorphism: nothing is lost, nothing is glued together.

Group (Fr. groupe, Ger. Gruppe):
Introduced by Galois (1830s) for "group of permutations of roots of an equation".
Originally: a collection of symmetries associated with an algebraic equation.

Ring (Ger. Ring):
Initially: cyclic structures of type $\mathbb{Z}/n\mathbb{Z}$, which "loop around".
Dedekind (1871) used for "rings of integers" in fields.

Field (Ger. Körper = body, Eng. field):
German term "body" = space for full-fledged arithmetic.
English "field" = domain/field of activity for all operations.
Field — where one can divide by everything nonzero.

Ideal (from Kummer, 1840s):
Originally "ideal numbers" — fictitious elements for restoring uniqueness of factorization into prime factors.
Dedekind formalized as a subset of a ring.

Geometric and topological terms

Topology (τόπος = place, λόγος = study):
"Science of places/position" — studies what is near what.
Introduced by Listing (1847): "Topology — geometry of position".

Manifold (from Ger. Mannigfaltigkeit):
A "manifold" can look different in different places, but locally always like $\mathbb{R}^n$. Riemann (1854).

Homotopy (ὁμός = same, τόπος = place):
Two paths are "same-placed" = can be continuously deformed one into the other.

Simplex (Lat. simplex = simple):
Simplest polytope in a given dimension:
point → segment → triangle → tetrahedron → ...

Compact (Lat. compactus = densely compressed):
A set is "tightly packed" — from any covering one can choose a finite subcovering. No "infinite holes" or "escape to infinity"

Connected:
A "whole" cannot be split into two disjoint open sets.
One can get from any point to any other without leaving the set.

Analytic terms

Continuous (from Lat. continuus = connected):
"Without breaks" – small changes in argument give small changes in value.

Differentiable (from Lat. differentia = difference):
One can compute the "difference quotient" and pass to the limit.
The function is "distinguishable" – one can see how it changes.

Integral (from Lat. integer = whole):
"Restoration of the whole" from parts (summation of infinitesimals).
Leibniz used ∫ as a stylized letter S (summa).

Convergence (from Lat. convergere = to incline toward):
Members of the sequence "incline toward" one point.

Limit (from Lat. limes = boundary):
"Boundary" toward which a sequence or function tends.

Category viewpoint – a language for everything that follows

Before moving on to specific structures (groups, spaces, manifolds),
let us introduce the language in which all of them are described uniformly.

Main idea:
Mathematical objects are important not in themselves, but through relations between them.

Category = Objects + Arrows (morphisms) between them

$$A \xrightarrow{f} B \xrightarrow{g} C$$

Objects: what we study (sets, groups, spaces)
Arrows: how objects are connected (functions, homomorphisms, continuous maps)

Key property: arrows can be composed ($g \circ f: A \to C$)

Examples of categories (which will be encountered below)

CATEGORY	OBJECTS	ARROWS
Set	Sets	Functions
Grp	Groups	Homomorphisms
Vect	Vector sp.	Linear mappings
Top	Top. sp.	Continuous mappings
Man	Manifolds	Smooth mappings

In all these cases the pattern is one:
- There are objects with some structure
- There are mappings preserving this structure
- Mappings can be composed

How a mathematician thinks — heuristics and methods of thinking

We have defined the basic objects: sets, numbers, relations, categories. Before building spaces on them (Part II) — a few words about how mathematical thinking itself is organized. What is an invariant? When to generalize? When to look for a counterexample? These heuristics permeate the entire atlas.

When to look for an invariant

An invariant is a quantity that does not change under transformations. If we rotate, deform, recompute in other coordinates — and some number or property remains the same — this is an invariant. Invariants separate the essential from the inessential.

If objects change, but something is preserved — look for an invariant.

Examples:
- Rotations change coordinates, but preserve length: $\|v\| = \mathrm{inv}$
- Deformations change shape, but preserve the number of holes: $\pi_1 = \mathrm{inv}$
- Time changes the system, but preserves energy: $H = \mathrm{inv}$ (if $\partial L/\partial t = 0$)

Heuristic: "What has not changed?" — the first question of a mathematician.

When to generalize, when to make concrete

Generalize, if:
- The proof works for a broader class
- Specific details are not used
- You want to understand the essence, discarding the "noise" of particulars

Make concrete, if:
- The general theorem does not give an explicit answer
- A computational result is needed
- The special case has additional structure

Example: Fixed point theorem (general) → Newton's method (concrete).
The general one says "exists", the concrete one says "how to find".

What to do when the proof does not go through

1. Check special cases
   Is the statement true for n=1,2,3? For the simplest examples?
   If not — look for a counterexample, not a proof.

2. Weaken the statement
   Perhaps it is true under additional conditions?
   Perhaps a weaker estimate is true?

3. Strengthen the statement
   Paradoxically, but sometimes a stronger statement is easier to prove.
   Induction often requires strengthening the hypothesis.

4. Reformulate
   The same problem in a different language (algebra ↔ geometry ↔ analysis).
   Sometimes a different view makes the solution obvious.

5. Study analogous theorems
   How were similar results proved? What ideas were used?

How to choose between formalizations

One problem — many languages (see introduction). How to choose?

Criteria:
- What operations are needed? (addition → linear alg, proximity → topology)
- What answer is needed? (existence → abstract, number → concrete)
- What is known? (symmetry → groups, smoothness → analysis)

Heuristic: Choose the language where the problem becomes standard.

Example: Heat equation
- Want to understand qualitative behavior → semigroups (e^{At})
- Want to compute a concrete solution → Fourier or numerically
- Want to prove existence → functional analysis

Principle of economy of structure

Do not introduce more structure than needed for the solution.

Bad: "Let V — be a Hilbert space." (if only norm is needed)
Good: "Let V — be a normed space."

Why this is important:
- The proof works for a broader class
- Easier to understand what exactly is being used
- The result is easier to apply in other contexts

Exception: If additional structure makes the proof simpler,
sometimes it is worth using it, and then generalizing the result.

Intuition vs rigor

A mathematician works in two stages:

Stage 1 (intuition): "Why should this be true?"
Pictures, analogies, physical reasoning, examples.
Goal: understand, not prove.

Stage 2 (rigor): "How to prove this?"
Formal definitions, logical steps, checking all cases.
Goal: convince (oneself and others).

Beginner's mistake: skipping stage 1.
Without intuition proof is blind enumeration.

Physicist's mistake: stopping at stage 1.
Intuition sometimes deceives (example: paradoxes of 19th century analysis).

PART II: SPACES

Introduction: Different views on one space

Three Problems on Spaces

What can be done with mathematical objects?

Any work with a space (and with a mathematical object in general)
Reduces to one of three problems:

Classification – what kind of object is this? what is possible in principle?
Computation – find a specific value within the known
Construction – create a new object from existing ones

These three problems exhaust everything that can be done in mathematics.

Classification – normative boundaries

Classification answers: What is this? What is possible in principle?

Result: "this is X, not Y", "this is possible / impossible", "there are exactly N of these"

Examples:

- Closed surfaces are exactly this many: spheres with handles + nonorientable.
- A rigid body has 6 degrees of freedom (not 5, not 7)
- Equation of 5th degree is not solvable in radicals (group S_5 is unsolvable)
- Field $F = \operatorname{grad} \varphi$ exists $\Longleftrightarrow$ $\operatorname{rot} F = 0$ and domain is simply connected
- Sphere and torus are not homeomorphic (different π_1, different χ)

Classification establishes boundaries – what can be sought, and what is pointless.

Tools: topological invariants, symmetry groups, existence/uniqueness theorems.

Computation – find the specific

Computation answers: Where exactly? How many? What value?

Result: number, coordinates, formula, specific object

Examples:

- Find roots of equation $x^3 - 2x + 1 = 0$
- Compute integral $\int_0^\infty e^{-x^2} dx = \sqrt{\pi}/2$
- Find eigenvalues of a matrix
- Determine shortest path on a surface (geodesic)
- Find minimum of function on a compact

Computation works within boundaries established by classification.
If classification says "does not exist" — there is nothing to compute.

Tools: algorithms, optimization methods, numerical methods,
analytical techniques.

Construction — create new

Construction answers: How to obtain a new object from existing ones?

Result: new object that did not exist before

Examples:

- a × b — new vector from two given
- V ⊗ W — new space from two given
- G/H — quotient group (new group from group and subgroup)
- ℚ → ℝ — completion (new space from old)
- Product M × N — new manifold
- Tangent bundle TM — new space over M
- Additional constructions in geometry (draw a line, drop ⊥)

Difference from computation: computation finds existing (root already exists,
we are searching for it). Construction creates — before operation a × b this vector did not exi

Tools: operations (×, ⊗, ∧, /, ×), constructions (completion,
covering, extension), universal properties.

Connection of three problems

```
                    Classification
                    "what is possible?"
                         |
            +------------+------------+
            ↓                         ↓
      Construction              Computation
    "create new"              "find within"
            |                         |
            +----------+-------------+
                       ↓
                 New objects
                       |
                       ↓
            Classification of new.
```

Classification establishes boundaries →
  Construction creates objects within boundaries →
    Computation finds specific values →
      Results may require new classification

Classification (what is possible, invariants, types):
  Groups – classification of symmetries and motions
  Topology – classification of spaces by shape
  Number theory – classification of numbers

Computation (find value, solve equation):
  Linear algebra – systems of equations, eigenvalues
  Analysis – derivatives, integrals, series
  Functional analysis – equations in infinite dimension

Construction (create new object):
  Vector products – ⟨,⟩, ×, ∧, ⊗
  Duality – V → V*
  Tensors – multilinear constructions
  Manifolds – gluing from charts
  Diff. forms – forms from vectors

Most fields include all three problems in different proportions.

Table of spaces – central table

Space – central object of mathematics

All mathematics studies spaces and structures on them:

- Topology: shape of space (holes, connectedness)
- Algebra: symmetries of space (groups)
- Analysis: functions on space
- Geometry: measurements on space (metric, curvature)

Hierarchy of Spaces – what is added at each level

```
Set (simply a collection of points)
    | + topology (notion of "closeness", open sets)
    ↓
Topological space
    | + local Euclideanness (looks like ℝⁿ locally)
    ↓
Manifold
    | + metric (way to measure distances)
    ↓
Riemannian manifold
    | + physical equations
    ↓
Spacetime (GR)

Parallel branch:

Set
    | + linear structure (addition, multiplication by number)
```

```
       ↓
   Vector space
       | + scalar product
       ↓
   Euclidean space ℝⁿ
       | + infinite dimension
       ↓
   Hilbert space (quantum mechanics)

===============================================================================
Main Spaces — catalog
===============================================================================

 Notation in table:

   dim = dimension (how many coordinates needed to describe a point)
   π₁  = fundamental group (which loops cannot be contracted to a point?)
         means "all loops contract", ℤ — "there is one unclosed"
   H₁  = first homology group (similar to π₁, but abelian version)
   χ   = Euler characteristic = V − E + F (vertices − edges + faces)
         Shape invariant: sphere χ=2, torus χ=0, projective plane χ=1

   Invariants — table

+-------------+-----+-------+-------+-----+-------------------------------------+
| SPACE       |dim  | π₁    | H₁    |  χ  | WHAT IT IS / WHERE IT APPEARS       |
```

Point {*}	0	0	0	1	Trivial; zero-dimensional "world"
Line $\mathbb{R}$	1	0	0	–	Noncompact; time, temperature
Circle S^1	1	$\mathbb{Z}$	$\mathbb{Z}$	0	$S^1 = \{\lvert z\rvert=1\}$; angles, phases, periods Group U(1).
Plane $\mathbb{R}^2$	2	0	0	–	Noncompact; ordinary geometry
Sphere S^2	2	0	0	2	Surface of ball; Earth, sky All loops contract.
Torus T^2	2	$\mathbb{Z}^2$	$\mathbb{Z}^2$	0	Donut = $S^1\times S^1$; two angles Periodic boundary conditions
Projective plane	2	$\mathbb{Z}/2$	$\mathbb{Z}/2$	1	$\mathbb{R}P^2$ = "directions of lines" Nonorientable.
Klein bottle K	2	$\mathbb{Z}\rtimes\mathbb{Z}$	$\mathbb{Z}\oplus\mathbb{Z}/2$	0	Nonorientable; 4D needed
3-sphere S^3	3	0	0	0	≅ SU(2). Space of rotations (up to ±)
SO(3)	3	$\mathbb{Z}/2$	$\mathbb{Z}/2$	0	All rotations in $\mathbb{R}^3$ ≅ $\mathbb{R}P^3$ (not S^3.)
$\mathbb{R}^4$ (space-time SR)	4	0	0	–	Spacetime (SR) Flat, Minkowski metric

Classification of closed surfaces (dim = 2)

Theorem: Any closed surface is:

Orientable: Sphere with g handles (genus g)

$g = 0$	$g = 1$	$g = 2$	
○	⊚	⊚⊚	$\chi = 2 - 2g$
sphere	torus	"pretzel"	

Non-orientable: Sphere with k "crosses" (Möbius glued in)

$k = 1$	$k = 2$	
$\mathbb{R}P^2$	K	$\chi = 2 - k$
project.	Klein	

This is a complete classification. There are no other closed surfaces.

Connection of spaces with groups

Many important spaces are simultaneously groups:

$S^1 \cong U(1) \cong SO(2)$ – circle = rotation group of the plane
$S^3 \cong SU(2)$ – 3-sphere = group (double cover of SO(3))
$SO(3) \cong \mathbb{R}P^3$ – rotations in 3D = projective space
GL(n), SL(n), O(n) – matrix groups = manifolds

These are Lie groups – groups which are simultaneously manifolds.

The fundamental group π_1 also connects:
$\pi_1(S^1) = \mathbb{Z}$ – integers as a group
$\pi_1(T^2) = \mathbb{Z}^2$ – lattice
$\pi_1(\vee_n S^1) = F_n$ – free group

Three attitudes toward space

Problem: define the subject of each area of mathematics

AREA	VERB	QUESTION
Differential geometry	Measures	How is it arranged inside? (curvature, metric, angles)

Algebraic topology	Distinguishes	How does it differ from others? (invariants: π_1, H_n)
Group theory	Transforms	What can be done with it? (symmetries, actions)

The word "algebraic":

ALGEBRAIC TOPOLOGY	Space → Group. Algebra as a tool.
ALGEBRAIC GEOMETRY	Equations → Space. Algebra as a source.

Group – symmetries of space

If a set is dust, and topology is fabric, then a group is mobility.
A group answers the question: how can one move in space without breaking it?
What transformations are allowed?

In terms of "object–observer" the group is the key concept.

Group is a catalog of observer's motions that do not change the object

If the observer turns – the object looks different, but the object itself has not changed.
The group SO(3) is all possible rotations. Each element of the group – a specific rotation. Composition – perform one rotation, then another.

Why this is fundamental: physical laws should not depend on how the observer stands. "Covariance" in physics is the requirement that equations look the same for all observers connected by a group.

An invariant of a group – that which does not change under any motions from the group.
For example, distance is an invariant of the rotation group. Two observers, rotated relative to each other, will measure the same distance.

Group as a view of space

A group is a set of transformations of space preserving its structure. Different groups "see" different things in one space:

- SO(3) sees rotations in $\mathbb{R}^3$ (preserves distances and orientation)
- GL(n) sees linearity (preserves lines and origin)
- Crystal symmetries see discrete lattice

Where did groups come from — history

Equations of degrees 1, 2, 3, 4 are solved by formulas (Cardano, Ferrari).
Equation of degree 5: $x^5 + ax^4 + bx^3 + cx^2 + dx + e = 0$

For almost 300 years people searched for a formula. Galois (at age 20) proved: it does not ex

How? He connected the equation with the group of permutations of its roots:
If this group is "solvable" → a formula exists.
The permutation group of 5 elements is not solvable → there is no formula.

Thus group theory was born — from the question about solving equations.

Physical view: degrees of freedom of a rigid body

Take any object — a book, a stone, a molecule.
What motions are possible with it, if it cannot be deformed?

Translations: can be moved to any point in space.
Three directions: forward-backward, left-right, up-down → 3 numbers.

Rotations: can be rotated around any axis.
Axis (2 parameters) + angle (1 parameter) → 3 numbers.

Total: 6 parameters. The position of a rigid body is described by 6 numbers.
Not 5, not 7 — exactly 6. This is a fact about the structure of space $\mathbb{R}^3$.

The set of all such motions is called SE(3) — the special Euclidean group.

Why this is a group — properties of motions

Motions possess certain properties:

1. Two motions in succession — also a motion
   First translate, then rotate — we get some motion.
   We don't go outside SE(3).

2. The order of grouping doesn't matter
   (A then B) then C = A then (B then C)
   This is a property of composition of any mappings.

3. There is "do nothing"
   The identity motion — leave everything in place.

4. Any motion can be undone
   Shifted 3 meters to the right → shift 3 meters to the left.
   Rotated by 30° → rotate by −30°.

These four properties are the axioms of a group.
They are not invented, but follow from the nature of the notion "motion".

Different constraints – different groups

What motions are "allowed" depends on what needs to be preserved:

WHAT WE PRESERVE	GROUP	WHERE IT OCCURS
Distances and angles	E(3)	Rigid body (with reflections)
+ orientation	SE(3)	Rigid body (without reflections)
Only angles	Conformal	Cartography, complex analysis
Parallelness	Aff(3)	Shadows in sunlight
Only straightness	PGL(3)	Perspective in painting
Volume	SL(3)	Incompressible fluid
Linearity	GL(3)	Any linear deformation

Fewer constraints → larger group:
  Linear: $SO(3) \subset O(3) \subset SL(3) \subset GL(3)$
  Affine: $SE(3) \subset E(3) \subset Aff(3)$
  Connection: $GL(3) \subset Aff(3)$, but $E(3) \not\subset GL(3)$ (isometries include translations)

This is a classification: the group describes what transformations are possible.

Only rotations – the group SO(3)

If an object is fixed at one point (top, gyroscope, satellite),
only rotations remain. This is the group SO(3).

$$SO(3) = \{ \text{rotations of } \mathbb{R}^3 \text{ around the origin} \} = \{ \text{orthogonal matrices } 3\times3 \text{ with } \det = +1 \}$$

Dimension: 3 (three Euler angles, or axis + angle).

Important fact: SO(3) – a non-abelian group.

  Rotate around X, then around Y ≠ rotate around Y, then around X

  You can check with a book:
  1) place the book, rotate by 90° around the vertical, then
     by 90° around the horizontal axis "away from you"
  2) do it in reverse order
  The results are different.

This is not an abstraction – satellite control must take this into account.

Discrete symmetries – finite groups

Not all objects allow any rotations.
A salt crystal (cube) looks the same only under certain rotations.

Symmetries of a square – the group D₄:

```
   1---2     Original         4---1    Rotation by 90°
   |   |     square           |   |    clockwise
   |   |                      |   |
   4---3                      3---2
```

Possible transformations:

- 4 rotations: by 0°, 90°, 180°, 270°
- 4 reflections: with respect to horizontal, vertical, two diagonals

Total 8 elements. Not 7, not 9 – exactly 8.
This is the complete answer to the question "what are the symmetries of a square".

Symmetries of regular polyhedra

POLYHEDRON	GROUP	NUMBER OF SYMMETRIES
Tetrahedron	Td	24
Cube / Octahedron	Oh	48 (they are dual)
Dodecahedron / Icosahedron	Ih	120

The cube and octahedron have the same symmetry group – they are "geometrically equivalent" in the sense of symmetries (dual polyhedra).

Group – motivation and examples

Visualization: symmetries of the square

```
   1---2     Initial          2---3    90° rotation     3---4   180° rotation
   |   |     square           |   |    counterclockwise |   |
   |   |                      |   |                     |   |
   4---3                      1---4                     2---1

   2---1     Reflection       4---3    Reflection       Total 8 symmetries:
   |   |     (vert. axis)     |   |    (horiz. axis)    • 4 rotations
   |   |                      |   |                     • 4 reflections
   3---4                      1---2
```

Any two symmetries can be combined → result is a symmetry.
Each symmetry has an inverse (return back).
There is an identity symmetry (do nothing).

This is the group D₄ – the dihedral group (symmetries of the square).

Group as a set of transformations

A group formalizes the notion of "set of invertible transformations".

AXIOM	MEANING IN TERMS OF TRANSFORMATIONS
Closure	Composition of two transformations — transform.
Associativity	$(f\circ g)\circ h = f\circ(g\circ h)$
Neutral el.	Identity transformation id
Inverse element	Each transformation is invertible

Definition: Symmetry group of object X — set of all bijections $X \to X$,
preserving the structure of X.

Necessity of axioms

AXIOM	WHY NECESSARY
Composition	Sequential application of symmetries — symmetry
Neutral	Identity mapping preserves structure
Inverse	Inverse of a symmetry — symmetry
Associativity	Follows from associativity of function composition

Formal definition

Definition of group

Group $(G, \cdot)$ — set G with binary operation $\cdot : G \times G \to G$,
satisfying axioms:

Notation remark: $(G, \cdot)$ denotes a tuple (ordered pair),
where the first element — carrier set, second — operation (signature).
This is not the same as the pair $\{\{G\}, \{G, \cdot\}\}$ from set theory.

Here brackets mean "structure = carrier + operations".

AXIOM	FORMULA	MEANING
G1. Closure	$\forall a,b \in G: a\cdot b \in G$	Result – element of G
G2. Associativity	$\forall a,b,c \in G: (a\cdot b)\cdot c = a\cdot(b\cdot c)$	Brackets don't matter
G3. Neutral element	$\exists e \in G: \forall a \in G: e\cdot a = a\cdot e = a$	"Do nothing"
G4. Inverse element	$\forall a \in G\ \exists a^{-1} \in G: a\cdot a^{-1} = a^{-1}\cdot a = e$	Everything cancels

Remarks:

- Neutral element is unique
- Inverse element for each a is unique
- $(a^{-1})^{-1} = a$
- $(a\cdot b)^{-1} = b^{-1}\cdot a^{-1}$ (order reverses)

Abelian vs non-abelian groups

Abelian group: $\forall a,b: a\cdot b = b\cdot a$ (commutativity)

ABELIAN (order irrelevant)	NON-ABELIAN (order matters)
$(\mathbb{Z}, +)$: $2+3 = 3+2$	D_4: rotation∘reflection ≠ reflection∘rotation
$(\mathbb{R}, +)$: $\pi+e = e+\pi$	S_n ($n\geq3$): permutations
$(\mathbb{R}^*, \times)$: $2\times3 = 3\times2$	GL(n): matrices $AB \neq BA$
$(\mathbb{Z}/n, +)$: cyclic	SO(3): rotations in 3D

==

Examples of Groups — in Detail

==

Example 1: Integers ($\mathbb{Z}$, +)

Set: $\mathbb{Z} = \{..., -2, -1, 0, 1, 2, 3, ...\}$
Operation: addition +

Checking axioms:

G1. Closure: $a + b \in \mathbb{Z}$ for any $a, b \in \mathbb{Z}$ ✓
(sum of integers — integer)

G2. Associativity: $(a+b)+c = a+(b+c)$ ✓
$(2+3)+4 = 5+4 = 9$
$2+(3+4) = 2+7 = 9$

G3. Identity: $e = 0$, because $a + 0 = 0 + a = a$ ✓

G4. Inverse: $a^{-1} = -a$, because $a + (-a) = 0$ ✓

Additionally: $a + b = b + a$ → Abelian group

Visualization: Shifts along the number line

```
←----------●----------→
... -2  -1  0  1  2  3 ...
```

+3 = "shift right by 3"
−2 = "shift left by 2"
0 = "stay in place"

Example 2: Cyclic group $\mathbb{Z}/n$ (or $\mathbb{Z}_n$)

Set: $\mathbb{Z}/n = \{0, 1, 2, ..., n-1\}$
Operation: addition modulo n

Example: $\mathbb{Z}/4 = \{0, 1, 2, 3\}$ — "clock arithmetic" with 4 divisions

Cayley table (group multiplication table):

+	0	1	2	3

0	0	1	2	3
1	1	2	3	0
2	2	3	0	1
3	3	0	1	2

0 — identity element
(row and column for 0
coincide with headers)

Check: $2 + 3 = 5 \bmod 4 = 1$ ✓ (see table)
$3 + 3 = 6 \bmod 4 = 2$ ✓

Visualization:

```
    0
 3     1
    2
```

Shifts around a circle of 4 points

Identity: $e = 0$
Inverses:
$0^{-1} = 0$ $(0+0=0)$
$1^{-1} = 3$ $(1+3=4\equiv 0)$
$2^{-1} = 2$ $(2+2=4\equiv 0)$
$3^{-1} = 1$ $(3+1=4\equiv 0)$

Generator: Element g whose powers give the entire group.
In $\mathbb{Z}/4$: $g = 1$, because $1, 1+1=2, 1+1+1=3, 1+1+1+1=0$
Also $g = 3$ works: $3, 3+3=2, 3+3+3=1, 3+3+3+3=0$
But $g = 2$ doesn't work: $2, 2+2=0, 2+2+2=2, \ldots$ (not all elements)

Example 3: Symmetries of equilateral triangle D_3

Symmetries of triangle — all ways to place it "in the same manner":

```
     1
    /\
   /  \
  /    \
 /______\
3        2
```

Rotations: e = don't rotate
r = by 120° counterclockwise
r^2 = by 240° ($= r\circ r$)

Reflections: s_1 = with respect to altitude from 1
s_2 = with respect to altitude from 2
s_3 = with respect to altitude from 3

Total 6 elements: $D_3 = \{e, r, r^2, s_1, s_2, s_3\}$

Cayley table (composition: first column, then row):

$\circ$	e	r	r^2	s_1	s_2	s_3
e	e	r	r^2	s_1	s_2	s_3
r	r	r^2	e	s_3	s_1	s_2
r^2	r^2	e	r	s_2	s_3	s_1
s_1	s_1	s_2	s_3	e	r	r^2
s_2	s_2	s_3	s_1	r^2	e	r
s_3	s_3	s_1	s_2	r	r^2	e

Observations:
- $r \circ s_1 = s_3$, but $s_1 \circ r = s_2$ → non-abelian
- Subgroup of rotations $\{e, r, r^2\} \cong \mathbb{Z}/3$ — abelian
- Each reflection: $s_i^2 = e$ (apply twice = nothing)

Physical meaning: D_3 describes the symmetry of a molecule with triangular structure
(for example, BF_3 — boron trifluoride)

Example 4: Permutation groups S_n

S_n = all permutations of n elements
$|S_n|$ = n! elements

Example: S_3 — all permutations of {1, 2, 3}

Notation: $\sigma = (\sigma(1), \sigma(2), \sigma(3))$

$e = (1, 2, 3)$ — identity permutation
$\sigma_1 = (1, 3, 2)$ — swaps 2 and 3
$\sigma_2 = (3, 2, 1)$ — swaps 1 and 3
$\sigma_3 = (2, 1, 3)$ — swaps 1 and 2
$\sigma_4 = (2, 3, 1)$ — cyclic shift $1\to2\to3\to1$
$\sigma_5 = (3, 1, 2)$ — cyclic shift $1\to3\to2\to1$

Composition: $(\sigma \circ \tau)(x) = \sigma(\tau(x))$ — first τ, then σ

Cycle notation (more compact):
$\sigma_3 = (1\ 2)$ — transposition (swaps $1\leftrightarrow2$, rest in place)
$\sigma_4 = (1\ 2\ 3)$ — 3-cycle $(1\to2,\ 2\to3,\ 3\to1)$

Fact: $S_3 \cong D_3$ (permutation group of 3 elements is isomorphic to symmetries of △)

Critically important:
S_5 — unsolvable group (its normal series does not reach {e} through abelian factors). By the Abel–Ruffini theorem this means that the quintic equation is not solvable by radicals.

Example 5: Nonzero real numbers $(\mathbb{R}^*, \times)$

Set: $\mathbb{R}^* = \mathbb{R} \setminus \{0\}$ = all reals except zero
Operation: multiplication ×

Checking axioms:

G1. Closure: $a \times b \in \mathbb{R}^*$ for $a, b \neq 0$ ✓
(product of nonzero is nonzero)

G2. Associativity: $(a\times b)\times c = a\times(b\times c)$ ✓

G3. Identity: $e = 1$, because $a \times 1 = 1 \times a = a$ ✓

G4. Inverse: $a^{-1} = 1/a$, because $a \times (1/a) = 1$ ✓
(this is why 0 is excluded — it has no inverse)

Why $(\mathbb{R}, \times)$ is not a group:
$0 \times (\text{anything}) = 0$, but $0 \times ? = 1$ has no solution.
Zero has no inverse element.

What is not a group – counterexamples

STRUCTURE	Why not a group	How to fix
$(\mathbb{N}, +)$ natural numbers	No inverses: $3 + ? = 0$ unsolvable	→ $(\mathbb{Z}, +)$ add negatives
$(\mathbb{Z}, \times)$ integers	Inverses not integer: $2^{-1} = \frac{1}{2} \notin \mathbb{Z}$	→ $(\mathbb{Q}^*, \times)$ switch to rationals
$(\mathbb{R}, \times)$ real numbers	0 has no inverse $0 \times ? = 1$ unsolvable	→ $(\mathbb{R}^*, \times)$ remove 0
n×n matrices with ×	$\det=0$ → no inverse	→ $GL(n)$ only invertible $(\det\neq 0)$

Subgroups

Definition:
$H \leq G$ is called a subgroup of G (notation $\leq$, not $\subseteq$), if H is a group with respect to the same operation (inherited from G).

Notation $H \leq G$ is standard for subgroups, $H \subseteq G$ – for subsets.

Subgroup criterion (convenient for verification):

$$\boxed{H \leq G \iff H \neq \varnothing \text{ and } \forall a,b \in H: a\cdot b^{-1} \in H}$$

(One condition instead of four axioms)

Why it works:

- $H \neq \varnothing \Rightarrow \exists a \in H \Rightarrow a\cdot a^{-1} = e \in H$ (neutral exists)
- $e \in H \Rightarrow e\cdot b^{-1} = b^{-1} \in H$ (inverses exist)
- $a, b^{-1} \in H \Rightarrow a\cdot(b^{-1})^{-1} = a\cdot b \in H$ (closure)

Examples of subgroups

Group	subgroup	notation
$(\mathbb{Z}, +)$	Even numbers $2\mathbb{Z}$	$2\mathbb{Z} < \mathbb{Z}$
	Multiples of 3: $3\mathbb{Z}$	$3\mathbb{Z} < \mathbb{Z}$
	Multiples of n: $n\mathbb{Z}$	$n\mathbb{Z} < \mathbb{Z}$
$(\mathbb{R}^*, \times)$	Positive $\mathbb{R}^+$	$\mathbb{R}^+ < \mathbb{R}^*$
	$\{1, -1\}$	$\{\pm 1\} < \mathbb{R}^*$
D_4 (symmetries □)	Rotations $\{e, r, r^2, r^3\}$	$\cong \mathbb{Z}/4$
	$\{e, r^2\}$	$\cong \mathbb{Z}/2$
	$\{e, s\}$ for any reflection	$\cong \mathbb{Z}/2$
GL(n) (invertible matrices)	$SL(n) = \{A : \det A = 1\}$	special linear group
	$O(n) = \{A : A^T A = I\}$	orthogonal group
	$SO(n) = O(n) \cap SL(n)$	special orthogonal

Trivial subgroups:
{e} – trivial subgroup (exists in any group)
G – the group itself (improper subgroup)

Lagrange's Theorem

```
+-------------------------------------------------------------------+
|  If G is a finite group and H is a subgroup of G, then:           |
|                                                                   |
|             |H| divides |G|                                       |
|                                                                   |
|  Moreover: |G| = |H| × [G : H], where [G : H] is the subgroup index|
+-------------------------------------------------------------------+
```

Corollaries:

- The order of an element divides the order of the group
  (order of element a = smallest n: $a^n = e$)

- A group of prime order p is cyclic
  (no other subgroups except {e} and G)

- In S_4 (24 elements) subgroups can have order
  1, 2, 3, 4, 6, 8, 12, 24 (divisors of 24)
  Subgroups of order 5 or 7 are impossible

Homomorphisms and Isomorphisms

Homomorphism – a mapping that preserves structure

A mapping $\varphi: G \to H$ is called a group homomorphism if:

$$\forall a, b \in G: \quad \varphi(a \cdot b) = \varphi(a) * \varphi(b)$$

"Image of product = product of images"

Properties (follow automatically):

- $\varphi(eG) = eH$ (image of neutral is neutral)
- $\varphi(a^{-1}) = \varphi(a)^{-1}$ (image of inverse is inverse)

Examples of homomorphisms:

HOMOMORPHISM	VERIFICATION: $\varphi(a \cdot b) = \varphi(a) * \varphi(b)$
exp: $(\mathbb{R},+) \to (\mathbb{R}^{+},\times)$	$\exp(a+b) = \exp(a) \times \exp(b)$ ✓
det: $(GL(n),\times) \to (\mathbb{R}^{*},\times)$	$\det(AB) = \det(A) \times \det(B)$ ✓
sign: $(S_n,\circ) \to (\{\pm 1\},\times)$	$\text{sign}(\sigma \circ \tau) = \text{sign}(\sigma) \times \text{sign}(\tau)$ ✓
mod n: $(\mathbb{Z},+) \to (\mathbb{Z}/n,+)$	$(a+b) \bmod n = (a \bmod n)+(b \bmod n)$ ✓

Kernel and image

For homomorphism $\varphi: G \to H$:

CONCEPT	DEFINITION	PROPERTIES
$\text{Ker}(\varphi)$ (kernel)	$\{a \in G : \varphi(a) = eH\}$ What maps to neutral	Normal subgroup of G φ inject. $\Longleftrightarrow$ Ker=$\{e\}$
$\text{Im}(\varphi)$ (image)	$\{\varphi(a) : a \in G\} \subseteq H$ Where G maps to	Subgroup of H φ surject. $\Longleftrightarrow$ Im=H

Example: $\varphi: \mathbb{Z} \to \mathbb{Z}/6$, $\varphi(k) = k \bmod 6$

$\text{Ker}(\varphi) = 6\mathbb{Z} = \{..., -12, -6, 0, 6, 12, ...\}$

$\text{Im}(\varphi) = \mathbb{Z}/6 = \{0, 1, 2, 3, 4, 5\}$

★ First Isomorphism Theorem — fundamental result

Theorem: For any homomorphism φ: G → H the following holds:

$$G / \mathrm{Ker}(\varphi) \cong \mathrm{Im}(\varphi)$$

Geometric intuition:

```
      G                           H
  +-------------+           +-------------+
  |  ○ ○ ○ ○    |     φ     |             |
  |  | | | |    | ---------▶ |   ●   ●     |   Im(φ) — where we landed
  |  ○ ○ ○ ○    |           |   |   |     |
  |  | | | |    |           |   ●   ●     |
  |  ○ ○ ○ ○    |           |             |
  +-------------+           +-------------+
      ↓
  Glue elements
  with the same image
      ↓
  +---------+
  |  ◉   ◉  |  G/Ker(φ)  — cosets
  |  ◉   ◉  |  (elements mapping to one point are glued)
  +---------+
```

Meaning: "Factorization by kernel removes everything redundant and leaves only the image"

Factor group G/H is the quotient set $G/\sim$, where
equivalence relation $\sim$ is defined as $g_1 \sim g_2 \iff g_1 g_2^{-1} \in H$.
Key condition: for the group operation to be well-defined on $G/\sim$,
subgroup H must be normal ($gHg^{-1} = H$ for all $g \in G$).

Example: $\varphi: \mathbb{Z} \to \mathbb{Z}/6$, $k \mapsto k \bmod 6$

$\mathrm{Ker}(\varphi) = 6\mathbb{Z} = \{..., -6, 0, 6, 12, ...\}$
$\mathrm{Im}(\varphi) = \mathbb{Z}/6$

Theorem: $\mathbb{Z} / 6\mathbb{Z} \cong \mathbb{Z}/6$ ✓
(factor group by kernel is isomorphic to image)

Isomorphism — when groups are "the same"

Isomorphism φ: G → H — homomorphism + bijection. We write $G \cong H$.

"G and H – one group, differ only by names of elements"

ISOMORPHISM	MAPPING	WHY IT WORKS
$(\mathbb{Z}, +) \cong (2\mathbb{Z}, +)$	$\varphi(n) = 2n$	Even ↔ all integers
$(\mathbb{R}, +) \cong (\mathbb{R}^+, \times)$	$\varphi = \exp$, $\varphi^{-1} = \ln$	Addition ↔ multiplication
$\mathbb{Z}/6 \cong \mathbb{Z}/2 \times \mathbb{Z}/3$	$k \mapsto (k \bmod 2, \bmod 3)$	Chinese remainder thm
$S^1 \cong U(1) \cong SO(2)$	$e^{i\theta}$ ↔ rotation by θ	Circle ≅ rotations

$\mathbb{Z}/2 \cong \{\pm 1\}$ (two elements, one structure)

$\varphi(0) = 1, \quad \varphi(1) = -1$

$S_3 \cong D_3$ (6 elements, symmetries of △)

Non-isomorphic:

$\mathbb{Z}/4 \ncong \mathbb{Z}/2 \times \mathbb{Z}/2$ (different structure)

In $\mathbb{Z}/4$ there is an element of order 4 (generator).

In $\mathbb{Z}/2 \times \mathbb{Z}/2$ all elements have order ≤ 2.

Groups in physics and life

Applications of groups

FIELD	GROUP	WHAT IT DESCRIBES
Crystallography	230 space groups	All crystal symmetries NaCl: cubic symm. → optics

Standard model of particle physics	U(1) SU(2) SU(3) U(1)×SU(2)×SU(3)	Electromagnetism (phase) Weak interaction Strong interaction (quarks) Entire Standard model
Relativity theory	SO(3,1)	Lorentz: preserves speed of light 3 space + 1 time
Quantum mechanics	SU(2)	Particle spin e^-: spin ½ → rotation 720°.
Music	$\mathbb{Z}/12$	12 semitones, transposition
Cryptography	$(\mathbb{Z}/n)^*$ Elliptic curves	RSA: multiplicative group Groups of points on curves
Thermal physics (example for engineer)	Similarity group (scaling)	Dimensional analysis = Lie groups Finding formula Nu = f(Re, Pr) = choosing Lie group orbit by values of invariants (Re, Pr) Phys.quantities (α, λ, v, L) — coordinates on manifold, symmetric with respect to action of scaling group
Combinatorics	Any G	Counting "up to symmetry" Burnside's lemma

Sylow theorems — structure of finite groups

Let $|G| = p^n \cdot m$, where p is prime and gcd(p, m) = 1.

Sylow subgroup: subgroup of order p^n (maximal p-power)

Theorem 1: Sylow subgroup exists

For any prime p dividing $|G|$, there exists a subgroup of order p^n.

Theorem 2: All Sylow subgroups are conjugate

Any two p-Sylow subgroups P and Q are related: $Q = gPg^{-1}$ for some g.

Theorem 3: The number of Sylow subgroups n_p satisfies

- $n_p \equiv 1 \pmod{p}$
- n_p divides $m = |G|/p^n$

Application – classification of small groups:

$|G| = 15 = 3\cdot5$: $n_3 \mid 5$ and $n_3 \equiv 1 \pmod{3} \Rightarrow n_3 = 1$
$n_5 \mid 3$ and $n_5 \equiv 1 \pmod{5} \Rightarrow n_5 = 1$
Unique Sylow subgroups ⇒ normal ⇒ $G \cong \mathbb{Z}_{15}$

Sylow theorems – powerful tool: from the size of a group one can derive its structure.

Where it leads – connection with other areas

DIRECTION	CONNECTION	IDEA
Lie groups	→ (manifolds)	Group + smooth structure SO(3), SU(2), GL(n)
Representation theory	→ (lin.alg.)	$g \mapsto$ matrix $\rho(g)$ Group through lin. algebra
Noether's theorem	→ (DE)	Symmetry → conservation law Shift t → energy Rotation → angular momentum
Fund. group π_1	→ (topology)	Loops form a group Classification of spaces

Summary: hierarchy of algebraic structures

```
 From simple to complex


   Set
       |
       | + one operation + group axioms
       ↓
   Group (G, ·)
       |
       | + commutativity
       ↓
   Abelian group
       |
       | + second operation with distributivity
       ↓
   Ring (R, +, ×)
       |
       | + inverses for ×
       ↓
   Field (F, +, ×) ← multiplication commutative

       Important: Field = commutative multiplication (ab = ba).
       If we remove commutativity ⇒ Division Ring.
       Example of division ring: Quaternions ℍ (ij ≠ ji).

   Further — another type of object (not "special case of field",
   but new set V with action of field F on it):

   Set V + field F + addition in V + multiplication by scalars from F
       ↓
   Vector space (V over F)
       |
       | + norm ‖·‖
       ↓
   Normed space
       |
       | + completeness (all limits exist)
       ↓
   Banach space
       |
       | + inner product (‖x‖² = ⟨x,x⟩)
       ↓
   Hilbert space

 Each level inherits structure from previous + adds new.
```

Main idea

Group = minimal structure for describing symmetries.

If you can:
- Combine transformations (composition)
- Do nothing (neutral)
- Undo actions (inverse)

— we have a group.

Symmetries of an object determine its properties.
Symmetry group is the "DNA" of an object.

Applied example: turbine rotor balancing

Problem: Turbine rotor with 6 blades. During rotation vibration occurs.
Need to understand which blade defects cause which vibration frequencies.

```
        ●
       /|\
     /  |  \
   ●    |    ●
   |    ○    |
   ●    |    ●
     \  |  /
       \|/
        ●
```

Symmetry group of rotor:
$C_6 = \{e, r, r^2, r^3, r^4, r^5\}$
where r = rotation by 60°

This is cyclic group of order 6

Key fact: Rotor vibration decomposes by representations of group C_6

REPRESENTATION	PHYSICAL MEANING	VIBRATION FREQUENCY
Trivial (symmetric)	All blades equally deflected	No vibration (perfect balance)
Alternating	Alternation ± "every other"	f = 3 × rev/s (3-fold)
2-dimensional representations	Imbalance "wave" around circumference	f = n × rev/s (1×, 2×)

Practical application:
- If vibration at frequency 1× rev/s → static imbalance (one blade)
- If vibration at frequency 2× rev/s → pair of opposite blades
- If vibration at frequency 3× rev/s → every second blade

Group theory allows classification of imbalance types before measurements

Another example: three-phase electrical network

Three phases: A, B, C with 120° shift

Symmetry group: $C_3 = \{e, r, r^2\}$ where r = phase shift by 120°

- Symmetric load (all phases identical) → current in neutral = 0
- Symmetry violation → current in neutral ≠ 0

Method of symmetrical components (Fortescue): decomposition of asymmetric system into symmetric components — this is decomposition by representations of C_3.

Hierarchy of algebraic structures

Why different structures are needed

Each structure — answer to the question: "What do we want to be able to do?"

Want to add → Semigroup
+ there is "zero" → Monoid
+ can subtract → Group
+ order doesn't matter → Abelian group
+ can multiply → Ring
+ can divide (except by 0) → Field

The more operations — the more we can do, but the fewer examples.

Hierarchy: what is added at each step

```
 Semigroup ------→ monoid ------→ group ------→ abelian group
      |                |              |                  |
  associativity    + identity    + inverses      + commutativity
  a(bc)=(ab)c        a·e=a       a·a⁻¹=e           ab=ba
      |                |              |                  |
  Example:         Example:        Example:          Example:
```

```
   (ℕ⁺, ·)          (ℕ, +, 0)     (ℤ, +, 0)         (ℤ, +)
```

```
  Abelian group ------→ ring ------→ field
         |                   |             |
     (one operation)   + multiplication  + division
                       distributive      (except 0)
                       a(b+c)=ab+ac    a≠0 ⇒ ∃a⁻¹
         |                   |             |
     Example:            Example:       Example:
     (ℤ, +)              (ℤ, +, ×)     (ℚ, +, ×)
```

```
 Where encountered in real life
+------------+---------------------------------------------------+
| STRUCTURE  | EXAMPLES                                          |
+------------+---------------------------------------------------+
|            |                                                   |
| Semigroup  | String concatenation "abc"+"def"="abcdef"         |
|            | (actually monoid — there is empty string "")      |
|            |                                                   |
+------------+---------------------------------------------------+
|            |                                                   |
| Monoid     | (ℕ, +, 0) — naturals with zero. Cannot subtract. |
|            | Functions with composition ∘ and identity id      |
|            |                                                   |
+------------+---------------------------------------------------+
|            |                                                   |
| Group      | Symmetries (everything can be undone)             |
|            | Cryptography: elliptic curves, RSA                |
|            |                                                   |
+------------+---------------------------------------------------+
|            |                                                   |
| Ring       | Polynomials ℤ[x] — can +,−,×, but not ÷          |
|            | Matrices n×n — not every one is invertible        |
|            | Integers ℤ — 5÷2 is not an integer.               |
|            |                                                   |
+------------+---------------------------------------------------+
|            |                                                   |
| Field      | ℚ, ℝ, ℂ — everything possible: +,−,×,÷          |
|            | Finite fields 𝔽ₚ — cryptography, codes             |
|            |                                                   |
+------------+---------------------------------------------------+
```

Summary table:

STRUCTURE	ASSOC.	IDENTITY	INVERSES	COMMUTAT.	×	DIVISION
Semigroup	✓					
Monoid	✓	✓				
Group	✓	✓	✓			
Abelian group	✓	✓	✓	✓		
Ring	✓	✓	✓	(+)	✓	
Field	✓	✓	✓	✓✓	✓	✓

Why this is important

When you see a new object, ask: "What structure is this?"

- Matrices — ring (can multiply, cannot always divide)
- Functions $[0,1]\to\mathbb{R}$ — vector space (over field $\mathbb{R}$)
- Permutations — group (everything invertible)
- Polynomials — ring (or even algebra over field)

Knowing the structure — you know which theorems are applicable.

Rings and fields — arithmetic + algebra

Why does an engineer need this?

Rings and fields — this is not abstraction for the sake of abstraction. This is the foundation:

- Cryptography: RSA works in ring $\mathbb{Z}_n$ (residues from division)
- Error codes: QR codes, CD, internet — Galois fields $GF(2^8)$
- Discrete mathematics: hash functions, checksums
- Signals: Z-transform — this is a ring of formal series

Main idea: sometimes arithmetic is needed where numbers "wrap around"
(like a clock: after 12 comes 1), or where division works differently.

Group — this is one operation. But in arithmetic there are two: addition and multiplication.
How to combine them?

Ring = two connected operations

Intuition: Ring — this is "arithmetic" where you can add, subtract, multiply, but not necessarily divide.

Formally: a ring $(R, +, \cdot)$ is a set R with two operations:

- $(R, +)$ — abelian group (addition works as usual)
- $(R, \cdot)$ — monoid (multiplication exists, but inverses may not be)
- $a \cdot (b + c) = a \cdot b + a \cdot c$ (distributivity — brackets expand)

Examples:

Ring	Why ring, not field
$\mathbb{Z}$ (integers)	2 not invertible: $1/2 \notin \mathbb{Z}$
$\mathbb{Z}_n$ (residues)	If n not prime, there are zero divisors
$\mathbb{Z}[x]$ (polynomials)	x not invertible: $1/x$ — not polynomial
$M_n(\mathbb{R})$ (matrices)	Degenerate matrices not invertible

Zero divisors — strangeness of rings:
In $\mathbb{Z}_6$: $2 \cdot 3 = 6 = 0 \pmod 6$
Both factors nonzero, but product = 0.
This does not happen in ordinary numbers — sign of structure "defect".

Field = Ring where you can divide

Intuition: Field — this is "full-fledged arithmetic" with division. Everything we were taught in school about numbers — these are properties of fields.

Formally: Field — ring where each $a \neq 0$ has inverse a^{-1}.

Examples of fields:

- $\mathbb{Q}$ (rationals) — minimal field containing $\mathbb{Z}$
- $\mathbb{R}$ (reals) — completion of $\mathbb{Q}$
- $\mathbb{C}$ (complex) — algebraically closed
- $\mathbb{Z}_p = \mathbb{Z}/p\mathbb{Z}$ for prime p — finite field (important for cryptography)
  (in modern literature often written $\mathbb{F}_p$; the notation ℤ_p — with underscore — is reserved for p-adic integers)

Why is $\mathbb{Z}_p$ — field for prime p?

In $\mathbb{Z}_5$: elements $\{0, 1, 2, 3, 4\}$

Inverses: $1^{-1} = 1$, $2^{-1} = 3$ (because $2 \cdot 3 = 6 = 1 \bmod 5$)
$3^{-1} = 2$, $4^{-1} = 4$ (because $4 \cdot 4 = 16 = 1 \bmod 5$)

Each nonzero element invertible. This is field.

Why is $\mathbb{Z}_6$ — not field?

In $\mathbb{Z}_6$: $2 \cdot 3 = 0$, so 2 and 3 — zero divisors.
Zero divisor cannot be invertible (otherwise $0 = 2^{-1} \cdot 0 = 2^{-1} \cdot 2 \cdot 3 = 3 \neq 0$).

Theorem: $\mathbb{Z}_n$ — field $\Longleftrightarrow$ n prime.

Ideals — "divisibility" in abstract ring

Intuition: Ideal — this is generalization of concept "all numbers divisible by n".

In $\mathbb{Z}$: set of all numbers divisible by 3 — this is {…, -6, -3, 0, 3, 6, …}
Notation: $3\mathbb{Z}$ or (3)

Key property: if a divisible by 3, then also $a \cdot k$ divisible by 3.
"Divisibility absorbs multiplication" — this is definition of ideal.

Formally: Ideal $I \subset R$ — this is subset such that:
- I — subgroup under addition
- $a \in I,\ r \in R \Rightarrow r \cdot a \in I$ (multiplication by any element stays in I)

Examples:

RING	EXAMPLES OF IDEALS
$\mathbb{Z}$	$n\mathbb{Z} = \{nk : k \in \mathbb{Z}\}$ — all ideals such
$\mathbb{R}[x]$	$(x^2 + 1)$ = all multiples $(x^2 + 1)$
C(X) (functions)	$\{f : f(x_0) = 0\}$ — functions with zero at x_0

Why ideals?
Ideals allow "gluing" elements of a ring — as normal subgroups allow gluing elements of a group. Result — quotient ring.

Quotient Ring — "Arithmetic of Remainders"

Intuition: Quotient ring R/I is a "gluing" of elements differing by an element from I. As if everything from I became zero.

Main example:

$\mathbb{Z}/3\mathbb{Z} = \{\bar{0}, \bar{1}, \bar{2}\}$ — remainders from division by 3

Here $\bar{0} = \{..., -6, -3, 0, 3, 6, ...\}$ (all multiples of 3 "glued" into 0)
$\bar{1} = \{..., -5, -2, 1, 4, 7, ...\}$
$\bar{2} = \{..., -4, -1, 2, 5, 8, ...\}$

Arithmetic: $\bar{2} + \bar{2} = \bar{4} = \bar{1}$, $\bar{2} \cdot \bar{2} = \bar{4} = \bar{1}$

More profound example — how to construct $\mathbb{C}$:

Problem: in $\mathbb{R}$ there is no root of -1.
Solution: $\mathbb{C} = \mathbb{R}[x]/(x^2 + 1)$

Take polynomials in x with coefficients from $\mathbb{R}$.
"Glue" all multiples of $(x^2 + 1)$, that is, set $x^2 + 1 = 0$.
Then $x^2 = -1$, and x plays the role of i.

Elements: a + bx (higher powers reduce: $x^2 \to -1$)
This is precisely complex numbers a + bi.

Finite Fields – Cryptography and Codes

Theorem: A finite field exists if and only if the number of elements = p^n (power of a prime). Notation: $GF(p^n)$ or $\mathbb{F}_{p^n}$.

- GF(2) = {0, 1} – binary arithmetic (XOR = addition)
- $GF(2^8)$ = 256 elements – used in AES, QR codes
- GF(p) = $\mathbb{Z}_p$ – simplest finite fields

Example: GF(4) – field of 4 elements

Cannot simply take $\mathbb{Z}_4$ – there $2 \cdot 2 = 0$, zero divisors.

Correct construction: $GF(4) = GF(2)[x]/(x^2 + x + 1)$
Elements: {0, 1, x, x+1} with arithmetic mod 2 and mod $(x^2 + x + 1)$

Multiplication table:

·	0	1	x	x+1
0	0	0	0	0
1	0	1	x	x+1
x	0	x	x+1	1
x+1	0	x+1	1	x

← $x^2 = x+1$ (from $x^2+x+1=0$)

Application: Reed-Solomon codes (CD, DVD, QR) work over $GF(2^8)$.
This allows correcting errors mathematically precisely.

Hierarchy: From Set to Field

```
Set (just elements)
     |
     ▼ + one operation
Semigroup (associativity)
     |
```

```
            ▼ + neutral element
        Monoid
            |
            ▼ + inverse elements
        Group
            |
            ▼ + commutativity
     Abelian Group
            |
            ▼ + second operation (multiplication) + distributivity
        Ring (ℤ, ℤ[x], matrices)
            |
            ▼ + no zero divisors
   Integral Domain (ℤ)
            |
            ▼ + every nonzero invertible
         Field (ℚ, ℝ, ℂ, ℤₚ)
```

At each step a property is added → structure becomes "better".
Field — the most "nice" arithmetic structure.

Number Theory — Arithmetic as Structure

Number Theory as a View on Space

Numbers are not just objects for computation. Numbers form spaces with rich structure.

SPACE	STRUCTURE	WHAT WE STUDY
$\mathbb{Z}$	Ring (+ and ×) + order	Divisibility, prime numbers
$\mathbb{Z}_n = \mathbb{Z}/n\mathbb{Z}$ (remainders mod n)	Finite ring When n=p — field.	Modular arithmetic Cryptography
$\mathbb{Q}_p$ (p-adic)	Field with ultrametric $\lvert x+y\rvert_p \le \max(\lvert x\rvert_p,\lvert y\rvert_p)$	Local analysis Diophantine equations
$\mathbb{Z}[i]$ (Gaussian)	Ring in $\mathbb{C}$ Euclidean	Sums of two squares

Key idea: The same number can be considered in different spaces

Number 7:
- In $\mathbb{Z}$: prime, irreducible
- In $\mathbb{Z}_7$: zero ($7 \equiv 0 \bmod 7$)
- In $\mathbb{Z}[i]$: still prime ($7 \equiv 3 \bmod 4$, not a sum of two squares)
- In $\mathbb{Z}[\sqrt{-5}]$: remains prime

Number 6 in $\mathbb{Z}[\sqrt{-5}]$: two different factorizations.

$= 2 \cdot 3 = (1+\sqrt{-5})(1-\sqrt{-5})$

This shows that in $\mathbb{Z}[\sqrt{-5}]$ there is no uniqueness of factorization.

Number theory studies how arithmetic properties depend on the algebraic structure of the space.

Number theory studies properties of integers. This is one of the most ancient areas of mathematics, but it is connected with the most modern ones: cryptography, algebraic geometry, representation theory.

Divisibility – basic concepts

$a \mid b$ means "a divides b" $\iff \exists k \in \mathbb{Z}: b = a \cdot k$

CONCEPT	DEFINITION
GCD(a,b) = gcd	Greatest common divisor $\max\{d : d \mid a \text{ and } d \mid b\}$
LCM(a,b) = lcm	Least common multiple $\min\{m > 0 : a \mid m \text{ and } b \mid m\}$
Coprime	$\gcd(a,b) = 1$
Prime number p	$p > 1$, divisors only 1 and p

Key relation: $\gcd(a,b) \cdot \mathrm{lcm}(a,b) = a \cdot b$

Fundamental theorem of arithmetic

Each natural number $n > 1$ uniquely (up to order) factors into a product of primes:

$$n = p_1^{a_1} \cdot p_2^{a_2} \cdot \ldots \cdot p_k^{a_k}$$

Examples:
$60 = 2^2 \cdot 3 \cdot 5$
$100 = 2^2 \cdot 5^2$
$2024 = 2^3 \cdot 11 \cdot 23$

Corollaries:

OPERATION	THROUGH FACTORIZATION
$\gcd(a,b)$	Product of $p^{\min(a_p, b_p)}$ over all p
$\text{lcm}(a,b)$	Product of $p^{\max(a_p, b_p)}$ over all p
$a \mid b$	$a_p \le b_p$ for all p
Number of divisors	$(a_1+1)(a_2+1)\cdots(a_k+1)$

Analogy with vectors:
  Number n ↔ vector $(a_1, a_2, a_3, \ldots)$ — exponents of primes
  Multiplication ↔ vector addition
  gcd ↔ componentwise min
  lcm ↔ componentwise max

Congruences modulo — arithmetic of remainders

$a \equiv b \pmod{n}$ means $n \mid (a - b)$ $\iff$ a and b give the same remainder

PROPERTIES (congruences can be added, multiplied, raised to powers)
$a \equiv b,\ c \equiv d \Rightarrow a + c \equiv b + d \pmod{n}$
$a \equiv b,\ c \equiv d \Rightarrow a \cdot c \equiv b \cdot d \pmod{n}$
$a \equiv b \Rightarrow a^k \equiv b^k \pmod{n}$

Ring $\mathbb{Z}_n = \{0, 1, 2, \ldots, n-1\}$ with operations mod n

n	structure of $\mathbb{Z}_n$
$n = p$ (prime)	Field. Every nonzero element is invertible. Example: $\mathbb{Z}_5$, where $2\cdot3 = 6 \equiv 1$, so $2^{-1} = 3$
$n = p^k$	Local ring (unique maximal ideal)
$n = p\cdot q$ (composite)	Has zero divisors. In $\mathbb{Z}_6$: $2\cdot3 = 0$ But by CRT: $\mathbb{Z}_n \cong \mathbb{Z}_p \times \mathbb{Z}_q$ if $\gcd(p,q)=1$

CRT (Chinese Remainder Theorem):
  If $\gcd(m,n) = 1$, then $\mathbb{Z}_{mn} \cong \mathbb{Z}_m \times \mathbb{Z}_n$
  Practically: system $x \equiv a \pmod{m}$, $x \equiv b \pmod{n}$ has unique solution

Fermat's little theorem and Euler's function

$\varphi(n)$ = Euler's function = count of numbers from 1 to n coprime with n

n	$\varphi(n)$	FORMULA
p (prime)	$p - 1$	All except 0 coprime with p
p^k	$p^k - p^{k-1}$	$= p^k(1 - 1/p)$
$m \cdot n$	$\varphi(m) \cdot \varphi(n)$	if $\gcd(m,n) = 1$ (multiplicative)
general	$n\prod(1 - 1/p)$	product over all primes $p \mid n$

Theorems:

Fermat's little theorem (p prime)	$a^p \equiv a \pmod p$ for any a If $\gcd(a,p)=1$: $a^{p-1} \equiv 1 \pmod p$
Euler's theorem (generalization)	$a^{\varphi(n)} \equiv 1 \pmod n$ if $\gcd(a,n)=1$ This generalizes Fermat to composite n

Application – RSA cryptography:

Choose large primes p, q. Let $n = p \cdot q$, $\varphi(n) = (p-1)(q-1)$.
Choose e coprime with $\varphi(n)$, find d: $e \cdot d \equiv 1 \pmod{\varphi(n)}$.
Encrypt: $c = m^e \bmod n$. Decrypt: $m = c^d \bmod n$.
Works by Euler's theorem: $m^{ed} = m^{1 + k\varphi(n)} \equiv m \pmod n$.

P-adic numbers – alternative completion of $\mathbb{Q}$

$\mathbb{R}$ is the completion of $\mathbb{Q}$ with respect to the usual metric $|x - y|$.
But there are other metrics on $\mathbb{Q}$.

p-adic norm:

$|x|_p = p^{-v_p(x)}$, where $v_p(x)$ = degree of p in the factorization of x

Examples (p = 5):

$|25|_5 = 5^{-2} = 1/25$ (25 = 5^2, many fives → small norm)
$|1/5|_5 = 5^1 = 5$ (few fives in numerator)
$|7|_5 = 5^0 = 1$ (no fives at all)
$|0|_5 = 0$

PROPERTY	USUAL NORM $|\cdot|$	p-ADIC $|\cdot|_p$
Large numbers	Far from 0	Can be close to 0!
Triangle	$|x+y| \le |x|+|y|$	$|x+y|_p \le \max(|x|_p,|y|_p)$ (Ultrametric — stronger)
Completion	$\mathbb{R}$	$\mathbb{Q}_p$ (p-adic numbers)
Alg. closure	$\mathbb{C}$ (dim 2 over $\mathbb{R}$)	$\mathbb{C}_p$ (infinite-dimensional)

Why is this needed:

- Local-global principle: an equation has a solution in $\mathbb{Q}$ $\Longleftrightarrow$ has a solution in $\mathbb{R}$ and in all $\mathbb{Q}_p$ (with caveats)
- Modern algebraic geometry works over all these fields
- Number theory: many problems are easier to solve "locally" in $\mathbb{Q}_p$

Connection with other areas

AREA	CONNECTION WITH NUMBER THEORY
Groups	$(\mathbb{Z}/n\mathbb{Z})^*$ — group of invertible elements Order = $\varphi(n)$, Lagrange's theorem → Euler
Rings	$\mathbb{Z}$ — principal ideal domain Ideal $(n) = n\mathbb{Z}$, quotient ring = $\mathbb{Z}_n$
Fields	$\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$ — finite field, extensions → codes $\mathbb{Q}_p$ — local field for arithmetic geometry
Topology	$\mathbb{Z}$_p = lim $\mathbb{Z}/p^n\mathbb{Z}$ — projective limit (p-adic integers; not to be confused with $\mathbb{F}_p$) Topology on $\mathbb{Z}$_p: basis = classes mod p^n
Complex analysis	$\zeta(s) = \Sigma\, n^{-s}$ — Riemann zeta function Connects prime numbers and complex analysis

Galois theory — why there is no formula for roots of degree 5

Quadratic equation is solved by a formula (known for ~2000 years).
Cubic and quartic — also (Cardano, Ferrari, 16th century).
For fifth degree Abel (1824) proved: there is no general formula.
Galois (1832) explained why — and created group theory.

Key idea:
To each polynomial p(x) corresponds a group Gal(p) — the group of permutations of roots preserving all algebraic relations.

Galois theorem:

A polynomial is solvable by radicals (roots are expressible via +, −, ×, ÷, √) if and only if its Galois group is solvable.

Why this works:

Root extraction $\sqrt[n]{}$ adds a "layer" of symmetry — a cyclic group $\mathbb{Z}/n\mathbb{Z}$.
Solvable group = can be decomposed into a "tower" of cyclic subgroups.
Symmetric group S_5 is not solvable (contains simple group A_5).
Therefore general polynomial of degree 5 is not solvable by radicals.

Galois correspondence:

SUBGROUPS Gal(p)	INTERMEDIATE FIELDS
Gal(p)	$\mathbb{Q}$ (base field)
{e} (trivial)	Splitting field
Normal subgroup $H \lhd G$	Normal extension

This bijective "dictionary" between groups and fields is one of the deepest ideas in mathematics: the problem of equations is solved through symmetries.

Table of groups — systematics of symmetries

Group = symmetries of an object
Problem: classify groups and their connections

Hierarchy of groups

```
                      All groups
                     /          \
             Finite              Infinite
           /        \          /           \
      Abelian  Non-abelian  Discrete    Continuous
         |          |          |         (Lie groups)
      ℤ/n, ...  Sₙ, Dₙ, ...  ℤ, Fₙ, ...        |
                                   Compact / Non-compact
                                       |            |
                                  SO(n), SU(n)   ℝⁿ, GL(n)
```

Finite groups — complete classification exists

GROUP	ORDER	WHAT IT IS GEOMETRICALLY / CONNECTIONS
{e}	1	Trivial: "do nothing"
$\mathbb{Z}/n$	n	Cyclic: rotations of n-gon $\cong$ roots of unity: $\{1, \omega, \omega^2, ...\}$, $\omega = e^{(2\pi i/n)}$ Abelian. $\mathbb{Z}/p$ (p prime) — simplest "atoms"
$\mathbb{Z}/2 \times \mathbb{Z}/2$ (Klein four-group)	4	Klein group: symmetries of rectangle not cyclic. (no element of order 4)
D_n	2n	Dihedral: rotations + reflections of n-gon $D_3 \cong S_3$ (unique case) Non-abelian for $n \geq 3$
S_n	n!	Symmetric: all permutations of n elements Any finite group $\subset S_n$ (Cayley's theorem)
A_n	n!/2	Alternating: even permutations A_5 — simplest non-abelian simple group $A_5 \cong$ symmetries of icosahedron
Q_8	8	Quaternion: $\{\pm 1, \pm i, \pm j, \pm k\}$ Non-abelian, but all subgroups are normal Connection with rotations in 3D (see quaternions)

Infinite discrete:

GROUP	WHAT IT IS / CONNECTIONS
$(\mathbb{Z}, +)$	Integers: shifts by integer along line $\pi_1(S^1) = \mathbb{Z}$ – fundamental group of circle. Unique infinite cyclic group
$\mathbb{Z}^n$	Lattice: vertices of integer lattice in $\mathbb{R}^n$ $\pi_1(T^n) = \mathbb{Z}^n$ – fundamental group of n-torus
F_n	Free on n generators: all words from n letters $\pi_1(\vee^n S^1) = F_n$ – wedge of n circles "Universal": any group is its quotient

Lie groups (continuous) — symmetries of physics:

GROUP	dim	WHAT IT IS / WHERE IT APPEARS
$U(1) \cong S^1$	1	Complex numbers $\lvert z\rvert=1$, rotations of plane Phase in quantum mechanics, electromagnetism
SO(2)	1	Rotations of plane ($\cong$ U(1) as Lie groups)
SO(3)	3	Rotations in $\mathbb{R}^3$: orientation of rigid body not simply connected $\pi_1(SO(3)) = \mathbb{Z}/2$
SU(2)	3	Unitary 2×2 with det=1, double cover of SO(3) Spin in quantum mechanics. $SU(2) \cong S^3$ (3-sphere) — simply connected Important for robotics: Quaternion q and −q give the same rotation. Rotation by 360° gives −1, need 720° to return to 1. This is not a bug, it's topology: $\pi_1(SO(3)) = \mathbb{Z}/2$.
SU(3)	8	Symmetry of strong interaction (quarks) Standard model: SU(3)×SU(2)×U(1)
SO(3,1)	6	Lorentz group: symmetries of spacetime Special theory of relativity
$GL(n,\mathbb{R})$	n^2	All invertible n×n matrices (det ≠ 0) "General linear group"
$SL(n,\mathbb{R})$	n^2-1	Matrices with det = 1 (preserve volume)

Key connections

1. Fundamental groups (topology → algebra):
   $\pi_1(S^1) = \mathbb{Z}$, $\pi_1(T^2) = \mathbb{Z}^2$, $\pi_1(\vee_n S^1) = F_n$

2. Coverings (connection between groups):
   $SU(2) \xrightarrow{2:1} SO(3)$, $\mathbb{R} \to S^1$ (exp), $SL(2,\mathbb{C}) \xrightarrow{2:1} SO(3,1)$

3. Classification of finite simple groups (completed ~1980):
   Cyclic $\mathbb{Z}/p$ + Alternating A_n ($n \geq 5$) +
   Groups of Lie type + 26 sporadic (including "Monster")

4. Classification theorem for finite abelian groups:
   Any $\cong \mathbb{Z}/n_1 \times \mathbb{Z}/n_2 \times \ldots \times \mathbb{Z}/n_k$ (unique decomposition)

Groups describe symmetries — what can be done with space. But to talk about continuity of transformations, we need the notion of closeness. What does "points nearby" mean? What does "transformation doesn't tear space" mean?

This is what topology deals with — the next fundamental view of space.

Topology — the study of nearness without distances

If a set is dust, then topology is fabric. We add to points the notion of "nearby": which points can be considered close, which — not. But we still don't know how close — there are no distance-numbers.

This is the minimal structure for continuity. To say "a function is continuous", it suffices to know which points are nearby. Specific distances are not needed.

In terms of "object—observer": topology is a property of the space itself, not of the observer. Two observers with different coordinate systems will see one and the same topology: the same holes, the same connectedness, the same boundaries. Topological properties are invariants, not depending on the method of description.

This is precisely why topology is so fundamental: it speaks about the shape of an object, not about how the observer records it.

Topology as a view of space

Let us recall from the introduction: each branch of mathematics is a way of looking at space. Topology sees in space only the nearness of points, but not distances between them.

```
  Set → topological space → metric
  (points)         (nearness)              (distances)
```

Topology is the "middle": more structure than a bare set, but less than a metric space.

Main example: donut = mug

```
      Donut (torus)     mug

         ╭--------╮                    ╭-╮
       ╭-╯  ╭--╮  ╰-╮               ╭--╯ ╰--╮
       │    │  │    │               │ ╭-╮   │
       │    │  │    │               │ │ │   │
       ╰-╮  ╰--╯  ╭-╯               │ ╰-╯   │
         ╰--------╯                 │       │
                                    ╰-------╯

  One hole                         One hole (handle)
```

Topologically identical. One can continuously deform one into the other.
But the donut cannot be transformed into a mug without a handle – the hole will disappear.

Why topology is needed

Problem: We want to talk about "nearness" and "continuity", but:

- Distance does not always exist (how to measure distance between functions?)
- Sometimes distance is redundant (shape matters to us, not sizes)
- Different distances can give the same "nearness"

Solution: Define "nearness" directly, without distance.
Instead of $d(x,y) < \varepsilon$ we say: "y in a neighborhood of x"

Analogy: Metro map vs city map
- City map: exact distances, scale
- Metro map: only connections between stations

For navigating the metro, distances are not needed

What topology studies

Geometry: angles, lengths, distances – metric invariants
Topology: connectedness, holes, number of components – topological invariants

> Topology studies properties invariant under continuous deformations (homeomorphisms). Allowed: stretching, compression, bending. Not allowed: tearing, gluing.

Homeomorphic objects:

OBJECTS	TOPOLOGICAL INVARIANT
○ ≅ □ ≅ △	No holes, $\pi_1 = 0$
Mug ≅ Torus	One hole, $\pi_1 = \mathbb{Z}$
O ≅ D, B ≅ 8, A ≅ R	Classification by number of holes

Physical view: frozen object

Topology looks at space as a frozen object.
Group looks as a dynamical system.

Four questions of topology:

1. How many holes are in the object?
   Donut — one. Mug with handle — also one. They are "the same".
   Sphere — none. It cannot be deformed into a donut.

2. Does the object have a boundary?
   A disk has one (circle). A sphere does not. A Möbius strip — one edge.

3. Can it be cut without breaking into parts?
   Cut a torus crosswise — a tube remains (one part).
   Cut a sphere — it breaks into two caps (two parts).

4. Can one distinguish "left" and "right" on it?
   On a Möbius strip — no. It is non-orientable.

These are all invariants — properties that do not change under deformation.

Contrast with groups:

Group asks: What motions are possible with this object?
Topology asks: What is the shape of this object?

Group: dynamics of motion → SE(3), SO(3), D_4
Topology: statics of shape → holes (π_1), connectedness (π_0), Euler (χ)

Both give classification — establish what is possible and what is not.
Both answer the question "what type is this object".

Example of connection: Fundamental group $\pi_1(X)$ is a group
that classifies loops in topological space X.
Topology and algebra meet.

Alphabet Letters: Classification by Holes

No holes (all the same):	C, G, I, J, L, M, N, S, U, V, W, Z
One hole (all the same):	A, D, O, P, Q, R
Two holes (all the same):	B, 8

This is a complete classification of letters from a topological point of view.
The letter "O" and the letter "D" – the same object (can be deformed).
The letter "O" and the letter "C" – different objects (C has no hole).

Open Sets – Basic Concept

Definition of open set:
A set $U \subset X$ is called open if for each point $x \in U$
there exists a neighborhood of x entirely contained in U.

Equivalent formulation (for metric space):
$\forall x \in U \ \ \exists \varepsilon > 0: \ \ B(x, \varepsilon) \subset U$

$$\boxed{x \in U \iff x \text{ – interior point} \iff \exists \varepsilon > 0: B(x,\varepsilon) \subset U}$$

Boundary points: $\forall \varepsilon > 0$ ball $B(x,\varepsilon)$ contains points both from U and from $X \setminus U$
An open set does not contain its boundary points.

Examples on the number line ℝ

INTERVAL TYPE	VISUALIZATION	PROPERTY
open (0, 1)	--○━━━━━━━━━━━━━━○-- 0 not in 1 not in	Open: each point has "margin" inside
closed [0, 1]	--●━━━━━━━━━━━━━━●-- 0 in 1 in	Closed: contains boundary No "margin" at edges
half-open [0, 1)	--●━━━━━━━━━━━━━━○-- 0 in 1 not in	Neither open nor closed This happens

Openness depends on ambient space

The same set can be open in one space
and not open in another.

Example 1: Set $[0, 1)$
- In $\mathbb{R}$: not open (point 0 is boundary)
- In $[0, \infty)$: open (no points to the left of 0, so 0 is interior)

Example 2: Set $[0, \frac{1}{2})$
In space $X = [0, 1]$ with induced topology from $\mathbb{R}$:
- $[0, \frac{1}{2})$ is open in X (because $[0, \frac{1}{2}) = X \cap (-1, \frac{1}{2})$)
- But point 0 intuitively seems like an "edge" — this is a trap

Conclusion: When saying "open", always specify — in which space

Definition of Metric

A metric on a set X is a function $d: X \times X \to [0, +\infty)$ such that:

(M1)	$d(x,y) = 0 \iff x = y$	(identity of indiscernibles)
(M2)	$d(x,y) = d(y,x)$	(symmetry)
(M3)	$d(x,z) \leq d(x,y) + d(y,z)$	(triangle inequality)

The pair (X, d) is called a metric space.

Examples of metrics:

SPACE	METRIC
$\mathbb{R}$	$d(x,y) = \lvert x - y\rvert$
$\mathbb{R}^n$ (Euclidean)	$d(x,y) = \sqrt{\Sigma_i (x_i-y_i)^2}$
$\mathbb{R}^n$ (Manhattan)	$d(x,y) = \Sigma_i \lvert x_i-y_i\rvert$
$\mathbb{R}^n$ (sup-metric)	$d(x,y) = \max_i \lvert x_i-y_i\rvert$
C[a,b] (functions)	$d(f,g) = \max_{x\in[a,b]} \lvert f(x)-g(x)\rvert$
Discrete	$d(x,y) = 0$ if $x=y$, otherwise 1

Why metric:
- Defines the notion of "closeness" quantitatively
- Generates topology (open sets via balls)
- Allows talking about convergence: $x_n \to x \iff d(x_n,x) \to 0$

Formal definition of open set (via metric)

Let (X, d) — metric space.

Open ball: $B(x, \varepsilon) = \{y : d(x,y) < \varepsilon\}$
(all points closer than ε to center x)

A set U is called open if:

```
+---------------------------------------------------------+
| ∀x ∈ U  ∃ε > 0:  B(x, ε) ⊆ U                        |
|                                                         |
| "For each point x from U there exists a ball centered at x, |
| entirely lying in U"                                    |
+---------------------------------------------------------+
```

Visually:

```
       U — open                          V — not open
    ╭─────────────────╮               ╭─────────────────╮
    |    ╭───╮        |               |            ●----+ ← point on
    |    | ● | ball   |               |           no    |   boundary,
    |    ╰───╯ inside |               |           ball  |   ball won't
    |                 |               |                 |   fit
    ╰─────────────────╯               ╰─────────────────╯
```

Closed sets

Closed set = complement of open

F closed ⟺ X \ F open

Equivalently: F contains all its limit points.
(If a sequence from F converges, the limit is also in F)

Examples in ℝ:

[0, 1]	closed	(ℝ \ [0,1] = (−∞,0) ∪ (1,+∞) — open)
{0}	closed	(single point — degenerate case)
[0, +∞)	closed	(half-line with included endpoint)
ℚ	not closed	(√2 — limit point, but √2 ∉ ℚ)

Attention: "Not open" ≠ "Closed".

[0, 1) — neither open nor closed
∅ — both open and closed (the only such: ∅ and entire X)
ℝ — both open and closed

Boundary, interior, closure

For any set $A \subseteq X$:

$\mathrm{Int}(A)$ = interior = largest open subset of A
$= \{x \in A : \exists\varepsilon > 0, B(x,\varepsilon) \subseteq A\}$

$\mathrm{Cl}(A)$ = closure = smallest closed superset of A
$= A \cup \{\text{all limit points of } A\}$

∂A = boundary = $\mathrm{Cl}(A) \setminus \mathrm{Int}(A)$
= points in any neighborhood of which there are both A and non-A

Example: $A = (0, 1]$

$\mathrm{Int}(A) = (0, 1)$ – removed point 1 (it's on boundary)
$\mathrm{Cl}(A) = [0, 1]$ – added point 0 (limit point)
$\partial A = \{0, 1\}$ – two boundary points

Connection:
A open $\iff$ $A = \mathrm{Int}(A)$ $\iff$ A does not contain its boundary
A closed $\iff$ $A = \mathrm{Cl}(A)$ $\iff$ A contains its boundary

Topological space – abstraction

Idea: forget about metric, keep only "open sets"

Observation: All properties of continuity can be expressed through open sets, without mentioning distance.

Idea: What if we directly specify which sets to consider "open"?
Not derive from metric, but simply declare.

Need rules so that "open" sets behave reasonably.

Definition of topology

A topology on set X is a family $\tau \subseteq P(X)$ of subsets of X, satisfying three axioms:

```
+---------------------------------------------------------+
| (T1)  ∅ ∈ τ  and  X ∈ τ                                 |
```

```
  | Empty and entire space are open                      |
  |                                                        |
  | (T2)  U₁, U₂ ∈ τ  ⇒  U₁ ∩ U₂ ∈ τ                       |
  | Intersection of two open sets is open                |
  | (by induction: finite intersection of open sets is open) |
  |                                                        |
  | (T3)  {Uᵢ}ᵢ∈I ⊆ τ  ⇒  ⋃ᵢ∈I Uᵢ ∈ τ                      |
  | Union of any family of open sets is open             |
  | (even infinite, even uncountable)                      |
  +--------------------------------------------------------+
```

The pair (X, τ) is called a topological space.
Elements of τ are called open sets.

Why such axioms

These axioms are derived from properties of open sets in metric space:

(T1) Obvious: everywhere there is "margin", nowhere there is "margin"

(T2) If in U_1 there is a ball of radius ε_1, and in U_2 — of radius ε_2, then in intersection there is a ball of radius $\min(\varepsilon_1, \varepsilon_2)$

Why only finite? For infinite intersection min can be 0.

Example: $\bigcap_n (-1/n, 1/n) = \{0\}$ — single point, not an open set.

(T3) If a point is in some U_i, it has a ball in this U_i, so in the union there is also a ball. Works for any family.

Examples of Topologies

Let $X = \{a, b, c\}$ — three points. Compare topologies:

TOPOLOGY	OPEN SETS	MEANING

Discrete (maximal)	All subsets: $\varnothing,\{a\},\{b\},\{c\},\{a,b\},$ $\{a,c\},\{b,c\},\{a,b,c\}$	Points "far" from each other, each isolated
Antidiscrete (minimal)	Only $\varnothing$ and X: $\{\varnothing, \{a,b,c\}\}$	Points "glued", indistinguishable
Intermediate (example)	$\{\varnothing, \{a\}, \{a,b\}, \{a,b,c\}\}$	a "open", c "closed"

Standard topology on ℝ

$\tau = \{U \subseteq \mathbb{R} : \forall x \in U\ \exists \varepsilon > 0: (x-\varepsilon, x+\varepsilon) \subseteq U\}$

Open sets = those where each point is surrounded by an interval.

Topology base:

No need to describe all open sets.
Sufficient to specify a base – a collection of "building blocks" from which all other open sets are obtained by unions.

For $\mathbb{R}$: base = all open intervals (a, b)
For $\mathbb{R}^n$: base = all open balls $B(x, \varepsilon)$

Important: Standard topology on $\mathbb{R}$ is generated by metric $d(x,y) = |x-y|$
But one can define other topologies on the same $\mathbb{R}$.

Metric → Topology (but not conversely)

Any metric d on X generates a topology:
$\tau_d = \{U : \forall x \in U\ \exists \varepsilon > 0: B_d(x,\varepsilon) \subseteq U\}$

But not every topology is generated by some metric.
(Such topologies are called "metrizable")

Example of non-metrizable topology:
Antidiscrete on X with $|X| > 1$: cannot separate points by balls.

Different metrics can give the same topology:

On $\mathbb{R}^2$: $d_1(x,y) = |x_1-y_1| + |x_2-y_2|$ (Manhattan)
$d_2(x,y) = \sqrt{(x_1-y_1)^2 + (x_2-y_2)^2}$ (Euclidean)

$$d_\infty(x,y) = \max(|x_1-y_1|, |x_2-y_2|) \quad \text{(Chebyshev)}$$

Balls of different shapes: ◇ (d_1)   ○ (d_2)   □ (d_∞)
But the topology is one. The same sets are open.

```
     ◇           ○           □
    / \         / \         +-+
   /   \       |   |        | |
   \   /       |   |        | |
    \ /         \ /         +-+
```

Topologically equivalent (homeomorphic as balls)

Separation axioms — how "good" the space is

Problem: Not all topological spaces are "good".
In antidiscrete topology two different points are indistinguishable.

Separation axioms define a hierarchy of "good" spaces:

```
+---------+---------------------------------------------------------------+
```

```
| AXIOM   | CONDITION                                                     |
+---------+---------------------------------------------------------------+
|         |                                                               |
| T₀      | For any x ≠ y there exists a neighborhood of one of the       |
|(Kolmog.)| points not containing the other.                              |
|         | (Can somehow distinguish)                                     |
|         |                                                               |
+---------+---------------------------------------------------------------+
|         |                                                               |
| T₁      | For any x ≠ y there exists a neighborhood of x not            |
|(Fréchet)| containing y, and a neighborhood of y not containing x.       |
|         | Equivalently: all singleton sets {x} are closed.              |
|         |                                                               |
+---------+---------------------------------------------------------------+
|         |                                                               |
| T₂      | For any x ≠ y there exist disjoint neighborhoods:             |
|(Hausd.) | U ∋ x, V ∋ y, U ∩ V = ∅                                       |
|         |                                                               |
|         |     x           y                                             |
|         |    (U)         (V)        ← U and V do not intersect          |
|         |   ╭---╮       ╭---╮                                           |
|         |   | ● |       | ● |                                           |
|         |   ╰---╯       ╰---╯                                           |
|         |                                                               |
+---------+---------------------------------------------------------------+
|         |                                                               |
| T₃      | T₁ + for a point x and closed F ∌ x there exist               |
|(regul.) | disjoint neighborhoods of x and F.                            |
|         |                                                               |
+---------+---------------------------------------------------------------+
|         |                                                               |
| T₄      | T₁ + for any disjoint closed F, G                             |
|(norm.)  | there exist disjoint neighborhoods.                           |
|         |                                                               |
+---------+---------------------------------------------------------------+
```

Hierarchy:

T_4 (normal) $\subset$ T_3 (regular) $\subset$ T_2 (Hausdorff) $\subset$ T_1 $\subset$ T_0

Why T_2 (Hausdorffness) is important:

1. Uniqueness of limits:
   In a Hausdorff space a sequence has
   at most one limit. (Without T_2 the limit may be non-unique)

2. Compact sets are closed:
   In a Hausdorff space a compact subset is closed.

3. Practice:
   Almost all spaces in analysis and geometry are Hausdorff.
   $\mathbb{R}^n$, manifolds, metric spaces — all T_2.

Examples:

- ✓ $\mathbb{R}^n$ with standard topology – T_4 (normal)
- ✓ Any metric space – T_4
- ✗ Antidiscrete topology ($|X| > 1$) – not even T_0
- ✗ Line with doubled point – T_1, but not T_2

Urysohn's Lemma (consequence of T_4):
In a normal space for any disjoint closed
F and G there exists a continuous $f: X \to [0,1]$ with $f|_F = 0$ and $f|_G = 1$.
(One can "smoothly" separate sets by a function)

Continuous mappings – preservation of proximity

Definition of continuity

Let (X, τ_X) and (Y, τ_Y) be topological spaces.

```
+----------------------------------------------------------+
| Definition: f: X → Y continuous ⟺ ∀V ∈ τ_Y: f⁻¹(V) ∈ τ_X |
| (preimage of an open set is open)                        |
+----------------------------------------------------------+
```

Here $f^{-1}(V) = \{x \in X : f(x) \in V\}$ – preimage (not inverse function).

Preimage is always defined, even if f is not invertible.

Visualization:

```
    X                            Y
  +-----------+    f       +-----------+
  | ╭------╮  |   ---→     |   ╭---╮   |
  | |f⁻¹(V)|  |            |   | V |   |   V open in Y
  | ╰------╯  |            |   ╰---╯   |
  +-----------+            +-----------+
    open?                    open ✓
```

f continuous ⟺ for any open V, its preimage $f^{-1}(V)$ is open

Connection with ε-δ definition

Two definitions – one idea (for metric spaces are equivalent):

DEFINITION	FORMULATION

ε-δ (analysis)	$\forall\varepsilon>0\ \exists\delta>0: d(x,x_0)<\delta \Rightarrow d(f(x),f(x_0))<\varepsilon$ "image arbitrarily close for close argument"
Topological	Preimage of an open set is open "neighborhood of image ← neighborhood of argument"

Why topological? Works without metric.

Examples and counterexamples

FUNCTION	CONTINUOUS?	WHY
$f(x) = x^2$	YES	Preimage of (a,b) open
$f(x) = \sin(x)$	YES	Smooth → continuous
$f(x) = \lvert x\rvert$	YES	At each point cont.
$f(x) = \lfloor x\rfloor$	no	Preimage not open
$f(x) = \theta(x)$ (Heavi)	NO	Discontinuity at x=0

Problem at point 0: jump.

Homeomorphism — topological equivalence

Definition:
Homeomorphism $f: X \to Y$ is a bijection, continuous in both directions:

1. f is a bijection (one-to-one correspondence)
2. f is continuous
3. f^{-1} is also continuous

If there exists a homeomorphism $X \to Y$, we write $X \cong Y$ ("X homeomorphic to Y")

Meaning: X and Y have the same topological structure.
They differ only in "names" of points.

Physical interpretation: very viscous flow

Imagine an object made of very viscous liquid (dough, plasticine, resin).
Homeomorphism is a slow deformation, in which:

- Material flows, changes shape
- But doesn't tear (can't create a hole)
- and doesn't glue together (can't fill a hole)

Donut → mug: dough "flows over", hole is preserved
Donut → sphere: impossible without tearing (hole must disappear)

This intuition connects topology with hydrodynamics:

- Homeomorphism = result of infinitely slow incompressible flow
- Topological invariants = that which is preserved under any flow
- Diffeomorphism = smooth flow (without "folds")

Why is continuity of f^{-1} needed?

A continuous bijection is not necessarily a homeomorphism.

Counterexample:

$X = [0, 2\pi)$ with usual topology (half-interval)
$Y = S^1$ circle
$f(t) = (\cos t, \sin t)$

```
[0------------2π)        f          ┌------┐
 ●                      ---→       /   ●    \
 |                                |    ↑     |
 +-→ stretches onto circle        |  start   |
                                   \        /
                                    └------┘
```

f is a bijection ✓
f is continuous ✓
f^{-1} is not continuous ✗

Why? When going around the circle, approaching the starting point, f^{-1} makes a jump: $. \to 2\pi-\varepsilon \to 0$ (discontinuity)

$[0, 2\pi)$ and S^1 are not homeomorphic, although there is a continuous bijection.

Examples of homeomorphisms

$(0, 1) \cong \mathbb{R}$	via $f(x) = \tan(\pi(x - 1/2))$ Open interval ↔ entire line
$(0, 1) \cong (0, \infty)$	via $f(x) = x/(1-x)$
Circle ≅ Square	"Inflate" square to circle
$\mathbb{R}^2 \setminus \{0\} \cong S^1\times\mathbb{R}$	Plane without origin ↔ Cylinder (polar coordinates: $(r,\theta) \mapsto (\theta, \ln r)$)

Not homeomorphic:

$[0,1] \ncong (0,1)$	Closed interval ≠ open (different number of endpoints)
$S^1 \ncong [0,1]$	Circle ≠ segment (segment has endpoints)
$S^2 \ncong T^2$	Sphere ≠ torus (different number of holes)

Hierarchy of equivalences

Isometry ⊂ Diffeomorphism ⊂ Homeomorphism ⊂ Homotopy equivalence

EQUIVALENCE	WHAT IT PRESERVES	SQUARE
Isometry (most strict)	Distances	≅ only to itself/rotation
Diffeomorphism	Smooth structure	≅ to circle (smooth deform.)
Homeomorphism	Topology	≅ to circle (continuous deform.)
Homotopy equiv. (most weak)	"Shape of holes"	≃ to point (can contract)

Connectedness — "in one piece"

Connectedness — informal description

Connected space: not representable as union of two disjoint
nonempty open sets

```
  Connected: not connected:
  ┌-----------┐              ┌-------┐   ┌-------┐
  |           |              |       |   |       |
  |     ●     |              |   ●   |   |   ●   |
  |           |              |       |   |       |
  └-----------┘              └-------┘   └-------┘
  one piece                 two separate pieces
```

Formal definition

Topological space X is called connected if
there does not exist a partition X = U ∪ V, where:

- $U, V \neq \varnothing$ (both nonempty)
- $U \cap V = \varnothing$ (disjoint)
- $U, V \in \tau$ (both open)

Equivalently: the only sets that are simultaneously open and closed are $\varnothing$ and X.

Meaning: Cannot "cut" X into two open pieces.
Between any two points there is a "topological path".

Examples

Connected:

$[0, 1]$	Any partition into open sets is impossible
$(0, 1)$	Also connected
$\mathbb{R}$	Connected
$\mathbb{R}^n$	Connected for any n
S^1, S^2, S^n	All spheres are connected
Disk, ball	Connected

Not connected:

$(0,1) \cup (2,3)$	Two intervals = two pieces $U = (0,1)$, $V = (2,3)$ – partition
$\mathbb{Q}$	Rational numbers are not connected Partition: $\{q < \sqrt{2}\}$ and $\{q > \sqrt{2}\}$ (both open in induced topology)
$\mathbb{R} \setminus \{0\}$	Line without zero = two rays $U = (-\infty, 0)$, $V = (0, +\infty)$
$\{0, 1\}$ (discrete)	Two isolated points (discrete topology) $U = \{0\}$, $V = \{1\}$ – both open

Path connectedness (stronger)

X is path connected if any two points can be connected by a path:

$$\forall x, y \in X \quad \exists \gamma: [0,1] \to X \text{ continuous}, \quad \gamma(0) = x, \quad \gamma(1) = y$$

Relationship:

Path connected ⇒ Connected (always)
Connected ⇒ Path connected (for "nice" spaces, but not always)

Counterexample (topologist's sine curve):

$$A = \{(x, \sin(1/x)) : x > 0\} \cup \{(0, y) : -1 \le y \le 1\}$$

```
    |∩   ∩   ∩
```

```
   ----+┘└---┘└---┘└---  ← oscillates infinitely
       |                   as x → 0
       |
     vertical
     segment
```

This is connected (in topological sense)
But not path connected: no continuous path from sine curve to segment

Connected Components

Connected component of a point x — maximal connected subset containing x.

Properties:

- X decomposes into disjoint connected components
- There can be finitely many, countably or uncountably many components

Examples:

$\mathbb{R} \setminus \{0\}$ — two components: $(-\infty, 0)$ and $(0, +\infty)$
$\mathbb{Z}$ (discr.) — countably many components: each point separately
Cantor set — uncountably many components (each point — a component)

H_0 in homology counts connected components:
$H_0(X) = \mathbb{Z}^k$, where k = number of components

Compactness — finiteness in infinity

Why compactness is needed

Problem: On infinite or non-closed sets many theorems break.

- $f(x) = x$ on $(0, +\infty)$: no maximum and minimum
- $f(x) = 1/x$ on $(0, 1)$: continuous, but unbounded
- A sequence in $\mathbb{R}$ can "escape" to infinity

Solution: Compactness — a property guaranteeing "finitely-like" behavior on infinite sets.

Analogy with thermal engineering:
Compact space — like a bounded system without "leaks".
Energy cannot "escape to infinity".

Definition of compactness (via coverings)

Covering of a set X – a family of sets $\{U_i\}_{i\in I}$ such that $X \subseteq \bigcup_i U_i$
Open covering – covering by open sets.

Definition:

```
+--------------------------------------------------------------+
| X is compact if from any open covering one can choose        |
| a finite subcovering.                                        |
|                                                              |
| ∀{Uᵢ}ᵢ∈I open covering of X                                 |
| ∃ finite J ⊂ I:  X ⊆ ⋃ⱼ∈J Uⱼ                               |
+--------------------------------------------------------------+
```

Compactness: any open covering has a finite subcovering
No matter how finely we cover it, finitely many suffice.

Example: Why [0,1] is compact, but (0,1) is not

(0, 1) is not compact:

Covering: $U_n = (1/n, 1)$ for n = 2, 3, 4, ...

```
U₂ = (1/2, 1)     +━━━━━━━━━━━━━+
U₃ = (1/3, 1)   +━━━━━━━━━━━━━━━━+
U₄ = (1/4, 1)  +━━━━━━━━━━━━━━━━━━+
⋮             ⋮
```

$\bigcup_n U_n = (0, 1)$ – covering ✓

But any finite subcovering $U_{n_1}, \dots, U_{n_k}$
covers only $(1/N, 1)$, where $N = \max(n_1,\dots,n_k)$.
Points $(0, 1/N)$ are not covered. ✗

[0, 1] is compact:

The same covering $\{(1/n, 1)\}$ does not cover [0,1] – does not contain 0.
Any covering of [0,1] must contain a neighborhood of 0 and a neighborhood of 1.
This "closes" the construction, allowing to choose a finite subcovering.

(Formal proof: Heine–Borel theorem)

Intuition: (0,1) is non-compact not because it's "large" – it is bounded.
The problem is absence of edges: coverings can "leak" to the boundary,
requiring infinitely many sets to "catch up" points near 0.
Closedness adds edges to which coverings "stick".

Heine–Borel theorem (for $\mathbb{R}^n$)

$$\boxed{\text{In } \mathbb{R}^n\text{: } X \text{ is compact} \iff X \text{ is closed and bounded}}$$

Warning: This is true only for $\mathbb{R}^n$. In a general topological space closedness + boundedness does not guarantee compactness.

Counterexample (critical for functional analysis):

Unit ball $B = \{f \in L^2 : \|f\| \leq 1\}$ in infinite-dimensional L^2:

- Closed? Yes
- Bounded? Yes
- Compact? No

The sequence $e_n(x) = \sin(n\pi x)$ is bounded, but has no convergent subsequence in L^2.

This breaks intuition from $\mathbb{R}^n$. In infinite-dimensional spaces weak compactness or additional conditions are needed.

Important: Although the ball B is non-compact in the strong (norm) topology, it is weakly compact (Banach–Alaoglu theorem). This is critically important for calculus of variations: minimum of energy exists precisely thanks to weak compactness.

Why this is a catastrophe for an engineer:

Weierstrass theorem: "continuous function on a compact attains min". But if the space is non-compact — the minimum may not exist.

In infinite-dimensional optimization problems (calculus of variations, neural network training) this means:

- We descend a "slope", but there is no bottom
- Minimizing sequence f_n "escapes to infinity"
- The function becomes ever thinner and taller, ceasing to be a function

Solution: add regularization ($\int |f'|^2 \leq C$), which makes the admissible set compact in the weak topology.

Examples in $\mathbb{R}^n$:

SET	CLOSED?	BOUNDED?	COMPACT?
$[0, 1]$	Yes	Yes	Yes
$(0, 1)$	No	Yes	No
$[0, +\infty)$	Yes	No	No
$\mathbb{N} = \{1, 2, 3, ...\}$	Yes	No	No
Unit ball D^2	Yes	Yes	Yes
Sphere S^2	Yes	Yes	Yes
$\mathbb{R}^n$	Yes	No	No
$\{1/n : n \in \mathbb{N}\}$	No	Yes	No
$\{0\} \cup \{1/n : n \in \mathbb{N}\}$	Yes	Yes	Yes (added 0!)

Properties of compact spaces

PROPERTY	FORMULATION
Image of compact	$f: X \to Y$ continuous, X compact $\Rightarrow f(X)$ compact
Weierstrass theorem (Key to optimization)	$f: X \to \mathbb{R}$ continuous, X compact $\Rightarrow f$ attains max and min
Closed ⊂ compact	X compact, $F \subseteq X$ closed $\Rightarrow F$ compact
Compact in Hausdorff	Y Hausdorff, $X \subseteq Y$ compact $\Rightarrow X$ closed in Y
Tychonoff theorem	X, Y compact $\Rightarrow X \times Y$ compact (works for any product)

Applications of compactness

AREA	HOW IT IS USED
Optimization	On compact extremum exists (Weierstrass)
Numerical methods	Convergence via subsequences
Physics	Compact phase space = closed system

Thermodynamics:
  Finite reservoir (compact domain) vs infinite medium.
  Properties of solutions of the heat equation depend essentially on compactness of the domain.

Topological invariants — how to distinguish spaces

Problem: how to prove that spaces are different?

Homeomorphism shows that spaces are the same.
But how to prove that a homeomorphism does not exist?

Idea: Associate with a space a number or group (invariant).
If invariants are different — spaces are definitely different.

Invariant = property preserved under homeomorphisms.

Simple invariants

INVARIANT	APPLICATION
Number of points $\|X\|$	$\|X\| \neq \|Y\| \Rightarrow X \not\cong Y$ (for finite)
Connectedness	X connected, Y not $\Rightarrow X \not\cong Y$ Example: $(0,1) \not\cong (0,1)\cup(2,3)$
Compactness	X compact, Y not $\Rightarrow X \not\cong Y$ Example: $[0,1] \not\cong (0,1)$
Number of components π_0	$\pi_0(X) \neq \pi_0(Y) \Rightarrow X \not\cong Y$

Problem: These invariants are too coarse.
S^1 and S^2 — both connected and compact, but different. Need more subtle invariants.

Main (subtle) invariants

INVARIANT	WHAT IT MEASURES
π_1 (fund. group)	"Group of loops" — distinguishes S^1 from S^2
χ (Euler char.)	V − E + F — "mesh" number
H_n (homology)	"Holes" of different dimensions

Fundamental group π_1 — detailed explanation

Intuition: Imagine that you tied a rope to a point and went walking around the space. You returned to the start. Can you shrink the rope to a point, without breaking and without leaving the space?

```
  On plane:                on plane with hole:

     ╭-----╮                  ╭-----╮
    /       \                /   ●   \   ← hole
   |    ●    |              |  /   \  |
   |   start |              | |     | |
    \       /                \ \   / /
     ╰-----╯                  ╰-----╯
```

Can shrink → loop   Cannot shrink → loop
"trivial"   "nontrivial"

Formal definition:

Loop in X with base point x_0 is a continuous map
$\gamma: [0,1] \to X$, where $\gamma(0) = \gamma(1) = x_0$

Two loops γ and δ are called homotopic ($\gamma \simeq \delta$), if one can be continuously deformed into the other, without breaking and without releasing x_0.

Homotopy = family of loops $H(s,t)$, where:

- $H(0,t) = \gamma(t)$ – start with γ
- $H(1,t) = \delta(t)$ – end with δ
- $H(s,0) = H(s,1) = x_0$ – base point is fixed

Fundamental group:

$$\pi_1(X, x_0) = \{\text{loops in X from } x_0\} / \{\text{homotopy}\}$$
$$= \text{set of homotopy classes of loops}$$

Group operation:

$$[\gamma] \cdot [\delta] = [\gamma * \delta]$$

where $\gamma * \delta$ = "first traverse γ, then δ":

$$(\gamma*\delta)(t) = \begin{cases} \gamma(2t) & \text{if } 0 \le t \le \tfrac{1}{2} \\ \delta(2t-1) & \text{if } \tfrac{1}{2} \le t \le 1 \end{cases}$$

- Identity element: constant loop $[x_0]$
- Inverse element: $[\gamma]^{-1} = [\gamma^{-1}]$, where $\gamma^{-1}(t) = \gamma(1-t)$

```
 Examples of fundamental groups
+---------------------+-------------------+-----------------------------------+
| SPACE               | π₁                | EXPLANATION                       |
+---------------------+-------------------+-----------------------------------+
|                     |                   |                                   |
| ℝⁿ (any n)          | {e} (trivial)     | Any loop can be shrunk to a point |
|                     |                   |                                   |
+---------------------+-------------------+-----------------------------------+
|                     |                   |                                   |
| Sⁿ for n ≥ 2        | {e} (trivial)     | On sphere (n≥2) any loop          |
| (sphere)            |                   | contracts (no "holes")            |
|                     |                   |                                   |
+---------------------+-------------------+-----------------------------------+
|                     |                   |                                   |
| S¹ (circle)         | ℤ                 | Loops are classified by winding   |
|                     |                   | number: ..., −2, −1, 0, 1, 2,...  |
|                     |                   | n>0: counterclockwise              |
|                     |                   | n<0: clockwise                     |
|                     |                   |                                   |
+---------------------+-------------------+-----------------------------------+
|                     |                   |                                   |
| T² (torus = donut)  | ℤ × ℤ             | Two independent loops:            |
|                     |                   | "around hole" and "through hole"  |
|                     |                   |                                   |
|                     |                   |   ╭----------╮                    |
|                     |                   |  /   ╭----╮    \                  |
|                     |                   | |  | ●→→→|     |  ← loop a        |
|                     |                   | |    ╰----╯    |                  |
|                     |                   |  \    ↑       /   ← loop b        |
|                     |                   |   ╰----------╯                    |
|                     |                   |                                   |
+---------------------+-------------------+-----------------------------------+
|                     |                   |                                   |
| ℝ² \ {0}            | ℤ                 | Plane without point ≃ circle      |
| (plane without 0)   |                   | Loops = windings around hole      |
|                     |                   |                                   |
+---------------------+-------------------+-----------------------------------+
|                     |                   |                                   |
| "Figure eight"      | F₂                | Free group on 2 generators        |
| S¹ ∨ S¹             | (free)            | Noncommutative: ab ≠ ba.          |
|                     |                   |                                   |
+---------------------+-------------------+-----------------------------------+
|                     |                   |                                   |
| ℝP² (projective     | ℤ/2               | There is a loop, not contractible,|
| plane)              |                   | but traversed twice — contracts   |
|                     |                   |                                   |
+---------------------+-------------------+-----------------------------------+
```

Why $\pi_1(S^1) = \mathbb{Z}$ — detailed explanation

Circle $S^1 = \{e^{i\theta} : \theta \in [0, 2\pi)\} \subset \mathbb{C}$
Base point: $x_0 = 1$ (at $\theta = 0$)

Loop with n windings: $\gamma_n(t) = e^{2\pi int}$, $t \in [0,1]$

$n = 0$: $\gamma_0(t) = 1$ (staying in place)
$n = 1$: $\gamma_1(t) = e^{2\pi it}$ (one winding counterclockwise)
$n = -1$: $\gamma_{-1}(t) = e^{-2\pi it}$ (one winding clockwise)
$n = 2$: $\gamma_2(t) = e^{4\pi it}$ (two windings counterclockwise)

Theorem: $\gamma_n \simeq \gamma_m \iff n = m$

Intuitively: cannot "unwind" a winding without breaking the loop.

Group operation: $[\gamma_n] \cdot [\gamma_m] = [\gamma_{n+m}]$

This is exactly the group $(\mathbb{Z}, +)$!

Corollary: $S^1 \not\cong S^2$ (sphere)
$\pi_1(S^1) = \mathbb{Z} \neq \{e\} = \pi_1(S^2)$
Different groups ⇒ spaces are not homeomorphic.

Applications of the fundamental group

1. Topology: Distinguishing spaces
   $\pi_1(X) \neq \pi_1(Y) \Rightarrow X \not\cong Y$

2. Borsuk–Ulam theorem:
   On the surface of the Earth there are two antipodal points with identical temperature and pressure. (Related to the fact that $\pi_1(\mathbb{R}P^2) = \mathbb{Z}/2$)

3. Knot theory:
   $\pi_1(\mathbb{R}^3 \setminus \text{knot})$ distinguishes knots (which knots can be untied?)

4. Physics:
   - Defects in crystals are classified by π_1
   - Vortices in superfluid helium: $\pi_1(\text{parameter space}) = \mathbb{Z}$
   - Magnetic monopoles: $\pi_2(\text{parameter space})$

5. Coverings:
   Covering spaces ↔ subgroups of π_1
   Universal covering ↔ $\pi_1 = \{e\}$

Higher homotopy groups π_n

π_1 measures "holes for loops". But what if we use spheres?

$\pi_1(X)$ — classes of maps $S^1 \to X$ (loops)
$\pi_2(X)$ — classes of maps $S^2 \to X$ (spheres)
$\pi_n(X)$ — classes of maps $S^n \to X$ (n-spheres)

Examples:

GROUP	VALUE
$\pi_1(S^1) = \mathbb{Z}$	Loop can go around circle n times
$\pi_1(S^2) = 0$	On sphere any loop contracts
$\pi_2(S^2) = \mathbb{Z}$	S^2 can "wrap" S^2 an integer number
$\pi_3(S^2) = \mathbb{Z}$	Hopf fibration. (unexpectedly $\neq 0$)
$\pi_n(S^n) = \mathbb{Z}$	Identity map generates

Deep fact: $\pi_3(S^2) = \mathbb{Z} \neq 0$

This means that a 3-sphere can be nontrivially mapped onto a 2-sphere.
This map is called the Hopf fibration: $S^3 \to S^2$ with fiber S^1.
Each point of S^2 is a circle in S^3.

In physics:
- π_2 – classification of magnetic monopoles
- π_3 – classification of instantons
- Hopf fibration – related to electron spin

Homology – algebra of "holes" in space

Intuition: Homology answers the question "what holes are in the space?"

H_0 counts connected components (how many separate pieces)
H_1 counts tunnels (holes through which one can thread a rope)
H_2 counts cavities (closed voids, like inside a ball)

Key idea: cycle vs boundary

Cycle = closed contour (beginning coincides with end)
Boundary = contour that bounds some region

Example on plane with hole:

```
       +----------------+         +----------------+
```

```
       |    ╭---╮        |          |                 |
       |   /     \   γ   |          |    ╭---╮        |
       |  |   ●   |←-----|          |   /     \   δ   |
       |   \     /       |          |  |       |←-----|
       |    ╰---╯        |          |   \     /       |
       +-----------------+          |    ╰---╯        |
       γ goes around hole ●         +-----------------+
                                    δ does not go around hole

  γ — cycle, but not boundary        δ — cycle and boundary
  (cannot "fill" — there's a hole)   (can be filled with disk)
```

H_1 = {cycles} / {boundaries} = "cycles that cannot be filled"

For plane with hole: $H_1 = \mathbb{Z}$ (one generator — going around hole)
For ordinary plane: $H_1 = 0$ (all cycles can be filled)

Construction: from gluing to algebra

The idea of homology is to translate geometry into linear algebra. The space is cut into simple pieces, the pieces are written as formal sums, and the question "is there a hole?" turns into the question "is a system of linear equations solvable?".

Step 1: Cutting into simplices

Simplex — minimal "brick" of each dimension:

- 0-simplex = point ●
- 1-simplex = segment ●---●
- 2-simplex = triangle △
- 3-simplex = tetrahedron ▲

Any "decent" space can be cut into simplices, glued along entire faces. Gluing of two simplices is their union, where the common part is an entire face of some dimension. The result is a simplicial complex.

Step 2: Orientation — where signs come from

For the algebra to work, simplices must be oriented:

- Point: sign + or −
- Segment: direction chosen (start → end)
- Triangle: direction of traversal (clockwise or counterclockwise)
- Tetrahedron: consistent orientation of faces

Orientation is not decoration, but necessity. Without it, it's impossible to define signs, and the whole construction falls apart. Changing orientation is equivalent to multiplication by −1.

Key property: when gluing two oriented triangles along a common edge, this edge enters each triangle with opposite orientations — and cancels in the sum. All internal edges disappear, only the external boundary remains.

```
      A                       A---D
     / \    gluing along     / \ /    ∂ = external contour
    / → \   edge AC:        / → \     (AC cancelled)
   B-----C                 B-----C
```

Step 3: Chains — formal sums

Chain of dimension k is a formal sum of oriented k-simplices with integer coefficients:

$$c = n_1\sigma_1 + n_2\sigma_2 + ... + n_m\sigma_m, \quad n_i \in \mathbb{Z}$$

Coefficient $n_i = -1$ means the same simplex with reverse orientation. The set of all k-chains is denoted C_k. This is an abelian group:

- (AB + BC) + (CD) = AB + BC + CD
- Zero chain: 0
- Inverse: −(AB) = BA (reverse orientation)

Step 4: Boundary operator ∂

Operator ∂ takes the oriented boundary:

$$\partial(\text{segment } AB) = B - A \quad \text{(end minus start)}$$

$$\partial(\text{triangle } ABC) = BC + CA + AB \quad \text{(traversal along boundary)}$$

```
          A
         / \          ∂△ = AB + BC + CA
        /   \         (closed contour)
       B-----C
```

$$\partial(\text{tetrahedron } ABCD) = BCD - ACD + ABD - ABC \quad \text{(four faces with signs)}$$

Properties of ∂:

- Linearity: $\partial(c_1 + c_2) = \partial c_1 + \partial c_2$
- Dimension reduction: $\partial: C_k \to C_{k-1}$
- $\partial\partial = 0$ (boundary of boundary is empty)

The last is not magic, but a consequence of orientation. Let's check:

$$\begin{aligned}\partial(\partial\,\triangle ABC) &= \partial(AB + BC + CA)\\ &= (B-A) + (C-B) + (A-C)\\ &= 0 \ \checkmark\end{aligned}$$

Each vertex enters exactly twice: as the start of one edge and as the end of another. The signs are opposite — everything cancels. This works in any dimension: $\partial_{k-1} \circ \partial_k = 0$.

Step 5: Chain complex and homology

In total we have a sequence of groups and operators:

$$\ldots \to C_{k+1} \to\partial_{k+1} C_k \to\partial_k C_{k-1} \to \ldots$$

From $\partial\partial = 0$ follows: $\mathrm{Im}(\partial_{k+1}) \subseteq \ker(\partial_k)$. Every boundary is a cycle.
Question: is the converse true? Is every cycle a boundary?

If yes – there are no holes. If no – "extra" cycles are the holes.

$Z_k = \ker(\partial_k)$ – cycles (chains without boundary)
$B_k = \mathrm{Im}(\partial_{k+1})$ – boundaries (chains that are someone's boundary)

$H_k = Z_k / B_k$ – k-th homology group

H_k measures the gap between "being a cycle" and "being a boundary".
If $H_k = 0$ – all k-cycles are boundaries (no k-holes).
If $H_k \neq 0$ – there are cycles that cannot be filled.

Connection with linear algebra: ∂_k is a matrix (huge, sparse, almost all elements 0, the rest ±1). Cycles = kernel of matrix. Boundaries = image of another matrix. Homology = kernel/image. The whole problem reduces to linear algebra, though the matrices are gigantic.

Example: torus T^2

$H_1(T^2) = \mathbb{Z} \oplus \mathbb{Z}$ – two independent generators: [meridian] and [parallel]. Any cycle on the torus is homologous to $n\cdot$[meridian] + $m\cdot$[parallel] for integers n, m.

Meridian – cycle (closed), but not boundary: it cannot be "filled" with a disk while remaining on the surface of the torus.

Examples:

SPACE	H_0	H_1	H_2	INTERPRETATION
Point	$\mathbb{Z}$	0	0	1 piece, no hole
Circle S^1	$\mathbb{Z}$	$\mathbb{Z}$	0	1 tunnel
Sphere S^2	$\mathbb{Z}$	0	$\mathbb{Z}$	1 cavity
Torus T^2	$\mathbb{Z}$	$\mathbb{Z}\oplus\mathbb{Z}$	$\mathbb{Z}$	2 tunnels,1 cav.
Donut (solid)	$\mathbb{Z}$	$\mathbb{Z}$	0	1 tunnel

Connection with Euler characteristic:

$$\chi = \mathrm{rank}(H_0) - \mathrm{rank}(H_1) + \mathrm{rank}(H_2) - \ldots$$

For sphere: $\chi = 1 - 0 + 1 = 2$ ✓
For torus: $\chi = 1 - 2 + 1 = 0$ ✓

Cohomology — a dual view through integration

Homology: "what closed surfaces are there in the space?"
Cohomology: "what forms can be integrated over these surfaces?"

Intuition:

Suppose we have a 1-form ω (like df = "gradient").
The integral $\int_\gamma \omega$ depends only on the class of γ in H_1.

If $\gamma_1 \simeq \gamma_2$ (homologous), then $\int_{\gamma_1} \omega = \int_{\gamma_2} \omega$.

Closed vs exact form:

Closed form: $d\omega = 0$
Exact form: $\omega = df$ for some function f

Exact ⇒ closed (because $d^2 = 0$)
But not conversely. On spaces with holes there are closed but not exact forms.

Example — $d\theta$ on the circle:

The form $d\theta$ on S^1 is closed ($d(d\theta) = 0$).
But $\int_{S^1} d\theta = 2\pi \neq 0$.

If $d\theta = df$, then $\int_{S^1} d\theta = f(\text{end}) - f(\text{start}) = 0$.
Contradiction. Therefore $d\theta$ is not exact on S^1.

de Rham cohomology:

$$H^k_{dR}(M) = \{\text{closed } k\text{-forms}\} / \{\text{exact } k\text{-forms}\}$$
$$= \text{"forms that cannot be represented as } df\text{"}$$

Pairing (de Rham's theorem):

$\langle \omega, \gamma \rangle = \int_\gamma \omega$ — integral of form over cycle

This gives an isomorphism: $H^k_{dR}(M) \cong H_k(M; \mathbb{R})^*$

Cohomology = "linear functionals on homology"
(duality in action)

CW-complexes — constructor for spaces

Idea: Build spaces from simple pieces — cells (disks).
Like LEGO: start with points, attach segments, then disks, etc.

Cells:
0-cell = point ●

```
  1-cell = open interval        ●-------●  (endpoints — 0-cells)
  2-cell = open disk            ○  (boundary — 1-cells)
  n-cell = open n-disk          (boundary — (n−1)-cells)
```

"Attaching" — what does it mean:

Take an n-disk D^n and map its boundary $\partial D^n = S^{n-1}$
to an already constructed skeleton X^(n-1).

New space = X^(n-1) ∪_f D^n (gluing along f: S^{n-1} → X^(n-1))

Example: how to construct the sphere S^2

Method 1 (minimal):

Step 0: One 0-cell (point) ●
Step 2: Attach a 2-cell, entire boundary goes to this point

```
        ●  ----------►   ○
     (point)         (entire boundary    (result = S²)
                      collapsed
                      to a point)
```

S^2 = 1 zero-dimensional cell + 1 two-dimensional cell

Method 2 (like a globe):

poles (0-cells) + 1 meridian (1-cell) +
hemispheres (2-cells)

Example: how to construct the torus T^2

Torus = square with opposite sides glued:

```
      a --------► a              After gluing:
      |           |                    ╭-------╮
    b |           | b                 /         \
      |           |                  |  ╭-----╮  |
      ▼           ▼                  |  |     |  |
      a --------► a                   \ ╰-----╯  /
                                       ╰--------╯
```

T^2 = 1 point + 2 loops (a and b) + 1 square
= 1·(0-cell) + 2·(1-cell) + 1·(2-cell)

Why CW-complexes:

1. Easy to compute homology: H_n depends only on n-cells and (n±1)-cells
2. Euler characteristic: χ = (number of 0-cells) − (number of 1-cells) + (number of 2-cells)
3. Many spaces have a simple cellular structure

For S^2: $\chi = 1 - 0 + 1 = 2$ ✓
For T^2: $\chi = 1 - 2 + 1 = 0$ ✓

Homotopy groups of spheres — why it's hard and interesting

Question: In how many essentially different ways can one map
one sphere to another?

$\pi_n(S^m)$ = classes of mappings $S^n \to S^m$ (up to deformation)

Simple cases (intuition works):

$\pi_1(S^1) = \mathbb{Z}$: A loop on the circle can wind 0, 1, 2, ... times.
The winding number is an integer.

$\pi_1(S^2) = 0$: Any loop on a sphere can be contracted to a point
(no "hole" to catch on).

$\pi_n(S^n) = \mathbb{Z}$: A sphere onto itself can be "wrapped" an integer number of times.
(Degree of mapping)

Surprise: $\pi_3(S^2) = \mathbb{Z} \neq 0$

It would seem: the 3-dimensional sphere is "larger" than the 2-dimensional one. Any mapping $S^3 \to S^2$ must compress the extra dimension, everything should contract.

But no. There exists a nontrivial mapping — the Hopf fibration:

$h: S^3 \to S^2$, where the preimage of each point is a circle S^1

Imagine: S^3 is "ruled" by non-intersecting circles,
and each circle corresponds to a point on S^2.

$S^3 = \bigsqcup$ (circles), indexed by points of S^2

These circles are linked with each other. Any two are like links of a chain.

Picture (two linked circles — fibers over two points of S^2):

```
              ┌──────────┐
             /            \
            /   ┌─────────╫──────┐
           |   /           \      \
           |  |    ●         |      \
            \  \   fiber    /       |
             \  \  over p₁ /        |         ← fiber over p₂ ∈ S²
              \──╫────────┘         |           passes through the
                  \                 |           fiber over p₁;
                   \       ●       /           they cannot be
                    \    fiber   /             separated
                     \   over p₂/              continuously
                      └────────┘
```

Fibers form torus-like surfaces nested inside each other like rings. Every pair of fibers (over distinct points $p, q \in S^2$) has linking number exactly 1 — this is what makes $[h] \in \pi_3(S^2)$ nontrivial: the map cannot be "untangled" to a constant.

Physics: The Hopf fibration describes the topology of electron spin.
Also: gauge fields, Dirac magnetic monopoles, instantons.

Table of groups $\pi_{n+k}(S^n)$:

k\n	1	2	3	4	5	∞
0	$\mathbb{Z}$	$\mathbb{Z}$	$\mathbb{Z}$	$\mathbb{Z}$	$\mathbb{Z}$	$\mathbb{Z}$
1	0	$\mathbb{Z}$	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}_2$
2	0	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}_2$
3	0	$\mathbb{Z}_2$	$\mathbb{Z}_{12}$	$\mathbb{Z}\oplus\mathbb{Z}_{12}$	$\mathbb{Z}_{24}$	$\mathbb{Z}_{24}$

Column ∞ — stable groups: for $n \geq k+2$ the answer doesn't change.

Why such strange numbers ($\mathbb{Z}_{12}$, $\mathbb{Z}_{24}$)?

This is a deep result. Where does 24 come from? Connection with:

- 24 = dimension of the Leech lattice (coding theory)
- 24 divisors of modular forms (number theory)
- 24 in Riemann's formula for $\zeta(-1) = -1/12$ (via $2 \cdot 12 = 24$)

Homotopy groups of spheres are a window into deep connections of different areas of mathemati

Stabilization and Suspension — Why the Table "Settles Down"

Suspension Σ: we take a space X and "suspend" it:

$$\Sigma X = (X \times [0,1]) \,/\, (X\times\{0\} \sim \text{point},\ X\times\{1\} \sim \text{point})$$
= "two cones, glued along X"

Example: $\Sigma S^0 = S^1$ (two cones over two points = circle)
$\Sigma S^1 = S^2$ (two cones over a circle = sphere)
$\Sigma S^n = S^{n+1}$

Freudenthal Theorem:

Suspension induces a homomorphism: $\Sigma: \pi_{n+k}(S^n) \to \pi_{n+k+1}(S^{n+1})$

For $n \geq k+2$ this homomorphism is an isomorphism.

Here k is the "stem" (k = index of homotopy group minus n).
Example: $\pi_3(S^2) \to \pi_4(S^3) \to \pi_5(S^4) \to \ldots$ — all have stem k=1.
Condition: $n \geq k+2 = 3$. Starting from n=3 (i.e. $\pi_4(S^3)$) the group stabilizes = $\mathbb{Z}_2$.

Stable groups π_k^s:

$$\pi_k^s = \lim \pi_{n+k}(S^n) \quad \text{as } n \to \infty$$

These are the "true" invariants, purified from dimensional effects. Computing π_k^s is one of the main problems of algebraic topology.

Computation tools:

- Hurewicz Theorem: relates π_n and H_n for "nice" spaces
- Exact sequences: relate π for fibrations $F \to E \to B$
- Spectral sequences: systematic computation method
- Eilenberg–MacLane spaces $K(G,n)$: standard building blocks

Euler Characteristic — A Number Determining Shape

Definition (for polyhedra):

$$\chi = V - E + F$$

V = number of vertices
E = number of edges
F = number of faces

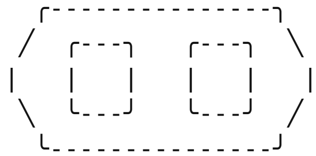
Examples:

FIGURE	V	E	F	χ
Tetrahedron	4	6	4	2
Cube	8	12	6	2
Octahedron	6	12	8	2
Dodecahedron	20	30	12	2
Icosahedron	12	30	20	2

Euler's Theorem: For any convex polyhedron $\chi = 2$.

Moreover: χ is a topological invariant. Depends only on the "shape" of the surface, not on the subdivision into faces.

Euler characteristic of surfaces:

SURFACE	χ	EXPLANATION
Sphere S^2	2	"Ball": no holes
Torus T^2 (donut)	0	One "hole"
Double torus 	−2	Two "holes"
g-torus (genus g)	2−2g	g "holes"
Projective plane $\mathbb{R}P^2$	1	Non-orientable
Klein bottle	0	Non-orientable, $\chi = 0$ like torus, but π_1 different.

Formula: $\chi = 2 - 2g$ (for orientable surfaces of genus g)

What is genus g?
Genus = "number of holes" in the surface.
Sphere: g=0 (no holes). Torus: g=1 (one hole, like in a donut).
g-torus is obtained from a sphere by "attaching g handles".
Formally: $g = (2 - \chi)/2$ for orientable closed surfaces.

Connection with other invariants:

$$\chi = \Sigma\ (-1)^n \cdot \mathrm{rank}(H_n) \quad \text{(via homology groups)}$$

Applications:

- Hairy ball theorem: on S^2 ($\chi=2$) you cannot comb a hedgehog (a vector field necessarily has zeros)
- On a torus ($\chi=0$) combing is possible.
- Gauss–Bonnet formula: $\iint K\, dA = 2\pi\chi$ (connection of curvature and topology)

Classification of Compact Surfaces

Connected sum M # N (operation on surfaces):
1. Cut out a small disk from M and N
2. Glue M and N along the boundaries of these disks

Result: a new surface. Example: torus = S^2 # (handle).

Theorem (classification):
Every connected compact surface is homeomorphic to exactly one of:

Orientable:

- Sphere S^2 (genus 0)
- Torus T^2 (genus 1)
- Double torus (genus 2)
- ... g-fold torus (genus g)

Non-orientable:

- Projective plane $\mathbb{R}P^2$
- Klein bottle
- ... (connected sums with $\mathbb{R}P^2$)

Orientability — can we define "right" and "left"?

On a sphere or torus: Yes — going around, right remains right.

On a Möbius strip: No — going around, right becomes left.

Möbius strip:

Take a strip of paper, twist by 180° and glue the ends.

Result: a surface with one side.

```
┌-----------------┐
|   →  →  →  →    | ← one side
└-----------------┘
     ↓ glue with a twist
┌-----------------┐
|   ←  →  →  ←    | ← the same side.
└-----------------┘
```

Klein bottle = Möbius strip, "sealed" into a tube
(cannot be embedded in $\mathbb{R}^3$ without self-intersection)

Invariants distinguishing surfaces:

SURFACE	χ	π_1	ORIENTABLE?
Sphere S^2	2	{e}	Yes
Torus T^2	0	$\mathbb{Z}\times\mathbb{Z}$	Yes
Klein bottle	0	other	No
$\mathbb{R}P^2$	1	$\mathbb{Z}/2$	No

Important: χ and π_1 together completely classify compact surfaces.

Table of invariants — how to distinguish spaces

Idea of an invariant

Problem: how to prove that two spaces are different?
Torus ≠ sphere — this is "visible", but a rigorous argument is needed.

Solution: assign to a space a number or a group.
If the numbers are different → the spaces are definitely different.

Invariant = that which does not change under deformation (without tears and gluing).

Main invariants with explanations

INVARIANT	WHAT IT MEANS IN SIMPLE WORDS
π_0 (connectedness)	How many pieces? If the object falls apart into parts. π_0(● ●) = 2 (two points), π_0(○) = 1 (one circle)
π_1 (fund. group)	Which loops cannot be contracted to a point? On the plane — all loops contract: $\pi_1 = 0$ On the circle — loop around doesn't contract: $\pi_1 = \mathbb{Z}$ On the torus — two independent loops: $\pi_1 = \mathbb{Z}\times\mathbb{Z}$
χ (Euler)	Formula $V - E + F$ for subdivision into polygons. Sphere: $\chi = 2$ (any subdivision) Torus: $\chi = 0$ Surface with g holes: $\chi = 2 - 2g$
Orientability	Are there "two sides"? Sphere, torus — YES (can be painted in two colors) Möbius strip, Klein bottle — no (one side)
Dimension	How many numbers are needed to specify a point? Line: dim = 1, plane: dim = 2, space: dim=3
Compactness	Bounded and closed? Sphere — compact (finite), $\mathbb{R}$ — no (infinite)

How to use

Invariant differs → spaces are definitely different.
$\chi(\text{sphere}) = 2 \neq 0 = \chi(\text{torus})$ → sphere ≇ torus ✓

Invariants coincide → not yet a fact that they are the same.
Need to check other invariants or look for explicit isomorphism.

Example: How to prove that mug = torus?
$\chi = 0$ for both, $\pi_1 = \mathbb{Z}$ for both, both orientable.
But this is not a proof. Need to construct a deformation.

Summary: central concepts of topology

Hierarchy of concepts

```
  Set X
      |
      | + topology τ (family of "open" sets)
      ↓
  Topological space (X, τ)
      |
      +--→ Open/closed sets
      |    Boundary, interior, closure
      |
      +--→ Continuous mappings f: X → Y
      |    (preimage of open is open)
      |
      +--→ Homeomorphism (topological equivalence)
      |    (continuous bijection with continuous inverse)
      |
      +--→ Connectedness (cannot be split into two open sets)
      |
      +--→ Compactness (finite subcovers)
      |
      +--→ Invariants (π₁, Hₙ, χ, ...)
           for distinguishing spaces
```

Connection with Other Structures

```
                    Topology (nearness)
                         |
            + metric     |
            (distance)   |
```

```
                     ↓
             Metric space
                     |
      + lin. structure  |
        + norm (‖x‖,    |
      d(x,y) = ‖x−y‖)  ↓
           Normed space
                     |
         + completeness |
       (all limits      |
        exist)          ↓
            Banach space
                     |
       + scal. product  |
       (‖x‖²=⟨x,x⟩)    ↓
            Hilbert space
```

Each level inherits topology from the previous one,
but adds additional structure.

Practical Consequences

For engineer / practitioner:

- Continuity of functions — key to numerical methods
  (close inputs give close outputs = stability)

- Compactness — guarantee of existence of solutions
  (extrema exist, sequences converge)

- Connectedness — "integrity" of system
  (is it possible to pass from one state to another?)

- Topological invariants — "shape" of parameter space
  (how many "holes" in admissible region?)

Complete Scheme of Embedding of Spaces

Each level adds structure and restricts the class of spaces:

```
+-----------------------------------------------------------------------+
|                                                                       |
|   SET                                                                 |
```

```
|       |                                                                  |
|       | + topology τ                                                     |
|       ↓                                                                  |
|   TOPOLOGICAL SPACE                                                      |
|       |                                                                  |
|       | + axiom T₂ (Hausdorffness)                                       |
|       ↓                                                                  |
|   Hausdorff space (uniqueness of limits)                                 |
|       |                                                                  |
|       | + metrizability (topology generated by metric)                   |
|       ↓                                                                  |
|   metric space (notion of distance)                                      |
|       |                                                                  |
|       | + completeness (all Cauchy seq-s converge)                       |
|       ↓                                                                  |
|   COMPLETE METRIC SPACE                                                  |
|       |                                                                  |
|       | + linear structure + norm                                        |
|       ↓                                                                  |
|   Banach space (complete normed)                                         |
|       |                                                                  |
|       | + norm from scalar product                                       |
|       ↓                                                                  |
|   Hilbert space (complete with scal. product)                            |
|                                                                          |
+--------------------------------------------------------------------------+
```

Examples at each level:

LEVEL	EXAMPLES
Topological	$\mathbb{R}^n$, any set with any topology
Hausdorff	$\mathbb{R}^n$, manifolds, but not antidiscrete
Metric	$\mathbb{R}^n$, $C[0,1]$, ℓ^p, any with metric
Complete metric	$\mathbb{R}^n$, $C[0,1]$, ℓ^p, but not $\mathbb{Q}$, not $(0,1)$
Banach	$\mathbb{R}^n$, $C[0,1]$, ℓ^p $(p\geq1)$, L^p
Hilbert	$\mathbb{R}^n$, L^2, ℓ^2, but not $C[0,1]$, not ℓ^1

Note: L^p – space of functions with finite $\int|f|^p$. Complete definition requires Lebesgue integral. In $\mathbb{R}^n$ one can use Riemann.

What is a norm

A norm $\|\cdot\|$ on a vector space V is the "length of a vector":

- $\|x\| \geq 0$, moreover $\|x\| = 0 \iff x = 0$
- $\|\alpha x\| = |\alpha|\cdot\|x\|$ (scaling)
- $\|x + y\| \leq \|x\| + \|y\|$ (triangle inequality)

Norm generates a metric: $d(x,y) = \|x - y\|$
Examples: $\|x\|_2 = \sqrt{(\Sigma x_i^2)}$ in $\mathbb{R}^n$, $\|f\|_2 = \sqrt{(\int|f|^2)}$ in L^2

Metrizability — how to prove that a space is (not) metrizable

Definition:
A topological space (X, τ) is called metrizable if there exists a metric d on X generating the topology τ.

Necessary conditions for metrizability (if violated — not metrizable)

1. Hausdorff property (T_2):
   A metric space is always T_2.
   ⇒ If a space is not Hausdorff — it is not metrizable.

   Example: Antidiscrete topology on {a,b} — points are inseparable.

2. First axiom of countability:
   In a metric space each point has a countable base of neighborhoods (balls of radius 1/n).
   ⇒ If there is no countable base of neighborhoods — not metrizable.

3. Sequentiality:
   In a metric space closedness is equivalent to sequential closedness (contains limits of its sequences).
   ⇒ If there exists sequentially closed but not closed — not metrizable.

Sufficient conditions (metrization theorems)

Two axioms of countability:

- First (A_1): each point has a countable base of neighborhoods (balls of radius 1/n in a metric space)

- Second (A_2): the entire topology has a countable base

(any open = union of elements of a countable family)

$A_2 \Rightarrow A_1$, but not conversely. Example: uncountable discrete space satisfies A_1 (base of a point = the point itself), but not A_2.

Urysohn's theorem:
Regular (T_3) + Second axiom of countability ⇒ metrizable.
(Countable base of the entire topology is needed, not just neighborhoods of points)

Nagata–Smirnov theorem:
T_3 + σ-locally finite base ⇒ metrizable.

Strategies for proving non-metrizability

Strategy 1: Violation of T_2
Show that two points cannot be separated by disjoint neighborhoods.
Example: X with antidiscrete topology ($\tau = \{\varnothing, X\}$).

Strategy 2: No countable base of neighborhoods
Find a point for which any countable system of neighborhoods is not a base.
Example: Uncountable product $\prod_i \mathbb{R}$ with product topology.

Strategy 3: Sequential argument
Find a set A which is sequentially closed but not topologically closed.

Strategy 4: Compactness + uncountability
In a compact metric space every open cover has a countable subcover (Lindelöf property).
⇒ If compact but no countable subcover – not metrizable.

Classical examples

SPACE	METRIZABLE?	WHY
$\mathbb{R}^n$ (standard topology)	YES	Euclidean metric

$\mathbb{R}$ (Zariski topology)	NO	Not T_2: closed = finite
$\mathbb{R}$ (arrow, Sorgenfrey topology)	NO	Separable, but no countable base (T_2, but not 2-countable)
$[0,1]^{[0,1]}$ (Tychonoff cube)	NO	No countable base of neighborhoods
$\beta\mathbb{N}$ (Stone-Čech)	NO	Uncountable discrete subspace in a compact
ω_1 (first uncountable ordinal)	NO	Sequentially compact, but not compact

Step-by-step example: Prove that $[0,1]^{\mathbb{R}}$ is not metrizable

1. This is an uncountable product of compacta ⇒ compact (Tychonoff).
2. Suppose it is metrizable.
3. Then it has a countable base of neighborhoods at each point.
4. Consider the point $x = (0, 0, 0, \ldots)$ – zero in all coordinates.
5. A basic neighborhood depends on finitely many coordinates.
6. Countable base ⇒ affects a countable set of coordinates.
7. But there are uncountably many coordinates ⇒ contradiction.

Conclusion: $[0,1]^{\mathbb{R}}$ is not metrizable, because there is no countable base of neighborhoods.

Applied example: connectivity of a pipeline network

Problem: City heating network. Need to understand what happens in case of failure.

```
Boiler house
   ●===========●===========● District A
   |           |           |
   |           |           |
   ●===========●===========● District B
   |           |           |
   |           |           |
   ●===========●===========● District C
```

Topological view:

1. Connectivity:
   Network is connected = from any point one can reach any other.
   If failure cuts network into two parts → network becomes disconnected.

2. Number of independent cycles = first Betti number β_1:

   $\beta_1 = E - V + 1$ (for connected graph)
   E = number of edges (pipelines)

V = number of vertices (nodes)

For our network: $V = 9, E = 12 \Rightarrow \beta_1 = 12 - 9 + 1 = 4$

Meaning: Can disconnect 4 pipes without loss of connectivity.
(with correct choice)

3. Practical conclusion:

- $\beta_1 = 0$: tree-like network, any failure disconnects part of consumers
- $\beta_1 > 0$: ring network, there are reserve paths
- The larger β_1, the more reliable the network (but more expensive construction)

Topological equivalence in heat exchangers:

Two heat exchangers with same flow topology — work identically.

```
  Pipe in pipe:      ===============     Counterflow: efficiency ~100%
                     ---------------

  Plate:             +===+ +===+        Counterflow with mixing
                     |   | |   |
                     +===+ +===+        efficiency ~95%

  Mixing:            liquids             Parallel flow (topologically
                     mix                 different) — efficiency ~50%
```

Flow topology (counterflow/parallel flow/crossflow) determines theoretical limit of heat exchange efficiency.

Topology gives notion of closeness, but does not say how to measure and compute. To work with space quantitatively — add vectors, solve equations, find projections — additional structure is needed.

Linear algebra adds to space operations of addition and multiplication by number. This makes computations possible — and this is the language of almost all physics.

Linear algebra — the language of linear transformations

Linear algebra adds to space flatness. Now space is not just "connected" (topology) — it is flat, without curvature. One can draw straight line, one can add two vectors, one can stretch vector by factor of two.

In terms of "object—observer" linear algebra — this is place where observer appears for real for first time.

Basis — this is choice of observer. Coordinates — his language.

One and the same vector v exists independently of basis. But to write it as numbers (3, 4, 5), need to choose basis. Another observer in different basis will write same vector as (1, 2, 6). Vector did not change – notation changed.

Transition matrix between bases – this is dictionary for translation between languages of two observers. Formula $A' = P^{-1}AP$ says: "here is how same operator A looks through eyes of new observer".

What is invariant (does not depend on observer):

- Rank of matrix – how many independent directions it uses
- Determinant – by what factor volume changes
- Trace – sum of eigenvalues
- Characteristic polynomial – "DNA" of operator

These invariants – real properties of the object. The concrete numbers in a matrix are merely a notational choice, depending on the chosen basis.

Linear algebra as view on space

Linear algebra sees in space two operations:

- Addition of vectors (can add points)
- Multiplication by number (can stretch/compress)

This is more structure than topology has (only closeness), but less than metric has (concrete distances).

Key fact: Almost all spaces in physics and engineering – vector.

- Forces can be added
- Velocities can be added
- States in quantum mechanics can be added

Linear algebra – second most important structure after sets. It describes everything that can be "added" and "stretched": vectors, functions, matrices.

Why this is central topic:

- Quantum mechanics lives in vector spaces
- Nonlinear problems are linearized (derivative)
- Data – these are vectors in $\mathbb{R}^n$
- Linear systems are solved explicitly

Place in the Overall Picture

Linear algebra is the study of vector spaces and linear mappings between them.

Why this is important:

- This is the simplest structure where one can add and stretch

- Linear problems are solvable (unlike nonlinear ones)
- Nonlinear problems are approximated by linear ones (derivative)
- Quantum mechanics lives in vector spaces

```
+--------------------------------+
| HIERARCHY OF TRANSFORMATIONS   |
+--------------------------------+
|                                |
| Arbitrary mappings             |
| ↓ preserve continuity          |
| Continuous (topology)          |
| ↓ preserve smoothness          |
| Smooth (analysis, manifolds)   |
| ↓ preserve linear structure    |
| Linear  ← we are here          |
| ↓ preserve angles and lengths  |
| Orthogonal (group O(n))        |
| ↓ preserve orientation         |
| Rotations (group SO(n))        |
|                                |
+--------------------------------+
```

Vector space – formal definition

```
+-------------------------------------------------------------------+
| DEFINITION: VECTOR SPACE                                          |
+-------------------------------------------------------------------+
|                                                                   |
| A vector space over a field F is a set V with two                 |
| operations: addition (+) and multiplication by scalar (·), such that: |
|                                                                   |
| addition axioms (V is an abelian group under +):                  |
| (V1) u + v = v + u                    (commutativity)             |
| (V2) (u + v) + w = u + (v + w)        (associativity)             |
| (V3) ∃ 0 ∈ V: v + 0 = v              (neutral element)            |
| (V4) ∀v ∃(−v): v + (−v) = 0          (inverse element)            |
|                                                                   |
| SCALAR MULTIPLICATION AXIOMS:                                     |
| (V5) α(βv) = (αβ)v                   (associativity)              |
| (V6) 1·v = v                         (field unity)                |
|                                                                   |
| DISTRIBUTIVITY AXIOMS:                                            |
| (V7) α(u + v) = αu + αv              (over vectors)               |
| (V8) (α + β)v = αv + βv              (over scalars)               |
|                                                                   |
+-------------------------------------------------------------------+
```

Characteristic property: A vector space admits:

- Adding objects (parallelogram)
- Stretching/compressing objects (multiplication by number)
- and these operations behave "nicely" (coherently)

Why exactly these axioms:

This is the minimal set of rules under which the intuition of "arrows that can be added and stretched" works.

Examples of vector spaces

SPACE	ELEMENTS	DIMENSION	COMMENT
$\mathbb{R}^n$	$(x_1, x_2, ..., x_n), x_i \in \mathbb{R}$	n	Standard example
$C([a,b])$	Continuous functions $f: \mathbb{R}\to\mathbb{R}$ $(f+g)(x) = f(x)+g(x)$	∞	Functions – also vectors
$\mathbb{R}[x]\leq n$	Polynomials of degree $\leq n$ $a_0 + a_1x + ... + a_nx^n$	$n + 1$	Basis: $1, x, x^2,...$
$M_{m\times n}(\mathbb{R})$	Matrices of size $m\times n$	$m\cdot n$	Basis: E_{ij}
$\mathbb{C}^n$	$(z_1, ..., z_n), z_i \in \mathbb{C}$	n (over $\mathbb{C}$)	Or $2n$ over $\mathbb{R}$

Counterexample: What is not a vector space

$\{(x,y) \in \mathbb{R}^2 : x \geq 0\}$ (right half-plane)	$(1,0) \in V$, but $(-1)\cdot(1,0) = (-1,0) \notin V$ No closure under ×scalar.

Linear independence and basis

DEFINITION: LINEAR COMBINATION
A linear combination of vectors $v_1, ..., v_k$ is an expression $\alpha_1v_1 + \alpha_2v_2 + ... + \alpha_kv_k$, where $\alpha_i \in F$ (scalars)

DEFINITION: LINEAR INDEPENDENCE

Vectors $v_1, \ldots, v_k$ are linearly independent if from the equality

$$\alpha_1 v_1 + \alpha_2 v_2 + \ldots + \alpha_k v_k = 0$$

it follows that $\alpha_1 = \alpha_2 = \ldots = \alpha_k = 0$.

In other words: the only way to obtain 0 is to take all $\alpha_i = 0$.

Geometric meaning and examples:

STATEMENT	WHY
2 vectors LI	$\Longleftrightarrow$ not on one line
3 vectors LI	$\Longleftrightarrow$ not in one plane
LD = "redundant"	One is expressed via others

EXAMPLE IN $\mathbb{R}^3$	LI OR LD?
$e_1=(1,0,0)$, $e_2=(0,1,0)$, $e_3=(0,0,1)$	LI: $\alpha_1 e_1+\alpha_2 e_2+\alpha_3 e_3=0 \Rightarrow$ all $\alpha_i=0$
$v_1=(1,0,0)$, $v_2=(0,1,0)$, $v_3=(1,1,0)$	LD: $v_1+v_2-v_3=0$ (found nonzero)

DEFINITION: BASIS

A basis of space V is a set of vectors $\{e_1, \ldots, e_n\}$, which:

(1) Is linearly independent
(2) Spans V (any $v \in V$ is expressed as $v = \Sigma\, \alpha_i e_i$)

Equivalently: a basis is a minimal spanning system,
or a maximal linearly independent system.

THEOREM ON DIMENSION

All bases of one space contain the same number of vectors.
This number is called the dimension of the space: dim(V).

Examples of dimensions:

- $\dim(\mathbb{R}^n) = n$
- $\dim(M_{m\times n}) = m\cdot n$
- $\dim(\mathbb{R}[x]_{\le n}) = n + 1$
- $\dim(\{\text{continuous functions}\}) = \infty$

Two types of bases in infinite-dimensional spaces

TYPE	DEFINITION	EXAMPLE
Hamel basis (algebraic)	$v = \Sigma\, \alpha_i e_i$ (finite sum) Each vector – finite linear combination	For C[0,1] such basis is uncountable. Cannot explicitly construct.
Schauder basis (topological)	$v = \Sigma\, \alpha_i e_i$ (infinite convergent series) Topology is required.	$\{1, x, x^2, ...\}$ for analytic functions $\{\sin nx, \cos nx\}$ – L^2

Why this is important:

- In algebra by default Hamel basis
- In Fourier series and functional analysis – Schauder basis
- $\dim_{Hamel}(L^2)$ = uncountable, $\dim_{Schauder}(L^2)$ = countable (if separable)

One and the same space has different "dimensions" depending on which type of basis is used.

Linear mapping – preserving structure

DEFINITION: LINEAR MAPPING

A mapping $T: V \to W$ is called linear if:

(L1) $T(u + v) = T(u) + T(v)$ (preserves addition)

(L2) $T(\alpha v) = \alpha T(v)$ (preserves scalar multiplication)

Equivalently (in one formula):

$T(\alpha u + \beta v) = \alpha T(u) + \beta T(v)$ (preserves linear combinations)

Consequences from the definition:

PROPERTY	WHY
$T(0) = 0$	$T(0\cdot v) = 0\cdot T(v) = 0$
$T(-v) = -T(v)$	$T((-1)\cdot v) = (-1)\cdot T(v)$
$T(\Sigma\alpha_i v_i) = \Sigma\alpha_i T(v_i)$	Multiple application of L1+L2

Key fact: T is completely determined by values on a basis.
If we know $T(e_1),...,T(e_n)$, then $T(v) = T(\Sigma\alpha_i e_i) = \Sigma\alpha_i T(e_i)$

Examples of linear mappings:

MAPPING	FORMULA	WHERE USED
Rotation by θ	$(x,y) \mapsto (x\cdot\cos\theta-y\cdot\sin\theta,$ $x\cdot\sin\theta+y\cdot\cos\theta)$	Geometry, physics
Projection on x-axis	$(x,y) \mapsto (x, 0)$	Shadow, components
Differentiation	$p \mapsto p'$ for $p \in \mathbb{R}[x]$	Analysis
Integration	$p \mapsto \int p$	Analysis
Multiplication by A	$v \mapsto Av$	Systems of equations

Counterexample: $T(x,y)=(x+1, y)$ — not linear. $T(0,0)=(1,0)\neq(0,0)$

Matrix — linear mapping in coordinates

Main mistake of engineers: a matrix is not a tensor.

A matrix is a table of numbers, a record of coordinates of something in a basis.
One and the same 3×3 matrix can represent:

- Linear operator $T: V \to V$ (tensor of type (1,1))
- Bilinear form $B: V\times V \to \mathbb{R}$ (tensor of type (0,2))
- Quadratic form $Q(v)$ (tensor of type (0,2))
- Just a set of coefficients (not a tensor at all)

without specifying what it is — a matrix is meaningless under basis change.
Different objects transform according to different laws:

$A' = P^{-1}AP$ (operator) vs $A' = P^{T}AP$ (form)

Confusion here is the source of half the errors in mechanics and physics.

Key idea: Matrix = record of T in chosen bases
Changed basis → matrix changed, but T is the same.

Construction: $T: V \to W$, basis $\{e_j\}$ in V, basis $\{f_i\}$ in W
$T(e_j) = \Sigma_i\, a_{ij} f_i$ → j-th column of A = coordinates of $T(e_j)$ in $\{f_i\}$

Example: Rotation by 90° in $\mathbb{R}^2$

BASIS VECTOR	IMAGE	MATRIX COLUMN
$e_1 = (1,0)$	$T(e_1) = (0,1)$	$(0, 1)^T$ – 1st column
$e_2 = (0,1)$	$T(e_2) = (-1,0)$	$(-1, 0)^T$ – 2nd column

$$\text{Matrix: } A = \begin{vmatrix} 0 & -1 \\ 1 & 0 \end{vmatrix} \quad \text{Check: } A\begin{vmatrix} 1 \\ 0 \end{vmatrix} = \begin{vmatrix} 0 \\ 1 \end{vmatrix} = T(1,0)\ ✓$$

THEOREM: $T \leftrightarrow A,\ S \leftrightarrow B \ \Rightarrow\ S\circ T \leftrightarrow BA$ (order reversed)

Isomorphism: Choice of bases gives bijection

$$\mathrm{Hom}(V, W) \cong M_{m\times n}(F)$$

where $\mathrm{Hom}(V, W)$ – space of linear mappings $V \to W$,
$M_{m\times n}(F)$ – space of matrices m×n over field F,
$m = \dim(W)$, $n = \dim(V)$.

Important: Isomorphism depends on choice of bases. Different bases → different matrices for one operator. Connection: $A' = P^{-1}AP$ (basis change).

Critical warning (common mistake)

The law $A' = P^{-1}AP$ holds for a linear operator (tensor of type (1,1)).

For a quadratic/bilinear form (tensor of type (0,2)) the law is different:

$$A' = P^TAP \quad (\text{not } P^{-1}, \text{ but } P^T.)$$

OBJECT	TRANSFORMATION LAW
Linear operator $T: V \to V$	$A' = P^{-1}AP$ (matrix similarity)
Bilinear form $B: V \times V \to \mathbb{R}$	$A' = P^TAP$ (matrix congruence)
Metric g_{ij} (tensor (0,2))	$g'_{ij} = (\partial x^k/\partial x'^i)(\partial x^l/\partial x'^j)g_{kl}$ (two lower indices)

Where this is critical: When computing eigenvalues of a quadratic form (moment of inertia, stress tensor) use P^TAP, not $P^{-1}AP$.

Types of tensors

Physical meaning: matrix connects different quantities

Matrix-vector multiplication $w = Av$ has two fundamentally different meanings:

Meaning 1: active transformation — the vector actually changes

Was vector v, became different vector w = Av
Physically: rotated, stretched, deformed the object itself

Examples:
- Rotated a rod by 30° (rotation matrix)
- Stretched a spring (deformation matrix)
- Changed velocity in a collision

Meaning 2: passive transformation — vector is the same, coordinates change

Was vector v with coordinates (3, 4) in old basis
Same vector has coordinates (1.5, 4) in new basis
Physically: nothing changed, just recalculated the numbers

Examples:
- Measured length in meters, then in feet
- Switched from Cartesian coordinates to polar
- Changed reference frame (from train → from ground)

Meaning 3: matrix as physical law — connection between different quantities

This is the main application in physics and engineering.
Matrix describes how one vector quantity generates another.

```
+--------------------------------------------------------------------+
| THERMAL CONDUCTIVITY MATRIX  q = −λ·∇T                              |
```

```
+-------------------------------------------------------------------+
|                                                                   |
| ∇T – temperature gradient (vector, K/m): "where and how fast      |
| temperature grows"                                                |
| q  – heat flux (vector, W/m²): "where and how much heat flows"    |
| λ  – thermal conductivity matrix                                  |
| (later we'll learn this is a "rank-2 tensor")                     |
|                                                                   |
| In isotropic material (metal, water):                             |
| λᵢⱼ = λ·δᵢⱼ  (scalar × identity matrix)                           |
| Heat flows along gradient (in direction of decreasing T)          |
|                                                                   |
| In anisotropic material (wood, crystal, layered insulation):      |
|     ⎛λₓₓ  λₓγ  λₓᵤ⎞                                                |
| λ = ⎜λγₓ  λγγ  λγᵤ⎟                                                |
|     ⎝λᵤₓ  λᵤγ  λᵤᵤ⎠                                                |
|                                                                   |
| Heat can flow not along gradient.                                 |
| Heat from one side – heat goes sideways.                          |
|                                                                   |
+-------------------------------------------------------------------+

+---------------------------------------------------------------+
| STRESS MATRIX  σ                                              |
+---------------------------------------------------------------+
|                                                               |
| n – normal to surface (vector, direction)                     |
| F = σ·n – force per unit area (vector, N/m²)                  |
| σ – stress matrix 3×3                                         |
|                                                               |
|                                                               |
| σ says: "if surface is oriented like this (n),                |
| then force on it will be like this (F)"                       |
|                                                               |
| This is not one force – this is rule for all possible surfaces.|
| σ contains complete information about stress state at point.  |
|                                                               |
```

```
+----------------------------------------------------------------+

+-------------------------------------------------------------------+
| INERTIA MATRIX  I                                                 |
+-------------------------------------------------------------------+
|                                                                   |
| ω — angular velocity (vector): "around which axis and how fast"  |
| L = I·ω — angular momentum (vector)                               |
| I — inertia matrix 3×3                                            |
|                                                                   |
|                                                                   |
| For symmetric body (sphere, cube):                                |
| L ∥ ω  (momentum parallel to angular velocity)                   |
|                                                                   |
| For asymmetric body (dumbbell at angle):                          |
| L and ω not parallel.                                             |
| Rotate around one axis — momentum directed in other direction.   |
| (This is why wheels are balanced — removes off-diagonal elements I)|
|                                                                   |
+-------------------------------------------------------------------+

+-----------------------------------------------------------------+
| ABOUT VECTOR FIELD                                              |
+-----------------------------------------------------------------+
|                                                                 |
| If there is a field of vectors v(x,y,z) (e.g., wind velocity):  |
|                                                                 |
| • one matrix A for all points: global transformation            |
| A = rotation → entire field rotated                             |
| A = stretching → entire field stretched                         |
| Physically: changed coordinate system                           |
|                                                                 |
| • different matrix A(x,y,z) at each point: local deformation    |
| This is a field of matrices (later we'll learn: "tensor field")  |
| Example: strain tensor in material under load                   |
| At each point its own stretching/shear                          |
|                                                                 |
+-----------------------------------------------------------------+

Bottom line: Matrix is not just "table of numbers", but:
  • Either transformation (active or passive)
  • Or physical law connecting cause and effect
```

```
Kernel and image – main subspaces

+---------------------------------------------------------------+
| DEFINITIONS                                                   |
+---------------------------------------------------------------+
|                                                               |
| For linear T: V → W:                                          |
|                                                               |
| kernel:  ker(T) = {v ∈ V : T(v) = 0}  (what maps to zero)     |
| image:   Im(T) = {T(v) : v ∈ V}       (where we land)         |
|                                                               |
| Both are subspaces (closed under + and ·).                    |
|                                                               |
+---------------------------------------------------------------+

+--------------------------------------------+
| DIMENSION THEOREM (Rank-Nullity)           |
+--------------------------------------------+
|                                            |
| dim(V) = dim(ker T) + dim(Im T)            |
| ↑              ↑                           |
| nullity         rank                       |
| (defect)         (rank)                    |
|                                            |
| "Dimension of domain = losses + result"    |
|                                            |
+--------------------------------------------+

Example: Projection T: ℝ³ → ℝ³, T(x,y,z) = (x,y,0)

  ker(T) = {(0,0,z) : z ∈ ℝ} = z-axis         dim(ker T) = 1
  Im(T) = {(x,y,0) : x,y ∈ ℝ} = xy-plane      dim(Im T) = 2

  Check: 3 = 1 + 2 ✓

Geometric intuition:

  V (3-dimensional)               W (target space)
  +-------------------+
  |      /\           |
  |     /  \ ker(T)   |
  |    /    \ = z-axis|         +-----------------+
  |   /      \        |    T      |                 |
  |  / plane  \       | --------▶ |    Im(T)        |
  | /   xy     \      |           |  = plane        |
  |/------------\     |           |                 |
  +-------------------+           +-----------------+

  dim(V) = 3
  dim(ker T) = 1  (what "collapses" to 0)
  dim(Im T) = 2   (what "survives")

  Theorem: 3 = 1 + 2  ✓
```

Meaning: Space V "splits" into two parts:

- ker T (what is lost)
- complement (what maps to image 1-1)

Connection to systems of equations:

CONCEPT	WHAT IT MEANS
ker(A)	Solutions of Ax = 0 (homogeneous system)
Im(A)	Linear span of columns of A
rank(A)	Number of nonzero rows in row echelon form

PROPERTY OF T	CONDITION
Injective (1-1)	ker(T) = {0}
Surjective (onto)	Im(T) = W
Bijective (isomorph.)	ker={0} and Im=W ⟺ dim V = dim W, det ≠ 0

Kernel and image – universal structure

The structure (ker, Im) is present for any morphism between structures:

BRANCH	OPERATOR	KER	IM
Lin. algebra	T: V → W	ker T ⊂ V	Im T ⊂ W
Group theory	φ: G → H	ker φ ⊲ G	Im φ ≤ H
Diff. forms	$d: \Omega^k \to \Omega^{k+1}$	ker d (closed)	Im d (exact)
XXI Diff. eq.	L: C∞ → C∞	ker L (solutions)	Im L (reachable)
QM (physics)	Ĥ−E	ker(Ĥ−E)=state	spectrum={E: ker≠0}

Dimension theorem: dim(V) = dim(ker T) + dim(Im T)

Cohomology: $H^k = \ker(d_k) / \mathrm{Im}(d_{k-1})$

Eigenvalues and eigenvectors

DEFINITION

Nonzero vector v is called eigenvector for T (or A), if

$T(v) = \lambda v$ (or $Av = \lambda v$)

Number λ is called eigenvalue.

Geometric meaning of eigenvalues:

VALUE λ	GEOM. ACTION	EXAMPLE
$\lambda > 1$	Stretching	λ=2: elongation by factor of 2
$0 < \lambda < 1$	Compression	λ=0.5: compression by half
$\lambda = 1$	Unchanged	Direction preserved
$\lambda = 0$	Collapse (to 0)	Projection destroys this dir.
$\lambda < 0$	Reflection + scale	λ=-1: pure reflection
$\lvert\lambda\rvert = 1$	Isometry/reflection	Preserves length

How to find:

$$Av = \lambda v \iff (A - \lambda I)v = 0 \iff v \in \ker(A - \lambda I)$$

For $\ker \neq \{0\}$, need $\det(A - \lambda I) = 0$.

Characteristic polynomial: $p(\lambda) = \det(A - \lambda I)$
Eigenvalues = roots of $p(\lambda)$.

Example: $A = \begin{vmatrix} 3 & 1 \\ 0 & 2 \end{vmatrix}$

$$\det(A - \lambda I) = \det\begin{vmatrix} 3-\lambda & 1 \\ 0 & 2-\lambda \end{vmatrix} = (3-\lambda)(2-\lambda) - 0 = \lambda^2 - 5\lambda + 6$$

Roots: $\lambda_1 = 3$, $\lambda_2 = 2$

For $\lambda_1 = 3$: $(A-3I)v = 0 \to \begin{vmatrix} 0 & 1 \\ 0 & -1 \end{vmatrix}\begin{vmatrix} v_1 \\ v_2 \end{vmatrix} = 0 \to v_2 = 0 \to v_1 = (1,0)$

For $\lambda_2 = 2$: $(A-2I)v = 0 \to \begin{vmatrix} 1 & 1 \\ 0 & 0 \end{vmatrix}\begin{vmatrix} v_1 \\ v_2 \end{vmatrix} = 0 \to v_1 = -v_2 \to v_2 = (-1,1)$

Invariants (independent of basis):

$\det(A) = \lambda_1 \cdot \lambda_2 \cdot \ldots \cdot \lambda_n$ (product of eigenvalues)
$\mathrm{tr}(A) = \lambda_1 + \lambda_2 + \ldots + \lambda_n$ (sum of eigenvalues)

Diagonalization — choosing the "right" basis

Idea:
In a basis of eigenvectors, the matrix becomes diagonal.

DIAGONALIZATION THEOREM

Matrix A is diagonalizable $\iff$ there exists a basis of eigenvectors.

Then $A = PDP^{-1}$, where:
- $P = [v_1 \mid v_2 \mid \dots \mid v_n]$ — matrix of eigenvectors
- $D = \mathrm{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ — diagonal of eigenvalues

Why is this needed:

PROBLEM	HOW IT HELPS
Matrix powers A^n	PD^nP^{-1}, where $D^n = \mathrm{diag}(\lambda_1^n,\dots,\lambda_n^n)$
Differential equations	e^{At} is computed via diagonal
Stability analysis	Behavior as $t\to\infty$ depends on $\lvert\lambda_i\rvert$

Diagonalizability conditions:

SITUATION	DIAGONALIZABLE?
All λ_i distinct	YES (sufficient, not necessary)
Multiple λ, but enough eigenvectors	YES (example: I, where $\lambda=1$ mult. n)
Not enough eigenvect.	no → Jordan form
Example: $A = [0\ 1;\ 0\ 0]$	NO ($\lambda=0$ mult. 2, only 1 vector)

Connection with groups

Invertible matrices form a group $GL(n)$ — general linear group

SUBGROUP	CONDITION	WHAT IT PRESERVES

GL(n)	$\det(A) \neq 0$	Linear structure
SL(n)	$\det(A) = 1$	+ volume
O(n)	$A^T A = I$	+ lengths and angles
SO(n)	$A^T A = I$, det=1	+ orientation
U(n)	$A^* A = I$ (complex)	Hermitian inner product
SU(n)	$A^* A = I$, det=1	+ volume (quantum mechanics)

$\det: GL(n) \to \mathbb{R}^*$ – this is a group homomorphism

$\ker(\det) = SL(n)$ – kernel of the homomorphism

Connection with Lie groups:

GL(n), O(n), SO(n), U(n), SU(n) – these are all Lie groups (simultaneously groups and manifolds)

Matrix exponential – key to systems of differential equations

Definition:

$$\boxed{\begin{array}{l} e^A = I + A + A^2/2! + A^3/3! + \ldots = \Sigma A^n/n! \\ \text{(series always converges for any matrix A)} \end{array}}$$

Why is it needed:

System of linear DEs $\dot{x} = Ax$ has solution $x(t) = e^{At} \cdot x(0)$

Computation (if A is diagonalizable):

$$A = PDP^{-1}, \text{ where } D = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$$

$$e^A = P \cdot \mathrm{diag}(e^{\lambda_1}, e^{\lambda_2}, \ldots, e^{\lambda_n}) \cdot P^{-1}$$

Properties:

- $e^0 = I$
- $e^{A+B} = e^A \cdot e^B$ only if AB = BA.
- $(e^A)^{-1} = e^{-A}$
- $d/dt\, e^{At} = A \cdot e^{At}$
- $\det(e^A) = e^{\mathrm{tr}(A)}$

Example for thermophysics:

System of n rooms with temperatures $T_1, T_2, \ldots, T_n$:

$\dot{T} = A \cdot T$, where A – matrix of thermal conductivities

Solution: $T(t) = e^{At} \cdot T(0)$

Eigenvalues of A (all negative) → decay rates

Eigenvectors of A → "modes" of the system (which rooms heat together)

Connection with Lie groups:
e^A maps Lie algebra gl(n) into Lie group GL(n)
This is the exponential map — foundation of Lie group theory

One object — three perspectives

ALGEBRA	GEOMETRY	ANALYSIS
Matrix A (table of numbers)	Transformation of space	System of linear equations $Ax = b$
det(A) (number)	Coefficient of volume change + sign of orient.	Condition for solvability ($\det\neq 0 \iff \exists!$ solution)
Eigenvector v $Av = \lambda v$	Invariant direction (does not rotate)	Solution of $(A-\lambda I)v = 0$
Eigenvalue λ	Stretching coefficient along v	Root of charact. polynomial $\det(A-\lambda I)=0$
Diagonalization $A = PDP^{-1}$	Choice of "right" basis (of eigen-vectors)	Separation of variables

---

Formulas for calculations

Determinant 2×2:
$$\det\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

Determinant 3×3: Expansion along the first row:

$$\det\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = a\cdot\det\begin{vmatrix} e & f \\ h & i \end{vmatrix} - b\cdot\det\begin{vmatrix} d & f \\ g & i \end{vmatrix} + c\cdot\det\begin{vmatrix} d & e \\ g & h \end{vmatrix}$$

Inverse 2×2:
$$\begin{vmatrix} a & b \\ c & d \end{vmatrix}^{-1} = \frac{1}{ad-bc}\begin{vmatrix} d & -b \\ -c & a \end{vmatrix}$$

Gaussian method: Reduction to row echelon form by elementary row operations

Eigenvalues: solve $\det(A - \lambda I) = 0$

Eigenvectors: for each λ solve $(A - \lambda I)v = 0$

Diagonalization: $A = PDP^{-1}$, where $P = [v_1|v_2|\dots|v_n]$ from eigenvectors

SVD — singular value decomposition (king of decompositions)

Problem: Diagonalization $A = PDP^{-1}$ works only for square matrices,
and not even for all of them (need a full set of eigenvectors).

Solution: SVD works for any m×n matrix!

```
+-----------------------------------------------------------------------+
| THEOREM (SVD): Any matrix A ∈ ℝᵐˣⁿ decomposes as                      |
|                                                                       |
| A = UΣVᵀ                                                              |
|                                                                       |
| where:                                                                |
| • U ∈ ℝᵐˣᵐ — orthogonal (UᵀU = I), columns = left sing. vectors      |
| • Σ ∈ ℝᵐˣⁿ — diagonal, σ₁ ≥ σ₂ ≥ ... ≥ 0 (singular values)           |
| • V ∈ ℝⁿˣⁿ — orthogonal (VᵀV = I), columns = right sing. vectors     |
+-----------------------------------------------------------------------+
```

Geometric meaning:
Any linear transformation = rotation U × stretching Σ × rotation V^T

```
    +---+   Vᵀ    +---+    Σ     /\      U     ◇
    |   |  ---→   |   |  ---→   /  \   ---→   / \
    +---+ rotation +---+ stretch \  / rotation \ /
    square        square          \/            ◇
                                ellipse      ellipse (rotated)
```

Connection with eigenvalues:

- σ_i^2 = eigenvalues of A^TA (or AA^T)
- Columns of V = eigenvectors of A^TA
- Columns of U = eigenvectors of AA^T

Applications (why SVD is the "king"):

PROBLEM	HOW TO USE SVD
Data compression (images)	$A \approx \Sigma_r \sigma_i u_i v_i^T$ (keep first r terms) Error = σ_{r+1} (Eckart–Young theorem)
PCA (data analysis)	Principal components = right sing. vectors V (for centered data)
Pseudoinversion (least squares)	$A^+ = V\Sigma^+U^T$, where $\Sigma^+_{ii} = 1/\sigma_i$ (for $\sigma_i \neq 0$) Solution of $Ax \approx b$: $x = A^+b$
Matrix rank	rank(A) = number of nonzero σ_i
Matrix norm	$\|A\|_2 = \sigma_1$ (maximum singular value) $\|A\|\_F = \sqrt{(\Sigma\sigma_i^2)}$ (Frobenius norm)
Condition number	$\text{cond}(A) = \sigma_1/\sigma_n$ (sensitivity to errors)

Comparison with diagonalization:

- Diagonalization: $A = PDP^{-1}$ (only square, doesn't always exist)
- SVD: $A = U\Sigma V^T$ (any matrices, always exists)

Special case: symmetric positive definite

For $A = A^T > 0$: SVD coincides with diagonalization.

- $U = V = P$ (matrix of eigenvectors)
- $\sigma_i = \lambda_i$ (singular values = eigenvalues)
- $A = P\Lambda P^T = U\Sigma V^T$ — this is the same thing.

For symmetric (not necessarily pos. def.): $\sigma_i = |\lambda_i|$

Applications

FIELD	HOW LINEAR ALGEBRA IS USED
diff. equations (stability)	$y' = Ay$ → solution $y(t) = e^{At}y_0$ $\mathrm{Re}(\lambda)>0$: growth, $\mathrm{Re}(\lambda)<0$: decay, $\mathrm{Re}(\lambda)=0$: oscill.
quantum mechanics	Observables = Hermitian operators ($A = A^*$) Eigenvalues = measurement results Eigenvectors = states
data analysis (PCA)	Covariance matrix $C = (1/n)X^TX$ Eigenvectors = principal components Eigenvalues = variance along directions
graphs and networks	PageRank = principal eigenvector Spectral clustering = e.v. of Laplacian
mechanics	Natural frequencies of oscillations = $\sqrt{\lambda}$ Modes of oscillations = eigenvectors

---

Dual space V* (brief introduction)

Definition:

$V^* = \{\varphi: V \to \mathbb{R} \mid \varphi \text{ linear}\}$ – set of all linear functions on V

Example: $V = \mathbb{R}^3$

$\varphi(x, y, z) = 2x + 3y - z$ – element of V* (linear function)

Action: $\varphi(1, 0, 2) = 2\cdot1 + 3\cdot0 - 1\cdot2 = 0$

Facts (for finite-dimensional V):

- $\dim(V^*) = \dim(V)$ Only for finite-dimensional.
- Basis of V*: functions $e^i(e_j) = \delta^i{}_j$ (1 if i=j, otherwise 0)
- Element of V* is called a covector or linear functional

For infinite-dimensional: $\dim(V^*)$ can be greater than $\dim(V)$.

Example: $V = \ell^1$ (absolutely summable: $\Sigma|x_n|<\infty$), $V^* = \ell\infty$ (bounded sequences)

Why:

- Tensors live on $V^* \times ... \times V^* \times V \times ... \times V$
- Differential df — element of cotangent space T^*_pM
- In detail (Duality)

Connection with other sections

Groups:
GL(n), O(n), SO(n) — matrix groups
Representation theory: group → matrices

Manifolds:
T_pM — tangent space = vector space
Locally linear algebra works.

Duality:
V^* — dual space (linear functionals on V)
Matrix rows ↔ columns of transpose

Tensors:
Tensor of rank 2 = linear map = matrix (in basis)

Series:
Functions = vectors of infinite-dimensional space
Fourier coefficients = coordinates in orthonormal basis

Applied example: thermal balance of heating system

Problem: Building with 4 rooms. Heat losses and heat transfer between rooms are known. Find temperatures in steady state.

```
       T₁ ======+          Heat losses to outside: αᵢ·(Tᵢ − Tₒᵤₜ)
       |        |          Heat exchange between: kᵢⱼ·(Tᵢ − Tⱼ)
   Qₕₑₐₜ₁  T₂ ==+== T₃
       |        |          Heat from heating: Qᵢ
       T₄ ======+
```

Balance equations (for each room):

$$Q_i = \alpha_i(T_i - T_{out}) + \Sigma_j\, k_{ij}(T_i - T_j)$$

"Supply = losses to outside + heat exchange with neighbors"

This is a linear system. Write in matrix form:

$$\begin{pmatrix} \alpha_1+k_{12}+k_{14} & -k_{12} & 0 & -k_{14} \\ -k_{12} & \alpha_2+k_{12}+k_{23} & -k_{23} & 0 \\ 0 & -k_{23} & \alpha_3+k_{23}+k_{34} & -k_{34} \\ -k_{14} & 0 & -k_{34} & \alpha_4+k_{14}+k_{34} \end{pmatrix} \cdot \begin{pmatrix} T_1 \\ T_2 \\ T_3 \\ T_4 \end{pmatrix} = \begin{pmatrix} Q_1 + \alpha_1 \cdot T_{out} \\ Q_2 + \alpha_2 \cdot T_{out} \\ Q_3 + \alpha_3 \cdot T_{out} \\ Q_4 + \alpha_4 \cdot T_{out} \end{pmatrix}$$

$$K \cdot T = Q$$

Numerical example:

$\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 100$ W/K (heat losses)
$k_{12} = k_{23} = k_{34} = k_{14} = 50$ W/K (heat exchange between rooms)
$T_{out} = 0°C$, $Q_1 = 3000$ W (heating only in room 1)

Matrix K:

$$\begin{bmatrix} 200 & -50 & 0 & -50 \\ -50 & 200 & -50 & 0 \\ 0 & -50 & 200 & -50 \\ -50 & 0 & -50 & 200 \end{bmatrix}$$

Solution: $T = K^{-1} \cdot Q$
$T_1 = 20°C$, $T_2 = 10°C$, $T_3 = 5°C$, $T_4 = 10°C$

Eigenvalues – relaxation modes

If we turn off heating, temperatures will drop.
How exactly? Solution: $T(t) = \Sigma\, c_i \cdot v_i \cdot e\text{^}(-\lambda_i t)$

λ_i = eigenvalues of K (relaxation rates)
v_i = eigenvectors of K (modes – like "shape" of cooling)

- Smallest λ_1 → slowest mode (entire building together)
- Largest λ_n → fastest mode (equalization between rooms)

Moral: Linear algebra is the language of thermal calculations.
Thermal conductivity matrix + eigenvalues = complete understanding of system.

We learned to add vectors and multiply them by numbers. But how to multiply vectors by each other? It turns out there is no single answer – different problems require different products. This is the key to understanding tensors.

Products of vectors – different questions, different answers

Why so many types of multiplication

Numbers: 2 × 3 = 6. One way of multiplication.

Vectors: What do we want to know about two vectors?

- How codirectional are they? → Scalar · → number
- What is perpendicular to both? → Cross × → vector
- All combinations of components? → Tensor ⊗ → matrix
- What area do they span? → Outer ∧ → bivector

Different questions require different operations.

PRODUCT	RESULT	QUESTION / GEOMETRIC MEANING
$u \cdot v$	number	"How codirectional are u and v?" = $\lvert u\rvert\lvert v\rvert\cos\theta$
$u \times v$	vector	"What is ⊥ to both?" (only in $\mathbb{R}^3$ and $\mathbb{R}^7$.)
$u \otimes v$	matrix	"All combinations of components?" $(u\otimes v)_{ij} = u_i v_j$
$u \wedge v$	bivector	"What area do they span?" (any dimension.)

Main distinction: inner vs outer product

	INNER (·, contraction)	OUTER (⊗, ∧)
Rank change	DECREASES vector×vector → scalar	INCREASES vector×vector → matrix
Meaning	CODIRECTIONALITY "how parallel"	NEW OBJECT in space of higher dimension
Formula	$a_i b^i$ (contraction on index)	$(a\otimes b)_{ij} = a_i b_j$

Remember: Inner – "contracts", outer – "expands".

Why does the vector product exist only in $\mathbb{R}^3$ and $\mathbb{R}^7$?

This is a deep result connected with division algebras over $\mathbb{R}$:

dim 1: $\mathbb{R}$ (real numbers)
dim 2: $\mathbb{C}$ (complex numbers)
dim 4: $\mathbb{H}$ (quaternions) – non-commutative
dim 8: $\mathbb{O}$ (octonions) – non-associative

and that's all. No other division algebras over $\mathbb{R}$ exist (Hurwitz's theorem).

What comes next? (Cayley–Dickson doubling)

Dim 16: $\mathbb{S}$ (sedenions) – have zero divisors.
$ab = 0$, but $a \neq 0$ and $b \neq 0$
Also lose alternativity

One can continue doubling infinitely (dim 32, 64, ...), but each time more and more properties are lost. Division algebras end at $\mathbb{O}$.

Hierarchy of losses:
$\mathbb{R} \to \mathbb{C}$: lost orderability
$\mathbb{C} \to \mathbb{H}$: lost commutativity ($ab \neq ba$)
$\mathbb{H} \to \mathbb{O}$: lost associativity ($(ab)c \neq a(bc)$)
$\mathbb{O} \to \mathbb{S}$: lost absence of zero divisors ($ab = 0$, $a,b \neq 0$)

Connection with vector product:

In $\mathbb{R}^3$: $u \times v$ is related to quaternions $\mathbb{H}$
(multiplication of imaginary parts of quaternions)

In $\mathbb{R}^7$: $u \times v$ is related to octonions $\mathbb{O}$
(multiplication of imaginary parts of octonions)

Difference: In $\mathbb{R}^7$ the vector product does not satisfy the Jacobi identity:

$$u \times (v \times w) + v \times (w \times u) + w \times (u \times v) \neq 0$$

Consequence: $(\mathbb{R}^7, \times)$ is not a Lie algebra, unlike $(\mathbb{R}^3, \times) = so(3)$

This emphasizes the uniqueness of dimensions 1, 2, 3, 4, 7, 8.

Scalar product — formal definition

DEFINITION: SCALAR PRODUCT

A scalar product on a vector space V (over $\mathbb{R}$) is a function $\langle\cdot,\cdot\rangle: V \times V \to \mathbb{R}$, satisfying:

(S1) $\langle u, v\rangle = \langle v, u\rangle$ (symmetry)

(S2) $\langle \alpha u + \beta v, w\rangle = \alpha\langle u, w\rangle + \beta\langle v, w\rangle$ (linearity in 1st arg.)

(S3) $\langle v, v\rangle \geq 0$, and $\langle v, v\rangle = 0 \iff v = 0$ (positive definiteness)

A space V with a scalar product is called Euclidean (or pre-Hilbert).

Why precisely these axioms are needed:

AXIOM	GEOMETRIC MEANING
S1 symmetry	Angle between u and v = angle between v and u
S2 linearity	Projection of sum = sum of projections
S3 positive	Length ≥ 0, and = 0 only for zero vector

Examples of scalar products:

SPACE	FORMULA ⟨u, v⟩
Standard $\mathbb{R}^n$	$\Sigma_i u_i v_i = u_1 v_1 + ... + u_n v_n$
With weights	$\Sigma_i w_i u_i v_i$, where $w_i > 0$
Functions on [a,b]	$\int_a^b f(x)g(x)\,dx$

Counterexample: $\langle u,v\rangle = u_1 v_1 - u_2 v_2$ – not a scalar product.
Check: $\langle(0,1),(0,1)\rangle = -1 < 0$ ✗ (this is a pseudo-Euclidean metric from GR)

Chain: scalar product → norm → metric

Scalar product generates all geometry:

```
  Scalar product ⟨u, v⟩
            |
            | ‖v‖ = √⟨v, v⟩
            ↓
        Norm ‖·‖ (vector length)
            |
            | d(u, v) = ‖u − v‖
            ↓
       Metric d(·,·) (distance between points)
            |
            ↓
      Topology (notion of closeness, open sets)
```

Important: This chain goes in one direction only.

- Not every metric is generated by a norm (ex: discrete metric)
- Not every norm is generated by a scalar prod. (ex: $\|v\| = \max|v_i|$)

CAUCHY–BUNYAKOVSKY–SCHWARZ INEQUALITY

$$|\langle u, v\rangle| \le \|u\| \cdot \|v\|$$

Equality ⟺ u and v are proportional (lie on the same line).

consequence: $|\cos\theta| = |\langle u,v\rangle|/(\|u\|\|v\|) \le 1$ – angle is well-defined.

```
+---------------------------------------------------------------------------+

+--------------------------------+
| TRIANGLE INEQUALITY            |
+--------------------------------+
|                                |
| ‖u + v‖ ≤ ‖u‖ + ‖v‖            |
|                                |
| "Shortest path – along a line"|
|                                |
+--------------------------------+


Orthogonality – geometry from algebra

+------------------------------------------------------------------+
| DEFINITION: ORTHOGONALITY                                        |
+------------------------------------------------------------------+
|                                                                  |
| Vectors u and v are orthogonal (perpendicular) if ⟨u, v⟩ = 0.  |
| Notation: u ⊥ v                                                  |
|                                                                  |
+------------------------------------------------------------------+

Key facts:
+--------------------------+---------------------------------+
| CONCEPT                  | FORMULA / MEANING               |
+--------------------------+---------------------------------+
| Orthogonality u ⊥ v      | ⟨u,v⟩ = 0 ⟺ θ = 90°             |
| Orthonormal basis        | ⟨eᵢ,eⱼ⟩ = δᵢⱼ (0 or 1)         |
| Coordinate in ONB        | vᵢ = ⟨v,eᵢ⟩ (simply projection)|
+--------------------------+---------------------------------+

Gram–Schmidt process (any basis → orthonormal):
+-----+---------------------------------+--------------+
| STEP| ORTHOGONALIZATION               | NORMALIZATION|
+-----+---------------------------------+--------------+
| 1   | u₁ = v₁                         | e₁ = u₁/‖u₁‖ |
| 2   | u₂ = v₂ − ⟨v₂,e₁⟩e₁              | e₂ = u₂/‖u₂‖ |
| 3   | u₃ = v₃ − ⟨v₃,e₁⟩e₁ − ⟨v₃,e₂⟩e₂    | e₃ = u₃/‖u₃‖ |
| k   | uₖ = vₖ − Σᵢ₌₁ᵏ⁻¹⟨vₖ,eᵢ⟩eᵢ          | eₖ = uₖ/‖uₖ‖ |
+-----+---------------------------------+--------------+

-------------------------------------------------------------------------------
Scalar product – geometric interpretation
-------------------------------------------------------------------------------

Formula in ℝⁿ:  u · v = u₁v₁ + u₂v₂ + ... + uₙvₙ = Σᵢ uᵢvᵢ

Geometry: u · v = |u| × (length of projection of v onto u)

             v
            /|
```

```
          / |
         /  |
        /   |         u·v = |u| × |v|cos θ
       / θ  |
   ---●-----+----→ u
      +-----+
       projection = |v|cos θ
```

Physics: Work $W = F \cdot s$ (force × displacement in direction of force)

Connection with metric:
  Scalar product defines a metric — a way to measure.
    $|v|^2 = v \cdot v$ (length)
    $\cos \theta = (u \cdot v)/(|u||v|)$ (angle)

On a manifold:
  Metric tensor g_{ij} — this is the scalar product of basis vectors:
  $g_{ij} = \langle \partial/\partial x^i, \partial/\partial x^j \rangle$

Vector product $u \times v$ — only in $\mathbb{R}^3$.

Formula: $u \times v = (u_2v_3-u_3v_2, u_3v_1-u_1v_3, u_1v_2-u_2v_1)$

Mnemonic via determinant:

$$u \times v = \begin{vmatrix} i & j & k \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix} = i(u_2v_3-u_3v_2) - j(u_1v_3-u_3v_1) + k(u_1v_2-u_2v_1)$$

Properties:

PROPERTY	CONSEQUENCE
$u \times v \perp u$, $u \times v \perp v$	Result is perpendicular to both
$\|u \times v\| = \|u\|\|v\|\sin \theta$	= area of parallelogram
$u \times v = -(v \times u)$	Anticommutative.
$u \times u = 0$	Vector × itself = 0
$(u \times v) \times w \neq u \times (v \times w)$	Not associative.

Applications in physics:

QUANTITY	FORMULA
Torque	$\tau = r \times F$
Magnetic force	$F = qv \times B$
Linear velocity	$v = \omega \times r$

Why only in $\mathbb{R}^3$:

Result must be a vector of the same dimension.
In $\mathbb{R}^2$ there is no "perpendicular direction" (it would be in $\mathbb{R}^3$)
In $\mathbb{R}^n$ $(n>3)$ there are too many perpendicular directions ($n-2$ dimensions)

Mathematically: $\dim(\wedge^2\mathbb{R}^n) = n(n-1)/2 = n$ only when $n = 3$

Generalization: Exterior product ∧ works in any dimension.

Exterior product $u \wedge v$ — works everywhere

Idea: $u \wedge v$ = oriented area of parallelogram on u and v
Result: Bivector $\in \wedge^2 V$ (not a vector)

Properties:

PROPERTY	GEOMETRIC MEANING
$u \wedge v = -v \wedge u$	Orientation change with swap
$u \wedge u = 0$	Degenerate parallelogram
$(u+v) \wedge w = u\wedge w + v\wedge w$	Linearity (like area)
$(\alpha u) \wedge v = \alpha(u \wedge v)$	Scaling of area

In coordinates: $u \wedge v = \Sigma_{i<j} (u_i v_j - u_j v_i)\, e_i \wedge e_j$
Dimension: $\dim(\wedge^2\mathbb{R}^n) = C(n,2) = n(n-1)/2$

Connection with × in $\mathbb{R}^3$: Components coincide. But $u\times v \in \mathbb{R}^3$, while $u\wedge v \in \wedge^2\mathbb{R}^3$
Transformation: $*(u \wedge v) = u \times v$ (Hodge star)

Important for engineers: normal vs bivector

Engineers are used to: flux through surface = $\iint F\cdot n\, dS$
where n — unit normal to surface (vector sticking outward).

Diff. forms replace this with: $\iint \omega$ (2-form on surface)
2-form = bivector = "oriented area element of tangent plane"

In $\mathbb{R}^3$ normal and bivector — dual descriptions of the same thing.
Hodge star (*) converts one to the other: $*(dx\wedge dy) = dz$.

But: bivector is more honest, because it doesn't require "ambient space". On a surface in $\mathbb{R}^3$ one can indicate a normal (exit into 3D).
On an abstract manifold without embedding — one cannot. But a bivector one can.

Applications:

WHERE	HOW IT'S USED
Diff. forms	Integration on manifolds
Determinant	det = $e_1\wedge e_2\wedge\ldots\wedge e_n$ (n-dim volume)
Maxwell's equations	Elegant notation via forms
Any dimension	Works everywhere (unlike ×)

Higher degrees: $u\wedge v\wedge w \in \wedge^3V$ — oriented volume of parallelepiped

Tensor product $u \otimes v$ — all combinations

Idea: Write all possible products of components u_iv_j → matrix

Example: $u=(u_1,u_2)$, $v=(v_1,v_2)$ → $u\otimes v = \begin{vmatrix} u_1v_1 & u_1v_2 \\ u_2v_1 & u_2v_2 \end{vmatrix}$

Properties:

PROPERTY	CONSEQUENCE
Not commutative: $u\otimes v \neq v\otimes u$	$u\otimes v = (v\otimes u)^T$ (transposition)
Bilinear	$(\alpha u+\beta w)\otimes v = \alpha(u\otimes v)+\beta(w\otimes v)$
$\mathrm{rank}(u\otimes v) = 1$ (if $u,v \neq 0$)	Any rank-1 matrix = $u\otimes v$
Matrix rank r = $\Sigma\, u_i\otimes v_i$	This is SVD decomposition.

Applications:

WHERE	HOW
Tensors from vectors	General construction method
Metric on manifold	$g = g_{ij}\, dx^i\otimes dx^j$
Quantum mechanics	$\|\psi\rangle = \|\psi_1\rangle\otimes\|\psi_2\rangle$ (2 particles)
Entanglement	not every $\psi = \psi_1\otimes\psi_2$.

Summary table of products

PRODUCT	RESULT	DIMENSION	MAIN PROPERTY
$u \cdot v$ (scalar)	scalar	1	Symmetric: $u \cdot v = v \cdot u$ Generates metric
$u \times v$ (cross)	vector	3 (and 7)	Antisymm: $u \times v = -v \times u$ Perpendic. to both
$u \wedge v$ (exterior)	bivector	$n(n-1)/2$	Antisymm: $u \wedge v = -v \wedge u$ Generalization × to any n
$u \otimes v$ (tensor)	matrix	n^2	Not commut: $u \otimes v = (v \otimes u)^T$ All combinations $u_i v_j$

Connection with other sections

SECTION	HOW PRODUCTS WORK THERE
Manifolds (metric)	$\langle \cdot,\cdot \rangle \rightarrow$ metric $g_{ij} = \langle \partial/\partial x^i, \partial/\partial x^j \rangle$ How to measure distances on curved surfaces
Manifolds (integration)	$\wedge \rightarrow$ diff. forms on manifold $dx \wedge dy$ = area, $dx \wedge dy \wedge dz$ = volume
Tensors	$\otimes \rightarrow$ construction of tensors from vectors $g = g_{ij}\, dx^i \otimes dx^j$, energy-momentum tensor
Lin. algebra	$\langle \cdot,\cdot \rangle \rightarrow$ orthogonality, Gram–Schmidt Orthonormal basis simplifies computations
Fourier series	$\langle f,g \rangle = \int f(x)g(x)dx$ — scalar prod. of functions Fourier coeff. = projections on orthonorm. bas

Hierarchy of spaces:

Vector space → [+⟨,⟩] → Euclidean → [+completeness] → Hilbert

Applied example: three problems — three products

Problem 1: pump work (scalar product)

Pump creates force $F = (0, 0, -1000)$ N (downward)
Water moves by $d = (0, 0, 5)$ m (upward)

$$\text{Work} = F \cdot d = 0 \cdot 0 + 0 \cdot 0 + (-1000) \cdot 5 = -5000 \text{ J}$$

Minus means: force is directed against motion (pump does work).
Scalar product answers the question: "How much energy?"

Problem 2: torque on shaft (vector product)

Wrench handle: $r = (0.3, 0, 0)$ m (from axis)
Applied force: $F = (0, 100, 0)$ N (perpendicular)

$$\begin{aligned}\text{Torque} = r \times F &= (0 \cdot 0 - 0 \cdot 100,\ \ 0 \cdot 0 - 0.3 \cdot 0,\ \ 0.3 \cdot 100 - 0 \cdot 0) \\ &= (0, 0, 30) \text{ N} \cdot \text{m}\end{aligned}$$

Torque is directed along rotation axis (z axis).
Vector product answers the question: "Where and how strongly does it rotate?"

Problem 3: flow through cross-section (exterior product / area)

Rectangular pipe cross-section: edges $a = (0.5, 0, 0)$ m, $b = (0, 0.3, 0)$ m

$$\text{Cross-sectional area} = |a \times b| = |(0, 0, 0.15)| = 0.15 \text{ m}^2$$

Or via exterior product: $a \wedge b = 0.15 \cdot (dx \wedge dy)$
This is a 2-form — an object that "eats" a pair of vectors and yields area.

If flow velocity $v = (0, 0, 2)$ m/s:
Volume flow rate $= v \cdot S = 2 \cdot 0.15 = 0.3 \text{ m}^3/\text{s}$

Summary: which product for which problem

QUESTION	PRODUCT	EXAMPLE FROM THERMOPHYSICS
How much energy?	Scalar ·	Pump work, power
Where rotates?	Vector ×	Torque on shaft, flow vortex
What area?	Exterior ∧	Pipe cross-section, H/E surface
All combinations?	Tensor ⊗	Stress tensor, conductivity

We saw different products: scalar gives number, tensor gives matrix. But where does this asymmetry come from — number vs object? Answer: scalar product is pairing of vector and covector. What is a covector?

This leads to fundamental idea of duality: every space has a "shadow" — space of linear functions on it.

Duality — fundamental symmetry of mathematics

Duality as view on space

Every space V generates dual space V* — space of linear functions on V. This is like "shadow" of original space.

V = space of vectors (directions, displacements)
V* = space of covectors (gradients, "prices per unit")

Two spaces — but connected by deep symmetry.
This symmetry manifests everywhere: in physics, geometry, algebra.

Duality — one of the deepest ideas of mathematics. Before giving formal definitions, let's start with physical intuition.

Physical intuition: covectors as "prices"

Key to understanding — analogy with physical dimensions.

Basis vector = Unit of measurement

What is a basis? It is a set of "standards", units of measurement.

e_1 = "1 meter in x direction"
e_2 = "1 second"
e_3 = "1 kilogram"

Vector is quantity measured in these units:

$v = 5e_1 + 3e_2$ = "5 meters and 3 seconds"

Numbers (5, 3) are components of vector. Vector v itself is physical quantity that exists independently of choice of units.

Covector = "price per unit" (density)

Covector ω is neither number nor vector.
A covector is a function that takes a vector and yields a number.

Physical analogy: price per unit.

ω = "10 rubles per meter"

This is not number 10 — this is rule: "take length in meters, multiply by 10".

How covector acts on vector:

v = 5 meters
$\omega(v) = (10 \text{ rub/m}) \times (5 \text{ m}) = 50$ rubles ← this is already number.

Formally:

$$\omega: V \to \mathbb{R}, \quad \omega(v) = \omega_i v^i = \omega_1 v^1 + \omega_2 v^2 + \ldots$$

Other examples of "prices":
- Pressure = "force per unit area" [N/m^2]
- Density = "mass per unit volume" [kg/m^3]
- Temperature gradient = "change in T per unit length" [K/m]

All these quantities — covectors: they "eat" extensive quantity (area, volume, displacement) and yield number.

Why components transform oppositely

Let's switch from meters to centimeters: new unit $e'_1 = e_1/100$.

Vector "5 meters":
Old units: $v = 5 \text{ m} = 5\, e_1$
New units: $v = 500 \text{ cm} = 500\, e'_1$

Component increased by factor of 100.
(Unit became smaller → need more units)

Covector "10 rubles per meter":
Old units: $\omega = 10$ rub/m
New units: $\omega = 0.1$ rub/cm

Component decreased by factor of 100.
(Unit became smaller → price per unit also smaller)

Check — result does not depend on units:
$\omega(v) = 10$ rub/m × 5 m = 50 rub
$\omega(v) = 0.1$ rub/cm × 500 cm = 50 rub ✓

This is the key property:
$\omega(v)$ — invariant, does not depend on choice of units (basis).
For this to work, components of ω and v must change oppositely.

Tangent space — the correct picture

Consider paraboloid $z = f(x,y) = x^2 + y^2$.
At point $p = (1, 0, 1)$ the tangent plane T_pM is literally
the plane that touches the surface at this point.

```
          z
          ↑          tangent plane at p
          |          z = 2x − 1  (marked with ≡≡≡)
          |
          |            ≡≡≡≡≡≡≡≡≡≡≡≡
          |           ≡          ●≡  ← point p = (1, 0, 1)
          |        /            /
          |       / paraboloid /
          |      /  z = x²+y² /
          |     /____________/
          |    /            /
          |   \____________/  ← curve of intersection
          |    \__________/      with plane z = 1
          +-------────────+-----→ x
         /                |
        /                 p projected to xy: (1, 0)
       y
                   (paraboloid shown only from above for clarity)
```

Equation of tangent plane:

$$z = f(1,0) + f_x(1,0)\cdot(x-1) + f_y(1,0)\cdot(y-0)$$

$$z = 1 + 2\cdot(x-1) + 0\cdot(y-0) = 2x - 1$$

Vector in T_pM is an arrow lying in the tangent plane.
Gradient ∇f is one such arrow (direction of greatest increase).

Cotangent space — what is it?

T^*_pM is the space of all "prices per unit displacement" at point p.

If T_pM answers the question "where can one go?" (directions),
then T^*_pM answers the question "how much does it cost to go?" (linear functions).

Specifically for the paraboloid at point $p = (1, 0, 1)$:

Differential df is a covector that says:
"If you displace by vector v, function f will change by $df(v)$".

$df = f_x\,dx + f_y\,dy = 2dx + 0dy = 2dx$

Take vector $v = (3, 4)$ in the tangent plane.
$df(v) = 2\cdot3 + 0\cdot4 = 6$

This means: with displacement by v function f will increase approximately by 6.

Geometrically: df is "level lines with marking".
df(v) = how many level lines vector v crosses.

```
==============  level lines f = const
==============
   →v          vector crosses several lines
==============
==============
```

Gradient vs differential — arrow vs marking

∇f (gradient) — vector:

- Lives in tangent space T_pM
- This is an arrow pointing in direction of greatest increase of f
- Components: $(\nabla f)^i = g^{ij}\,\partial f/\partial x^j$ (need metric)
- Dimension: [units of f / units of length], but it is a vector

Df (differential) — covector:

- Lives in cotangent space T^*_pM
- This is a function: "how much f will change with displacement by v"
- Components: $(df)_i = \partial f/\partial x^i$ (metric not needed)
- Dimension: [units of f / units of length], but it is a covector

Connection:
$df(v) = \langle \nabla f, v\rangle = g(\nabla f, v)$
∇f = "raised" df using metric: $(\nabla f)^i = g^{ij}(df)_j$

Why they're confused:
In Euclidean space with orthonormal basis $g^{ij} = \delta^{ij}$,
therefore components of ∇f and df numerically coincide.
But conceptually these are different objects.

∇f — arrow (direction)
df — marking (function on directions)

Key point: df exists always (for differentiable f).
∇f exists only if metric g is given.
Without metric one can talk about df, but not about ∇f.

Dual basis = prices for basis units

If $\{e_1, e_2, ..., e_n\}$ — basis of V (units of measurement),
then dual basis $\{\varepsilon^1, \varepsilon^2, ..., \varepsilon^n\}$ are "prices":

ε^1 = "1 ruble per unit of e_1, 0 for the rest"
ε^2 = "1 ruble per unit of e_2, 0 for the rest"

Formally:

$$\varepsilon^i(e_j) = \delta^i{}_j = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases}$$

Then ε^i "selects" the i-th component of a vector:

$v = v^1e_1 + v^2e_2 + ... + v^ne_n$
$\varepsilon^2(v) = v^2$ ← second component

Example:

Basis: e_1 = "1 meter", e_2 = "1 second"
Dual basis: ε^1 = "select meters", ε^2 = "select seconds"
Vector: $v = (5\text{ m}, 3\text{ s}) = 5e_1 + 3e_2$
$\varepsilon^1(v) = 5$, $\varepsilon^2(v) = 3$

On a manifold:

Basis of T_pM: $\{\partial/\partial x, \partial/\partial y, ...\}$ — directions of coordinate lines
Dual basis of $T^*{}_pM$: $\{dx, dy, ...\}$ — differentials of coordinates
$dx(\partial/\partial y) = 0$, $dx(\partial/\partial x) = 1$

Final table: vector vs covector

	VECTOR v	COVECTOR ω
Physical analogy	Quantity (5 meters)	Price per unit (10 rub/meter)
Dimension	[m], [s], [kg]	[1/m], [1/s], [1/kg]
Where it lives	Tangent T_pM	Cotangent T^*_pM
Geometric image	Arrow	Marking (level lines)
When measurement units become smaller	Components grow (m→cm: ×100)	Components fall (m→cm: ÷100)
Examples	Velocity, displacement, gradient ∇f	Differential df, momentum p, pressure
Index	Upper: v^i	Lower: ω_i
Basis	$\{\partial/\partial x^1, \partial/\partial x^2, ...\}$	$\{dx^1, dx^2, ...\}$
Product	$\omega(v) = \omega_i v^i$ = NUMBER (invariant)	

Examples of Covectors in Physics

1. differential of a function df
   "Change of f per unit displacement"
   df(v) = how much f will change when displaced by v

2. momentum p in mechanics
   $p = \partial L/\partial\dot{q}$ — this is a covector, not a vector.
   Work = $p\cdot v = p_i v^i$ — must be a scalar (number)
   Therefore if v — vector, then p — covector

3. wave covector k
   Wave phase $\varphi = k\cdot x = k_i x^i$ — must be a scalar
   k defines "direction" through surfaces of equal phase

4. pressure, stress
   "Force per unit area" — price per area

General principle:
If quantity A·B must be an invariant (number),
and A — vector, then B — covector.

Duality — what is it

Duality — definition and applications

Two ways to describe the same thing:

Points of space	Functions on space
"Where is it located?"	"What can be measured?"

Example 1: Map of terrain

- Points on map (places)
- Contour lines of height (height functions)

One and the same mountain is described both by points of the peak and by level lines.

Example 2: Vector vs linear function

- Vector $v = (3, 4)$ — "arrow" in space
- Function $\varphi(x,y) = 3x + 4y$ — "measurer", gives number for any point

Important: Identification $v \leftrightarrow \varphi$ requires scalar product (metric).
Formula: $\varphi(u) = \langle v, u\rangle$. Without metric vector and covector — different objects.
Isomorphism $V \cong V^*$ (Riesz–Fréchet theorem) exists only in Hilbert spaces. In general, $V^* \neq V$ as structures.

Example 3: Time and frequency

- Signal in time: $f(t)$
- Spectrum (frequencies): $\hat{f}(\omega)$

Fourier transform — transition between dual descriptions

Key idea:
To study space = to study functions on it
These two approaches are equivalent and complement each other

What is duality

Duality is a correspondence between two mathematical structures, in which:

- Each object A corresponds to a "dual" object A^*
- Each operation corresponds to a "dual" operation
- $(A^*)^* \cong A$ (dual to dual = original)

Deep meaning:
Space and functions on it are equal descriptions.
One can study points, or one can study "tests" (what can be measured).

Dual vector space

Definition: dual space

Let V — vector space over field F.

Dual space $V^* = \mathrm{Hom}(V, F)$ is the set of all linear functionals on V:

$$V^* = \{\varphi: V \to F \mid \varphi \text{ linear}\}$$

Structure:
- $(\varphi + \psi)(v) = \varphi(v) + \psi(v)$
- $(\alpha\varphi)(v) = \alpha\cdot\varphi(v)$

V^* itself is a vector space.

Dual basis

Let $\{e_1, ..., e_n\}$ — basis of V.

Dual basis $\{e^1, ..., e^n\}$ of space V^* is defined:

$$e^i(e_j) = \delta^i{}_j = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases}$$

Properties:
- $\dim(V^*) = \dim(V)$ only for finite-dimensional (for counterexample)
- Any $\varphi \in V^*$ expands: $\varphi = \varphi(e_i)\cdot e^i = \varphi_i e^i$
- Action on vector: $\varphi(v) = \varphi_i v^i$ (contraction)

Example: $V = \mathbb{R}^3$ with basis $e_1 = (1,0,0)$, $e_2 = (0,1,0)$, $e_3 = (0,0,1)$

$e^1(x,y,z) = x$ (projection onto first coordinate)
$e^2(x,y,z) = y$ (projection onto second coordinate)
$e^3(x,y,z) = z$ (projection onto third coordinate)

Canonical isomorphism $V \cong V^{**}$

For finite-dimensional spaces there exists a canonical isomorphism:

$$\iota: V \to V^{**}$$
$$\iota(v)(\varphi) = \varphi(v)$$

"Vector v is a functional on functionals, which evaluates them at v"

Important:

- $V \cong V^*$ requires choice of basis (or scalar product)
- $V \cong V^{**}$ is canonical (does not require choice)

In infinite-dimensional case:

- For Banach spaces: $V \hookrightarrow V^{**}$ (embedding, not isomorphism)
- $V = V^{**}$ is called reflexive space
- L^p is reflexive for $1 < p < \infty$, but L^1 and L^∞ are not

Duality in different areas

Preliminary explanations

In the table below there are terms that are worth clarifying:

$H_k(X)$ — homology — algebraic invariants, counting "holes" in space
H_0 counts connected components (how many "pieces")
H_1 counts "tunnels" (can one go around and return nontrivially)
H_2 counts "cavities" (closed voids inside)
Formally: $H_k = \ker(\partial_k)/\mathrm{Im}(\partial_{k+1})$ — "cycles that are not boundaries"

$H^k(X)$ — cohomology — dual to homology (functions on cycles)
de Rham cohomology: classes of closed forms ω ($d\omega=0$) modulo exact

Intuition via potential: closed form is a field that locally looks like gradient (one can find potential on small piece). Exact form is a field that globally is gradient (potential exists everywhere). Cohomology measures obstructions (holes) preventing local potential from becoming global.

Kronecker pairing: $\langle\cdot,\cdot\rangle: H^k(X) \times H_k(X) \to \mathbb{R}$
Intuitively: cohomology "measures" homology — integral of form over cycle

$S^1 = \{z \in \mathbb{C} : |z| = 1\}$ — unit circle in complex plane

Table of Dualities

Important: These are different types of duality, united by a common pattern.

TYPE OF DUALITY	EXAMPLES FROM TABLE BELOW

Algebraic (lin. algebra)	$V \leftrightarrow V^*$, rows ↔ columns
Pontryagin (top. groups)	$G \leftrightarrow \hat{G}$ (group ↔ character group)
Poincaré (topology)	$H_k \leftrightarrow H^{n-k}$ (homology ↔ cohomology)
Categorical (category theory)	$F \leftrightarrow F^{op}$ (covariant ↔ contravariant)

Common: everywhere there is a pairing $\langle\cdot,\cdot\rangle \to$ number. But the details are substantially differe

OBJECT	DUAL	CONNECTING MAPPING
Vector $v \in V$	Covector $\varphi \in V^*$	$\varphi(v) \in F$ (pairing)
Matrix row	Matrix column	Transposition A^T
Point $x \in X$	Function $f: X \to \mathbb{R}$	$f(x)$ (value at point)
Abelian group G	Character group $\hat{G} = \mathrm{Hom}(G, S^1)$	$\chi: G \to S^1$ (homomorphism to unit circle)
Time t	Frequency ω	Fourier: $e^{i\omega t}$
Position q	Momentum p	Hamiltonian mechanics
T_pM (tangent)	T^*_pM (cotangent)	$df(v) = v(f)$
$H_k(X)$ (homology) "cycles"	$H^k(X)$ (cohomology) "forms"	$\langle \omega, c \rangle = \int_c \omega$ (integral)
$\cup$ (union) $\vee$ (or)	$\cap$ (intersection) $\wedge$ (and)	De Morgan's laws: $\neg(A \cup B) = \neg A \cap \neg B$
Point (in project. geometry)	Hyperplane	Point ↔ set of lines through it
Functor $F: C \to D$ (covariant)	Contravar. functor $F: C^{op} \to D$	$F(f: A \to B)$ gives $F(f): F(B) \to F(A)$ (arrow reverses)

Main idea of duality

In all examples above one object "measures" another through pairing:

$\langle \text{dual}, \text{original} \rangle \to \text{number}$

This is like thermometer and temperature, ruler and length, scales and mass.
The dual object is a "measuring instrument" for the original.

Pontryagin Duality

Explanation: What is a "locally compact" group

A locally compact group is a topological group (group + topology, compatible with each other), in which every point has a compact neighborhood.

Compactness — boundedness + closedness (in $\mathbb{R}^n$). Intuitively: can be "covered by a finite number of small balls".

Examples of locally compact groups:
- $\mathbb{R}$, $\mathbb{R}^n$ — yes (any point has a closed ball around it)
- $\mathbb{Z}$ — yes (discrete topology, each point is itself compact)
- S^1 — yes (compact as a whole)
- $GL(n,\mathbb{R})$ — yes (open subset of $\mathbb{R}^{n^2}$)

Examples of non-locally compact:
- Infinite-dimensional Banach spaces (unit ball is not compact)

Why this is needed:
On locally compact groups there exists a Haar measure (invariant measure), which allows integration and defining the Fourier transform.

Definition: dual group

For a locally compact abelian group G its dual group:

$$\hat{G} = \mathrm{Hom}(G, S^1) = \{\chi: G \to S^1 \mid \chi \text{ continuous homomorphism}\}$$

Elements of $\hat{G}$ are called characters of the group G.

Pontryagin's Theorem:
For a locally compact abelian group G: $\hat{\hat{G}} \cong G$

Examples of Pontryagin duality

GROUP G	DUAL $\hat{G}$	CHARACTERS
$\mathbb{Z}$	$S^1 \cong \mathbb{R}/\mathbb{Z}$	$\chi_t(n) = e^{2\pi int}, t \in [0,1)$
$S^1 \cong \mathbb{R}/\mathbb{Z}$	$\mathbb{Z}$	$\chi_n(t) = e^{2\pi int}, n \in \mathbb{Z}$
$\mathbb{R}$	$\mathbb{R}$	$\chi_\xi(x) = e^{2\pi i\xi x}, \xi \in \mathbb{R}$ (self-dual)
$\mathbb{Z}/n\mathbb{Z}$	$\mathbb{Z}/n\mathbb{Z}$	$\chi_k(m) = e^{2\pi ikm/n}$ (self-dual)
$\mathbb{R}^n$	$\mathbb{R}^n$	$\chi_\xi(x) = e^{2\pi i\langle\xi,x\rangle}$

Connection with the Fourier transform

Main idea: the Fourier transform is precisely the transition to the dual group. This is not an analogy, this is literally the same thing.

- Time t lives in the group $\mathbb{R}$
- Frequency ω lives in the dual group $\hat{\mathbb{R}} \cong \mathbb{R}$
- Fourier image $\hat{f}(\omega)$ = "coordinates of the function in the basis of characters"

Subtlety: The isomorphism $\hat{\mathbb{R}} \cong \mathbb{R}$ is not canonical – it depends on the choice of normalization (2π in the exponential). This is like $V^* \cong V$: true, but requires a choice. Different textbooks use $e^{i\omega t}$, $e^{2\pi i\xi t}$ or $e^{-i\omega t}$ – hence the confusion with coefficients 2π in Fourier formulas.

When an engineer performs FFT (fast Fourier transform), they implicitly use Pontryagin duality for the group $\mathbb{Z}/n\mathbb{Z}$.

The Fourier transform is the decomposition of a function in characters:

$$\hat{f}(\xi) = \int_G f(x)\, \chi_\xi(x)^{-1}\, dx$$

Correspondence table:

GROUP	FOURIER	APPLICATION
$\mathbb{R}$	Ordinary Fourier	Signal processing
S^1	Fourier series	Periodic functions
$\mathbb{Z}$	DTFT	Discrete signals
$\mathbb{Z}/n\mathbb{Z}$	DFT (FFT)	Digital processing
Finite group	Representation th.	Chemistry, physics

Gelfand–Naimark Theorem

Space ↔ algebra of functions

Theorem (Gelfand–Naimark):

A compact Hausdorff space X is completely determined by the algebra of continuous functions C(X).

$X \leftrightarrow C(X)$ (equivalence of categories)

Corollary:
One can study the space X by studying functions on it.
This is the basis of noncommutative geometry.

Generalization (noncommutative geometry):
Noncommutative C*-algebras = "functions on noncommutative spaces"
Quantum mechanics = noncommutative geometry of phase space

Physical Meaning of Duality

Position ↔ Momentum

	POSITION q	MOMENTUM p
Meaning	"Where?"	"Where is it moving?"
Mathematically	$q \in Q$ (configuration)	$p \in T^*_qQ$ (cotangent)
Quantum mechanics	$\psi(q)$ – wave function	$\tilde{\psi}(p) = F[\psi(q)]$ – Fourier
Uncertainty	$\Delta q \cdot \Delta p \geq \hbar/2$	

Similarly: Time ↔ Energy ($\Delta E \cdot \Delta t \geq \hbar/2$)

Summary: two ways to look at space

```
+--------------------------------------------------------------+
|                                                              |
| SPACE                              FUNCTIONS ON IT           |
| (points, objects)                  (measurements, tests)     |
|                                                              |
| X            ◄==========►            C(X)                    |
|                                                              |
| • Points x ∈ X                    • Functions f: X → ℝ       |
| • Subsets                         • Ideals                   |
| • Maps X → Y                      • Homomorphisms C(Y) → C(X)|
| (direction reversed)                                         |
|                                                              |
+--------------------------------------------------------------+
```

Both descriptions are equivalent.
The choice depends on the problem and convenience.

Applied Example: Duality in Thermodynamics

Thermodynamics is full of dual pairs — this is no coincidence.

EXTENSIVE (additive)	INTENSIVE (measurer)	RELATION
Volume V	Pressure p	δW = −p dV
Entropy S	Temperature T	δQ = T dS
Amount of sub n	Chem. potential μ	δG = μ dn
Charge q	Potential φ	δW = φ dq

Mathematical meaning:

Extensive quantities = coordinates of state space
Intensive quantities = functions on this space (covectors)

Work/heat = pairing ⟨p, dV⟩, ⟨T, dS⟩
This is exactly the action of a covector on a vector.

Legendre Transform:

Transition between descriptions: U(S,V) ↔ H(S,p) ↔ F(T,V) ↔ G(T,p)

This is a change of basis between dual variables.

- U(S,V) — internal energy (natural: S, V)
- H(S,p) = U + pV — enthalpy (replaced V with p)
- F(T,V) = U − TS — free energy (replaced S with T)
- G(T,p) = U + pV − TS — Gibbs potential (replaced S→T, V→p)

Practical meaning:

- At $p = \text{const}$ enthalpy H is convenient (pumps, compressors)
- At $T = \text{const}$ free energy F is convenient (isothermal processes)
- At $T, p = \text{const}$ Gibbs potential G is convenient (chemical reactions)

Conclusion: Duality is not an abstraction, but a working tool.
Thermodynamic potentials are a choice of convenient description for a specific problem (what is fixed: T or S? p or V?)

Delta Function – Duality in Action

Problem: What is a "point charge" or an "instantaneous impulse"?

Engineers constantly use the "delta function" $\delta(x)$:
- Point charge: $\rho(x) = q\cdot\delta(x - x_0)$
- Impulsive force: $F(t) = P\cdot\delta(t - t_0)$
- Initial condition: $u(x,0) = \delta(x)$

Formally they write:

$$\delta(x) = \begin{cases} 0, & x \neq 0 \\ \infty, & x = 0 \end{cases} \quad \text{and} \quad \int\delta(x)dx = 1$$

But this is not a function. There is no function with value ∞ at a point and integral 1.

Solution: δ is a functional (element of the dual space)

Correct definition:

$$\delta: C^\infty_0(\mathbb{R}) \to \mathbb{R}, \quad \delta[\varphi] = \varphi(0)$$

δ is a linear functional that takes a test function φ and returns its value at zero.

The "integral" $\int\delta(x)\varphi(x)dx$ is understood as $\delta[\varphi] = \varphi(0)$.

Connection with duality:
- Space of probe (test) functions:
  $\mathcal{D}(\mathbb{R}) = C^\infty_0(\mathbb{R}) = \{\varphi: \text{smooth with compact support}\}$
  (φ vanishes outside some bounded interval)
- Dual space: $\mathcal{D}'(\mathbb{R}) = \{\text{generalized functions/distributions}\}$
- $\delta \in \mathcal{D}'(\mathbb{R})$ – element of the dual space.

Another variant: Schwartz class $\mathcal{S}(\mathbb{R})$ – rapidly decreasing functions.
Then $\mathcal{S}'(\mathbb{R})$ – tempered distributions (including δ).

The derivative δ' is also a functional:

$$\delta'[\varphi] = -\varphi'(0) \quad \text{(integration by parts)}$$

Practical properties

$\int f(x)\delta(x-a)dx = f(a)$ "selects" the value of f at point a

$\int f(x)\delta'(x-a)dx = -f'(a)$ "selects" the derivative at point a

Important: These formulas make sense only for smooth (test) functions f. Applying δ to an arbitrary function from L^2 is incorrect – such functions have no "value at a point" (only an equivalence class).

$\delta(ax) = (1/|a|)\delta(x)$ scaling of the argument

$x\delta(x) = 0$ $x\cdot 0 = 0$ everywhere, including $x=0$

Fourier transform:

$F[\delta(x)](\omega) = 1$ $\delta(x) \leftrightarrow \text{const}$

$F[1](\omega) = 2\pi\delta(\omega)$ $\text{const} \leftrightarrow \delta(\omega)$

Convolution:

$f * \delta = f$ δ – neutral element of convolution

$f * \delta' = f'$ differentiation through convolution

Forbidden operations with δ

In the theory of generalized functions (Schwartz distributions), it is not allowed:

- $\delta(x)\cdot\delta(x) = \delta^2(x)$ – undefined.
- $\delta(x)\cdot|x|^{-1}$ – undefined.
- Any nonlinear operations with δ

Why: The product of two distributions is generally undefined. Distributions are linear functionals, and the product $\delta(x)\cdot\delta(x)$ would require a "value of δ at a point", which does not exist.

Where this is a problem:

- Quantum field theory: "divergences" = attempts to compute δ^2
- Nonlinear equations with singular data
- Squaring a signal (power of a signal with a δ-pulse)

What is allowed: Linear operations (addition, differentiation, convolution with a smooth function) – all of this is well-defined.

Applications in engineering

Impulse response:

Linear system L with input $u(t)$ and output $y(t) = L[u]$

Impulse response: $h(t) = L[\delta(t)]$

Then: $y(t) = (h * u)(t) = \int h(\tau)u(t-\tau)d\tau$

Transfer function:

$H(s) = L[h(t)]$ (Laplace transform of the impulse response)

If $H(s) = 1/(s+a)$, then $h(t) = e^{-at}$

Green's function:

The solution of the equation $LG(x,\xi) = \delta(x-\xi)$ is called the Green's function

Solution of $Lu = f$: $u(x) = \int G(x,\xi)f(\xi)d\xi$

We have become acquainted with duality: V and V*. Now we can construct objects that "eat" several vectors and covectors simultaneously. These are tensors — multilinear functions on products of V and V*.

Tensors — the language of physics: stresses, strains, electromagnetic field.

Tensors — multilinear objects on spaces

A tensor is the culmination of the theme "object vs observer".

A tensor is an object that is indifferent to the choice of observer.

A vector exists independently of a basis. But to write it down with numbers, one can only by choosing a basis — that is, by becoming an "observer". Different observers will write one vector with different numbers. The transformation law of components is the rule for conversion between the languages of different observers.

A tensor is a generalization: an object that "eats" several vectors and covectors and produces a number. This number is an invariant, the same for all observers. The components of a tensor (numbers in a table) change when the basis changes, but according to law that guarantees that the result of the calculation will remain the same.

This is not abstract mathematics — this is the language of physics. Stress in a material, the metric of spacetime, the electromagnetic field — all of these are tensors. They describe a reality that does not depend on how we decided to write it down.

The main mistake: confusing a tensor (geometric object) with its components (numbers in a specific basis). A 3×3 matrix can be a record of completely different tensors — or not be a tensor at all.

The key principle: a tensor does not change when coordinates change — our description (components) changes to compensate for the change of ruler.

Tensors as a view of space

Tensors describe how properties of space change from point to point and from direction to direction.

- Scalar (tensor of rank 0): temperature $T(x)$ — one number at each point
- Vector (rank 1): velocity $v(x)$ — direction at each point
- Tensor (rank 2): stresses $\sigma_{ij}(x)$ — depend on point and on plane

Tensors — the language of describing anisotropy: when space is "different" in different directions (thermal conductivity of a crystal, elasticity of a composite).

Now that we understand duality, we can define tensors.

Why tensors are needed — motivation

Why would anyone need to change coordinate systems? — motivation

Before talking about tensors and transformations, let us answer the main question: Why would anyone need to change coordinates?

Reason 1: different observers

An observer stands on a platform. A train passes by. A passenger sits in the train.

- For the observer: the passenger moves at a speed of 100 km/h
- For the passenger: he is at rest, and the observer moves at a speed of −100

Who is right? Both. Velocity is a vector, and its components depend on the reference frame. But the passenger himself is the same physical object.

Problem: The laws of physics must be the same for all observers.
$F = ma$ must work both on the platform and in the train.

Reason 2: convenience of solving a problem

Problem: find the area of the ellipse $x^2/a^2 + y^2/b^2 = 1$

In Cartesian coordinates — a complicated integral.
Make a substitution: $u = x/a$, $v = y/b$
The ellipse turns into a circle $u^2 + v^2 = 1$, area = π.
Area of the ellipse = πab (taking into account the stretching of coordinates).

Other examples:
- Problem with rotation → polar coordinates
- Problem with a sphere → spherical coordinates
- Oscillations → normal modes (eigenvectors)

The right choice of coordinates turns a difficult problem into a simple one.

Reason 3: measurements in different units

American engineer: pipe of length 10 feet, diameter 6 inches
Russian engineer: the same pipe — 3.048 m, diameter 152.4 mm

The pipe is the same. The numbers are different.

If the formula for calculating pressure losses gives different answers in feet and in meters — the formula is incorrect. Physics doesn't know about feet and meters.

Reason 4: curvilinear coordinates on surfaces

On the surface of the Earth there are no "natural" Cartesian coordinates. We use latitude and longitude — but they behave strangely:

- At the pole longitude is undefined
- "A meter to the east" has different length at the equator and in Moscow

At different points the basis vectors $\partial/\partial\varphi$ and $\partial/\partial\theta$ have different length. This forces us to think about coordinate transformations.

Conclusion: what a "correct" mathematical object must be able to do

A physical quantity (velocity, force, stress, metric) must:

1. exist independently of the choice of coordinates
2. have components that change when coordinates change
3. change according to a definite law, so that the quantity remains the same

An object satisfying these requirements is called a tensor.

The transition matrix A is simply a record of how the new basis vectors are expressed in terms of the old ones. It is forced to be exactly this way, because the basis vectors are concrete geometric objects.

Concrete example: thermal conductivity in a layered material

There is a layered thermal insulation: the layers go horizontally.

- Along the layers thermal conductivity λ‖ = 0.5 W/(m·K)
- Across the layers thermal conductivity λ⊥ = 0.05 W/(m·K)

If the coordinate axes coincide with the layers:

$$\lambda = \begin{pmatrix} 0.5 & 0 \\ 0 & 0.05 \end{pmatrix}$$

Diagonal matrix — simple.

Now rotate the coordinates by 45°. The layers remained the same. But the thermal conductivity matrix will change:

$$\lambda' = R^{-1} \lambda R = \begin{pmatrix} 0.275 & 0.225 \\ 0.225 & 0.275 \end{pmatrix}$$

Off-diagonal elements appeared.

What does this mean physically?

In rotated coordinates: if the temperature gradient is directed along the x' axis, then the heat flux will not be along x', but at an angle. Because the material "wants" to conduct heat along the layers.

$\lambda' \neq \lambda$ — the components are different
But the physical material is the same
λ and λ' describe one object in different coordinates

The transformation law $\lambda' = R^{-1}\lambda R$ is not an arbitrary choice. It is forced to be this way, so that the heat flux $q = -\lambda\nabla T$ is the same physical vector in any coordinates.

Why tensors are needed

Problem 1: Physical quantities must not depend on coordinates.

- Temperature is one, no matter how you rotate the axes
- Velocity is one vector, but the components change
- Stress in a material — how to describe?

Problem 2: Objects "more complicated" than vectors are needed

- A scalar takes a point → gives a number
- A vector. What does it take? What does it give?
- A matrix acts on a vector → gives a vector
- A metric takes two vectors → gives a number (length/angle)

Solution: A tensor = a multilinear map with a definite law
of coordinate transformation

Multilinearity — the key idea

Definition: multilinear mapping

A mapping $T: V_1 \times V_2 \times \ldots \times V_k \to W$ is called multilinear (or k-linear), if it is linear in each argument separately:

$$T(v_1, \ldots, \alpha u_i + \beta w_i, \ldots, v_k) = \alpha T(v_1,\ldots,u_i,\ldots,v_k) + \beta T(v_1,\ldots,w_i,\ldots,v_k)$$

for all $i = 1, \ldots, k$.

Explanation: what does "linear in each argument" mean

Imagine a function $f(x, y)$ of two variables.

"Linear in x" means:
If we fix y, then f behaves as a linear function of x:
$f(\alpha x_1 + \beta x_2, y) = \alpha \cdot f(x_1, y) + \beta \cdot f(x_2, y)$

"Linear in y" means:
If we fix x, then f behaves as a linear function of y:
$f(x, \alpha y_1 + \beta y_2) = \alpha \cdot f(x, y_1) + \beta \cdot f(x, y_2)$

"Bilinear" = linear in x and linear in y (simultaneously).

Concrete numerical example: Scalar product in $\mathbb{R}^2$

Scalar product: $\langle u, v\rangle = u_1v_1 + u_2v_2$

Take concrete vectors:
$u = (1, 2),\ \ v = (3, 0),\ \ w = (0, 1)$

Checking linearity in the first argument:

Compute $\langle 2u + 3w, v\rangle$ in two ways:

Method 1 (direct):
$2u + 3w = 2\cdot(1,2) + 3\cdot(0,1) = (2,4) + (0,3) = (2, 7)$
$\langle(2,7), (3,0)\rangle = 2\cdot3 + 7\cdot0 = 6$

Method 2 (via linearity):

$$\begin{aligned} 2\cdot\langle u, v\rangle + 3\cdot\langle w, v\rangle &= 2\cdot\langle(1,2),(3,0)\rangle + 3\cdot\langle(0,1),(3,0)\rangle \\ &= 2\cdot(1\cdot3 + 2\cdot0) + 3\cdot(0\cdot3 + 1\cdot0) \\ &= 2\cdot3 + 3\cdot0 = 6 \quad ✓ \end{aligned}$$

The results coincided – linearity in the first argument works.

Why this is important:
Multilinearity allows us to "decompose" complex expressions into parts.
Instead of computing T(complex_combination) we can expand into
simple summands and add the results.

Counterexample – the norm is not linear:
$\|2u\| = \|(2,4)\| = \sqrt{(4+16)} = \sqrt{20} \approx 4.47$
$2\cdot\|u\| = 2\cdot\|(1,2)\| = 2\cdot\sqrt{5} \approx 4.47$ ✓ (coincided here)

But: $\|(-1)\cdot u\| = \|(-1,-2)\| = \sqrt{5} \approx 2.24$
$(-1)\cdot\|u\| = -\sqrt{5} \approx -2.24$ ✗ (did not coincide)

The norm always gives a positive number, whereas $(-1)\cdot\|u\|$ is negative.
Therefore the norm is not linear, and consequently is not a tensor.

Examples of multilinear mappings

MAPPING	TYPE	MULTILINEARITY CHECK
Scalar product $\langle u, v\rangle$	$V \times V \to \mathbb{R}$ bilinear ($k=2$)	$\langle \alpha u, v\rangle = \alpha\langle u,v\rangle$ ✓ $\langle u, \alpha v\rangle = \alpha\langle u,v\rangle$ ✓
Determinant $\det(v_1,...,v_n)$	$V \times V \times ... \times V \to \mathbb{R}$ n-linear	$\det(..., \alpha v, ...) =$ $= \alpha\cdot\det(..., v, ...)$ ✓
Cross product $u \times v$	$V \times V \to V$ bilinear	$(\alpha u) \times v = \alpha(u \times v)$ ✓ $u \times (\alpha v) = \alpha(u \times v)$ ✓
Metric $g(u, v)$	$T_pM \times T_pM \to \mathbb{R}$ bilinear	$g(\alpha u, v) = \alpha g(u,v)$ ✓ $g(u, \alpha v) = \alpha g(u,v)$ ✓
Riemann curvature $R(X,Y)Z$	$T_pM \times T_pM \times T_pM \to T_pM$ trilinear	$R(\alpha X,Y)Z = \alpha R(X,Y)Z$ ✓ (linear in all three)

Counterexample: what is not multilinear

The norm $\|v\| = \sqrt{\langle v,v\rangle}$ is not linear.

$$\|\alpha v\| = |\alpha|\cdot\|v\| \neq \alpha\cdot\|v\| \quad (\text{for } \alpha < 0)$$

The norm is not a tensor. It is a function, but not a linear one.

Formal definition of a tensor

Definition: tensor of type (p, q)

Let V be a vector space over $\mathbb{R}$, V^* its dual.

A tensor of type (p, q) on V is a multilinear mapping:

$$T: \underbrace{V^* \times ... \times V^*}_{p \text{ copies}} \times \underbrace{V \times ... \times V}_{q \text{ copies}} \to \mathbb{R}$$

p is called the contravariant valence (upper indices)
q is called the covariant valence (lower indices)

Total valence (rank) of the tensor = p + q

Hierarchy of tensors

TYPE	(p,q)	WHAT IT IS / HOW MANY COMPONENTS
Scalar	(0,0)	T: {} → $\mathbb{R}$ (just a number) 1 component: T
Vector	(1,0)	T: V* → $\mathbb{R}$ (linear function on covectors) n components: T^i Identified with an element of V via V ≅ V**
Covector (1-form)	(0,1)	T: V → $\mathbb{R}$ (linear function on vectors) n components: T_i This is an element of V*
Bilinear form	(0,2)	T: V × V → $\mathbb{R}$ n^2 components: T_{ij} Example: metric g_{ij}, scalar product
Linear operator	(1,1)	T: V* × V → $\mathbb{R}$ or T: V → V n^2 components: $T^i{}_j$ Example: transformation matrix
Tensor (2,0)	(2,0)	T: V* × V* → $\mathbb{R}$ n^2 components: T^{ij} Example: inverse metric g^{ij}
Riemann tensor	(1,3)	R: V* × V × V × V → $\mathbb{R}$ n^4 components: $R^i{}_{jkl}$ (with symmetries: 20 independent in 4D)

Tensor components in a basis

Let $\{e_1, ..., e_n\}$ – basis of V, $\{e^1, ..., e^n\}$ – dual basis of V*.

Components of tensor T of type (p,q):

$$T^{i_1...i_p}_{j_1...jq} = T(e^{i_1}, ..., e^{i_p}, e_{j_1}, ..., e_{jq})$$

The tensor itself is reconstructed:

$$T = T^{i_1...i_p}_{j_1...jq} \cdot e_{i_1} \otimes ... \otimes e_{i_p} \otimes e^{j_1} \otimes ... \otimes e^{jq}$$

(summation over repeated indices – Einstein summation convention)

Key distinction: matrix-table vs matrix-tensor

It is often said "a tensor is a multidimensional table of numbers". This is incorrect.
A table is only a representation of a tensor in a specific basis.

Matrix as table:
Just 4 numbers in a definite order. Do not change under change of basis.
Example: data in Excel, image pixels.

Matrix as tensor:
Numbers + type + transformation rule.
Under change of basis the numbers change according to a definite law.
Example: metric, linear operator, bilinear form.

Example: one matrix – three different tensors

Take the identity matrix 2×2:

$$M = \begin{pmatrix} 1 & 0 \\ & \\ 0 & 1 \end{pmatrix}$$

This matrix can represent three different tensors:

Tensor type	Indices	What it is
(1,1) – operator	$M^i{}_j$	Identity transformation v ↦ v
(0,2) – bilin.form	M_{ij}	Standard scalar product ⟨u,v⟩
(2,0) – bivector	M^{ij}	Inverse metric (raises indices)

Now apply a change of basis: stretch the x axis by a factor of 2.

Transition matrix: $A = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$, $A^{-1} = \begin{pmatrix} 1/2 & 0 \\ 0 & 1 \end{pmatrix}$

Transformation results:

Type	Formula	Result
(1,1) operator	$M' = A^{-1} M A$	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ (unchanged)
(0,2) form	$M' = A^{T} M A$	$\begin{pmatrix} 4 & 0 \\ 0 & 1 \end{pmatrix}$ (grew in x)
(2,0) bivector	$M' = (A^{-1})M(A^{-1})^{T}$	$\begin{pmatrix} 1/4 & 0 \\ 0 & 1 \end{pmatrix}$ (shrank in x)

Conclusion:

One and the same table of numbers $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ turned into three different matrices.

A tensor is not numbers, but numbers + transformation rule.

Numerical example: how to compute the transformation

Let's verify for type (0,2) – bilinear form (metric):

$$M' = A^{T} M A$$

$$= \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}^{T} \cdot \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 4 & 0 \\ 0 & 1 \end{pmatrix} \quad ✓$$

Physical meaning:

The basis vector $e'_1 = 2e_1$ became twice as long.

The coordinate $x' = x/2$ (in the new units fewer steps are needed).

Length: $s^2 = g'_{11}(x')^2 = 4\cdot(x/2)^2 = x^2 = g_{11}x^2$ ✓

The metric increased by a factor of 4 to compensate for the decrease in coordinates. This is like price: if a unit of goods became twice as large, the price per unit increases fourfold (so that the price per gram remains the same).

Criterion: how to understand whether it's a tensor or just a table?

Ask the question: "Should these numbers change under change of coordinate system to describe the same physical/geometric object?"

JUST A TABLE	TENSOR
Image pixels Data in database Equation coefficients (in fixed coordinates)	Space metric g_{ij} Stress tensor in material Moment of inertia tensor Electromagnetic tensor F_{ij}
Change of coordinates: numbers stay the same	Change of coordinates: numbers change according to a definite law

Conclusion:
Tensor = abstract object, independent of coordinates.
Tensor components = numbers that depend on coordinates in such a way that the tensor itself remains the same.

Tensor Product

Universal Property (canonical definition)

The tensor product $V \otimes W$ is a space with a bilinear map $\otimes: V \times W \to V \otimes W$, through which any bilinear map factors:

$$\begin{tikzcd}
V \times W \arrow[r, "\otimes"] \arrow[dr, "B\ (\text{bilin})"'] & V \otimes W \arrow[d, "\exists!\ \text{linear}"] \\
& U
\end{tikzcd}$$

For any bilinear $B: V \times W \to U$ there exists a unique linear map $\tilde{B}: V \otimes W \to U$ such that $B = \tilde{B} \circ \otimes$.

Meaning: $V \otimes W$ is the "most general" space into which one can map bilinearly.

Construction (for finite-dimensional spaces)

In the finite-dimensional case $V \otimes W$ can be identified with the space of bilinear forms on $V^* \times W^*$:

For $v \in V$ and $w \in W$ we define $v \otimes w \in V \otimes W$:

$$(v \otimes w)(\varphi, \psi) = \varphi(v) \cdot \psi(w) \quad \text{for } \varphi \in V^*,\ \psi \in W^*$$

Properties:

- $(\alpha v) \otimes w = v \otimes (\alpha w) = \alpha(v \otimes w)$
- $(v_1 + v_2) \otimes w = v_1 \otimes w + v_2 \otimes w$
- $v \otimes (w_1 + w_2) = v \otimes w_1 + v \otimes w_2$
- $\dim(V \otimes W) = \dim(V) \cdot \dim(W)$

Basis:

If $\{e_i\}$ is a basis of V, $\{f_j\}$ is a basis of W, then $\{e_i \otimes f_j\}$ is a basis of $V \otimes W$.

Tensor Space as Tensor Product

Tensor space of type (p,q) on V:

$$T^p_q(V) = \underbrace{V \otimes \dots \otimes V}_{\text{p copies}} \otimes \underbrace{V^* \otimes \dots \otimes V^*}_{\text{q copies}}$$

Examples:

- $T^0_0(V) = \mathbb{R}$ (scalars)
- $T^1_0(V) = V$ (vectors)
- $T^0_1(V) = V^*$ (covectors)
- $T^1_1(V) = V \otimes V^* \cong \mathrm{End}(V)$ (linear operators)
- $T^0_2(V) = V^* \otimes V^*$ (bilinear forms)

Example: tensor product in components

Let $v = (v^1, v^2, v^3)$, $w = (w^1, w^2)$ be vectors in $\mathbb{R}^3$ and $\mathbb{R}^2$.

Then $v \otimes w$ is a 3×2 matrix:

$$(v \otimes w)^{ij} = v^i w^j$$

$$= \begin{pmatrix} v^1w^1 & v^1w^2 \\ v^2w^1 & v^2w^2 \\ v^3w^1 & v^3w^2 \end{pmatrix}$$

This is a rank 1 matrix (all rows are proportional).
Not every matrix is a tensor product of two vectors.

Coordinate Transformation

Transition Matrix A – what is it

The transition matrix A describes how new basis vectors are expressed in terms of old ones:

$$e'_1 = A^1{}_1 e_1 + A^2{}_1 e_2 + \ldots$$
$$e'_2 = A^1{}_2 e_1 + A^2{}_2 e_2 + \ldots$$
$$\ldots$$

Or compactly: $e'_j = A^i{}_j\, e_i$ (j-th new = linear combination of old)

Example: Stretching the x-axis by a factor of 2

Old basis: $e_1 = (1,0)$, $e_2 = (0,1)$
New basis: $e'_1 = (2,0) = 2e_1$, $e'_2 = (0,1) = e_2$

Transition matrix: $A = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$

Columns of A are the coordinates of new basis vectors in the old basis.

Physical meaning:

- A describes how the "units of measurement" changed
- A = rotation → rotated coordinate axes
- A = scaling → enlarged/reduced the scale
- This is a passive transformation: objects are the same, only description changes

Important:

A^{-1} – inverse matrix – transforms from new coordinates to old
$\det(A) \neq 0$ – bases must be non-degenerate

Attention: convention for transition matrix

In this atlas A is the basis transition matrix: $e' = Ae$ (new through old).

In physics literature (Landau–Lifshitz, MTW) the coordinate transition matrix is often used: $x' = \Lambda x$. Then the basis transforms as $e' = e\Lambda^{-1}$.

These matrices are inverse to each other: $A = \Lambda^{-1}$.

Don't confuse them. The formula $g' = A^T g A$ (our convention) becomes $g' = (\Lambda^{-1})^T g \Lambda^{-1}$ in the physics convention.

Tensor Transformation Law

Under basis change $e'_i = A_{ji} e_j$, tensor components transform:

Vector (contravariant, upper index):

$v'^i = (A^{-1})^i{}_j v^j$ — transforms oppositely to basis

Covector (covariant, lower index):

$\omega'_i = A^j{}_i \omega_j$ — transforms like basis

Tensor of type (p,q):

$$T'^{i_1\ldots i_p}_{j_1\ldots jq} = (A^{-1})^{i_1}_{k_1}\ldots(A^{-1})^{i_p}_{k_p} \cdot A^{l_1}_{j_1}.A^{lq}_{jq} \cdot T^{k_1\ldots k_p}_{l_1\ldots lq}$$

Key point:

Upper indices transform through A^{-1} (contravariantly)
Lower indices transform through A (covariantly)

Why "contra" and "co" — visual comparison

Situation: doubled the length of basis vector ($e' = 2e$)
New "unit of measurement" = 2 old ones

	VECTOR v^i (contravariant)	COVECTOR ω_i (covariant)
Geom. meaning	"Where to move" (arrow)	"Price per unit" (differential df)
Example	$v = 3e_1$ (displacement)	$\omega = 10\varepsilon^1$ (price)
In old basis	$v^1 = 3$	$\omega_1 = 10$
In new basis ($e' = 2e$)	$v'^1 = 1.5$ (components ÷2)	$\omega'_1 = 20$ (components ×2)
Direction of transformation	AGAINST basis (basis ×2 → comp. ÷2)	TOGETHER with basis (basis ×2 → comp. ×2)
Why so	Vector v is the same, just measured in other units	Covector "price per unit" Unit became larger → price for it also larger
Transformation formula	$v'^i = (A^{-1})^i{}_j v^j$	$\omega'_i = A^j{}_i \omega_j$
Index	UPPER	LOWER

Invariant check:
Old basis: $v^1\omega_1 = 3 \times 10 = 30$
New basis: $v'^1\omega'_1 = 1.5 \times 20 = 30$ ✓

$v^i\omega_i$ is an invariant (doesn't depend on basis, because contra × co)

Examples of contraction (invariants)

CONTRACTION	MEANING
$v^iv_i = g_{ij}v^iv^j$	Square of vector length $\|v\|^2$
$v^i\omega_i$	Action of covector on vector (number)
$T^i{}_i$	Trace of matrix Tr(T)
$R^i{}_{jij} = R$	Scalar curvature (from Riemann tensor)
$F_{ij}F^{ij}$	Electromagnetic field invariant

Operations on Tensors

Table of Operations

OPERATION	ACTION ON TYPES	EXAMPLE

Addition	$(p,q) + (p,q) \to$ $\to (p,q)$	$(S^i{}_j + T^i{}_j)$ Only same types.
Tensor product	$(p_1,q_1) \otimes (p_2,q_2) \to$ $\to (p_1+p_2, q_1+q_2)$	$(v^i)(w^j) = v^i w^j$ Rank adds up
Contraction (trace)	$(p,q) \to (p-1,q-1)$ Rank decreases by 2	$T^i{}_{jk} \to T^i{}_{ji} = S_j$ Sum over repeated index
Raising index	$g^{ij}T_{jk} = T^i{}_k$ $(0,2) \otimes (1,1) \to$ → contraction → (1,1)	Lower → upper Using inverse metric
Lowering index	$g_{ij}T^j = T_i$	Upper → lower Using metric
Symmetrization	$T_{(ij)} = \frac{1}{2}(T_{ij}+T_{ji})$	Symmetric part
Antisymmetrization	$T[_{ij}] = \frac{1}{2}(T_{ij}-T_{ji})$	Antisymmetric part → differential forms.

Tensors in Physics

Main Physical Tensors

TENSOR	TYPE	WHAT IT DESCRIBES

Metric g_{ij}	(0,2)	How to measure distances $ds^2 = g_{ij} dx^i dx^j$ Euclidean: $g_{ij} = \delta_{ij}$ Minkowski: $g = \mathrm{diag}(-1,1,1,1)$
Riemann tensor $R^i{}_{jkl}$	(1,3)	Curvature of space Measures rotation of vector around a contour $R = 0 \iff$ space is flat
Ricci tensor $R_{ij} = R^k{}_{ikj}$	(0,2)	"Trace" of Riemann tensor Enters Einstein's equations
Energy-momentum tensor T_{ij}	(0,2)	Distribution of matter T_{00} = energy density T_{0i} = energy flux = momentum density T_{ij} = pressure/stress
Electromagnetic tensor F_{ij}	(0,2)	E and B fields as one object Antisymmetric: $F_{ij} = -F_{ji}$ → differential 2-form.
Inertia tensor I_{ij}	(0,2)	How a body resists rotation $L = I\omega$ (angular momentum = inertia × ang.vel)
Stress tensor σ_{ij}	(0,2)	Forces inside material σ_{ij} = force in direction j on surface ⊥ i

Einstein's Equations — Tensors in Action

$$\underbrace{R_{ij} - \tfrac{1}{2}g_{ij}R + \Lambda g_{ij}}_{\substack{\text{geometry}\\ \text{(curvature of space)}}} = 8\pi G \cdot \underbrace{T_{ij}}_{\text{matter}}$$

Left side:

- R_{ij} — Ricci tensor (curvature)
- $R = g^{ij}R_{ij}$ — scalar curvature
- Λ — cosmological constant
- g_{ij} — metric of spacetime

Right side:
- T_{ij} – distribution of energy and momentum
- G – gravitational constant

Meaning: Matter curves space. Curvature determines motion.

Connection to Other Sections

Section Connection Graph

```
  (Lin. Algebra)
       |
       | vector spaces, V*
       ▼
  (Products)
       |
       | scalar ⊗ tensor
       ▼
  (Duality)
       |
       | V* – dual space
       ▼
  ▶ (Tensors) ◀
       |
       +--------------► (Diff. Forms) – antisymmetric tensors
       |
       +--------------► (Manifolds) – tensor fields on M
       |
       +--------------► Applications – GR, electromagnetism
```

Summary: What is a Tensor

Three equivalent definitions:

1. Algebraic:
   A tensor of type (p,q) is a multilinear map
   $T: V^* \times ... \times V^* \times V \times ... \times V \to \mathbb{R}$

2. Via basis:
   A collection of numbers $T^{i_1...i_p}_{j_1...jq}$ with a specific law of transformation under change of basis

3. Geometric:
   A tensor is an object that exists independently of coordinates, but can be expressed in components in any coordinate system

All three definitions are equivalent.

Important to remember:

- Tensor ≠ table of numbers. Table is merely a representation in a specific basis
- The same matrix can be a tensor of different types
- The type of tensor determines how the numbers change under change of basis
- Covector (element of V^*) is a linear function on vectors, not a vector

Bonus: π-theorem and Similarity Criteria

This is not directly about tensors, but about dimensions — the foundation of physics.

Problem: Pressure loss in a pipe ΔP depends on ρ, v, D, L, μ, ε.
That's 7 quantities. How to find the dependence?

Buckingham's Π-theorem:

> If n quantities are related by a physical law, and these quantities are expressed through k basic dimensions (M, L, T, ...), then the law can be written as a relation of $(n - k)$ dimensionless combinations.

Example: Losses in a pipe

$n = 7$ (ΔP, ρ, v, D, L, μ, ε), $k = 3$ (M, L, T) → 4 dimensionless numbers

$\Pi_1 = \Delta P/(\rho v^2)$ — pressure coefficient (Euler)
$\Pi_2 = \rho v D/\mu$ — Reynolds number Re
$\Pi_3 = L/D$ — relative length
$\Pi_4 = \varepsilon/D$ — relative roughness

Law: $\Pi_1 = f(\Pi_2, \Pi_3, \Pi_4)$, or: $\Delta P/(\rho v^2) = f(\mathrm{Re}, L/D, \varepsilon/D)$

This gives the Darcy–Weisbach formula: $\Delta P = \lambda(\mathrm{Re}, \varepsilon/D)\cdot(L/D)\cdot(\rho v^2/2)$

Main dimensionless numbers of thermophysics:

NUMBER	FORMULA	PHYSICAL MEANING
Reynolds	$\mathrm{Re} = \rho v L/\mu$	Inertia / Viscosity $\mathrm{Re} < 2300$: laminar, $\mathrm{Re} > 4000$: turb.
Prandtl	$\mathrm{Pr} = \nu/a = \mu c_p/\lambda$	Viscous diffusion / Thermal $\mathrm{Pr} \approx 0.7$ gases, $\mathrm{Pr} \approx 7$ water
Nusselt	$\mathrm{Nu} = \alpha L/\lambda$	Convection / Thermal conductivity $\mathrm{Nu} = 1$: pure conduction
Grashof	$\mathrm{Gr} = g\beta\Delta T L^3/\nu^2$	Buoyancy / Viscosity Free convection
Mach	$\mathrm{Ma} = v/c$	Velocity / Speed of sound $\mathrm{Ma} > 1$: supersonic, shock waves

Why this is needed:
- Modeling: if Re_model = Re_original → flows are similar
- Correlations: $Nu = f(Re, Pr)$ – universal heat transfer formulas
- Estimates: $Re \sim 10^6$ → definitely turbulence, calculate by turb. formulas

Deep connection: π-theorem as a consequence of the Lie group of scalings

Why does the Π-theorem work? Behind it lies symmetry – a Lie group.

Scaling group:

$G = (\mathbb{R}^+, \times)$ – positive real numbers with multiplication
This is a Lie group: continuous, smooth, one-dimensional

Action on dimensional quantities:

$[L] \to \lambda[L]$ (length scales by a factor of λ)
$[T] \to \mu[T]$ (time scales by a factor of μ)
$[M] \to \nu[M]$ (mass scales by a factor of ν)

Key idea:

```
+--------------------------------------------------------------------+
| A physical law must be invariant with respect to the choice of units.|
| This means: the law is invariant with respect to the scaling group G.|
|                                                                    |
| Dimensionless numbers (Re, Pr, Nu) are invariants of the action of G.|
+--------------------------------------------------------------------+
```

Lie algebra and dimension:

Generator of scalings: $D = x\cdot\partial/\partial x + y\cdot\partial/\partial y + z\cdot\partial/\partial z + \ldots$

A quantity f has dimension [L]^a[T]^b[M]^c if:

$D\cdot f = (a + b + c)\cdot f$ (eigenfunction of operator D)

Dimensionless quantity: $D\cdot\Pi = 0$ (invariant)

Example: Why is $Re = \rho vL/\mu$ dimensionless?

$[\rho] = M/L^3$, $[v] = L/T$, $[L] = L$, $[\mu] = M/(L\cdot T)$

$[\rho vL/\mu] = (M/L^3)\cdot(L/T)\cdot L / (M/(L\cdot T)) = (M\cdot L\cdot L)/(L^3\cdot T) \times (L\cdot T)/M = 1$

Re is invariant with respect to $\lambda \to \lambda L$, $\mu \to \mu T$, $\nu \to \nu M$

Generalization – Noether's theorem:

Symmetry	↔	Conservation law
Scaling	↔	Dimensional analysis, Π-theorem
Time translation	↔	Energy conservation
Spatial translation	↔	Momentum conservation
Rotation	↔	Angular momentum conservation

All these symmetries form Lie groups, and their Lie algebras give generators.

Tensors live on vector spaces. But real spaces are curved.
Earth's surface, spacetime, configuration space of a robot.
How to define tensors on curved space?

For this we need manifolds — spaces that locally look like $\mathbb{R}^n$,
but globally can have a complex shape. At each point there is a tangent
space — and that's where tensors live.

Manifolds — spaces that can be mapped

A manifold is the culmination of the idea "locally flat, globally curved".
The Earth is round, but each map in an atlas is flat. The atlas covers the entire Earth,
although no single map covers it completely without distortions.

In "object–observer" terms, a manifold is a place where the observer
is fundamentally local. There is no global coordinate system for the entire space.

A map is the view of one local observer.
An atlas is a coordinated collection of such views.

On a sphere there is no "correct" coordinate system covering everything without singularities.
There are many maps, and in regions of overlap they are coordinated: if a point falls on
two maps, one can convert coordinates from one to the other.

Transition functions between maps are "dictionaries" between the languages of neighboring
observers. The requirement of smoothness of transition functions guarantees that the notion of
"derivative" does not depend on the choice of map.

What is invariant (does not depend on the map):

- Dimension — how many coordinates are needed locally
- Topology — holes, connectedness, orientability
- Curvature (in the Riemannian case) — intrinsic property of the metric

Coordinates are merely a means of description. The manifold itself exists without them.

Why manifolds are needed

Problem: We know how to work with $\mathbb{R}^n$ (coordinates, derivatives, integrals).
But real spaces are curved.

- Earth's surface — not a plane
- Space-time — curved by mass (GR)
- Configuration space of a robot — complex shape
- Space of all rotations — not $\mathbb{R}^3$

Solution: Manifold = space that is locally like $\mathbb{R}^n$
Globally can be any shape, but at each point — "flat"

Analogy: World atlas. Earth is round, but each page of the atlas is flat.
Sufficiently many maps cover the entire Earth.

Key idea:

We cannot work with curved space directly.
But we can break it into pieces, each of which "straightens out" into $\mathbb{R}^n$.
Then coordinate these pieces with each other.

Critical distinction: topology vs geometry

Manifold (topological + smooth) = only shape

- Can talk about tangent spaces, derivatives, differential forms
- Cannot talk about lengths, angles, distances.
- Sphere and ellipsoid — identical manifolds (diffeomorphic)

Riemannian manifold = manifold + metric g_{ij}

- Can measure lengths, angles, areas
- Sphere and ellipsoid — different (different metrics)

Common engineer's mistake: Curvilinear coordinates in $\mathbb{R}^3$
(cylindrical, spherical) — this is not curvature of space.
Space is flat ($\mathbb{R}^3$), just coordinates are "curved".
Curvature is a property of the metric, not coordinates.

Chart — local coordinates

```
+-----------------------------------------------------------------+
| DEFINITION: CHART                                               |
+-----------------------------------------------------------------+
|                                                                 |
| A chart on a topological space M is a pair (U, φ), where:       |
|                                                                 |
| • U ⊆ M — open set (domain of the chart)                        |
| • φ: U → ℝⁿ — homeomorphism onto an open subset of ℝⁿ           |
|                                                                 |
| φ is called the coordinate map (or parameterization).           |
|                                                                 |
+-----------------------------------------------------------------+
```

Geometric meaning:

```
       M (manifold)                 ℝⁿ (coordinate space)

          +-----+                     y
         /   U   \        φ          ↑
        |    ●p   |   ---------→     |  ●φ(p)
         \       /                   |
          +-----+                    +------→ x
```

A chart "straightens out" a piece of the manifold into flat space.
A point $p \in U$ receives coordinates $\varphi(p) = (x^1, x^2, ..., x^n) \in \mathbb{R}^n$.

Why homeomorphism:

PROPERTY	WHAT IT MEANS
φ bijective	Each point of U ↔ exactly one point in $\mathbb{R}^n$
φ continuous	Close points remain close
φ^{-1} continuous	Can return back

Examples of charts:

MANIFOLD	CHART	PECULIARITY
$\mathbb{R}$	$U=\mathbb{R}$, $\varphi=id$	One chart suffices
S^1	$U_1=S^1\setminus\{(-1,0)\}$, φ_1=angle θ	One chart does not cover.
	Need second: $U_2=S^1\setminus\{(1,0)\}$	→ need atlas of 2 charts

Visualization for S¹:

```
        ●--------------●                  -----●-----
       /    removed     \      φ₁           −π   0   π
      ●        ×         ●    ----→
       \     (−1,0)     /
        ●--------------●
```

Atlas – a set of charts covering the entire manifold

DEFINITION: ATLAS

An atlas on M is a set of charts $\mathcal{A} = \{(U_\alpha, \varphi_\alpha)\}_{\alpha\in A}$ such that:

$$\bigcup_{\alpha} U_\alpha = M$$

(charts cover the entire manifold)

Example: Atlas for S¹ (minimal)

Chart 1: $U_1 = S^1 \setminus \{(-1, 0)\}$, $\varphi_1(\cos\theta, \sin\theta) = \theta \in (-\pi, \pi)$

Chart 2: $U_2 = S^1 \setminus \{(1, 0)\}$, $\varphi_2(\cos\theta, \sin\theta) = \theta \in (0, 2\pi)$

```
         Chart 1                         Chart 2
     ●--------------●               ●--------------●
    /                \             /                \
   ●        ×         ●           ●        ×         ●
    \    (−1,0)      /             \     (1,0)      /
     ●--------------●               ●--------------●
```

$U_1 \cup U_2 = S^1$ ✓ (together cover the entire circle)

Analogy with geography:

- Map of Europe covers Europe, but not Australia
- Map of Australia covers Australia, but not Europe
- Together = entire world (atlas)
- Where maps overlap — one can transition from one to another

Transition functions — how to glue charts

Problem: If point p lies in two charts (U_1, φ_1) and (U_2, φ_2), it has two sets of coordinates: $\varphi_1(p)$ and $\varphi_2(p)$. How are they related?

DEFINITION: TRANSITION FUNCTION

Transition function from chart (U_1, φ_1) to chart (U_2, φ_2):

$$\psi_{12} = \varphi_2 \circ \varphi_1^{-1} : \varphi_1(U_1 \cap U_2) \to \varphi_2(U_1 \cap U_2)$$

This is a map $\mathbb{R}^n \to \mathbb{R}^n$ (between coordinate spaces)

Diagram:

```
                    M
                 /     \
              U₁        U₂
                 \  p  /
             φ₁ ↓ ● ↓ φ₂
                 /     \
          φ₁(p)          φ₂(p)
            ∈ℝⁿ ------→ ∈ℝⁿ
                  ψ₁₂
```

ψ_{12} translates coordinates from the first system to the second.

Example: Transition function on S^1

Intersection: $U_1 \cap U_2 = S^1 \setminus \{(-1,0), (1,0)\}$ — two arcs

ARC	φ_1 gives	φ_2 gives	$\psi_{12}(\theta)$ =
Upper (sin>0)	$\theta \in (0, \pi)$	$\theta \in (0, \pi)$	θ (identity)
Lower (sin<0)	$\theta \in (-\pi, 0)$	$\theta \in (\pi, 2\pi)$	$\theta + 2\pi$

Key fact: Transition function $\psi_{12}: \mathbb{R}^n \to \mathbb{R}^n$ — an ordinary map.
We know how to work with it. If we require smoothness of ψ_{12} → smooth structure.

Smooth structure — what transitions are allowed

```
+-------------------------------------------------------------+
| DEFINITION: SMOOTH ATLAS                                    |
+-------------------------------------------------------------+
|                                                             |
| An atlas is called Cᵏ-smooth (or smooth of class Cᵏ), if   |
| all transition functions ψαβ are Cᵏ-smooth (k times        |
| continuously differentiable).                               |
|                                                             |
| C∞-smooth atlas: all transitions are infinitely differentiable. |
| Such an atlas is called simply "smooth".                    |
|                                                             |
+-------------------------------------------------------------+
```

```
+-------------------------------------------------------------------------+
| DEFINITION: SMOOTH MANIFOLD                                             |
+-------------------------------------------------------------------------+
|                                                                         |
| A smooth manifold of dimension n is a topological space M               |
| with a maximal C∞-smooth atlas.                                         |
|                                                                         |
| technical requirements (usually assumed):                               |
| • Hausdorff property: distinct points are separable by neighborhoods    |
| • Second-countable: topology has a countable basis                      |
| (Without them pathologies arise, for example "long line")               |
|                                                                         |
| (Maximal = contains all charts compatible with the given ones)          |
|                                                                         |
+-------------------------------------------------------------------------+
```

Hierarchy of structures:

TYPE OF MANIFOLD	TRANSITION FUNCTIONS	WHAT CAN BE DONE
Topological	Homeomorphisms (continuous)	Talk about closeness, connectedness, but not about derivatives
C^1-smooth	Differentiable (1 derivative)	Define tangent vectors
C^∞-smooth (simply "smooth")	Infinitely differentiable	Differentiate as many times as desired
Analytic (C^ω)	Analytic (Taylor series)	Expand into Taylor series

Why smoothness is needed:

Problem: define the derivative of a function $f: M \to \mathbb{R}$ on a manifold

Answer: We work in coordinates.
1. Choose a chart (U, φ) around point p
2. Consider $f \circ \varphi^{-1}: \mathbb{R}^n \to \mathbb{R}$ (an ordinary function)
3. Differentiate as usual

Problem: What if we choose a different chart?

Solution: If transition functions are smooth, then:
$\partial f/\partial x' = (\partial x/\partial x') \cdot (\partial f/\partial x)$ (chain rule)
Derivatives in different charts are coordinated.

Detailed example: atlas for the sphere S^2

$S^2 = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 1\}$

Method 1: stereographic projection (minimal atlas – 2 charts)

Chart 1: Projection from the north pole $N = (0, 0, 1)$

N ●

```
          /|\
         / | \           Project point P onto the plane z = 0
        /  |  \          through the line NP
       /   |   \
      ●----+----●
     P     |    φ₁(P)
      \    |   /
       \   |  /
        ---+---  plane z = 0
```

$U_1 = S^2 \setminus \{N\}$ (entire sphere without the north pole)

$\varphi_1(x, y, z) = (x/(1-z), y/(1-z))$

$\varphi_1^{-1}(u, v) = (2u/(1+u^2+v^2), 2v/(1+u^2+v^2), (u^2+v^2-1)/(1+u^2+v^2))$

Chart 2: Projection from the south pole $S = (0, 0, -1)$

$U_2 = S^2 \setminus \{S\}$ (entire sphere without the south pole)

$\varphi_2(x, y, z) = (x/(1+z), y/(1+z))$

Transition function $\psi_{12} = \varphi_2 \circ \varphi_1^{-1}$:

On the intersection $U_1 \cap U_2 = S^2 \setminus \{N, S\}$:

$\psi_{12}(u, v) = (u, v)/(u^2 + v^2) = (u/(u^2+v^2), v/(u^2+v^2))$

This is inversion with respect to the unit circle.

Smoothness check: ψ_{12} is infinitely differentiable on $\mathbb{R}^2 \setminus \{0\}$. ✓
(The origin corresponds to the poles, which are not in the intersection)

Method 2: projections onto coordinate planes (6 charts)

CHART	DOMAIN $U_k\pm$	MAP φ	WHICH PART
U_1^+	$z > 0$	$\varphi(x,y,z) = (x,y)$	Upper hemisphere
U_1^-	$z < 0$	$\varphi(x,y,z) = (x,y)$	Lower hemisphere
U_2^+	$y > 0$	$\varphi(x,y,z) = (x,z)$	Front hemisphere
U_2^-	$y < 0$	$\varphi(x,y,z) = (x,z)$	Back hemisphere
U_3^+	$x > 0$	$\varphi(x,y,z) = (y,z)$	Right hemisphere
U_3^-	$x < 0$	$\varphi(x,y,z) = (y,z)$	Left hemisphere

Analogy: a world atlas uses different projections for different regions.

Remark: There does not exist an atlas of S^2 from a single chart.
This follows from the compactness of S^2 and the fact that $\mathbb{R}^2$ is non-compact.
Topologically $S^2 \neq \mathbb{R}^2$ (one is compact, the other is not).

Hierarchy of spaces — from simple to complex

```
  Topological space (X, τ)
    | add: local Euclideanness + Hausdorffness
    ↓
  Topological manifold
    | add: smooth structure (C∞ transition functions)
    ↓
  Smooth manifold ← main object of differential geometry
    | add: metric gᵢⱼ (way to measure distances)
    ↓
  Riemannian manifold
    | add: physics (Einstein equations)
    ↓
  Spacetime of GR
```

At each level structure is added, but the previous is not lost.

Tangent space T_pM — local linearity

Problem: define the notion of velocity on a curved space

On $\mathbb{R}^n$ it's simple: velocity = vector = arrow.
On a sphere: where does the arrow point? It "flies out" from the sphere.

Definition (intuitive):
A tangent vector at point p is the velocity of a curve passing through p.

Definition (formal via curves):
Let $\gamma: (-\varepsilon, \varepsilon) \to M$ be a smooth curve, $\gamma(0) = p$.
Tangent vector: $v = \gamma'(0) = d\gamma/dt|_{t=0}$

The set of all tangent vectors at point p forms the tangent space T_pM.

Three equivalent definitions of tangent vector

1. Geometric (equivalence classes of curves):
   Two curves γ_1, γ_2 with $\gamma_1(0) = \gamma_2(0) = p$ are equivalent if
   $(\varphi \circ \gamma_1)'(0) = (\varphi \circ \gamma_2)'(0)$ in any chart φ.
   Tangent vector = equivalence class of curves.
   Intuition: the curve itself doesn't matter, only its "velocity" at p matters.

2. Algebraic (via derivations):
   Tangent vector = linear operator $v: C^\infty(M) \to \mathbb{R}$ with the Leibniz rule.

3. Coordinate:
   Tangent vector = set of numbers $(v^1, ..., v^n)$, transforming
   by the rule $v'^i = (\partial x'^i/\partial x^j)v^j$ under coordinate change.

All three definitions are equivalent — they give the same $T_pM \cong \mathbb{R}^n$.

Visualization:

```
                   TₚM (plane tangent to the sphere at point p)
                  / \
                 /   \  ← velocity vectors live here
                /  ●  \
               /   p   \
           ---●---------●---
             /    ◠      \   ← sphere (manifold M)
            /             �\
```

In coordinates:

Let (U, φ) be a chart near p with coordinates $(x^1, ..., x^n)$.

Basis of T_pM: $\partial/\partial x^1|_p, \partial/\partial x^2|_p, ..., \partial/\partial x^n|_p$

Any vector $v \in T_pM$: $v = v^i \, \partial/\partial x^i|_p$ (sum over i)

Important: Why $\partial/\partial x^i$ are these "vectors"?

In linear algebra a vector is an "arrow" or a "column of numbers".
On a manifold this doesn't work: an arrow "flies out" from a curved surface.

Solution: Define a vector as differentiation.

A tangent vector v at point p is a linear operator $v: C\infty(M) \to \mathbb{R}$, satisfying the Leibniz rule:

$$v(fg) = v(f)\cdot g(p) + f(p)\cdot v(g)$$

$\partial/\partial x^i|_p$ is an operator: $f \mapsto \partial f/\partial x^i|_p$ (partial derivative at point p).
It satisfies Leibniz, therefore — a tangent vector.

Intuition: A vector is a "direction of change of functions".
Arrow → "how fast and where does the value of a function change when moving".

Key facts:

FACT	CONSEQUENCE
$T_pM \cong \mathbb{R}^n$	Linear algebra works.
$\dim(T_pM) = \dim(M)$	Dimension is preserved
Different points → different T_pM	Cannot add $v \in T_pM$ and $w \in T_\gamma M$
Basis: $\partial/\partial x^i$	$v = v^i \, \partial/\partial x^i$ (sum over i)

PHYSICAL QUANTITY	MATHEMATICAL OBJECT
Position	Point $p \in M$
Velocity	Vector $v \in T_pM$
Acceleration	Requires connection (comparison of different T_pM)

Smooth mappings of manifolds

> **DEFINITION: SMOOTH MAPPING**
>
> A mapping $f: M \to N$ between smooth manifolds is called smooth, if for any charts (U, φ) on M and (V, ψ) on N the composition $\psi \circ f \circ \varphi^{-1} : \mathbb{R}^m \to \mathbb{R}^n$ is smooth (as an ordinary function).

Diagram:

$$\begin{tikzcd}
M \arrow[r, "f"] \arrow[d, "\varphi"'] & N \arrow[d, "\psi"] \\
\mathbb{R}^m \arrow[r, "\psi\circ f\circ\varphi^{-1}", "\text{(smooth)}"'] & \mathbb{R}^n
\end{tikzcd}$$

Differential of a mapping:

A smooth $f: M \to N$ generates a linear mapping $df_p: T_pM \to T_{f(p)}N$ (differential or pushforward)

This is a "linear approximation" of f near point p.

> **DEFINITION: DIFFEOMORPHISM**
>
> A diffeomorphism is a smooth bijection $f: M \to N$ such that f^{-1} is also smooth. This is an "isomorphism" in the category of smooth manifolds.
>
> If there exists a diffeomorphism $M \to N$, we write $M \cong N$ (diffeomorphic).

Hierarchy of "sameness":

```
  Homeomorphism: M ≈ N (topologically)  — same "shape"
  Diffeomorphism: M ≅ N (smoothly)      — same "smooth structure"
  Isometry: M = N (metrically)          — same "distances"

Example: S² ≈ ellipsoid (homeomorphic)
        S² ≅ ellipsoid (diffeomorphic)
        S² ≠ ellipsoid (not isometric — different curvature)
```

Examples of manifolds

MANIFOLD	dim	WHAT IT IS / WHERE IT OCCURS
Circle S^1	1	Angle of rotation, wave phase, time on a clock Minimal atlas: 2 charts $\pi_1(S^1) = \mathbb{Z}$ (loops wind around)
Sphere S^2	2	Surface of Earth, direction in space Minimal atlas: 2 charts (stereographic projections) $\pi_1(S^2) = 0$, $H_2(S^2) = \mathbb{Z}$
Torus T^2	2	Donut, periodic boundary conditions $= S^1 \times S^1$ (two independent angles) $\pi_1(T^2) = \mathbb{Z} \times \mathbb{Z}$ (two independent loops)
$\mathbb{R}P^n$ (project.)	n	"Directions of lines" (line = pair of opposite points) $\mathbb{R}P^2$ non-orientable. (like Möbius strip) $\pi_1(\mathbb{R}P^n) = \mathbb{Z}/2$ for $n \geq 2$
SO(3)	3	All rotations in 3D (orientation of a body) $\cong \mathbb{R}P^3$, not a sphere. $\pi_1(SO(3)) = \mathbb{Z}/2$ (belt trick)
$GL(n,\mathbb{R})$	n^2	Invertible matrices n×n Open subset of $\mathbb{R}^{n^2}$ ($\det \neq 0$) Lie group — manifold + group.
Spacetime	4	3 spatial + 1 temporal dimension Curved by mass → gravity. Pseudo-Riemannian (metric not positive definite)

Connectedness — how to compare vectors at different points

Problem:

On $\mathbb{R}^n$ one can compare vectors at different points — simply "translate".
On a curved space this is impossible without additional structure.

A vector in Moscow and a vector in Sydney — how to compare if Earth is round?

Solution: Connection

Connection = rule defining "parallel transport" of a vector along a curve.

Visualization of parallel transport:

Plane (curvature = 0): sphere (curvature > 0):

```
    B ●----------● C                  B ●-----● C
      |  ----→   |                     /   →   \
      |  ----→   |                    / ↗   ↘   \
      |  ----→   |                   /           \
    A ●----------● D               A ●-----------● D
      ↑          ↑                   ↖           ↗
    Vector    Same                 Vector    different
    at A      vector.              at A      vector.
```

On plane: transport along contour A→B→C→D→A returns the same vector.
On sphere: vector rotates. Angle of rotation = curvature × area.

Example on sphere:

```
                  North pole
                       ●
                      /|\
                     / | \
                    /  ↓  \
                   /   |   \      1. Start at equator with vector →
    Equator ------●----|----●     2. Go north to pole
                  |    |    |     3. Turn 90° and go south
                  |    |    |     4. Vector now points ↑
                  |    |    |        (rotated by 90°.)
                  ●----+----●
```

Explanation of terms (before reading formulas):

Γ(TM) — "sections of tangent bundle" — these are simply vector fields.
Vector field X is a choice of vector $X(p) \in T_pM$ at each point $p \in M$.

Example: wind field on Earth — at each point a velocity vector is given.

(Xf) or $X(f)$ — action of vector field X on function $f: M \to \mathbb{R}$.
This is derivative of f in direction X: $(Xf)(p)$ = rate of change of f at point p, if moving in direction $X(p)$.
In coordinates: if $X = X^i \partial/\partial x^i$, then $Xf = X^i(\partial f/\partial x^i)$.
Example: X = "eastward", f = temperature. Xf = "how fast it warms when moving eastward".

$[X,Y]$ — Lie bracket (commutator) of two vector fields.
Definition: $[X,Y](f) = X(Y(f)) - Y(X(f))$
Meaning: $[X,Y]$ measures how much the operations "shift by X" and "shift by Y" do not commute.
If $[X,Y] = 0$, then X and Y are "compatible" — order of movement doesn't matter.
Example: $[\partial/\partial x, \partial/\partial y] = 0$ (shift by x and by y commute).

Mathematical definition:

Connection ∇ is a mapping
$\nabla: \Gamma(TM) \times \Gamma(TM) \to \Gamma(TM)$
$(X, Y) \mapsto \nabla_X Y$ (derivative of Y in direction X)

satisfying:
- $\nabla_{fX}Y = f\nabla_X Y$ (linear in X with coefficients from functions)
- $\nabla_X(Y+Z) = \nabla_X Y + \nabla_X Z$ (linear in Y)
- $\nabla_X(fY) = (Xf)Y + f\nabla_X Y$ (Leibniz rule)

In coordinates:

$$\nabla_{\partial/\partial x^i}(\partial/\partial x^j) = \Gamma^k{}_{ij} (\partial/\partial x^k)$$

$\Gamma^k{}_{ij}$ — Christoffel symbols (n^3 functions)

Meaning of Christoffel symbols:

$\Gamma^k{}_{ij}$ shows: when I move in direction i, how does basis vector $\partial/\partial x^j$ "rotate" in direction k?

On plane: $\Gamma = 0$ everywhere (basis vectors don't change when moving)
On sphere: $\Gamma \neq 0$ (when moving "direction to north" rotates)

Analogy: Imagine carrying a long pole along a surface.
On plane pole preserves direction by itself.
On sphere one needs to constantly "steer" to preserve parallelness.
$\Gamma^k{}_{ij}$ = magnitude of this "steering".

For metric connection (Levi–Civita):
$$\Gamma^k{}_{ij} = \tfrac{1}{2}g^{kl}(\partial g_i{}^l/\partial x^j + \partial g_j{}^l/\partial x^i - \partial g_{ij}/\partial x^l)$$

This formula says: Christoffel symbols are completely determined by metric g_{ij} and its derivatives. Geometry → connection automatically.

Critically important: $\Gamma^k{}_{ij}$ — not a tensor.

Under change of coordinates an additional term with second derivatives appears:

$$\Gamma'^{k}{}_{ij} = \underset{\overset{\uparrow}{\text{tensor part}}}{(\partial x'^{k}/\partial x^{l})(\partial x^{m}/\partial x'^{i})(\partial x^{n}/\partial x'^{j})\Gamma^{l}{}_{mn}} + \underset{\overset{\uparrow}{\text{NONtensor addition}}}{(\partial x'^{k}/\partial x^{l})(\partial^{2}x^{l}/\partial x'^{i}\partial x'^{j})}$$

This is precisely why Γ depends on choice of coordinates.
On plane in Cartesian $\Gamma = 0$, but in polar $\Gamma \neq 0$.
But plane remained flat – only coordinates changed.

Consequence: Cannot say "$\Gamma = 0$ means space is flat".
Can say: "There exist coordinates in which $\Gamma = 0$" $\Longleftrightarrow$ flat.

Types of connections:

TYPE	PROPERTY
Metric	Preserves scalar product under transport
Torsion-free	$\nabla_X Y - \nabla_Y X = [X,Y]$
Levi–Civita	Metric + torsion-free (unique)

Curvature – measure of non-flatness

How to detect curvature (without going into the ambient space):

METHOD	WHAT HAPPENS
Parallel transport along closed path	Plane: vector returned the same Curved: vector rotated
Sum of angles of triangle	Plane: $\alpha+\beta+\gamma = 180°$ Sphere: $\alpha+\beta+\gamma > 180°$ (positive curvature) Saddle: $\alpha+\beta+\gamma < 180°$ (negative curvature)
Riemann tensor (formal def.)	$R^{i}{}_{jkl}$ = rotation of v^{i} when traversing inf. small parallelogram in directions k,l

Physics: curvature = gravity (GR)

CAUSE	CONSEQUENCE
Mass/energy	Curves spacetime
Bodies move along geodesic	Looks like attraction = geometry.
Einstein equation	$R\mu\nu - \frac{1}{2}g\mu\nu R = 8\pi G\cdot T\mu\nu$ (curvature = matter)

Connection with other areas

AREA	HOW IT CONNECTS TO MANIFOLDS
Topology	Manifold = top. space + smooth structure Invariants π_1, H_n, χ work for manifolds
Groups	Lie groups = manifolds + group structure SO(3), SU(2), GL(n) — both groups and manifolds Lie algebra = T_e (tangent space at identity)
Lin. algebra	T_pM — vector space at each point Locally all linear algebra works. Connection = way to "join" T_pM at different points
Duality	T^*_pM — cotangent (where df, dp live) Vector: where to move. Covector: how to measure Differential forms ∈ $\Lambda^k T^*M$
Tensors	Tensor field = tensor at each point Metric g_{ij} — field (0,2). Curvature $R^i{}_{jkl}$ — field (1,3)
Noether	Symmetries of manifold → conservation laws Shifts → momentum. Rotations → angular momentum
Categories	Man = category of smooth manifolds Functor T: Man → VectBund (tangent bundle)

Summary: manifold as central object

```
                         Manifold M
                              |
         +--------------------+--------------------+
```

```
         |                     |                     |
         ↓                     ↓                     ↓
    Locally              Globally             Structures
    ≈ ℝⁿ                 (topology)           on manifold
    (charts, atlas)      (π₁, Hₙ, χ)          (metric, connection)
         |                     |                     |
         ↓                     ↓                     ↓
    Can                  Distinguish          Measure
    differentiate        spaces               distances
    (TₚM, df)            (invariants)         (gᵢⱼ, curvature)
```

Main idea:
Manifold is a way to locally use familiar tools (coordinates, derivatives, integrals) on globally complex spaces.

Practical applications:

- Physics: spacetime, configuration space
- Robotics: state space (positions and orientations)
- Machine learning: data manifold (manifold hypothesis)
- Optimization: optimization on manifolds (Riemannian optimization)

Applied example: turbine blade surface

Problem: Calculate heat transfer on the surface of a gas turbine blade. The surface has a complex curvilinear shape — no "natural" x, y.

```
           ╭───────────────╮
          /                 \        Surface S — 2D manifold,
         /                   \       embedded in ℝ³
        |      BLADE          |
         \                   /       Need local coordinates (u, v)
          \                 /        on the surface
           ╰───────────────╯
```

Local coordinates:

u — along the profile chord (0 = leading edge, 1 = trailing edge)
v — along the span (0 = root, 1 = tip)

Point on the surface: $r(u, v) = (x(u,v), y(u,v), z(u,v))$

Metric on the surface:
How to measure distance between two points on the blade?

$$ds^2 = g_{11}\, du^2 + 2g_{12}\, du\, dv + g_{22}\, dv^2$$

where $g_{ij} = \partial r/\partial u^i \cdot \partial r/\partial u^j$ — components of the metric tensor

Surface element area:
$dA = \sqrt{(g_{11}g_{22} - g_{12}^2)}\, du\, dv = \sqrt{\det(g)}\, du\, dv$

Heat flux integral:
$Q = \iint_S q(u,v)\, dA = \iint q(u,v) \sqrt{\det(g)}\, du\, dv$

The metric "knows" how areas are distorted in the transition from flat (u,v)-space to the real curved surface.

Numerical example:
If in flat coordinates $q = 50$ kW/m^2 = const,
but the blade is curved such that $\sqrt{\det(g)} \approx 1.3$ (area is "stretched"),
then the real flux $Q = 50 \times 1.3 \times A_{(uv)} = 65$ kW per unit (u,v)-area

Moral: Manifold = way to work with curvilinear surfaces through local coordinates. The metric g_{ij} "translates" formulas from the flat case to the curvilinear one. Without this language — calculations for turbines, heat exchangers of complex shape, aerodynamic surfaces are impossible.

Connection and curvature — how to compare vectors at different points

Problem: how to compare vectors at different points?

In $\mathbb{R}^n$ this is trivial: parallel transport and compare.

On a curved surface — unclear.

```
             vector at point A
                  ↗
                A ------------------- B
               /                       \      vector at point B
              /    curved               \          ↘
             /      surface              \
```

Vector at A and vector at B live in different tangent spaces.
They cannot be directly added or subtracted.

Need a rule: how to "transport" a vector from A to B for comparison.

Connection = rule of parallel transport

Connection ∇ is a rule for how to differentiate vector fields.

$\nabla_X Y$ = "derivative of field Y in direction X"

Covariant derivative:
In coordinates: $\nabla_{\partial_i} \partial_j = \Gamma^k_{ij} \partial_k$

Γ^k_{ij} — Christoffel symbols (connection coefficients)

For a metric (Levi–Civita connection):

$$\Gamma^k_{ij} = \tfrac{1}{2}g^{kl}(\partial_i g_j^l + \partial_j g_i^l - \partial^l g_{ij})$$

This connection:

- Preserves the metric: $\nabla g = 0$
- Torsion-free: $\Gamma^k_{ij} = \Gamma^k_{ji}$
- Unique with these properties.

Parallel transport — visualization

A vector is "parallel transported" along a curve $\gamma(t)$ if:

$$\nabla_{\dot\gamma} V = 0 \quad \text{(covariant derivative along the curve = 0)}$$

On a plane: Parallel transport is obvious — direction is preserved.

On a sphere — surprise.

```
       North pole
             N
            /|\
  vector → / | \
          /  |  \   Transport vector along a closed path:
         A---+---B  N → A → B → N
             |
             |      Vector has rotated
```

Rotation angle = triangle area / R^2 (where R — sphere radius)

This rotation — manifestation of curvature.

Curvature — measure of "non-flatness"

Riemann curvature tensor:

$$R(X, Y)Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z$$

What does this measure:
Transport vector Z first in direction X, then Y.
Then in direction Y, then X.
Difference = R(X, Y)Z.

If R = 0 everywhere — space is flat (locally like $\mathbb{R}^n$).

In coordinates:

$$R^k{}_{lij} = \partial_i\Gamma^k_{jl} - \partial_j\Gamma^k_{il} + \Gamma^k_{im}\Gamma^m_{jl} - \Gamma^k_{jm}\Gamma^m_{il}$$

Important contractions:

- Ricci tensor: $R_{ij} = R^k{}_{ikj}$
- Scalar curvature: $R = g^{ij}R_{ij}$

For a surface (2d):

$K = R/2$ – Gaussian curvature
$K > 0$: sphere (positive curvature)
$K = 0$: plane, cylinder
$K < 0$: saddle (negative curvature)

Geodesics – "straight lines" on curved space

Definition: Geodesic – curve that parallel transports its velocity vector:

$$\nabla_{\dot\gamma} \dot\gamma = 0$$

In coordinates (geodesic equation):

$$d^2x^k/dt^2 + \Gamma^k_{ij} (dx^i/dt)(dx^j/dt) = 0$$

Examples:

- On a plane: straight lines
- On a sphere: great circles (equator, meridians)
- In GR: trajectories of free particles in a gravitational field

Equivalent definitions:

1. Curve of shortest length (locally)
2. Curve extremizing $\int ds$
3. Curve with zero geodesic acceleration

Important: geodesic is a local minimum of length, not global.
On a sphere from A to B one can go by a "short" or "long" path – both are geodesics.

Gauss' Theorema Egregium – a profound result

Gaussian curvature K depends only on the metric g_{ij}, but not on how the surface is embedded in $\mathbb{R}^3$.

Consequence: It is impossible to draw a map of a sphere on a plane without distortions.

Sphere: $K = 1/R^2 > 0$
Plane: $K = 0$

Any world map necessarily distorts either angles or areas.
(Mercator distorts areas, equal-area projections – angles)

But a cylinder can be unrolled onto a plane without distortions.
Cylinder: $K = 0$ (one of the principal curvatures equals zero)

Connection with physics: general relativity

Einstein's equations:

$$R_{ij} - \frac{1}{2}Rg_{ij} = (8\pi G/c^4)\,T_{ij}$$

Left side: geometry (curvature of spacetime)
Right side: matter (energy-momentum tensor)

Trajectories of freely falling bodies = geodesics in curved spacetime. Gravitation is not a force, but geometry.

Bundles — spaces over spaces

Motivation: tangent vectors to all points

On a manifold M at each point p there is a tangent space T_pM.

Question: How to organize all tangent spaces into a single structure?

Answer: We unite them into a bundle:

$$TM = \bigsqcup_{p\in M} T_pM$$

This is a new manifold — the tangent bundle.
$\dim(TM) = 2\cdot\dim(M)$

Definition: bundle

A bundle is a triple (E, π, B), where:

- E — total space (the entire bundle)
- B — base (space over which the bundle is fibered)
- $\pi: E \to B$ — projection

For each point $b \in B$ the preimage $\pi^{-1}(b) = F$ is called the fiber.

Visualization:

```
E (total space)
|
|    F₁    F₂    F₃    ← fibers (all isomorphic to F)
|    |     |     |
|    |     |     |
▼    ▼     ▼     ▼
----●-----●-----●----  B (base)
    b₁    b₂    b₃
```

Locally: $E \cong U \times F$ (product of neighborhood $U \subset B$ and fiber F)
Globally: may be twisted.

Trivial vs nontrivial bundle

Trivial: $E = B \times F$ globally (simply a product)

Example: cylinder = $S^1 \times \mathbb{R}$

```
+-----------+
|           |
|  cylinder |     Base: circle S¹
|           |     Fiber: line ℝ
+-----------+
```

Nontrivial: globally does not decompose into a product

Example: Möbius strip.

```
+-----╲╱-----+
|     ╲╱     |    Base: S¹ (circle)
|      ╳     |    Fiber: segment [-1, 1]
|     ╱╲     |    But globally — twisted.
+-----╱╲-----+
```

Going around the base in a circle, the fiber flips.

Principal types of bundles

BUNDLE	FIBER F	EXAMPLE
Tangent TM	$\mathbb{R}^n$	All velocities at a point
Cotangent T*M	$\mathbb{R}^n$	All momenta at a point
Tensor	Tensors	All tensors of type (p,q) at pt
Principal (G-bundle)	Group G	Gauge fields in physics
Frame bundle	GL(n)	All bases of tangent space

Section of a bundle = field on a manifold

A section $s: B \to E$ is a map where $\pi(s(b)) = b$ for all b.
(We choose one point from each fiber)

- Section of TM = vector field on M
- Section of T*M = 1-form on M
- Section of tensor bundle = tensor field

Example: Wind velocity field — section of tangent bundle TS^2
(S^2 = surface of Earth)

Hairy Ball Theorem:

On S^2 there does not exist a continuous nonzero vector field.
(Cannot comb a hedgehog — somewhere there will be a "cowlick")

Reason: tangent bundle TS^2 is nontrivial.

Connection on a bundle — how to transport between fibers

Connection on a bundle = rule for how to transport an element of a fiber along a path on the base.

For tangent bundle:
Connection = Christoffel symbols $\Gamma^k{}_{ij}$
This is what we defined above

For principal g-bundle:
Connection = 1-form A with values in Lie algebra g
In physics: A — gauge potential.

Curvature of bundle:
$F = dA + A \wedge A$ (curvature 2-form)
In physics: F — field strength.

Electromagnetism: $G = U(1)$, A = potential, F = E, B
Standard model: $G = SU(3)\times SU(2)\times U(1)$

Characteristic classes — topology of bundles

Idea: measuring "twistedness" of bundle

Bundle can be trivial ($E = B \times F$) or nontrivial.
Characteristic classes are numerical invariants measuring
"how much the bundle is twisted".

They live in cohomology of base: $c_i \in H^*(B)$

If class is nonzero — bundle is definitely nontrivial.
If zero — may be trivial, or may not (necessary, not sufficient)

Chern classes — for complex bundles

For complex vector bundle $E \to B$ are defined:

$c_0(E) = 1$, $c_1(E)$, $c_2(E)$, ., $c_n(E)$ where $n = \dim_{\mathbb{C}}(E)$

$c_k(E) \in H^{2k}(B; \mathbb{Z})$ — Chern class of degree k

Key properties:

- $c(E \oplus F) = c(E) \cup c(F)$ (additivity with respect to direct sum)
- $c(E^*) = \bar{c}(E)$ (classes of conjugate bundle)
- For line bundle L: $c_1(L)$ = Euler class

Example: Tangent bundle of sphere

$c_1(TS^2) = 2 \in H^2(S^2) \cong \mathbb{Z}$

This is related to "hairy ball" theorem and Euler characteristic.

Pontryagin and Euler classes — for real bundles

For real bundle $E \to B$:

$p_k(E) \in H^{4k}(B; \mathbb{Z})$ — Pontryagin class

$e(E) \in H^n(B; \mathbb{Z})$ — Euler class (if E is oriented, $n = \operatorname{rank} E$)

Gauss–Bonnet–Chern theorem:

$\int_M e(TM) = \chi(M)$ (integral of Euler class = Euler characteristic)

This is generalization of Gauss–Bonnet theorem to any dimensions.

Applications in physics

Quantization of magnetic monopole (Dirac):

Magnetic charge g is quantized: $eg = n\hbar/2$

Reason: first Chern class of line bundle over S^2 is integral.

Anomalies in quantum field theory:

Gauge anomalies are related to Chern classes of bundle of gauge group.

Topological insulators:

Chern classes of electron bands determine quantum Hall effect.

$c_1 = 0, \pm 1, \pm 2, \ldots \to$ Hall conductivity $= c_1 \cdot e^2/h$

Projective spaces — geometry of "directions"

Intuition: What is projective space

Imagine we stand at origin and look in different directions.
Each direction of gaze is a point of projective space.

Formally: Points (x, y, z) and $(2x, 2y, 2z)$ define one direction.
We "glue" all points on one ray from origin.

Definition

Real projective space $\mathbb{R}P^n$:

$$\mathbb{R}P^n = (\mathbb{R}^{n+1} \setminus \{0\}) / \sim, \quad \text{where} \quad x \sim y \Longleftrightarrow x = \lambda y \text{ for } \lambda \neq 0$$

In other words: $\mathbb{R}P^n$ = set of lines in $\mathbb{R}^{n+1}$ passing through 0.

Homogeneous coordinates:
  Point of $\mathbb{R}P^n$ is written as $[x_0 : x_1 : . : x_n]$
  This is equivalence class: $[x_0 : x_1 : . : x_n] = [\lambda x_0 : \lambda x_1 : . : \lambda x_n]$

Dimensions:
- $\mathbb{R}P^1$ — projective line (topologically = circle S^1)
- $\mathbb{R}P^2$ — projective plane (non-orientable surface)
- $\mathbb{R}P^3 \cong SO(3)$ — space of rotations in 3D

Visualization of $\mathbb{R}P^1$ and $\mathbb{R}P^2$

$\mathbb{R}P^1$ — projective line:

  Take $\mathbb{R}^2 \setminus \{0\}$ and glue opposite points $(x,y) \sim (-x,-y)$.

```
           y
           |    /
           |   /  ← this line = one point of ℝP¹
           |  /
     ------●------ x
          /|
         / |
```

  Can parametrize by slope: $[1 : t]$ for $t \in \mathbb{R}$, plus $[0 : 1]$ (vert.)
  This is $\mathbb{R} \cup \{\infty\}$ — line with added "infinity". Topologically = S^1.

$\mathbb{R}P^2$ — projective plane:

Take hemisphere and glue opposite points on equator:

```
        ┌-------┐
      /    ●      \      Interior of hemisphere — ordinary plane
     |      ●      |     Boundary (equator) — "infinity"
     |   ●    ●    |     But points A and A' on equator are glued.
      \   A    A' /
        └-------┘
```

Result: non-orientable surface (contains Möbius strip).
Cannot be embedded in ℝ³ without self-intersections.

Why Projective Spaces

1. parallel lines intersect "at infinity"

   In ordinary geometry, parallel lines do not intersect.
   In projective geometry — they intersect at an "infinitely distant point".

   This simplifies many theorems: Pappus' theorem, Desargues' theorem, duality.

2. computer graphics

   Homogeneous coordinates [x : y : z : w] — standard in 3D graphics.
   Point in 3D: [x : y : z : 1] → (x/1, y/1, z/1)
   Point "at infinity": [x : y : z : 0] — direction

   All transformations (translation, rotation, scaling, perspective) —
   are multiplication by a 4×4 matrix.

   Without homogeneous coordinates, translation is not a linear transformation.
   With them — linear. This is a huge simplification for GPUs.

3. algebraic geometry

   Many theorems are simpler in projective space:

   Bézout's theorem: Two curves of degrees m and n in $\mathbb{P}^2$ intersect
   in exactly mn points (counting multiplicities and points at infinity).

   In the ordinary plane: a line and parabola may not intersect (0, 1, or 2).
   In projective: a line (degree 1) and conic (degree 2) — always 2 points.

4. physics: state space

   In quantum mechanics, a pure state is a ray in Hilbert space.
   $|\psi\rangle$ and $\lambda|\psi\rangle$ — same state (global phase is not observable).

   For a qubit: state space = $\mathbb{C}P^1$ = Bloch sphere (S^2).

Complex Projective Space $\mathbb{C}P^n$

Analogous to the real case, but over $\mathbb{C}$:

$$\mathbb{C}P^n = (\mathbb{C}^{n+1} \setminus \{0\}) / \sim, \quad \text{where} \quad z \sim w \iff z = \lambda w \text{ for } \lambda \in \mathbb{C} \setminus \{0\}$$

Important cases:

- $\mathbb{C}P^1$ = Riemann sphere = $\mathbb{C} \cup \{\infty\}$
  This is the compactification of the complex plane.

- $\mathbb{C}P^n$ — compact Kähler manifold
  Central object of algebraic geometry

Connection to physics:
State space of a system with n+1 levels = $\mathbb{C}P^n$
Bloch sphere (qubit) = $\mathbb{C}P^1 \cong S^2$

Projective Spaces as Equivalence Classes

This is another example of construction via quotient:

$X = \mathbb{R}^{n+1} \setminus \{0\}$ (all nonzero vectors)
$x \sim y \iff x = \lambda y$ (lie on the same ray)
$\mathbb{R}P^n = X/\sim$ (set of classes = set of rays = set of lines)

Each equivalence class is a line through the origin (without the origin itself). We have "collapsed" each line into one point.

On manifolds we can define functions, vector fields, tensors.
But how to integrate? The ordinary integral $\int f\, dx$ requires coordinates, but we want an invariant definition.

Tensor Field — a Tensor at Each Point

We studied tensors as algebraic objects on a single vector space. But on a manifold, at each point there is its own tangent space T_pM. How to "glue together" tensors from different points?

A tensor field is a choice of tensor at each point of the manifold, varying smoothly from point to point.

EXAMPLE	WHAT IT IS
Scalar field	Function f: M → ℝ (temperature at each point)
Vector field	Vector v(p) ∈ T_pM at each point (velocity)
Metric g_{ij}	Scalar product at each T_pM
Curvature tensor	Measures "curvature" at each point

In terms of "object–observer": a tensor field is an object on the manifold. In each chart (observer's coordinate system) it is written in components $g_{ij}(x)$. Under change of chart, components transform by the tensor transformation law — but the field itself remains the same.

The metric g_{ij} — main example: it defines distances and angles, but its components depend on the choice of coordinates. Distance is an invariant.

Differential forms are objects that can be integrated on manifolds. They unify all integration theorems (Green's, Stokes', Gauss') into one — and this is the language of modern physics.

Differential Forms — Language of Modern Physics

Forms as a View of Space

Differential forms are objects that "measure" pieces of space:

- 1-form measures curves (work of a force along a path)
- 2-form measures surfaces (flux through an area)
- 3-form measures volumes (mass in a region)

Forms are "dimension detectors": a k-form senses k-dimensional objects. They allow integration on manifolds without coordinates.

Why Differential Forms Are Needed

Problem: Integrals depend on what we are integrating

- Along a curve we integrate work: $\int F\cdot dr$ — this is something one-dimensional
- Over a surface — flux: $\iint F\cdot dS$ — this is something two-dimensional
- Over a volume — density: $\iiint \rho\, dV$ — this is something three-dimensional

Task: formalize the notion of "oriented area"

Answer: differential forms — objects created for integration

- 0-form — function (integral = value at a point)
- 1-form — integrated over curves
- 2-form — integrated over surfaces
- n-form — integrated over n-dimensional regions

Bonus: Stokes' theorem, Green's theorem, Gauss' theorem — all become one theorem.

What a k-form measures — physical table

A k-form is a "detector" for k-dimensional objects in space.

k	WHAT IT MEASURES	PHYSICAL EXAMPLE	MATHEMATICALLY
0-form	point (scalar at it)	Temperature $T(x,y,z)$ Pressure $p(x,y,z)$ Concentration c	Simply a function $f: M \to \mathbb{R}$
1-form	curve (integral along path)	Work of force $\int F \cdot dr$ Circulation Voltage $\int E \cdot dl$	$\omega = F_x dx + F_\gamma dy + F_z dz$ Acts on tangent vector to curve
2-form	surface (integral over area)	Flux $\iint F \cdot dS$ Magnetic flux Φ Fluid flow rate Q	$\omega = F_x dy \wedge dz + F_\gamma dz \wedge dx + F_z dx \wedge dy$ Acts on pair of tangent vectors
3-form	volume (integral over region)	Mass $\iiint \rho \, dV$ Charge $Q = \iiint \rho \, dV$ Energy in volume	$\omega = \rho \, dx \wedge dy \wedge dz$ In 3D this is maximal dimension of form

Mnemonic: a k-form "eats" k vectors and outputs a number.

0-form: $f()$ — eats nothing, immediately a number
1-form: $\omega(v)$ — eats 1 vector
2-form: $\sigma(u,v)$ — eats 2 vectors
3-form: $\mu(u,v,w)$ — eats 3 vectors

Visualization: what a k-form "looks like"

0-form (function): Scalar field — color/height at each point

```
high
   ↑   ████
   | ██████████
   |████████████████
low
```

Temperature map: to each point — a number.

1-form: "Stacks of planes" — level lines with direction

```
| | | | | |
| | | | | |   ← intersection with vector = "how many lines crossed"
| | |→| | |
| | | | | |
```

1-form dx "counts" how many times a vector crossed lines x = const.
Denser lines = larger value of form.

2-form: "Tubes" — measure flux through area

```
| | | | |
| | | | |   ← area perpendicular to tubes = max. value
| +===+ |      area along tubes = zero value
| |   | |
| +===+ |
| | | | |
```

2-form dx∧dy "counts" how many tubes pass through area.
Orientation of area matters (sign).

3-form: "Density of points" — how many in volume

```
+---------+
| ● ● ● ● |
| ● ● ● ● |   ← integral over volume = total quantity
| ● ● ● ● |
+---------+
```

3-form ρ dx∧dy∧dz = density. Integral = mass/charge/energy.

Exterior derivative d — transition between levels

The operator d raises the degree of a form by 1:

```
  0-form  --d--▶  1-form  --d--▶  2-form  --d--▶  3-form  --d--▶  0
  function        gradient         curl            divergence
     f               df              dω                dσ
```

Physical meaning:

TRANSITION	WHAT IT MEANS PHYSICALLY
df (0→1)	How f changes along path = gradient $df(v) = \nabla f \cdot v$ = directional derivative
dω (1→2)	"Curledness" of field = curl Nonzero dω = field has vortices
dσ (2→3)	"Sources" of field = divergence Nonzero dσ = there are sources/sinks

Magical property: $d \circ d = 0$

Applying d twice, we get 0. This is a deep fact:

- $\mathrm{rot}(\mathrm{grad}\ f) = 0$ – gradient has no vortices
- $\mathrm{div}(\mathrm{rot}\ F) = 0$ – vortex has no sources
- $\partial^2 M = \varnothing$ – boundary of boundary is empty

---

Differential as a 1-Form

Rethinking the Differential

In analysis: $df = f'(x)dx$ – "infinitesimal increment"

New view: df is a linear function on tangent vectors.

$df: T_pM \to \mathbb{R}$
$df(v)$ = "derivative of f in direction v" = $\nabla f \cdot v$

Example: $f(x,y) = x^2 + y^2$, point $p = (1, 2)$

$df = 2x\,dx + 2y\,dy$
At point p: $df = 2dx + 4dy$

Vector $v = (3, 1)$ at point p:
$df(v) = 2 \cdot 3 + 4 \cdot 1 = 10$

This is the derivative of f in direction v: $\nabla f \cdot v = (2,4)\cdot(3,1) = 10$ ✓

Conclusion: dx, dy are not "small quantities", but basis 1-forms.

Definition: 1-Form

A 1-form on manifold M is a smooth mapping

$$\omega: TM \to \mathbb{R}$$

which is linear on each tangent space T_pM.

In coordinates:

$$\omega = \omega_1 dx^1 + \omega_2 dx^2 + \ldots + \omega_n dx^n = \omega_i dx^i$$

where $\omega_i = \omega_i(x^1,\ldots,x^n)$ — smooth functions (components of the form).

Action on a vector:

$$\omega(v) = \omega_i v^i = \omega_1 v^1 + \omega_2 v^2 + \ldots + \omega_n v^n$$

1-form = covector = element of T^*_pM

Connection with previous sections:

- (Lin. algebra): covector $\in V^*$ — linear function on V
- (Tensors): 1-form — tensor of type (0,1)
- (Duality): 1-form is dual to a vector

At each point $p \in M$:

$$\omega_p \in T^*_pM \quad \text{(cotangent space)}$$

1-form on M = smooth family of covectors $\{\omega_p\}_{p \in M}$
= section of cotangent bundle T^*M

$Df \neq \nabla f$: iron-clad argument via dimensions

Let f = temperature in degrees, x = coordinate in meters.

OBJECT	DIMENSION	TYPE
f	[°C]	Scalar (function)
df (differential)	[°C]	1-form (covector)
$\partial f/\partial x$ (partial deriv.)	[°C/m]	Gradient component
∇f (gradient)	[°C/m]	Vector (with metric)

Evident: df and ∇f have different dimensions.

df: acts on displacement vector [m], gives temperature change [°C].
∇f: itself has dimension [°C/m], indicates direction of growth.

Connection via metric: $(\nabla f)^i = g^{ij}(\partial f/\partial x^j)$

In Cartesian: $g^{ij} = \delta^{ij}$, therefore ∇f "looks like" $(\partial f/\partial x, \partial f/\partial y)$.

In curvilinear: metric is nontrivial, and confusing them is forbidden.

Bridge: dx in calculus vs dx in forms

An engineer is accustomed to dx as "infinitesimal increment" from Riemann integral:

$$\int f(x)dx = \lim \Sigma f(x_i)\Delta x_i$$

Here we say: dx is a linear functional.
How is this connected?

CONTEXT	What is dx
Calculus (intuition)	"Infinitesimal piece" of x-axis $\Delta x \to 0$ in the limit
Diff. forms (precise)	Linear function $dx: T_pM \to \mathbb{R}$ $dx(v) = v^1$ (projection onto x-axis)
Integration of forms	$\int_\gamma \omega = \lim \Sigma\, \omega(\Delta\gamma_i)$ Form acts on tangents to the curve

Key to understanding:

In the integral $\int f(x)dx$ the 1-form $f(x)dx$ is "hidden", which is integrated along a curve (segment). The old notation is a simplification, hiding that a linear function on tangent vectors is being integrated.

When we write $\int f(x)dx$, we are actually pairing the form $f(x)dx$ with the tangent vector to the integration curve. This is not "sum of infinitesimals", but a limit of sums of form values on small vectors.

$\int_a^b f(x)dx$ (calculus) = $\int_{[a,b]} f\cdot dx$ (forms)

The left notation is a special case of the right for a curve in $\mathbb{R}^1$.

Practical conclusion:
- In 1D: almost no difference, can use familiar notation
- In nD: form $dx\wedge dy$ — area, $dV = dx\wedge dy\wedge dz$ — volume
- On manifolds: without forms one cannot correctly define an integral

Exterior Product — Key Operation

Definition: exterior (wedge) product

For 1-forms α and β their exterior product $\alpha \wedge \beta$ is a 2-form:

$$(\alpha \wedge \beta)(u, v) = \alpha(u)\beta(v) - \alpha(v)\beta(u)$$

Why exactly this formula?

Suppose α and β are "measuring devices" for vector components:

- $\alpha(u)$ = "how much of vector u in direction α"
- $\beta(v)$ = "how much of vector v in direction β"

The product $\alpha(u)\beta(v)$ is the "area of a rectangle" with sides $\alpha(u)$, $\beta(v)$
But a rectangle is not the correct measure for the area of a parallelogram.

To obtain the oriented area of the parallelogram on u and v:

- Take $\alpha(u)\beta(v)$ — one order of measurements
- Subtract $\alpha(v)\beta(u)$ — the other order
- Get the 2×2 determinant = oriented area.

Formula = determinant:

$$(\alpha \wedge \beta)(u, v) = \det\begin{vmatrix} \alpha(u) & \alpha(v) \\ \beta(u) & \beta(v) \end{vmatrix}$$

Properties:

- Bilinearity: $(a\alpha + b\beta) \wedge \gamma = a(\alpha\wedge\gamma) + b(\beta\wedge\gamma)$
- Antisymmetry: $\alpha \wedge \beta = -\beta \wedge \alpha$
- Consequence: $\alpha \wedge \alpha = 0$
- Associativity: $(\alpha \wedge \beta) \wedge \gamma = \alpha \wedge (\beta \wedge \gamma)$

Example: $dx \wedge dy$

Let $u = (u^1, u^2)$, $v = (v^1, v^2)$ — vectors in $\mathbb{R}^2$.

$$\begin{aligned} (dx \wedge dy)(u, v) &= dx(u)\cdot dy(v) - dx(v)\cdot dy(u) \\ &= u^1v^2 - v^1u^2 \\ &= \det\begin{vmatrix} u^1 & v^1 \\ u^2 & v^2 \end{vmatrix} \end{aligned}$$

$dx \wedge dy$ computes the area of the parallelogram spanned by u and v.

Sign = orientation (positive or negative)

k-forms

A k-form is an antisymmetric tensor of type (0, k).

In coordinates on $\mathbb{R}^n$:

$$\omega = \sum_{i_1<\dots<i_k} \omega_{i_1\dots i_k} \, dx^{i_1} \wedge \dots \wedge dx^{i_k}$$

Dimension of the space of k-forms on $\mathbb{R}^n$:

$$\dim \Omega^k(\mathbb{R}^n) = C(n,k) = n!/(k!(n-k)!)$$

Table for $n = 3$:

k	BASIS	dim
0	1 (functions)	1
1	dx, dy, dz	3
2	$dy \wedge dz$, $dz \wedge dx$, $dx \wedge dy$	3
3	$dx \wedge dy \wedge dz$	1
>3	0 (no nonzero forms)	0

Note: $\dim \Omega^k = \dim \Omega^{n-k}$ — symmetry. (related to Hodge $*$)

Concrete example: gas dynamics — three types of integrals

In gas dynamics, three types of "infinitesimals" constantly appear:

- dx, dl — element of length along the flow (1-form)
- $d\omega$, dA, dS — element of cross-sectional area (2-form)
- dV — element of volume (3-form)

These are not simply "small quantities" — they are objects of different nature.

Example 1: work of pressure force (1-form, line integral)

A piston moves in a cylinder. Work $= \int F \cdot dl = \int p \cdot A \cdot dl$

Here dl is the path element (1-form). We integrate along the trajectory.
Result: scalar (number of joules).

Example 2: mass flow rate through a cross-section (2-form, surface integral)

Gas flows through a pipe. Flow rate = $\iint \rho v \cdot dA = \iint \rho v_n \, d\omega$

Here $d\omega$ (or dA) is the area element (2-form).
We integrate over the cross-section of the pipe.
Result: kg/s (mass flow rate).

Important: $d\omega = dy \wedge dz$ — this is not a number, but a 2-form.
It "eats" two vectors and outputs an oriented area.

Example 3: mass of gas in a volume (3-form, volume integral)

Mass of gas in a reservoir: $m = \iiint \rho \, dV$

Here $dV = dx \wedge dy \wedge dz$ is the volume element (3-form).
We integrate over the entire reservoir.
Result: kg (mass).

Stokes' theorem in gas dynamics: continuity equation

Integral form: $\partial/\partial t \iiint_V \rho \, dV + \iint_S \rho v \cdot dA = 0$

"Change of mass in volume = mass flux through boundary"

Transition to differential form (via Gauss–Ostrogradsky theorem):

$$\partial\rho/\partial t + \nabla\cdot(\rho v) = 0$$

Or in the language of forms: $\partial(\rho \cdot dV)/\partial t + d(\rho v \lrcorner dV) = 0$
where $\lrcorner$ is the interior product (contraction of a vector with the volume n-form). Equivalently, via the codifferential $\delta = \pm{*d*}$:
$\partial\rho/\partial t + \delta(\rho v) = 0$ with the appropriate sign.

Conclusion: When you see an integral, ask yourself:
- Over what are we integrating? (line / surface / volume)
- What form is under the integral? (1-form / 2-form / 3-form)

dx, dA, dV — these are not "just small quantities", but objects of different types.

Exterior derivative d

Definition: exterior derivative

The exterior derivative $d: \Omega^k \to \Omega^{k+1}$ is defined:

For a 0-form (function) f:

$$df = (\partial f/\partial x^1)dx^1 + ... + (\partial f/\partial x^n)dx^n = (\partial f/\partial x^i)dx^i$$

For a k-form $\omega = \omega_{i_1...i_k} dx^{i_1} \wedge ... \wedge dx^{i_k}$:

$$d\omega = d\omega_{i_1...i_k} \wedge dx^{i_1} \wedge ... \wedge dx^{i_k}$$

Properties:

- $d(\alpha + \beta) = d\alpha + d\beta$ (linearity)
- $d(\alpha \wedge \beta) = d\alpha \wedge \beta + (-1)^k \alpha \wedge d\beta$ (Leibniz rule)
- $d(d\omega) = 0$ (key property)

Examples of the exterior derivative

Example 1: d of function $f(x,y,z)$

$$df = (\partial f/\partial x)dx + (\partial f/\partial y)dy + (\partial f/\partial z)dz$$

This is the differential of f — a covector (1-form).
Not to be confused with the gradient ∇f, which is a vector.
(Components coincide numerically only in an orthonormal basis)

Example 2: d of 1-form $\omega = P\, dx + Q\, dy + R\, dz$

$$d\omega = dP \wedge dx + dQ \wedge dy + dR \wedge dz$$

$$= (\partial P/\partial y\, dy + \partial P/\partial z\, dz) \wedge dx + ...$$

$$= (\partial R/\partial y - \partial Q/\partial z)\, dy \wedge dz + (\partial P/\partial z - \partial R/\partial x)\, dz \wedge dx + (\partial Q/\partial x - \partial P/\partial y)\, dx \wedge dy$$

This is the curl rot(P,Q,R), written as a 2-form.

Example 3: d of 2-form $\eta = A\, dy \wedge dz + B\, dz \wedge dx + C\, dx \wedge dy$

$$d\eta = (\partial A/\partial x + \partial B/\partial y + \partial C/\partial z)\, dx \wedge dy \wedge dz$$

This is the divergence div(A,B,C), written as a 3-form.

Unifying table: vector analysis as forms

CLASSICAL	FORMS	RELATION
f (scalar)	f (0-form)	identical
∇f (gradient)	df (1-form)	components coincide
F (vector field)	$\omega = F_i dx^i$ (1-f.) or $\eta = *\omega$ (2-form)	via metric: $\omega_i = g_{ij}F^j$ via Hodge $*$
rot F	$d\omega$ (2-form)	d(1-form) = 2-form
div F	$d{*}\omega$ (3-form)	d(2-form) = 3-form or $*d*\omega$ (function)
$\text{rot}(\nabla f) = 0$ $\text{div}(\text{rot } F) = 0$	$d(df) = 0$ $d(d\omega) = 0$	$d^2 = 0$ $d^2 = 0$

A.t.2 property $d^2 = 0$ – realizations in different areas

AREA	OPERATOR	IDENTITY
Vector analysis	∇, rot, div	rot(grad f)=0, div(rot F)=0
Diff. forms	$d: \Omega^k \to \Omega^{k+1}$	$d(d\omega) = 0$
Homology	$\partial: C_k \to C_{k-1}$	$\partial(\partial c) = 0$
Cohomology	$\delta: C^k \to C^{k+1}$	$\delta(\delta f) = 0$
Complex analysis	$\bar{\partial}$ (Dolbeault)	$\bar{\partial}^2 = 0$
Homol. algebra	d (chain complex)	$d_{k+1} \circ d_k = 0$

Consequence: $H^k = \ker(d)/\text{Im}(d)$ – cohomology
- Closed forms ker(d) / Exact forms Im(d) = topological invariant

Stokes' Theorem — one theorem instead of three

Generalized Stokes' Theorem

```
+--------------------------+
|                          |
| ∫   dω  =  ∫   ω         |
| M         ∂M             |
|                          |
| integral      integral   |
| over domain   over boundary |
|                          |
+--------------------------+
```

Let M — oriented manifold with boundary ∂M,
ω — differential (n−1)-form on M.

Then integral of derivative over domain = integral of form over boundary.

Special cases — all classical theorems = one

dim M	NAME	FORMULA
1 segment	Newton–Leibniz (fund. thm. calculus)	$\int_a^b df = f(b) - f(a)$
2 domain	Green	$\oint_{\partial D} Pdx+Qdy = \iint(\partial Q/\partial x-\partial P/\partial y)dA$
2 surface	Classical Stokes	$\oint_{\partial S} F\cdot dr = \iint_S \text{rot } F\cdot dS$
3 solid	Gauss–Ostrogradsky	$\iint_{\partial V} F\cdot dS = \iiint_V \text{div } F\, dV$

All this — one theorem: $\int_M d\omega = \int_{\partial M} \omega$

$$\iint_{\partial V} F\cdot dS = \iiint_V \text{div } F\, dV$$

This is the Gauss–Ostrogradsky theorem.

Meaning: boundary of boundary is empty

From $d^2 = 0$ and Stokes' theorem it follows:

$$\int_{\partial\partial M} \omega = \int_{\partial M} d\omega = \int_M d^2\omega = 0$$

Therefore $\partial\partial M = \varnothing$ (boundary of boundary is empty)!

Geometric intuition:

```
  dim 1:    ●━━━━━━━━━━●          Segment
            ↓          ↓
            ●          ●          Boundary = 2 points
            ↓          ↓
            ∅          ∅          Boundary of point = ∅

  dim 2:    +----------+          Disk
            | ┌------┐ |
            | |      | | ---▶     ┌------┐  Boundary = circle
            | └------┘ |          └------┘
            +----------+               ↓
                                     ∅   Boundary of circle = ∅

  dim 3:       +----+
              /    /|              Cube
             +----+ |                  ↓
             |    |/              6 faces (surface)
             +----+                    ↓
                                     ∅   Boundary of closed surface = ∅
```

Formula: $\partial^2 = 0 \leftrightarrow d^2 = 0$ (dual statements)

Orientation consistency — source of 50% sign errors

Stokes' theorem $\int_M d\omega = \int_{\partial M} \omega$ works only with consistent orientation of M and its boundary ∂M. Without this — sign error.

```
------------------------------------------------------------------------------
Rule: how orientation is induced on boundary
------------------------------------------------------------------------------
```

dim 1 → dim 0 (segment → points):

Segment [a, b] is oriented "left to right".
Boundary: point a with "−", point b with "+".

```
      −●━━━━━━━━━━━━━━━━━━━●+
       a                   b
```

Therefore: $\int_a^b df = f(b) - f(a)$ (not $f(a) - f(b)$.)

```
-------------------------------------------------------------------------------

 Dim 2 → dim 1 (surface → contour): "right-hand rule"

   Surface oriented by normal n (chose "up").
   Boundary traversed so that n points "left" of motion.

          n ↑
            |    +------------+
            |    |            |
                 |    ---→    |  ← direction of traversal ∂M
                 |            |
                 +------------+

   Mnemonic: Stand on surface, head along normal → boundary on left.

-------------------------------------------------------------------------------

 Dim 3 → dim 2 (volume → surface): "outward normal"

   Volume V oriented standardly (dx∧dy∧dz).
   Boundary ∂V oriented by outward normal.

              +---------+
             /|        /|
            / |  →n   / |   ← normal outward
           +---------+  |
           |  |      |  |
           |  +------|--+
           | /       | /
           |/        |/
           +---------+

   Gauss' theorem: ∬_{∂V} F·dS = ∭_V div F dV
   dS = n dS, where n — outward normal.

-------------------------------------------------------------------------------
Practical advice:

   If you got the "wrong" sign — check:
   1. Where is the normal to the surface directed?
   2. In which direction is the contour traversed?
   3. Are they coordinated by the "corkscrew rule"?
```

Maxwell's Equations in the Language of Forms

Electromagnetic Tensor as a 2-Form

Electric field E and magnetic field B are unified into a 2-form F:

$$F = E_x\, dx\wedge dt + E_\gamma\, dy\wedge dt + E_z\, dz\wedge dt$$
$$+ B_x\, dy\wedge dz + B_\gamma\, dz\wedge dx + B_z\, dx\wedge dy$$

In matrix form (tensor $F_{\mu\nu}$):

$$F = \begin{pmatrix} 0 & -E_x & -E_\gamma & -E_z \\ E_x & 0 & B_z & -B_\gamma \\ E_\gamma & -B_z & 0 & B_x \\ E_z & B_\gamma & -B_x & 0 \end{pmatrix}$$

Maxwell's Equations

Classical form (4 equations):

$$\mathrm{div}\, E = \rho/\varepsilon_0 \qquad \mathrm{rot}\, E = -\partial B/\partial t$$
$$\mathrm{div}\, B = 0 \qquad \mathrm{rot}\, B = \mu_0 J + \mu_0\varepsilon_0\, \partial E/\partial t$$

In the language of forms (2 equations):

```
+-------------------------------------------------------------+
|                                                             |
| dF = 0         (homogeneous: div B = 0, rot E = −∂B/∂t)     |
|                                                             |
| d∗F = J        (inhomogeneous: div E = ρ, rot B = J + ∂E/∂t)|
|                                                             |
+-------------------------------------------------------------+
```

where:
- F — electromagnetic 2-form
- ∗F — its Hodge dual 2-form
- J — current 3-form (current + charge density)

Beauty:
- $dF = 0$ automatically follows from $F = dA$ (potential)
- $d(d{*}F) = 0$ gives charge conservation: $dJ = 0$
- Lorentz invariance is obvious (no separation of E and B)

de Rham Cohomology — Topology Through Forms

Closed and Exact Forms

A form ω is called:

Closed: $d\omega = 0$
Exact: $\omega = d\eta$ for some η

From $d^2 = 0$ it follows: exact ⇒ closed

$d(d\eta) = 0$ ✓

Question: Is the converse true? Is every closed form exact?

Answer: not always. This depends on the topology of the manifold.

de Rham Cohomology

$$H^k_{dR}(M) = \{\text{closed } k\text{-forms}\} / \{\text{exact } k\text{-forms}\}$$

$$= \mathrm{Ker}(d: \Omega^k \to \Omega^{k+1}) / \mathrm{Im}(d: \Omega^{k-1} \to \Omega^k)$$

The dimension of $H^k(M)$ is called the k-th Betti number: $b_k = \dim H^k(M)$

de Rham Theorem:
$H^k_{dR}(M) \cong H^k(M; \mathbb{R})$ (de Rham cohomology = singular cohomology)

This is a bridge between analysis (forms) and topology (holes)!

Examples

$\mathbb{R}^n$: $H^0 = \mathbb{R}$, $H^k = 0$ for $k > 0$
Every closed form is exact (space is "trivial")

S^1: $H^0 = \mathbb{R}$, $H^1 = \mathbb{R}$
The form $d\theta$ is closed but not exact.
$\oint d\theta = 2\pi \neq 0$ (integral over cycle is nonzero)
This "feels" the hole in the circle

T^2 (torus): $H^0 = \mathbb{R}$, $H^1 = \mathbb{R}^2$, $H^2 = \mathbb{R}$
Two independent 1-forms (two independent cycles)

S^2: $H^0 = \mathbb{R}$, $H^1 = 0$, $H^2 = \mathbb{R}$
No 1-dimensional holes, there is a 2-dimensional "shell"

Homology vs Cohomology – Duality Intuition

Both theories "count holes", but from different sides:

Homology H_k:

- Objects: cycles (k-dimensional "contours" without boundary)
- Question: "Which cycles are not boundaries?"
- Geometric objects: curves, surfaces, ...

Example: On a torus there are two cycles (around the hole and through the hole), which cannot be "filled" by a surface inside the torus.
$\Rightarrow H_1(T^2) = \mathbb{Z} \times \mathbb{Z}$

Cohomology H^k:

- Objects: forms (functions on cycles)
- Question: "Which forms are closed but not exact?"
- Functional objects: measure cycles

Example: The form $d\theta$ on a circle is closed but not exact.
It "measures" how many times a cycle goes around the hole.
$\Rightarrow H^1(S^1) = \mathbb{R}$

Analogy:

```
+-------------------------------------------------------------------+
|                                                                   |
| homology – these are "holes" (geometric objects)                  |
| cohomology – these are "hole measurers" (functions on holes)      |
|                                                                   |
| Like vector and covector: one is an object, the other is a function on an object |
|                                                                   |
+-------------------------------------------------------------------+
```

Connection (duality):

$H^k(M; \mathbb{R}) \cong \mathrm{Hom}(H_k(M), \mathbb{R})$ (cohomology = functionals on homology)

Pairing: $\langle [\omega], [c] \rangle = \int_c \omega$
(integral of form over cycle)

Why two theories:

- Homology is simpler for geometric intuition
- Cohomology has multiplication (ring structure)
- de Rham cohomology is related to analysis (differential forms)

Summary and Connections

Main Ideas of the Section

1. Differential forms = "objects for integration"
   k-form is integrated over k-dimensional surfaces

2. Exterior product ∧ = antisymmetrization
   $\alpha \wedge \beta = -\beta \wedge \alpha$, $\alpha \wedge \alpha = 0$

3. Exterior derivative d generalizes grad, rot, div
   $d^2 = 0$ — key property

4. Stokes' theorem unifies N-L, Green, Stokes, Gauss
   $\int_M d\omega = \int_{\partial M} \omega$

5. de Rham cohomology connects analysis and topology
   Closed/exact forms "sense" holes in space

Connection Graph

```
  (Lin.alg)        (Tensors)        (Manifolds)
       |               |                   |
       |   covectors   | antisymm.tensors  |  TₚM, T*ₚM
       +---------------+-------------------+
                       |
                       ▼
                 ▶ (Forms) ◀
                       |
       +---------------+---------------+
       ▼               ▼               ▼
  (Topology)       Physics          Analysis
  cohomology      Maxwell         integrals
                   GTR theor. Stokes
```

Until now we have been building structures on continuous spaces: topology, linearity, smoothness, forms. But mathematics also works with discrete objects — finite sets, graphs, logical statements.

The next two sections — about discrete structures. They are no less fundamental: order lies at the foundation of logic and set theory, graphs describe networks and algorithms.

Order and Lattices — Three Faces of One Structure

Key table "Logic = Sets = Order":

Partial Order — Foundation

A relation ≤ is called a partial order if:

- a ≤ a (reflexivity)
- a ≤ b and b ≤ a ⇒ a = b (antisymmetry)
- a ≤ b and b ≤ c ⇒ a ≤ c (transitivity)

"Partial" = not all elements are comparable
Example: sets {1} and {2} are incomparable under ⊆

Order Everywhere — Examples

SET	ORDER ≤	WHERE IT APPEARS
Numbers ℝ	Usual ≤	Analysis
Subsets 2^X	Inclusion ⊆	Logic, topology
Natural numbers ℕ	Divisibility a|b	Number theory
Words	Prefix	Computer science
Data types	Inheritance	OOP
Open sets	Inclusion ⊆	Topology
Subspaces	Inclusion ⊆	Linear algebra
Subgroups	Inclusion ⊆	Group theory

Lattice — Order with Operations

Lattice = partial order where any two elements have:
a ∨ b = sup{a, b} (least upper bound)
a ∧ b = inf{a, b} (greatest lower bound)

```
 Example: Subsets {a,b,c}             Example: Divisors of 12

         {a,b,c}                                 12
        /   |   \                              / |  \
    {a,b} {a,c} {b,c}                        4   6   ← LCM
       \ X     X /                            \ / \ /
       {a}  {b}  {c}                           2   3
          \ | /                                 \ /
            ∅                                    1   ← GCD

   {a}∨{b} = {a,b}                           2∨3 = LCM(2,3) = 6
   {a}∧{b} = ∅                               2∧3 = GCD(2,3) = 1
```

Boolean Algebra = Lattice with Complement

Additional requirement: for each a there exists $\neg a$ such that

$a \vee \neg a = 1$ and $a \wedge \neg a = 0$

Divisors of 12 – lattice, but not Boolean. (no complement for 2, 3, 4, 6)
Subsets – Boolean algebra (complement always exists)

Connection with Topology

Open sets of a topology form a lattice (but not a Boolean algebra)

- $U_1 \vee U_2 = U_1 \cup U_2$ (union of open sets – open)
- $U_1 \wedge U_2 = U_1 \cap U_2$ (intersection of open sets – open)
- BUT: complement of open – closed, not open.

This leads to intuitionistic logic (without law of excluded middle)

CLASSICAL	INTUITIONISTIC	EXAMPLE
$P \vee \neg P = \top$ (always true)	Not always.	"x is rational or not" – need to constructively show

Applied example: launching a technological line

Problem: Boiler room. Need to start equipment in the correct order.
Some operations can be done in parallel, some – strictly afterwards.

Partial order of operations:

```
Open gas supply (a)
       |
       +-------------------+
       |                   |
       ▼                   ▼
Turn on smoke exhauster (b)    Turn on fan (c)
       |                   |
       +-------+-----------+
               |
               ▼
Burner ignition (d)     ← only after b and c.
       |
       ▼
Enter operating mode (e)
```

This is a partial order:

- $a \le b$, $a \le c$ (gas – before smoke exhauster and fan)
- $b \le d$, $c \le d$ (smoke exhauster and fan – before ignition)
- $d \le e$ (ignition – before entering operating mode)
- but: b and c are incomparable (can be in any order)

Operations on the lattice:

- $b \wedge c = a$ (GLB – greatest operation that is before both)
- $b \vee c = d$ (LUB – least operation that is after both)

Practical meaning:
$b \vee c = d$ means: "d – first operation that can be done only after completion of both b and c"

Application: Gantt chart = visualization of partial order in time.
Critical path = longest chain in the partial order.

Place in the overall picture

Graph = visualization of relation

Relation $R \subseteq A \times A$ – abstract concept
Graph – its picture: vertices = elements, edge = pair in relation

Since relations – foundation of mathematics, graphs are everywhere.

One object – many representations

RELATION	GRAPH	MATRIX
$R \subseteq V \times V$ (set of pairs)	$G = (V, E)$ (picture)	A: $A_{ij} \in \{0,1\}$ (table of numbers)
$(a,b) \in R$	Edge a--b	$A_{a\beta} = 1$
R symmetric $(a,b)\in R \Rightarrow (b,a)\in R$	Undirected	$A = A^T$
Path from a to b in n steps	Chain of edges	$(A^n)_{a\beta} > 0$ (number of paths)
Transitive closure	"Reachability"	$(I+A)^n$ or $(I-A)^{-1}$

Graphs everywhere – examples

FIELD	VERTICES	EDGES
Social networks	People	"Friends"
Internet	Pages	Links
Chemistry	Atoms	Bonds
City map	Crossroads	Streets
Electrical circuit	Nodes	Conductors
Finite automaton	States	Transitions
Task ordering	Tasks	Dependencies
Group (Cayley graph)	Elements	Multiplication by gen.
Simpl. complex (1-skel.)	Vertices	1-simplices

Connection with linear algebra

Adjacency matrix A turns graph theory into linear algebra.

- $(A^n)_{ij}$ = number of paths of length n from i to j
- Eigenvalues of A – spectrum of graph (invariant)
- λ_1 (largest) is related to connectivity
- Number of connected components = multiplicity of $\lambda = 0$ for Laplacian $L = D - A$

Google PageRank = principal eigenvector of transition matrix.

Connection with Topology

Euler characteristic: $\chi = V - E + F$

For a planar graph (on a sphere): $\chi = 2$
For a graph on a torus: $\chi = 0$

This is the same invariant χ as in homology.
A graph is the simplest "simplicial complex" (only 0- and 1-simplices)

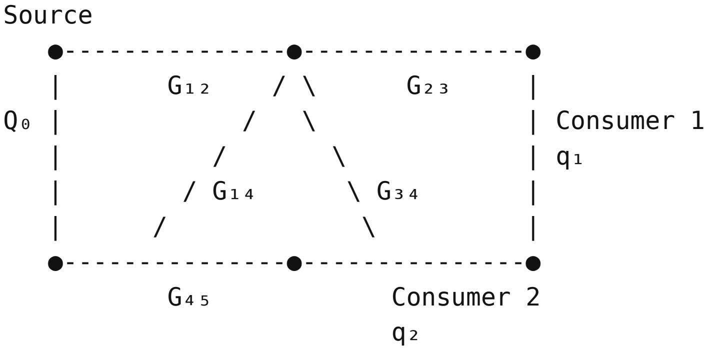
$H_0(\text{graph}) = \mathbb{Z}^k$, where k = number of connected components
$H_1(\text{graph}) = \mathbb{Z}^m$, where m = number of independent cycles = $E - V + k$

Types of graphs

TYPE	PROPERTY and VALUE
Connected	∃ path between any two vertices; $H_0 = \mathbb{Z}$
Tree	Connected + without cycles; $\lvert E\rvert = \lvert V\rvert - 1$; $H_1 = 0$
Complete K_n	All connected; $\lvert E\rvert = n(n-1)/2$
Bipartite	Vertices = 2 classes, edges between classes
Planar	Embeddable in plane without crossings; $\chi = 2$
Cayley graph of group G	Encodes group structure; V=G, edges=generators

Applied example: hydraulic calculation of a heating network

Problem: City heating network. Find flow rates and pressure losses.



Network graph:
  Vertices = nodes (sources, consumers, junctions)
  Edges = pipelines (with flow rate G and resistance S)

First Kirchhoff's law (mass conservation at nodes):

At each node: $\Sigma G_{in} = \Sigma G_{out}$

This is homology. $H_0 = 0$ means: flow can "flow" through the network.

Second Kirchhoff's law (pressure conservation in loops):

In each closed loop: $\Sigma \Delta p_i = 0$

Number of independent loops = $\beta_1 = E - V + 1$ (first Betti number).

Matrix form:

$A \cdot G = q$ (A – incidence matrix "nodes × edges")
$B \cdot \Delta p = 0$ (B – loop matrix)
$\Delta p = S \cdot G^2$ (quadratic resistance law)

Total: nonlinear system, solved by Newton's method.

Numerical example:
Network with $V=5$, $E=6 \Rightarrow \beta_1 = 6-5+1 = 2$ independent loops
loop equations + 4 node equations = 6 equations for 6 flow rates

Conclusion: Hydraulics of networks is graph theory + linear algebra.
Betti number β_1 = number of independent loops = dimension of H_1.

Order is a relation "greater/less" without the requirement of comparability of all pairs.
Graphs are another discrete structure: relation "connected/not connected".

Graphs are discrete spaces. On them work analogs of continuous concepts: Laplacian, diffusion, homology. And they are everywhere: networks, algorithms, data.

Graphs – discrete structures of connections

Graph as a discrete space

A graph is a discrete analog of a space, where:

- Vertices – "points" of the space
- Edges – "neighborhood" (who is next to whom)
- Path – "movement" from point to point
- Distance = number of edges in the shortest path

Many concepts of continuous mathematics have discrete analogs:

Continuous	Discrete (on a graph)
Laplace operator ∇^2	Graph Laplacian $L = D - A$
Heat conduction	Diffusion on a graph: $u' = -Lu$
Potential flow	Electrical circuit (Kirchhoff's laws)

Graph as a structure

Graph $G = (V, E)$ is:
- V – set of vertices (nodes)
- $E \subseteq V \times V$ – set of edges (connections between vertices)

Types of graphs:

- Directed: edges have direction ($a \to b \neq b \to a$)
- Undirected: $\{a, b\} = \{b, a\}$
- Weighted: each edge is assigned a weight $w(e) \in \mathbb{R}$

Examples:

- Social network: vertices = people, edges = acquaintances
- Pipeline: vertices = nodes, edges = pipes, weights = resistances
- Internet: vertices = servers, edges = communication channels

Key Concepts

Path: sequence of vertices $v_0 \to v_1 \to \dots \to v_n$, where $(v_i, v_{i+1}) \in E$

Cycle: path where $v_0 = v_n$ (closed path)

Connected graph: between any two vertices there exists a path

Tree: connected graph without cycles

INVARIANT	FORMULA / PROPERTY
Number of vertices	$\lvert V\rvert = n$
Number of edges	$\lvert E\rvert = m$
Number of components	k (connected pieces of graph)
Euler characteristic	$\chi = n - m + k$
Number of cycles	$\beta_1 = m - n + k$ (first Betti number)

For a tree: $m = n - 1$ (minimum edges for connectivity), $\beta_1 = 0$

Connection with other areas

Topology	Graph — 1-dimensional simplicial complex $\beta_1 = \mathrm{rank}(H_1)$ — topological invariant
Linear algebra	Adjacency matrix, Laplacian, spectrum of graph
Lattices	Hasse diagram — graph of partial order
Manifolds	Discretization: mesh = graph on manifold

Adjacency matrix and Laplacian

Adjacency matrix a:

$A_{ij} = 1$ if there is an edge (i,j), otherwise 0

```
1---2                 ⎛ 0 1 1 0 ⎞
|   |        →    A = ⎢ 1 0 0 1 ⎥
3---4                 ⎢ 1 0 0 1 ⎥
                      ⎝ 0 1 1 0 ⎠
```

Degree matrix d:

$D_{ii} = \deg(i)$ = number of edges from vertex i

Graph Laplacian:

$L = D - A$

Properties of Laplacian:

- L is symmetric, positive semidefinite
- Eigenvalues: $0 = \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$
- Multiplicity of $\lambda = 0$ equals the number of connected components
- λ_2 (algebraic connectivity) — measure of "how connected" the graph is

Analogy with physics:

Graph Laplacian — discrete analogue of Laplace operator ∇^2.

Heat equation on graph: $du/dt = -Lu$

Why $L \approx \nabla^2$ (detailed explanation)

Continuous Laplacian on the line:

$(\nabla^2 f)(x) = f''(x) \approx [f(x+h) - 2f(x) + f(x-h)] / h^2$

Graph as discretization: vertices = points, edges = adjacency.

Action of graph Laplacian on function $f: V \to \mathbb{R}$ (values at vertices):

$$(Lf)(i) = \underset{\substack{\uparrow \\ \text{sum over neighbors } j \text{ of vertex } i}}{\Sigma_{j\sim i}} [f(i) - f(j)] = \deg(i)\cdot f(i) - \Sigma_{j\sim i} f(j)$$

This is exactly the discrete version of ∇^2.

- At a point: value minus average of neighbors
- Measures "deviation from local average"
- Harmonic function: $Lf = 0 \iff$ at each point f = average of neighbors

Heat equation:

Continuous: $\partial u/\partial t = \nabla^2 u$ → Discrete: $du/dt = -Lu$

Heat flows from hot vertices to cold neighbors.

Spectral graph theory — linear algebra in action

Eigenvalues and eigenvectors of Laplacian L — these are "frequencies" and "modes" of network oscillations, analogous to resonant frequencies of a string or membrane.

Spectrum of Laplacian: $0 = \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$

EIGENVALUE	WHAT IT MEANS
$\lambda_1 = 0$	Always eigenvector = (1,1,...,1)
Multiplicity $\lambda = 0$	= number of connected components of graph
λ_2 (Fiedler value)	Algebraic connectivity: the larger, the "stronger" the graph is connected (harder to cut)
λ_n	Maximum "oscillation frequency" of network

Fiedler vector:
  Eigenvector v_2 corresponding to λ_2.

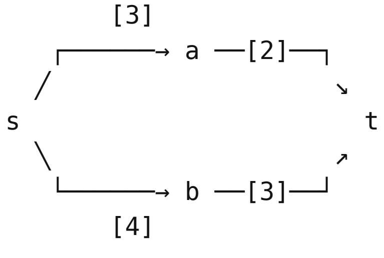
  Magic: Signs of components of v_2 divide the graph into two parts.
  Vertices with $v_{2i} > 0$ — one group, with $v_{2i} < 0$ — another.
  This is optimal partitioning (minimizes number of edges between groups).

Applications:
- Spectral clustering: grouping vertices by eigenvectors
- PageRank: principal eigenvector of transition matrix (Google)
- Synchronization: connection of λ_2 with stability of synchronous regimes
- Graph neural networks (GNN): convolution through spectrum of Laplacian

This is pure magic of linear algebra in the discrete world.
Local structure (edges) → global property (spectrum).

Engineering Applications: Network Flows

Maximum Flow Problem:
  Given a network with source s and sink t.
  Edges have capacity c(e).
  Find maximum flow from s to t.



Max flow = 2 + 3 = 5

(bottlenecks: a→t is 2,
b→t is 3)

Cut: {s} | {a, b, t}. Sum of capacities of cut edges:
$c(s \to a) + c(s \to b) = 3 + 4 = 7$. But this is not the minimum cut.
Minimum cut: {s,a,b} | {t}, sum = $c(a \to t) + c(b \to t) = 2 + 3 = 5$.

Theorem (Ford–Fulkerson):
Maximum flow = minimum cut

Applications:

- Pipeline networks (maximum throughput)
- Transportation problems (optimal logistics)
- Electrical networks (Kirchhoff's laws)

Kirchhoff's Laws as Graph Theory

Electrical circuit = weighted graph, where:

- Vertices = circuit nodes
- Edges = resistors/conductors
- Weights = conductances $g = 1/R$

Kirchhoff's First Law (nodal):
$\Sigma I_k = 0$ at each node
↔ Current vector $I \in \ker(\partial_1)$ – kernel of boundary operator.

Kirchhoff's Second Law (loop):
$\Sigma U_k = 0$ around any closed loop
↔ Voltage vector $U \in \mathrm{im}(\partial_0^*)$ – image of co-boundary operator.

Ohm's Law: $I = GU$ (G – conductance matrix)

Result: Circuit solution = solution of system $L\varphi = I\_ext$
where L – weighted Laplacian, φ – node potentials

This is the same mathematics as heat conduction, diffusion, hydraulics.

Shortest Paths and Algorithms

Problem: Find shortest path between vertices in weighted graph.

ALGORITHM	COMPLEXITY	WHEN TO USE
BFS (breadth-first)	$O(V + E)$	Unweighted graph
Dijkstra	$O(E + V \log V)$	Non-neg. weights, one source
Bellman–Ford	$O(VE)$	Any weights, one source
Floyd–Warshall	$O(V^3)$	All pairs, any weights

Applications in engineering:

- Routing in networks
- Optimal pipeline routes
- Project planning (CPM/PERT – task graph)

Combinatorics – the Art of Counting

Combinatorics answers the question "how many?": how many ways to choose, arrange, partition objects. These formulas are needed everywhere: from probability theory to quantum mechanics.

Combinatorics as a View of Space

Combinatorics studies finite discrete spaces.

If topology asks "which points are close?", and metric – "how close?", then combinatorics asks: "how many points/paths/configurations?"

This is zero-dimensional geometry: structure without continuity, only counting.

Combinatorial objects = points in spaces

OBJECT	WHAT KIND OF SPACE IS THIS	SIZE
Permutation of n elements	Set of all ordered arrangements = group S_n	n! points
k-subset from n elements	Set of all selections of k from n = Grassmannian Gr(k,n)	C(n,k) points
Path in graph from A to B	Space of paths (discrete manifold)	Counted by graph
Partition of number n	Set of ways to represent n as sum	p(n) points (number partitions)

Connection to Other Structures

Groups:

Set of permutations S_n is a group.

- Operation: composition of permutations
- Group order $|S_n| = n!$
- Subgroups of S_n are symmetry groups of finite objects

Linear algebra:

Binomial coefficient C(n,k) = dimension of space.

- Set of all k-subsets of {1,...,n} – basis
- dim(space of k-forms on $\mathbb{R}^n$) = C(n,k)
- Pascal's triangle = dimensions of exterior powers

Probability:

Combinatorics — foundation of discrete probability.

- P(event) = (favorable outcomes) / (all outcomes)
- "All outcomes" are counted combinatorially

Graphs:

Number of paths in graph = combinatorics on discrete space

- Adjacency matrix $A^n[i,j]$ = number of paths of length n from i to j

---

Two basic principles

Addition principle:

If a task can be performed by method A or method B (mutually exclusive),
and A can be done in m ways, B — in n ways, then there are m + n ways total.

Multiplication principle:

If a task consists of step A and step B (sequential),
and A can be done in m ways, B — in n ways, then there are m · n ways total.

Pigeonhole principle (box principle):

If n+1 objects are placed into n boxes, at least one box contains ≥ 2 objects.

Application: Among 367 people there will be two with the same birthday.

Factorial

```
+-----------------------------------------------------------------+
| n! = n · (n−1) · (n−2) · ... · 2 · 1                            |
|                                                                 |
| n! = number of ways to arrange n distinct objects in a row      |
+-----------------------------------------------------------------+
```

Values:

0! = 1 (by definition)
1! = 1
2! = 2
3! = 6
4! = 24
5! = 120
10! = 3 628 800
20! ≈ 2.4×10^{18}

Stirling's formula (asymptotics for large n):

$n! \approx \sqrt{2\pi n} \cdot (n/e)^n$

More precisely: $n! = \sqrt{2\pi n} \cdot (n/e)^n \cdot (1 + O(1/n))$

Permutations, arrangements, combinations

	ORDER MATTERS?	REPETITION	FORMULA
Permutations (all n)	YES	NO	$P_n = n!$
Arrangements (k from n)	YES	NO	$A_n^k = n!/(n-k)!$ $= n(n-1)...(n-k+1)$
Combinations (k from n)	NO	NO	$C_n^k = n!/(k!(n-k)!)$ $=$ (n choose k)
Arrangements with repet.	YES	YES	n^k
Combinations with repet.	NO	YES	C_{n+k-1}^k

Examples:

- How many ways to arrange 5 books on a shelf? $P_5 = 5! = 120$
- How many three-digit numbers from digits 1-9 without repetition? $A_9^3 = 504$
- How many ways to choose 3 people from 10? $C_{10}^3 = 120$
- How many three-letter words from alphabet {a,b,c}? $3^3 = 27$

Binomial coefficients

DEFINITION:

$$C_n^k = \binom{n}{k} = \frac{n!}{k!(n-k)!} = \text{"n choose k"}$$

= number of ways to choose k objects from n (order doesn't matter)

Properties:

$C_n^0 = C_n^n = 1$

$C_n^k = C_n^{n-k}$ (symmetry)

$C_n^k = C_{n-1}^{k-1} + C_{n-1}^k$ (recurrence relation, Pascal's triangle)

Pascal's triangle:

```
           1                         n=0
         1   1                       n=1
       1   2   1                     n=2
     1   3   3   1                   n=3
   1   4   6   4   1                n=4
 1   5  10  10   5   1              n=5
```

Each number = sum of the two above it

Newton's binomial theorem

$$\boxed{\begin{aligned}(a+b)^n &= \sum_{k=0}^{n} C_n^k \cdot a^{n-k} \cdot b^k \\ &= C_n^0a^n + C_n^1a^{n-1}b + C_n^2a^{n-2}b^2 + \dots + C_n^nb^n\end{aligned}}$$

Special cases:

$(a+b)^2 = a^2 + 2ab + b^2$

$(a+b)^3 = a^3 + 3a^2b + 3ab^2 + b^3$

$(1+x)^n = 1 + nx + n(n-1)x^2/2 + \dots$

Corollaries (substituting specific a, b):

$a=b=1$: $2^n = \Sigma C_n^k = C_n^0 + C_n^1 + \dots + C_n^n$

$a=1, b=-1$: $0 = \Sigma(-1)^kC_n^k$ (sum of evens = sum of odds)

Inclusion-Exclusion Principle

For two sets:

$|A \cup B| = |A| + |B| - |A \cap B|$

For three sets:

$$\begin{aligned}|A \cup B \cup C| = &|A| + |B| + |C| \\ &- |A\cap B| - |A\cap C| - |B\cap C| \\ &+ |A\cap B\cap C|\end{aligned}$$

General formula for n sets:

$|A_1 \cup \dots \cup A_n| = \Sigma|A_i| - \Sigma|A_i\cap A_j| + \Sigma|A_i\cap A_j\cap A_k| - \dots$

Application: How many numbers from 1 to 100 are not divisible by 2, 3, or 5?

$|A_2\cup A_3\cup A_5| = 50+33+20 - 16-10-6 + 3 = 74$

Answer: $100 - 74 = 26$

Why the formulas are exactly this way – geometric intuition

Factorial n! = volume of "discrete cube"

Imagine n! as the number of vertices in the space of ordered sets.
Each choice "first element, second, ..." – this is a path through a tree:

```
●----+----●----+----●----●
```

```
              |         |
              +----●    +----●
              |         |
              +----●    +----●
```

At each level: n, then n-1, then n-2, ... variants.
Total paths: $n \times (n-1) \times ... \times 1 = n!$

C(n,k) = "volume" of the space of unordered k-subsets

Paths in ordered selection: $n(n-1)...(n-k+1) = n!/(n-k)!$
But k! orderings give one subset.
Therefore: $C(n,k) = n! / (k!(n-k)!)$

This is the same pattern as in the subspace theorem:
$\dim(V/W) = \dim(V) - \dim(W)$ — factorization by symmetry.

We examined discrete structures (order, graphs). Now — an important supplement to continuous mathematics: complex numbers.

Why here? Complex numbers are not "yet another number system".
This is a place where algebra (multiplication) and geometry (rotations) coincide.
They are critical for analysis, which will come later.

Complex numbers — algebra meets geometry

Complex numbers as a view of space

$\mathbb{C}$ — this is two-dimensional space $\mathbb{R}^2$ with additional structure: multiplication, which encodes rotations and dilations.

As space: $\mathbb{C} \cong \mathbb{R}^2$ (plane)
As algebra: $\mathbb{C}$ has multiplication ($\mathbb{R}^2$ — does not)

This makes $\mathbb{C}$ unique: simultaneously geometry and algebra.

Main discovery

Complex numbers — this is a place where algebra and geometry merge.

Algebra: numbers that can be added and multiplied
Geometry: points of the plane and rotations

Key fact: multiplication by $e^{i\theta}$ = rotation by angle θ

One object — three views

ALGEBRA	GEOMETRY	ANALYSIS
$z = a + bi$ $(a, b \in \mathbb{R}, i^2 = -1)$	Point (a, b) on the plane	Pair of functions Re(z), Im(z)
$z = \|z\|e^{i\theta}$ (polar form)	Polar coordinates (r, θ)	$e^{i\theta} = \cos\theta + i \sin\theta$ (Euler's formula)
Multiplication by w $z \mapsto wz$	Scaling by $\|w\|$ + rotation by θ_w	Linear mapping $\mathbb{R}^2 \to \mathbb{R}^2$
$\|z\| = 1$ (numbers of form $e^{i\theta}$)	Unit circle S^1	Group $U(1) \cong SO(2)$ (rotations of plane)

Group U(1) — key to everything

Numbers with $|z| = 1$ form a group under multiplication:

- Closure: $|z_1| = |z_2| = 1 \Rightarrow |z_1 z_2| = 1$
- Identity: $e^0 = 1$
- Inverse: $(e^{i\theta})^{-1} = e^{-i\theta}$

This is the group $U(1) \cong S^1 \cong SO(2) \cong \mathbb{R}/\mathbb{Z}$ — all of them are the same:

NOTATION	HOW WE VIEW
$U(1)$	Unitary 1×1 matrices (= complex numbers $\|z\|=1$)
S^1	Circle as topological space
$SO(2)$	Group of plane rotations
$\mathbb{R}/\mathbb{Z}$	Real numbers modulo 1 (angles $0 \equiv 2\pi$)

Connection to Fourier: basis e^{inx} = representations of group U(1).

Euler's Formula — a Bridge Between Worlds

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

Special case ($\theta = \pi$): $e^{i\pi} + 1 = 0$ (Euler's identity)

Connects: e (analysis), i (algebra), π (geometry), 1 and 0 (arithmetic)

Proof (Taylor series):

$$\begin{aligned} e^{ix} &= \Sigma\, (ix)^n/n! = \Sigma\, i^n x^n/n! \\ &= (1 - x^2/2! + x^4/4! - \ldots) + i(x - x^3/3! + x^5/5! - \ldots) \\ &= \cos x + i\sin x \end{aligned}$$

Why ℂ "completes" algebra

Fundamental theorem of algebra:
Any polynomial of degree n has exactly n roots in ℂ (counting multiplicity)

This means: in ℂ any algebraic equation is solvable.
No need to extend numbers further (for solving equations).

Connection to I-bis: ℂ — "end of the path" $\mathbb{N} \to \mathbb{Z} \to \mathbb{Q} \to \mathbb{R} \to \mathbb{C}$

Formulas (for calculations)

Multiplication: $z_1 z_2 = |z_1||z_2|\, e^{i(\theta_1+\theta_2)}$ (moduli ×, angles +)

Power: $z^n = |z|^n\, e^{in\theta}$ (de Moivre's formula)

Roots: $\sqrt[n]{z} = \sqrt[n]{|z|} \cdot e^{i(\theta+2pk)/n}$, $k = 0,1,\ldots,n-1$ (n roots)

Trig. identities: $\cos\theta = (e^{i\theta} + e^{-i\theta})/2$, $\sin\theta = (e^{i\theta} - e^{-i\theta})/(2i)$

Concrete example: damped temperature oscillations

Problem: Building wall. Outside temperature oscillates (day/night).
How do oscillations propagate into the wall?

Heat conduction equation:

$$\partial T/\partial t = a \cdot \partial^2 T/\partial x^2 \quad \text{(a — thermal diffusivity coefficient)}$$

Boundary condition (outside): $T(0,t) = T_0 + \Delta T \cdot \cos(\omega t)$
(daily oscillations with period 24 hours)

Solution through complex numbers:

Seek $T(x,t) = \mathrm{Re}[\tilde{T}(x)\cdot e^{i\omega t}]$ – complex amplitude $\tilde{T}$

Substitute: $i\omega\cdot\tilde{T} = a\cdot d^2\tilde{T}/dx^2$

Characteristic equation: $\lambda^2 = i\omega/a$

Solution: $\lambda = \pm(1+i)\sqrt{\omega/2a} = \pm(1+i)/\delta$

Where $\delta = \sqrt{2a/\omega}$ – penetration depth of temperature wave

Physical meaning:

$$T(x,t) = T_0 + \Delta T\cdot \underbrace{e^{-x/\delta}}_{\text{damping}}\cdot \underbrace{\cos(\omega t - x/\delta)}_{\text{phase shift}}$$

- Oscillation amplitude decays with depth as $e^{-x/\delta}$
- Phase lags: maximum inside occurs later than outside

Numerical example:
Brick wall: $a \approx 0.5\times10^{-6}$ m²/s
Daily oscillations: $\omega = 2\pi/(24\cdot3600) \approx 7.3\times10^{-5}$ rad/s
Penetration depth: $\delta = \sqrt{2\cdot0.5\times10^{-6}/7.3\times10^{-5}} \approx 0.12$ m

At depth 30 cm: amplitude drops by factor of $e^{0.3/0.12} \approx 12$.

Moral: Complex numbers are not "imaginary". They naturally arise when solving equations with oscillations and damping.

Where it leads

Quantum mechanics: states are complex, phase $e^{i\theta}$ is physically important
Wave function $\psi(x) \in \mathbb{C}$, probability = $|\psi|^2$

Electrical engineering: impedance $Z = R + iX$, phase relations

Control theory: transfer functions, stability analysis

Quaternions $\mathbb{H}$: generalization to 4D, three "imaginary units" i,j,k
Rotations in 3D = SU(2) = unit quaternions

Why categories are needed – motivation

Problem and solution

Problem: In mathematics there are many similar constructions

- Group homomorphism, ring homomorphism, continuous map.
- Product of sets, direct product of groups, product of topol.
- Kernel of homomorphism, kernel of linear operator.

Goal: extract common pattern from special cases

Answer: category theory – a language for describing "structure of structures"

Slogan: "Mathematics is what remains when you forget the specifics"

Definition of category

Definition: category

Category C consists of:

1. Objects: class $\mathrm{Ob}(C)$
   (sets, groups, topological spaces, ...)

2. Morphisms: for each pair of objects A, B – set $\mathrm{Hom}(A, B)$
   (functions, homomorphisms, continuous maps, ...)

3. Composition: for $f: A \to B$ and $g: B \to C$ defined $g \circ f: A \to C$

Axioms:
- (C1) Associativity: $(h \circ g) \circ f = h \circ (g \circ f)$
- (C2) Identity: for each A there exists $id\_A: A \to A$ such that
  $f \circ id\_A = f$ and $id\_B \circ f = f$ for any $f: A \to B$

Equality "=" vs isomorphism "≅" – critical distinction

In category theory this distinction is fundamental:

Equality (=): Objects literally coincide, it's one object.
Isomorphism (≅): Objects are "structured identically", but they are different objects.

Examples:
- $V \cong V^{**}$ (isomorphism), but $V \neq V^{**}$ (different sets)
  Canonical isomorphism exists, but it's not equality.

- $\mathbb{R}^2 \cong \mathbb{C}$ as vector spaces, but $\mathbb{R}^2 \neq \mathbb{C}$
  ($\mathbb{C}$ has multiplication, $\mathbb{R}^2$ – doesn't)

- All singleton sets are isomorphic, but $\{0\} \neq \{1\}$

In this atlas we sometimes write "=" where a mathematician would write "≅". This is a conscious simplification for readability, but remember:

Isomorphism means "can be identified without loss of information"
Equality means "it's literally the same thing"

Visualization

Category is a directed graph with composition:

```
      f           g
A ------► B ------► C
 \                 /
  \               /
   \    g∘f      /
    \---------/
```

- Objects = vertices
- Morphisms = arrows
- Composition = gluing of paths
- At each vertex there is a loop id

Examples of categories

Basic categories

CATEGORY	OBJECTS	MORPHISMS
Set	Sets	Functions

Grp	Groups	Group homomorphisms
Ab	Abelian groups	Group homomorphisms
Ring	Rings	Ring homomorphisms
Vect_k	Vector spaces over field k	Linear maps
Top	Topological spaces	Continuous maps
Man	Smooth manifolds	Smooth maps
Pos	Partially ordered sets	Monotone functions
hTop	Topological spaces	Homotopy classes of maps

"Small" categories

Category from one object = monoid

- One object ●
- Morphisms ● → ● = elements of monoid
- Composition = multiplication in monoid
- id = unit of monoid

If all morphisms are invertible → group

Category from partial order:

- Objects = elements of set
- Hom(a,b) = {unique arrow}, if $a \le b$, otherwise ∅
- Composition: $a \le b \le c \Rightarrow a \le c$ (transitivity)

Universal constructions: product in different categories

"Product A × B" is defined identically in all categories:

Object P with projections $\pi_1: P \to A$ and $\pi_2: P \to B$ such that for any Q with arrows $f: Q \to A$ and $g: Q \to B$ there exists unique arrow $h: Q \to P$ with $\pi_1 \circ h = f$ and $\pi_2 \circ h = g$.

$$\begin{tikzcd}
 & Q \arrow[dl, "f"'] \arrow[d, "h"] \arrow[dr, "g"] & \\
A & P \arrow[l, dashed, "\pi_1"] \arrow[r, dashed, "\pi_2"'] & B
\end{tikzcd}$$

What this gives in concrete cases:

CATEGORY	What is A × B
Set	Cartesian product $\{(a,b) : a \in A, b \in B\}$
Grp	Direct product of groups
Top	Product with product topology
Vect	Direct sum $V \oplus W$
Sets with ≤	Componentwise order $(a_1,b_1) \le (a_2,b_2)$

Conclusion: One definition → many realizations.
Categorical language reveals common structure of different constructions.

Functors – mappings between categories

Key idea

"To see analogies between analogies means to be able to map mappings between different categories"

A functor is precisely this: a way to transfer structure from one area of mathematics to another, preserving connections between objects.

Definition: functor

Everyday analogy: A functor is a translator between languages of mathematics.

Imagine: there is a book in Russian (category C) and its translation to English (category D). A good translation (functor F):
- Associates to each word/concept a word/concept: A ↦ F(A)
- To each connection between concepts – a connection: (A→B) ↦ (F(A)→F(B))
- Preserves logic: if A→B→C, then F(A)→F(B)→F(C)

Example: π_1 – a "translator" from the language of topology to the language of algebra.
Associates to a topological space a group,
to a continuous map – a group homomorphism.

A functor F: C → D between categories C and D consists of:

1. A mapping of objects: A ↦ F(A)
2. A mapping of morphisms: f ↦ F(f), where if f: A → B, then F(f): F(A) → F(B)

Axioms:

(F1) $F(g\circ f) = F(g)\circ F(f)$ (preserves composition)
(F2) $F(id_A) = id_{F(A)}$ (preserves identity)

Visualization:

$$\begin{tikzcd}
C & & & & D & \\
A \arrow[r, "f"] & B \arrow[dr, "g"'] & {} \arrow[r, "F"] & {} & F(A) \arrow[r, "F(f)"] & F(B) \arrow[d, "F(g)"'] \\
 & & C & & & F(C)
\end{tikzcd}$$

A functor is a "translator" between categories,
preserving structure (arrows and their composition).

Examples of functors

TYPE	FUNCTOR	WHAT IT DOES
forgetful (lose structure)	Grp → Set	Group ↦ set (forgot operation)
	Vect → Set	Space ↦ set of vectors
	Top → Set	Space ↦ set of points
free (add structure)	Set → Grp	X ↦ free group F(X)
	Set → Vect	X ↦ space with basis X
topological invariants	π_1: Top → Grp	Space ↦ fundamental group
	H_n: Top → Ab	Space ↦ n-homology
	χ: Top → $\mathbb{Z}$	Space ↦ Euler characteristic
algebraic	*: Vect → Vect	V ↦ V* (dual)
	T: Man → VBund	M ↦ TM (tangent bundle)

Covariant vs contravariant

COVARIANT	$f: A \to B$ gives $F(f): F(A) \to F(B)$ (arrow same way)
CONTRAVARIANT	$f: A \to B$ gives $F(f): F(B) \to F(A)$ (arrow reversed)
Example: V ↦ V*	$T: V \to W$ gives $T^*: W^* \to V^*$ (reversed)

Natural transformations

Definition: natural transformation

Let $F, G: C \to D$ be two functors. A natural transformation $\eta: F \Rightarrow G$ is a family of morphisms $\{\eta_A: F(A) \to G(A)\}_{A \in Ob(C)}$, such that for any $f: A \to B$ the diagram commutes:

$$\begin{tikzcd}
F(A) \arrow[r, "\eta_A"] \arrow[d, "F(f)"'] & G(A) \arrow[d, "G(f)"] \\
F(B) \arrow[r, "\eta_B"'] & G(B)
\end{tikzcd}$$

That is: $G(f) \circ \eta_A = \eta_B \circ F(f)$

Clarification:
A natural transformation is a "morphism between functors", which is compatible with all morphisms in the source category.

Example: Double Dual

Functors Id, **: Vect → Vect
$\mathrm{Id}(V) = V$
$**(V) = V^{**}$

Natural transformation η: Id ⇒ **
$\eta_V: V \to V^{**}$
$\eta_V(v)(\varphi) = \varphi(v)$

Naturality:
For any $T: V \to W$ the diagram commutes:

$$\begin{tikzcd}
V \arrow[r, "\eta_V"] \arrow[d, "T"'] & V^{**} \arrow[d, "T^{**}"] \\
W \arrow[r, "\eta_W"'] & W^{**}
\end{tikzcd}$$

This is the canonical isomorphism $V \cong V^{**}$ (does not require choice of basis)

Universal Properties

Idea of Universal Property

Many constructions in mathematics are defined not by an explicit formula, but by a property: "the unique object with such-and-such property".

This is the universal property.

Advantage: The definition works the same way in all categories.

Product (universal definition)

The product of objects A and B is an object $A\times B$ with projections π_1, π_2:

$$\begin{tikzcd}
A & A\times B \arrow[l, "\pi_1"'] \arrow[r, "\pi_2"] & B
\end{tikzcd}$$

such that for any X with maps $f: X\to A$, $g: X\to B$
there exists a unique $h: X \to A\times B$ with $\pi_1\circ h = f$, $\pi_2\circ h = g$:

$$\begin{tikzcd}
& X \arrow[dl, "f"'] \arrow[d, "h"] \arrow[dr, "g"] & \\
A & A\times B \arrow[l] \arrow[r] & B
\end{tikzcd}$$

Examples:

- Set: $A\times B$ = Cartesian product
- Grp: $A\times B$ = direct product of groups
- Top: $A\times B$ = product with product topology
- Vect: $A\times B$ = direct sum $A\oplus B$

One definition — all these constructions.

Table of universal constructions

CONSTRUCTION	UNIVERSAL PROPERTY

Product A×B	Pair of projections factorizing any pair of arrows
Coproduct A⊔B (sum)	Pair of injections factorizing any pair of arrows (dual to product)
Terminal object 1	Unique arrow from any object (singleton set, trivial group)
Initial object 0	Unique arrow to any object (empty set, trivial group)
Kernel f	Equalizer of f and 0
Cokernel f	Coequalizer of f and 0
Pullback	"Product over an object"
Pushout	"Coproduct under an object"
Limit	Generalization of product to diagrams
Colimit	Generalization of sum to diagrams

Universal constructions in different categories

One definition via universal property has different realizations:

CONSTRUCTION	Set	Grp	Vect	Top
Product (A,B)→A×B	A×B (Cartesian)	G×H (direct)	V⊕W (direct ⊕)	X×Y with product topology
Coproduct (sum)	A⊔B (disjoint)	G*H (free)	V⊕W (same)	X⊔Y with sum topology
Terminal object	{*} (singleton)	{e} (trivial)	{0} (zero)	{*} (point)
Initial object	∅ (empty)	{e} (trivial)	{0} (zero)	∅ (empty)

Remark: In Vect product = coproduct = ⊕ (abelian category)
In Grp they are different: ×≠* (direct ≠ free)

Definition: adjunction

Functors F: C → D and G: D → C are called adjoint (F ⊣ G), if there exists a natural isomorphism:

$$\mathrm{Hom\_D}(F(A), B) \cong \mathrm{Hom\_C}(A, G(B))$$

F — left adjoint, G — right adjoint.

Examples of adjoint functors

Free ⊣ forgetful:

F: Set → Grp (free group)
U: Grp → Set (forgetful functor)

$$\mathrm{Hom\_Grp}(F(X), G) \cong \mathrm{Hom\_Set}(X, U(G))$$

"A homomorphism from the free group F(X) to G is determined by a map of the set X to G"

Tensor ⊣ hom:

- ⊗ B: Vect → Vect
Hom(B, −): Vect → Vect

Hom(A ⊗ B, C) ≅ Hom(A, Hom(B, C))

This is currying from functional programming.

Programming analogy:
A function of two arguments f(a, b) → c
is equivalent to a function returning a function: a → (b → c)

```
In Haskell: curry :: ((a, b) -> c) -> (a -> (b -> c))
            uncurry :: (a -> (b -> c)) -> ((a, b) -> c)
```

This is not just a coincidence — this is the same universal principle, which in mathematics is called adjunction of functors.

Summary

Hierarchy of categorical mathematics

```
Level 0:  Objects (sets, groups, spaces)
                |
Level 1:  Morphisms between objects (functions, homomorphisms)
                |
Level 2:  Functors between categories
                |
Level 3:  Natural transformations between functors
                |
Level 4:  Modifications between nat. transformations (2-categories)
                |
    .      (∞-categories, homotopy type theory)
```

Why this is needed — practical applications

Mathematics:
- Unification: one proof works in all categories
- Homological algebra: Ext, Tor functors
- Algebraic geometry: sheaves, schemes
- Topos theory: generalization of set theory

Programming:
- Functional programming: functors, monads
- Type theory: Curry–Howard–Lambek correspondence
- Haskell, Scala: categorical constructions in the language

Physics:
- TQFT: functors from the category of cobordisms
- Quantum gravity: spin foams, ∞-categories

Logic:
- Proofs = morphisms
- Types = objects
- Equivalence = isomorphism

Algebraic geometry — a bridge between algebra and geometry

Algebraic geometry — one of the central areas of modern mathematics. It studies geometric objects defined by algebraic equations.

Main idea: duality geometry ↔ algebra

GEOMETRY	ALGEBRA
Set of points X	Ring of functions on X (or ideal of equations defining X)
Point $p \in X$	Maximal ideal $m_p \subset k[X]$
Subvariety $Y \subset X$	Ideal $I(Y) \subset k[X]$
Map $X \to Y$	Ring homomorphism $k[Y] \to k[X]$ (in the OPPOSITE direction)
X irreducible (one piece)	$I(X)$ — prime ideal

This is like a dictionary between two languages — to each geometric notion corresponds an algebraic one, and vice versa.

Examples of algebraic varieties

EQUATION	GEOMETRY	PROPERTIES
$x^2 + y^2 = 1$	Circle	Genus 0, rational curve
$y^2 = x^3 + ax + b$	Elliptic curve	Genus 1, group of points. Cryptography (ECDSA)
$x^2 + y^2 + z^2 = 1$	Sphere S^2	Quadric, rational surface
$xy = 1$	Hyperbola	Two components in $\mathbb{R}$, one in $\mathbb{C}$
$y^2 = x^3$	Curve with cusp (cuspidal curve)	Singular at (0,0)

Hilbert's Nullstellensatz (theorem on zeros)

This is the foundation of algebraic geometry — the connection between ideals and points.

Let k — algebraically closed field (for example, $\mathbb{C}$).

THEOREM: For ideal $I \subset k[x_1,...,x_n]$

$I(V(I)) = \sqrt{I}$

where $V(I)$ = {points vanishing all $f \in I$}
$I(X)$ = {polynomials vanishing on X}
$\sqrt{I}$ = $\{f : f^n \in I$ for some $n\}$ (radical)

Corollary: Maximal ideals in $k[x_1,...,x_n] \leftrightarrow$ points of k^n
Ideal $m = (x_1 - a_1, ..., x_n - a_n)$ corresponds to point $(a_1,...,a_n)$

Categorical View: Schemes

Classical variety = points + structure.
Scheme (Grothendieck) = space + structural sheaf of rings.

CLASSICAL	SCHEMES

Points = solutions over k	Points = prime ideals (Spec) (including "generic points".)
Only over a field	Over any ring (ℤ, $\mathbb{F}_p$, ...)
Reduced ($x^2 = 0 \Rightarrow x = 0$)	Nilpotents preserve "infinitesimal" information
Contravariant functor	Category AffSch ≃ CRing^{op} (affine schemes = rings^{op})

Why schemes:

- Work uniformly over ℚ, ℝ, ℂ, $\mathbb{F}_p$, ℤ
- Number theory: equation $x^2 + y^2 = 1$ over ℤ — this is a scheme
- Moduli: parametrize families of objects

Connection to Other Fields

FIELD	CONNECTION TO AG
Rings	Ring of functions = "algebra of geometry" Ideals = subvarieties
Topology	Zariski topology: closed = V(I) Coarser than Euclidean, but algebraically natural
Categories	Sheaves, cohomology, functors Schemes = contravariant functors
Number theory (5.4)	Diophantine equations = rational points Curves over $\mathbb{F}_p$, Weil conjecture
Complex analysis	Algebraic curves = Riemann surfaces Genus = topological invariant

Table of Categories

Category = (Objects, Morphisms, Composition)
Task: define criterion of "sameness" of objects

CATEGORY	OBJECTS	MORPHISMS	"SAME" =

Set	Sets	Functions	Bijection (equinumerous)
Grp	Groups	Homomorphisms	Group isomorphism
Ab	Abelian groups	Homomorphisms	Isomorphism
Ring	Rings	Ring hom.	Ring isomorphism
Vect_K	Vect. sp. / K	Lin. maps	Isomorphism (dim equal)
Top	Topol. sp.	Continuous	Homeomorphism
Man	Manifolds	Smooth	Diffeomorphism
Met	Metric sp.	Continuous	Isometry
Pos	Partial orders	Monotone	Order isomorphism

CATEGORY	FEATURE
0	Empty: no objects, no morphisms
1	One object *, only identity morphism
2	Two objects 0 → 1, one nontrivial arrow
BG	Group G as category: one object, morphisms = elements of G
Cat	Category of categories: objects = categories, morphisms = functors

Conclusion of Part II: Unity of Structures

Hierarchy of Structures — Main Table with Examples and History

This table shows mathematics as a hierarchy of levels of abstraction.
Each level builds on the previous one.

Horizontal = level of abstraction (bottom to top — from emptiness to physics)
Vertical = type of structure (what information we preserve)

LEVEL	DISCRETE (sets)	ALGEBRAIC (operations)	TOPOLOGICAL (closeness)	ANALYTIC (change)
LEVEL 0 Source ~XIII BC (philos.)	EMPTINESS ∅ "Universe is emptiness"			
LEVEL 1 Choice prelang.	BOUNDARIES/FORMS "Drawing boundaries" Pattern manipulations	(operations will appear later)	(closeness will appear later)	(change will appear later)

```
| LEVEL 2  | SET THEORY      |  RELATIONS      |                 |                 |
| Categori-|  A, B, ∈, ⊆     |  R ⊆ A × B      |    (none yet)   |   (none yet)    |
| zation   |  ∪, ∩, ×, ∅     |  Equivalence    |                 |                 |
| ~1870    |  𝒫(A)           |  Order          |                 |                 |
| Cantor   |  Natural ℕ      |  Function f:A→B |                 |                 |
|          |  from ∅         |                 |                 |                 |
|          |                 |                 |                 |                 |
| Examples:| • ℕ,ℤ,ℚ,ℝ as    | • = (equality)  |                 |                 |
|          |   sets          | • < (order)     |                 |                 |
|          | • Boolean 𝒫(A)  | • f: A → B      |                 |                 |
|          |                 |   (functions)   |                 |                 |
|          |                 |                 |                 |                 |
| tasks:   | • Prove A⊆B     | • Prove that    |                 |                 |
|          | • |𝒫(A)| = 2^|A||   R transitive  |                 |                 |
|          | • Construct ℕ   | • Find f⁻¹(B)   |                 |                 |
|          |   from ∅        |                 |                 |                 |
|          |                 |                 |                 |                 |
| why:     | Basic language  | Describing links|                 |                 |
|          | for everything  | between objects |                 |                 |
|          |                 |                 |                 |                 |
+----------+-----------------+-----------------+-----------------+-----------------+
|          |                 |                 |                 |                 |
| LEVEL 3  |  COMMUNICATION  |      LOGIC      |                 |                 |
| Language | Natural         |  ∀, ∃, ∧, ∨, ⇒  |   (notion of    |  (absent        |
| ~300 BC  | languages       |  Inference rules|   truth)        |   at this       |
|Aristotle | Symbols         |  Proofs         |                 |   level)        |
|          |                 |                 |                 |                 |
| Examples:| • Greek,        | • Syllogisms    |                 |                 |
|          |   Latin, ...    | • Modus ponens  |                 |                 |
|          |                 | • Predicate     |                 |                 |
|          |                 |   logic         |                 |                 |
|          |                 |                 |                 |                 |
| tasks:   | • Translation   | • Check         |                 |                 |
|          |   between       |   correctness   |                 |                 |
|          |   languages     |   of inference  |                 |                 |
|          |                 |                 |                 |                 |
| why:     | Knowledge       | Compensating    |                 |                 |
|          | transmission    | language losses |                 |                 |
|          |                 |                 |                 |                 |
+----------+-----------------+-----------------+-----------------+-----------------+
|          |                 |                 |                 |                 |
| LEVEL 4  |   NUMBERS       |   GROUPS        |  METRIC         |   FUNCTIONS     |
| Basic    |  ℤ, ℚ, ℝ, ℂ     |   (G, ∗)        |  spaces         |   f: ℝ → ℝ      |
| structures| (from ℕ)       |   Subgroups     |   (X, d)        |   Continuous    |
| ~300 BC  |  +, ×, <        |   Isomorphism   |   Open sets     |   Differentiable|
| Euclid   |                 |                 |   Completeness  |                 |
| ~1830    |                 |   rings         |  ~1900 Fréchet  |  ~1670 Newton   |
| Galois   |                 |   (R, +, ×)     |                 |  ~1680 Leibniz  |
| ~1670    |                 |   FIELDS        |                 |                 |
| Newton   |                 |   (F, +, ×)     |                 |                 |
|          |                 |                 |                 |                 |
| Examples:| • ℤ (integers)  | • (ℤ, +)        | • ℝ with |x-y|  | • f(x) = x²     |
|          | • ℚ (rationals) | • (ℝ*, ×)       | • ℝⁿ with Euclid| • sin, cos, exp |
|          | • ℝ (reals)     | • S₃ (permutations)|metric        | • polynomials   |
|          | • ℂ (complex)   | • GL(n) (matrices)|• C[0,1] (functions)|            |
|          |                 |                 |                 |                 |
```

```
| tasks:   | • Solve         | • Find          | • Show          | • Find limit    |
|          |   equation      |   subgroups     |   completeness  | • Find          |
|          | • Prove         | • Check         | • Prove         |   derivative    |
|          |   irrationality |   isomorphism   |   convergence   | • Compute       |
|          |                 |                 |                 |   integral      |
|          |                 |                 |                 |                 |
| why:     | Computation,    | Symmetries,     | Notion of       | Modeling        |
|          | measurement     | cryptography    | distance        | processes       |
|          |                 |                 |                 |                 |
+----------+-----------------+-----------------+-----------------+-----------------+
|          |                 |                 |                 |                 |
| LEVEL 5  |  GRAPHS         |  VECTOR         |  TOPOLOGICAL    |  MEASURES       |
| Complex  |   G = (V, E)    |  SPACES         |  SPACES         |   (X, 𝒜, μ)     |
| structures|  Trees          |   (V, +, ·)     |   (X, τ)        |   σ-algebra     |
| ~1730    |   Networks      |   Basis, dim    |   Compactness   |   Lebesgue      |
| Euler    |                 |   Norm ‖·‖      |   Connectedness |   integral      |
| ~1850    |  combinatorics  |                 |   Homeomorphism |                 |
| Cayley   |   Permutations  |  linear         |  ~1900 Hausdorff|  probability    |
| ~1880    |   Combinations  |  maps           |                 |   (Ω, ℱ, P)     |
| Grassmann|                 |   T: V → W      |                 |   Random var.   |
```

```
| ~1900    |                |   Matrices     |                |   E[X], Var[X] |
| Lebesgue |                |   ker, Im      |                |                |
|          |                |   Eigenvalues  |                |  ~1933 Kolmogorov|
|          |                |                |                |                |
| Examples:| • K₅ (complete)| • ℝⁿ           | • ℝ (standard top.)| • Lebesgue measure|
|          | • Tree         | • C[0,1] (functions)| • Circle S¹  | • Probability   |
|          | • Planar graph | • L² (Hilbert) | • Torus T²     |   measure      |
|          | • Road network | • Matrices m×n | • Cantor set   | • Counting measure|
|          |                |                |                |                |
| tasks:   | • Find         | • Find basis   | • Show         | • Compute      |
|          |   shortest     | • Check linear |   compactness  |   measure      |
|          |   path         |   independence | • Prove        | • Construct    |
|          | • Color        | • Solve linear |   connectedness|   measurable   |
|          |   graph        |   system       | • Find         |   function     |
|          | • Find spanning| • Find eigen-  |   homeomorphism|                |
|          |   tree         |   values       |                |                |
|          |                |                |                |                |
| why:     | Networks, links,| Linear        | "Shape" without| Generalization |
|          | path           | algebra for    | distance,      | of integral,   |
|          | optimization   | everything     | deformations   | probability    |
|          |                |                |                |                |
+----------+----------------+----------------+----------------+----------------+
|          |                |                |                |                |
| LEVEL 6  |     PHYSICS    |  LIE GROUPS    |  MANIFOLDS     |  FUNCTIONAL    |
| Applica- |  "Mathematics =|  Continuous    |   M (smooth)   |  spaces        |
| tion to  |   experimental |  symmetries    |   TₚM (tangent)|   Cᵏ, Lᵖ       |
| world    |   physics"     |                |   Riemannian   |   Banach       |
| ~1915    |                |  ~1870 Lie     |   Curvature    |   Hilbert      |
| Einstein |  categories    |                |  ~1850 Riemann |                |
| ~1930    |   Objects      |  ALGEBRAS      |                |  OPERATORS     |
| categories|   Morphisms   |   over fields  |  bundles       |   Spectrum σ(T)|
| ~1920    |   Functors     |                |   Connections  |   Adjoint      |
| Banach   |                |  Lie algebras  |   Forms        |                |
| ~1945    |                |                |                |  ~1920 Banach  |
| Eilenberg|                |                |                |                |
|          |                |                |                |                |
| Examples:| • GR (geometric)| • SO(3) (rotations)| • Sphere S²  | • L²(ℝ)        |
|          | • QM (operators)| • SU(2) (spin) | • Torus T²     | • Cᵏ[0,1]      |
|          | • Category Set | • U(1) (electr.)| • Spacetime   | • Operators    |
|          | • Functors     |                |   (4-manifold) |   in Hilbert sp.|
|          |                |                |                |                |
| tasks:   | • Find         | • Find Lie     | • Compute      | • Find spectrum|
|          |   symmetries   |   algebra      |   curvature    |   of operator  |
|          | • Construct    | • Find         | • Construct    | • Solve        |
|          |   functor      |   representation|   tangent     |   equation     |
|          | • Apply to     | • Use          |   bundle       |   in Hilbert sp.|
|          |   physics      |   in physics   |                |                |
|          |                |                |                |                |
| why:     | Unified        | Continuous     | Curved         | Quantum        |
|          | structure of all| symmetries,   | spaces,        | mechanics,     |
|          | mathematics,   | gauge          | gravity,       | infinite-dim.  |
|          | application    | theories       | GR             | analysis       |
|          |                |                |                |                |
+----------+----------------+----------------+----------------+----------------+

Legend (bottom to top — from emptiness to physics):
```

[Level 0] = emptiness — source of everything
[level 1] = boundaries/forms — drawing boundaries, pattern manipulations
[Level 2] = sets — categorization, construction of ℕ from ⌀
[level 3] = language and logic — communication, proofs
[Level 4] = basic structures — one kind of structure
[Level 5] = combined — several structures simultaneously
[Level 6] = physics — application to the world, category theory

Vertical connections (what is preserved):

- Discrete: elements, subsets (without operations and distances)
- Algebraic: operations and their properties (∗, +, ×)
- Topological: notion of "closeness" (without distance)
- Analytic: measurement and change (integrals, derivatives)

Key principle:

> Mathematics — a natural hierarchy arising from the fundamental act of drawing boundaries in emptiness. Each level builds on the previous one.

Important observations:

1. History goes bottom to top: ~300 BC → ~1670 → ~1830 → ~1900 → ~1930 → ~1945

2. Each level builds on previous ones: one cannot understand manifolds without topology. One cannot understand topology without sets.

3. Four pillars developed in parallel: but in the 20th century they united

---

Table of Analogies — one pattern in different areas

This table shows: the same pattern manifests in all areas.

"Having learned to understand the pattern, you see it everywhere"

CONCEPT	SETS	ALGEBRA	TOPOLOGY	ANALYSIS
SUBOBJECT	$A \subseteq B$ Subset	$H \leq G$ Subgroup Subring	$U \subseteq X$ Subspace (with induc. topology)	$W \subseteq V$ Subspace (closed)
MORPHISM (mapping, preserving structure)	$f: A \to B$ Function	$\varphi: G \to H$ Homomorphism	$f: X \to Y$ Continuous map	$T: V \to W$ Linear operator

isomorphism	Bijection	G ≅ H	X ≈ Y	V ≅ W

(structures "the same")	f: A ↔ B	Isomorph. of groups	Homeomorph. of spaces	Isomorphism of vect. sp.
PRODUCT (combination)	$A \times B$ Cartesian product Pairs (a,b)	$G \times H$ Direct product Pairs (g,h)	$X \times Y$ Product of topologies Pairs (x,y)	$V \oplus W$ Direct sum Sum v + w
QUOTIENT (gluing)	A / ~ Factor set Equivalence classes	G / H Factor group Cosets	X / ~ Factor space Equivalence classes	V / W Factor space Cosets
duality	$\mathcal{P}(A) \leftrightarrow 2^A$ Boolean ↔ Ind. func.	$G \leftrightarrow \hat{G}$ Group ↔ Characters	Homology ↔ Cohomology	$V \leftrightarrow V^*$ Dual space
basis (minimal description)	Minimal generating	Generating elements	Basis of topology Minimal cover of open sets	Hamel basis (lin. indep.) Minimal set of vectors
KERNEL	$f^{-1}(\{b\})$ Preimage of element	$\ker(\varphi) = \{g: \varphi(g)=e\}$ That which goes to unity	$f^{-1}(\{y\})$ Preimage of point	$\ker(T) = \{v: T(v)=0\}$ That which is killed
IMAGE	f(A) Set of values	$\mathrm{Im}(\varphi) = \{\varphi(g): g \in G\}$ Subgroup in H	f(X) Subspace in Y	$\mathrm{Im}(T) = \{T(v): v \in V\}$ Subspace in W

Additional table — basic operations:

OPERATION	ARITY	EXAMPLES	APPLICATION
Negation	Unary (1)	$\neg P$, A^c, $-x$	Complement, opposite
Addition	Binary (2)	$+$, $\cup$, $\vee$	Union, disjunction
Multiplication	Binary (2)	$\times$, $\cap$, $\wedge$	Intersection, conjunction
Composition	Binary (2)	$f \circ g$, AB	Sequential application
Application	Binary (2)	$f(x)$, $T(v)$	Functions, operators

Philosophy of the table of analogies:

One and the same pattern manifests in all areas of mathematics.

If you understood "substructure" in sets ($A \subseteq B$),
you automatically understand subgroup ($H \leq G$),
subspace ($U \subseteq X$) etc.

Learn horizontally (one concept in all areas),
not vertically (all concepts of one area).

This is the principle of categorical mathematics: to study general patterns, not concrete realizations.

We examined structures on spaces: algebraic (groups, rings), topological (proximity), linear (addition, stretching), differential (smoothness, forms). Categories showed how all these structures are connected.

Now — part III: analysis. This is the art of measuring spaces.

We constructed a space and learned to move through it (groups), understand its shape (topology), work with flat approximations (linear algebra), describe curvature (manifolds). But how to measure what is happening?

A function is a scanner with which an observer scans the space.

Temperature $T(x)$ — a function that assigns a number to each point of the room.
Velocity $v(x,t)$ — a function giving a vector at each point and at each moment.
A function is an observer's tool for obtaining numerical data from space.

Analysis studies how these functions change:
- Derivative — rate of change (local scanner)
- Integral — accumulated effect (global scanner)
- Limit — what happens "at the boundary" of measurements

In terms of duality: to study a space of points is the same as to study the space of functions on it (this is the deep Gelfand–Naimark theorem).

PART III: ANALYSIS OF SPACES

In this part the observer begins to compute for the first time. Everything that was structure before (topology, groups, tensors), now becomes a tool for concrete calculations: limits, derivatives, integrals, series.

Why after manifolds – limits? In Part II we looked at spaces globally: topology, curvature, connectedness. Now we need a microscope – a tool for studying the structure of space near a point. Analysis gives this microscope: the derivative shows local behavior, the integral collects local data into a global result.

But remember: computation depends on the choice of coordinates (observer). The result does not. The integral along a contour does not depend on parametrization. The derivative in a direction is a geometric object.

Analysis – from a concrete space to general structures

Mathematical analysis – calculus of infinitesimals

Analysis as a view of the space $\mathbb{R}^n$

Calculus – this is the study of the concrete space $\mathbb{R}^n$ with all its structure: metric, linearity, order.

We have already seen general structures:

- Topology: what "close" means in general
- Metric: what "distance" means
- Linearity: what "add" and "multiply by a number" mean

Now we look at how all this works together on $\mathbb{R}^n$:

- Limit = topology (convergence in metric)
- Derivative = linear approximation (best linear approximation)
- Integral = measure (generalized "size")

Calculus – not a separate island. This is an example of how general ideas are embodied in a concrete space.

Historical note

Mathematical analysis studies continuous changes:

- How a function changes (derivative)
- How change accumulates (integral)
- What happens "in the limit" (limit)

This is the language of physics, engineering, economics – everything where there is change.

XVII century: Newton and Leibniz independently created analysis
XIX century: Cauchy, Weierstrass made it rigorous (ε-δ definitions)
XX century: generalization to manifolds and infinite-dimensional spaces

Limit — fundamental concept

Below — definitions of limits for metric spaces (via $|x-a|$).
General topological definition (via neighborhoods).
These definitions are equivalent: ε-neighborhood — a special case of neighborhood.

Definitions of limits

TYPE	FORMAL DEFINITION
Limit of a sequence	$\lim x_n = a$ means: $\forall\varepsilon>0\ \exists N\in\mathbb{N}: \forall n>N \Rightarrow \lvert x_n-a\rvert < \varepsilon$ Meaning: after number N all terms in ε-neighborhood of a
Limit of a function at a point	$\lim_{x\to a} f(x) = L$ means: $\forall\varepsilon>0\ \exists\delta>0: 0<\lvert x-a\rvert<\delta \Rightarrow \lvert f(x)-L\rvert < \varepsilon$ Meaning: f(x) close to L when x close to a
Limit at infinity	$\lim_{x\to\infty} f(x) = L$ means: $\forall\varepsilon>0\ \exists M: x>M \Rightarrow \lvert f(x)-L\rvert < \varepsilon$
Infinite limit	$\lim_{x\to a} f(x) = \infty$ means: $\forall M>0\ \exists\delta>0: 0<\lvert x-a\rvert<\delta \Rightarrow f(x) > M$

Examples of limits of sequences

SEQUENCE	LIMIT	EXPLANATION
$x_n = 1/n$	0	1, 0.5, 0.33, 0.25, ... → 0
$x_n = (n+1)/n$	1	2, 1.5, 1.33, ... → 1
$x_n = (1+1/n)^n$	$e \approx 2.718$	Definition of number e
$x_n = n$	$+\infty$	Unboundedly grows
$x_n = (-1)^n$	does not exist	Jumps: 1, −1, 1, −1, ...
$x_n = \sin(n)$	does not exist	Chaotically oscillates

Most important limits of functions

LIMIT	VALUE	APPLICATION
$\lim_{x\to0} \sin(x)/x$	1	First remarkable limit
$\lim_{x\to0} (1+x)^{(1/x)}$	e	Second remarkable limit
$\lim_{x\to0} (e^x-1)/x$	1	Derivative of e^x at zero
$\lim_{x\to0} \ln(1+x)/x$	1	Derivative of ln at one
$\lim_{x\to0} (1-\cos x)/x^2$	1/2	Expansion of cosine
$\lim_{x\to+\infty} x^n/e^x$	0	e^x grows faster than power
$\lim_{x\to+\infty} \ln(x)/x^\alpha$ $(\alpha>0)$	0	Power grows faster than ln

Rules for computing limits

RULE	FORMULA
Sum	$\lim(f+g) = \lim f + \lim g$
Product	$\lim(f\cdot g) = \lim f \cdot \lim g$
Quotient	$\lim(f/g) = \lim f / \lim g$ (if $\lim g \neq 0$)
Constant	$\lim(c\cdot f) = c \cdot \lim f$
Composition	$\lim f(g(x)) = f(\lim g(x))$ if f continuous
Squeeze (two policemen)	$g\leq f\leq h$, $\lim g=\lim h=L \Rightarrow \lim f=L$

Indeterminate forms – When rules don't work

TYPE	EXAMPLE	METHOD OF RESOLUTION
$0/0$	$\sin(x)/x$ at $x\to 0$	L'Hôpital, expansion, change
∞/∞	x^2/e^x at $x\to\infty$	L'Hôpital
$0\cdot\infty$	$x\cdot\ln(x)$ at $x\to 0^+$	Reduce to $0/0$ or ∞/∞
$\infty-\infty$	$x-\sqrt{(x^2+1)}$	Multiply by conjugate
1^∞	$(1+1/x)^x$ at $x\to\infty$	Take logarithm
0^0	x^x at $x\to 0^+$	Take logarithm
∞^0	$x^{(1/x)}$ at $x\to\infty$	Take logarithm

Important: Why division by 0 is not defined (and it is not infinity)

Many think: "$1/0 = \infty$, after all $\lim(1/x)$ at $x\to 0^+$ equals $+\infty$"
This is a mistake. Limit and value of a function – different things.

Why 1/0 is undefined – algebraic argument

Division is the inverse operation of multiplication:

$$a/b = c \quad \text{means} \quad b\cdot c = a$$

Let $1/0 = c$. Then by definition: $0\cdot c = 1$.
But $0\cdot c = 0$ for any c (field axiom).
We get $0 = 1$ – contradiction.

Let $0/0 = c$. Then $0 \cdot c = 0$ – true for any c.
Which c to choose? 0? 1? 1000? No unique answer.

Why limit ≠ value

$\lim_{x \to a} f(x) = L$ means: "f(x) becomes arbitrarily close to L"

This speaks about behavior near point a, but not about value at point a.

Examples:

- $\lim_{x \to 1} (x^2-1)/(x-1) = 2$, but $(x^2-1)/(x-1)$ at $x=1$ is undefined
- $\lim_{x \to 0^+} 1/x = +\infty$, but $1/0$ is undefined (not even "equals $+\infty$")

Additional problem: different limits from different sides

$\lim_{x \to 0^+} 1/x = +\infty$ (from the right)
$\lim_{x \to 0^-} 1/x = -\infty$ (from the left)

Which "infinity" to choose? $+\infty$ or $-\infty$?
If $1/0 = \infty$, then we would get $+\infty = -\infty$, which is absurd.

Conclusion

- $1/0$ – undefined (there does not exist a number c such that $0 \cdot c = 1$)
- $0/0$ – indeterminate (there are infinitely many solutions to $0 \cdot c = 0$)
- $\lim 1/x = \pm\infty$ – this is about behavior, not about value
- In standard arithmetic ∞ – not a number, but a symbol for limit

(In extended number systems ∞ can be introduced as an element,
but then some familiar properties of arithmetic are lost.
Example: on the Riemann sphere $\hat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ the operation $1/0 = \infty$ is valid,
but then $0 \cdot \infty$ and $\infty - \infty$ are undefined. See about complex analysis)

Continuity – absence of discontinuities

Definitions of continuity

FORMULATION	DEFINITION
Via limit	f continuous at a $\iff \lim_{x \to a} f(x) = f(a)$
ε-δ definition	$\forall\varepsilon>0\ \exists\delta>0: \lvert x-a\rvert<\delta \Rightarrow \lvert f(x)-f(a)\rvert<\varepsilon$
Via sequences	$x_n \to a \Rightarrow f(x_n) \to f(a)$ for any (x_n)
Topological definition	Preimage of open set is open $f^{-1}(U)$ is open for any open U

Three conditions for continuity at point a

№	CONDITION	WHAT IS VIOLATED (IF NOT MET)
1	f(a) is defined	Point is punctured
2	$\lim_{x \to a} f(x)$ exists	Jump or oscillation
3	$\lim_{x \to a} f(x) = f(a)$	Removable discontinuity

Types of discontinuities

TYPE	SIGN	EXAMPLE
Removable	lim exists, but f(a)≠lim or f(a) undefined	$f(x)=(x^2-1)/(x-1)$ at x=1 lim=2, but f(1) undef.
Jump (1st kind)	One-sided limits exist, but not equal	sign(x) at x=0 $\lim_- = -1$, $\lim_+ = +1$
Essential (2nd kind)	At least one one-sided limit doesn't exist	sin(1/x) at x=0 Infinite oscillations

Properties of continuous functions

OPERATION	RESULT
f, g continuous at a	f+g, f−g, f·g continuous at a
f, g continuous, g(a)≠0	f/g continuous at a
f cont. at a, g cont. at f(a)	g∘f continuous at a
|f| when f continuous	|f| continuous
max(f,g), min(f,g)	Continuous if f and g are continuous

Fundamental theorems

THEOREM	FORMULATION and COROLLARY
Weierstrass (on extrema)	f cont. on [a,b] ⇒ f attains max and min Corollary: ∃c,d∈[a,b]: f(c)≤f(x)≤f(d) ∀x
Intermediate value (Bolzano)	f cont. on [a,b], f(a)<y<f(b) ⇒ ∃c∈(a,b): f(c)=y "Continuous function doesn't skip values"
On zero of function (corollary)	f cont., f(a)<0<f(b) ⇒ ∃c: f(c)=0 Application: proving existence of roots
Cantor (uniform continuity)	f cont. on [a,b] ⇒ f uniformly continuous δ depends only on ε, not on point

Standard continuous functions

FUNCTION	DOMAIN OF CONTINUITY
Polynomial p(x)	ℝ (everywhere)
Rational p(x)/q(x)	ℝ \ {roots of q(x)}
√x	[0, +∞)
sin x, cos x	ℝ
e^x	ℝ
ln x	(0, +∞)
|x|	ℝ (but not differentiable at 0!)

Derivative – instantaneous rate of change

Visualization: secant → tangent

```
        y
        |             /
        |          /
        |       /     • (a+h, f(a+h))      secant (slope = Δf/Δx)
        |     •/
        |   /
        | •---------------------          tangent (slope = f'(a))
        |  (a, f(a))
        |
        |
        +------------------------------ x
              a    a+h
```

When h → 0 the secant "rotates" and becomes the tangent.
Slope of tangent = derivative = lim[h→0] (f(a+h)−f(a))/h

Definition of derivative

FORMULA	INTERPRETATION
$f'(a) = \lim_{h\to 0} (f(a+h)-f(a))/h$	Limit of the ratio of increments
$f'(a) = \lim_{x\to a} (f(x)-f(a))/(x-a)$	Equivalent form
GEOMETRICALLY	Slope of tangent to graph at point a
PHYSICALLY	Instantaneous velocity (if f = position)
ANALYTICALLY	Best linear approximation: $f(a+h) \approx f(a) + f'(a)\cdot h$

Notations for derivative

NOTATION	WHEN USED
$f'(x)$	General notation (Lagrange)
df/dx	Emphasizes variable of differentiation (Leibniz)
Df	Operator notation
$\dot{f}$ (dot)	Derivative with respect to time (physics)
f'', f''' $f^{(n)}$	Second, third derivatives n-th derivative

Relationship between differentiability and continuity

STATEMENT	TRUE?
Differentiable ⇒ Continuous	YES (always)
Continuous ⇒ Differentiable	NO. Counterexample: $f(x)=\|x\|$ at point 0
Continuous but nowhere diff.	Exists. (Weierstrass function)

Table of derivatives

FUNCTION $f(x)$	DERIVATIVE $f'(x)$
c (constant)	0
x	1
x^n	$n\cdot x^{n-1}$
e^x	e^x
a^x	$a^x \cdot \ln(a)$
$\ln(x)$	$1/x$
$\log_a(x)$	$1/(x\cdot\ln(a))$
$\sin(x)$	$\cos(x)$
$\cos(x)$	$-\sin(x)$
$\tan(x)$	$1/\cos^2(x) = \sec^2(x)$
$\arcsin(x)$	$1/\sqrt{(1-x^2)}$
$\arccos(x)$	$-1/\sqrt{(1-x^2)}$
$\arctan(x)$	$1/(1+x^2)$
$\sqrt{x}$	$1/(2\sqrt{x})$

Rules of differentiation

RULE	FORMULA
Linearity	$(\alpha f + \beta g)' = \alpha f' + \beta g'$
Product (Leibniz)	$(f\cdot g)' = f'\cdot g + f\cdot g'$
Quotient	$(f/g)' = (f'\cdot g - f\cdot g')/g^2$
Chain rule (composition)	$(f(g(x)))' = f'(g(x))\cdot g'(x)$
Inverse function	$(f^{-1})'(y) = 1/f'(x)$, where $y = f(x)$

Examples of applying the chain rule

FUNCTION	DERIVATIVE
$\sin(x^2)$	$\cos(x^2)\cdot 2x = 2x\cdot\cos(x^2)$
$e^{(3x)}$	$e^{(3x)}\cdot 3 = 3e^{(3x)}$
$\ln(x^2+1)$	$2x/(x^2+1)$
$\sqrt{(1-x^2)}$	$-x/\sqrt{(1-x^2)}$
$(x^2+1)^{10}$	$10(x^2+1)^9\cdot 2x = 20x(x^2+1)^9$

Geometric and physical meaning

CONTEXT	MEANING OF DERIVATIVE
Geometry	Slope of tangent
Motion	Velocity ($v = dx/dt$)
Acceleration	Second derivative ($a = d^2x/dt^2$)
Economics	Marginal profit/cost
Heat	Heat flux ($q = -k\cdot dT/dx$)
Optimization	$f'(a) = 0$ – candidate for extremum

Derivative as linear approximation:

$$f(a + h) \approx f(a) + f'(a)\cdot h \quad \text{(for small h)}$$

This is the tangent. Near point a the function behaves almost linearly.

Jacobian matrix – multidimensional derivative

For function $f: \mathbb{R}^n \to \mathbb{R}^m$ the Jacobian matrix is the matrix of partial derivatives:

$$J = \begin{pmatrix} \partial f_1/\partial x_1 & \partial f_1/\partial x_2 & \cdots & \partial f_1/\partial x_n \\ \partial f_2/\partial x_1 & \partial f_2/\partial x_2 & \cdots & \partial f_2/\partial x_n \\ \vdots & \vdots & \ddots & \vdots \\ \partial f_m/\partial x_1 & \partial f_m/\partial x_2 & \cdots & \partial f_m/\partial x_n \end{pmatrix}$$

Geometric Meaning: Local Change of Coordinates

Recall: $f'(a)$ is the slope of the tangent, i.e. linear approximation.

In the multidimensional case: $J(a)$ is a linear mapping which approximates f near the point a:

$$f(a + h) \approx f(a) + J(a)\cdot h \quad \text{(for small h)}$$

This means:
Near point a the transformation f behaves like a linear change of coordinates.

Nonlinear f: linear approximation j:

```
     ╭----╮                          ╭----╮
    /      \                        /      \
   |   ●    |  --→  ╭------╮       |   ●    |  --→  ╭---------╮
   |   a    |       |curved|       |   a    |       |linear   |
    \      /        ╰------╯        \      /        ╰---------╯
     ╰----╯                          ╰----╯
```

The Jacobian says: "locally, on a small scale, f is a linear transf."

Determinant of Jacobian = volume stretching coefficient

$|\det J|$ is by how many times the transformation f stretches volume

Example: Polar coordinates $(r, \theta) \to (x, y) = (r \cos \theta, r \sin \theta)$

$$J = \begin{pmatrix} \partial x/\partial r & \partial x/\partial \theta \\ \partial y/\partial r & \partial y/\partial \theta \end{pmatrix} = \begin{pmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{pmatrix}$$

$$\det J = r\cdot\cos^2\theta + r\cdot\sin^2\theta = r$$

This means: area of "element" $dr\cdot d\theta$ in polar coordinates corresponds to area $r\cdot dr\cdot d\theta$ in Cartesian.

Change of variables in integral:

$$\iint f(x,y)\, dxdy = \iint f(x(u,v), y(u,v)) \cdot |\det J|\, dudv$$

Jacobian — "area conversion coefficient" under change of coordinates.

Inverse Function Theorem

If $\det J(a) \neq 0$, then f is locally invertible near a.

Intuition: If Jacobian is nondegenerate (does not "flatten" space),
then transformation can be inverted in a small neighborhood.

Jacobian of inverse mapping: $J(f^{-1}) = J(f)^{-1}$

Practice: If you want to check whether you can "reverse" transformation
of coordinates — compute determinant of Jacobian. If $\neq 0$ — you can.

Implicit Function Theorem

Given: $F(x, y) = 0$ defines y as a function of x.

Question: When can we express $y = y(x)$?

Answer: If $\partial F/\partial y \neq 0$, then locally $y = y(x)$ exists and

$$\frac{dy}{dx} = -\frac{\partial F/\partial x}{\partial F/\partial y}$$

Example: Circle $x^2 + y^2 = 1$
$F(x,y) = x^2 + y^2 - 1 = 0$
$\partial F/\partial y = 2y \neq 0$ (except $y = 0$, i.e. points $(\pm 1, 0)$)
$dy/dx = -2x/(2y) = -x/y$

At points $(\pm 1, 0)$ tangent is vertical — y is not expressed through x!

Intuition: Curve $F = 0$ locally — graph of function, if not vertical.

Multiple Integrals and Fubini's Theorem

Double integral: $\iint_D f(x,y)\, dA$ — «volume under surface $z = f(x,y)$»

Fubini's Theorem: If f is continuous on rectangle $[a,b]\times[c,d]$:

$$\iint_D f(x,y)\, dA = \int_a^b \left(\int_c^d f(x,y)\, dy\right) dx = \int_c^d \left(\int_a^b f(x,y)\, dx\right) dy$$

Order of integration can be changed.

For more complex domains: limits of inner integral depend on
outer variable.

Example: $\iint_D xy\, dA$, where D: $0 \leq x \leq 1$, $0 \leq y \leq x$

$$= \int_0^1 \left(\int_0^x xy\, dy\right) dx = \int_0^1 x\cdot[y^2/2]_0^x\, dx = \int_0^1 x^3/2\, dx = 1/8$$

Change of variables in multiple integrals:

$$\iint_D f(x,y)\,dxdy = \iint_{D'} f(x(u,v), y(u,v)) \cdot |\det J|\, dudv$$

Polar: $dxdy = r\,dr\,d\theta$
Cylindr.: $dxdydz = r\,dr\,d\theta\,dz$
Spheric.: $dxdydz = r^2 \sin\varphi\,dr\,d\varphi\,d\theta$

Material (Convective) Derivative — Key to Gas Dynamics

Problem: There is a temperature field $T(x, y, z, t)$. How does temperature of fluid particle change, which moves with velocity v?

```
+--------------------------------------------------------------------------+
|                                                                          |
|     t = 0           t = dt                                               |
|                                                                          |
|      ●------------→ ●                                                    |
|   particle       particle                                                |
|   at point x     at point x + v·dt                                      |
|                                                                          |
|   T(x, 0)        T(x + v·dt, dt)                                       |
|                                                                          |
+--------------------------------------------------------------------------+
```

Material derivative:

$$\frac{DT}{Dt} = \frac{\partial T}{\partial t} + (v\cdot\nabla)T = \frac{\partial T}{\partial t} + v_x \frac{\partial T}{\partial x} + v_\gamma \frac{\partial T}{\partial y} + v_z \frac{\partial T}{\partial z}$$

TERM	PHYSICAL MEANING
$\partial T/\partial t$	Local change: field changes at given point
$(v\cdot\nabla)T$	Convective: particle is transported to other point

Example: River with nonuniform temperature

Standing on shore, lowered thermometer: T changes ($\partial T/\partial t \neq 0$) — this is local change.

Floating in boat: even if $\partial T/\partial t = 0$, temperature around us changes, because we move into region with different T — this is convection.

Navier–Stokes equation uses precisely D/Dt:

$$\rho\, Dv/Dt = -\nabla p + \mu\nabla^2 v + \rho g$$

Left side — acceleration of particle (not of point in space)

Integral — accumulation of change

Intuition: area under the graph

Definite integral $\int_a^b f(x)dx$ is the area under the graph f from x = a to x = b (taking into account sign: below axis — negative).

```
          |
       f  |    /\
          |   /██\
          |  /████\
          | /██████\        ← shaded area
        --+-------------
          a         b
```

Physical meaning:

- If f(t) = velocity, then $\int_a^b f(t)dt$ = distance traveled
- If f(t) = power, then $\int_a^b f(t)dt$ = work
- If f(x) = density, then $\int_a^b f(x)dx$ = mass

Definition of Riemann integral

Idea: Approximate area by sum of rectangles, then let width of rectangles tend to zero.

Partition: $a = x_0 < x_1 < x_2 < ... < x_n = b$
Width: $\Delta x_i = x_i - x_{i-1}$
Choose point: $\xi_i \in [x_{i-1}, x_i]$

Integral sum: $S_n = \Sigma_i f(\xi_i)\cdot\Delta x_i$

$$\boxed{\int_a^b f(x)dx = \lim_{\max(\Delta x_i)\to 0} \Sigma_i f(\xi_i)\cdot\Delta x_i}$$

(limit must exist and not depend on choice of partition and ξ_i)

Function is Riemann integrable if this limit exists.

Sufficient condition: Continuous on [a,b] ⇒ integrable.

Concrete example: $\int_0^1 x\,dx$

Geometrically: area of triangle with vertices (0,0), (1,0), (1,1).
Answer: $\frac{1}{2} \cdot 1 \cdot 1 = 1/2$

Via integral sums:

Partition $[0,1]$ into n equal parts: $\Delta x = 1/n$

Take right endpoints: $\xi_i = i/n$

$$S_n = \Sigma_{i=1}^n (i/n)\cdot(1/n) = (1/n^2)\, \Sigma_{i=1}^n i = (1/n^2)\cdot n(n+1)/2$$
$$= (n+1)/(2n) = 1/2 + 1/(2n)$$

$$\lim_{n\to\infty} S_n = \lim_{n\to\infty} (1/2 + 1/(2n)) = 1/2 \quad ✓$$

Properties of integral

1. Linearity:
   $\int_a^b (\alpha f + \beta g)dx = \alpha\int_a^b f\, dx + \beta\int_a^b g\, dx$

2. Additivity over interval:
   $\int_a^b f\, dx + \int_b^c f\, dx = \int_a^c f\, dx$

3. Monotonicity:
   $f(x) \le g(x)$ on $[a,b] \Rightarrow \int_a^b f\, dx \le \int_a^b g\, dx$

4. Estimate:
   $m \le f(x) \le M$ on $[a,b] \Rightarrow m(b-a) \le \int_a^b f\, dx \le M(b-a)$

5. Mean value theorem:
   If f continuous on $[a,b]$, then $\exists c \in (a,b)$: $\int_a^b f\, dx = f(c)\cdot(b-a)$

Fundamental theorem of calculus — connection of derivative and integral

Newton–Leibniz theorem

If $F'(x) = f(x)$, then:

$$\int_a^b f(x)dx = F(b) - F(a)$$

Notation: $F(x)|_a^b = F(b) - F(a)$

Meaning: To compute integral, it suffices to find antiderivative F (function whose derivative equals f) and take difference.

Example:

$\int_0^2 x^2\, dx = ?$

$F(x) = x^3/3$ (check: $(x^3/3)' = 3x^2/3 = x^2$ ✓)

$\int_0^2 x^2\, dx = F(2) - F(0) = 8/3 - 0 = 8/3$

Second part of fundamental theorem

Let f be continuous. Define:

$F(x) = \int_a^x f(t)dt$ ("integral with variable upper limit")

$$\boxed{\frac{d}{dx} \int_a^x f(t)dt = f(x)}$$

Meaning: Derivative of integral = integrand.
Integration and differentiation — inverse operations.

d/dx and $\int dx$ — like exponentiation and logarithm, or + and −.
One "cancels" the other.

Table of antiderivatives (indefinite integrals)

FUNCTION $f(x)$	ANTIDERIVATIVE $\int f(x)dx$
x^n $(n \neq -1)$	$x^{n+1}/(n+1) + C$
$1/x$	$\ln\lvert x\rvert + C$
e^x	$e^x + C$
a^x	$a^x/\ln(a) + C$
$\sin(x)$	$-\cos(x) + C$
$\cos(x)$	$\sin(x) + C$
$1/\cos^2(x)$	$\tan(x) + C$
$1/\sin^2(x)$	$-\cot(x) + C$
$1/(1+x^2)$	$\arctan(x) + C$
$1/\sqrt{(1-x^2)}$	$\arcsin(x) + C$
$1/(x^2+a^2)$	$(1/a)\arctan(x/a) + C$
$1/\sqrt{(x^2 \pm a^2)}$	$\ln\lvert x + \sqrt{(x^2 \pm a^2)}\rvert + C$

Methods of integration

1. change of variable (substitution):
   $\int f(g(x))g'(x)dx = \int f(u)du$, where $u = g(x)$

   Example: $\int 2x \cdot e^{x^2} dx = \int e^u du = e^u + C = e^{x^2} + C$ $(u = x^2)$

2. integration by parts:
   $\int u\, dv = uv - \int v\, du$

   Example: $\int x \cdot e^x dx$
   $u = x$, $dv = e^x dx \Rightarrow du = dx$, $v = e^x$
   $= x \cdot e^x - \int e^x dx = x \cdot e^x - e^x + C = e^x(x-1) + C$

3. partial fraction decomposition:
   For rational functions $P(x)/Q(x)$

4. trigonometric substitutions:
   $\sqrt{(a^2-x^2)}$: $x = a\cdot\sin(\theta)$
   $\sqrt{(a^2+x^2)}$: $x = a\cdot\tan(\theta)$
   $\sqrt{(x^2-a^2)}$: $x = a\cdot\sec(\theta)$

Example for thermophysics: Carnot cycle as contour integral

Work of gas per cycle = $\oint P\,dV$ = area inside cycle on P-V diagram

```
      P
      |     1
      |    /\  isotherm T₁ (heater)
      |   /  \
      |  /    \ 2
      | | work  |  adiabat
      | |  W>0  |
      |  \    / 3
      |   \  /
      |    \/  isotherm T₂ (cooler)
      |     4
      +------------------ V
```

Why $\oint P\,dV \neq 0$?
  If $P = P(V)$ were function of only V, then $\oint P\,dV = 0$ always.
  But $P = P(V, T)$ — depends on T also.
  On upper isotherm pressure is higher at same V → work is positive.

Efficiency of Carnot cycle:

$$\eta = W/Q_1 = (Q_1 - Q_2)/Q_1 = 1 - T_2/T_1$$

  This is maximum efficiency for any engine between T_1 and T_2.
  (Follows from 2nd law of thermodynamics)

Connection with entropy:
  For reversible process: $\oint dQ/T = 0$
  This means: $\oint dS = 0$ (entropy — state function)
  Mathematically: dS — exact differential, but dQ — not.

Moral: Contour integral $\oint = 0$ only for exact differential.
       dQ not exact → can extract work. dS exact → returns.

Applications of Analysis

Applications of Derivatives

1. finding extrema:
   $f'(x) = 0$ – necessary condition for local extremum
   $f''(x) < 0 \Rightarrow$ local maximum
   $f''(x) > 0 \Rightarrow$ local minimum

2. function analysis:
   $f' > 0$: increasing
   $f' < 0$: decreasing
   $f'' > 0$: concave up (∪)
   $f'' < 0$: concave down (∩)

3. L'Hôpital's rule (for indeterminate forms 0/0 or ∞/∞):
   $\lim_{x\to a} f(x)/g(x) = \lim_{x\to a} f'(x)/g'(x)$ (if the right limit exists)

4. Taylor series (approximation of function by polynomial):
   $f(x) = f(a) + f'(a)(x-a) + f''(a)(x-a)^2/2! + f'''(a)(x-a)^3/3! + \dots$

   Reference table of expansions (a = 0, Maclaurin series):

FUNCTION	EXPANSION	RADIUS R
e^x	$1 + x + x^2/2! + x^3/3! + \dots$	∞
$\sin x$	$x - x^3/3! + x^5/5! - x^7/7! + \dots$	∞
$\cos x$	$1 - x^2/2! + x^4/4! - x^6/6! + \dots$	∞
$\sinh x$	$x + x^3/3! + x^5/5! + x^7/7! + \dots$	∞
$\cosh x$	$1 + x^2/2! + x^4/4! + x^6/6! + \dots$	∞
$1/(1-x)$	$1 + x + x^2 + x^3 + \dots$ (geometric)	1
$\ln(1+x)$	$x - x^2/2 + x^3/3 - x^4/4 + \dots$	1
$\arctan x$	$x - x^3/3 + x^5/5 - x^7/7 + \dots$	1
$(1+x)^\alpha$	$1 + \alpha x + \alpha(\alpha-1)x^2/2! + \dots$ (binomial)	1
Special cases:		
$\sqrt{(1+x)}$ ($\alpha=\frac{1}{2}$)	$1 + x/2 - x^2/8 + x^3/16 - \dots$	1
$1/(1+x)$ ($\alpha=-1$)	$1 - x + x^2 - x^3 + \dots$	1
$1/(1+x)^2$ ($\alpha=-2$)	$1 - 2x + 3x^2 - 4x^3 + \dots$	1

   R = radius of convergence: series converges for $|x| < R$, diverges for $|x| > R$.

Applications of Integrals

1. area between curves:
   $S = \int_a^b |f(x) - g(x)|\, dx$

2. volume of solid of revolution:
   $V = \pi \int_a^b [f(x)]^2\, dx$ (rotation around x-axis)

3. arc length:
   $L = \int_a^b \sqrt{(1 + [f'(x)]^2)}\, dx$

4. surface area of revolution:
   $S = 2\pi \int_a^b f(x)\sqrt{(1 + [f'(x)]^2)}\, dx$

5. physical applications:
   - Work: $W = \int F\, dx$
   - Center of mass: $\bar{x} = \int x\cdot\rho(x)dx / \int\rho(x)dx$
   - Moment of inertia: $I = \int r^2\cdot dm$

Connection with Heat Engineering

Heat conduction equation:

$$\partial T/\partial t = \alpha \cdot \partial^2 T/\partial x^2$$

This is a partial differential equation.
Temperature T depends on time t and coordinate x.

Fourier's law (heat flux):
$q = -k \cdot dT/dx$
Flux is proportional to temperature gradient.
Minus: heat flows from hot to cold.

Stationary case ($\partial T/\partial t = 0$):
$d^2T/dx^2 = 0 \Rightarrow T(x) = ax + b$ (linear profile)

Integral = heat accumulation:
$Q = \int_0^t P(\tau)d\tau$ (energy = integral of power over time)

Convergence of Series — When Infinite Sum Makes Sense

Problem: what does "infinite sum" mean?

Expression $a_1 + a_2 + a_3 + ... = \Sigma_n a_n$ is not an ordinary sum.
We cannot add infinitely many numbers directly.

Definition via limit of partial sums:

$$S_n = a_1 + a_2 + ... + a_n \quad \text{(n-th partial sum)}$$

```
+-----------------------------------------------------+
```

Series Σa_n converges if a finite limit exists:

$$\sum_{n=1}^{\infty} a_n = \lim_{n\to\infty} S_n = S$$

Otherwise the series diverges.

Key Examples

1. geometric series:
   $\sum_{n=0}^{\infty} r^n = 1 + r + r^2 + r^3 + \ldots$

   - $|r| < 1$: converges to $1/(1-r)$
     Example: $1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \ldots = 2$

   - $|r| \geq 1$: diverges
     Example: $1 + 1 + 1 + \ldots = \infty$

2. harmonic series:
   $\sum_{n=1}^{\infty} 1/n = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \ldots = \infty$ (diverges).

   Although terms → 0, sum is still infinite.
   (Proof: group $1/3+1/4 > 1/2$, $1/5+\ldots+1/8 > 1/2$, etc.)

3. p-series:
   $\sum_{n=1}^{\infty} 1/n^p$ converges $\iff p > 1$

   - $\Sigma\, 1/n^2 = \pi^2/6$ (Basel problem, Euler 1734)
   - $\Sigma\, 1/n = \infty$ (harmonic, $p = 1$)

4. alternating series:
   $1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \ldots = \ln(2)$ converges (conditionally)

Convergence Tests – Brief Table

TEST	FORMULATION
Necessary (but not sufficient)	If Σa_n converges, then $a_n \to 0$ Converse is FALSE: $a_n \to 0 \nRightarrow$ convergence
Comparison	$0 \leq a_n \leq b_n$, Σb_n conv. $\Rightarrow \Sigma a_n$ conv.
d'Alembert (ratio)	$\lim \|a_{n+1}/a_n\| = L$: $L<1$ conv., $L>1$ div.
Cauchy (root)	$\lim \sqrt[n]{\|a_n\|} = L$: $L<1$ conv., $L>1$ div.
Integral	$\Sigma f(n)$ conv. $\iff \int f(x)dx$ conv. ($f\downarrow$, $f>0$)
Leibniz (alternating)	$(-1)^n a_n$, $a_n \downarrow 0 \Rightarrow$ converges

Absolute vs Conditional Convergence

- Σa_n converges absolutely if $\Sigma|a_n|$ converges
- Σa_n converges conditionally if Σa_n conv. but $\Sigma|a_n|$ div.

Important: Absolute convergence ⇒ convergence (converse is false)

Example:
$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + ...$ converges conditionally
(because $1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + ... = \infty$)

Riemann's theorem: A conditionally convergent series can by rearrangement of terms be made to converge to any number (or diverge)!

An absolutely convergent series can be rearranged arbitrarily — the sum does not change.

Differential equations — equations with derivatives

Differential equations as a view on space

Everything that changes is described by a differential equation.

Temperature of a rod, oscillation of a pendulum, population growth, nuclear decay, asset price, tank heating — all these processes are united by one thing: the rate of change of a quantity depends on the quantity itself.

A differential equation is a law relating the state of a system to how fast it changes. The solution of a DE is not a number, but a function: the complete trajectory of the system in time.

In terms of spaces: an ODE defines a vector field on the space of states. At each point — an arrow "where to move". The solution is an integral curve of this field.

What is a differential equation

A differential equation (DE) — an equation containing an unknown function $y(x)$ and its derivatives y', y'', ...

Examples:
- $y' = ky$ (exponential growth/decay)
- $y'' + y = 0$ (harmonic oscillator)
- $y' = y(1-y)$ (logistic growth)

The solution of a DE is a function $y(x)$ satisfying the equation.

The order of a DE is the highest derivative in the equation.

$y' = ky$ – first order

$y'' + y = 0$ – second order

Classification

TYPE	DESCRIPTION
ODE (ordinary)	One independent variable: $y(x)$ Example: $y' = f(x, y)$
PDE (partial deriv.)	Several independent variables: $u(x,y,t)$ Example: $\partial u/\partial t = \alpha \cdot \partial^2 u/\partial x^2$ (heat conduction)
Linear	y and derivatives enter linearly Example: $y'' + p(x)y' + q(x)y = f(x)$
Nonlinear	y or derivatives in power > 1, in products Example: $y' = y^2$, $(y')^2 + y = 1$
Homogeneous	$f(x) = 0$ (right-hand side equals zero)
Nonhomogeneous	$f(x) \neq 0$

How to independently formulate a differential equation

Differential equations don't fall from the sky – they are constructed from physical principles. Here is an algorithm for constructing any meaningful DE:

```
  +---------------------+
  |   Physical          |
```

```
  |   object            |  →  What are we studying?
  +---------+-----------+
            ↓
  +---------------------+
  |   Physical          |
  |   process           |  →  What is happening?
  +---------+-----------+
            ↓
  +---------------------+
  |   Elementary        |
  |   volume            |  →  dx, dy, dz, dt
  +---------+-----------+
            ↓
  +---------------------+
  |   Conservation      |
  |   law               |  →  What doesn't change?
  +---------+-----------+
            ↓
  +---------------------+
  |  Differential       |
  |  equation           |  →  Mathematical model
  +---------------------+
```

Example: heat conduction equation

---

1. object: rod
2. process: heat propagation along the rod
3. elementary volume: dx — small piece of rod
4. conservation law: energy (heat flowing in = heat flowing out + accumulation)

- Heat flux at point x: $q(x) = -k \cdot \partial T/\partial x$ (Fourier's law)
- Heat flux at point x+dx: $q(x+dx) = -k \cdot \partial T/\partial x|_{x+dx}$
- Heat accumulation in volume: $\rho c \cdot \partial T/\partial t \cdot dx$
- Balance: $q(x) - q(x+dx) = \rho c \cdot \partial T/\partial t \cdot dx$
- Expansion: $q(x+dx) \approx q(x) + \partial q/\partial x \cdot dx$
- Result: $-\partial q/\partial x = \rho c \cdot \partial T/\partial t$
- Substitution: $k \cdot \partial^2 T/\partial x^2 = \rho c \cdot \partial T/\partial t$

Result: $\partial T/\partial t = \alpha \cdot \partial^2 T/\partial x^2$, where $\alpha = k/(\rho c)$ — thermal diffusivity

---

Important note

Practically any DE used in physics (Navier–Stokes, Maxwell, diffusion equation) can technically be solved even in Excel.

Each cell corresponds to one linear equation, referencing neighboring cells. At the boundary, constants are given (boundary conditions).

This is the essence of numerical methods: replace derivatives with differences, continuous — with discrete.

Most important DEs and their solutions

1. exponential growth/decay:
   $y' = ky$
   Solution: $y = Ce^{kx}$

   Examples: radioactive decay ($k<0$), population growth ($k>0$),
   heating/cooling (Newton's law)

2. harmonic oscillator:
   $y'' + \omega^2 y = 0$
   Solution: $y = A\cdot\cos(\omega x) + B\cdot\sin(\omega x) = C\cdot\cos(\omega x + \varphi)$

   Examples: pendulum, spring, oscillations in circuits

3. damped oscillations:
   $y'' + 2\gamma y' + \omega^2 y = 0$
   Characteristic equation: $\lambda^2 + 2\gamma\lambda + \omega^2 = 0$

   - $\gamma < \omega$: damped oscillations (underdamping)
   - $\gamma = \omega$: critical damping
   - $\gamma > \omega$: aperiodic regime (overdamping)

4. heat conduction equation (PDE):
   $\partial T/\partial t = \alpha\cdot\partial^2 T/\partial x^2$

   Solution by separation of variables or Fourier series

5. wave equation (PDE):
   $\partial^2 u/\partial t^2 = c^2\cdot\partial^2 u/\partial x^2$

   Solution: $u = f(x-ct) + g(x+ct)$ (waves traveling left and right)

Methods for solving first-order ODEs

1. separation of variables:
   If $y' = f(x)g(y)$, then $dy/g(y) = f(x)dx$ → integrate both sides

   Example: $y' = xy$
   $dy/y = x\,dx \rightarrow \ln|y| = x^2/2 + C \rightarrow y = Ae^{x^2/2}$

2. linear ODE of 1st order:
   $y' + p(x)y = q(x)$

   Integrating factor: $\mu(x) = e^{\int p(x)dx}$
   Solution: $y = (1/\mu)\int \mu q\,dx$

3. homogeneous equation:
   $y' = f(y/x)$
   Substitution: $v = y/x$, then $y = vx$, $y' = v + xv'$

Initial and boundary conditions

The general solution of a DE contains arbitrary constants.
To find a particular solution, additional conditions are needed:

Initial conditions (Cauchy problem):
  Values of the function and derivatives at the initial moment
  $y(0) = y_0$, $y'(0) = v_0$, ...

Boundary conditions:
  Values at the boundaries of the domain
  $y(a) = A$, $y(b) = B$

Existence and uniqueness theorem (Picard):
  For $y' = f(x,y)$, $y(x_0) = y_0$, if f is continuous and satisfies
  the Lipschitz condition in y, then the solution exists and is unique.

Methods for solving second-order ODEs

Linear ODE of 2nd order with constant coefficients:
  $ay'' + by' + cy = f(x)$

Step 1: Solve the homogeneous equation ($f(x) = 0$)

  Seek solution of the form $y = e^{lx}$. Substitute:
  $a\cdot\lambda^2 e^{lx} + b\cdot\lambda e^{lx} + c\cdot e^{lx} = 0$
  $e^{lx}(a\lambda^2 + b\lambda + c) = 0$

CHARACTERISTIC EQUATION: $a\lambda^2 + b\lambda + c = 0$

General solution of the homogeneous depending on roots λ_1, λ_2:

DISCRIMINANT	GENERAL SOLUTION
$D > 0$ (two real)	$y = C_1e^{l_1x} + C_2e^{l_2x}$ ($\lambda_1 \neq \lambda_2$ – different real)
$D = 0$ (one repeated)	$y = (C_1 + C_2x)e^{lx}$ ($\lambda_1 = \lambda_2 = \lambda$)
$D < 0$ (complex)	$y = e^{\alpha x}(C_1\cos(\beta x) + C_2\sin(\beta x))$ where $\lambda = \alpha \pm i\beta$

Step 2: Find a particular solution of the nonhomogeneous (method of variation of parameters
       or method of undetermined coefficients)

Step 3: general solution = general homogeneous + particular nonhomogeneous

Example: $y'' + 4y = 0$

$\lambda^2 + 4 = 0 \rightarrow \lambda = \pm 2i \rightarrow y = C_1\cos(2x) + C_2\sin(2x)$

Laplace transform — powerful tool

DEFINITION:

$$\mathcal{L}\{f(t)\} = F(s) = \int_0^\infty e^{-st} f(t)\, dt$$

Transforms a time function $f(t)$ into a complex frequency function $F(s)$

Why needed:

- Transforms DEs into algebraic equations
- Automatically accounts for initial conditions
- Ideal for linear systems, control theory

Key property — derivative:

$\mathcal{L}\{f'(t)\} = sF(s) - f(0)$

$\mathcal{L}\{f''(t)\} = s^2F(s) - sf(0) - f'(0)$

Differentiation → multiplication by s (plus initial conditions)

Table of Laplace transforms

$f(t)$	$F(s) = \mathcal{L}\{f(t)\}$
1	$1/s$
t	$1/s^2$
t^n	$n!/s^{n+1}$
e^{at}	$1/(s-a)$
$\sin(\omega t)$	$\omega/(s^2+\omega^2)$
$\cos(\omega t)$	$s/(s^2+\omega^2)$
$e^{at}\sin(\omega t)$	$\omega/((s-a)^2+\omega^2)$
$e^{at}\cos(\omega t)$	$(s-a)/((s-a)^2+\omega^2)$
$\delta(t)$ (delta function)	1
$u(t)$ (unit step)	$1/s$

Example: solving DE by Laplace transform

```
Problem: y'' + 4y = 0,  y(0) = 1,  y'(0) = 0

Step 1: Apply ℒ to both sides
  ℒ{y''} + 4ℒ{y} = 0
  [s²Y(s) − sy(0) − y'(0)] + 4Y(s) = 0
  [s²Y(s) − s·1 − 0] + 4Y(s) = 0

Step 2: Solve algebraic equation
  (s² + 4)Y(s) = s
  Y(s) = s/(s² + 4)

Step 3: Inverse transform (from table)
  Y(s) = s/(s² + 2²) → y(t) = cos(2t)

Answer: y(t) = cos(2t)


Table of Fourier transforms

Definition: ℱ{f(t)} = F(ω) = ∫_{−∞}^{∞} f(t)·e^{−iωt} dt

+-----------------------------+------------------------------+
| f(t)                        | F(ω) = ℱ{f(t)}               |
+-----------------------------+------------------------------+
| δ(t)                        | 1                            |
+-----------------------------+------------------------------+
| 1                           | 2π·δ(ω)                      |
+-----------------------------+------------------------------+
| e^{iω₀t}                    | 2π·δ(ω − ω₀)                 |
+-----------------------------+------------------------------+
| cos(ω₀t)                    | π[δ(ω−ω₀) + δ(ω+ω₀)]         |
+-----------------------------+------------------------------+
| sin(ω₀t)                    | πi[δ(ω+ω₀) − δ(ω−ω₀)]        |
+-----------------------------+------------------------------+
| e^{−a|t|}  (a>0)            | 2a/(a² + ω²)                 |
+-----------------------------+------------------------------+
| e^{−at²}  (Gaussian)        | √(π/a)·e^{−ω²/(4a)}          |
+-----------------------------+------------------------------+
| rect(t/τ) (rectangle)       | τ·sinc(ωτ/2)                 |
+-----------------------------+------------------------------+
| u(t)·e^{−at} (a>0)          | 1/(a + iω)                   |
+-----------------------------+------------------------------+
| t·u(t)·e^{−at}              | 1/(a + iω)²                  |
+-----------------------------+------------------------------+

Key properties:
  Linearity:    ℱ{af + bg} = aF + bG
  Shift:        ℱ{f(t−t₀)} = e^{−iωt₀}F(ω)
  Modulation:   ℱ{e^{iω₀t}f(t)} = F(ω − ω₀)
  Convolution:  ℱ{f * g} = F · G
  Derivative:   ℱ{f'(t)} = iω·F(ω)
  Parseval:     ∫|f(t)|²dt = (1/2π)∫|F(ω)|²dω
```

Summary table: basic PDEs of mathematical physics

EQUATION	FORMULA	TYPE	PHYSICS
Heat conduction	$\partial u/\partial t = \alpha\nabla^2 u$	Parabolic	Diffusion, heat exchange
Wave	$\partial^2 u/\partial t^2 = c^2\nabla^2 u$	Hyperbolic	Oscillations, waves, sound
Laplace	$\nabla^2 u = 0$	Elliptic	Stationary fields, potentials
Poisson	$\nabla^2 u = f$	Elliptic	Electrostatics, gravitation
Schrödinger	$i\hbar\partial\psi/\partial t = \hat{H}\psi$	Parabolic (in t)	Quantum mechanics
Navier–Stokes	$\partial v/\partial t + (v\cdot\nabla)v = -\nabla p/\rho + \nu\nabla^2 v$	Nonlinear.	Hydrodynamics

Classification by characteristics (for linear 2nd order):

TYPE	DISCRIMINANT	BEHAVIOR OF SOLUTIONS
Elliptic (Laplace)	$B^2 - 4AC < 0$ (no real char.)	Smooth, max inside Boundary value problems
Parabolic (Heat cond.)	$B^2 - 4AC = 0$ (one char.)	Smoothing, diffusion Initial + boundary cond.
Hyperbolic (Wave)	$B^2 - 4AC > 0$ (two char.)	Waves, discontinuities Cauchy problem

Boundary conditions – how to pose a problem

PDEs without boundary conditions have infinitely many solutions.

Boundary conditions select a unique physically meaningful one.

TYPE	FORMULA	PHYSICAL MEANING
Dirichlet	$u\|_{\partial\Omega} = g$	Temperature specified at boundary (thermostat)
Neumann	$\partial u/\partial n\|_{\partial\Omega} = h$	Heat flux specified through boundary (insulation when $h=0$)
Robin (third kind)	$(\partial u/\partial n + \alpha u)\|_{\partial\Omega} = h$	Convective heat exchange with environment (Newton–Richmann law): $q = \alpha(T_{surf} - T_{env})$

For an engineer: Dirichlet – thermostat, Neumann – insulation/heater,
Robin – free heat exchange with air.

Method of separation of variables – example

Problem: $\partial T/\partial t = \alpha \cdot \partial^2 T/\partial x^2$ on $[0, L]$, $T(0,t) = T(L,t) = 0$, $T(x,0) = f(x)$

Idea: seek $T(x,t) = X(x)\cdot\theta(t)$ – product of function of x and function of t.

Substitution: $X\cdot\theta' = \alpha\cdot X''\cdot\theta \;\rightarrow\; \theta'/(\alpha\theta) = X''/X = -\lambda$ (= const)

Two ODEs instead of one PDE:

$$X'' + \lambda X = 0, \quad X(0) = X(L) = 0 \quad \rightarrow \quad X_n = \sin(n\pi x/L), \quad \lambda_n = (n\pi/L)^2$$

$$\theta' + \alpha\lambda_n\theta = 0 \quad \rightarrow \quad \theta_n = e^{-\alpha(n\pi/L)^2 t}$$

General solution (superposition):

$$T(x,t) = \sum_n b_n \cdot \sin(n\pi x/L) \cdot e^{-\alpha(n\pi/L)^2 t}$$

Coefficients from initial condition (Fourier series):

$$b_n = (2/L) \int_0^L f(x)\cdot\sin(n\pi x/L)\, dx$$

Physical meaning: each harmonic decays exponentially,
high frequencies (large n) decay faster – heat "smooths out".

```
================================================================================
Uniform Convergence — Key Concept
================================================================================

 Problem: When can we interchange limits?


 Suppose fₙ(x) → f(x) as n → ∞. Are the following equalities valid?

   lim ∫fₙ(x)dx  =?  ∫ lim fₙ(x) dx     (limit of integral = integral of limit)
   n→∞                 n→∞

   lim fₙ'(x)    =?  (lim fₙ(x))'        (limit of derivatives = derivative)
   n→∞                 n→∞

 Answer: In general — no.

 Counterexample:
   fₙ(x) = xⁿ on [0,1]
   fₙ(x) → f(x) = { 0, if x < 1
                  { 1, if x = 1
   ∫₀¹ fₙ dx = 1/(n+1) → 0
   ∫₀¹ f dx  = 0       ✓ (coincided accidentally)

   But f is discontinuous, although all fₙ are continuous.
   The limit of continuous functions need not be continuous.


 Two Types of Convergence


 +--------------------------------------------------------------------------+
 | POINTWISE CONVERGENCE                                                    |
 | fₙ → f pointwise, if ∀x: lim fₙ(x) = f(x)                               |
 |                           n→∞                                            |
 |                                                                          |
 | Formally: ∀x ∀ε>0 ∃N(x,ε): n>N ⇒ |fₙ(x)−f(x)| < ε                      |
 |                    ↑                                                     |
 |           N depends on x!                                                |
 +--------------------------------------------------------------------------+

 +--------------------------------------------------------------------------+
 | UNIFORM CONVERGENCE                                                      |
 | fₙ ⇉ f uniformly, if convergence is "equally fast" for all x           |
 |                                                                          |
 | Formally: ∀ε>0 ∃N(ε): ∀x ∀n>N ⇒ |fₙ(x)−f(x)| < ε                       |
 |                  ↑                                                       |
 |         N does not depend on x!                                          |
 +--------------------------------------------------------------------------+

 Equivalent condition:
   fₙ ⇉ f  ⟺  sup|fₙ(x) − f(x)| → 0
              x
```

Visually:

Uniform convergence = graph of f_n entirely fits in an ε-tube around the graph of f (for sufficiently large n)

```
      f(x)+ε -------------------------
             ~~~~~~~~fₙ(x)~~~~~~~~~   ← fₙ inside the tube
      f(x)   =========================
             ~~~~~~~~~~~~~~~~~~~~~~~~
      f(x)−ε -------------------------
```

Why Uniform Convergence is Important

Theorems on interchange of limits:

```
+-------------------------------------------------------+
| If fₙ ⇉ f UNIFORMLY on [a,b]:                         |
|                                                       |
| 1. LIMIT PRESERVES CONTINUITY:                        |
| fₙ continuous ⇒ f continuous                          |
|                                                       |
| 2. ONE CAN INTERCHANGE LIMIT and INTEGRAL:            |
| lim ∫fₙ dx = ∫ lim fₙ dx = ∫f dx                      |
| n→∞         n→∞                                       |
|                                                       |
| 3. ONE CAN INTERCHANGE LIMIT and DERIVATIVE:          |
| If fₙ' ⇉ g uniformly and fₙ(x₀) converges, then f' = g |
+-------------------------------------------------------+
```

Without uniformity these assertions are false.

Criterion for Uniform Convergence of Series (Weierstrass Test)

Let $\Sigma u_n(x)$ be a functional series.

```
+--------------------------------------------------------------------------+
| WEIERSTRASS TEST (M-test):                                               |
|                                                                          |
| If there exist numbers Mₙ ≥ 0 such that:                                 |
|   1. |uₙ(x)| ≤ Mₙ for all x                                              |
|   2. Σ Mₙ converges (numerical series)                                   |
|                                                                          |
| Then Σuₙ(x) converges UNIFORMLY and ABSOLUTELY.                          |
+--------------------------------------------------------------------------+
```

Example: Series $\Sigma\, x^n/n^2$ on $[-1, 1]$

$|x^n/n^2| \leq 1/n^2 = M_n$

$\Sigma\, 1/n^2$ converges (p-series, $p=2>1$)

⇒ series converges uniformly on $[-1, 1]$

Power Series and Uniform Convergence

Power series $\Sigma a_n x^n$ has radius of convergence r:

- $|x| < R$: series converges absolutely
- $|x| > R$: series diverges
- $|x| = R$: need to check separately

Cauchy–Hadamard formula: $1/R = \limsup_{n\to\infty} \sqrt[n]{|a_n|}$

Key fact:

On any interval $[-r, r]$ with $r < R$ the power series converges uniformly.

Therefore one can differentiate and integrate power series term-by-term inside the circle of convergence.

Applied Example: Storage Tank Dynamics

Problem: Hot water storage tank. Inflow and outflow vary with time. How does water temperature change?

```
    Gᵢₙ, Tᵢₙ
       ↓
  +===========+
  |           |      V – tank volume
  |    T(t)   |      T(t) – water temperature
  |           |      Gᵢₙ, Gₒᵤₜ – flow rates (kg/s)
  +===========+
       ↓
    Gₒᵤₜ, T
```

Heat balance equation:

```
ρVc · dT/dt = Gᵢₙ·c·(Tᵢₙ − T) + Qₗₒₛₛ
    ↑            ↑                  ↑
accumulation   heat inflow      losses to outside
```

Derivative – rate of change:

dT/dt = rate of temperature change [°C/s]

- $dT/dt > 0$: tank is heating up
- $dT/dt < 0$: tank is cooling down
- $dT/dt = 0$: steady-state regime (temperature is constant)

Integral — accumulation:

Accumulated heat over time [0, t]:

$$Q = \int_0^t G_{in} \cdot c \cdot (T_{in} - T)\, d\tau \quad [J]$$

If $G_{in} = 1$ kg/s, $T_{in} = 80°C$, $T = 60°C$, $c = 4200$ J/(kg·K):
Over 1 hour (3600 s): $Q = 1 \times 4200 \times 20 \times 3600 = 302$ MJ

Numerical example:

$V = 1$ m³, $\rho = 1000$ kg/m³, $c = 4200$ J/(kg·K)
$G_{in} = 0.1$ kg/s, $T_{in} = 90°C$, initial $T_0 = 20°C$
Losses: $Q_{losses} = -100 \cdot (T - T_{am}\beta)$, $T_{am}\beta = 20°C$

Time constant: $\tau = \rho Vc / (G_{in} \cdot c + k) \approx 4.2 \times 10^6 / (420 + 100)$
≈ 8000 s ≈ 2.2 hours

Over time τ temperature reaches ~63% of steady-state value.

Moral: Derivative = instantaneous rate of change.
Integral = accumulated effect over time. This is the language of thermal dynamics.

Calculus works with functions as objects. But the set of functions — is also a space, and an infinite-dimensional one at that. Is it possible to do linear algebra on it?

Yes. Series are coordinates of a function in the basis $\{1, x, x^2, ...\}$ or $\{e^{inx}\}$. This is a bridge to functional analysis — linear algebra on spaces of functions.

Series and functional spaces — bridge to functional analysis

Series as a view of the space of functions

Functions form an infinite-dimensional space. Fourier and Taylor series — are coordinates of a function in this space relative to a chosen basis.

- Taylor basis: $1, x, x^2, x^3, ...$ (power functions)
- Fourier basis: $e^{inx} = \cos(nx) + i \cdot \sin(nx)$ (harmonics)

Series coefficients = function coordinates = projections onto basis vectors.
This turns analysis into linear algebra of infinite dimension.

Prerequisite knowledge: For Fourier series the basis e^{inx} is used.
Reminder: Euler's formula $e^{ix} = \cos x + i \sin x$.

Why this section is here

This section is a bridge between calculus and functional analysis.

Key idea: functions form a vector space.
This thought turns analysis into linear algebra of infinite dimension.

Taylor and Fourier series are expansions of functions in a basis.
Fourier coefficients are coordinates of the function in this basis.

Understanding this section makes functional analysis natural.

Key discovery

Functions form a vector space.

- Can be added: $(f + g)(x) = f(x) + g(x)$
- Can be multiplied by a number: $(cf)(x) = c\cdot f(x)$
- Has a zero element: $f(x) = 0$

This means all linear algebra works for functions.
- Basis
- Expansion in a basis
- Scalar product
- Orthogonality

One pattern — three realizations

FINITE-DIMENSIONAL (ordinary vectors)	TAYLOR SERIES	FOURIER SERIES
Space $\mathbb{R}^n$	Smooth functions	Periodic functions
Basis: $e_1, e_2, \ldots, e_n$	Basis: $1, x, x^2, x^3, \ldots$	Basis: e^{inx} ($n \in \mathbb{Z}$)
Vector $v = \Sigma\, v_i e_i$	$f(x) = \Sigma\, a_n x^n$	$f(x) = \Sigma\, c_n e^{inx}$
Coefficient $v_i = v\cdot e_i$ (scalar product)	$a_n = f^{(n)}(0)/n!$ (derivative)	$c_n = \langle f, e^{inx}\rangle$ (integral)
Dimension: n	Dimension: ∞	Dimension: ∞

Scalar product of functions

For vectors: $\langle u, v\rangle = \Sigma_i\, u_i v_i$

For functions: $\langle f, g\rangle = \int f(x)\bar{g}(x)\, dx$ ← sum is replaced by integral.

(For complex functions we take the conjugate $\bar{g}$, so that $\langle f,f\rangle \geq 0$)

Orthogonality: $\langle f, g\rangle = 0$

Important: This scalar product turns functions into a Hilbert space L^2. The entire construction of Fourier series is an application of orthogonal projections in an infinite-dimensional Hilbert space.

Fact: functions e^{inx} are orthogonal to each other.

$$\langle e^{imx}, e^{inx}\rangle = \int_{-\pi}^{\pi} e^{i(m-n)x}\, dx = 0 \quad \text{for } m \neq n$$

Therefore Fourier coefficients are found in the same way as in the finite-dimensional case:

$$c_n = \langle f, e^{inx}\rangle / \langle e^{inx}, e^{inx}\rangle \quad \text{(projection onto basis vector)}$$

Comparison: local vs global

TAYLOR	FOURIER
Expansion AROUND POINT (local information)	Expansion OVER ENTIRE PERIOD (global information)
Coefficients from DERIVATIVES at point	Coefficients from INTEGRALS over period
Requires infinite differentiability	Works even for discontinuous functions.
Application: • Approximate calc. • Behavior analysis • Solving DEs by series	Application: • Signal analysis • Sound/image processing • Compression (JPEG, MP3)

Concrete example: pressure pulsations in a pipeline

Problem: A pump creates pressure pulsations. A sensor takes a signal p(t).
Question: what frequencies are present? Is there resonance with the pipe?

Signal from sensor (conditionally):

```
  p(t) |    /\    /\    /\    /\
       |   /  \  /  \  /  \  /  \   ← main pump frequency
       |  /~~~~\/~~~~\/~~~~\/~~~~\  ← "ripple" on peaks
       |--------------------------→ t
```

Fourier expansion:

$$p(t) = p_0 + A_1\sin(\omega t) + A_2\sin(2\omega t) + A_3\sin(3\omega t) + \dots$$

- p_0 = average pressure (constant component)
- ω = pump rotation frequency (fundamental harmonic)
- 2ω, 3ω, ... = higher harmonics ("ripple")

Spectrum (amplitudes):

```
  |Aₙ| | █
       | █
       | █  █
       | █  █  █
       | █  █  █  █
       |-----------------→ n (harmonic number)
         2  3  4
```

What the spectrum gives:
- Peak at frequency 47 Hz → this is the pump shaft rotation frequency
- Peak at 94 Hz (2×47) → two-bladed impeller
- Unexpected peak at 120 Hz → resonance with pipe's natural frequency.

Conclusion: Fourier turns "mush" in time into a clear picture by frequencies.

Formulas (for calculations)

Taylor around a:

$$f(x) = f(a) + f'(a)(x-a) + f''(a)(x-a)^2/2! + f'''(a)(x-a)^3/3! + \dots$$

Important series (a = 0):

$$e^x = 1 + x + x^2/2! + x^3/3! + \dots$$
$$\sin x = x - x^3/3! + x^5/5! - \dots$$
$$\cos x = 1 - x^2/2! + x^4/4! - \dots$$
$$1/(1-x) = 1 + x + x^2 + x^3 + \dots \quad (|x| < 1)$$

Fourier:

$$f(x) = \Sigma_n c_n e^{inx}, \quad \text{where} \quad c_n = (1/2\pi) \int_{-\pi}\hat{}\pi\, f(x)\, e^{-inx}\, dx$$

Magic of Fourier: convolution ↔ multiplication

Convolution — the most important operation in signal processing:

$$(f * g)(t) = \int f(\tau)\, g(t - \tau)\, d\tau$$

Meaning: "smearing" function f with "kernel" g
- Signal filtering
- Smoothing (Gaussian blur)
- Response of linear system

Problem: Convolution — complex operation (integral over all shifts)

Convolution theorem:

```
+-------------------------------------------------------+
|                                                       |
| F[f * g] = F[f] · F[g]                                |
|                                                       |
| Fourier transform turns convolution into multiplication. |
|                                                       |
+-------------------------------------------------------+
```

Visualization:

Time domain (complex) — Frequency domain (simple)

$$\begin{tikzcd} f(t), g(t) \arrow[r, "F"] \arrow[d, "\text{convolution}"] & F(\omega), G(\omega) \arrow[d, "\text{multiplication}"] \\ (f * g)(t) & F(\omega) \cdot G(\omega) \arrow[l, "F^{-1}"'] \end{tikzcd}$$

Practical significance:
Instead of complex integral ($O(n^2)$) we do:

1. FFT of input — $O(n \log n)$
2. Multiplication — $O(n)$
3. Inverse FFT — $O(n \log n)$

Total: $O(n \log n)$ instead of $O(n^2)$ — huge gain.

Application examples:

- Image processing (filters in Photoshop)
- Speech recognition
- Solving differential equations
- Neural networks (convolutional layers)
- Multiplication of large numbers (Schönhage–Strassen algorithm)

Inverse theorem:
$F[f \cdot g] = F[f] * F[g]$ (multiplication ↔ convolution in reverse direction)

Deep meaning:
Fourier transform is an isomorphism between two algebras:
(functions, convolution) ≅ (functions, multiplication)
Complex structure in one world = simple in another.

Where it leads

Functional analysis: study of infinite-dimensional spaces of functions
Hilbert space $L^2 = \{f: \int|f|^2 < \infty\}$ with scalar product

Quantum mechanics: state $\psi \in L^2$, observables — operators
Expansion of ψ in eigenfunctions = superposition of states

Signal processing: FFT (fast Fourier transform)
$O(n \log n)$ instead of $O(n^2)$ — revolution in computations

Representation theory: basis e^{inx} = representation of group $U(1)$.
Generalization to other groups → harmonic analysis

Series showed: functions are points in infinite-dimensional space.
Now we need a theory of such spaces. What does "norm of a function" mean?
When does a series converge? Which operators are continuous?

Functional analysis is linear algebra in infinite dimension.
All the same (norms, scalar products, projections), but with new effects.

Functional analysis — infinite-dimensional spaces

Main idea: functions are "points" in infinite-dimensional space

This section is not about new formulas. It's about changing perspective.

In school: function $f(x) = x^2$ is a "rule", formula, graph.

New view: function f is a point in the space of all functions.

Analogy: finite → infinite

In $\mathbb{R}^3$ a point is given by 3 coordinates:

$v = (v_1, v_2, v_3)$ — three numbers determine a vector

Function f on [0,1] is given by infinitely many coordinates:

$f = (f(0), f(0.01), f(0.02), ..., f(0.99), f(1))$

Each value f(t) is a "coordinate" of the function in "direction t".

```
y
↑     ╭--╮
|    /    \
|   /      \
|  /        \     f(t)
| /          \
+----------------→ t
   0.5    1
```

This curve is not a graph. It's a single point f in infinite-dimensional space.

Why this is useful

If functions are points of a space, then:

FINITE-DIMENSIONAL	INFINITE-DIMENSIONAL (functions)
Vector $v \in \mathbb{R}^n$	Function $f \in L^2$
Length $\|v\| = \sqrt{(\Sigma v_i^2)}$	Norm $\|f\| = \sqrt{(\int\lvert f\rvert^2 dx)}$
Distance $\|v-w\|$	Distance $\|f-g\|$ between functions
Angle $\cos \theta = (v \cdot w)/\|v\|\|w\|$	"Angle" via $\langle f,g \rangle = \int fg\, dx$
Orthogonality $v \perp w$	Orthogonality $\langle f,g \rangle = 0$
Basis $\{e_1,...,e_n\}$	Basis $\{\sin(nx), \cos(nx)\}$ or $\{e^{inx}\}$
Decomposition $v = \Sigma v_i e_i$	Fourier series $f = \Sigma c_n e^{inx}$
Projection onto subsp.	Best approximation (least squares)
Linear operator A	Differential operator d/dx
Eigenvalues	Spectrum of operator (resonances)

All of linear algebra works for functions.
But there are subtleties: infinity creates new effects (see below).

Traps of infinite-dimensionality — where intuition from $\mathbb{R}^n$ breaks

Property	In $\mathbb{R}^n$	In L^2 / ℓ^2
Closed ball $\{\|x\| \leq 1\}$	Compact (Heine–Borel)	not compact. (infinite seq. without conv. subseq)
All norms	Equivalent	not equivalent. $\|f\|_1 \neq \|f\|_2 \neq \|f\|\_\infty$
Convergence $x_n \to x$	By any norm the same	Different convergences. Pointwise ≠ in norm
Orthogonal basis	FINITE	Countable or uncountable (sin, cos, ...)
Operator with eigenval.	Diagonalizable (for symm.)	May not have eigenvect. Spectrum ≠ eigenvalues

Main thought: In infinite-dimensional space there is much "freedom", and pathologies impossible in $\mathbb{R}^n$ become the norm. Functional analysis teaches to recognize when finite-dimensional intuition works, and when it doesn't.

Why functional analysis is needed

Linear algebra works with $\mathbb{R}^n$ — finite-dimensional spaces.
But spaces of functions are infinite-dimensional.

Functional analysis = linear algebra + analysis + topology for infinite-dimensional spaces

Applications:

- Quantum mechanics (states = vectors in L^2)
- Differential equations (operator methods)
- Signal processing (Fourier transform)
- Machine learning (kernel methods, RKHS)

Hierarchy of spaces

```
+----------------------------------------------------------------------+
|                                                                      |
|  Vector space                                                        |
|         |                                                            |
|         | + norm ‖·‖                                                 |
|         ▼                                                            |
|  Normed space                                                        |
|         |                                                            |
|         | + completeness (all Cauchy converge)                       |
|         ▼                                                            |
|  Banach space                                                        |
|         |                                                            |
|         | + norm from scalar product: ‖x‖ = √⟨x,x⟩                   |
|         ▼                                                            |
|  Hilbert space                                                       |
|                                                                      |
+----------------------------------------------------------------------+
```

Critical difference from the finite-dimensional case

In $\mathbb{R}^n$: X compact $\Longleftrightarrow$ X closed and bounded (Heine–Borel theorem)

In infinite-dimensional space this is false.

Unit ball $B = \{f \in L^2 : \|f\| \leq 1\}$:

- Closed ✓
- Bounded ✓
- not compact. ✗

Why: sequence $e_n = (0,...,0,1,0,...)$ with one in the n-th place lies in the unit ball, but has no convergent subsequence ($\|e_n - e_m\| = \sqrt{2}$ for $n \neq m$).

Consequence for an engineer:

- In $\mathbb{R}^n$: minimum of a continuous function on a compact is attained
- In L^2: minimum of a functional may not be attained on closed bounded
- Therefore PDE are harder than ODE: special methods are needed (weak solutions)

Compactness in the infinite-dimensional case:
Additional conditions are needed – for example, equicontinuity (Arzelà-Ascoli theorem) or weak compactness.

Norm – generalization of length

Norm $\lVert\cdot\rVert: V \to \mathbb{R}$ must satisfy:

AXIOM	MEANING
$\lVert x\rVert \geq 0$	Length is non-negative
$\lVert x\rVert = 0 \iff x = 0$	Only the zero vector has zero length
$\lVert\alpha x\rVert = \lvert\alpha\rvert\cdot\lVert x\rVert$	Scaling
$\lVert x+y\rVert \leq \lVert x\rVert+\lVert y\rVert$	Triangle inequality

Norm induces a metric: $d(x,y) = \lVert x-y\rVert$

Examples of norms and spaces

SPACE	NORM	COMPLETENESS
$\mathbb{R}^n$ (Euclidean)	$\lVert x\rVert_2 = \sqrt{(\Sigma x_i^2)}$	Yes (Banach and Hilbert)
$\mathbb{R}^n$ (Manhattan)	$\lVert x\rVert_1 = \Sigma\lvert x_i\rvert$	Yes (Banach)
$\mathbb{R}^n$ (sup-norm)	$\lVert x\rVert_\infty = \max\lvert x_i\rvert$	Yes (Banach)
C[a,b] (cont. functions)	$\lVert f\rVert_\infty = \max\lvert f(x)\rvert$	Yes (Banach)
$L^p[a,b]$ $(1 \leq p < \infty)$	$\lVert f\rVert_p = (\int\lvert f\rvert^p)^{(1/p)}$	Yes (Banach) When p=2: Hilbert
$L^2[a,b]$ (square integr.)	$\lVert f\rVert_2 = \sqrt{(\int\lvert f\rvert^2)}$ $\langle f,g\rangle = \int f\cdot\bar{g}$	Yes (Hilbert) Main in quantum mechanics
ℓ^2 (sequences)	$\lVert x\rVert = \sqrt{(\Sigma\lvert x_n\rvert^2)}$	Yes (Hilbert) Infinite-dimensional analog $\mathbb{R}^n$

Important: elements of L^p are not functions.

Elements of L^p are equivalence classes of functions.
Two functions f and g are equivalent if f = g almost everywhere
(i.e. differ only on a set of measure zero).

Consequence: For $f \in L^2$ the value $f(x_0)$ at a specific point is not defined.

- One can change f at a single point – it's the same function in L^2
- "$f(0) = 3$" has no meaning for an element of L^2
- The integral $\int f$ is defined, but the value $f(x)$ is not.

When one can speak of values:
- If f is continuous (then there is a unique continuous representative)
- Through embedding theorems (Sobolev): $W^{1,p} \subset C$ when $p > n$

Where this is critical:
- Boundary conditions in PDE: $u|_{\partial\Omega} = g$ requires "trace" of function
- Delta function $\delta(x-a)$ "selects" value – but this is not $f(a)$.

Unit ball – "portrait" of the norm

Unit ball $B = \{x : \|x\| \leq 1\}$ shows the geometry of the norm.
The shape of the ball determines what the norm considers "close to zero".

```
  L¹ (Manhattan)           L² (Euclidean)        L∞ (sup-norm)
  ‖x‖₁ = |x| + |y|         ‖x‖₂ = √(x²+y²)       ‖x‖_∞ = max(|x|,|y|)

         ◆                      ●●●                  ■■■■■■■
        ◆ ◆                   ●●   ●●                ■     ■
       ◆   ◆                 ●       ●               ■     ■
      ◆     ◆               ●         ●              ■     ■
     ◆-------◆             ●-----------●            ■-------■
      ◆     ◆               ●         ●              ■     ■
       ◆   ◆                 ●       ●               ■     ■
        ◆ ◆                   ●●   ●●                ■     ■
         ◆                      ●●●                  ■■■■■■■

     Diamond                  Circle                Square
  "Taxi in Manhattan"       "Usual distance"      "Chess king"
```

Interpretation:

- L^1: "How many blocks to walk?" – sum of deviations along axes
  Points (1,0), (0,1), (0.5, 0.5) are equidistant from zero.

- L^2: "As the crow flies" – usual Euclidean distance
  Pythagoras: diagonal $\sqrt{2}$, not 2

- L^∞: "Worst case" – maximum deviation along any axis
  Points (1,0), (1,1), (1, 0.5) are equidistant from zero.

General case L^p: $\|x\|_p = (|x|^p + |y|^p)^{(1/p)}$

$p = 1$:   diamond (sharp angles)
$p = 2$:   circle
$p \to \infty$:   square

As p increases, the ball "inflates" from diamond to square through circle.

Application:
- L^1 in optimization: gives sparse solutions (LASSO regression)
- L^2 in physics: energy, least squares
- L^∞ in engineering: control of maximum deviation

Completeness – key property

CONCEPT	DEFINITION / Example
Fundamental (Cauchy) sequence	$\forall\varepsilon>0\ \exists N: m,n>N \Rightarrow \lVert x_m-x_n\rVert<\varepsilon$ "Terms approach each other"
Complete space	Every fund. seq. converges in this space
Example of incompl.	$\mathbb{Q}$: seq. 3, 3.1, 3.14, 3.141... → $\pi \notin \mathbb{Q}$
Completion	$\mathbb{Q} \to \mathbb{R}$ (added all limits) $C[0,1]$ with $\lVert\cdot\rVert_2 \to L^2[0,1]$

Why completeness is needed

APPLICATION	WHY COMPLETENESS IS NEEDED
Iteration method $x_{n+1} = f(x_n)$	Guarantee that the limit exists
Fourier series	$\Sigma c_n e^{inx}$ converges in L^2, not pointwise
Banach fixed point theorem	Contraction in complete space has fix. pt.
Solving DE by Picard method	Iterations converge to solution

Completeness vs closedness – don't confuse.

Closedness is a property of a subset (within some space)
Completeness is a property of a space (in itself)

Key example: $\mathbb{Q}$ (rational numbers)

- $\mathbb{Q}$ is closed IN itself (as a topological space)
- $\mathbb{Q}$ is not complete. (seq. 3, 3.1, 3.14... → $\pi \notin \mathbb{Q}$)

Closedness says: "contains all its limit points"
But if the limit point doesn't exist in the space – it doesn't count.

Connection:
- Closed subset of a complete space is complete
- Complete subset of a metric space is closed

Example for an engineer:
Space C[0,1] with norm $\|f\|_2 = \sqrt{\int |f|^2}$ is not complete.
The limit may be a discontinuous function (not in C[0,1]).
Completion: $L^2[0,1]$ — already complete (with Lebesgue integral).

Why this is important:
With Riemann integral the space of functions is "holey", like $\mathbb{Q}$.
L^2 is complete only with Lebesgue integral — that's why Lebesgue is needed.

Hilbert space

Definition: Banach + norm from scalar product $\|x\| = \sqrt{\langle x,x\rangle}$

PROPERTY	CONSEQUENCE
Has scalar product $\langle\cdot,\cdot\rangle$	Can speak of angles, orthogonality
Pythagorean theorem	$x \perp y \Rightarrow \|x+y\|^2 = \|x\|^2 + \|y\|^2$
Orthogonal complement	$H = M \oplus M^\perp$ for closed M
Orthonormal basis	$x = \Sigma_n \langle x,e_n\rangle e_n$ (generalized Fourier series)
Riesz–Fréchet theorem	Any functional $f(x) = \langle x,y\rangle$ for unique y Consequence: $H \cong H^*$ (isomorphic to dual)

Operators in Hilbert space

Operator $T: H \to H$ is linear if $T(\alpha x+\beta y) = \alpha Tx + \beta Ty$

TYPE	DEFINITION	PROPERTIES
Bounded	$\|Tx\| \leq C\|x\|$ for all x $\|T\| = \sup\{\|Tx\|: \|x\|=1\}$	$\iff$ continuous Operator norm
Compact	Image of bounded set is precompact	"Almost finite-dimensional" Spectrum is discrete + 0
Self-adjoint ($T^* = T$)	$\langle Tx,y\rangle = \langle x,Ty\rangle$	Eigenvalues are real Eigenvectors are orthogonal
Unitary ($U^*U = I$)	$\langle Ux,Uy\rangle = \langle x,y\rangle$	Preserves norm and angles $\lvert\lambda\rvert = 1$ for eigenvalues
Projector	$P^2 = P$	If $P^*=P$: orthogonal Projection onto subspace

Spectral theory

Spectrum: $\sigma(T) = \{\lambda\in\mathbb{C}: (T-\lambda I) \text{ not invertible}\}$

PART OF SPECTRUM	DEFINITION
Point σ_p	λ is eigenvalue: $\exists x\neq 0: Tx=\lambda x$
Continuous σ_c	$(T-\lambda I)$ injective, image dense, but not closed
Residual σ_r	$(T-\lambda I)$ injective, image not dense

Why spectrum ≠ eigenvalues (concrete example)

In the finite-dimensional case (matrices) spectrum = eigenvalues.
In the infinite-dimensional case this is false.

Example: Right shift operator on ℓ^2

$S: (x_1, x_2, x_3, ...) \mapsto (0, x_1, x_2, x_3, ...)$

Eigenvalues: none.
Let $Sx = \lambda x$. Then $(0, x_1, x_2, ...) = (\lambda x_1, \lambda x_2, \lambda x_3, ...)$
From the first component: $0 = \lambda x_1$
If $\lambda \neq 0$, then $x_1 = 0$, hence $x_2 = 0$, ..., $x = 0$. Contradiction.
If $\lambda = 0$, then $(0, x_1, x_2, ...) = 0$, hence $x = 0$. Contradiction.

Spectrum: $\sigma(S) = \{\lambda : |\lambda| \leq 1\}$ – the entire unit disk.
Because $(S - \lambda I)$ is not invertible for all $|\lambda| \leq 1$.
When $|\lambda| < 1$: image not dense (residual spectrum σ_r).
When $|\lambda| = 1$: image dense, but not closed (continuous spectrum σ_c).

Moral: In infinite-dimensional space an operator can be non-invertible
Not because of eigenvectors, but because of "almost eigen" directions.

Simpler example: Multiplication operator $Tf(x) = x\cdot f(x)$ on $L^2[0,1]$
Eigenfunctions: none (δ-function not in L^2).
Spectrum: $\sigma(T) = [0,1]$ – the entire interval (continuous spectrum).
Physically: this is the position operator in quantum mechanics.

This is critical in quantum mechanics: the energy spectrum of a particle in a potential can be continuous (free particle) or discrete (atom).

Spectral theorem (for self-adjoint compact operator)

STATEMENT	CONSEQUENCE
Spectrum = real eigenvalues + possibly 0	All $\lambda_n \in \mathbb{R}$
Eigenvectors are orthonormal basis	$\{e_n\}$ is basis in H
$T = \Sigma_n \lambda_n \langle\cdot,e_n\rangle e_n$	Decomposition of operator by eigenvect.

Why self-adjointness is critical for physics

Fact: Spectrum of a self-adjoint operator is always real.

Physical meaning:

In quantum mechanics observable quantities (energy, momentum, position) are represented by self-adjoint operators. The spectrum of the operator is the set of possible measurement results.

Energy must be a real number — we cannot measure "complex energy". That is precisely why the Hamiltonian H must be self-adjoint: $H = H^*$.

If an operator is not self-adjoint — its spectrum may be complex, and it does not describe an observable physical quantity.

Conclusion: Self-adjointness is not a mathematical abstraction.
It is a requirement of physical meaningfulness of measurements.

Connection with quantum mechanics

QUANTUM MECHANICS	FUNCTIONAL ANALYSIS
State of system	Vector $\psi \in H$, $\|\psi\| = 1$
Observable (energy, etc.)	Self-adjoint operator A
Possible measurement results	Spectrum $\sigma(A)$
Probability of result λ	$\lvert\langle\psi,e\lambda\rangle\rvert^2$ where $Ae\lambda = \lambda e\lambda$
State after measurement	Projection onto eigen-subspace

Concrete example: cooling of a rod (heat equation)

Problem: Metal rod of length L. Ends are maintained at T=0.

Initial temperature distribution $T(x,0) = f(x)$. How does the rod cool?

```
+--------------------------------------------------------------------------+
|    T=0                                        T=0                        |
|     |                                          |                         |
|     ▼         initial T(x,0) = f(x)           ▼                         |
|   --█████████████████████████████████████████--                         |
|     x=0                                      x=L                        |
+--------------------------------------------------------------------------+
```

Equation:

$$\partial T/\partial t = a\cdot\partial^2 T/\partial x^2 \qquad (a \text{ – thermal diffusivity})$$

Boundary conditions: $T(0,t) = T(L,t) = 0$

Initial condition: $T(x,0) = f(x)$

Solution via Eigenfunctions

Step 1: Find eigenfunctions of operator d^2/dx^2 with given boundary conditions

$$d^2\varphi/dx^2 = \lambda\varphi, \quad \varphi(0) = \varphi(L) = 0$$

Solution: $\varphi_n(x) = \sin(n\pi x/L)$, $\lambda_n = -(n\pi/L)^2$, $n = 1, 2, 3, \ldots$

Step 2: These functions form an orthonormal basis in $L^2[0,L]$.

$$\langle\varphi_m, \varphi_n\rangle = \int_0^L \sin(m\pi x/L)\cdot\sin(n\pi x/L)\,dx = (L/2)\cdot\delta_{mn}$$

Step 3: Expand initial condition in this basis

$$f(x) = \sum_n c_n \sin(n\pi x/L), \qquad c_n = (2/L)\int_0^L f(x)\sin(n\pi x/L)\,dx$$

Step 4: Each mode decays independently with exponential $e^{(\lambda_n\cdot a\cdot t)}$

$$T(x,t) = \sum_n c_n \cdot \underset{\substack{\uparrow\\ \text{decay}}}{e^{(-a(n\pi/L)^2 t)}} \cdot \underset{\substack{\uparrow\\ \text{spatial form}}}{\sin(n\pi x/L)}$$

Physical meaning

- First mode ($n=1$): $\tau_1 = L^2/(\pi^2 a)$ – slowest decay
- Higher modes decay faster: $\tau_n = \tau_1/n^2$
- After time $\sim\tau_1$ only first mode remains (sinusoid)

Numerical example (copper rod $L=1$m, $a\approx1.1\times10^{-4}$ m²/s):

$$\tau_1 = 1^2/(\pi^2\cdot1.1\times10^{-4}) \approx 920 \text{ s} \approx 15 \text{ minutes}$$

Moral: Functional analysis is not abstraction.
Operators, eigenfunctions, basis expansions — these are working
tools for solving equations of heat conduction, diffusion, oscillations.

Important theorems of functional analysis

THEOREM	FORMULATION and APPLICATION
Banach fixed-point theorem	T contraction in complete $(X,d) \Rightarrow \exists!\ x: Tx=x$ Appl: existence of solutions of DE
Hahn–Banach	Functional from subspace extends to entire space preserving norm
Banach–Steinhaus (uniform boundedness)	Pointwise bounded family of operators uniformly bounded
Open mapping	Surjective bounded operator — open mapping
Closed graph	Operator with closed graph is bounded

Weak convergence — key to infinite-dimensionality

In $\mathbb{R}^n$ there is only one way of convergence. In infinite-dimensional — many.

TYPE OF CONVERGENCE	DEFINITION
STRONG (by norm)	$x_n \to x$ means $\|x_n - x\| \to 0$ "Distance to limit tends to zero"
WEAK	$x_n \rightharpoonup x$ means $f(x_n) \to f(x)$ for all $f \in X^*$ "All functionals converge"
WEAK-* (in dual)	$f_n \rightharpoonup^* f$ means $f_n(x) \to f(x)$ for all $x \in X$ "Pointwise convergence of functionals"

Relations:
  Strong ⇒ Weak ⇒ Weak-* (converses false)

Key example: $e_n = (0,...,0,1,0,...)$ in ℓ^2

- $\|e_n - e_m\| = \sqrt{2}$ for $n \neq m$ — no strong limit
- BUT: $e_n \rightharpoonup 0$ weakly. (for any $f \in (\ell^2)^*$, $f(e_n) \to 0$)

Why weak convergence is needed:

- Unit ball not compact in strong topology
- but compact in weak topology. (Banach–Alaoglu theorem)
- This allows one to find minima of functionals

Application in PDE:
Sequence of approximate solutions u_n may not have strong limit, but have weak limit u – this is a weak solution.

Riesz representation theorem

There are two Riesz theorems. Both fundamental.

Riesz–Fréchet (for Hilbert spaces)

Any continuous linear functional $f: H \to \mathbb{R}$ on Hilbert space H has form:

$$f(x) = \langle x, y \rangle \quad \text{for unique } y \in H$$

Corollary: $H \cong H^*$ (Hilbert space is isomorphic to its dual). This is not true for general Banach spaces.

Riesz–Markov–Kakutani (for measures)

Any positive linear functional $I: C(X) \to \mathbb{R}$ on the space of continuous functions on a compact X is represented by an integral:

$$I(f) = \int_X f \, d\mu \quad \text{for a unique measure } \mu$$

Meaning: Every method of "weighted averaging" is an integral.

Application: If there is an operator that assigns a number to a function (and preserves linearity and order) – it is an integral with respect to some measure.

Measure and Lebesgue integral – the correct notion of "size"

Functional analysis uses integrals everywhere: norm $\|f\| = \sqrt{\int|f|^2}$, scalar product $\langle f,g \rangle = \int fg$. But which integral? The Riemann integral breaks down on limits – one cannot interchange lim and $\int$.

The Lebesgue integral solves this problem. It is based on the notion of measure – a more general way of measuring the "size" of sets. This is the foundation of probability theory and modern analysis.

Measure as a view of space

Measure answers the question: what "size" does a subset of a space have?

- On the line $\mathbb{R}$: measure of an interval = its length
- On the plane $\mathbb{R}^2$: measure of a region = its area
- In $\mathbb{R}^3$: measure of a body = its volume
- In a space of functions: measure = probability

Lebesgue measure is the "correct" way to measure the size of sets, which works even for very complex (fractal, discontinuous) sets. Without it, L^2 spaces and probability would have no rigorous meaning.

Connection with topology — Hausdorff dimension:
Fractals have fractional dimension. The Cantor set $\subset [0,1]$:
- Topological dimension = 0 (totally disconnected)
- Hausdorff dimension = $\log(2)/\log(3) \approx 0.631$

This is a measure of "complexity" of a set — how much space it occupies.
The coastline of Britain: $\dim_H \approx 1.25$ (more than a line, less than a plane).

The main reason for transitioning to Lebesgue (for an engineer)

The Riemann integral breaks down on limits.
The Lebesgue integral does not.

This is not an abstraction. This is critical for:

- Fourier series: $\lim\int f_n = \int\lim f_n$ — Lebesgue's theorem is needed
- Solving PDEs: limit of approximate solutions → solution
- Probability: $E[\lim X_n] = \lim E[X_n]$ — dominated convergence

The Dirichlet function (1 on $\mathbb{Q}$, 0 on $\mathbb{R}\setminus\mathbb{Q}$) is an artificial example.
The real reason: one needs to interchange limit and integral.
For Riemann — impossible. For Lebesgue — possible (under certain conditions).

The problem of the Riemann integral

PROBLEM	DESCRIPTION
Dirichlet function	$f(x) = 1$ if $x\in\mathbb{Q}$, otherwise 0 not Riemann integrable Though intuitively: $\mathbb{Q}$ "occupies 0%"
Limit does not preserve integrability	$\lim f_n$ can be non-integrable, even if all f_n are integrable
Cannot interchange $\int$ and lim	$\lim\int f_n \neq \int\lim f_n$ in general

Measure — axiomatics

Measure $\mu: \Sigma \to [0,+\infty]$ on a σ-algebra Σ of subsets of X

AXIOM	MEANING
$\mu(\varnothing) = 0$	Empty set has zero size
$\mu(\bigsqcup_n A_n) = \Sigma_n \mu(A_n)$ $(A_i \cap A_j = \varnothing)$	Countable additivity (for disjoint) Size of union = sum of sizes

σ-algebra — family of "measurable" sets

PROPERTY	FORMULA
Closed under complements	$A \in \Sigma \Rightarrow X \setminus A \in \Sigma$
Closed under countable unions	$A_n \in \Sigma \Rightarrow \bigsqcup_n A_n \in \Sigma$
Contains the whole space	$X \in \Sigma$

Why precisely σ-algebra? (key question)

Why can't we restrict ourselves to finite unions (ordinary algebra)?

Problem: The limit of measurable sets must be measurable.

Let $A_n = [0, 1 - 1/n]$. Each A_n is an interval with length $(1 - 1/n)$.
Limit: $\bigsqcup_n A_n = [0, 1)$ — must also have a measure.

If the algebra is closed only under finite unions,
then $[0, 1) = \bigsqcup_{n=1}^\infty A_n$ may turn out to be non-measurable.

Solution: We require closure under countable unions.

STRUCTURE	CLOSED UNDER
Algebra of sets	Finite ∪, ∩, complements
σ-algebra	COUNTABLE ∪, ∩, complements

σ is the standard prefix for "countable" operations (σ-additivity, σ-algebra, σ-compactness — all admitting countable versions).

Deep reason:
Analysis works with limits, and a limit is a countable construction.
For integration and limits to be compatible, a σ-algebra is needed.
Convergence theorems (Levi, Lebesgue) require countable additivity.

Lebesgue measure on $\mathbb{R}^n$

SET	LEBESGUE MEASURE λ
Interval $(a,b) \subset \mathbb{R}$	$b - a$ (length)
Rectangle in $\mathbb{R}^2$	width × height (area)
Parallelepiped in $\mathbb{R}^3$	$a \cdot b \cdot c$ (volume)
Single point $\{x\}$	0
Countable set ($\mathbb{Q}$, $\mathbb{Z}$)	0 (measure zero)
Cantor set C	0, though C is uncountable.
Non-measurable sets	Exist (Vitali), but "unnatural"

Lebesgue integral

COMPARISON	RIEMANN vs LEBESGUE
Partition	Riemann: domain $[a,b]$ Lebesgue: range $\mathbb{R}$
Sum	Riemann: $\Sigma f(x_i) \cdot \Delta x_i$ Lebesgue: $\Sigma y_j \cdot \mu(\{x: f(x) \approx y_j\})$
Dirichlet fn	Riemann: not integrable Lebesgue: $\int f \, d\lambda = 1 \cdot 0 + 0 \cdot 1 = 0$

Construction of Lebesgue integral

STEP	DEFINITION
1. Simple function	$s = \Sigma c_k \cdot \chi\_\{A_k\}$ (step function)
2. Integral of simple	$\int s \, d\mu = \Sigma c_k \cdot \mu(A_k)$
3. Integral of $f \geq 0$	$\int f \, d\mu = \sup\{\int s \, d\mu: s \text{ simple}, s \leq f\}$
4. General integral	$\int f \, d\mu = \int f^+ \, d\mu - \int f^- \, d\mu$ where $f^+ = \max(f,0)$, $f^- = \max(-f,0)$

Convergence theorems – main advantage of Lebesgue

THEOREM	FORMULATION
Monotone convergence (B. Levi)	$f_n \uparrow f$ (monotonically) $\Rightarrow \int f_n \to \int f$
Dominated convergence (Lebesgue)	$\lvert f_n \rvert \le g$, $\int g < \infty$, $f_n \to f$ a.e. $\Rightarrow \int f_n \to \int f$ Also: $\int \lvert f_n - f \rvert \to 0$
Fatou	$f_n \ge 0 \Rightarrow \int(\liminf f_n) \le \liminf \int f_n$
"a.e." (almost everywhere)	= except for a set of measure zero

Why these theorems are important

SITUATION	CONSEQUENCE
Series of functions Σf_n	Can integrate term by term
Diff. under integral sign	$d/dt \int f(x,t)dx = \int \partial f/\partial t \, dx$ (under cond.)
Do not work for Riemann.	This is the main advantage of Lebesgue

Types of convergence – important distinction

In measure theory there are several types of convergence, and they are not equivalent.

TYPE OF CONVERGENCE	DEFINITION	WHAT IT MEANS
Pointwise	$\forall x$: $f_n(x) \to f(x)$	At each point separately
Almost everywhere (a.e.)	$f_n(x) \to f(x)$ for almost all x	Pointwise, except for a set of measure 0
Uniform	$\sup_x \lvert f_n(x) - f(x) \rvert \to 0$	Equally fast everywhere
In measure	$\mu(\{\lvert f_n - f \rvert > \varepsilon\}) \to 0$ for all $\varepsilon > 0$	Measure of "bad" set → 0
In L^p	$\int \lvert f_n - f \rvert^p \, d\mu \to 0$	Integral error → 0

Hierarchy (for finite measure):

```
  Uniform ⇒ Pointwise ⇒ Almost everywhere ⇒ In measure
      ↓
  In L∞        In Lᵖ ⇒ In measure (but NOT almost everywhere)
```

Counterexample: Convergence in measure does not imply convergence almost everywhere.
"Traveling hump": $f_n = \chi_{[k/2^m, (k+1)/2^m]}$ on $[0,1]$
Converges in measure to 0, but at each point a subsequence = 1.

L^p spaces

Definition: $L^p(X,\mu) = \{f: \int|f|^p d\mu < \infty\} / \{f=0 \text{ a.e.}\}$

p	NORM	PROPERTIES and APPLICATIONS
$p = 1$	$\|f\|_1 = \int\|f\|$	Integrable functions
$p = 2$	$\|f\|_2 = \sqrt{(\int\|f\|^2)}$ $\langle f,g\rangle = \int f\overline{g}$	Hilbert. Quantum mechanics, Fourier Scalar product
$p = \infty$	$\|f\|_\infty = \text{ess sup}\|f\|$	Essentially bounded functions

Key inequalities and theorems

RESULT	FORMULATION
Hölder's inequality	$\|fg\|_1 \le \|f\|_p \cdot \|g\|_q$ where $1/p+1/q=1$
Minkowski's inequality	$\|f+g\|_p \le \|f\|_p + \|g\|_p$
Riesz–Fischer theorem	L^p is complete (Banach) space

Applied example: heat flow through an inhomogeneous wall

Problem: A wall of thickness L with inhomogeneous structure. Thermal conductivity k(x) varies strongly across thickness (different materials, inclusions, cracks).

```
      T₁                                      T₂
-------------------------------------------------------------------------------
      |░░░|████|░░|████████|░░░░░░|██|░░░░|
      |░░░|████|░░|████████|░░░░░░|██|░░░░|
-------------------------------------------------------------------------------
      x=0                                    x=L
```

░░ = insulation ($k = 0.05$ W/(m·K))
██ = concrete ($k = 1.5$ W/(m·K))
Cracks – breaks in structure

Thermal resistance:

$$R = \int_0^L dx/k(x) \quad [\text{m}^2\cdot\text{K/W}]$$

Problem: $k(x)$ – discontinuous function with possible "pathologies".

Why the Lebesgue measure is needed

---

1. Discontinuities don't interfere:
   $k(x)$ can have a finite number of discontinuities (layer boundaries).
   The Lebesgue integral "doesn't notice" this – the set of discontinuity points has measure zero.

2. Cracks = set of measure zero:
   Microcracks are "thin" regions with $k \to \infty$ (air).
   If their total "thickness" = 0, they don't affect the integral.

3. Order of integration can be changed:
   When calculating 2D/3D heat transfer (integrals over area, volume)
   Fubini's theorem guarantees: $\iint = \int(\int)$.

Numerical example

Wall: 3 layers with total thickness 40 cm

- Plaster: 2 cm, $k = 0.8$ W/(m·K) → $R_1 = 0.02/0.8 = 0.025$
- Brick: 25 cm, $k = 0.7$ W/(m·K) → $R_2 = 0.25/0.7 = 0.357$
- Insulation: 10 cm, $k = 0.04$ W/(m·K) → $R_3 = 0.10/0.04 = 2.500$
- + 3 microcracks (total 0.5 mm) → $R = 0$ (measure zero)

$R\_total = \int_0^{0.40} dx/k(x) = R_1 + R_2 + R_3 = 2.88$ m²·K/W

Heat flux: $q = \Delta T/R = (20-(-10))/2.88 = 10.4$ W/m²

Moral: The Lebesgue measure allows integrating "bad" functions.
Discontinuities, jumps, singularities on sets of measure zero – not a problem.
For an engineer: one doesn't need to think about mathematical subtleties in calculations.
The integral exists for any "reasonable" physical quantity.

Measures — generalization of the concept of "size"

Idea of measure

Measure = way to assign "size" to sets

Length, area, volume, probability — all are examples of measures.

Measure μ: sets → numbers ≥ 0 (or $+\infty$)

Axioms of measure

1. $\mu(\varnothing) = 0$	Empty set has zero size
2. $\mu(A) \geq 0$	Size is non-negative
3. σ-additivity: $\mu(\sqcup_i A_i) = \Sigma_i \, \mu(A_i)$ (for countable number)	Size of union of disjoint = sum of sizes

Examples of measures

MEASURE	DESCRIPTION
Lebesgue measure on $\mathbb{R}^n$	Ordinary n-dimensional volume $\mu([a,b]) = b - a$ (length of interval)
Counting measure	$\mu(A) = \lvert A\rvert$ (number of elements) On $\mathbb{Z}$, on finite sets
Probability measure	$\mu(\Omega) = 1$, $\mu(A) = P(A)$ Measure of whole = 1
Dirac measure δ_x	$\delta_x(A) = 1$ if $x \in A$, else 0 "Concentrated at point x"
Haar measure	Invariant with respect to shifts On Lie groups (volume in group space)

Lebesgue integral

Ordinary Riemann integral: partition by x, sum $f(x_i)\cdot\Delta x$

Lebesgue integral: partition by y, sum $y\cdot\mu(\{x: f(x) \approx y\})$

Advantages:
- Can integrate more "bad" functions
- Theorems about passing to the limit under the integral sign
- Natural connection with probability theory

Formula: $\int f\, d\mu = \int_0^\infty \mu(\{x: f(x) > t\})\, dt$ (for $f \geq 0$)

Sets of measure zero

Set A has measure zero: $\mu(A) = 0$

Examples (in the sense of Lebesgue measure on $\mathbb{R}$):
- Any point $\{x\}$
- Any countable set ($\mathbb{Q}$, $\mathbb{Z}$, $\mathbb{N}$)
- Cantor set (uncountable, but of measure 0!)

Paradox: $\mathbb{Q}$ is dense in $\mathbb{R}$, but $\mu(\mathbb{Q}) = 0$, and $\mu(\mathbb{R}\setminus\mathbb{Q}) = \infty$

"Almost everywhere" = "except on a set of measure zero"

Measure and spaces

Measure is a way to measure size on a space.

Different spaces – different natural measures:
- On $\mathbb{R}^n$: Lebesgue measure (ordinary volume)
- On a Lie group: Haar measure (invariant under shifts)
- On a manifold: measure induced by metric

Integral is "summation over space":
$\int_M f \, d\mu$ = "average value of f with respect to measure"

Probability space = space + measure with $\mu(\Omega) = 1$

Differential Equations – How to Formulate

Main Principle

DEs are not written out of thin air. They are based on conservation laws.

(And conservation laws, by Noether's theorem, follow from symmetries)

Formulation algorithm:

1. Object → What are we studying? (body, fluid, field)
2. Process → What is happening? (flows, heats up, moves)
3. Elem. volume → Isolate dx, dy, dz, dt
4. Conservation → What doesn't change? (mass, energy, momentum)
5. Balance → Inflow − Outflow = Change → DE.

Examples: conservation law → equation

CONSERVATION LAW	EQUATION
Mass conservation	$\partial\rho/\partial t + \nabla\cdot(\rho v) = 0$ (continuity)
Energy conservation	$\partial T/\partial t = \alpha\nabla^2 T$ (heat conduction)
Momentum conservation	Navier–Stokes equations
Charge conservation	$\partial\rho/\partial t + \nabla\cdot j = 0$
Newton's second law	$m(d^2x/dt^2) = F$

Classification of partial differential equations

TYPE	EXAMPLE and PHYSICS
Elliptic	$\nabla^2 u = f$ (Poisson equation) Stationary problems, electrostatics
Parabolic	$\partial u/\partial t = \alpha\nabla^2 u$ (heat conduction) Diffusion, smoothing
Hyperbolic	$\partial^2 u/\partial t^2 = c^2\nabla^2 u$ (wave) Wave propagation

Practical fact

Practically any DE (Navier–Stokes, Maxwell, heat conduction) can be numerically solved in Excel.

- Each cell = value at one grid point
- Cell formula = references to neighbors (finite differences)
- Boundary conditions = constants at edges
- Iterations → solution

This is the finite difference method in its simplest form.

Conclusion

Central idea of the document

Mathematics is the language of spaces.

Space = set + structure (way to connect points).

- Topology studies the shape of space (holes, connectivity)
- Algebra studies symmetries of space (transformation groups)
- Analysis studies functions on space (change, extrema)
- Geometry studies measurements in space (distances, curvature)

Functors translate between areas, preserving structure.

Key facts by sections — minimum you need to know

Below are not questions, but answers. If something is unclear, return to the section.

Logic

"From false follows anything" ($F \to P = T$ for any P):
Implication $P \to Q$ is false only when P is true and Q is false.
If P is false, then $P \to Q$ is true for any Q.
Analogy: "If I'm a millionaire, I'll buy you an island" — not lying,
because I'm not a millionaire.

Sets

Bijection $\mathbb{N} \leftrightarrow \mathbb{Z}$: $f(n) = n/2$ if n is even, $-(n+1)/2$ if odd
$0 \mapsto 0,\ 1 \mapsto -1,\ 2 \mapsto 1,\ 3 \mapsto -2,\ 4 \mapsto 2, \ldots$

Why $|\mathbb{N}| < |\mathbb{R}|$: Cantor's diagonal argument.
Suppose $f: \mathbb{N} \to [0,1]$ is a bijection. Construct a number x that
differs from f(n) in the n-th digit: $x \notin \mathrm{Im}(f)$. Contradiction.

Groups

Why $(ab)^{-1} = b^{-1}a^{-1}$:
Check: $(ab)(b^{-1}a^{-1}) = a(bb^{-1})a^{-1} = aea^{-1} = aa^{-1} = e$ ✓
Example: put on socks, then shoes → take off shoes, then socks.

Kernel of homomorphism $\varphi: G \to H$:
$\ker \varphi = \{g \in G : \varphi(g) = e_h\}$ — everything that maps to neutral.
Example: $\varphi: \mathbb{Z} \to \mathbb{Z}/6$, $k \mapsto k \bmod 6$. Then $\ker \varphi = 6\mathbb{Z} = \{.,-6,0,6,12,.\}$

Topology

Set that is neither open nor closed:
$[0, 1)$ in $\mathbb{R}$. Contains boundary point 0, but doesn't contain 1.

Why mug ≅ donut:
Both surfaces have one hole (genus 1). Can be continuously deformed one into the other without tearing or gluing.

$\pi_1(S^1) = \mathbb{Z}$: Loops on a circle are classified by winding number.
$\pi_1(S^2) = 0$: Any loop on a sphere contracts to a point.

Linear Algebra

Why $\det(AB) = \det(A)\det(B)$, but $\det(A+B) \neq \det(A)+\det(B)$:
Determinant is multiplicative, but NOT additive.
Counterexample: $A = B = I$ (identity). $\det(I) = 1$.
$\det(I+I) = \det(2I) = 2^n \neq 1+1 = 2$ (for $n > 1$).

Vector that is not an "arrow":
Function $f(x) = x^2$ is a vector in space $C[0,1]$.
$(f+g)(x) = f(x)+g(x)$, $(\alpha f)(x) = \alpha f(x)$ – axioms are satisfied.

Manifolds

Why multiple charts for sphere:
S^2 is compact, $\mathbb{R}^2$ is not. Homeomorphism preserves compactness.
Therefore, $S^2 \not\cong \mathbb{R}^2$. Need at least 2 charts (stereographic projections).

What lives in T_pM:
Tangent vectors = velocities of curves through p = directions of motion.
$T_pM \cong \mathbb{R}^n$, but these are different spaces for different p.
$v \in T_pM$ and $w \in T\gamma M$ cannot be added directly – need a connection.

Differential Forms

Why $d^2 = 0$ is related to $\partial^2 = 0$:
These are dual statements. ∂ acts on chains (regions),
d acts on forms (integrands). Stokes theorem: $\int_{\partial M} \omega = \int_M d\omega$.

$\partial^2 M = \varnothing$ (boundary of boundary is empty) ↔ $d^2\omega = 0$ (exterior derivative twice gives zero). This is the same fact from two sides.

Stokes unifies classics:
- dim=1: $\int_a^b f'dx = f(b)-f(a)$ (Newton–Leibniz)
- dim=2: $\iint \operatorname{rot} F\cdot dA = \oint F\cdot dr$ (Green/Stokes)
- dim=3: $\iiint \operatorname{div} F\, dV = \oiint F\cdot dS$ (Gauss–Ostrogradsky)

Categories

Group as category with one object:

- One object: ●
- Morphisms: group elements $g \in G$ (arrows ● → ●)
- Composition: group operation $g \circ h = gh$
- Identity morphism: neutral element e
- All morphisms are invertible (that's what "group" means.)

Example of functor: $\pi_1: \mathrm{Top}^* \to \mathrm{Grp}$

- Objects: spaces with marked point → groups
- Morphisms: continuous maps → group homomorphisms
- Composition is preserved: $\pi_1(f \circ g) = \pi_1(f) \circ \pi_1(g)$

> Main criterion of understanding:
>
> You understand mathematics when you see one pattern in different sections:
>
> $\ker \varphi$ (groups) = $\ker T$ (lin.alg.) = $\ker d$ (forms) = "what collapses"
> $G/\ker \varphi \cong \operatorname{Im} \varphi$ — everywhere the same isomorphism theorem.

We encountered complex numbers in Part II as an algebraic construction.
But analysis on $\mathbb{C}$ — derivatives, integrals, series is the territory of Part III.
And here something surprising is discovered.

Complex Analysis — The Magic of Holomorphic Functions

Complex numbers are algebra. But when we start differentiating and integrating functions $f: \mathbb{C} \to \mathbb{C}$, a miracle happens.

Holomorphy = complex differentiability

A function $f: \mathbb{C} \to \mathbb{C}$ is holomorphic at point z_0 if the limit exists:

$$f'(z_0) = \lim_{h\to 0} \frac{f(z_0 + h) - f(z_0)}{h} \qquad (h \in \mathbb{C},\ h \to 0 \text{ from any direction})$$

This is much stronger than differentiability in $\mathbb{R}^2$.

Cauchy–Riemann conditions:

If $f(x + iy) = u(x,y) + iv(x,y)$, then:

$$\partial u/\partial x = \partial v/\partial y \quad \text{and} \quad \partial u/\partial y = -\partial v/\partial x$$

These two equations connect the Re and Im parts of a holomorphic function.

Why holomorphy is magic

If f is holomorphic, then automatically:

- f is infinitely differentiable (all derivatives exist)
- f is analytic (expands in a power series)
- f is determined by its values on any curve
- Integral over a closed contour = 0 (if there are no singularities inside)

In $\mathbb{R}$: differentiability of f at a point does not guarantee existence of f'' (or even continuity of f'). In $\mathbb{C}$: one derivative → infinitely many.

Why? The Cauchy–Riemann conditions are a system of equations that "propagates" information about the function in all directions.

Visualization: what complex functions "look like"

A complex function $f: \mathbb{C} \to \mathbb{C}$ is a mapping from plane to plane.
This is a 4D object (2D input + 2D output), difficult to visualize directly.

Method: grid deformation

Draw a grid in the z-plane, see where it maps to in the w-plane.

```
z-plane                  w = z²              w-plane

| | | | |                                   \ | /
-+-+-+-+-+-        ------------▶             -+-
| | | | |                                   / | \
-+-+-+-+-+-
| | | | |              Squares → curvilinear quadrilaterals
                       Right angles → right angles (conformality)
```

Examples of mappings

FUNCTION	WHAT WE SEE IN THE PICTURE
$f(z) = z$	Identity — grid doesn't change
$f(z) = z^2$	Doubling of angles: ray θ → ray 2θ Circle $\lvert z\rvert=r$ → circle $\lvert w\rvert=r^2$
$f(z) = e^z$	Vertical x=const → circle $\lvert w\rvert=e^x$ Horizontal y=const → ray $\arg(w)=y$ Strip $0 < \mathrm{Im}(z) < 2\pi$ → entire plane (.)
$f(z) = 1/z$	Inversion: large ↔ small $\lvert z\rvert>1$ → $\lvert w\rvert<1$, angles change sign

Conformality — the main property of holomorphic functions

A holomorphic function $f(z)$ preserves angles at points where $f'(z) \neq 0$.

What this means:

- Two curves intersect at angle α
- Their images intersect at the same angle α
- Infinitesimal circles → infinitesimal circles (not ellipses)

Locally a holomorphic function is a rotation + scaling:

$$f(z) \approx f(z_0) + f'(z_0)\cdot(z - z_0)$$

- The factor $f'(z_0) = |f'|\cdot e^{i\theta}$
- $|f'|$ — scaling coefficient
- $\theta = \arg(f')$ — rotation angle

Visually: Small squares of the grid → small squares,
but rotated and scaled (not skewed)

Exception: At points where $f'(z) = 0$ (critical points)
angles can be multiplied. For $f(z) = z^n$ at zero: angle × n.

Cauchy's integral theorem

If f is holomorphic inside and on a closed contour γ:

$$\oint_\gamma f(z)\, dz = 0$$

Corollary (Cauchy's integral formula):

```
             ∮   f(ζ)
f(z) = ----- ·  ------- dζ
       2πi   γ  ζ − z
```

The value of the function inside a contour is determined by values on the contour.
This is as if the temperature inside a room were determined only by the walls.

Residues — technique for computing integrals

If f has an isolated singularity at point z_0:

$$f(z) = \dots + a_{-2}/(z-z_0)^2 + \underset{\substack{\uparrow \\ \text{Residue} = a_{-1}}}{a_{-1}/(z-z_0)} + a_0 + a_1(z-z_0) + \dots$$

Residue theorem:

$$\oint_\gamma f(z)\,dz = 2\pi i \cdot \sum_{z_k \text{ inside } \gamma} \operatorname{Res}(f, z_k)$$

Application: Computation of real integrals.

$$\int_{-\infty}^{\infty} dx/(1+x^2) = \pi \quad \text{(closing contour in upper half-plane)}$$

Many integrals impossible in $\mathbb{R}$ are trivial in $\mathbb{C}$.

Connection with topology

Winding number (winding number):
How many times does curve γ wind around point z_0?

$$n(\gamma, z_0) = \frac{\oint}{2\pi i} \cdot_{\gamma} \frac{dz}{z - z_0} = \text{integer.}$$

This connects complex analysis with algebraic topology (π_1).

Example: Proof of Fundamental theorem of algebra (every polynomial over $\mathbb{C}$ has a root) uses winding number.

Complex potential — hydro- and aerodynamics

One of the most beautiful applications of complex analysis — potential flow of ideal fluid (or gas at low velocities).

Why this is needed

Problem: Find how fluid flows around obstacle.
Issue: Navier–Stokes equations are complex, no analytical solutions.
Simplification: If fluid is "ideal" (without viscosity) and flow without vortices, problem reduces to one function of complex variable.

Step 1: What is potential flow

"Irrotational" means: fluid does not rotate locally.
Imagine matchstick-boat on water:

- In ordinary flow: matchstick floats and spins around its axis
- In irrotational: matchstick floats but does NOT spin (always points north)

Mathematically: rot $v = 0$, which means $v = \nabla\varphi$ for some function φ.
Φ is called velocity potential.

For incompressible fluid (div v = 0):
  $\nabla^2\varphi = 0$ ← Laplace equation.

Step 2: Why complex numbers help

Key fact: If $w(z) = \varphi + i\psi$ is holomorphic, then:
  $\partial\varphi/\partial x = \partial\psi/\partial y$ and $\partial\varphi/\partial y = -\partial\psi/\partial x$ (Cauchy–Riemann conditions)

Consequence: $\partial^2\varphi/\partial x^2 + \partial^2\varphi/\partial y^2 = 0$ ← this is Laplace equation.

Conclusion: Any holomorphic function w(z) automatically gives solution of flow equation. No need to solve differential equations.

Step 3: what φ and ψ mean physically

Φ – velocity potential:
  Velocity = gradient of potential: $v_x = \partial\varphi/\partial x$, $v\gamma = \partial\varphi/\partial y$
  Lines φ = const – equipotentials (like elevation lines on map)

Ψ – stream function:
  Lines ψ = const are streamlines (trajectories of fluid particles)
  Fluid flows along these lines, not crossing them.

Why they are orthogonal:
  Gradient of φ (direction of steepest increase of φ) is perpendicular to lines φ = const. But gradient φ = velocity vector. And velocity is directed along streamlines ψ = const. Hence, lines ⊥ to each other.

```
     φ = 3   φ = 2   φ = 1   φ = 0   ← equipotentials (vertical)
        |       |       |       |
 ψ=1 ---+-------+-------+-------+--→  ← streamlines (horizontal)
        |       |       |       |
 ψ=0 ---+-------+-------+-------+--→
        |       |       |       |
 ψ=−1---+-------+-------+-------+--→
        |       |       |       |
```

  Fluid flows to the right (along ψ = const), crossing equipotentials.

Examples of complex potentials – with calculations

Example 1: uniform flow ($w = Uz$)

$w(z) = Uz = U(x + iy) = Ux + iUy$

Hence: $\varphi = Ux$, $\psi = Uy$

Velocity: $v_x = \partial\varphi/\partial x = U$, $v_y = \partial\varphi/\partial y = 0$

Interpretation: Uniform flow with velocity U to the right.

Specifically: $U = 5$ m/s, point $z = 2 + 3i$ ($x=2$, $y=3$)
$\varphi = 5\times2 = 10$, $\psi = 5\times3 = 15$
Velocity = (5, 0) m/s everywhere the same

Example 2: source at origin ($w = (Q/2\pi) \ln z$)

$z = re^{i\theta}$, $\ln z = \ln r + i\theta$
$w = (Q/2\pi)(\ln r + i\theta)$

Hence: $\varphi = (Q/2\pi) \ln r$, $\psi = (Q/2\pi)\,\theta$

Velocity (in polar coordinates):
$v_r = \partial\varphi/\partial r = Q/(2\pi r)$, $v_\theta = 0$

Interpretation: Fluid flows radially outward from point.
The farther from source – the slower (velocity ∝ 1/r).

Specifically: $Q = 10$ m²/s (discharge per unit depth), $r = 2$ m
Velocity = $10/(2\pi\times2) = 0.8$ m/s (radially outward)

```
     ↖  ↑  ↗
   ←    ●    →    Fluid flows out
     ↙  ↓  ↘     radially from center
```

Example 3: flow around a cylinder ($w = U(z + a^2/z)$)

This is a sum of uniform flow and a dipole.

Let's verify that $|z| = a$ is a streamline ($\psi = 0$):

On the circle $z = ae^{i\theta}$:
$w = U(ae^{i\theta} + ae^{-i\theta}) = U \cdot 2a \cos\theta = 2Ua \cos\theta$ (purely real)

Therefore $\psi = \mathrm{Im}(w) = 0$ on the entire circle.
The circle $|z| = a$ is a streamline. The fluid does not cross it.

This is the surface of the cylinder.

```
                    →→→→→→→→→→→→
                  →   ╭-----╮    →
                →   /         \    →
              →   |     ●     |    →    Flow around cylinder
                →   \         /    →
                  →   ╰-----╯    →
                    →→→→→→→→→→→→
```

Specifically: $U = 10$ m/s, $a = 0.5$ m, point $z = 1$ (on x-axis, outside cylinder)
$w = 10(1 + 0.25/1) = 10 \times 1.25 = 12.5$
$dw/dz = U(1 - a^2/z^2) = 10(1 - 0.25) = 7.5$ m/s (local velocity)

Superposition: constructing complex flows

Key idea: The Laplace equation is linear.

If w_1 and w_2 are solutions, then $w_1 + w_2$ is also a solution.
Complex flows = sums of simple ones.

Example: Source + sink + uniform flow

$$w = \underset{\overset{\uparrow}{\text{flow}}}{Uz} + \underset{\overset{\uparrow}{\text{source at a}}}{(Q/2\pi) \ln(z-a)} - \underset{\overset{\uparrow}{\text{sink at }-a}}{(Q/2\pi) \ln(z+a)}$$

This gives flow around an oval body.

```
→→→   ↗ source       sink ↘   →→→
→→→ →  ●---------------------●  → →→→
→→→   ↘                     ↗   →→→
```

Joukowski transformation — how wings are calculated

Problem: We know flow around a cylinder. But a wing is not a cylinder.

Joukowski's idea (1910): Conformal mapping transforms a circle into a wing.

Why this works

A holomorphic function $\zeta = f(z)$ has the property of conformality:

- Preserves angles between curves
- Streamlines remain streamlines.
- If $w(z)$ is a potential for a cylinder, then $w(f^{-1}(\zeta))$ is a potential for what the cylinder has been transformed into

Joukowski transformation:

$$\zeta = z + a^2/z$$

```
 Step 1: Take a circle |z − z₀| = r (with offset center)

 Step 2: Apply the transformation — obtain a wing profile.

        z-plane                        ζ-plane

           ╭---╮
          /  ●   \
         | center |        ζ = z + a²/z
          \     /         -------------→        ╭----------╮
           ╰---╯                               /            \
          circle with                     ----●--------------\---
          offset                               \            /
          center                                ╰----------╯
                                              wing profile
                                              (sharp trailing edge)

-------------------------------------------------------------------------------
 Joukowski's lift theorem
-------------------------------------------------------------------------------

   L = ρ U Γ   (lift force per unit wingspan)

   where: ρ — air density
          U — freestream velocity
          Γ — circulation around the profile

 Specific calculation:

   Data: ρ = 1.2 kg/m³, U = 50 m/s, Γ = 20 m²/s
   L = 1.2 × 50 × 20 = 1200 N/m

   For a wing of length 10 m: L_total = 12000 N = 1.2 tons of lift.

 Where circulation comes from:
   Kutta–Joukowski condition: flow leaves the sharp trailing edge smoothly.
   This fixes the magnitude of circulation — it is not arbitrary.

 Historical significance:
   This is not an abstraction — this is how first airplanes were actually designed.
   Complex analysis literally lifted humanity into the air.
```

We have considered many structures: groups, topologies, vector spaces, manifolds, complex numbers. Everywhere there are "objects" and "mappings between them" (homomorphisms, continuous functions, linear operators).

Categories are a language for describing this general pattern. They allow seeing analogies between different areas of mathematics as precise statements.

Categories and functors — mathematics about mathematics

> "Understanding consists in reducing one type of reality to another."
> — Claude Lévi-Strauss
>
> Category theory does this literally: functors translate structures of one type into structures of another.

Summary Tables

T.1 great correspondence table

	SET sets	GRP groups	VECT vect.sps	TOP topology	MAN manifolds
OBJECTS	sets	groups	sps over F	spaces	smooth

					manifolds
morphism	function	homomorphism	linear mapping	continuous mapping	smooth mapping
ISOMORPHISM	bijection	isomorphism of groups	isomorphism of sps	homeomorphism	diffeomorphism
INVARIANT	$\lvert A \rvert$ cardinality	order, table	dim V	π_1, χ, H_n	dim, curvature
SUBOBJECT	subset $A \subseteq B$	subgroup $H \leq G$	subspace $W \subseteq V$	subspace (open/clsd)	submanifold
QUOTIENT	$A/\sim$	G/H ($H \triangleleft G$)	V/W	$X/\sim$	M/G (orbits)
PRODUCT	$A \times B$	$G \times H$	$V \oplus W$	$X \times Y$	$M \times N$
COPRODUCT	$A \sqcup B$	$G * H$ (free)	$V \oplus W$ (same)	$X \sqcup Y$	$M \sqcup N$
KERNEL	—	$\ker \varphi \triangleleft G$	$\ker T \subseteq V$	—	—
IMAGE	$f(A)$	$\mathrm{Im}\ \varphi \leq H$	$\mathrm{Im}\ T \subseteq W$	$f(X)$	$f(M)$
INITIAL	$\varnothing$	$\{e\}$	$\{0\}$	$\varnothing$	$\varnothing$
TERMINAL	$\{*\}$	$\{e\}$	$\{0\}$	$\{*\}$	$\{*\}$

T.2 universal patterns

PATTERN	MANIFESTATIONS
$d^2 = 0$	$d \circ d = 0$ (forms), $\partial \circ \partial = 0$ (boundaries), rot∘grad=0, div∘rot=0
dim = dim ker + dim Im	Linear algebra, groups (Lagrange's theorem), diff. forms (cohomology)
Duality	$V \leftrightarrow V^*$, $G \leftrightarrow \hat{G}$, $q \leftrightarrow p$, points ↔ hyperplanes
Local → Global	Stokes' thm, de Rham thm, bundles
Symmetry → Conservation	Noether's theorem: every continuous symmetry yields a conservation law
Classification	Finite groups, surfaces, simple algebras

T.3 basic principles of the standard

NUMBER	PRINCIPLE
1	Mathematics – discovery of structure inherent in the act of distinction
2	From ∅ through categorization all mathematics arises
3	Logic – superstructure over set theory for communication
4	Proof – path through the graph of set embeddings
5	Understanding = ability to visualize
6	One pattern manifests in all branches

T.4 summary correspondence table between branches

CONCEPT	REALIZATIONS IN BRANCHES
Isomorphism	Set: bijection isomorphism of groups isomorphism of spaces homeomorphism equivalence of categories
Invariant (what is preserved)	Set: cardinality $\|A\|$ multiplication table dimension dim V π_1, χ, H_n
Kernel of morphism	$\ker(\varphi) \triangleleft G$ $\ker(T) \subseteq V$ $\ker(d)$ = closed forms
Duality	$G \leftrightarrow \hat{G}$ $V \leftrightarrow V^*$ $\Omega^k \leftrightarrow \Omega^{n-k}$ $q \leftrightarrow p$ (symplectic)
Property $d^2=0$	Vector analysis: $\mathrm{rot} \circ \mathrm{grad}=0$, $\mathrm{div} \circ \mathrm{rot}=0$ $d \circ d=0$ $\partial \circ \partial=0$ (homology)

PART IV: APPLIED MATHEMATICS

Probability — Mathematics of Uncertainty

Remember the section on Lebesgue measure? We said that the Riemann integral is bad for limits — a sequence of integrable functions can converge to a non-integrable one. Lebesgue measure solved this problem by generalizing the notion of "lengt

Probability uses the same upgrade. Instead of "length on a line" we measure "probability of an event" — and this is also a measure, obeying the same axioms. All of probability theory is measure theory on a space of outcomes.

$E[X] = \int X \, dP$ — this is not an analogy with an integral, this is literally a Lebesgue integral with respect to probability measure P. Everything we know about integrals (monotone convergence theorem, Fubini, etc.), works here too.

Probability as a view of space

Probability theory is measure theory on special spaces.

PROBABILISTIC TERM	GEOMETRIC MEANING
Sample space Ω	SET of all possible states of the world Each point $\omega \in \Omega$ — one "scenario" (what came up, what weather, what path)
Event $A \subseteq \Omega$	SUBSET of sample space ("in which scenarios did A occur")
Probability $P(A)$	MEASURE of set A (generalized "volume" of region)
Random variable X	MAPPING $X: \Omega \to \mathbb{R}$ (function on sample space)
Expected value $E[X]$	INTEGRAL with respect to measure P $E[X] = \int_\Omega X(\omega) \, dP(\omega)$
Independent experiments	PRODUCT of spaces $\Omega = \Omega_1 \times \Omega_2$, $P = P_1 \times P_2$

Examples:

- Coin toss: Ω = {heads, tails} — two worlds
- Random walk: Ω = all possible trajectories
- Thermal fluctuations: Ω = infinite-dimensional (temperature functions)

Conclusion: Probability is NOT "chances" in everyday sense.
This is a rigorous mathematical structure: space + measure.

Concrete example: roll of two dice

Sample space:

$\Omega = \{(1,1), (1,2), (1,3), ..., (6,6)\} = \{1,...,6\} \times \{1,...,6\}$
$|\Omega| = 36$ points

This is a finite discrete space — a 6×6 "lattice":

```
    j
  6 | ● ● ● ● ● ●
  5 | ● ● ● ● ● ●
  4 | ● ● ● ● ● ●      Each point ● — one outcome
  3 | ● ● ● ● ● ●      Measure: P({(i,j)}) = 1/36 for all
  2 | ● ● ● ● ● ●
  1 | ● ● ● ● ● ●
    +------------- i
      1 2 3 4 5 6
```

Event "sum = 7":

$A = \{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)\}$

This is a subset (diagonal in our "lattice").
$P(A) = |A|/|\Omega| = 6/36 = 1/6$

Random variable X = "sum of points":

$X: \Omega \to \mathbb{R}, \quad X((i,j)) = i + j$

$E[X] = \Sigma\, X(\omega)\cdot P(\{\omega\}) = \Sigma_{ij}\, (i+j)/36 = 7$

(Sum of all values with their probabilities = integral with respect to measure P)

Connection with other sections of the atlas

Set theory (1.1):
Ω — set, events — subsets, σ-algebra — system of subsets
Operations: $A \cup B$ (or), $A \cap B$ (and), A^c (not)

Measure:
P is a measure with property $P(\Omega) = 1$ (normalization)
All properties of measure work: countable additivity, continuity

Functional analysis:
Random variables with $E[X^2] < \infty$ form a Hilbert space $L^2(\Omega,P)$
Covariance $\mathrm{Cov}(X,Y) = \langle X - E[X], Y - E[Y]\rangle$ — inner product.
Uncorrelatedness ⟺ orthogonality

Linear algebra:
  Covariance matrix Σ – positive semidefinite
  Principal components (PCA) = eigenvectors of Σ
  Mahalanobis metric $d^2 = (x-\mu)^T\Sigma^{-1}(x-\mu)$ – distance accounting for Σ

Probability = measure

Probability theory is measure theory with condition $\mu(\Omega) = 1$

PROBABILITY SPACE	(Ω, Σ, P)
Ω	Sample space
Σ	σ-algebra of events
P	Probability measure: $P(\Omega) = 1$

Dictionary: probability ↔ measure theory

PROBABILITY	MEASURE THEORY
Event A	Measurable set $A \in \Sigma$
Probability $P(A)$	Measure $\mu(A)$, normalized $\mu(\Omega)=1$
Random variable X	Measurable function $X: \Omega \to \mathbb{R}$
Probability density $f(x)$	Radon–Nikodym derivative $dP/d\lambda$ (ratio of measure P to Lebesgue measure λ)
Expected value $E[X]$	Lebesgue integral $\int X \, dP$
Independence A, B	$P(A \cap B) = P(A) \cdot P(B)$
Conditional probability $P(A\|B)$	$P(A \cap B)/P(B)$

Kolmogorov axioms (1933)

AXIOM	FORMULATION
Non-negativity	$P(A) \geq 0$ for all $A \in \Sigma$
Normalization	$P(\Omega) = 1$
Countable additivity	$P(\bigcup_n A_n) = \Sigma_n P(A_n)$ for disjoint A_n

Consequences of axioms

PROPERTY	FORMULA
Impossible event	$P(\varnothing) = 0$
Complement	$P(A^c) = 1 - P(A)$
Monotonicity	$A \subseteq B \Rightarrow P(A) \leq P(B)$
Inclusion-exclusion form.	$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Random variables and distributions

Random variable $X: \Omega \to \mathbb{R}$ — measurable function

DISTRIBUTION FUNCTION	$F(x) = P(X \leq x)$
Properties of $F(x)$	Non-decreasing, $F(-\infty)=0$, $F(+\infty)=1$, right-cont.
Discrete r.v.	Countable set of values: $P(X=x_k)=p_k$, $\Sigma p_k=1$
Continuous r.v.	$F(x)=\int_{-\infty}^{x} f(t)dt$, f — density, $\int f=1$

How a histogram becomes a probability density

This is key intuition for understanding continuous distributions.

Step 1: ordinary histogram

There are N measurements. We divide the axis into bins (intervals) of width Δx. In each bin we count the number of hits n_k.

Bar height = n_k (absolute frequency)

Problem: if we take a different bin width, the histogram will change.

Step 2: normalize by bin width

Bar height = $n_k / (N \cdot \Delta x)$

Now the height is the frequency density: "how many hits per unit x"

Dimension: [1/x], for example 1/meter, 1/second, 1/degree

Important: Bar area = $(n_k/N \cdot \Delta x) \cdot \Delta x = n_k/N$ = relative frequency
Sum of areas of all bars = 1

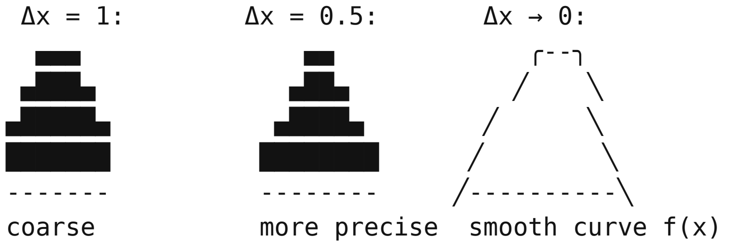
Step 3: let $\Delta x \to 0$ and $N \to \infty$

As $\Delta x \to 0$ the stepwise histogram transforms into a smooth curve.
This curve is the probability density function f(x).

$$f(x) = \lim_{\substack{\Delta x\to 0\\ N\to\infty}} \frac{n_k}{N \cdot \Delta x}$$

Properties of f(x):
- $f(x) \geq 0$ (frequency cannot be negative)
- $\int f(x)dx = 1$ (sum of all probabilities = 1)
- $P(a \leq X \leq b) = \int_a^b f(x)dx$ (area under curve = probability)

Visualization



Practical meaning

For temperature in a pipeline over a year (millions of measurements):

- Histogram with bins of 1°C gives a rough picture
- Histogram with bins of 0.1°C – more detailed
- Density f(T) – ideal limit with infinite data

f(25°C) = 0.15 [1/°C] means:
"In a small interval around 25°C the probability ≈ 0.15 · ΔT"

Important: f(x) is NOT probability (can be > 1)!
Probability is the area: $P = f(x) \cdot \Delta x$

Most important distributions – discrete

DISTRIBUTION	P(X=k)	APPLICATION
Bernoulli(p)	$P(1)=p$, $P(0)=1-p$	One yes/no trial
Binomial Bin(n,p)	$C(n,k)p^{k}(1-p)^{n-k}$	k successes in n trials
Poisson Pois(λ)	$e^{-}\lambda\cdot\lambda^{k}/k!$	Rare events
Geometric Geom(p)	$(1-p)^{k-1}\cdot p$	Number of first success

Most important distributions – continuous

DISTRIBUTION	DENSITY f(x)	APPLICATION
Uniform U(a,b)	$1/(b-a)$ on $[a,b]$	Random point
Exponential Exp(λ)	$\lambda e^{-}\lambda^{x}$, $x\geq 0$	Time until event
Normal N(μ,σ²)	$\exp(-(x-\mu)^{2}/2\sigma^{2})$ $/(\sigma\sqrt{2\pi})$	Errors, growth, finance EVERYWHERE.
Standard N(0,1)	$e^{-x^{2}/2}/\sqrt{2\pi}$	$Z = (X-\mu)/\sigma$

Numerical characteristics

CHARACTERISTIC	DEFINITION
Expected value E[X] (μ)	$\int X\,dP = \Sigma x_{k}p_{k}$ (discr.) or $\int xf(x)dx$ (cont.)
Variance Var(X) = σ²	$E[(X-\mu)^{2}] = E[X^{2}] - (E[X])^{2}$
Std. deviation σ	$\sqrt{Var(X)}$, in the same units as X
Covariance Cov(X,Y)	$E[(X-\mu_{x})(Y-\mu\gamma)] = E[XY] - E[X]E[Y]$
Correlation ρ(X,Y)	$Cov(X,Y)/(\sigma_{x}\sigma\gamma) \in [-1, 1]$
Moments E[Xⁿ]	n-th moment about zero

Properties of characteristics

PROPERTY	FORMULA
Linearity of E	$E[aX+b] = aE[X]+b$
Additivity of E (always)	$E[X+Y] = E[X]+E[Y]$
Multiplicativity (independent)	$E[XY] = E[X]E[Y]$
Scaling of Var	$Var(aX+b) = a^2Var(X)$
Additivity of Var (independent)	$Var(X+Y) = Var(X)+Var(Y)$
General formula	$Var(X+Y) = Var(X)+Var(Y)+2Cov(X,Y)$

Law of Large Numbers and CLT

Condition: $X_1,X_2,...$ – i.i.d. (indep. identically distr.) with $E[X_i]=\mu$, $Var=\sigma^2$
Sample mean: $\bar{X}_n = (X_1+...+X_n)/n$

THEOREM	FORMULATION and MEANING
Law of large numbers (LLN)	$\bar{X}_n \to \mu$ as $n\to\infty$ "Sample mean → true mean" Example: proportion of heads → 0.5
Central limit theorem (CLT)	$\sqrt{n}(\bar{X}_n-\mu)/\sigma \to N(0,1)$ in distribution Corollary: $\bar{X}_n \approx N(\mu, \sigma^2/n)$ "Sum of independent ≈ normal" Explains why $N(\mu,\sigma^2)$ is EVERYWHERE

Important limitation of CLT

CLT requires finite variance ($\sigma^2 < \infty$)!

Distributions with "heavy tails" DO NOT converge to normal:
- Cauchy: $E[X]$ does not exist, $Var = \infty$
- Pareto with $\alpha \le 2$: $Var = \infty$
- Stable Lévy distributions

Application: In finance and reliability analysis heavy tails are critical.
Normal distribution underestimates the probability of extreme events.

Sum of Cauchy: mean of n Cauchy numbers is distributed as one Cauchy number.
(averaging does NOT reduce scatter)

Conditional probability and Bayes

CONCEPT	FORMULA / MEANING
Conditional probability	$P(A \mid B) = P(A \cap B)/P(B)$ "Probability of A given B"
Bayes' theorem	$P(A \mid B) = P(B \mid A) \cdot P(A) / P(B)$
Via total probability	$P(A \mid B) = P(B \mid A)P(A) / [P(B \mid A)P(A)+P(B \mid A^c)P(A^c)]$

Bayes terminology

TERM	MEANING
$P(A)$ – prior	Probability BEFORE observing data
$P(A \mid B)$ – posterior	Probability AFTER observing B
$P(B \mid A)$ – likelihood	How likely is B under hypothesis A

Example: medical test (false positive paradox)

DATA	VALUE
$P(\text{sick})$	0.01 (1% of population is sick)
$P(+ \mid \text{sick})$	0.99 (sensitivity 99%)
$P(+ \mid \text{healthy})$	0.05 (5% false positives)
$P(\text{sick} \mid +) = ?$	$(0.99 \cdot 0.01)/(0.99 \cdot 0.01+0.05 \cdot 0.99) \approx 0.17$
Conclusion	With + test probability of disease is only ~17%.

Connection of probability with other branches

BRANCH	CONNECTION WITH PROBABILITY
Lebesgue measure	Probability = measure with $\mu(\Omega)=1$, $E[X]=\int X\, dP$
Functional analysis	$L^2(\Omega,P)$ – space of r.v. with finite Var
Linear algebra	Covariance matrix, PCA Correlation = cos of angle in L^2
Fourier series	Characteristic function $E[e^{itx}]$ = Fourier
Physics	Stat. mechanics, quantum theory, entropy
ML / Statistics	Bayesian inference, regression, testing
B1 Programming	PROBABILISTIC PROGRAMMING: Variables = distributions, code = model Languages: Stan, PyMC, Edward, Pyro

Characteristic function = Fourier transform of density

Definition:

$\varphi_X(t) = E[e^{itx}] = \int f(x) e^{itx} dx$

This is exactly the Fourier transform of density f(x).

Why this is important:

Convolution of densities ↔ Multiplication of characteristic functions:

If X and Y are independent, then $Z = X + Y$ has density $f_Z = f_X * f_Y$.
But convolution is complex. And in frequency domain:

$$\varphi_{X+Y}(t) = \varphi_X(t) \cdot \varphi_Y(t)$$

Multiplication is simpler than convolution – this is the power of Fourier.

Examples:

DISTRIBUTION	CHARACTERISTIC FUNCTION
$N(\mu, \sigma^2)$	$\varphi(t) = \exp(i\mu t - \sigma^2 t^2/2)$
$\mathrm{Exp}(\lambda)$	$\varphi(t) = \lambda/(\lambda - it)$
$\mathrm{Poisson}(\lambda)$	$\varphi(t) = \exp(\lambda(e^{it} - 1))$

Corollary (CLT via Fourier):
Why does sum of independent r.v. → normal distribution?
Because $\varphi_N(t) = \exp(-t^2/2)$ is the only function satisfying
$\varphi(t)^n \to \varphi(t)$ under proper normalization (fixed point).

Probabilistic programming — bridge between Bayes and code

In ordinary code: $x = 5$ (deterministic value)
In probab. code: $x \sim \text{Normal}(\mu, \sigma)$ (variable = distribution)

Example (Stan pseudocode):

```
data { vector[N] y; }                       // observations
parameters { real mu; real<lower=0> sigma; }
model {
  mu ~ Normal(0, 10);                       // prior distribution
  sigma ~ Cauchy(0, 5);
  y ~ Normal(mu, sigma);                    // data model
}
```

System automatically computes $P(\mu, \sigma \mid y)$ — posterior.

Why: Declarative model description instead of manual derivation of Bayes formulas.
Algorithms (MCMC, variational inference) work "under the hood".

Applied example: pump station reliability

Problem: Pump station with 3 pumps. For operation at least 2 out of 3 are needed.
Each pump fails independently with probability $p = 0.1$ per year.
What is the probability of station failure?

```
        +----------+
   -----| Pump 1   |-----+
       +----------+     |     Need ≥2 working.
        +----------+     |
   -----| Pump 2   |-----+---►  output
        +----------+     |
        +----------+     |
   -----| Pump 3   |-----+
        +----------+
```

Model: $X \sim \text{Binomial}(n=3, p=0.1)$ — number of failed pumps

$$P(X = k) = C(3,k) \cdot 0.1^k \cdot 0.9^{3-k}$$

$P(X = 0) = 0.9^3 = 0.729$ (all working)
$P(X = 1) = 3 \cdot 0.1 \cdot 0.81 = 0.243$ (one failed)
$P(X = 2) = 3 \cdot 0.01 \cdot 0.9 = 0.027$ (two failed — failure)
$P(X = 3) = 0.001$ (all failed — failure)

$P(\text{failure}) = P(X \geq 2) = 0.027 + 0.001 = 0.028 = 2.8\%$

Comparison of redundancy schemes

SCHEME	P(failure)	COMMENT
1 pump (no redundancy)	0.1 = 10%	Base variant
2 of 2 (both needed)	$1-0.9^2$ = 19%	Worse. Series connection
1 of 2 (any sufficient)	0.1^2 = 1%	Parallel connection
2 of 3 (our case)	2.8%	Majority voting

Exponential distribution – time to failure

If λ = failure rate (1/year), then time to failure:

$T \sim \text{Exp}(\lambda)$, $P(T > t) = e^{-\lambda t}$

Mean time to failure: $E[T] = 1/\lambda$

Example: $\lambda = 0.1$/year $\Rightarrow$ $E[T]$ = 10 years – average pump lifetime

Moral: Probability – a tool for calculating system reliability.
Independence of events + formulas for connection schemes = redundancy analysis.

Information theory – mathematics of uncertainty and communication

Information as a view on space

Information theory studies the space of messages and their probabilities.

Key idea: messages form a space, and one can measure "distance" between probability distributions on this space:

- Entropy $H(X)$ – "size" of uncertainty
- KL-divergence $D(P||Q)$ – "distance" between distributions
- Mutual information $I(X;Y)$ – "intersection" of uncertainties

This geometry on the space of probabilities – foundation of machine learning.

Visualization: entropy = measure of disorder

```
  Initial state after mixing
  (low entropy)                (high entropy)

  +---------+---------+            +-------------------+
  | ● ● ● ● | ○ ○ ○ ○ |            | ○ ● ○ ● ● ○ ○ ● |
  | ● ● ● ● | ○ ○ ○ ○ |     →      | ● ○ ● ○ ○ ● ● ○ |
  | ● ● ● ● | ○ ○ ○ ○ |            | ○ ○ ● ● ○ ● ○ ● |
  | ● ● ● ● | ○ ○ ○ ○ |            | ● ● ○ ○ ● ○ ● ○ |
  +---------+---------+            +-------------------+
   Hot        Cold                  Warm (uniform)
     gas        gas

  Information: "know where which molecules"  →  "don't know where which"
  Entropy:   low (order)            →  high (disorder)
  Process:    irreversible (2nd law of thermodynamics)

  Quantitatively: S = kβ ln W, where W – number of ways to arrange molecules
  Mixed state: W much larger → S much larger
```

Shannon entropy – measure of uncertainty

Motivation: How much "information" does a message carry?

- "The sun rose in the east" – little information (expected)
- "The sun rose in the west" – much information (unexpected)

Shannon's idea (1948): Information = measure of unexpectedness

Definition of entropy:

```
+--------------------------------------------------------------+
|                                                              |
| H(X) = −Σᵢ pᵢ log₂ pᵢ     (bits)                              |
|                                                              |
| H(X) = −Σᵢ pᵢ ln pᵢ       (nats, if natural logarithm)        |
|                                                              |
+--------------------------------------------------------------+
```

Meaning: Entropy = average number of "yes/no questions" to determine the outcome of random variable X.

Examples:

DISTRIBUTION	ENTROPY	INTUITION
Coin ($\frac{1}{2}$, $\frac{1}{2}$)	$H = 1$ bit	One question: "heads?"
Die ($\frac{1}{6}$, $\frac{1}{6}$, ..., $\frac{1}{6}$)	$H = \log_2 6 \approx 2.58$	~2.58 questions on avg.
Deterministic (1,0,0)	$H = 0$	No uncertainty
Biased (0.99,0.01)	$H \approx 0.08$ bit	Almost always known

Properties of entropy:

- $H(X) \geq 0$ (non-negativity)
- $H(X) = 0 \iff X$ is deterministic (probability 1 at one outcome)
- $H(X)$ is maximal for uniform distribution (max = log n)
- $H(X,Y) \leq H(X) + H(Y)$, equality for independence

Differential entropy (for continuous distributions):

$$h(X) = -\int f(x) \ln f(x)\, dx$$

Important: This is NOT a direct generalization of discrete entropy.

- $h(X)$ can be negative (e.g., for narrow distributions)
- $h(X)$ depends on units of measurement (scaling)
- For normal: $h(X) = \frac{1}{2} \ln(2\pi e\sigma^2)$, negative when $\sigma^2 < 1/(2\pi e)$

Engineers are often confused: discrete $H(X) \geq 0$, but diff. $h(X) \in \mathbb{R}$.

Connection with thermodynamic entropy

Thermodynamics (Boltzmann, ~1870):

$$S = k\beta \ln W$$

where W – number of microstates, kβ – Boltzmann constant

Information theory (Shannon, 1948):

$$H = -\Sigma\, p_i \ln p_i$$

These are the same. (up to the constant kβ)

PHYSICS	INFORMATION THEORY
System entropy S	Amount of unknown information H
2nd law: S increases	Information is lost during transmission
Heat death	Maximum uncertainty
Maxwell's demon	Information has thermodynamic cost

Landauer's principle (1961):
Erasing 1 bit of information requires at least $k_\beta T \ln 2$ energy
→ Information is physical.

Mutual information and communication channel

Conditional entropy:
$H(Y|X)$ = "uncertainty of Y, given X is known"
$H(Y|X) = \Sigma_x\, p(x)\, H(Y|X=x) = -\Sigma_{x,y}\, p(x,y) \log p(y|x)$

Mutual information:

$$\boxed{I(X;Y) = H(X) + H(Y) - H(X,Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)}$$

How much information X and Y have in common

Visualization (Venn diagram for entropies):

```
+---------------------------------------+
|               H(X,Y)                  |
|    +-----------------------+          |
|    |          ╭-----╮      |          |
|    |   H(X)  |I(X;Y)  H(Y) |          |
|    |          ╰-----╯      |          |
|    +-----------------------+          |
+---------------------------------------+

H(X|Y)    I(X;Y)    H(Y|X)
←-----→   ←----→    ←-----→
```

Shannon's channel theorem:
Channel capacity: $C = \max I(X;Y)$
At rates < C transmission without errors is possible.

Kullback–Leibler divergence

KL-divergence (relative entropy):

$$D\_KL(P \parallel Q) = \Sigma_i\ p_i \log(p_i/q_i)$$

Meaning: "Distance" from distribution Q to distribution P
(how many bits are lost when using Q instead of P)

Properties:
- $D\_KL \geq 0$ (Gibbs' inequality)
- $D\_KL = 0 \Longleftrightarrow P = Q$
- NOT symmetric: $D\_KL(P \| Q) \neq D\_KL(Q \| P)$ in general
- Does NOT satisfy triangle inequality

Applications:
- Machine learning: loss function for classification
- Variational inference: approximation of complex distributions
- Statistics: goodness-of-fit tests, model selection

Coding and compression

Shannon's source coding theorem:
Minimum average code length = H(X)
Cannot compress better than to H(X) bits per symbol.

Example: English text
Uniformly: $\log_2(26) \approx 4.7$ bits/letter
Actually: $H \approx 1.0\text{-}1.5$ bits/letter (due to frequency imbalance)
→ Can compress by ~3 times.

Optimal codes:
- Huffman code – optimal prefix code
- Arithmetic coding – approaches H(X)

Connection with spaces:
The set of all probability distributions on n outcomes –
is an (n−1)-dimensional simplex: $\Delta^{n-1} = \{(p_1,...,p_n): \Sigma p_i=1,\ p_i \geq 0\}$
Entropy H is a smooth function on this simplex.

Connection of information theory with other areas

AREA	CONNECTION
Thermodynamics	$S = k\beta H$ (entropy = informational entropy)
Statistics	Fisher information, ML estimates
Machine learning	Cross-entropy loss, VAE, information bottleneck
Coding theory	Error-correcting codes, compression
Cryptography	Perfect secrecy, key entropy
Quantum mechanics	Von Neumann entropy, entanglement

Deep meaning:

Entropy — universal measure of "disorder" or "ignorance".
It appears wherever there are probabilities and uncertainty.
It is a bridge between discrete (bits) and continuous (thermodynamics).

Noether's Theorem — symmetry ↔ conservation law

The deepest connection between mathematics and physics

Emmy Noether's Theorem (1918):

> To each continuous symmetry of a physical system there corresponds a conservation law of some quantity.

and conversely: to each conservation law there corresponds a symmetry.

Main examples

SYMMETRY	CONSERVED QUANTITY
Shift in TIME (physics is the same yesterday and today)	ENERGY $E = \text{const}$
Shift in SPACE (physics is the same here and there)	MOMENTUM $p = \text{const}$
ROTATION (physics doesn't depend on direction)	ANGULAR MOMENTUM $L = \text{const}$
Gauge symmetry (phase of wave function)	electric charge $Q = \text{const}$
Lorentz invariance (physics doesn't depend on reference frame)	Relativistic 4-momentum $p^u = (E/c, p)$

Mathematical formulation

System is described by Lagrangian $L(q, \dot{q}, t)$

If L is invariant under transformation $q \to q + \varepsilon\delta q$:

$$Q = \Sigma_i \frac{\partial L}{\partial \dot{q}_i} \cdot \delta q_i = \text{const} \quad \text{(conserved)}$$

Examples:

- L doesn't depend on t $\Rightarrow$ $\partial L/\partial \dot{q} \cdot \dot{q} - L = H$ (Hamiltonian) is conserved
- L doesn't depend on x $\Rightarrow$ $\partial L/\partial \dot{x} = p$ (momentum) is conserved
- L doesn't depend on angle θ $\Rightarrow$ $\partial L/\partial \dot{\theta} = L_z$ (angular momentum) is conserved

Philosophical meaning

Conservation laws are not "divine commandments", but consequences of the geometry of spacetime.

Space is homogeneous (no "special points") ⇒ momentum is conserved
Space is isotropic (no "special directions") ⇒ angular momentum is conserved
Time is homogeneous (laws don't change) ⇒ energy is conserved

Lie groups → Laws of physics.

Hamiltonian mechanics — symplectic geometry in action

Lagrangian describes system through coordinates q and velocities $\dot{q}$.
Hamiltonian — through coordinates q and momenta $p = \partial L/\partial\dot{q}$.

Hamilton's canonical equations:

$$\dot{q}_i = \partial H/\partial p_i, \qquad \dot{p}_i = -\partial H/\partial q_i$$

Phase space (q, p) has symplectic structure:
2-form $\omega = \Sigma\, dp_i \wedge dq_i$ (closed, nondegenerate)

Liouville's theorem: phase volume is conserved under Hamiltonian flow.
$\int dq_1...dq_n\, dp_1...dp_n = \text{const}$

Physical meaning: cloud of initial conditions in phase space
can change shape, but not volume. Fluid is incompressible.

Poisson brackets:

$$\{f, g\} = \Sigma_i (\partial f/\partial q_i \cdot \partial g/\partial p_i - \partial f/\partial p_i \cdot \partial g/\partial q_i)$$

Properties: antisymmetry, Jacobi identity, Leibniz rule.
Equation of motion: $\dot{f} = \{f, H\} + \partial f/\partial t$
Integral of motion: $\{f, H\} = 0 \iff f$ is conserved.

Connection with quantum mechanics:
$\{f, g\} \to (1/i\hbar)[\hat{f}, \hat{g}]$ (Poisson brackets → commutator of operators)
This is not an analogy — this is exact correspondence (Dirac quantization).

==

Optimization — gradient, extrema, Lagrangian

==

Optimization problem

Given function $f(x_1, x_2, ..., x_n)$
Find point where f attains minimum (or maximum)

Two cases:
- Unconstrained: seek extremum over entire space
- Constrained: $g(x) = 0$ (on surface)

Convexity — guarantee of global optimum

Main theorem: For convex function on convex set
local minimum = global minimum.

This is fundamental fact that makes problem "solvable".

Convex set

Set C is convex if for any $x, y \in C$ and any $t \in [0,1]$:

$$tx + (1-t)y \in C$$

(Segment between any two points lies entirely in set)

```
Convex:   non-convex:  non-convex:
+---------+       +---┐               +-----+
|    ●----+---●   |   |               |     +--+
|         |       |   +---+           |        |
+---------+       +-------+           +--------+
Circle, square    "Horseshoe"         With "hole"
```

Convex function

Function f is convex if for any x, y and $t \in [0,1]$:

$$f(tx + (1-t)y) \le t \cdot f(x) + (1-t) \cdot f(y)$$

(Chord lies above graph)

```
f(x)●------------------●f(y)   ← chord
    \                 /
     \               /
      \------●------/          ← graph of f
       f(tx+(1-t)y)
```

Convexity criteria

CRITERION	FORMULATION
First order (tangent UNDER graph)	$f(y) \ge f(x) + \nabla f(x) \cdot (y-x)$ for all x, y
Second order (for C^2 functions)	$H(f) \succcurlyeq 0$ (Hessian matrix is positive semidefinite at all points)
For f(x) single variable	$f''(x) \ge 0$ for all x

Examples of convex and non-convex functions

CONVEX	NON-CONVEX
x^2 (parabola)	$-x^2$ (inverted parabola) — CONCAVE
e^x (exponential)	$\sin(x)$, $\cos(x)$ — oscillate
$\lvert x\rvert$ (absolute value)	x^3 — changes curvature
$x \log x$ $(x > 0)$	$\sqrt{x}$ — concave (but $-\sqrt{x}$ is convex)
$\lVert x\rVert$ (any norm)	$x^4 - x^2$ — multiple minima
$\max(f_1, f_2)$ if f_i convex	

Applied example: insulation thickness optimization

Problem: Pipe with heat carrier. Find optimal insulation thickness δ.

Cost = Insulation cost + Heat loss cost

$$C(\delta) = C\_insul \cdot \underset{\substack{\uparrow \\ \text{linear in } \delta}}{\delta} + C\_heat \,/\, \underset{\substack{\uparrow \\ \text{decreases as } 1/(R_0 + \delta/\lambda)}}{(R_0 + \delta/\lambda)}$$

Convexity analysis:

First term: linear (convex)
Second term: $1/(R_0 + \delta/\lambda)$ — convex for $\delta > 0$
Sum of convex = convex.

$$C''(\delta) = 2 \cdot C\_heat \cdot (1/\lambda)^2 / (R_0 + \delta/\lambda)^3 > 0 \quad ✓$$

Conclusion: Cost function is convex ⇒ found minimum is global.
Gradient descent guaranteed to find optimum.

If non-convex: Could get stuck in local minimum.
Need global optimization methods (expensive)

Preservation of convexity (combination rules)

OPERATION	RESULT
αf, where $\alpha \geq 0$	Convex (scaling)
$f + g$	Convex (sum of convex)
$\max(f, g)$	Convex (maximum of convex)
$f(Ax + b)$	Convex (affine substitution)
∩ convex sets	Convex (intersection)

Caution: $\min(f, g)$ NOT necessarily convex.
Product fg NOT necessarily convex.
Composition $f(g(x))$ requires conditions on f and g.

Gradient — direction of steepest ascent

$$\nabla f = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n} \right)$$

Properties:

- ∇f points in direction of greatest growth of f
- $|\nabla f|$ = rate of growth in this direction
- $\nabla f \perp$ level curves f = const

Example: $f(x,y) = x^2 + y^2$
$\nabla f = (2x, 2y)$ — points outward from center

Subtlety: ∇f and df — NOT the same thing.

Why this matters:
When writing $\nabla f = (\partial f/\partial x, \partial f/\partial y)$, this works in $\mathbb{R}^n$ with standard metric.
But in general, the gradient and the differential are different objects.

Differential df is covector (1-form):
df: tangent vectors → numbers
df(v) = "derivative of f in direction v" = rate of change of f

$df = (\partial f/\partial x)\, dx + (\partial f/\partial y)\, dy$ ← linear combination of basis 1-forms

Gradient grad f = ∇f is vector:
∇f is vector corresponding to covector df through metric

Connection through metric:
$df(v) = \langle \nabla f, v \rangle$ for any vector v
or: $\nabla f = g^{-1}(df)$ where g — metric tensor

In Euclidean space (g = i):
Metric is trivial, so components coincide:
$df = (\partial f/\partial x, \partial f/\partial y)$ and $\nabla f = (\partial f/\partial x, \partial f/\partial y)$ — look identical

In curvilinear coordinates (spherical, cylindrical):
Components differ. Metric is non-trivial.

Visualization:

df — "price tag" (what does shift cost?) ← covector, lives in V*
∇f — "arrow" (where to go?) ← vector, lives in V

```
Level curves f = const
    /  /  /  /  /
   /  /  /  /  /
  /  /--→/  /   ← ∇f (vector, perpendicular to level curves)
 /  /  /  /  /
/  /  /  /  /
```

df specifies "distance" between level curves
∇f shows direction of this distance (requires metric)

Why $\nabla f \perp$ level curves:
Along level curve $f = \text{const}$: $df(v) = 0$ for tangent v
Therefore $\langle \nabla f, v \rangle = 0$, i.e. ∇f is perpendicular to tangents

Practical conclusion:
In $\mathbb{R}^n$ with standard metric one can not distinguish df and ∇f.
In general relativity, on manifolds — distinction is critical.

Necessary condition for extremum

Unconstrained:

$$\nabla f = 0$$

(all partial derivatives equal zero — critical point)

To determine type of point — Hessian matrix:

$$H(f) = \begin{pmatrix} \partial^2 f/\partial x_1^2 & \partial^2 f/\partial x_1 \partial x_2 & \cdots & \partial^2 f/\partial x_1 \partial x_n \\ \partial^2 f/\partial x_2 \partial x_1 & \partial^2 f/\partial x_2^2 & \cdots & \partial^2 f/\partial x_2 \partial x_n \\ \vdots & \vdots & \ddots & \vdots \\ \partial^2 f/\partial x_n \partial x_1 & \partial^2 f/\partial x_n \partial x_2 & \cdots & \partial^2 f/\partial x_n^2 \end{pmatrix}$$

$H > 0$ (positive definite) → minimum
$H < 0$ (negative definite) → maximum
$\det(H) < 0$ → saddle

Geometric meaning of Hessian — Why does this work?

Reminder (single variable):

$f''(a) > 0 \Rightarrow$ minimum (function "curves upward" ◡)
$f''(a) < 0 \Rightarrow$ maximum (function "curves downward" ◠)

Second derivative is curvature of graph.
Positive curvature = "bowl" = minimum
Negative curvature = "dome" = maximum

Multidimensional case:

At point a function f is approximated by quadratic form:

$$f(a + h) \approx f(a) + \underbrace{\nabla f(a) \cdot h}_{=0 \text{ at crit. point}} + \underbrace{\tfrac{1}{2} h^T H(a)\, h}_{\text{quadratic form}}$$

At critical point $\nabla f = 0$, so:

$f(a + h) - f(a) \approx \frac{1}{2} h^T H h$

Sign of this expression determines where "surface" faces.

Eigenvalues of Hessian = curvatures in principal directions:

H — symmetric matrix, so diagonalizable:

$H = Q \Lambda Q^T$, where $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, ..., \lambda_n)$

λ_i is curvature of function along i-th principal direction.

```
+--------------------------------------------------------------------------+
|                                                                          |
|   All λᵢ > 0:  "bowl" in all directions = minimum                        |
|                                                                          |
|         z                                                                |
|         |    ╭---╮                                                       |
|         |   /     \                                                      |
|         |  /   ●   \  ← point at bottom of bowl                          |
|         | /---------\                                                    |
|         +------------→ x,y                                               |
|                                                                          |
|   All λᵢ < 0:  "dome" in all directions = maximum                        |
|                                                                          |
|         z                                                                |
|         | \---------/                                                    |
|         |  \   ●   /  ← point at top of dome                             |
|         |   \     /                                                      |
|         |    ╰---╯                                                       |
|         +------------→ x,y                                               |
|                                                                          |
|   Different signs λᵢ:  "saddle" — up in one, down in other             |
|                                                                          |
|         z                                                                |
|         |    /\                                                          |
|         |   /  \                                                         |
|         |  -----● ← point in saddle                                      |
|         |   \  /                                                         |
|         |    \/                                                          |
|         +------------→ x,y                                               |
|                                                                          |
+--------------------------------------------------------------------------+
```

Why bordered Hessian works:

Under constrained optimization we move not over entire space,
but only along surface $g(x) = 0$.

Regular Hessian shows curvature in all directions.
But we're interested only in feasible directions (along constraint).

Bordering "subtracts" infeasible directions:
- Zero row/column — fixes constraint
- Interior part — Hessian of Lagrange function

Minors of bordered Hessian show curvature
In space perpendicular to ∇g (i.e. along surface).

Analogy:
Walking along mountain ridge (constraint = stay on ridge).
You're interested in: is this local minimum or maximum along ridge,
not that there are cliffs on sides.

Method of Lagrange multipliers — extremum with constraint

Problem: minimize $f(x)$ subject to $g(x) = 0$

Idea: At extremum point $\nabla f \parallel \nabla g$ (gradients parallel)

$$\nabla f = \lambda \cdot \nabla g$$

where λ — Lagrange multiplier (unknown scalar)

System of equations:

- $\partial f/\partial x_i = \lambda \cdot \partial g/\partial x_i$ for all i
- $g(x) = 0$ (constraint)

Equivalent: extremum of Lagrange function $L = f - \lambda g$

Example: Find rectangle of maximum area with $P = 20$
$f = xy$ (area), $g = 2x + 2y - 20 = 0$ (perimeter)
$\nabla f = (y, x)$, $\nabla g = (2, 2)$
$y = 2\lambda$, $x = 2\lambda \Rightarrow x = y \Rightarrow$ square 5×5

Bordered Hessian matrix — determining type in constrained optimization

Problem: Critical point found by Lagrange method.
Is this minimum, maximum or saddle? Regular Hessian doesn't work.

Idea: We're interested in sign of second derivative not in entire space,
but only along constraint surface $g(x) = 0$.

Bordered Hessian (for single constraint $g = 0$):

$$\bar{H}(L) = \begin{pmatrix} 0 & \partial g/\partial x_1 & \partial g/\partial x_2 & \cdots & \partial g/\partial x_n \\ \partial g/\partial x_1 & \partial^2 L/\partial x_1^2 & \partial^2 L/\partial x_1 \partial x_2 & \cdots & \partial^2 L/\partial x_1 \partial x_n \\ \partial g/\partial x_2 & \partial^2 L/\partial x_2 \partial x_1 & \partial^2 L/\partial x_2^2 & \cdots & \partial^2 L/\partial x_2 \partial x_n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \partial g/\partial x_n & \partial^2 L/\partial x_n \partial x_1 & \partial^2 L/\partial x_n \partial x_2 & \cdots & \partial^2 L/\partial x_n^2 \end{pmatrix}$$

where $L = f - \lambda g$ — Lagrange function

Extremum conditions (check corner minors $\bar{h}$, starting with 3rd):

Denote: $|\bar{H}_k|$ – principal minor of size k × k

TYPE OF POINT	CONDITION ON MINORS ($k = 3, 4, \ldots, n+1$)
minimum	All $|\bar{H}_k| < 0$ (negative)
maximum	Signs alternate: $|\bar{H}_3| > 0$, $|\bar{H}_4| < 0$, $|\bar{H}_5| > 0$, ... (first > 0, then sign alternation)
SADDLE	Otherwise

Example: $f(x,y) = xy$, $g(x,y) = x + y - 10 = 0$

$\nabla f = (y, x)$, $\nabla g = (1, 1)$
Critical point: $x = y = 5$, $\lambda = 5$

Lagrange function: $L = xy - \lambda(x + y - 10)$
$\partial^2 L/\partial x^2 = 0$, $\partial^2 L/\partial y^2 = 0$, $\partial^2 L/\partial x\partial y = 1$

$$\bar{H}(L) = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$

$$|\bar{H}_3| = \det(\bar{H}) = 0 + 1 + 1 - 0 - 0 - 0 = 2 > 0$$

Sign $|\bar{H}_3| > 0$ → maximum (xy is maximal when $x = y = 5$)

Multiple constraints $g_1 = 0, g_2 = 0, \ldots, g_m = 0$:

Bordering extends to m rows/columns:

$$\bar{H}(L) = \begin{pmatrix} 0 & 0 & \ldots & \nabla g_1^T \\ 0 & 0 & \ldots & \nabla g_2^T \\ \vdots & \vdots & \vdots & \vdots \\ \nabla g_1 & \nabla g_2 & \ldots & H \\ & & & \end{pmatrix}$$

Check minors starting with $(2m+1)$-th.

Connection with regular Hessian:
Bordering "projects" curvature onto tangent space
to constraint surface. Zeros in corner – contribution from constraints.

KKT – conditions for problems with inequalities

Lagrange multipliers work for equalities: $g(x) = 0$
But engineering problems often have inequalities: $g(x) \le 0$

Problem: min $f(x)$ subject to $g_i(x) \le 0$, $h_j(x) = 0$

Karush–Kuhn–Tucker (KKT) conditions:

```
+-----------------------------------------------------------+
| Necessary conditions for x* to be optimum:                |
|                                                           |
| 1. STATIONARITY:  ∇f(x*) + Σμᵢ∇gᵢ(x*) + Σλⱼ∇hⱼ(x*) = 0   |
|                                                           |
| 2. FEASIBILITY:    gᵢ(x*) ≤ 0,  hⱼ(x*) = 0                |
|                                                           |
| 3. DUAL FEASIBILITY:  μᵢ ≥ 0                              |
|                                                           |
| 4. COMPLEMENTARY SLACKNESS:  μᵢ · gᵢ(x*) = 0  for all i  |
| (either constraint is active gᵢ=0, or multiplier μᵢ=0)   |
+-----------------------------------------------------------+
```

Meaning of condition 4 (complementary slackness):
- If x^* inside region ($g_i < 0$), constraint doesn't affect → $\mu_i = 0$
- If x^* on boundary ($g_i = 0$), constraint is active → $\mu_i \ge 0$

Sufficiency: For convex problems (f convex, g_i convex, h_j linear)
KKT conditions are not only necessary, but also sufficient.

Example: min $x^2 + y^2$ subject to $x + y \ge 1$ (i.e. $g(x,y) = 1 - x - y \le 0$)
  $\nabla f = (2x, 2y)$, $\nabla g = (-1, -1)$
  From stationarity: $2x - \mu = 0$, $2y - \mu = 0$ → $x = y$
  From compl. slackness: either $\mu=0$ (then $x=y=0$, but $0+0<1$ – infeasible),
                        or $g=0$ (then $x=y=\frac{1}{2}$, $\mu=1>0$ – feasible)
  Answer: $x^* = y^* = \frac{1}{2}$, $f^* = \frac{1}{2}$

Calculus of variations – When function is unknown

Find function $y(x)$ minimizing functional:

$$J[y] = \int_a^b F(x, y, y')\,dx$$

Euler–Lagrange equation:

```
           ∂F     d  ⎛ ∂F  ⎞
```

```
          ---- − --- ⎜----⎟ = 0
          ∂y     dx ⎝ ∂y' ⎠
          ∂y    dx  ⎝ ∂y' ⎠
```

Examples:
- Shortest path → straight line
- Brachistochrone (fastest descent) → cycloid
- Principle of least action → laws of mechanics

Connection with spaces

Optimization is the search for special points on a space.

Function space is infinite-dimensional.
  Each function y(x) is a "point" in this space.
  Functional J[y] is the "height" of this point.
  Calculus of variations = search for extrema in ∞-dimensional space.

Geodesics on a manifold = extrema of path length.
  On a sphere: great circles
  In GR: trajectories in gravitational field

Gradient descent — numerical optimization method

Idea: Move in the direction of steepest decrease of the function.

$$x_{n+1} = x_n - \alpha \cdot \nabla f(x_n)$$

α — step (learning rate), ∇f — gradient

Visualization:

```
             /\          level lines f = const
            /  \
           /   ●---→ ∇f (direction of growth)
          /     \
         /       \    Descent goes against gradient: −∇f
        /     ●←------\
                ↓
                ● minimum
```

Choice of step α:
- α too small: slow convergence
- α too large: divergence, "jumps" over minimum
- Adaptive α: Armijo method, Wolfe conditions

Variants:

METHOD	FEATURE
Gradient descent	$x \leftarrow x - \alpha\nabla f$
Newton's method	$x \leftarrow x - H^{-1}\nabla f$ (H = Hessian) Faster, but needs 2nd derivative
Stochastic (SGD)	Gradient on random subsample For big data (machine learning)
Adam, RMSprop	Adaptive step for each coordinate Standard in deep learning

Problems:
- Local minima (global not guaranteed)
- Saddle points ($\nabla f=0$, but not extremum)
- Poor conditioning (elongated "ravines")

Connection with physics:
Gradient descent ≈ motion of a ball on surface f(x,y)
with viscous friction (without inertia)

Stability and control theory

Why stability is needed

Main question: If the system is slightly pushed, will it return or fly away?

Examples:
- Pendulum: at bottom – stable, at top – unstable
- Satellite: stable and unstable orbits
- Heat network: will it reach regime after disturbance?
- Economy: will the market return to equilibrium?

Mathematical model:
System is described by differential equation:

$$\dot{x} = f(x)$$

where x – state vector, f – vector field
Point x^* is called equilibrium position if $f(x^*) = 0$

Phase space as "river of trajectories"

Imagine phase space as a pool with flowing water:

- Vector field $f(x)$ – this is flow velocity at each point
- Trajectory $x(t)$ – this is path of a chip thrown in water
- Equilibrium x^* – this is point where water stands still ($f = 0$)
- Stable equilibrium – this is sink (water flows in)
- Unstable equilibrium – this is source (water flows out)
- Limit cycle – this is vortex into which trajectories are drawn

This analogy with hydrodynamics is deep:
- $\mathrm{div}(f) < 0$ in sink, $\mathrm{div}(f) > 0$ in source
- $\mathrm{rot}(f) \neq 0$ around center/focus
- Liouville's theorem: incompressible flow preserves volume in phase space

In Hamiltonian mechanics the phase flow is literally incompressible.

Types of stability

TYPE	DEFINITION
Lyapunov (stability)	$\forall\varepsilon>0\ \exists\delta>0: \lvert x(0)-x^*\rvert<\delta \Rightarrow \lvert x(t)-x^*\rvert<\varepsilon\ \forall t\geq 0$ "Started close – will remain close forever"
Asymptotic	Lyapunov stable + $x(t) \to x^*$ as $t \to \infty$ "Not just close, but tends to equilibrium"
Exponential	$\lvert x(t)-x^*\rvert \leq C\cdot e^{(-\alpha t)}\lvert x(0)-x^*\rvert$ for $C,\alpha > 0$ "Converges exponentially fast"
Global	Asymptotic for any initial conditions (not only in neighborhood of x^*)
BIBO	Bounded Input → Bounded Output Bounded input gives bounded output

Visualization:

```
  Stable              Asymptotically       Unstable
  (Lyapunov)          stable

      ●                   ●                    ○
     /|\                 /|\                  /|\
    / | \               / | \                / | \
   ●--●--●             ●→-●-←●              ●--●--●→
  does not go         tends to             diverges
  far                 center
```

Linearization and eigenvalues

Idea: Near equilibrium x^* we replace nonlinear system with linear

$$\dot{x} = f(x) \approx f(x^*) + Df(x^*)(x - x^*) = A(x - x^*)$$

where $A = Df(x^*)$ — Jacobian matrix (matrix of partial derivatives)

Lyapunov's first approximation theorem:

```
+----------------------------------------------------------------------+
|                                                                      |
| If all eigenvalues λᵢ of matrix A have Re(λᵢ) < 0,                 |
| then x* is ASYMPTOTICALLY STABLE.                                    |
|                                                                      |
| If at LEAST one Re(λᵢ) > 0, then x* is unstable.                    |
|                                                                      |
| If Re(λᵢ) ≤ 0 and there exists λ with Re(λ) = 0 — need additional analysis. |
|                                                                      |
+----------------------------------------------------------------------+
```

Geometry on complex plane:

```
           Im(λ)
             |
      ○      |      ×     ← × unstable (Re > 0)
             |
  -----------+-----------  Re(λ)
             |
      ○      |      ×     ← ○ stable (Re < 0)
             |
```

All λ in left half-plane ⟺ asymptotic stability

Physical meaning:

$\mathrm{Re}(\lambda) < 0$: exponential decay $e^{(\mathrm{Re}(\lambda)t)}$

$\mathrm{Re}(\lambda) > 0$: exponential growth

$\mathrm{Im}(\lambda) \neq 0$: oscillations with frequency $|\mathrm{Im}(\lambda)|$

Phase portrait — visualization of dynamics

For system $\dot{x} = Ax$ (2D) behavior type is determined by eigenvalues.
Six basic types of fixed points:

```
+--------------------------------------------------------------------------+
|                                                                          |
|   NODE (stable)            FOCUS (stable)           SADDLE               |
|   λ₁, λ₂ < 0, real         λ = α ± iβ, α < 0        λ₁ < 0 < λ₂          |
|                                                                          |
|     ↘  ↓  ↙                  ╭--←--╮                   ↑               |
|       ↘↓↙                   ↙      ↖           ←  ←  ●  →  →          |
|   → → ●  ← ←               ↓     ●     ↑                ↓               |
|       ↗↑↖                   ↘      ↗                                  |
|     ↗  ↑  ↖                  ╰--→--╯             (2 separatrices,   |
|   (all trajectories         (damped inward        attracting       |
|    pulled to 0)              spiral)                and repelling)     |
|                                                                          |
|                                                                          |
|   CENTER                   NODE (unstable)         FOCUS (unstable)     |
|   λ = ±iβ (purely imag.)   λ₁, λ₂ > 0, real        λ = α ± iβ, α > 0    |
|                                                                          |
|         ╭-→-╮                ↖  ↑  ↗                ╭--→--╮           |
|       ╭╯  ╭-╮  ╰╮              ↖↑↗                ↗        ↘         |
|      ↑  ╭╯  ╰╮  ↑           ← ← ● → →              ↑     ●     ↓      |
|      ↑  ╰╮  ╭╯  ↑             ↙↓↘                  ↖        ↙       |
|       ╰╮  ╰-╯  ╭╯              ↙  ↓  ↘                ╰--←--╯         |
|         ╰-←-╯              (all diverge           (outward            |
|   (nested closed orbits,   from zero)              expanding spiral)   |
|    conservative)                                                       |
|                                                                          |
+--------------------------------------------------------------------------+
```

Classification in (det A, tr A) plane:

```
                         tr A
                          ↑
                          |       unstable focus
            unstable node |       (det > (tr/2)²)
          (det > 0,      /|\
           tr > 0)      / | \    ← parabola det = (tr A / 2)²
                       /  |  \
                      /   |   \
         ------------ ----●----- ← centers (tr = 0, det > 0)
                      \   |   /
                       \  |  /
            stable node \ | /   stable focus
          (det > 0,      \|/    (det > (tr/2)²)
           tr < 0)        |
                         ----→ det A
                          |
                 SADDLES: det A < 0 (λ of opposite signs)
```

Example: Pendulum (linearized)

$$\ddot{\theta} + (b/m)\dot{\theta} + (g/L)\theta = 0 \;\rightarrow\; \dot{x} = Ax, \text{ where } x = (\theta, \dot{\theta})$$

$$A = \left| \begin{matrix} 0 & 1 \\ -g/L & -b/m \end{matrix} \right| \qquad \text{Characteristic equation: } \lambda^2 + (b/m)\lambda + g/L = 0$$

$b = 0$: center (ideal pendulum, oscillations without damping)
$b > 0$ small: focus (damped oscillations)
b large: node (aperiodic damping)

Lyapunov function — generalized "energy"

Idea: Find function $V(x)$ that decreases along trajectories

Definition:

$V: \mathbb{R}^n \to \mathbb{R}$ is called Lyapunov function for $\dot{x} = f(x)$ in neighborhood of x^*, if:

1. $V(x^*) = 0$ and $V(x) > 0$ for $x \neq x^*$ (positive definite)
2. $\dot{V}(x) = \nabla V \cdot f(x) \leq 0$ along trajectories (decreases or constant)

Lyapunov theorems:

```
+-------------------------------------------------------+
|                                                       |
| V̇ ≤ 0  ⇒  x* Lyapunov stable                          |
|                                                       |
| V̇ < 0 (except x*)  ⇒  x* asymptotically stable      |
|                                                       |
+-------------------------------------------------------+
```

Example: Pendulum with friction $\ddot{x} + b\dot{x} + \sin(x) = 0$

$V = \frac{1}{2}\dot{x}^2 + (1 - \cos x)$ ("energy" = kinetic + potential)
$\dot{V} = -b\dot{x}^2 \leq 0$ (energy decreases due to friction)

Since $\dot{V} \leq 0$ and $\dot{V} = 0$ only when $\dot{x} = 0$, point $(0, 0)$ is asymptotically stable (LaSalle's invariance principle).

Advantage:
No need to solve equations. Sufficient to find suitable V.
Works for nonlinear systems.

Difficulty:
No general method for constructing V. This is art + physical intuition.

Routh–Hurwitz criterion — stability without computing roots

Problem: Check whether all roots of characteristic polynomial
$p(\lambda) = \lambda^n + a_{n-1}\lambda^{n-1} + \ldots + a_1\lambda + a_0$ lie in left half-plane

Hurwitz matrix:

$$H = \begin{pmatrix} a_{n-1} & a_{n-3} & a_{n-5} & \cdots & 0 \\ 1 & a_{n-2} & a_{n-4} & \cdots & 0 \\ 0 & a_{n-1} & a_{n-3} & \cdots & 0 \\ 0 & 1 & a_{n-2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & a_0 \end{pmatrix}$$

Criterion:

$$\boxed{\text{System stable} \iff \text{all principal minors } \Delta_1, \Delta_2, \ldots, \Delta_n > 0}$$

Special cases:

$n = 2$: $\lambda^2 + a_1\lambda + a_0$

Stable $\iff a_1 > 0$ and $a_0 > 0$

$n = 3$: $\lambda^3 + a_2\lambda^2 + a_1\lambda + a_0$

Stable $\iff a_2 > 0, a_0 > 0, a_2a_1 > a_0$

Practical significance:

No need to find roots. Only check signs of determinants.

Can analyze stability with parameters ($a_i = f(k, \tau, \ldots)$)

State space – modern approach to control

State-space model:

$\dot{x} = Ax + Bu$ (state equation)

$y = Cx + Du$ (output equation)

$x \in \mathbb{R}^n$ – state vector (internal variables)

$u \in \mathbb{R}^m$ – input vector (control inputs)

$y \in \mathbb{R}^p$ – output vector (measured quantities)

A, B, C, D – system matrices

Example: Heat exchanger

$x = (T_1, T_2, T_3)^T$ – temperatures at points

$u = (Q_{in}, G)^T$ – thermal power, flow rate

$y = T_{out}$ – output temperature

Connection with transfer function:

$$G(s) = C(sI - A)^{-1}B + D$$

Poles of $G(s)$ = eigenvalues of A

Stability: all poles in left half-plane

Controllability and observability

Controllability: Is it possible to reach any x_1 from any x_0?

Controllability matrix: $\mathcal{C} = [B \;\; AB \;\; A^2B \;\; \cdots \;\; A^{n-1}B]$

| System CONTROLLABLE $\Longleftrightarrow$ $\mathrm{rank}(\mathcal{C}) = n$ (full rank) |

Observability: Is it possible to reconstruct $x(t)$ from $y(t)$?

Observability matrix: $\mathcal{O} = [C; CA; CA^2; \cdots; CA^{n-1}]^T$

| System OBSERVABLE $\Longleftrightarrow$ $\mathrm{rank}(\mathcal{O}) = n$ (full rank) |

Duality:

(A, B) controllable $\Longleftrightarrow$ (A^T, C^T) observable

Practical meaning:

- Uncontrollability: there are "hidden" modes that cannot be influenced
- Unobservability: there are "invisible" modes that are not visible in output

P, PI, PID controllers — workhorses of automation

Problem: Maintain quantity $y(t)$ at setpoint level $r(t)$

$e(t) = r(t) - y(t)$ — control error

---

1. P-controller (proportional)

$u(t) = K_p \cdot e(t)$

Essence: Control action is proportional to the error.
The larger the deviation — the stronger the reaction.

Example (boiler):
Setpoint 50°C, current temperature 45°C, $K_p = 2$
Error $e = 50 - 45 = 5$°C
Control $u = 2 \times 5 = 10$ units of power

✓ Pros: Reacts quickly
✗ Cons: Static error remains (doesn't "push through" to setpoint)

2. PI-controller (proportional-integral)

$$u(t) = K_p \cdot e(t) + K_i \cdot \int e(\tau)d\tau$$

Essence: "Memory" is added — accumulated sum of past errors.
Even a small constant error will accumulate over time
and increase the control action.

Example (same boiler, $K_p = 2$, $K_i = 0.5$):

Step 1: T = 45°C, e = 5°C
Integral = 5
$u = 2\times5 + 0.5\times5 = 10 + 2.5 = 12.5$

Step 2: T = 47°C, e = 3°C
Integral = 5 + 3 = 8
$u = 2\times3 + 0.5\times8 = 6 + 4 = 10$

Step 3: T = 49°C, e = 1°C
Integral = 8 + 1 = 9
$u = 2\times1 + 0.5\times9 = 2 + 4.5 = 6.5$

The integral "remembers" past errors and brings to setpoint.

✓ Pros: Eliminates static error
✗ Cons: Can overregulate (overshoot the setpoint)

3. PID-controller (proportional-integral-derivative)

$$u(t) = K_p \cdot e(t) + K_i \cdot \int e(\tau)d\tau + K_d \cdot \dot{e}(t)$$

Essence: Reaction to the rate of change of error is added.
If temperature rapidly approaches the setpoint —
the derivative part "brakes", to avoid overshooting.

Example ($K_p = 2$, $K_i = 0.5$, $K_d = 1$):

Step 1: T = 45°C, e = 5°C, $\dot{e} = 5$ (error appeared)
$u = 2\times5 + 0.5\times5 + 1\times5 = 10 + 2.5 + 5 = 17.5$ ← aggressive start

Step 2: T = 47°C, e = 3°C, $\dot{e} = 3-5 = -2$ (error decreasing)
$u = 2\times3 + 0.5\times8 + 1\times(-2) = 6 + 4 - 2 = 8$ ← braking

Step 3: T = 49°C, e = 1°C, $\dot{e} = 1-3 = -2$
$u = 2\times1 + 0.5\times9 + 1\times(-2) = 2 + 4.5 - 2 = 4.5$ ← smooth approach

✓ Pros: Faster and more accurate, minimum overregulation
✗ Cons: More difficult to tune, sensitive to noise

Comparison using cruise control example

P: The more the car lags behind the set speed, the harder the gas. But if the road goes uphill – constant lag.

PI: "Remembers" that it hasn't reached the speed for a long time – adds gas until complete match. But can accelerate and overshoot.

PID: If the car rapidly accelerates – releases gas in advance, to avoid overshooting the desired speed.

Summary table

REGULATOR	RESPONSE	STATIC ERROR	OVERREGULATION
P	Fast	EXISTS	None or small
PI	Slower	none	Possible
PID	Optimal	NONE	Minimal

Tuning (Ziegler–Nichols method and others):

1. Find critical coefficient K_p_crit (at stability boundary)
2. Find period of critical oscillations T_crit
3. Calculate K_p, K_i, K_d using empirical formulas

~95% of industrial control systems use PID.

Feedback and stabilization

Without feedback (open-loop system):

```
r(t) ---→ [Regulator] ---→ [Object] ---→ y(t)
```

Problem: Does not compensate for disturbances.

With feedback (closed-loop system):

```
r(t) --→⊕--→ [Regulator] ---→ [Object] ---→ y(t)
        |                                    |
        +------------- [−1] ←----------------+
```

Stabilization theorem:

If (A, B) is controllable, then one can always find a matrix K such that the feedback system $u = -Kx$ has any desired poles.

Closed-loop system: $\dot{x} = (A - BK)x$

By choosing K one can make all eigenvalues of $(A - BK)$ negative.

LQR (Linear Quadratic Regulator):

Optimal choice of K, minimizing the quality functional:

$J = \int_0^\infty (x^T Q x + u^T R u)\, dt$

Q — penalty for state deviation

R — penalty for control expenditure

Connection with other sections

SECTION	CONNECTION WITH STABILITY and CONTROL
Linear algebra	Eigenvalues determine stability Rank of matrices → controllability/observabi
Complex numbers	Poles on the complex plane Nyquist criterion (contour in ℂ)
Series (Fourier)	Frequency analysis, Bode diagrams Transfer function $G(j\omega)$
XIX Optimization	LQR, optimal control Pontryagin's maximum principle
Diff. equations	Phase portraits, bifurcations Nonlinear dynamics, chaos

---

Deep analogy: PID controller ↔ arima model

PID (continuous control) and arima (time series analysis) — this is the same pattern from different worlds.

COMPONENT	PID	ARIMA
Current state "Where am I now?"	P (proportion.) $e(t)$	AR (autoregression) x_{t-1}, x_{t-2}, ...
Accumulated history	I (integral) $\int e(\tau)d\tau$	I (integration) ∇^{-1} = accumulation
Rate of change	D (different.) de/dt	MA (moving average) ε_{t-1}, ε_{t-2}, ... (shocks)

Both approaches answer the question: how to account for the past, present and tendency of change for prediction/control of the future?

Deep meaning:

Stability is a topological property of phase space.
System trajectories form a "flow" on a manifold.
Attractors (stable points/cycles) — singularities of this flow.

Dynamical systems — nonlinear dynamics and chaos

Nonlinear systems as a view of space

Linear systems $\dot{x} = Ax$ are predictable: nodes, foci, saddles. The real world is nonlinear: $\dot{x} = f(x)$. Nonlinearity gives rise to qualitatively new phenomena: limit cycles, bifurcations, chaos.

A dynamical system is a flow on phase space.
Instead of solving equations we study the geometry of this flow.

Classification of singular points of nonlinear systems

For $\dot{x} = f(x)$ near a fixed point $f(x^*) = 0$ behavior is determined by the Jacobian $A = Df(x^*)$ — if all $\mathrm{Re}(\lambda_i) \neq 0$ (hyperbolic point).

Grobman–Hartman theorem:

If all eigenvalues of A have nonzero real part, then the nonlinear system is topologically equivalent to the linear $\dot{x} = Ax$ near x^*.

When $\mathrm{Re}(\lambda) = 0$ — linearization is insufficient, bifurcations arise.

Limit cycles — stable oscillations

Limit cycle — an isolated closed trajectory, to which neighboring trajectories converge (or from which they flee).

Key difference from linear center: limit cycle is stable (attracts), center is neutral (neither attracts nor repels).

Example: Van der Pol oscillator

$$\ddot{x} - \mu(1 - x^2)\dot{x} + x = 0 \quad (\mu > 0)$$

For small x: negative damping (pumps energy)
For large x: positive damping (brakes)
Result: stable limit cycle — self-oscillations.

Physics: heartbeat, generators, predator-prey.

Bifurcations — qualitative change of behavior

Bifurcation — change in topology of phase portrait when parameter changes.

BIFURCATION	WHAT HAPPENS
Saddle-node (fold)	Two fixed points (stable + unstable) merge and disappear. $\dot{x} = \mu - x^2$: $\mu>0$: two roots, $\mu<0$: none.
Pitchfork (pitchfork)	One point splits into three. $\dot{x} = \mu x - x^3$: for $\mu>0$ appear $\pm\sqrt{\mu}$. Example: loss of stability of a rod under compression (Euler problem).
Hopf	Fixed point loses stability, limit cycle is born. Eigenvalues cross the imaginary axis. Example: onset of self-oscillations in a circuit.
Period doubling	Limit cycle loses stability, cycle of double period is born. Cascade of doublings → chaos (Feigenbaum scenario). Universal constant $\delta \approx 4.669$.

Chaos — deterministic unpredictability

Chaotic system:
- Deterministic (no randomness)
- Sensitive to initial conditions (butterfly effect)
- Has strange attractor (fractal structure)

Lyapunov exponent λ:

$$|\delta x(t)| \approx |\delta x(0)| \cdot e^{\lambda t}$$

$\lambda > 0$: chaos (close trajectories diverge exponentially)
$\lambda < 0$: stability (converge)
$\lambda = 0$: on the boundary

Example: Lorenz system (convection model)

$$\dot{x} = \sigma(y - x)$$
$$\dot{y} = x(\rho - z) - y$$
$$\dot{z} = xy - \beta z$$

For $\sigma=10$, $\beta=8/3$, $\rho=28$: strange attractor — "Lorenz butterfly".

Practical significance for an engineer:

Chaos means that long-term prediction is impossible even with an ideal model. But short-term prediction and statistical properties — are predictable. Control of chaotic systems is possible with small perturbations (OGY method).

Machine Learning — Geometry of Data

Main Idea

Machine Learning = geometry in feature space

- Object = point in $\mathbb{R}^n$ (n features)
- Learning = search for structure in a cloud of points
- Classification = separation of points by a hypersurface
- Regression = projection onto a subspace
- Dimensionality reduction = mapping $\mathbb{R}^n \to \mathbb{R}^k$

All ML methods are geometric operations on points in space.

Dictionary: ml ↔ mathematics

Neural network weights W = coordinates of a point in parameter space $\mathbb{R}^d$
Training = movement along the manifold of loss function L(W)
Gradient descent = following the antigradient $-\nabla L$
L2-regularization = constraint $\|W\|^2 \le C$ (ball in weight space)
Bayesian approach = prior measure on parameter space
Overfitting = going beyond the "typical set" of data

Connection with other sections of the atlas

ML is not an isolated area. It is an application of the mathematics of the atlas.

ML CONCEPT	WHERE IN THE ATLAS
Linear regression	Projection onto subspace: $\hat{y} = X(X^TX)^{-1}X^Ty$ — orthogonal projection
Backpropagation	Chain rule of differentiation: $\partial L/\partial w_i = \partial L/\partial y \cdot \partial y/\partial w_i$
Regularization	Prior distribution = measure: L2-regularization ↔ Gaussian prior on weights
SVM with kernels	Hilbert space: $K(x,y)$ — reproducing kernel, $\langle\varphi(x),\varphi(y)\rangle$
PCA	Eigenvectors: Spectral decomposition of covariance matrix
Gradient descent	Optimization on manifold: $-\nabla L \in T^*M$ → movement in direction of decrease
Softmax	Exponential family of distributions: $p(y=k) \propto \exp(z_k)$ — maximum entropy
Cross-entropy loss	KL-divergence: $H(p,q) = -\Sigma p \log q$ — information measure
Neural network	Composition of functions: $f = f_n \circ ... \circ f_1$ Universal approximation

Understanding these connections, we see: ML is not a "black box", but geometry.

Fundamental Concepts

Data Splitting

Analogy with an Exam

Bad approach:
You memorize answers to specific questions from last year's exams.
On the same questions you get 100%. On the real exam — failure.

Good approach:
You learn the material deeply. You test yourself on new problems.
If you do well on new problems → you really understood the material.

A model should work on data it hasn't seen during training.

Train / Validation / Test:

```
Original data (100%):
+- 70% → Training
|         Model "sees" this data and "learns" from it
|
+- 10% → Validation
|         Hyperparameter tuning, model comparison
|         Can be looked at multiple times
|
+- 20% → Test
          Final check once at the end
          Honest quality assessment
```

What you should never do:

✗ Train model on test
✗ Tune hyperparameters based on test results
✗ Check on test multiple times (then it becomes validation)

Overfitting and Underfitting

Overfitting

Model "memorizes" training data instead of learning patterns.

Symptoms:

- Excellent quality on train (99%)
- Poor quality on test (70%)
- Large gap between train and test

Causes: model too complex, little data, no regularization

Geometrically:

```
   •  •  class 1
 /-|-\/-|-\  ← boundary "bent" to pass through every point
--•---•---•--
   •  •  class 0
```

Model fit to noise in data, not to true dependency.

Underfitting

Model too simple and cannot learn patterns.

Symptoms:
- Poor quality on train (70%)
- Poor quality on test (68%)
- both poor, gap small

Causes: model too simple, regularization too strong

Geometrically:

```
       •  •  class 1
     •  ○○○  •
       •  •  class 0 (○) surrounds class 1
    ----+---- ← straight line cannot separate such data
```

Cross-Validation (K-Fold CV)

Problem: validation set can be "unlucky" — too simple or too complex.

Solution: average over K different splits.

```
+----+----+----+----+----+
| F1 | F2 | F3 | F4 | F5 |
+----+----+----+----+----+

Iteration 1: train on F2,F3,F4,F5  →  test on F1
Iteration 2: train on F1,F3,F4,F5  →  test on F2
...
Iteration 5: train on F1,F2,F3,F4  →  test on F5

Final score = average over 5 iterations
Standard deviation = measure of model stability

+==========================================================+
| Low std = model works the same on different data        |
| High std = model is sensitive to train choice           |
+==========================================================+
```

Regularization

Regularization = limiting model complexity to combat overfitting.

L2 (Ridge): $L(\theta) = Loss(\theta) + \lambda\|\theta\|^2$
    Shrinks weights toward zero, but doesn't zero them out

L1 (Lasso): $L(\theta) = Loss(\theta) + \lambda\|\theta\|_1$
    Zeros out some weights → feature selection

Geometric interpretation:

```
θ₂                              θ₂
↑                               ↑
|    /---\                      |    /---\
|   /  ●  \ ← OLS solution      |   /  ●  \
|  |   \   |                    |  |   /| |
|  |    ○  | ← Ridge (on circle)|  |  / ○ | ← Lasso (on diamond)
|   \     /                     |   \/   /
|    ⊙    | ← constraint        |    ◆   |
+------------→ θ₁               +------------→ θ₁
     Ridge                           Lasso
```

Ridge: ellipse touches circle → weights shrink, but $\neq 0$

Lasso: ellipse touches diamond → touches at corner → some weights $= 0$

Linear Regression — Projection Geometry

Formulation:

Given: $X \in \mathbb{R}^{n \times p}$ (n objects, p features), $y \in \mathbb{R}^n$ (response)

Find: $\theta \in \mathbb{R}^p$ such that $\hat{y} = X\theta$ minimizes $\|y - \hat{y}\|^2$

Geometric interpretation:

```
+=============================================+
| Linear regression = projection of vector y |
| onto subspace spanned by columns of X      |
+=============================================+
```

```
     y (actual vector)
     ●
    /|\
   / | \
  /  |  \
 /   |90°\
/----●----\--------→ subspace L = span(X)
     ŷ (projection)
     ↑
     This is perpendicular.
```

Optimality condition: residual $(y - \hat{y}) \perp$ subspace L

$$X^T(y - X\theta) = 0$$
$$X^T y = X^T X\theta$$
$$\theta^* = (X^T X)^{-1} X^T y \leftarrow \text{normal equation}$$

Logistic Regression — Probabilistic Classification

Formulation:

Given: $X \in \mathbb{R}^{n \times p}$, $y \in \{0,1\}^n$ (binary labels)
Find: $P(y=1|x)$ for new objects

Why not linear regression?

Problem 1: predictions go outside [0,1]
Patient with very high pressure: $\hat{y} = 1.8$ — but this is not a probability.

Problem 2: outliers break the model
Linear regression "pulls" toward outliers

Solution — sigmoid:

Linear part		Sigmoid		Probability
$z = \theta^T x$	---→	$\sigma(z) = 1/(1+e^{-z})$	---→	$P(y=1)$
$z \in (-\infty, +\infty)$				$\in [0,1]$

$z = -10$ → $P \approx 0$ (almost certain: class 0)
$z = 0$ → $P = 0.5$ (uncertainty)
$z = +10$ → $P \approx 1$ (almost certain: class 1)

Loss function (Cross-Entropy):

$$L(\theta) = -\Sigma_i[y_i \log(\hat{p}_i) + (1-y_i) \log(1-\hat{p}_i)]$$

where $\hat{p}_i = \sigma(\theta^T x_i)$

Why cross-entropy, not MSE? (critically important)

Gradients for model $\hat{y} = \sigma(z)$, $z = \theta^T x$:

MSE: $\partial L/\partial z = 2(\hat{y} - y) \cdot \sigma(z)(1-\sigma(z))$
Cross-entropy: $\partial L/\partial z = \hat{y} - y$

Key fact: $\sigma'(z) = \sigma(z)(1-\sigma(z)) \rightarrow$ zero as $|z| \rightarrow \infty$

Sigmoid saturation:

z	σ(z)	σ'(z)

```
        -----   -----   -----
        -10     0.00005 0.00005  ← almost zero.
          0     0.50000 0.25000  ← maximum
        +10     0.99995 0.00005  ← almost zero.
```

Critical example:

```
        Correct answer: y = 1
        Model predicts: z = -10 → σ(-10) ≈ 0

        Error is huge. (predicted 0 instead of 1)

        MSE gradient:  2(0 - 1) · 0.00005 ≈ -0.0001  ← microscopic.
        CE gradient:   0 - 1 = -1.0                   ← normal.
```

```
    +=================================================================+
    | With large error MSE gives small gradient → getting stuck.    |
    | Cross-entropy: derivative σ'(z) cancels during differentiation|
    | leaving clean gradient (ŷ - y)                                |
    +=================================================================+
```

Mathematically: CE = negative log-likelihood for Bernoulli:
$P(y|x) = \sigma(z)^y \cdot (1-\sigma(z))^{1-y} \rightarrow -\log P = L\_CE$

Decision boundary:

Boundary = $\{x : P(y=1|x) = 0.5\} = \{x : \theta^T x = 0\}$ ← hyperplane.

```
         •  •  class 1
         •  •
    -----------------  ← boundary θᵀx = 0
        •  •
         •  •  class 0
```

SVM — maximum margin

Idea: find a hyperplane with maximum margin between classes.

```
         •  •  class 1
        ===========  ← margin
    ---------------  ← separating hyperplane wᵀx + b = 0
        ===========
         •  •  class 0
```

Optimization problem:

$\max \ \text{margin} = 2/\|w\|$
s.t. $y_i(w^T x_i + b) \geq 1$ for all i

Equivalently: $\min \frac{1}{2}\|w\|^2$

Support Vectors:

Points lying on the margin boundary ($y_i(w^T x_i + b) = 1$).
Only they determine the solution.

● ← support vector (on margin boundary)
• ← ordinary point (doesn't affect solution)

Kernel trick:

For non-linearly separable data: project into higher-dimensional space where data becomes linearly separable.

Trick: don't compute $\varphi(x)$ explicitly, but use kernel $K(x,x') = \langle\varphi(x),\varphi(x')\rangle$

Popular kernels:
- Linear: $K(x,x') = x^T x'$
- RBF (Gaussian): $K(x,x') = \exp(-\gamma\|x-x'\|^2)$
- Polynomial: $K(x,x') = (x^T x' + c)^d$

KNN — classification by neighbors

Idea: "Tell me who your neighbors are — and I'll tell you who you are."

Algorithm:

1. For new object x find K nearest neighbors in training set
2. Class of x = majority vote among K neighbors

K=1: boundary = Voronoi diagram
K large: boundary smooths out

Features:

✓ No training (lazy learning)
✓ Natural non-linear boundary
✓ Easy to add new data

✗ Slow prediction O(nN) where N — training set size
✗ Curse of dimensionality (in high dimensions all points are "far")
✗ Sensitive to feature scale → normalization needed

PCA — dimensionality reduction

Formulation:

Given: $X \in \mathbb{R}^{n \times p}$ (centered data)
Find: directions $w_1, w_2, \ldots, w_k$ of maximum variance

Main theorem:

```
+=============================================================+
| Directions of maximum variance =                            |
| = eigenvectors of covariance matrix S = XᵀX/(n-1)           |
|                                                             |
| Variance along PCᵢ = eigenvalue λᵢ                          |
+=============================================================+
```

Derivation:

$\max \ \mathrm{Var}(Xw) = w^\top S w$
s.t. $\|w\| = 1$

Lagrangian: $\mathcal{L} = w^\top S w - \lambda(w^\top w - 1)$
$\partial\mathcal{L}/\partial w = 2Sw - 2\lambda w = 0$

$Sw = \lambda w$ ← eigenvector problem.

Geometric intuition:

```
x₂
 ↑     PC1 (direction of max. variance)
 |    /●
 |  ●/   ●
 | ●/ ●  ●    ← PC1 goes along "length" of cloud
 |●/ ●   ●       PC2 ⊥ PC1
 +/----------→ x₁
  \
   PC2
```

Projection onto PC1 preserves maximum information about differences between points.

Why variance = information?

Large variance → points well distinguishable → much information
Small variance → points "stuck together" → little information
Zero variance = constant → no information

Connection with SVD:

$X = U\Sigma V^\top$

- Columns of V = directions of principal components (loadings)
- $\Sigma_{ii}^2 / (n-1)$ = eigenvalues of S = variances
- Columns of $U \cdot \Sigma$ = projections onto principal components (scores)

Comparison of classification methods

Aspect	Logistic regr.	SVM	KNN
Approach	Probabilistic	Geometric (margin)	Neighbors

Boundary	Linear	Linear/non-linear	Any
Training	Iterative	Quadr. optimization	None
Prediction	Fast O(p)	Fast O(#SV·p)	Slow
Probabilities	✓ Natural	✗ None	✓ Approx.
Interpretability	✓ High	~ Medium	✓ High
High dimensions	✓ Good	✓ Good	✗ Bad
Large data	✓ Good	✗ Bad	✗ Bad

Recommendations:

Logistic regression:
✓ Need probabilities, interpretability important, linear boundary sufficient
✓ Always start with it as baseline.

SVM:
✓ Medium data volume ($n < 10{,}000$), high dimensionality
✓ Non-linear boundary (with kernel), need robustness to outliers

KNN:
✓ Small data, complex boundary, no time to tune model
✓ Recommender systems, finding similar objects

Neural Networks — Composition of Nonlinear Transformations

Idea and Structure

Why are neural networks needed?

Linear methods are limited: if the dependency is $y = \sin(x_1) + x_2^3$, a linear model won't cope.

Neural network = composition of linear transformations and nonlinear activations.

Architecture (one hidden layer):

$$f(x) = W_2 \cdot \sigma(W_1 x + b_1) + b_2$$

where:
- $W_1 \in \mathbb{R}^{h \times n}$, $b_1 \in \mathbb{R}^h$ (input → hidden layer, h neurons)
- $\sigma: \mathbb{R} \to \mathbb{R}$ (nonlinear activation, elementwise)
- $W_2 \in \mathbb{R}^{k \times h}$, $b_2 \in \mathbb{R}^k$ (hidden → output layer)

$$\underset{\text{input}}{x \in \mathbb{R}^n} \rightarrow \underset{\text{affine}}{[W_1, b_1]} \rightarrow \underset{\text{linear}}{z \in \mathbb{R}^h} \rightarrow \underset{\text{nonlinearity}}{[\sigma] \rightarrow a \in \mathbb{R}^h} \rightarrow \underset{\text{affine}}{[W_2, b_2]} \rightarrow \underset{\text{output}}{y \in \mathbb{R}^k}$$

Geometric interpretation:

- $W_1x + b_1$ = affine transformation $\mathbb{R}^n \to \mathbb{R}^h$ (rotation, stretching, shift)
- σ = nonlinear deformation ("folding" of space)
- W_2 = linear combination $\mathbb{R}^h \to \mathbb{R}^k$

Composition yields nonlinear mapping $\mathbb{R}^n \to \mathbb{R}^k$.

Activation functions:

$\mathrm{ReLU}(z) = \max(0, z)$ ← most popular, "folds" half of space
$\sigma(z) = 1/(1+e^{-z})$ ← sigmoid, compresses into (0,1)
$\tanh(z) = (e^z-e^{-z})/(e^z+e^{-z})$ ← compresses into (-1,1)

Universal Approximation Theorem:

A neural network with one hidden layer and a sufficient number of neurons can approximate any continuous function.

Two spaces in neural networks — key idea

In ML there are two different spaces, and they must not be confused:

Data space (input space)

- Each point = one example (image, text, measurement)
- Dimensionality = number of features (784 for MNIST, millions for LLM)
- Neural network deforms this space so that classes become separable

Picture: cloud of points of two colors. Initially mixed.
After several layers — separated by hyperplane.

```
Before training: after training:
• ○ • ○ •                • • • | ○ ○ ○
○ • ○ • ○                • • • | ○ ○ ○
• ○ • ○ •                • • • | ○ ○ ○
(mixed)                  (linearly separable)
```

Weight space (parameter space) — this is where training happens

- Each point = one set of weights $\theta = (W_1, b_1, W_2, b_2, ...)$
- Dimensionality = total number of parameters (millions or billions)
- Loss function $L(\theta)$ — this is the "height" at each point

Loss function landscape:

```
            /\      /\
           /  \    /  \         ← local maxima (bad)
      ----/----\--/----\----
             \/                 ← saddle points (many)
          glob.min               ← want to get here
```

Training = journey through this landscape in search of minimum.
Gradient descent = "rolling down the slope".

Why is this difficult:
- Landscape in millions of dimensions — cannot be visualized
- Many local minima and saddle points
- Gradient can be huge or vanishingly small

Surprising fact:
In high dimensionality most critical points are saddles, not minima.
Therefore gradient descent usually finds a path to a good solution.

Tensor in ML ≠ tensor in mathematics

In PyTorch/TensorFlow "tensor" is simply a multidimensional array of numbers:

```
  torch.tensor([1, 2, 3])           # "1D-tensor" = vector
  torch.tensor([[1,2], [3,4]])      # "2D-tensor" = matrix
  torch.randn(3, 4, 5)              # "3D-tensor" = 3×4×5 array
```

In mathematics/physics tensor is an object that:
- Transforms according to a specific law under change of coordinates
- Covariant indices: $T'_i = (\partial x^j/\partial x'^i)\, T_j$
- Contravariant: $T'^i = (\partial x'^i/\partial x^j)\, T^j$

	ML "TENSOR"	MATHEMATICAL TENSOR
What is it?	Data container	Geometric object
Depends on basis?	Yes (just numbers)	No (invariant)
Transformation law	No	Yes (co/contravariance)
Example	Batch of images	Metric tensor g_{ij}

Why this is important:
In ordinary ML covariance is not needed — there is no "change of coordinates".
But in Geometric Deep Learning (graph networks, equivariant CNN)
true tensor structure is required: output must correctly
transform under rotation/reflection of input.

Backpropagation — error backpropagation

Intuition: "error echo"

Imagine: you shoot an arrow from a bow and miss.

Miss (output error) = you hit 10 cm to the left.

Who's to blame?
- Maybe your hand trembled during the shot (last "layer")
- Maybe you aimed poorly (middle "layer")
- Maybe you stood incorrectly (first "layer")

Backpropagation = distribution of blame from end to beginning.

Error "reflects" from the output and goes backward through the network, like an echo, decaying or amplifying at each layer.

```
input → [layer 1] → [layer 2] → [layer 3] → output
                                              ↓
                                          Error
                                              ↓
∂L/∂W₁ ← [δ₁] ← [δ₂] ← [δ₃] ← ∂L/∂output
          ↑        ↑        ↑
     decays  (or amplifies) at each layer
```

δ (delta) = "how much blame" this neuron bears for the final error.

Vanishing gradient problem:
If each layer multiplies δ by a number < 1, then $\delta_1 \approx 0$.
First layers don't learn. (echo faded)

ReLU solves this problem: gradient is either 0 or 1 (doesn't vanish).

Task: compute $\nabla L(\theta)$ for all parameters $\theta = \{W_1, b_1, W_2, b_2, ...\}$

Idea: apply chain rule for composition of functions.

For $f = g \circ h$: $\partial f/\partial x = (\partial g/\partial h) \cdot (\partial h/\partial x)$

Algorithm:

```
Forward pass (save intermediate values):
for i = 1 to L:
    $a^i = W_i z^{i-1} + b_i$        (linear transformation)
    $z^i = \sigma_i(a^i)$            (activation)

Backward pass (compute gradients from end to beginning):

$\delta^L = \partial L/\partial a^L$              (gradient at last layer)
```

```
for i = L-1 to 1:
    $\delta^i = (W_{i+1}^\top \cdot \delta^{i+1}) \odot \sigma'_i(a^i)$    (recurrent computation)

Gradients with respect to parameters:
    $\partial L/\partial W_i = \delta^i \cdot (z^{i-1})^\top$    (outer product)
    $\partial L/\partial b_i = \delta^i$
```

Computational complexity: same as forward pass. $O(\Sigma_i n_{i-1} \cdot n_i)$

Numerical example:

Network: $\mathbb{R}^1 \to \mathbb{R}^1$, one hidden neuron, ReLU
$f(x) = w_2 \cdot \mathrm{ReLU}(w_1 x + b_1) + b_2$

Parameters: $w_1=1$, $b_1=-1$, $w_2=2$, $b_2=0$
Data: $x=2$, $y=5$

Forward:
$a^1 = w_1 \cdot x + b_1 = 1 \cdot 2 + (-1) = 1$
$z^1 = \mathrm{ReLU}(1) = 1$
$\hat{y} = w_2 \cdot z^1 + b_2 = 2 \cdot 1 + 0 = 2$
$L = \frac{1}{2}(\hat{y}-y)^2 = \frac{1}{2}(2-5)^2 = 4.5$

Backward:
$\partial L/\partial \hat{y} = \hat{y} - y = -3$

$\partial L/\partial w_2 = \partial L/\partial \hat{y} \cdot z^1 = -3 \cdot 1 = -3$
$\partial L/\partial b_2 = \partial L/\partial \hat{y} \cdot 1 = -3$

$\delta^1 = \partial L/\partial \hat{y} \cdot w_2 \cdot \mathrm{ReLU}'(a^1) = -3 \cdot 2 \cdot 1 = -6$

$\partial L/\partial w_1 = \delta^1 \cdot x = -6 \cdot 2 = -12$
$\partial L/\partial b_1 = \delta^1 \cdot 1 = -6$

Optimization

Gradient descent:

$$\theta_{t+1} = \theta_t - \eta \cdot \nabla L(\theta_t)$$

η = learning rate

Variants:

Batch GD:	gradient over entire sample — accurate, but slow
SGD:	gradient over one example — noisy, but fast
Mini-batch:	gradient over batch of B examples — compromise (B=32,64,128)

Momentum:

$v_t = \beta \cdot v_{t-1} + \nabla L(\theta_t)$ (accumulation of "velocity")
$\theta_{t+1} = \theta_t - \eta \cdot v_t$

Analogy: ball rolling down a hill with inertia

Adam (Adaptive Moment Estimation):

$m_t = \beta_1 \cdot m_{t-1} + (1-\beta_1) \cdot \nabla L$ (gradient mean)
$v_t = \beta_2 \cdot v_{t-1} + (1-\beta_2) \cdot (\nabla L)^2$ (gradient square mean)

$\hat{m}_t = m_t / (1-\beta_1^t), \quad \hat{v}_t = v_t / (1-\beta_2^t)$ (bias correction for zero init)

$\theta_{t+1} = \theta_t - \eta \cdot \hat{m}_t / (\sqrt{\hat{v}_t} + \varepsilon)$

Each parameter has its own adaptive learning rate.
Standard values: $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\varepsilon = 10^{-8}$.

Convolutional Neural Networks (CNN)

Problem of fully connected networks for images:

MNIST 28×28: 784 inputs → $W_1 \in \mathbb{R}^{128\times784}$ = 100k parameters
ImageNet 224×224×3: 150,528 inputs → $W_1 \in \mathbb{R}^{128\times150528}$ = 19M parameters.

and this is only one layer.

CNN ideas:

1. Locality: each neuron looks only at k×k neighborhood
2. Weight sharing: one filter is used for all positions
3. Hierarchy: edges → textures → parts → objects

Convolution:

$$Y[i,j] = \Sigma_m \Sigma_n K[m,n] \cdot X[i+m, j+n]$$

Geometrically: dot product of kernel K with image patch

Example (vertical edge detector):

$$X = \begin{bmatrix} 1 & 2 & 3 & 0 \\ 0 & 1 & 2 & 3 \\ 3 & 0 & 1 & 2 \\ 4 & 2 & 0 & 1 \end{bmatrix} \qquad K = \begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix}$$

$Y[0,0] = (-1)\cdot1 + 1\cdot2 + (-1)\cdot0 + 1\cdot1 = 2$ (vertical edge present)

Pooling:

Max Pooling: $Y[i,j] = \max\{X[2i+m, 2j+n] \mid m,n \in \{0,1\}\}$

- Reduces dimensionality by factor of 2 along each axis
- Provides local shift invariance
- No trainable parameters

Typical architecture:

Input → [Conv-ReLU-Pool]×N → Flatten → FC → Output

LeNet-5 example:
INPUT 28×28×1 → CONV1 24×24×20 → POOL 12×12×20 →
→ CONV2 8×8×50 → POOL 4×4×50 → FLATTEN 800 → FC 500 → FC 10

Parameter savings:

Convolutional layer k=3, $C_{in}=64$, $C_{out}=128$:
$\quad$Parameters = $3^2 \cdot 64 \cdot 128 + 128 \approx 74\text{k}$

Fully connected layer for image 32×32×64 → 32×32×128:
$\quad$Parameters = $(32\cdot32\cdot64) \cdot (32\cdot32\cdot128) = 8.6$ billion.

CNN saves parameters by 4-5 orders of magnitude.

Feature hierarchy:

Layer 1-2: Edges, corners, simple gradients (receptive field 3-7 px)
Layer 3-5: Textures, repeating patterns (RF 20-50 px)
Layer 6-8: Object parts (eyes, wheels) (RF 100+ px)
Final: Classification by high-level features

Comparison of classical methods and neural networks

Criterion	Linear methods	Neural networks
Parameters	O(n)	O(h·n) or more
Solution	Closed form or convex	Iterative, non-convex
Expressiveness	Only linear	Arbitrary continuous
Interpretability	High	Low
Data	Works on small	Requires a lot of data
Computation	Easy on CPU	Requires GPU

When to use what:

Linear methods:
✓ n < 1000 features
✓ Dependency close to linear
✓ Interpretability important
✓ Little data (N < 10,000)

```
Neural networks:
✓ n > 1000 features (images, text)
✓ Nonlinear dependencies
✓ A lot of data (N > 100,000)
✓ GPU available
```

PART V: NUMERICAL METHODS — TOOLS FOR COMPUTATION

Why this section

The Atlas so far has been about understanding mathematics.
This section is about computation: ready-made algorithms for practical problems.

Each method is described as follows:

1. Problem (what we solve)
2. Algorithm (step by step)
3. Example with numbers
4. When to use / when not to use
5. Code (pseudocode or Python-like)

Connection with theory — why methods work

Numerical methods are not just "recipes". Behind each one stands a theorem.

METHOD	THEORETICAL JUSTIFICATION
Bisection	Intermediate value theorem: continuous function takes all values between $f(a)$ and $f(b)$
Newton	Banach fixed-point theorem: iterations $x_{n+1} = g(x_n)$ converge if g is a contraction. Newton: $g(x) = x - f(x)/f'(x)$
Gauss (LU-decomp.)	Factorization in group $GL(n)$: any invertible matrix = product of elementary (permutation, scale, shear)
Quadratures	Integral as linear functional: $\int\omega$ is a covector on function space. Quadratures — finite-dimensional approx.
FFT	Orthogonality of $\exp(2\pi ikn/N)$ in L^2: discrete analog of Fourier expansion. Speed = recursion over group $\mathbb{Z}/N\mathbb{Z}$
Runge–Kutta	Taylor expansion + coefficient matching for maximum order
Gradient descent	Derivative as covector: $\nabla f \in T^*M$ points in direction of growth. Metric g converts it to vector $-g^{-1}\nabla f$

Understanding theory helps:

- Predict when method will NOT work
- Choose right method for problem
- Estimate error without experiment

Solving nonlinear equations $f(x) = 0$

Bisection method

Problem: Find root $f(x) = 0$ on interval $[a, b]$, where $f(a)$ and $f(b)$ have different signs.

Idea: Divide interval in half, choose half with different signs, repeat.

Algorithm:

1. $c = (a + b) / 2$
2. If $f(c) \approx 0$ — found root
3. If $f(a)$ and $f(c)$ have different signs — root in $[a, c]$, take $b = c$
4. Otherwise — root in $[c, b]$, take $a = c$
5. Repeat until desired accuracy

Example: $f(x) = x^3 - x - 1$, find root on $[1, 2]$

Check: $f(1) = 1 - 1 - 1 = -1 < 0$
$f(2) = 8 - 2 - 1 = 5 > 0$ ✓ Different signs

Step 1: $c = 1.5$, $f(1.5) = 3.375 - 1.5 - 1 = 0.875 > 0$
Root in $[1, 1.5]$

Step 2: $c = 1.25$, $f(1.25) = 1.953 - 1.25 - 1 = -0.297 < 0$
Root in $[1.25, 1.5]$

Step 3: $c = 1.375$, $f(1.375) = 2.600 - 1.375 - 1 = 0.224 > 0$
Root in $[1.25, 1.375]$

Step 4: $c = 1.3125$, $f(1.3125) \approx -0.051 < 0$
Root in $[1.3125, 1.375]$

… Continue until desired accuracy. Answer: $x \approx 1.3247$

Convergence: Linear. Each iteration halves the interval.
After n iterations: error $\leq (b-a)/2^n$
For 6 decimal places need ≈ 20 iterations.

✓ When to use: Always works if there is sign change
✗ When not to use: Slow, doesn't work for multiple roots

Code:

```
  while b - a > eps:
      c = (a + b) / 2
      if f(a) * f(c) < 0:
          b = c
      else:
          a = c
  return c
```

Newton's method (tangent method)

Problem: Find root $f(x) = 0$, given initial approximation x_0.

Idea: Replace curve with tangent, find intersection with x-axis.

Formula:

$$x_{n+1} = x_n - f(x_n) / f'(x_n)$$

Geometrically:

```
      y
      |     /
      |    / f(x)
      |   ●----------------- tangent at xₙ
      |  /|
      | / |
      |/  |
  ----●---+---------------→ x
    x*   xₙ₊₁  xₙ
```

Tangent line: $y = f(x_n) + f'(x_n)(x - x_n)$

Intersection with $y = 0$: $x_{n+1} = x_n - f(x_n)/f'(x_n)$

Example: $f(x) = x^2 - 2$ (finding $\sqrt{2}$), $x_0 = 1$

$f(x) = x^2 - 2$, $f'(x) = 2x$

Step 1: $x_1 = 1 - (1 - 2)/(2 \cdot 1) = 1 - (-1)/2 = 1.5$

Step 2: $x_2 = 1.5 - (2.25 - 2)/(2 \cdot 1.5) = 1.5 - 0.25/3 = 1.4167$

Step 3: $x_3 = 1.4167 - (2.007 - 2)/(2 \cdot 1.4167) = 1.4142$

Already 4 digits after 3 iterations. ($\sqrt{2} \approx 1.41421356$)

Convergence: Quadratic. Number of correct digits doubles each iteration.

✓ When to use: Fast, if good initial approximation

✗ When not to use:
- $f'(x) = 0$ near root (division by 0)
- Bad x_0 (may go wrong way or cycle)
- Need to be able to compute derivative

Code:

```
x = x0
for i in range(max_iter):
    x = x - f(x) / df(x)
    if abs(f(x)) < eps:
        break
return x
```

Secant method

Idea: Like Newton, but replace derivative with difference quotient.
Don't need to know f'(x)!

Formula:

$$x_{n+1} = x_n - f(x_n) \cdot (x_n - x_{n-1}) / (f(x_n) - f(x_{n-1}))$$

Convergence: Superlinear (order ≈ 1.618, golden ratio)
Slower than Newton, but no derivative needed.

✓ When to use: When f'(x) is difficult to compute

Comparison of methods

METHOD	CONVERGENCE	REQUIRED	RELIABILITY
Bisection	Linear (slow)	[a,b] with diff signs	100% if root exists
Newton	Quadratic (very fast)	f'(x), good x_0	May diverge
Secant	Superlinear (~1.618)	Two initial points	May diverge

Engineering trade-offs: accuracy vs cost

Computer doesn't know which method is "better". Engineer must choose.

Question: Why not always use the most accurate method (Newton)?

Answer: Each call to f'(x) is computation.
If f(x) is a complex function (simulation, DB query), it's expensive.

Example: Finding optimal reactor temperature

f(T) = product yield at temperature T
One call to f(T) = running simulation for 10 minutes

Bisection: 20 iterations × 1 call = 20 calls = 200 minutes
Newton: 5 iterations × 2 calls (f and f') = 10 calls = 100 minutes

But: if f' needs to be computed numerically (2 more calls to f), then:
Newton: 5 × 3 = 15 calls = 150 minutes ← not so advantageous anymore.

Rule: Choose method based on cost of function call.

- f cheap (formula) → Newton
- f expensive (simulation) → bisection or derivative-free methods
- f with noise (experiment) → methods stable to noise

Practical recommendation:
Bisection for reliability → Newton for accuracy (combined method)

Numerical integration

Trapezoidal rule

Problem: Compute $\int_a^b f(x)\,dx$

Idea: Replace curve with piecewise linear (trapezoids).

```
y|    /\
 |   /  \
 |  /    \
 | /------\-------- ← piecewise linear (trapezoids)
 |/        \
 +---------------→ x
  a  x₁ x₂  ...  b
```

Formula (n equal intervals, $h = (b-a)/n$):

$$\int_a^b f(x)\,dx \approx h \cdot [f(a)/2 + f(x_1) + f(x_2) + \dots + f(x_{n-1}) + f(b)/2]$$

Example: $\int_0^1 x^2\,dx$, $n = 4$ (exact answer = $1/3 \approx 0.333$)

$h = 1/4 = 0.25$
Points: $x_0 = 0$, $x_1 = 0.25$, $x_2 = 0.5$, $x_3 = 0.75$, $x_4 = 1$
Values: $f = 0, 0.0625, 0.25, 0.5625, 1$

$I \approx 0.25 \times [0/2 + 0.0625 + 0.25 + 0.5625 + 1/2]$
$= 0.25 \times [0 + 0.0625 + 0.25 + 0.5625 + 0.5]$
$= 0.25 \times 1.375 = 0.34375$

Error: $0.34375 - 0.333... = 0.01$ (about 3%)

Error: $O(h^2)$ – halving h reduces error by factor of four

Simpson's rule (parabolas)

Idea: Replace curve with parabolas (more accurate than trapezoids).
Requires even number of intervals n.

Formula:

$$\int_a^b f\,dx \approx (h/3) \cdot [f(a) + 4f(x_1) + 2f(x_2) + 4f(x_3) + ... + f(b)]$$

Coefficients: 1, 4, 2, 4, 2, 4, ..., 2, 4, 1

Example: $\int_0^1 x^2\,dx$, $n = 4$

$$\begin{aligned} I &\approx (0.25/3) \times [0 + 4\times0.0625 + 2\times0.25 + 4\times0.5625 + 1] \\ &= (0.25/3) \times [0 + 0.25 + 0.5 + 2.25 + 1] \\ &= (0.25/3) \times 4 = 0.3333... \end{aligned}$$

Exact answer. (Simpson is exact for polynomials up to degree 3)

Error: $O(h^4)$ — much more accurate than trapezoids.

✓ When to use: Almost always better than trapezoids
✗ When not to use: Discontinuous functions, strong oscillations

Code:

```
h = (b - a) / n
I = f(a) + f(b)
for i in range(1, n):
    x = a + i * h
    if i % 2 == 1:
        I += 4 * f(x)
    else:
        I += 2 * f(x)
return I * h / 3
```

Gaussian quadrature — maximum accuracy

Idea: Choose points x_i and weights w_i optimally (not uniformly)

$$\int_{-1}^1 f(x)\,dx \approx \sum_i w_i\, f(x_i)$$

Table of nodes and weights (on [−1, 1]):

n	NODES x_i	WEIGHTS w_i
1	0	2
2	$\pm1/\sqrt{3} \approx \pm0.577$	1, 1
3	0, $\pm\sqrt{(3/5)} \approx \pm0.775$	8/9, 5/9, 5/9
4	±0.340, ±0.861	0.653, 0.653, 0.348, 0.348

```
 Accuracy: n Gaussian points integrate exactly polynomials up to degree 2n−1!
           (Simpson with 3 points — only up to degree 3)

 For arbitrary [a, b]:
   Substitution: x = (b−a)t/2 + (a+b)/2,  t ∈ [−1, 1]
   dx = (b−a)/2 dt


===============================================================================
Solving Systems of Linear Equations Ax = b
===============================================================================

 Gaussian Elimination (direct method)


 Idea: Reduce the matrix to triangular form, solve by back substitution.

 Algorithm:

 1. forward elimination: Zero out elements below the diagonal

    + a₁₁  a₁₂  a₁₃ | b₁ +      + a₁₁  a₁₂  a₁₃ | b₁  +
    | a₂₁  a₂₂  a₂₃ | b₂ |  →   |  0   a'₂₂ a'₂₃| b'₂ |
    + a₃₁  a₃₂  a₃₃ | b₃ +      +  0    0   a"₃₃| b"₃ +

 2. back substitution: Find xₙ, then xₙ₋₁, ..., x₁

    x₃ = b"₃ / a"₃₃
    x₂ = (b'₂ − a'₂₃ x₃) / a'₂₂
    x₁ = (b₁ − a₁₂ x₂ − a₁₃ x₃) / a₁₁

 Example:
    2x + y − z = 8
   −3x − y + 2z = −11
   −2x + y + 2z = −3

   Augmented matrix:
   +  2   1  −1 |  8  +
   | −3  −1   2 |−11  |
   + −2   1   2 | −3  +

   Step 1: R₂ → R₂ + (3/2)R₁,  R₃ → R₃ + R₁
   +  2    1    −1  |   8   +
   |  0   0.5   0.5 |   1   |
   +  0    2     1  |   5   +

   Step 2: R₃ → R₃ − 4R₂
   +  2    1    −1  |   8   +
   |  0   0.5   0.5 |   1   |
   +  0    0    −1  |   1   +

   Back substitution:
   z = 1/(−1) = −1
   y = (1 − 0.5×(−1)) / 0.5 = 1.5/0.5 = 3
   x = (8 − 1×3 − (−1)×(−1)) / 2 = 4/2 = 2
```

Answer: $x = 2$, $y = 3$, $z = -1$

Complexity: $O(n^3)$ operations

Important: Pivoting for numerical stability.

LU decomposition

Idea: Decompose $A = LU$, where L is lower triangular, U is upper triangular.

Why: If we need to solve many systems with the same A but different b.
$Ax = b \rightarrow LUx = b \rightarrow Ly = b$ (easy), $Ux = y$ (easy).

Decomposition is done once ($O(n^3)$), then each solution – $O(n^2)$.

Simple iteration method (for large sparse systems)

Idea: Transform $Ax = b$ into $x = Bx + c$ and iterate: $x_{n+1} = Bx_n + c$

Convergence: If $\|B\| < 1$ (spectral radius < 1)

Jacobi and Gauss–Seidel methods – specific ways to construct B and c.

✓ When to use: Very large sparse matrices (thousands × thousands)
✗ When not to use: Dense matrices – Gaussian elimination is faster

Numerical Solution of ODEs

Cauchy problem: $y' = f(t, y)$, $y(t_0) = y_0$

Find function $y(t)$, knowing the initial condition and differential equation.

Euler's method (simplest)

Idea: Replace derivative with difference: $y' \approx (y_{n+1} - y_n)/h$

Formula:

$$y_{n+1} = y_n + h \cdot f(t_n, y_n)$$

Geometrically: Go along the tangent for step h

```
      y
      |      ●--→ where we go (Euler)
      |     / \
      |    /   \ where we actually go
      |   ●
      |  yₙ
      +-------------→ t
         tₙ   tₙ₊₁
```

Example: $y' = y$, $y(0) = 1$ (solution: $y = e^t$)

$h = 0.1$, $f(t, y) = y$

$$\begin{aligned}
y_0 &= 1 \\
y_1 &= 1 + 0.1\times1 = 1.1 \\
y_2 &= 1.1 + 0.1\times1.1 = 1.21 \\
y_3 &= 1.21 + 0.1\times1.21 = 1.331 \\
&\ldots \\
y_{10} &= y(1) \approx 2.594 \quad (\text{exact: } e^1 \approx 2.718)
\end{aligned}$$

Error ≈ 5% over 10 steps. Not very accurate.

Error: $O(h)$ per step, $O(h)$ globally — first-order method

✓ When to use: Quick prototype, learning
✗ When not to use: When accuracy is needed

4th-order Runge–Kutta method (rk4) — workhorse

Idea: Estimate slope at several points and average.

Formulas:

$$\begin{aligned}
k_1 &= f(t_n, y_n) \\
k_2 &= f(t_n + h/2, y_n + h\cdot k_1/2) \\
k_3 &= f(t_n + h/2, y_n + h\cdot k_2/2) \\
k_4 &= f(t_n + h, y_n + h\cdot k_3)
\end{aligned}$$

$$y_{n+1} = y_n + (h/6)(k_1 + 2k_2 + 2k_3 + k_4)$$

Intuition:

- k_1 — slope at the beginning
- k_2 — slope at the midpoint (by estimate k_1)
- k_3 — slope at the midpoint (by estimate k_2, more accurate)
- k_4 — slope at the end
- Result — weighted average: $(1\times k_1 + 2\times k_2 + 2\times k_3 + 1\times k_4)/6$

Example: $y' = y$, $y(0) = 1$, $h = 0.1$

$$\begin{aligned}
k_1 &= 1 \\
k_2 &= 1 + 0.1\times1/2 = 1.05 \\
k_3 &= 1 + 0.1\times1.05/2 = 1.0525 \\
k_4 &= 1 + 0.1\times1.0525 = 1.10525
\end{aligned}$$

$$y_1 = 1 + (0.1/6)(1 + 2\times1.05 + 2\times1.0525 + 1.10525)$$
$$= 1 + (0.1/6)\times6.31025 = 1.10517.$$

Exact: $e^{0.1} = 1.10517$. ← Matches to 5 digits in 1 step.

Error: $O(h^4)$ per step, $O(h^4)$ globally — 4th-order method

✓ When to use: Universal method, suitable for most problems
✗ When not to use:
- Stiff equations (implicit methods needed)
- Very high accuracy (adaptive step needed)

Code:

```
def rk4_step(f, t, y, h):
    k1 = f(t, y)
    k2 = f(t + h/2, y + h*k1/2)
    k3 = f(t + h/2, y + h*k2/2)
    k4 = f(t + h, y + h*k3)
    return y + h/6 * (k1 + 2*k2 + 2*k3 + k4)
```

Comparison of ODE methods

METHOD	ORDER	f EVALUATIONS PER STEP	When to USE
Euler	1	1	Learning, prototype
RK2 (Heun)	2	2	Fast rough estimates
RK4	4	4	Universal
Adaptive RK (Dormand-Prince)	4-5	6	High accuracy, variable step
Implicit Euler	1	iterations	Stiff equations

Interpolation

Problem: Construct a function passing through given points (x_i, y_i)

Polynomial interpolation:

$n+1$ points → polynomial of degree n (unique)

Lagrange interpolation

Formula:

$$P(x) = \Sigma_i\ y_i \cdot L_i(x)$$

where $L_i(x) = \prod_{j\neq i} (x - x_j)/(x_i - x_j)$

Property of L_i: equals 1 at point x_i, equals 0 at remaining nodes.

Example: Points (0, 1), (1, 3), (2, 2)

$$L_0(x) = (x-1)(x-2)/((0-1)(0-2)) = (x-1)(x-2)/2$$
$$L_1(x) = (x-0)(x-2)/((1-0)(1-2)) = x(x-2)/(-1) = -x(x-2)$$
$$L_2(x) = (x-0)(x-1)/((2-0)(2-1)) = x(x-1)/2$$

$$\begin{aligned} P(x) &= 1\cdot(x-1)(x-2)/2 + 3\cdot(-x(x-2)) + 2\cdot x(x-1)/2 \\ &= (x^2-3x+2)/2 - 3x^2 + 6x + x^2 - x \\ &= -1.5x^2 + 3.5x + 1 \end{aligned}$$

Verification: $P(0) = 1$ ✓, $P(1) = -1.5 + 3.5 + 1 = 3$ ✓, $P(2) = -6 + 7 + 1 = 2$ ✓

Problem: With large number of points polynomial oscillates strongly.
(Runge phenomenon)

Spline interpolation (cubic spline)

Idea: Between each pair of points — own cubic polynomial,
but they are "smoothly stitched" (values, first and second derivatives match)

```
y
|    ●
|   / \        ● ← nodes
|  /   ●-------/
| ●
+----------------→ x
  smooth curve, does not oscillate
```

Advantages:

- No Runge oscillations
- Smooth curve (C^2 – continuous up to 2nd derivative)
- Locality: changing one point affects only neighboring segments

✓ When to use: Practically always better than polynomial

Optimization (function minimization)

Gradient descent

Problem: Find minimum of $f(x)$, $x \in \mathbb{R}^n$

Idea: Move in direction of steepest descent = against gradient

Algorithm:

$$x_{n+1} = x_n - \alpha \cdot \nabla f(x_n)$$

where α – step (learning rate)

```
       ╭────────────────────╮
      /    ↘                 \
     |       ↘                |      Descent along "slope"
     |         ↘    ●        |      to minimum
      \          ↘          /
       ╰────────────────────╯
```

Example: $f(x, y) = x^2 + y^2$, start (3, 4), $\alpha = 0.1$

$\nabla f = (2x, 2y)$

Step 1: (3, 4) − 0.1×(6, 8) = (3−0.6, 4−0.8) = (2.4, 3.2)
Step 2: (2.4, 3.2) − 0.1×(4.8, 6.4) = (1.92, 2.56)
...
Converges to (0, 0)

Choice of α:

- Too large – diverges, "jumps" over minimum
- Too small – converges very slowly
- Adaptive methods (Adam, RMSprop) – automatically adjust α

✓ When to use: Smooth functions, machine learning
✗ When not to use: Many local minima, discontinuous functions

Newton's method for optimization

Idea: Approximate f by quadratic function, find its minimum.

Formula:

$$x_{n+1} = x_n - H^{-1}(x_n) \cdot \nabla f(x_n)$$

where H — Hessian matrix (second derivatives)

Convergence: Quadratic (very fast)

✓ When to use: Function is smooth, dimension is small
✗ When not to use:
- Large dimension (H^{-1} expensive to compute)
- H is degenerate or negative definite (saddle points)

Quasi-Newton methods (BFGS, L-BFGS):
Approximate H^{-1} without explicit computation — compromise between convergence speed and iteration cost.

Fast Fourier Transform (FFT)

Problem and idea

Problem: Compute discrete Fourier transform (DFT):

$$X_k = \Sigma^{n-1}_{j=0}\ x_j \cdot e^{-2\pi ijk/n}$$

Direct calculation: $O(n^2)$ operations
FFT (Cooley-Tukey): $O(n \log n)$ operations.

Speedup: For n = 1024: direct = 1 million operations
FFT = 10 thousand operations (100 times faster)

Idea: Divide and conquer. Split into even and odd indices, recursively compute, combine in $O(n)$.

Applications:
- Signal and audio processing
- Image compression (JPEG)
- Spectral analysis
- Fast polynomial and large number multiplication
- Solving PDEs by spectral methods

Code (recursive, for understanding):

```
def fft(x):
    n = len(x)
    if n == 1: return x
    even = fft(x[0::2])   # even indices
    odd = fft(x[1::2])    # odd indices
    W = [exp(-2j*pi*k/n) for k in range(n//2)]
    return [even[k] + W[k]*odd[k] for k in range(n//2)] + \
           [even[k] - W[k]*odd[k] for k in range(n//2)]
```

```
===============================================================================
Quick reference: which method to choose
===============================================================================

 Solving equations f(x) = 0
+-------------------------------+-------------------------------------+
| SITUATION                     | RECOMMENDATION                      |
+-------------------------------+-------------------------------------+
| Have interval with sign change| Bisection (reliable) + Newton (accurate) |
| Have good initial approximation| Newton                              |
| No derivative available       | Secant or Brent                     |
| System of nonlinear equations | Multidimensional Newton             |
+-------------------------------+-------------------------------------+

 Numerical integration
+-----------------------+---------------------------------------+
| SITUATION             | RECOMMENDATION                        |
+-----------------------+---------------------------------------+
| Smooth function       | Simpson or Gauss                      |
| Need adaptivity       | Adaptive Simpson (refine where needed)|
| Multidimensional integral | Monte Carlo                       |
| Oscillating function  | Special methods (Filon)               |
+-----------------------+---------------------------------------+

 Linear systems Ax = b
+------------------------------+------------------------------------+
| SITUATION                    | RECOMMENDATION                     |
+------------------------------+------------------------------------+
| Small dense matrix           | LU decomposition                   |
| Many systems with same A     | LU once, then fast solutions       |
| Symmetric positive definite  | Cholesky (A = LLᵀ)                 |
| Large sparse                 | Iterative (CG, GMRES)              |
| Very large (>10⁶)         | Multigrid methods                  |
+------------------------------+------------------------------------+

 Ordinary differential equations
+--------------------------------+--------------------------------+
| SITUATION                      | RECOMMENDATION                 |
+--------------------------------+--------------------------------+
| Ordinary problem, need accuracy| RK4 or adaptive RK (dopri5)    |
| Stiff equation                 | Implicit methods (BDF, Radau)  |
| Energy conservation (Hamiltonian)| Symplectic methods (Verlet) |
| Long-term integration          | Multistep (Adams-Bashforth)    |
+--------------------------------+--------------------------------+
```

Optimization

SITUATION	RECOMMENDATION
Smooth, small dimension	BFGS
Large dimension (ML)	Adam, SGD
Constraints (inequalities)	Interior point, SQP
Global optimum	Genetic algorithms, annealing
Convex function	Any gradient – will converge to global

PART VI: REFERENCE

Dictionary: programming ↔ mathematics

Basic programming concepts in the language of set theory.

Basic concepts

Variable

Programming: `x = 5`

Mathematics:
  Variable = name (identifier) for an element of some set
  Formally: function from name space to set of values
  Var: Names → Values, Var("x") = 5 ∈ ℤ

Key difference:
  In mathematics: x = 5 → assertion (x always equals 5)
  In programming: x = 5 → assignment (put 5 in box "x")

Assignment

Programming: `x = 5;  x = x + 1  # now x = 6`

Mathematics:
  Assignment = updating mapping (sequence of states)
  $\text{Var}_0(\text{"x"}) = 5$
  $\text{Var}_1(\text{"x"}) = \text{Var}_0(\text{"x"}) + 1 = 6$

NOT mathematical equality.
  x = x + 1 (in programming) ≠ x = x + 1 (in mathematics — contradiction)

Type

Programming: `x: int = 5;  y: float = 3.14;  s: str = "text"`

Mathematics:
  Type = set of possible values

  int ≈ ℤ (integers)
  float ≈ ℝ (real numbers, with precision limitations)
  str = set of all finite sequences of characters
  bool = {True, False} ≈ {⊤, ⊥}

  Typing = restriction of domain of values: x: int means x ∈ ℤ

Logical operations

Condition (if)

Programming:

```
if condition:
    action_1
else:
    action_2
```

Mathematics (piecewise function):

$$f(x) = \begin{cases} g_1(x), & \text{if } P(x) \\ g_2(x), & \text{if } \neg P(x) \end{cases}$$

where $P(x)$ – predicate (logical condition)

Logical operators

Programming	Mathematics (logic)	Geometry (sets)
and	$\wedge$ (conjunction)	$\cap$ (intersection)
or	$\vee$ (disjunction)	$\cup$ (union)
not	$\neg$ (negation)	^{c} (complement)

Flow control

for loop

Programming: `for x in S: f(x)`

Mathematics:

Application of function f to each element of set S

$\forall x \in S: f(x)$

If collecting results: $\bigcup_{x \in S} \{f(x)\}$

---

while loop

Programming: `while condition: action`

Mathematics:

Iteration until stopping condition is reached

$x_{n+1} = f(x_n)$ while $P(x_n) = \top$

Stopping when $P(x_n) = \bot$

Function

Programming:

```
def f(x, y):
    return x + y
```

Mathematics:

$f: X \times Y \to Z, \quad f(x, y) = x + y$

Key:

- Function in programming can have side effects
- Mathematical function — pure mapping

Pure function (pure): `def pure(x): return x * 2`

Impure (impure): `def impure(x): global counter; counter += 1`

Recursion

Programming:

```
def factorial(n):
    if n == 0: return 1
    else: return n * factorial(n-1)
```

Mathematics (recurrence definition):

$f: \mathbb{N} \to \mathbb{N}$

$f(0) = 1$ (base)

$f(n) = n \cdot f(n-1)$ (recurrence step)

Data structures

List (list)

Programming: `L = [1, 2, 3]`

Mathematics:

Finite sequence = function from $\{0, 1, ..., n-1\}$ to set

$L: \{0, 1, 2\} \to \mathbb{Z}, \quad L(0)=1, L(1)=2, L(2)=3$

Ordered tuple: $(1, 2, 3) \in \mathbb{Z}^3$

Set (set)

Programming: `S = {1, 2, 3}`

Mathematics:

Set = collection without order and repetitions
$S = \{1, 2, 3\} \subset \mathbb{Z}$

Operations: $S \cup T$ (union), $S \cap T$ (intersection), $S \setminus T$ (difference)

Dictionary (dict)

Programming: `d = {'a': 1, 'b': 2}`

Mathematics:

Dictionary = partial function from Keys to Values
$d: K \rightharpoonup V$, $d('a') = 1$, $d('b') = 2$

As set of pairs: $d \subset K \times V$

Class and object

Programming:

```
  class Point:
      def __init__(self, x, y):
          self.x = x
          self.y = y
  p = Point(3, 4)
```

Mathematics:

Class = algebraic structure (set + operations + axioms)
Point = ($\mathbb{R} \times \mathbb{R}$, {distance, move, ...})
Object = element of this set: $p = (3, 4) \in \mathbb{R}^2$

Inheritance = nesting of structures (substructure inherits operations)

Functional programming

Lambda function

Programming: `f = lambda x: x**2`
Mathematics: $f = \lambda x.\ x^2$ (λ-calculus)

```
--------------------------------------------------------------------------------

 Map
 ---
 Programming:  map(f, [1,2,3]) → [f(1), f(2), f(3)]
 Mathematics:        (f(x₁), f(x₂), f(x₃)) — pointwise application of function

--------------------------------------------------------------------------------

 Filter
 ------
 Programming:  filter(P, L) — elements for which P(x) = True
 Mathematics:        {x ∈ L : P(x)} — selection of subset by predicate

--------------------------------------------------------------------------------

 Reduce (fold)
 -------------
 Programming:  reduce(⊕, [a,b,c], init) → ((init ⊕ a) ⊕ b) ⊕ c
 Mathematics:        Fold: composition of binary operation over list
                    reduce(+, [1,2,3], 0) = 0+1+2+3 = 6

--------------------------------------------------------------------------------

 List comprehension
 ------------------
 Programming:  [f(x) for x in S if P(x)]
 Mathematics:        {f(x) : x ∈ S, P(x)} — set definition

--------------------------------------------------------------------------------

 Function composition
 ------------------
 Programming:  compose(f, g)(x) = f(g(x))
 Mathematics:        (f ∘ g)(x) = f(g(x))

--------------------------------------------------------------------------------

 Decorator
 ---------
 Programming:
   @decorator
   def f(x): ...

 Mathematics:
   Higher-order function: decorator: (X → Y) → (X → Y)
   f̃ = decorator(f)

   Geometrically — wrapper:
        +-----------+
        | decorator |
```

```
        | +-------+ |
    x --|→ |   f   | -|→ y
        |  +-------+  |
        +-------------+
```

Special concepts

Global state

Programming:

```
counter = 0
def increment(): global counter; counter += 1
```

Mathematics:

Function with side effect: f: X × Env → Y × Env

Input + state → output + new state

State monad: State s a = s → (a, s)

Iterator

Programming:

```
it = iter([1, 2, 3])
next(it)  # 1
next(it)  # 2
```

Mathematics:

Iterator = pair (state, transition function)

Iterator = (S, next: S → S × V ∪ {Stop})

Automaton: s_0 -next→ (v_1, s_1) -next→ (v_2, s_2) → ...

Lazy evaluation

Programming:

```
g = (x**2 for x in range(10**9))  # NOT computed immediately.
next(g)  # computed on demand
```

Mathematics:

Thunk = () → Value — delayed computation

Instead of value we store function that will compute it

Infinite sequences:

```
def fibonacci():
    a, b = 0, 1
    while True:
        yield a
```

```
        a, b = b, a+b

  Mathematically: fib: ℕ → ℕ (function on all naturals)
  But stored as computational process
```

None / null

Programming: x = None

Mathematics: Maybe T = T ∪ {Nothing}

x: Maybe Int → x = Nothing or x = Just(5)

Summary correspondence table

Programming	Mathematics	Geometry
x = 5	Var("x") = 5 ∈ ℤ	Point at coordinate 5
if P: A else: B	{A, if P; B, if ¬P}	Branching of trajectory
for x in S: f(x)	∀x∈S: f(x)	Traversal of set points
while P: f()	$x_{n+1}=f(x_n)$ while $P(x_n)$	Movement to boundary
def f(x): return y	f: X → Y	Mapping
$[x_1, x_2, x_3]$	(x_1,x_2,x_3) – tuple	Discrete curve
$\{x_1, x_2, x_3\}$	$\{x_1,x_2,x_3\}$ – set	Set of points
{'k': v}	k ↦ v – partial function	Discrete mapping
class C	(S, F) – algebraic structure	Manifold with operations
object	element S	Point in space
map(f, L)	$(f(x_1),...,f(x_n))$	Pointwise transformation
filter(P, L)	{x∈L : P(x)}	Region selection
reduce(⊕, L)	$x_1⊕x_2⊕...⊕x_n$	Composition of operations
lambda x: e	λx.e	Anonymous function
[f(x) for x in S if P(x)]	{f(x) : x∈S, P(x)}	Definition via condition

Programming paradigms

SQL (declarative): Relational algebra (operations over sets)
Python (imperative): Step-by-step commands (algorithm)
Haskell (functional): Function composition (λ-calculus)
Prolog (logical): Inference rules (predicate logic)

Python libraries ↔ mathematics:

NumPy array: Vector/matrix (elements $\mathbb{R}^n$, $\mathbb{R}^{n\times m}$)
Pandas DataFrame: Relation (table from relational algebra)
sklearn.Model: Statistical model (mapping X → Y with parameters)
NetworkX Graph: Graph G = (V, E) (set of vertices + edges)

Dictionary: theory of random processes ↔ time series analysis

Bridge between rigorous theory (measure theory, stochastic analysis) and practice (econometrics, data analysis, forecasting).

Basic objects

Theory of random processes	Time series analysis
Random process $\{X(t), t \in T\}$ Mapping $T \times \Omega \to \mathbb{R}$	Time series $\{x_1, x_2, ..., x_n\}$ Sequence of observations
Trajectory (realization) $X(\cdot, \omega)$ for fixed ω	Observed series Concrete sequence of numbers
Probability space $(\Omega,\mathcal{F},P)$	DGP (Data Generating Process) Population
Filtration $\{\mathcal{F}_t\}$ σ-algebras of the past	Information set History of process up to time t

Properties of processes

Theory	Practice
Strict stationarity $F(x_{t_1},...,x_{t_n}) = F(x_{t_1+h},...)$	(no direct analog) Too strong requirement
Wide-sense stationarity (weak) $E[X(t)] = \mu = const$ $Cov(X(t),X(s)) = R(\lvert t-s \rvert)$	(Weak) stationarity Constant mean and variance, ACF depends only on lag $\mu_t = \mu, \sigma_t^2 = \sigma^2, \rho(\tau)$
Ergodicity $\lim(T\to\infty)\ 1/T \int X(t)dt = E[X]$	Time average = expectation $(1/n)\Sigma x_i \to \mu$ as $n \to \infty$ Can estimate from single realization

Characteristics of processes

Theory	Practice

Expectation	Mean level / trend
$m(t) = E[X(t)]$	$\mu = (1/n)\Sigma x_i$
For stationary: $m = const$	$\hat{\mu}$ – sample mean
Autocovariance function	ACVF (autocovariance function)
$\gamma(s,t) = Cov(X(s),X(t))$	$\gamma(k) = Cov(x_t, x_{t+k})$
For stationary:	$\hat{\gamma}(k) = (1/n)\Sigma(x_i-\bar{x})(x_{i+k}-\bar{x})$
$\gamma(\tau) = Cov(X(t),X(t+\tau))$	
Autocorrelation function	ACF (autocorrelation function)
$\rho(\tau) = \gamma(\tau)/\gamma(0)$	$\rho(k) = \gamma(k)/\gamma(0)$
Normalized covariance	$\hat{\rho}(k)$ – sample ACF
Partial autocorrelation	PACF
φ_{kk} from Yule–Walker equations	For determining AR model order
Spectral density	Power spectrum / periodogram
$S(\omega) = (1/2\pi)\Sigma\gamma(k)e^{(-ik\omega)}$	$I(\omega) = \lvert(1/\sqrt{n})\Sigma x_t e^{(-i\omega t)}\rvert^2$
Wiener-Khinchin theorem:	Connection of ACF and spectrum via FFT
$S(\omega) \leftrightarrow \gamma(\tau)$ through Fourier	

Types of processes and models

Theory	Practice
White noise	White noise / Innovations
$\{\xi_t\} \sim WN(0, \sigma^2)$	$\varepsilon_t \sim iid(0, \sigma^2)$
$E[\xi_t] = 0, E[\xi_t\xi_s] = \sigma^2\delta_{ts}$	Uncorrelated, $E[\varepsilon_t] = 0$
Gaussian process	Normally distributed series
All finite-dimensional	Often assumed for
distributions Gaussian	statistical inference
Markov process	AR(1) process
$P(X_{t+1} \mid X_t,X_{t-1},...) = P(X_{t+1} \mid X_t)$	$x_t = \varphi x_{t-1} + \varepsilon_t$
Future depends only on	First-order autoregression
present	

Wiener process	Random walk
(Brownian motion)	in continuous time
$W(t) \sim N(0,t)$, continuous,	Limit of $\Delta x_t = \varepsilon_t$ as $\Delta t \to 0$
non-differentiable	I(1) process
Process with independent	Random walk
increments	$x_t = x_{t-1} + \varepsilon_t$
$X(t)-X(s) \perp X(s)-X(r)$	Discrete analog of Wiener
Ornstein–Uhlenbeck process	AR(1) in continuous time
$dX_t = -\theta(X_t-\mu)dt + \sigma dW_t$	Mean-reverting process
Mean reversion	
Martingale	Unpredictable process
$E[X_{t+1}\mid\mathcal{F}_t] = X_t$	$E[\varepsilon_{t+1}\mid\text{past}] = 0$
Best prediction = current	Correctly fitted model
value	yields martingale residuals
Long memory process	ARFIMA / Long memory process
$\gamma(k) \sim Ck\hat{}(-\alpha)$, $\alpha \in (0,1)$	Slowly decaying ACF
	Fractionally integrated process

Time series models

Stochastic DE	Discrete model
Linear SDE	ARMA(p,q) model
$dX_t = a\cdot X_t dt + \sigma dW_t$	$\varphi(L)x_t = \theta(L)\varepsilon_t$
	$\varphi(L) = 1 - \varphi_1 L - \varphi_2 L^2 - \cdots - \varphi_p L^p$
	$\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta\psi L\varphi$
Ornstein–Uhlenbeck process	AR(1): $x_t = \varphi x_{t-1} + \varepsilon_t$
$dX_t = -\theta(X_t-\mu)dt + \sigma dW_t$	where $\varphi = e\hat{}(-\theta\Delta t)$
Wiener process	Random walk
$W(t) - W(s) \sim N(0, t-s)$	$x_t = x_{t-1} + \varepsilon_t$
	Integrated process I(1)
Geometric Brownian	Logarithmic random
motion	walk
$dS_t = \mu S_t dt + \sigma S_t dW_t$	$\ln(x_t) = \ln(x_{t-1}) + \mu + \sigma\varepsilon_t$
$S(t) = S(0)\exp((\mu-\sigma^2/2)t+\sigma W(t))$	Asset price model

Operators and estimation

Theory | practice

Shift operator $U_h f(t) = f(t+h)$ Shift group	Lag operator $Lx_t = x_{t-1}$, $L^2 x_t = x_{t-2}$ $L^k x_t = x_{t-k}$
Stochastic integral $\int_0^t f(s)dW(s)$ (Itô definition)	Accumulated sum of shocks $\Sigma_{i=1}^t \varepsilon_i$
Conditional expectation $E[X_{t+h}\|\mathcal{F}_t]$	Model forecast $\hat{x}_{t+h\|t} = E[x_{t+h}\|x_1,...,x_t]$
Maximum likelihood estimation	MLE (Maximum Likelihood) $\hat{\theta} = \text{argmax } L(\theta\|\text{data})$
Asymptotic normality	$\sqrt{n}(\hat{\theta}-\theta) \xrightarrow{d} N(0,V)$ Confidence intervals

Key formulas

AR(1): $X_t = \varphi X_{t-1} + \varepsilon_t$
- Stationarity: $|\varphi| < 1$
- Variance: $\text{Var}[X_t] = \sigma^2/(1-\varphi^2)$
- ACF: $\rho(k) = \varphi^k$
- Spectral density: $S(\omega) = \sigma^2/|1-\varphi e^{(-i\omega)}|^2$

AR(p): $X_t = \varphi_1 X_{t-1} + ... + \varphi_p X_{t-p} + \varepsilon_t$
- Stationarity: roots of $\varphi(z) = 0$ outside unit circle
- Yule–Walker equations: $\gamma(k) = \varphi_1\gamma(k-1) + ... + \varphi_p\gamma(k-p)$
- PACF: $\varphi_{kk} \neq 0$ for $k \leq p$, $\varphi_{kk} = 0$ for $k > p$

MA(q): $X_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + ... + \theta\psi\varepsilon_{t-\psi}$
- Invertibility: roots of $\theta(z) = 0$ outside unit circle
- ACF: $\rho(k) = 0$ for $k > q$ (cuts off)

ARIMA(p,d,q): $\varphi(L)\nabla^d X_t = \theta(L)\varepsilon_t$
- d times differencing for stationarity
- $\nabla x_t = x_t - x_{t-1}$ (difference operator)

Kalman filter:

State: $x_t = Ax_{t-1} + w_t$, $w_t \sim N(0,Q)$

Observ: $y_t = Cx_t + v_t$, $v_t \sim N(0,R)$

Predict: $\hat{x}_{t|t-1} = A\hat{x}_{t-1|t-1}$

Update: $K_t = P_{t|t-1}C^T(CP_{t|t-1}C^T + R)^{-1}$ (Kalman gain)

$\hat{x}_{t|t} = \hat{x}_{t|t-1} + K_t(y_t - C\hat{x}_{t|t-1})$

Key differences in approaches

Stochastic process theory:
- Questions: existence, uniqueness, structure
- Language: measure theory, functional analysis, topology
- Goal: prove general theorems
- Audience: mathematicians
- Emphasis: rigor, generality, abstraction

Time series analysis:
- Questions: identification, estimation, forecasting
- Language: statistics, regression analysis
- Goal: obtain numerical forecast
- Audience: statisticians, econometricians, engineers
- Emphasis: practicality, applicability, computation

Connection:

Time series analysis = discretization + estimation + diagnostics
of stochastic process theory

Process theory explains Why time series methods work.
Time series shows how to apply these methods in practice.

Dictionary: continuous → discrete (How mathematics gets into the computer)

Why this section

Mathematics operates with continuous objects: ℝ, integrals, derivatives.
Computer works with discrete objects: arrays, sums, differences.

Engineer must understand how one transforms into the other — and what is lost.

Main table: analog → digital

CONTINUOUS	DISCRETE	WHAT IS LOST
Function $f(x)$ $x \in [a,b]$	Array f[i] i = 0,1,...,N	Values between nodes (need interpolation)
Derivative df/dx	Difference (f[i+1]−f[i])/h or (f[i+1]−f[i-1])/2h	Accuracy $O(h)$ Accuracy $O(h^2)$
Second derivative d^2f/dx^2	$(f[i+1]-2f[i]+f[i-1])/h^2$	Accuracy $O(h^2)$
Integral $\int f(x)dx$	Sum $\Sigma f[i]\cdot h$ Trapezoids: $\Sigma(f[i]+f[i+1])h/2$ Simpson	Rectangles: $O(h)$ Trapezoids: $O(h^2)$ Simpson: $O(h^4)$
$\mathbb{R}$ (real numbers)	float64	Precision ~15 digits Overflow, underflow
Operator L: V→V (infinite-dim)	Matrix A: $\mathbb{R}^n \to \mathbb{R}^n$	Finite dimension

Kotelnikov-Shannon-Nyquist theorem — When nothing is lost

Fundamental result: a continuous signal can be exactly reconstructed from discrete samples — but only under certain conditions.

Theorem:
If signal $f(t)$ contains no frequencies above F_max (band-limited), then it is completely determined by samples with frequency $\geq 2\cdot F\_max$.

$$f(t) = \Sigma\, f(n\cdot T) \cdot \mathrm{sinc}((t - n\cdot T)/T), \quad \text{where } T \leq 1/(2\cdot F\_max)$$

$$\mathrm{sinc}(x) = \sin(\pi x)/(\pi x)$$ — sampling function (cardinal sine)

Corollaries:
- Nyquist frequency: F_Nyquist = 2·F_max — minimum sampling frequency
- Aliasing: if sampling is rarer than F_Nyquist, high frequencies "pretend" to be low (stroboscopic effect — wheel spins "backward" in movies)
- Anti-aliasing filter needed before discretization

Practice:
- CD Audio: F_max = 20 kHz → sampling 44.1 kHz > 40 kHz ✓
- When calculating structures: mesh step $h < \lambda\_min/2$ (wavelength)

This is the bridge between continuous and discrete — theoretical justification that digital signal processing is even possible without information loss.

Derivative → matrix

```
Operator d/dx on functions f(x) — infinite-dimensional.
But if f(x) is given at N points: f[0], f[1], ..., f[N-1], then:

First derivative (central difference):

  (df/dx)[i] ≈ (f[i+1] − f[i-1]) / (2h)

  Matrix D₁:
  +                              +
  |  0   1   0   0   0   ⋯  0 |
  | -1   0   1   0   0   ⋯  0 |   1
  |  0  -1   0   1   0   ⋯  0 |  ─── ·
  |  0   0  -1   0   1   ⋯  0 |  2h
  |  ⋮   ⋮   ⋮   ⋮   ⋱   ⋮  |
  |  0   0   0   ⋯  -1   0  1 |
  |  0   0   0   ⋯   0  -1  0 |
  +                              +

Second derivative (heat equation):

  (d²f/dx²)[i] ≈ (f[i+1] − 2f[i] + f[i-1]) / h²

  Matrix D₂:
  +                              +
  | -2   1   0   0   0   ⋯  0 |
  |  1  -2   1   0   0   ⋯  0 |   1
  |  0   1  -2   1   0   ⋯  0 |  ─── ·
  |  0   0   1  -2   1   ⋯  0 |  h²
  |  ⋮   ⋮   ⋮   ⋮   ⋱   ⋮  |
  |  0   0   0   ⋯   1  -2  1 |
  |  0   0   0   ⋯   0   1 -2 |
  +                              +

This is the same matrix as in the example (thermal balance of rooms)!
Heat conduction → matrix → eigenvalues → solution.


Graph Laplacian = discrete Laplace operator


On continuous space: Δf = d²f/dx² + d²f/dy² + d²f/dz²

On a graph (e.g., heat network):

  (Lf)[i] = Σⱼ wᵢⱼ (f[i] − f[j])

  where wᵢⱼ — edge weight between nodes i and j (pipe thermal conductivity)

Laplacian matrix:

  L = D − W

  D = diag(d₁, d₂, ..., dₙ) — vertex degrees (dᵢ = Σⱼwᵢⱼ)
  W = weight matrix (wᵢⱼ)

Eigenvalues of Laplacian:
```

- $\lambda_1 = 0$ always (for connected graph — unique zero)
- $\lambda_2 > 0 \iff$ graph is connected
- The larger λ_2, the "better connected" the graph

Application: Spectral clustering, network analysis, PageRank

Finite element method (fem) — briefly

Problem: Solve heat equation on complex domain Ω

$-\nabla\cdot(k\nabla T) = q$ (in domain Ω)
$T = T_0$ (on boundary $\partial\Omega$)

Step 1: Partition domain into elements (triangles, tetrahedra)

```
        ●-------●-------●
       /\      /\      /\
      /  \    /  \    /  \
     /    \  /    \  /    \
    ●------●------●------●
     \    /\    /\    /
      \  /  \  /  \  /
       \/    \/    \/
        ●-------●-------●
```

Step 2: Inside each element $T(x,y) \approx$ linear function

Step 3: Write variational form:

Minimize: $\int_\Omega [\frac{1}{2}k|\nabla T|^2 - qT]\, dA$

Step 4: Obtain system of linear equations:

$K\cdot T = F$

K — stiffness matrix (depends on geometry and k)
F — load vector (depends on q and boundary conditions)
T — vector of temperatures at nodes

Result: Differential equation → Matrix equation
Infinite-dimensional problem → Finite-dimensional linear algebra

Pitfalls of discretization

Problem 1: Loss of precision in float

In $\mathbb{R}$: $(a + b) + c = a + (b + c)$ — associativity
In float: may be false.

Example: $a = 1.0$, $b = 1e\text{-}16$, $c = 1e\text{-}16$
$(a + b) + c = 1.0 + 1e\text{-}16 = 1.0$ (b is lost)
$a + (b + c) = 1.0 + 2e\text{-}16 \approx 1.0$ (but slightly larger)

Problem 2: Instability of schemes

Explicit scheme for heat conduction:
$T[i,n+1] = T[i,n] + (a\Delta t/h^2)(T[i+1,n] - 2T[i,n] + T[i-1,n])$

Stable only when: $a\Delta t/h^2 \le 0.5$ (Courant condition)
Otherwise — oscillations and solution blow-up.

Problem 3: Poor conditioning of matrix

$\mathrm{cond}(A) = \|A\|\cdot\|A^{-1}\|$ — condition number
If $\mathrm{cond}(A) \gg 1$, small errors in data → large errors in solution

When decreasing h: $\mathrm{cond}(D_2) \sim 1/h^2$ → bad.

Z-transform — Laplace for discrete systems

Definition:
$X(z) = Z[x_n] = \sum_{n=0}^{\infty} x_n z^{-n}$

Connection with Laplace:
- Laplace: $L[x(t)] = \int_0^{\infty} x(t) e^{-st} dt$ (continuous time)
- Z: $Z[x_n] = \sum x_n z^{-n}$ (discrete time)

With discretization step T: $z = e^{sT}$

Key properties:

Time	Z-domain
Delay: x_{n-1}	$z^{-1} X(z)$
Difference: $x_n - x_{n-1}$	$(1 - z^{-1}) X(z)$
Convolution: $(x*y)_n$	$X(z)\cdot Y(z)$
Initial value	$\lim_{z\to\infty} X(z) = x_0$
Final value	$\lim_{z\to 1} (1-z^{-1})X(z) = \lim x_n$

Application in DSP (digital signal processing):
- Digital filter: $Y(z) = H(z)\cdot X(z)$
- Transfer function: $H(z) = (b_0 + b_1z^{-1} + ...)/(1 + a_1z^{-1} + ...)$
- Stability: all poles of $H(z)$ inside unit circle $|z| < 1$

Analogy with Laplace:

CONCEPT	s-domain (Laplace)	z-domain
Stability	Re(s) < 0	|z| < 1
Stability boundary	imaginary axis	unit circle
Frequency response	s = iω	z = e^{iωT}

Summary: connection of sections

```
  +----------------------+      discretization     +-----------------------+
  | Calculus             |  ---------------------► | Linear algebra        |
  | (derivatives, intgr.)|                         | (matrices, systems)   |
  +----------------------+                         +-----------------------+

  +----------------------+      discretization     +-----------------------+
  | Manifolds            |  ---------------------► | Graphs                |
  | (smooth surfaces)    |        (FEM)            | (meshes, nodes, edges)|
  +----------------------+                         +-----------------------+

  +----------------------+      discretization     +-----------------------+
  | Func. analysis       |  ---------------------► | Numerical linear      |
  | (operators)          |                         | algebra               |
  +----------------------+                         +-----------------------+
```

Continuous mathematics gives understanding (why it works).
Discrete mathematics gives computation (how to calculate).

Dictionary: engineering jargon ↔ mathematical terminology

Introduction

Engineers and mathematicians often talk about the same thing in different words.
This dictionary helps to "translate" between languages.

Control theory and signals

ENGINEERING TERM	MATHEMATICAL EQUIVALENT

Signal $x(t)$	Function $x \in L^2(\mathbb{R})$ or distribution
System, "black box"	Operator $T: X \to Y$ between functional spaces
Linear system	Linear operator $T(ax+by) = aT(x)+bT(y)$
Stationary (LTI) system	Shift-invariant: $T \circ \tau_s = \tau_s \circ T$ where τ_s – shift operator by s
Impulse response $h(t)$	Convolution kernel: $y = h * x$, i.e. $y(t)=\int h(\tau)x(t-\tau)d\tau$
Transfer function $H(s)$	Symbol of operator / Laplace transform of kernel $H(s) = \mathcal{L}\{h\}(s)$, relation: $Y(s) = H(s)X(s)$
Frequency response $H(i\omega)$	Fourier transform: $H(i\omega) = \mathcal{F}\{h\}(\omega)$
Poles of transfer function	Eigenvalues of operator (roots of characteristic polynomial)
Stable system (BIBO-stability)	Spectrum of operator in left half-plane $\mathrm{Re}(s)<0$ or: all eigenvalues with $\mathrm{Re}(\lambda)<0$
Resonance	Forcing frequency close to imaginary part of eigenvalue (pole)
Feedback	Modification of operator: $T_{fb} = T/(1 + KT)$ Change of eigenvalues.
Controllability	Reachability of any point in state space $\mathrm{rank}[B, AB, A^2B, ...] = n$ (Kalman criterion)
Observability	Recoverability of state from output $\mathrm{rank}[C; CA; CA^2; \cdots] = n$

Mechanics and thermal engineering

ENGINEERING TERM	MATHEMATICAL EQUIVALENT
Distributed system (heat conduction, waves)	PDE (partial differential equation) Solution — function $u(x,t)$ of several variables
Lumped parameters (point masses, capacities)	ODE (ordinary differential equation) Solution — vector function $x(t)$
Boundary conditions	Values of function (or derivatives) on $\partial\Omega$ Trace $u\vert_{\partial\Omega}$ — element of Sobolev space
Natural frequency ω_n	$\sqrt{\lambda_n}$, where λ_n — eigenvalue of operator (Laplacian with boundary conditions)
Mode of vibration $\varphi_n(x)$	Eigenfunction: $-\nabla^2\varphi_n = \lambda_n\varphi_n$
Mode expansion	Expansion in orthogonal basis of eigenfunctions $u(x,t) = \Sigma c_n(t)\varphi_n(x)$
Damping	Dissipative term in operator Shifts eigenvalues left (Re↓)
Stress tensor σ_{ij}	Symmetric tensor of rank 2 $\sigma: T_pM \otimes T_pM \to \mathbb{R}$ (bilinear form)
Temperature gradient	1-form $dT \in \Omega^1(M)$, or vector $\nabla T = g^{-1}(dT)$
Heat flux q	Vector or 2-form (integrated over surface) $q = -k\nabla T$ (Fourier's law = k relates them)

Statistics and machine learning

ENGINEERING TERM	MATHEMATICAL EQUIVALENT
Data (sample)	Empirical measure: $\mu\_n = (1/n)\Sigma\delta_{x_i}$
Features	Coordinates in feature space $\mathbb{R}^p$
Principal components(PCA)	Eigenvectors of covariance matrix = directions of maximal variance
Clustering	Partition of space by metric (Voronoi, k-means, ...)
Kernel methods	Reproducing kernel Hilbert space $K(x,y) = \langle\varphi(x), \varphi(y)\rangle$ for embedding $\varphi: X \to H$
Regularization (L1, L2, ...)	Constraint on norm of solution $\min \|Ax-b\|^2 + \lambda\|x\|_p$ — problem in Banach space
Overfitting	Approximation of noise rather than signal Violation of bias-variance tradeoff
Gradient descent	Iterative method: $x_{k+1} = x_k - \alpha\nabla f(x_k)$ Convergence: convexity + Lipschitz condition

How to Use the Tables When Solving Problems

Step-by-step algorithm:

Step 1: determine the level (Main table)
↓
At what level of abstraction is the problem?

Step 2: determine the type of structure (pillars in the main table)
↓
Discrete? Algebraic? Topological? Analytic?

Step 3: identify relations
↓
What relations are given? $\in$, $\subseteq$, $\sim$, $\leq$, $\to$?

Step 4: search for analogy (Analogy table)
↓
Have we seen a similar pattern in another area?

Step 5: choose proof method
↓
Direct? By contradiction? Induction? Contraposition?

```
Step 6: construct path (Philosophy)
        ↓
       Proof = path of set inclusions

Step 7: Check
        ↓
       Are all inclusions correct?

Concrete example:

Problem: Prove that the sum of two even numbers is even

Step 1: Level — Numbers (ℤ), Level 4
Step 2: Type — Algebraic (operation +) + Discrete (divisibility)
Step 3: Relations — "Even" = {n ∈ ℤ : ∃k, n = 2k}
Step 4: Analogy — Closure of subgroup.
Step 5: Method — Direct proof
Step 6: Path:
        1. a = 2k, b = 2m
        2. a + b = 2k + 2m = 2(k+m)
        3. ⇒ a + b even ✓
Step 7: Check — ✓ all steps correct

-------------------------------------------------------------------------------

Second example: More complex problem

Problem: Prove that the set of irrational numbers is uncountable

Step 1: determine the level
         • Sets + cardinality → Level 2 (set theory)
         • But use facts about ℝ and ℚ → Level 4 (numbers)

Step 2: determine the type
         • Discrete (cardinalities of sets)
         • Working with "sizes" of infinite sets

Step 3: identify relations
         Given:
           • ℝ uncountable (Cantor's theorem)
           • ℚ countable (known fact)
           • Irrationals = ℝ \ ℚ

         Required:
           • |ℝ \ ℚ| uncountable

Step 4: search for analogy
         Recall arithmetic of cardinalities:
           • If A finite, B finite ⇒ A ∪ B finite
           • Analogy for infinite: countable + countable = countable
           • But: uncountable ≠ countable + countable
```

Step 5: choose method

- By contradiction (reductio ad absurdum)
- Assume that irrationals are countable → contradiction

Step 6: construct path (proof)

1. Assume that ℝ \ ℚ countable [by contradiction]
2. ℚ countable (known) [fact]
3. ℝ = ℚ ∪ (ℝ \ ℚ) [partition]
4. Countable ∪ Countable = Countable [theorem]
5. ⇒ ℝ countable [from 1,2,3,4]
6. But ℝ uncountable (Cantor's diagonal method) [contradiction]
7. ⇒ Assumption false [reductio]
8. ⇒ ℝ \ ℚ uncountable ✓

Geometry (sets):

```
  If ℝ\ℚ were countable:

  ℝ = ℚ ∪ (ℝ\ℚ)
      ↓     ↓
   countable + countable = countable  (should be)

  But ℝ uncountable. → contradiction
```

Philosophy:

We tried to represent ℝ as union of two countable sets.
Got absurdity: ℝ simultaneously countable and uncountable.
Therefore, one of the sets is uncountable. Since ℚ countable, then ℝ\ℚ uncountable.

Step 7: Check

✓ Used facts are true (ℚ countable, ℝ uncountable)
✓ Theorem about countable union is true
✓ Logic by contradiction applied correctly
✓ Contradiction is real (not imaginary)

Typical problems — from formulation to method

FORMULATION	TYPE OF PROBLEM	METHOD
"Prove x ∈ A"	Membership	Direct: show definition
"Prove A ⊆ B"	Inclusion	Take arbitrary x ∈ A
"Prove A = B"	Equality of sets	Show A ⊆ B and B ⊆ A

"Does not exist"	Nonexistence	By contradiction
"For all n ∈ ℕ"	For all natural	Induction
"P ⇒ Q"	Implication	Direct or contraposition
"P ⇔ Q"	Equivalence	Two implications
"There exists unique"	Existence + uniqueness	Construction + uniqueness separately

Historical atlas — how mathematics grew by levels

PERIOD	WHAT APPEARED (by levels of main table)
~3000 BC - 300 BC	LEVEL 2-3: ℕ, ℤ, ℚ, Euclidean geometry, Aristotelian logic • Euclid "Elements" — first axiomatic system • Aristotle — formalization of logic
1500-1700	LEVEL 4: ℂ, calculus • Complex numbers ℂ (Cardano, ~1545) • Calculus (Newton, Leibniz, ~1670-1680)
1800-1850	LEVEL 4-5: Groups, rigor in analysis • Grassmann — vector spaces (~1844) • Galois — group theory (~1830) • Weierstrass — ε-δ definitions (~1850)
1850-1900	LEVEL 2, 5: Sets, topology • Cantor — set theory (~1870-1880) [foundation]
1900-1930	LEVEL 5-6: Topology, functional analysis • Lebesgue — measure theory (~1902) • Fréchet — metric spaces (~1906) • Hausdorff — general topology (~1914) • Banach — functional analysis (~1920-1930) • Gödel — incompleteness theorems (~1931)
1930-1960	LEVEL 6: Categories, algebraic topology • Eilenberg, Mac Lane — category theory (~1945) [meta-level] • Bourbaki — structuralist approach (~1935) • Kolmogorov — axiomatic probability theory (~1933)

Key observation:

History of mathematics goes bottom-up through the main table.
Each era builds on previous levels.

Typical Misconceptions

Almost everyone makes these mistakes. Better to learn about them in advance.

Misconceptions in Logic

ERROR	TRUTH
"From false follows only false"	From false follows ANYTHING. $F \to P$ is true for any P
"∀ and ∃ can be swapped"	$\forall x \exists y\ P(x,y) \not\equiv \exists y \forall x\ P(x,y)$ "Everyone has their own mom" ≠ "One mom for all"
"To prove = to check examples"	Examples never prove ∀-statements. But one counterexample refutes

Misconceptions in Algebra

ERROR	TRUTH
"A group is numbers with an operation"	A group is a set of SYMMETRIES of an object. Numbers are just one special case
"In a group always $ab = ba$"	Only in ABELIAN groups. Rotations + reflections: NOT commutative
"$(ab)^{-1} = a^{-1}b^{-1}$"	$(ab)^{-1} = b^{-1}a^{-1}$ – order is REVERSED. $(\text{put on socks, shoes})^{-1}$ = take off shoes, take off socks

Misconceptions in Topology

ERROR	TRUTH
"Open = without boundary"	Open = "inside each point there is a ball" $\mathbb{R}$ is open in $\mathbb{R}$, although "there is no boundary"
"A set is either open or closed"	Can be both ($\varnothing$, X) Can be NEITHER: $[0,1)$ in $\mathbb{R}$
"A continuous function cannot go to infinity"	$x \mapsto 1/x$ is continuous on $(0,\infty)$ and $\to \infty$ Continuity ≠ boundedness
"Donut and cup – it's a joke"	This is an EXACT theorem: both have genus 1 (one hole), therefore homeomorphic

Misconceptions in Linear Algebra

ERROR	TRUTH
"A vector is an arrow"	A vector is an element of a vector space. Functions, matrices, series — also vectors.
"A matrix is a table of numbers"	A matrix is a REPRESENTATION of a linear operator in a specific basis. Change basis — matrix changes
"det(A+B) = det(A) + det(B)"	FALSE. det is multiplicative: det(AB) = det(A)·det(B)
"If det A = 0, the matrix doesn't exist"	The matrix exists, but it is NONINVERTIBLE. Some vectors "collapse" into the kernel

Misconceptions in Differential Geometry

ERROR	TRUTH
"dx is an infinitesimal quantity"	dx is a basis element of the cotangent space T*M. A linear function.
"∫ is just an antiderivative"	$\int_m \omega$ is an independent concept (integral of a form over a manifold). Connection via Stokes.
"Curvature is bendedness"	Intrinsic curvature is defined through parallel transport of a vector along a contour. Cylinder: K=0 (flat). Sphere: K>0.
"Connection is abstract nonsense"	Connection defines what "parallel" means in curved space. Without it one cannot compare vectors at different points.

Misconceptions in Mathematical Analysis

ERROR	TRUTH
"Derivative is velocity"	Derivative is LINEAR APPROXIMATION. Velocity is just one physical example. $f(x+h) \approx f(x) + f'(x)\cdot h$ – that's the essence.
"Integral is area"	Area is a special case. Integral is a measure: generalized "size" of a set or a functional on $C(X)$.
"∞ is a number"	∞ is a SYMBOL for writing limits. Cannot: $\infty - \infty$, ∞/∞, $0\cdot\infty$ – undefined.
"$1/0 = \infty$"	$1/0$ is NOT DEFINED in $\mathbb{R}$. $\lim(1/x) = \infty$ as $x\to 0^+$ – this is behavior, not value at a point.
"A convergent series can be rearranged"	Only ABSOLUTELY convergent. Conditionally convergent → Riemann's theorem: by rearrangement get any sum.

Misconceptions about Tensors

ERROR	TRUTH
"A tensor is a multidimensional array of numbers"	A tensor is a MULTILINEAR FUNCTION. Array is just a representation in a basis. Tensor exists independently of coordinates.
"A matrix is a rank-2 tensor"	A matrix is a representation of operator $V \to W$. A tensor of type (1,1) is $V^* \otimes V$. Matrix changes under change of basis, tensor – doesn't.
"Upper/lower indices are just notation"	These are different objects: vector (↑) vs covector (↓). Related through metric. $v^i = g^{ij}v_j$ – raising index requires g.

Misconceptions about Probability

ERROR	TRUTH
"Probability is frequency"	Frequency is one INTERPRETATION. Mathematically probability is a measure with property $P(\Omega)=1$. Bayesian probability is degree of confidence.
"$P(A\|B) = P(B\|A)$"	FALSE. $P(\text{disease}\|\text{test}^+) \neq P(\text{test}^+\|\text{disease})$ Even doctors confuse this. Use Bayes.
"Independent events cannot occur together"	Independence: $P(A \cap B) = P(A) \cdot P(B)$. They can occur together. Just knowledge of one doesn't change probability of other.
"After 10 heads chance of tails is higher"	GAMBLER'S FALLACY. Coin doesn't remember past. $P(\text{tails}) = 0.5$ on each toss.

Misconceptions about Measures and Integrals

ERROR	TRUTH
"Any set has a measure"	There exist NON-MEASURABLE sets. (Vitali example, consequence of axiom of choice)
"Riemann integral is the only integral"	Lebesgue: integrate over function values. Riemann doesn't integrate characteristic function of $\mathbb{Q}$, Lebesgue – integrates (= 0).
"A set of measure zero is finite or countable"	Cantor set: uncountable, but $\mu = 0$! Measure and cardinality are different concepts.

Main Misconception

"Branches of mathematics are different subjects"

Truth:
It's one subject, viewed from different sides.
A group is a category with one object.
Topology is a category of open sets.
A vector space is a module over a field.
A differential form is a section of a bundle.

When you understand the deep structure, boundaries between branches disappear.

Why X is not Y – Negative Examples

Understanding what an object is NOT is no less important than knowing what it is. Here are key distinctions:

Algebraic Structures

STATEMENT	REASON
$\mathbb{Z}$ is a ring, but NOT a field	No division: $1/2 \notin \mathbb{Z}$
$\mathbb{Z}/6\mathbb{Z}$ is a ring, but NOT an integral domain	Has zero divisors: $2 \cdot 3 = 0 \pmod 6$
$(\mathbb{R}{>}0, \times)$ is a group, but $(\mathbb{R}{>}0, +)$ is NOT a group	No neutral element for + ($0 \notin \mathbb{R}{>}0$)
Quaternions $\mathbb{H}$ are a division algebra, but NOT a field	$ab \neq ba$ (non-commutativity of multiplication)
SPACES and NORMS	
$L^1[0,1]$ is Banach, but NOT Hilbert	Norm $\lVert f\rVert_1 = \int \lvert f\rvert$ is not generated by inner product (parallelogram law violated)
$C[0,1]$ with $\lVert\cdot\rVert_\infty$ is normed, but NOT complete under $\lVert\cdot\rVert_1$	There is a Cauchy sequence without limit in $C[0,1]$: limit is a discontinuous function
Metric $d(x,y) = \lvert x-y\rvert/(1+\lvert x-y\rvert)$ is not generated by a norm	$d(x,y) \leq 1$, but original $\mathbb{R}$ is unbounded No linear structure at the level of metric
TOPOLOGICAL PROPERTIES	
$\mathbb{Q}$ is NOT connected	$\mathbb{Q} = (-\infty,\sqrt{2})\cap\mathbb{Q} \cup (\sqrt{2},\infty)\cap\mathbb{Q}$ – two open sets
$[0,1)$ is NOT compact (unlike $[0,1]$)	Cover $(1/n, 1)$ has no finite subcover
Cylinder is NOT homeomorphic to plane	$\pi_1(\text{cylinder}) = \mathbb{Z} \neq 0 = \pi_1(\text{plane})$ (although both are locally Euclidean)
FUNCTIONAL ANALYSIS	
Weak convergence does NOT imply norm convergence	$e_n \to 0$ weakly in L^2, but $\lVert e_n\rVert = 1$ (orthonormal basis)
Compactness in ∞-dimensional is NOT equivalent to closed + bounded	Closed unit ball in L^2 is closed and bounded, but NOT compact.
A continuous linear operator does NOT necessarily have eigenvalues	Shift in ℓ^2: $T(x_1,x_2,...) = (0,x_1,x_2,...)$ Injective, but $Tx = \lambda x \Rightarrow x = 0$

Notation Reference

All symbols used in this standard:

Note: some symbols appear in different sections with different meanings — context always disambiguates which meaning is intended:

- ≅ – isomorphism (in algebra) / homeomorphism (in topology)
- ∧ – exterior product (in algebra) / conjunction (in logic)
- → – implication (in logic) / mapping (in set theory)
- ∂ – boundary of a set (in topology) / partial derivative (in analysis)
- ⊥ – orthogonality (in lin. algebra) / perpendicularity (in geometry)

Logical symbols

SYMBOL	NAME	MEANING
∀	Universal quantifier	"for all", "for any"
∃	Existential quant.	"there exists", "there is at least one"
∃!	Uniqueness	"there exists a unique"
¬	Negation	"not", "it is false that"
∧	Conjunction	"and", "simultaneously"
∨	Disjunction	"or" (inclusive)
→	Implication	"if…, then…", "implies"
↔	Equivalence	"if and only if", "equivalent"
⊢	Derivability	"right follows from left"
⊨	Semantic conseq.	"right semantically follows from left"

Set-theoretic symbols

SYMBOL	NAME	MEANING
∈	Membership	x ∈ A: "x is an element of A"
∉	Non-membership	x ∉ A: "x is not an element of A"
⊂, ⊆	Subset	A ⊂ B: "every element of A is element of B"
⊃, ⊇	Superset	A ⊃ B: equivalent to B ⊂ A
∪	Union	A ∪ B = {x : x∈A or x∈B}
∩	Intersection	A ∩ B = {x : x∈A and x∈B}
\	Difference	A \ B = {x : x∈A and x∉B}
∅	Empty set	Set with no elements
℘(A)	Boolean	Set of all subsets of A
|A|	Cardinality	Number of elements (for finite)
×	Cartesian product	A×B = {(a,b) : a∈A, b∈B}

Mapping symbols

SYMBOL	NAME	MEANING
$\to$	Mapping	$f: X \to Y$: "f from X to Y"
$\mapsto$	Mapping rule	$x \mapsto f(x)$: "x maps to f(x)"
$\circ$	Composition	$(g\circ f)(x) = g(f(x))$
f^{-1}	Inverse / Preimage	Inverse mapping or preimage of a set
id	Identity	$\mathrm{id}(x) = x$
Im, im	Image	$\mathrm{Im}\, f = f(X) = \{f(x) : x\in X\}$
ker	Kernel	$\ker f = \{x : f(x) = e\}$ (e – neutral)
Hom	Set of morphisms	Hom(A,B) = {morphisms from A to B}

Number sets

SYMBOL	NAME	ELEMENTS
$\mathbb{N}$	Natural numbers	$\{0, 1, 2, 3, ...\}$ or $\{1, 2, 3, ...\}$
$\mathbb{Z}$	Integers	$\{..., -2, -1, 0, 1, 2, ...\}$
$\mathbb{Q}$	Rational numbers	$\{p/q : p\in\mathbb{Z}, q\in\mathbb{Z}, q\neq 0\}$
$\mathbb{R}$	Real numbers	Number line (including irrationals)
$\mathbb{C}$	Complex numbers	$\{a + bi : a,b\in\mathbb{R}, i^2 = -1\}$
$\mathbb{H}$	Quaternions	4-dimensional extension of $\mathbb{C}$ (noncomm.)*
$\mathbb{O}$	Octonions	8-dimensional extension of $\mathbb{H}$ (nonassoc.)
*$\mathbb{H}$: $x^2+1=0$ has ∞ roots (±i, ±j, ±k and all their combinations on S^2)		
$\mathbb{R}^n$	n-dimensional space	Set of n-tuples $(x_1, ..., x_n)$, $x_i\in\mathbb{R}$
S^1	Unit circle	$\{z \in \mathbb{C} : \|z\| = 1\} = \{e^{i\theta} : \theta \in [0,2\pi)\}$
S^n	n-dimensional sphere	$\{x \in \mathbb{R}^{n+1} : \|x\| = 1\}$. S^2 – ordinary sphere
T^n	n-dimensional torus	$S^1 \times S^1 \times ... \times S^1$ (n times). T^2 – donut
$\mathbb{R}P^n$	Projective space	Lines through 0 in $\mathbb{R}^{n+1}$. $\mathbb{R}P^1 \cong S^1$

Algebraic symbols

SYMBOL	NAME	MEANING
$\cdot$, $*$	Group operation	Binary operation (multiplication)
e, 1	Neutral element	$e\cdot g = g\cdot e = g$ for all g
g^{-1}	Inverse element	$g\cdot g^{-1} = g^{-1}\cdot g = e$
$\leq$, $<$	Subgroup	$H \leq G$: "H is a subgroup of G"
$\lhd$, $\trianglelefteq$	Normal subgroup	$H \lhd G$: "H is a normal subgroup of G"
G/H	Quotient group	Set of cosets gH
$\cong$	Isomorphism	$G \cong H$: "G is isomorphic to H"
$\oplus$	Direct sum	$V \oplus W$ (for vector spaces)
$\otimes$	Tensor product	$V \otimes W$ (tensor product)
$\wedge$	Exterior product	$v \wedge w$ (antisymmetric)

Topological symbols

SYMBOL	NAME	MEANING
τ	Topology	Family of open sets
$B(x,\varepsilon)$	Open ball	$\{y : d(x,y) < \varepsilon\}$
$\bar{A}$	Closure	Smallest closed set containing A
int(A)	Interior	Largest open set contained in A
∂A	Boundary	$\bar{A} \setminus \mathrm{int}(A)$
$\cong$	Homeomorphism	Topological equivalence
$\simeq$	Homotopy equiv.	Homotopy equivalence
π_1	Fundamental group	$\pi_1(X)$ – group of loops in X
H_n	Homology group	n-dimensional "holes" in space

Linear algebra symbols

SYMBOL	NAME	MEANING
dim	Dimension	dim V = number of basis vectors
rank	Rank	rank A = dim Im A
det	Determinant	$\det A \in F$ (for square matrices)
tr	Trace	$\mathrm{tr}\, A = \Sigma\, a_{ii}$ (sum of diagonal)
$\langle\cdot,\cdot\rangle$	Scalar product	$\langle u,v\rangle \in F$ (bilinear form)
$\\|\cdot\\|$	Norm	$\\|v\\| = \sqrt{\langle v,v\rangle}$
$\perp$	Orthogonality	$u \perp v \iff \langle u,v\rangle = 0$
V^*	Dual space	$V^* = \mathrm{Hom}(V, F)$ – linear functionals
A^T	Transpose	$(A^T)_{ij} = A_{ji}$
$A^\dagger$, A^*	Hermitian conjugate	$(A^\dagger)_{ij} = \bar{A}_{ji}$

Differential geometry symbols

SYMBOL	NAME	MEANING
T_pM	Tangent space	Vector space at point p
T^*_pM	Cotangent space	Dual to T_pM
d	Exterior derivative	$d: \Omega^k \to \Omega^{k+1}$, $d^2 = 0$
$\Omega^k(M)$	k-forms	Antisymmetric (0,k)-tensors
∇	Connection / Nabla	Covariant derivative
$\partial/\partial x^i$	Coordinate basis	Basis of tangent space
dx^i	Cobasis	Basis of cotangent space
g_{ij}	Metric tensor	$\langle \partial/\partial x^i, \partial/\partial x^j\rangle$
Γ^i_{jk}	Christoffel symbols	Connection coefficients
R^i_{jkl}	Riemann tensor	Measure of space curvature

Category theory symbols

SYMBOL	NAME	MEANING
Ob(C)	Objects	Class of objects of category C
Mor(C)	Morphisms	Class of morphisms of category C
Hom(A,B)	Hom-set	Morphisms from A to B
$F: C\to D$	Functor	Mapping between categories
$\eta: F\Rightarrow G$	Natural transf.	Family of morphisms $\eta_a: F(A)\to G(A)$
$\dashv$	Adjunction	$F \dashv G$: F left adjoint to G
lim, ←	Limit	Universal cone over diagram
colim,→	Colimit	Universal cocone under diagram

Standard categories

SYMBOL	NAME	OBJECTS / MORPHISMS
Set	Category of sets	Sets / Functions
Grp	Category of groups	Groups / Group homomorphisms
Ab	Abelian groups	Commutative groups / Homomorphisms
Ring	Category of rings	Rings / Ring homomorphisms
Vect_F	Vector spaces	Spaces over F / Linear mappings
Top	Topological spaces	Spaces / Continuous mappings
Man	Manifolds	Smooth manifolds / Smooth mappings

Reference: fundamental inequalities

These inequalities are used throughout all of mathematics. Memorize them — they will appear in analysis, linear algebra, probability theory, physics.

Cauchy–Bunyakovsky–Schwarz inequality (CBS)

$$|\langle u, v\rangle| \leq \|u\| \cdot \|v\|$$

Modulus of scalar product ≤ product of lengths

Special cases:

For vectors in $\mathbb{R}^n$:

$$|\Sigma a_i b_i|^2 \leq (\Sigma a_i^2)(\Sigma b_i^2)$$

For integrals:

$$|\int fg\, dx|^2 \leq (\int f^2\, dx)(\int g^2\, dx)$$

For sums:

$$(a_1 b_1 + a_2 b_2)^2 \leq (a_1^2 + a_2^2)(b_1^2 + b_2^2)$$

Geometric meaning:

$\cos \theta = \langle u,v\rangle/(\|u\|\cdot\|v\|)$, and $|\cos \theta| \le 1$

When equality holds:

u and v are collinear (one is a multiple of the other): $v = \lambda u$

Triangle inequality

```
+-------------------------+
| ‖u + v‖ ≤ ‖u‖ + ‖v‖      |
|                         |
| Length of sum ≤ sum of lengths |
+-------------------------+
```

For numbers: $|a + b| \le |a| + |b|$

For metrics: $d(x,z) \le d(x,y) + d(y,z)$

Reverse inequality: $|\|u\| - \|v\|| \le \|u - v\|$

Inequality of means (AM-GM)

```
+--------------------------------------------------------+
| For nonnegative numbers a₁, ..., aₙ:                    |
|                                                        |
| (a₁ + a₂ + ... + aₙ)/n  ≥  ⁿ√(a₁ · a₂ · ... · aₙ)      |
| AM              ≥          GM                          |
| (arithmetic mean)              (geometric mean)        |
+--------------------------------------------------------+
```

Special case ($n = 2$):

$(a + b)/2 \ge \sqrt{(ab)}$ equality when $a = b$

Application:

Minimizing sum with fixed product (and vice versa)

Example: Which rectangle with perimeter 20 has max area?

$2(a+b) = 20 \Rightarrow a+b = 10$

$S = ab \le (a+b)^2/4 = 25$ (max when $a = b = 5$, i.e. square)

Jensen's inequality

```
+--------------------------------------------------+
| For CONVEX function f and weights λᵢ ≥ 0 with Σλᵢ = 1: |
|                                                  |
| f(Σ λᵢxᵢ)  ≤  Σ λᵢf(xᵢ)                          |
|                                                  |
| Function of mean ≤ mean of function              |
+--------------------------------------------------+
```

Important: For concave function sign changes to ≥

Examples of convex functions: x^2, e^x, $|x|$, $-\ln x$ (on $x>0$)
Examples of concave functions: $\sqrt{x}$, $\ln x$, $-x^2$

Special case: AM-GM follows from Jensen for $f(x) = -\ln(x)$

Hölder's and Minkowski's inequalities

Hölder's inequality (generalization of CBS):

For $p, q > 1$ with $1/p + 1/q = 1$:

$$\boxed{\begin{gathered} \Sigma|a_i b_i| \le (\Sigma|a_i|^p)^{(1/p)} \cdot (\Sigma|b_i|^q)^{(1/q)} \\ \|ab\|_1 \le \|a\|_p \cdot \|b\|_q \end{gathered}}$$

When $p = q = 2$ we get CBS.

Minkowski's inequality (triangle inequality for L^p):

$$\boxed{\begin{gathered} (\Sigma|a_i+b_i|^p)^{(1/p)} \le (\Sigma|a_i|^p)^{(1/p)} + (\Sigma|b_i|^p)^{(1/p)} \\ \|a + b\|_p \le \|a\|_p + \|b\|_p \end{gathered}}$$

This proves that $\|\cdot\|_p$ is indeed a norm.

Table: When to use which inequality

SITUATION	INEQUALITY
Estimate scalar product	CBS: $|\langle u,v\rangle| \le \|u\|\cdot\|v\|$
Estimate norm of sum	Triangle: $\|u+v\| \le \|u\|+\|v\|$
Compare sum and product	AM-GM: $(a+b)/2 \ge \sqrt{(ab)}$
Convex function of mean	Jensen: $f(Ex) \le E[f(x)]$
Work in L^p spaces	Hölder, Minkowski

Applied example: inequalities in heat transfer

Problem 1: Rectangle of maximum area at fixed perimeter (AM-GM)

Designing a fence. Given perimeter $P = 2(a + b) = 20$.
Which rectangle shape gives the maximum area $S = ab$?

By AM-GM: $(a + b)/2 \geq \sqrt{ab}$, i.e. $ab \leq ((a+b)/2)^2 = (P/4)^2$

Maximum $ab = (P/4)^2$ is achieved when $a = b$ (square).
For $P = 20$: $S_{max} = 25$, at $a = b = 5$.

This is the classical isoperimetric inequality for rectangles.

Note on fin design: the dual problem "maximize a+b at fixed ab" has NO maximum ($a+b \to \infty$ as $a \to 0$, $b \to \infty$). That is why heat-sink fins are made thin and long, not square — the actual constraint is "strength" or "minimum manufacturing thickness", not "fixed cross-section area".

Problem 2: Fan power estimate (CBS)

Air velocity $v(x,y)$ is inhomogeneous across the duct cross-section.
We need to estimate the kinetic energy of the flow.

$$E = \tfrac{1}{2}\rho \iint v^3 \, dA = \tfrac{1}{2}\rho \iint v \cdot v^2 \, dA$$

By Cauchy–Bunyakovsky:

$$\left(\iint v \cdot v^2 \, dA\right)^2 \leq \left(\iint v^2 \, dA\right) \cdot \left(\iint v^4 \, dA\right)$$

This allows estimating energy through simpler integrals.

Problem 3: Temperature averaging (Jensen)

Emissivity ~ T^4 (Stefan-Boltzmann law).
T^4 — convex function.

By Jensen: $(\text{average } T)^4 \leq \text{average}(T^4)$

Practical meaning:
One cannot replace temperature distribution with average temperature when calculating radiative heat transfer — this gives underestimation.

If $T_1 = 300$ K, $T_2 = 500$ K:
$(400)^4 = 2.56 \times 10^{10}$ — using average temperature
$\tfrac{1}{2}(300^4 + 500^4) = 3.53 \times 10^{10}$ — correct calculation (38% higher)

Vector analysis — engineer's handbook

Intuition: what grad, div, rot do

∇f (gradient) — "where to go up?"

Imagine a height map (temperatures, pressures).
∇f — this is an arrow pointing in the direction of greatest increase of f.
Length |∇f| = steepness of ascent.

```
         high
           ○
          /|\
         / | \     ∇f points upward along slope
        /  |  \    +---→
       /   |   \   |
      -----●-----  (point where you stand)
             low
```

Example: $T(x,y) = 20 - x^2 - y^2$ (temperature hill in center)
$\nabla T = (-2x, -2y)$ — directed toward center (to hill peak)

Div F (divergence) — "source or sink?"

Imagine vector field as fluid flow.
div F > 0: at this point fluid appears (source, faucet)
div F < 0: at this point fluid disappears (sink, drain)
div F = 0: as much flows in as flows out (incompressible flow)

```
   Source (div > 0) sink (div < 0) div = 0

         ↑                         ↓                    → → →
      ←--●--→                   →--●--←                 → → →
         ↓                         ↑                    → → →
   (arrows diverge)        (arrows converge)     (uniform flow)
```

Example: $F = (x, y, z)$ (radial field)
$\operatorname{div} F = \partial x/\partial x + \partial y/\partial y + \partial z/\partial z = 3 > 0$ — source everywhere.

Rot F (curl) — "is there a vortex?"

Imagine a small propeller thrown into flow.
rot F ≠ 0: propeller spins (there is a vortex)
rot F = 0: propeller does NOT spin (potential flow)

```
   Vortex (rot ≠ 0) potential (rot = 0)

        ←                            ↑
     ↙     ↖                         ↑
    ↓   ●    ↑                       ↑
     ↘     ↗                         ●
        →                            ↑
  (swirls)                   (rectilinear)
```

Direction of rot F — rotation axis (by right-hand rule)
Length |rot F| — rotation speed

Example: $F = (-y, x, 0)$ (circular flow)
rot $F = (0, 0, 2)$ — vortex along z axis

Formulas in Cartesian coordinates (x, y, z)

$$\nabla f = (\partial f/\partial x, \partial f/\partial y, \partial f/\partial z)$$

$$\text{div } F = \partial F_x/\partial x + \partial F_y/\partial y + \partial F_z/\partial z$$

$$\text{rot } F = (\partial F_z/\partial y - \partial F_y/\partial z,\ \ \partial F_x/\partial z - \partial F_z/\partial x,\ \ \partial F_y/\partial x - \partial F_x/\partial y)$$

$$\nabla^2 f = \partial^2 f/\partial x^2 + \partial^2 f/\partial y^2 + \partial^2 f/\partial z^2 \quad \text{(Laplacian)}$$

Formulas in cylindrical coordinates (r, φ, z)

$$x = r \cos \varphi,\ \ y = r \sin \varphi,\ \ z = z$$

$$\nabla f = (\partial f/\partial r,\ \ (1/r)\partial f/\partial \varphi,\ \ \partial f/\partial z)$$

$$\text{div } F = (1/r)\partial(rF_r)/\partial r + (1/r)\partial F_\varphi/\partial \varphi + \partial F_z/\partial z$$

$$\begin{aligned}\text{rot } F = (&(1/r)\partial F_z/\partial \varphi - \partial F_\varphi/\partial z,\\ &\partial F_r/\partial z - \partial F_z/\partial r,\\ &(1/r)[\partial(rF_\varphi)/\partial r - \partial F_r/\partial \varphi])\end{aligned}$$

$$\nabla^2 f = (1/r)\partial/\partial r(r\partial f/\partial r) + (1/r^2)\partial^2 f/\partial \varphi^2 + \partial^2 f/\partial z^2$$

When to use: pipes, cylinders, axisymmetric problems

Formulas in spherical coordinates (r, θ, φ)

$x = r \sin \theta \cos \varphi$, $y = r \sin \theta \sin \varphi$, $z = r \cos \theta$
(θ – angle from z axis, φ – azimuthal angle)

$$\nabla f = (\partial f/\partial r,\ (1/r)\partial f/\partial \theta,\ (1/(r \sin \theta))\partial f/\partial \varphi)$$

$$\mathrm{div}\, F = (1/r^2)\partial(r^2 F_r)/\partial r + (1/(r \sin \theta))\partial(\sin \theta\, F\theta)/\partial \theta + (1/(r \sin \theta))\partial F\varphi/\partial \varphi$$

$$\nabla^2 f = (1/r^2)\partial/\partial r(r^2 \partial f/\partial r) + (1/(r^2 \sin \theta))\partial/\partial \theta(\sin \theta\, \partial f/\partial \theta) + (1/(r^2 \sin^2 \theta))\partial^2 f/\partial \varphi^2$$

When to use: balls, spheres, point sources

Most important identities

$\mathrm{rot}(\nabla f) = 0$ ← "Gradient is irrotational"
(potential field does not rotate)

$\mathrm{div}(\mathrm{rot}\, F) = 0$ ← "Vortex has no sources"
(vortex lines are closed)

$\mathrm{rot}(\mathrm{rot}\, F) = \nabla(\mathrm{div}\, F) - \nabla^2 F$

$\mathrm{div}(fF) = f\, \mathrm{div}\, F + \nabla f \cdot F$

$\mathrm{rot}(fF) = f\, \mathrm{rot}\, F + \nabla f \times F$

Connection with physics (equations in vector form)

Heat conduction:
$q = -\lambda \nabla T$ (heat flux ~ minus temperature gradient)
$\partial T/\partial t = a \nabla^2 T$ (heat equation)

Hydrodynamics:
$\mathrm{div}\, v = 0$ (incompressible fluid)
$\partial v/\partial t + (v \cdot \nabla)v = -\nabla p/\rho + \nu \nabla^2 v$ (Navier–Stokes)

Electrodynamics (Maxwell):
$\mathrm{div}\, E = \rho/\varepsilon_0$ (source of E – charges)
$\mathrm{div}\, B = 0$ (no magnetic monopoles)
$\mathrm{rot}\, E = -\partial B/\partial t$ (change in B generates vortex E)
$\mathrm{rot}\, B = \mu_0 J + \mu_0 \varepsilon_0 \partial E/\partial t$ (current and change in E generate B)

The Grand Unification

All of mathematics is one pattern, viewed from different perspectives.

```
+---------------------------------------------------------------------+
|                                                                     |
|                          Structure                                 |
|                              |                                     |
|            +-----------------+-----------------+                   |
|            |                 |                 |                   |
|            ▼                 ▼                 ▼                   |
|      +----------+      +----------+      +----------+              |
|      | Algebra  |      |Topology  |      | Analysis |              |
|      |          |      |          |      |          |              |
|      |symmetry  |◄----►|  shape   |◄----►|  change  |              |
|      | G · g⁻¹  |      |  τ, π₁   |      |  d, ∫    |              |
|      +----------+      +----------+      +----------+              |
|            |                 |                 |                   |
|            +-----------------+-----------------+                   |
|                              |                                     |
|                              ▼                                     |
|                     +----------------+                             |
|                     |   Categories   |                             |
|                     |                |                             |
|                     | Objects        |                             |
|                     | Morphisms      |                             |
|                     | Functors       |                             |
|                     +----------------+                             |
|                              |                                     |
|                              ▼                                     |
|                    Universal language                              |
|                                                                     |
+---------------------------------------------------------------------+
```

A group is a category with one object, where all morphisms are invertible.
Topology is the category Top: objects are topological spaces,
morphisms are continuous mappings.
Linear algebra is the category Vect: objects are vector spaces,
morphisms are linear maps.
It also carries ⊕ (direct sum) and ⊗ (tensor product).

When you understand one structure deeply — you understand all.

One language — many dialects

```
+---------------------------------------------------------------------+
|                                                                     |
|    ALGEBRA         TOPOLOGY        LIN.ALGEBRA      DIFF.GEOM.      |
|       |                |                |               |          |
|       ▼                ▼                ▼               ▼          |
|    +-----+          +-----+          +-----+        +-----+       |
|    |ker φ|          | π₁  |          |ker T|        |ker d|       |
|    +--+--+          +--+--+          +--+--+        +--+--+       |
|       |                |                |               |          |
|       +----------------+----------------+---------------+          |
|                        |                |                          |
|                        ▼                ▼                          |
|                 one and the same pattern:                          |
|                                                                     |
|                     "WHAT GOES TO ZERO"                            |
|                                                                     |
|   The kernel is what "collapses", "is lost", "becomes trivial"     |
|   when applying a morphism.                                        |
|                                                                     |
|   dim V = dim ker + dim Im  ← THIS WORKS EVERYWHERE.               |
|                                                                     |
+---------------------------------------------------------------------+
```

Why this matters

Mathematics is not a set of techniques for passing an exam.
Mathematics is the language in which the Universe is written.

Maxwell's laws: $dF = 0$, $d{*F} = J$ (two lines)
Gravity: $G_{\mu\nu} = 8\pi T_{\mu\nu}$ (one equation)
Quantum mechanics: $i\hbar\partial\psi/\partial t = \hat{H}\psi$ (one equation)

The structures described in this atlas underlie:

- Elementary particle physics (symmetry groups)
- General relativity (manifolds, tensors)
- Quantum mechanics (Hilbert spaces, operators)
- Cryptography (number theory, elliptic curves)
- Machine learning (linear algebra, optimization)

By studying mathematics, you are studying the structure of reality itself.

> "The book of nature is written in the language of mathematics"
>
> — Galileo Galilei

> "God is a geometer"
>
> — Plato

> "Mathematics is the music of reason"
>
> — James Joseph Sylvester

Mathematics is not a set of formulas. It is a way of seeing the world: finding invariants amid change, structure amid chaos, unity amid differences.

Emptiness → boundaries → space → structure → measurement.

This path continues.

What's Next

This atlas covers the core of mathematics. Beyond it lie territories where active research is being conducted. A map for orientation:

Algebra:
- Representation theory — how groups act on spaces
- Homological algebra — chain complexes, derived functors
- Algebraic K-theory — generalization of the notion of dimension

Geometry and topology:
- Homotopy type theory — new foundation of mathematics
- Knot theory — Jones invariants, connection to quantum physics
- Algebraic topology — spectral sequences
- Symplectic topology — Arnold's conjecture, Gromov's theorem

Analysis:
- Stochastic analysis — Itô integral, Feynman-Kac formula
- Microlocal analysis — operators, wave fronts
- Noncommutative geometry (Connes) — generalization of manifolds

Mathematical physics:
- Quantum field theory — functional integrals, renormalization
- General relativity — Einstein's equations, black holes
- String theory — gauge/gravity duality

Discrete mathematics and computer science:
- Complexity theory — P vs NP, cryptography
- Game theory — equilibria, mechanisms
- Combinatorics on infinite structures — Ramsey theory

Each of these territories uses the language we already know: groups, spaces, forms, measures, functors. The atlas is a compass.