Topology Primer

Manas Patil

1 Concepts

- Topology — open sets, bases, subspace/product/quotient topologies
- Metric spaces and the metric topology (how metrics induce topologies)
- Simplices and abstract simplicial complexes (vertices, simplices, faces)
- Geometric realization of a simplicial complex
- Nerve of a cover (nerve complex) and the Nerve Lemma / nerve theorem
- Čech complex and Vietoris–Rips complex (how they are formed from point-clouds)
- Minkowski sum and support functions (Minkowski constructions for sets / offsets)
- Filtrations (in topology) and sublevel / offset filtrations
- Homology: chains, boundary operators, cycles, boundaries, homology groups, Betti numbers
- Persistent homology: persistence modules, birth/death, barcodes and persistence diagrams
- Algorithmic computation of persistent homology (boundary matrix reduction / pairing)
- Stability of persistence diagrams and distances between diagrams (bottleneck, Wasserstein)
- Nerve $\leftrightarrow$ union-of-balls relationship (how Čech relates to unions of balls via Nerve Lemma)
- Computational issues: combinatorial explosion of simplices, sparsification, and practical approximations (e.g., Rips vs Čech)
- Relative homology, Mayer–Vietoris, and multi-parameter / zigzag persistence (advanced topics)

2 What is a topology?

Definition (informal). A topology on a set X is a choice of which subsets of X we call open, subject to three rules:

1. The empty set $\varnothing$ and the whole set X are open.
2. Any union (even infinite) of open sets is open.
3. Any finite intersection of open sets is open.

A pair $(X, \mathcal{T})$ — where $\mathcal{T}$ is a family of subsets of X satisfying these rules — is called a topological space.

Why this abstraction? A topology captures a notion of "nearness" or "continuity" without resorting to distances. It allows us to talk about continuous maps, convergence, connectedness, compactness, and holes — the basic qualitative features of shape.

Basic examples

- Discrete topology: every subset is open ($\mathcal{T} = 2^X$).
- Trivial (indiscrete) topology: only $\varnothing$ and X are open.

- Standard topology on $\mathbb{R}$: opens are unions of open intervals (a, b).
- Metric topology: if d is a metric on X, open balls $B_r(x) = \{y : d(x, y) < r\}$ generate a topology (all unions of such balls).

3 Bases, continuity, subspace and product topologies

Base of a topology. A collection $\mathcal{B}$ of subsets of X is a basis if:

- For every $x \in X$ there is some $B \in \mathcal{B}$ with $x \in B$.
- If $x \in B_1 \cap B_2$ with $B_1, B_2 \in \mathcal{B}$, then there exists $B_3 \in \mathcal{B}$ with $x \in B_3 \subset B_1 \cap B_2$.

The sets generated by arbitrary unions of basis elements form a topology.

Continuity in topological terms. A function $f : (X, \mathcal{T}_X) \to (Y, \mathcal{T}_Y)$ is continuous iff the preimage of every open set of Y is open in X. This generalizes ε-δ continuity without distances.

Subspace topology. If $A \subset X$, the subspace topology on A is $\{U \cap A : U \in \mathcal{T}_X\}$.

Product topology. For product $X \times Y$, the product topology is the one generated by basis sets $U \times V$ with U open in X, V open in Y.

4 Metric spaces — how they connect to topology

A metric space (X, d) gives open balls $B_r(x)$. Those balls generate a topology: the metric topology. Many intuitive geometric statements (like "two points are close") come from metrics; topology records the same continuity/connectedness notions but without exact numeric distances.

5 Simplices and simplicial complexes — discrete models of topology

k-simplex. The convex hull of $k + 1$ affinely independent points in $\mathbb{R}^n$ (or an abstract copy) is a k-simplex. Examples:

- 0-simplex = point (vertex),
- 1-simplex = edge (segment between two vertices),
- 2-simplex = filled triangle,
- 3-simplex = filled tetrahedron.

Abstract simplicial complex. A set K of finite subsets of a vertex set V such that if $\sigma \in K$ and $\tau \subset \sigma$ then $\tau \in K$. Each σ is a simplex. This is combinatorial: it records which sets of vertices form a simplex.

Geometric realization. One can embed an abstract complex into Euclidean space to get an actual geometric simplicial complex (gluing simplices along common faces). The geometric realization is a topological space.

Simplicial complexes are the combinatorial playground for computing topology: connected components, loops, voids show up as patterns of simplices.

6 The nerve of a cover (nerve complex) and the Nerve Lemma

Nerve of a cover. Suppose $\mathcal{U} = \{U_i\}_{i \in I}$ is a cover of a space X (each $U_i \subset X$, union equals X). The nerve $N(\mathcal{U})$ is an abstract simplicial complex whose vertices correspond to the sets U_i; a finite set of vertices $\{U_{i_0}, \ldots, U_{i_k}\}$ spans a k-simplex precisely when the intersection $U_{i_0} \cap \cdots \cap U_{i_k}$ is nonempty.

Nerve Lemma (informal). If the cover $\mathcal{U}$ is a good cover (all finite intersections are either empty or contractible — i.e., have trivial topology like a ball), then the nerve $N(\mathcal{U})$ has the same homotopy type as the union $\bigcup_i U_i$. In particular, they have isomorphic homology groups.

Why important? The Nerve Lemma lets us replace a continuous union-of-sets (maybe complicated) with a combinatorial simplicial complex (the nerve) and still capture the same topological features (holes, components) — under the "good cover" condition.

7 Čech and Vietoris–Rips complexes (how they arise from points)

Given a finite point cloud $P = \{p_1, \dots, p_n\} \subset \mathbb{R}^d$ and a scale parameter $r > 0$:

- Union of balls (offsets): $U(r) = \bigcup_i B_r(p_i)$. The topology of $U(r)$ changes with r.
- Čech complex $\check{C}(P, r)$: vertices = points in P. A simplex $[p_{i_0}, \dots, p_{i_k}]$ is included iff $\bigcap_{j=0}^{k} B_r(p_{i_j}) \neq \varnothing$. Equivalently, the balls around those points have a common intersection. The Čech complex is exactly the nerve of the cover by balls of radius r. By the Nerve Lemma (balls in Euclidean space are contractible), the Čech complex has the same homotopy type as $U(r)$.
- Vietoris–Rips (Rips) complex $R(P, r)$: vertices = points in P. A simplex $[p_{i_0}, \dots, p_{i_k}]$ is included iff every pair of its vertices is at distance $\leq 2r$ (or $\leq r$, depending on convention). So Rips depends only on pairwise distances. Rips is computationally easier (only needs pairwise distances) but can have simplices that do not correspond to an actual overlap of all balls; generally there's an inclusion chain: $\check{C}(P, r) \subset R(P, r) \subset \check{C}(P, \alpha r)$ for some constant α depending on the metric (in Euclidean space often $\alpha = \sqrt{2}$ or similar for certain relations).

Takeaway: Čech complexes faithfully represent the union-of-balls topology (via Nerve Lemma). Rips complexes are useful approximations built from distances.

8 Minkowski sum and support functions — set operations used for offsets

Minkowski sum. For sets $A, B \subset \mathbb{R}^d$, the Minkowski sum is

$$A + B = \{a + b \mid a \in A, b \in B\}.$$

If B is a ball of radius r, $A + B$ is the radius-r offset (the union of balls of radius r centered on A).

Support function of a convex set. For a convex compact set $K \subset \mathbb{R}^d$ and direction $u \in S^{d-1}$,

$$h_K(u) = \sup_{x \in K} \langle u, x \rangle.$$

It encodes the farthest extent of K in direction u. The support function is linear with respect to Minkowski sums: $h_{A+B}(u) = h_A(u) + h_B(u)$.

Minkowski functional / gauge. If K is a convex set containing the origin, the Minkowski functional $p_K(x) = \inf\{\lambda \geq 0 : x \in \lambda K\}$ measures scaled membership in K. These concepts are handy for geometric computations of offsets, shape approximations, and support-based summaries of shapes.

9 Filtrations — the one-parameter family of nested spaces

A filtration is a nested sequence of spaces (or complexes)

$$X_0 \subseteq X_1 \subseteq X_2 \subseteq \cdots$$

Common filtrations in TDA:

- Scale filtration (offsets): $U(r)$ as r increases.
- Sublevel filtration: for a function $f : X \to \mathbb{R}$, sets $\{x : f(x) \leq t\}$ as t increases.
- Simplicial filtrations: complexes built at increasing thresholds (e.g., Čech or Rips complexes as r grows).

Filtrations let us track when topological features appear and disappear as we "zoom" through scales.

10 Homology in detail: chains, boundaries, cycles, homology groups

Chain groups. For a simplicial complex K, the k-chains $C_k(K)$ are formal linear combinations of k-simplices with coefficients in a field (typically $\mathbb{Z}_2$ or $\mathbb{Q}$). For $\mathbb{Z}_2$, coefficients are 0 or 1 and addition is XOR.

Boundary operator. $\partial_k : C_k \to C_{k-1}$ maps a k-simplex to the formal sum of its $(k-1)$-faces (with signs if using $\mathbb{Z}$). Key property: $\partial_{k-1} \circ \partial_k = 0$ (the boundary of a boundary is zero).

Cycles and boundaries.

- $Z_k = \ker \partial_k$ — k-cycles (chains with zero boundary).
- $B_k = \operatorname{im} \partial_{k+1}$ — k-boundaries (boundaries of $(k+1)$-chains).

Homology group.

$$H_k(K) = Z_k/B_k.$$

Elements of H_k are equivalence classes of cycles modulo boundaries. Intuitively they represent k-dimensional holes: for $k = 0$ components, $k = 1$ loops, $k = 2$ voids, etc.

Betti numbers. $\beta_k = \operatorname{rank}(H_k)$ is the number of independent k-dimensional holes.

11 Persistent homology: birth, death, barcodes, and diagrams

When a simplicial complex varies through a filtration $\{K_r\}$, the homology groups $H_k(K_r)$ change with r. Persistent homology tracks when a homology class appears (birth) and when it becomes trivial or merges into an older class (death).

Formalism. The sequence of homology groups and linear maps induced by inclusions gives a persistence module:

$$\cdots \to H_k(K_{r_i}) \to H_k(K_{r_j}) \to \cdots$$

Under mild finite conditions (finite complexes, field coefficients), this module decomposes into intervals $[b, d)$ — each interval corresponds to a generator born at scale b and dying at d. The collection of intervals is the barcode. The same information can be encoded as a multiset of points (b, d) in the plane — the persistence diagram.

Interpretation. Long intervals (far from the diagonal $b = d$) are persistent features (robust holes). Short intervals often correspond to noise.

12 How to compute persistent homology (algorithmic core)

Simplicial filtration input. We have an ordered list of all simplices $\sigma_1, \sigma_2, \ldots, \sigma_m$ respecting the filtration (if σ_i appears before σ_j then $i < j$).

Boundary matrix (D). Build a matrix whose columns correspond to simplices (ordered), rows correspond to faces of one lower dimension, and entries indicate whether that face is in the boundary of the simplex (over $\mathbb{Z}_2$ entries $0/1$). The matrix is upper-triangular if columns are ordered by increasing filtration index.

Column reduction / matrix reduction algorithm.

- Reduce the boundary matrix by column operations (adding one column to another, over the chosen field) until each column's lowest 1 (the pivot) is unique.
- The pairing rule: a column that reduces to zero corresponds to the birth of a feature that never dies within the filtration (or pairs with something at infinity); a nonzero column with lowest 1 at row i pairs the simplex of column j (birth at j) with the simplex corresponding to row i (death at j). Implementation detail: use standard persistent homology column-reduction (similar to Gaussian elimination but tailored).

Complexity. Worst-case cubic-ish in number of simplices; in practice, data structure optimizations, sparse representations, and clever orderings speed things up.

0-dimensional case (connected components) is easy. Use union-find (disjoint set union): as you add edges, components merge and you can pair births/deaths cheaply. For higher k one needs matrix reduction.

13 Stability: why persistence diagrams are meaningful

Bottleneck distance. Distance between two persistence diagrams D_1, D_2 is the minimal ε such that there's a bijection between points (allowing matching to the diagonal) with each pair at distance $\leq \varepsilon$. This metric formalizes closeness of diagrams.

Stability theorem (Cohen-Steiner, Edelsbrunner, Harer). Small perturbations of the input function (or point positions in Hausdorff distance) cause only small changes in the persistence diagram (with respect to bottleneck distance). This explains why persistent features are robust to noise.

14 How the Nerve + Čech + Minkowski ideas tie together (conceptual pipeline)

- Start with a point cloud P.
- For each scale r consider offset $U(r) = \bigcup B_r(p)$. This is the Minkowski sum $P + B_r(0)$ if you view P as a 0-dimensional set or as union of delta masses.
- The Čech complex at radius r is the nerve of the cover $\{B_r(p)\}$. By the Nerve Lemma (balls are contractible), the homology of the Čech complex equals that of $U(r)$.
- Therefore, computing persistent homology on the Čech filtration is, for our purposes, computing persistent homology of the union-of-balls filtration.
- Rips is a computationally cheaper surrogate, derived purely from pairwise distances. It approximates Čech and thus approximates the topology of $U(r)$.

So Minkowski sums/offsets, nerves (Čech), and persistence diagrams are the natural chain: offsets $\to$ nerve (Čech) $\to$ simplicial complex $\to$ algebraic homology $\to$ persistence diagram.

15 Advanced algebraic details: persistence modules, decomposition, and algebraic classification

Persistence module. Formally, a persistence module over $\mathbb{R}$ is a family of vector spaces $\{V_t\}_{t\in\mathbb{R}}$ with linear maps $\varphi_{s,t} : V_s \to V_t$ for $s \leq t$ satisfying functorial properties. In the finite, tame case, the module decomposes into a direct sum of interval modules $k_{[b,d)}$.

Interval decomposition theorem. For pointwise finite-dimensional, tame persistence modules over a field, we have a unique decomposition into intervals — this is the algebraic reason barcodes exist.

Zigzag / multi-parameter persistence. If the filtration is not a single parameter (e.g., two parameters vary independently), then interval decomposition may fail: modules may not decompose into intervals uniquely. Multi-parameter persistence is an active area of research and is more algebraically complicated.

16 Practical computation notes and pitfalls

- Combinatorial explosion: For n points, Rips complex at dimension k may have $\binom{n}{k+1}$ simplices — infeasible for large n or high k.
- Approximations: Use sparse filtrations, witness complexes, alpha-complexes (based on Delaunay triangulations), or landmarks to reduce size.

- Numeric robustness: Using $\mathbb{Z}_2$ coefficients avoids sign bookkeeping; floating point errors in distances can change adjacency and hence topology — robust heuristics and epsilon-tolerances help.
- Choice of filtration metric: The filtration you choose (Čech vs Rips vs sublevel of a function) affects the computed diagram; understanding the relation between them is crucial.

17 A worked mini example (0D and 1D) — concrete, step-by-step

0D (connected components) for a point cloud P:

- At $r = 0$: each point is its own component ($\beta_0 = n$).
- As r increases, when balls of two points touch (distance $\leq 2r$), their components merge. Using union-find, each merge pairs the birth of one component (the later-born point) with death at that scale. The persistence diagram for H_0 records component births at 0 and deaths at merge radii.
- Long-lived H_0 points: components that persist to large r (could be outliers if they die late).

1D (loops) — sketch:

- As r grows, three or more balls may overlap to create cycles (e.g., points arranged in a ring).
- In the Čech complex, a 1-cycle (loop) is represented by a cycle of edges not bounded by any filled triangle (2-simplex). The birth scale is when the cycle appears; the death is when a 2-simplex appears that fills it in (or when it merges with an older class).
- The boundary matrix reduction pairs simplices that create and kill homology classes: e.g., an edge may create an H_0 death but a triangle typically kills an H_1 cycle.

(If you want, I can write a tiny numeric example with 4 points in a square and show the filtration and resulting pairs — say so and I'll compute it.)

18 Connections you might have seen in the "Ablation analysis causality" chat

Based on the earlier messages you referenced (they asked for code computing Minkowski supports, nerve complex and persistence diagrams), the important conceptual links are:

- Minkowski support / sums provide a way to form offsets (union of balls) around geometric primitives.
- Nerve complex (Čech) is the combinatorial object encoding overlaps of those offsets.
- Persistence diagram summarizes how homological features of these nerves (or directly of offsets) appear/disappear across scales.
- Implementation-wise: compute offsets $\rightarrow$ compute nerve (or approximate via Rips) $\rightarrow$ build filtration $\rightarrow$ reduce boundary matrices $\rightarrow$ produce persistence diagram (points (b, d)).

19 Final advanced remarks and pointers (mathematical depth)

- Dualities and cohomology. Cohomology offers algebraic operations (cup product) and can sometimes be used to compute persistent cohomology more efficiently via matrix transpose tricks — practical libraries exploit this.
- Nerve Lemma assumptions. The Nerve Lemma requires intersections to be contractible. For unions of Euclidean balls that's satisfied; for arbitrary covers, caution is needed.

- Algebraic stability and inference. Persistence stability ensures that for reasonable small perturbations of input (Hausdorff distance, sup-norm on function), persistence diagrams vary continuously (in bottleneck metric). That's why long intervals are meaningful.

- Multiparameter persistence and invariants. When data has more than one parameter (e.g., spatial scale + density threshold), straightforward barcodes fail; one instead studies rank invariants, Hilbert series, or uses derived invariants — much more complicated.

20 Worked Example

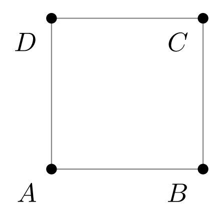
20.1 Point Configuration

We consider four points in $\mathbb{R}^2$ forming a unit square:

$$A(0,0), \quad B(1,0), \quad C(1,1), \quad D(0,1).$$

Pairwise distances:

$$d(A,B) = d(B,C) = d(C,D) = d(D,A) = 1, \quad d(A,C) = d(B,D) = \sqrt{2}.$$



20.2 Čech and Rips Complexes

For a scale parameter $r > 0$:

- **Čech complex** $\check{C}(P,r)$: includes simplex $[p_{i_0}, \ldots, p_{i_k}]$ if $\bigcap_{j=0}^{k} B_r(p_{i_j}) \neq \varnothing$.
- **Rips complex** $R(P,r)$: includes simplex $[p_{i_0}, \ldots, p_{i_k}]$ if every pair of its vertices is at distance $\leq 2r$.

20.3 Filtration of the Rips Complex

We increase r and note when edges and triangles appear.

r	Edges appear	Triangles appear	Homology changes
0	none	none	$\beta_0 = 4$
0.5	AB, BC, CD, DA	none	$\beta_0 = 1$
0.71	AC, BD	ABCD cycle now filled	$\beta_1 = 1$
0.8	ABC, BCD, CDA, DAB (2-simplices)	fills hole	$\beta_1 = 0$

For the above, the description of what happens is:

1. 4 isolated vertices
2. All points connect (square cycle)
3. Loop born (1-hole in middle)
4. Loop dies

Here $0.71 \approx \frac{\sqrt{2}}{2}$ is when diagonals appear in the Rips complex. At that scale, all 4 points are pairwise within $2r = \sqrt{2}$ distance.

20.4 Homology Computation

0-dimensional Homology (H_0):

$$\text{At } r = 0 : \beta_0 = 4, \quad r = 0.5 : \beta_0 = 1.$$

Thus, three connected components die at $r = 0.5$ (merged).

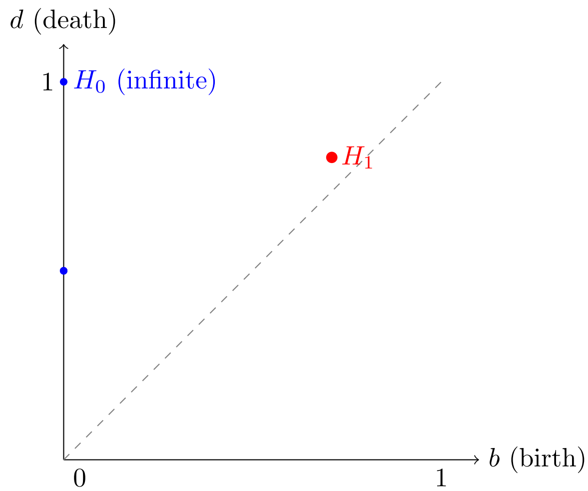
1-dimensional Homology (H_1):

$$\text{At } r = 0.5 : \text{no loops}, \quad r = 0.71 : 1 \text{ loop born}, \quad r = 0.8 : \text{loop dies}.$$

20.5 Persistence Diagram

We represent each topological feature (birth, death) as a point (b, d).

$$H_0 : (0, 0.5), (0, 0.5), (0, 0.5), (0, \infty) H_1 : (0.71, 0.8)$$



The long blue point $(0, \infty)$ represents the single persistent connected component (the whole square). The red point corresponds to the transient 1-dimensional hole (the square's center void).

20.6 Summary Interpretation

- At small r, four isolated points $\Rightarrow$ four components.
- As r grows, edges connect components $\Rightarrow$ merge to one component.
- Before diagonals fill, a single 1-dimensional loop appears (square boundary).
- At larger r, the square becomes filled $\Rightarrow$ loop dies.

Thus, the persistence diagram compactly captures the evolution of topology across scales.