Mathematics

http://www.aimspress.com/journal/Math

AIMS Mathematics, 11(9): 29754–29773.
DOI: 10.3934/math.20261180
Received: 19 June 2026
Revised: 03 September 2026
Accepted: 04 September 2026
Published: 14 September 2026

Research article

Filtered groupoid persistence: A homotopy-theoretic framework for symmetry-aware topological data analysis

Samirah Alsulami*

Department of Mathematics and Statistics, Faculty of Science, University of Jeddah, Jeddah 21589, Saudi Arabia

* **Correspondence:** Email: shalsulami@uj.edu.sa.

Abstract: Classical persistent homology extracts multiscale topological features from data but treats every point as an independent entity, making it insensitive to symmetries and equivalence relations intrinsic to physical, biological, and network datasets. We introduce filtered groupoid persistence (FGP), a mathematically rigorous framework that encodes both metric proximity and group-symmetry identifications within a single, unambiguous categorical structure: The *action groupoid* $(\mathcal{G}_\varepsilon)_{\varepsilon\geq 0}$ of a finite group G acting simplicially on the Vietoris–Rips filtration $(\mathrm{VR}_\varepsilon(X))_{\varepsilon\geq 0}$ of a finite metric space X. The classifying space of this action groupoid is, by a standard identity, the *Borel construction* $B\mathcal{G}_\varepsilon \simeq EG \times_G |\mathrm{VR}_\varepsilon(X)|$, and persistent homology of this filtered space defines the *FGP persistence module* $M_k(\varepsilon) = H_k(B\mathcal{G}_\varepsilon; \mathbb{F})$. Grounding the construction in the classical action groupoid removes the ambiguity present in earlier drafts of this framework and lets us prove, rather than merely assert, that FGP specializes to classical persistent homology when G is trivial. We establish four principal results: An exact *triviality proposition* recovering classical persistent homology PH as the $G = \{e\}$ case; a *functoriality theorem* showing that FGP is a well-defined functor from filtered G-spaces to persistence modules; a *symmetry-collapse theorem*, proved via the standard fact that the Borel construction of a free G-action is homotopy equivalent to the genuine quotient, characterizing precisely when group actions cause homological cycles to vanish; and a *quantitative stability theorem*, now proved directly for the orbit-quotient pseudometric space that the practical algorithm consumes, so that it holds unconditionally at every scale rather than only below a free threshold. We further prove that, when G acts freely below an explicit scale threshold, the practical algorithm (Vietoris-Rips persistence on an orbit-identified distance matrix) has a provable one-sided relationship to the categorical FGP module in degree one, as it can only under-report loops, never fabricate them. This closes part of the gap between the categorical definition and the computational pipeline while stating plainly, and corrects a previous overclaim, that the two do not coincide as simplicial complexes. The flag complex built from the quotient pseudometric can contain higher simplices with no consistent lift to the original space, so equality is not asserted in general and is left open in degree two and above. For actions that are not free (approximate or soft symmetries), we are explicit that the algorithm is a heuristic extension without a proven equivalence. We validate FGP on four synthetic datasets: A noisy circle under antipodal $\mathbb{Z}_2$-symmetry, a figure-eight under reflection symmetry, a flat torus under $(\mathbb{Z}_2)^2$-symmetry, and a Lorenz-attractor point cloud with approximate $\mathbb{Z}_2$-symmetry, reporting bottleneck distances, repeated-trial statistics, and the free-action scale threshold $\varepsilon_{\text{free}}$ itself alongside the qualitative comparison. In the two experiments where the relevant symmetry is genuinely free away from an easily avoided small locus (circle, figure-eight), FGP substantially reduces the spurious cycle count predicted by classical persistent homology, though we show that $\varepsilon_{\text{free}}$ is small enough in the figure-eight case that the observed dominant loop is only partially covered by the free-action guarantee. In the torus and Lorenz cases, where a fixed locus or only approximate symmetry is present, FGP reduces but does not eliminate spurious features, and we report these residuals honestly rather than as clean zeros.

Keywords: filtered topological groupoid; classifying space; persistent homology; symmetry-aware TDA; homotopy quotient; stability; Segal nerve; dynamical systems
Mathematics Subject Classification: 55N31, 18B40, 55P35, 55U10, 62R40

1. Introduction

Topological data analysis (TDA) has become a principled tool for extracting shape information from complex datasets. Its central instrument, persistent homology, builds a nested sequence of simplicial complexes on a point cloud and tracks the birth and death of homological features as the scale parameter ε grows, encoding the result as a *persistence barcode* [6, 13, 14]. The power of persistent homology derives from its metric stability [9] and applicability to data from biology [25], physics [12], sensor networks [11], and machine learning [22].

Despite its versatility, classical persistent homology has a fundamental limitation: It is constructed on a *point cloud*, a metric space with no additional structure, and therefore cannot distinguish data points related by symmetries or equivalence relations. In many physical and biological systems, such symmetries are not incidental; they are intrinsic. Phase spaces of Hamiltonian systems carry symplectic symmetry groups; molecular configurations have crystallographic symmetry; brain networks exhibit bilateral organization; and attractors of chaotic dynamical systems contain patterns invariant under scaling and rotation. When persistent homology is applied to such data without acknowledging the underlying symmetry, it systematically *over-counts* topological features: Orbits of the same topological cycle appear as independent generators of homology.

Contribution. We propose *filtered groupoid persistence (FGP)*, a framework that resolves this limitation by encoding symmetry directly as morphisms in a topological groupoid. For a dataset X equipped with a metric d and a group G acting on X by isometries, we define the *filtered data groupoid* $(\mathcal{G}_\varepsilon)_{\varepsilon\geq 0}$: A categorical structure whose objects are data points and whose morphisms are generated by both proximity arrows (when $d(x, y) < \varepsilon$) and symmetry arrows (from the G-action). Applying the Segal nerve construction [23] yields a filtered classifying space $B\mathcal{G}_\varepsilon$, and $H_k(B\mathcal{G}_\varepsilon)$ defines the FGP persistence module.

This construction is closely related to, but distinct from, a small existing literature on symmetry-aware persistence, and we position it carefully against each strand rather than claiming an empty field:

- *Equivariant homology* [5, 26] modifies coefficient systems but does not define a filtration of classifying spaces.

- *Persistent (Borel) equivariant cohomology* [16] is the closest prior work: It studies $H^*_G(\mathrm{VR}_\varepsilon(X)) = H^*(EG \times_G |\mathrm{VR}_\varepsilon(X)|)$ for a *continuous* Lie group G (the circle group acting on itself) and derives explicit formulas via the Serre spectral sequence and the Gysin homomorphism for that specific example. FGP specializes the same Borel-construction idea to *finite* symmetry groups acting on finite point clouds, proves an exact free-action correspondence with a concrete orbit-identified Vietoris–Rips algorithm (Section 4 and Proposition 4.3), and targets the applied pipeline rather than the closed-form computation for one continuous-group example.
- *Persistent symmetry parametrization* [17] measures the *degree of symmetry* of a point configuration itself (via span categories and symmetry barcodes that track how close a fixed configuration is to admitting a given symmetry type) rather than computing the homology of a symmetry quotient; FGP instead assumes a known group action and asks what topology remains after quotienting by it. The two are complementary: Theirs quantifies symmetry and ours exploits a stipulated symmetry to correct persistent homology.
- *Categorical persistence* [10, 15] enriches the target category but does not incorporate group actions as groupoid morphisms.
- *Group equivariant non-expansive operators (GENEOs)* [3] construct operators between *function spaces* and produce operator pseudo-distances, not persistence barcodes; they use no groupoids or classifying spaces.

As a baseline in our experiments (Section 5), we also compare against the simpler heuristic of *orbit-space persistent homology(PH)*. We stress that this is deliberately *not* the same construction as Algorithm 4.1: Orbit-space PH fixes, once and for all and independently of ε, a single representative $r(x) \in G \cdot x$ per orbit via an arbitrary measurable section $r : X \to X$ of the quotient map (e.g. "take the point of the orbit with positive first coordinate, breaking ties arbitrarily"), and then runs classical Vietoris-Rips persistence on the reduced point set $r(X)$ with the *original*, unquotiented metric d. Algorithm 4.1, in contrast, never chooses a representative: It runs Vietoris–Rips persistence directly on the quotient pseudometric $\bar{d}$ (defined in Step 2 of Algorithm 4.1 below and studied formally in Lemma 4.2), which by construction assigns the correct, symmetric, scale-independent distance to every pair of orbits regardless of where the arbitrary cut r happens to fall. This is why the two produce different diagrams (Table 2): r introduces an artificial discontinuity at the boundary of its fundamental domain, so points on opposite sides of that boundary can be judged far apart by orbit-space PH even when their orbits are genuinely close in $\bar{d}$; this is not itself a published framework but a natural ablation that isolates the effect of retaining groupoid/Borel structure (and the representative-free pseudometric) versus a naive coordinate-wise collapse.

Main results.

(i) **Triviality** (Proposition 3.6): When $G = \{e\}$, $B\mathcal{G}_\varepsilon$ is isomorphic to $|\mathrm{VR}_\varepsilon(X)|$ *exactly*, so classical persistent homology is literally the $G = \{e\}$ case of FGP, not merely an informal special case.

(ii) **Functoriality** (Theorem 3.7): FGP is a well-defined functor from filtered G-spaces to persistence modules.

(iii) **Symmetry collapse** (Theorem 3.9): If G acts freely on $|\mathrm{VR}_\varepsilon(X)|$, then $H_k(B\mathcal{G}_\varepsilon; \mathbb{F}) \cong H_k(|\mathrm{VR}_\varepsilon(X)|/G; \mathbb{F})$; in particular, this vanishes whenever the quotient complex does.

(iv) **Algorithm-theory correspondence** (Proposition 4.3): Below the free-action threshold $\varepsilon_{\mathrm{free}}$, the orbit-identified Vietoris–Rips algorithm of Section 4 computes an H_1-rank that is a *provable lower* *bound* for $\dim M_1(\varepsilon)$, never an overcount; we no longer claim exact equality in general, or any bound in degree ≥ 2, correcting an earlier overclaim (Section 4).

(v) **Stability** (Theorem 3.10): $d_{\mathrm{B}}(\mathrm{dgm}_k(X/G, \bar{d}), \mathrm{dgm}_k(X'/G, \bar{d}')) \leq d_{\mathrm{H}}(X, X')$ unconditionally, for the algorithm's own output diagram; a matching bound for $M_k(\varepsilon)$ itself remains open.

Organization. Section 2 reviews persistent homology and groupoid theory. Section 3 defines FGP and proves the three main theorems. Section 4 presents the computational pipeline. Section 5 reports numerical experiments. Section 6 compares FGP with prior frameworks. Section 7 discusses applications. Section 8 concludes.

2. Background

2.1. Persistent homology

Let $K_0 \subset K_1 \subset \cdots \subset K_N$ be a filtration of simplicial complexes. Applying singular homology $H_k(-;\mathbb{F})$ over a field $\mathbb{F}$ yields a sequence of vector spaces and linear maps:

$$H_k(K_0) \to H_k(K_1) \to \cdots \to H_k(K_N).$$

A *persistence module* is such a sequence; its *barcode* is the multiset of half-open intervals $[b, d)$ (birth–death pairs) characterizing its isomorphism type [27]. The most common construction on a point cloud X is the *Vietoris–Rips filtration* $\mathrm{VR}(X, \varepsilon)$, which includes all simplices with a pairwise vertex distance below ε.

The *bottleneck distance* between barcodes A and B is

$$d_{\mathrm{B}}(A, B) = \inf_{\gamma} \sup_{(b,d)\in A} |(b, d) - \gamma(b, d)|_\infty,$$

where the infimum is over matchings γ of A to B. The fundamental stability theorem [9] states $d_{\mathrm{B}}(\mathrm{dgm}(X), \mathrm{dgm}(X')) \leq d_{\mathrm{H}}(X, X')$ for classical VR persistence.

2.2. Topological groupoids and classifying spaces

A *topological groupoid* $\mathcal{G}$ consists of topological spaces $\mathrm{Ob}(\mathcal{G})$ (objects) and $\mathrm{Mor}(\mathcal{G})$ (morphisms) with continuous source and target maps $s, t : \mathrm{Mor} \to \mathrm{Ob}$, continuous composition, unit, and inversion maps satisfying the groupoid axioms [20].

Example 2.1 (Action groupoid). *Let G act on X. The action groupoid $X \rtimes G$ has* $\mathrm{Ob} = X$, $\mathrm{Mor} = X \times G$, *source $s(x, g) = x$, target $t(x, g) = g \cdot x$, and composition* $(g \cdot x, h) \circ (x, g) = (x, hg)$. *When G acts freely, its classifying space a $B(X \rtimes G) \simeq EG \times_G X \simeq X/G$ [26].*

The *nerve* of $\mathcal{G}$ is the simplicial space

$$N_k(\mathcal{G}) = \{(f_1, \ldots, f_k) \in \mathrm{Mor}^k : s(f_i) = t(f_{i+1})\},$$

and the *classifying space* is $B\mathcal{G} = |N_\bullet(\mathcal{G})|$ [23]. For a discrete group G (one-object groupoid), BG is the Eilenberg-MacLane space $K(G, 1)$.

2.3. Related symmetry-aware TDA approaches

Existing approaches include: *Equivariant homology* [5] (which requires a closure-finite weak(CW) structure, no filtration defined); *persistent Borel equivariant cohomology* [16] (the closest prior work, restricted to a continuous Lie group acting on the Vietoris-Rips thickenings of the circle; see the discussion in Section 1), *orbit-space PH* (which collapses orbits before PH without retaining any groupoid or Borel structure, which we show in Section 5.1 can overcollapse a genuinely noncontractible quotient), *categorical persistence* [10, 15] (which enriches the target category, with no group-action morphisms), *persistent symmetry parametrization* [17] (which quantifies how symmetric a configuration is, rather than the homology of a symmetry quotient); and *GENEOs* [3] (function-space operators, not barcodes). None of these applies the Segal nerve Borel construction to a *filtered, finite-group* action groupoid with a proven algorithmic correspondence.

3. Filtered groupoid persistence

3.1. Filtered data groupoids

Definition 3.1 (Symmetry-metric dataset). *A* symmetry-metric dataset *is a triple* (X, d, G) *where* (X, d) *is a finite metric space, and G is a finite group acting on X by isometries; in particular, X is G-invariant as a set* ($g \cdot x \in X$ *for all* $x \in X$, $g \in G$).

Remark 3.2 (From a raw sample to a symmetry-metric dataset). *A finite sample* X_0 *drawn from an ambient metric space on which G acts by isometries is, in general,* not *itself G-invariant: For the noisy circle of Section 5.1, a randomly drawn* $x \in X_0$ *satisfies* $-x \in X_0$ *only by coincidence. Definition 3.1 is therefore not applied directly to* X_0*; it is applied to its G-*saturation $X := \widetilde{X_0} = \bigcup_{g \in G} g \cdot X_0$, *which is G-invariant by construction and is exactly the point set Step 1 of Algorithm 4.1 computes. Every theorem below (Lemma 3.8, Theorem 3.9, Proposition 4.3) is stated for a symmetry-metric dataset* (X, d, G) *in this sense, so when applied to sampled data, the object under study is always* $X = \widetilde{X_0}$, *not the raw sample* X_0*; we make this substitution explicit here so that the theoretical statements and Algorithm 4.1 are inputs to the same object rather than merely analogous ones. This also fixes* $\varepsilon_{\text{free}}$ *as a property of* $\widetilde{X_0}$ *(Lemma 3.8), and we report its computed value for each experiment in Section 5 rather than leaving it implicit.*

We deliberately do *not* define $\mathcal{G}_\varepsilon$ as an abstractly "generated" groupoid on the point set X. That approach admits several inequivalent readings: A free groupoid on the proximity graph and symmetry arrows, a thin groupoid that remembers only connectivity, or an action groupoid of a simplicial complex, and these give different, and in two of the three cases homologically unsuitable, answers. Concretely: The *free* groupoid on the 1-skeleton (proximity edges plus symmetry arrows) has as a classifying space a wedge of circles whose H_1-rank is the cycle rank of the raw proximity graph, which is generically much larger than the number of "real" topological loops, as it is not corrected by the 2-simplices that a Vietoris-Rips complex uses to fill in spurious cycles. The *thin* equivalence-relation groupoid (at most one arrow between related objects) has each connected component equivalent to a point, so $B\mathcal{G}_\varepsilon$ is always homotopy discrete and $H_k = 0$ for all $k \geq 1$ regardless of the data giving no information whatsoever about loops such as the figure-eight case of Section 5.2. Neither reduces to classical persistent homology when G is trivial, and neither is what the algorithm of Section 4 computes. We instead adopt the third, standard option: The *action groupoid of a group acting on a* *filtered simplicial complex*, exactly the classical construction of Example 2.1 applied to the geometric realization of the Vietoris–Rips filtration. This is unambiguous, is the object whose classifying space is a well-known homotopy invariant (the Borel construction), and is the object our algorithm actually computes (Proposition 4.3).

Definition 3.3 (Data groupoid $\mathcal{G}_\varepsilon$)**.** *For (X, d, G) and $\varepsilon \geq 0$, let $\mathrm{VR}_\varepsilon(X)$ be the Vietoris–Rips simplicial complex on X at scale ε and let $Y_\varepsilon = |\mathrm{VR}_\varepsilon(X)|$ be its geometric realization. Because G acts on X by isometries, $d(x, y) < \varepsilon \iff d(g \cdot x, g \cdot y) < \varepsilon$, so G acts simplicially on $\mathrm{VR}_\varepsilon(X)$ and hence continuously on Y_ε. The* data groupoid *$\mathcal{G}_\varepsilon$ is the action groupoid $Y_\varepsilon \rtimes G$ of Example 2.1:* $\mathrm{Ob}(\mathcal{G}_\varepsilon) = Y_\varepsilon$, $\mathrm{Mor}(\mathcal{G}_\varepsilon) = Y_\varepsilon \times G$, *$s(y, g) = y$, $t(y, g) = g \cdot y$, composition $(g \cdot y, h) \circ (y, g) = (y, hg)$.*

Definition 3.4 (Filtered data groupoid)**.** *For $\varepsilon \leq \varepsilon'$,* $\mathrm{VR}_\varepsilon(X) \subseteq \mathrm{VR}_{\varepsilon'}(X)$ *as a G-equivariant inclusion of simplicial complexes, inducing a groupoid inclusion $\mathcal{G}_\varepsilon \hookrightarrow \mathcal{G}_{\varepsilon'}$. The family $(\mathcal{G}_\varepsilon)_{\varepsilon \geq 0}$ is the* filtered data groupoid *of (X, d, G).*

Definition 3.5 (FGP persistence module)**.** *By the standard identity for classifying spaces of action groupoids [20, 26], $B\mathcal{G}_\varepsilon \cong EG \times_G Y_\varepsilon$, the Borel construction of G acting on Y_ε. The* FGP persistence module *in dimension k is*

$$M_k(\varepsilon) = H_k(B\mathcal{G}_\varepsilon; \mathbb{F}) = H_k(EG \times_G Y_\varepsilon; \mathbb{F}),$$

with structure maps induced by $B\mathcal{G}_\varepsilon \hookrightarrow B\mathcal{G}_{\varepsilon'}$. The barcode is $\mathrm{dgm}_k(\mathcal{G})$.

Proposition 3.6 (Trivial group recovers classical PH exactly)**.** *When $G = \{e\}$, $E\{e\}$ is a single point and the $\{e\}$-action on Y_ε is trivial, so $B\mathcal{G}_\varepsilon = E\{e\} \times_{\{e\}} Y_\varepsilon \cong \{\mathrm{pt}\} \times Y_\varepsilon \cong Y_\varepsilon = |\mathrm{VR}_\varepsilon(X)|$, an isomorphism of spaces, not merely a homotopy equivalence. Consequently $M_k(\varepsilon) = H_k(|\mathrm{VR}_\varepsilon(X)|; \mathbb{F})$ for every k and ε: Classical Vietoris-Rips persistent homology is* literally *the $G = \{e\}$ case of Definition 3.5, not an informal limit.*

3.2. Functoriality theorem

Theorem 3.7 (Functoriality of FGP)**.** *Every morphism $F = \{F_\varepsilon : \mathcal{G}_\varepsilon \to \mathcal{H}_\varepsilon\}$ of filtered data groupoids induces a morphism $F_* : M_k^{\mathcal{G}}(\varepsilon) \to M_k^{\mathcal{H}}(\varepsilon)$ of persistence modules. If each F_ε induces a homotopy equivalence $BF_\varepsilon : B\mathcal{G}_\varepsilon \xrightarrow{\simeq} B\mathcal{H}_\varepsilon$, then F_* is an isomorphism.*

Proof. An equivariant map $F_\varepsilon : Y_\varepsilon \to Z_\varepsilon$ of G-spaces (i.e., a functor of the corresponding action groupoids) induces, by functoriality of the Borel construction, a continuous map $EG \times_G F_\varepsilon : EG \times_G Y_\varepsilon \to EG \times_G Z_\varepsilon$; that is, $BF_\varepsilon : B\mathcal{G}_\varepsilon \to B\mathcal{H}_\varepsilon$ [20, 26]. The commutativity condition $F_{\varepsilon'} \circ \iota_{\varepsilon,\varepsilon'} = \iota'_{\varepsilon,\varepsilon'} \circ F_\varepsilon$ passes to $EG \times_G (-)$, so the induced homology maps commute with the structure maps, yielding a well-defined morphism of persistence modules. If each BF_ε is a homotopy equivalence it, induces isomorphisms on all homology groups; hence, F_* is an isomorphism of persistence modules. □

3.3. Symmetry-collapse theorem

We first isolate the freeness condition at the level of the simplicial complex, which is what the proof actually needs (freeness of G on the vertex set X is necessary but not sufficient, as g could fix a higher simplex setwise while permuting its vertices).

Lemma 3.8 (Free scale threshold)**.** *Let (X, d, G) be a symmetry-metric dataset, and set $\varepsilon_{\text{free}} = \frac{1}{2}\min_{x\in X,\, e\neq g\in G} d(x, g\cdot x)$ (finite and positive because X is finite; if G does not act freely on X, set $\varepsilon_{\text{free}} = 0$ and the statement is vacuous). For every $0 < \varepsilon < \varepsilon_{\text{free}}$, G acts freely on $Y_\varepsilon = |\mathrm{VR}_\varepsilon(X)|$: Every $g \neq e$ has $g\cdot y \neq y$ for all $y \in Y_\varepsilon$.*

Proof. Let $\sigma = \{x_0, \ldots, x_p\}$ be a simplex of $\mathrm{VR}_\varepsilon(X)$, so $d(x_i, x_j) < \varepsilon$ for all i, j. For $g \neq e$ and any $x_i \in \sigma$, $d(x_i, g\cdot x_i) \geq 2\varepsilon_{\text{free}} > 2\varepsilon > \varepsilon$ by triangle inequality considerations forces $g\cdot x_i \notin \sigma$: If it were, then $d(x_i, g\cdot x_i) < \varepsilon$ since both lie in σ, contradicting $d(x_i, g\cdot x_i) \geq 2\varepsilon_{\text{free}}$. Hence $g\cdot\sigma \neq \sigma$ as a set of vertices for every simplex σ, so g fixes no point of $|\mathrm{VR}_\varepsilon(X)|$ (points of the interior of σ are only moved into $g\cdot\sigma \neq \sigma$). □

Theorem 3.9 (Symmetry-induced collapse)**.** *Let (X, d, G) be a symmetry-metric dataset and $0 < \varepsilon < \varepsilon_{\text{free}}$. Then*

$$H_k(B\mathcal{G}_\varepsilon; \mathbb{F}) \;\cong\; H_k(Y_\varepsilon/G; \mathbb{F}) \quad \textit{for every } k,$$

where $Y_\varepsilon/G = |\mathrm{VR}_\varepsilon(X)|/G$ is the genuine topological quotient. In particular, if Y_ε/G has vanishing $\mathbb{F}$-homology in degree k (e.g. is $\mathbb{F}$-acyclic in that degree), then $M_k(\varepsilon) = 0$.

Proof. By Lemma 3.8, G acts freely on the G-CW complex Y_ε. For a free action of a discrete group on a CW complex, the canonical projection $EG \times_G Y_\varepsilon \to Y_\varepsilon/G$, $[e, y] \mapsto [y]$, is a fibre bundle with contractible fibre EG, hence a weak (and, since both spaces are CW, an honest) homotopy equivalence [26, Ch. II]. Since $B\mathcal{G}_\varepsilon \cong EG \times_G Y_\varepsilon$ (Definition 3.5), homotopy invariance of singular homology gives $H_k(B\mathcal{G}_\varepsilon; \mathbb{F}) \cong H_k(Y_\varepsilon/G; \mathbb{F})$. □

This replaces the earlier, unjustified appeal to a Künneth splitting $B\mathcal{G}_\varepsilon \simeq (X/G)\times BG$: That splitting is only valid when G acts trivially on π_1 of each component and, more basically, conflates the *orbit space of the finite point set X/G* with the orbit space of the *whole filtered complex Y_ε/G*, which is the object that actually carries the ε-dependent homotopy type. Theorem 3.9 avoids this by working with Y_ε/G directly and gives a concrete, checkable target: $H_k(B\mathcal{G}_\varepsilon)$ equals the homology of an explicit finite quotient simplicial complex, computed in Sections 5.1 and 5.2.

3.4. Stability theorem

We restate stability for the object that is actually computed and compared throughout Section 5: The persistence diagram $\mathrm{dgm}_k(X/G, \bar{d})$ returned by Algorithm 4.1. Our earlier proof proved an interleaving for the wrong pair of objects (the raw metric d on the augmented set $\widetilde{X}$, rather than the quotient pseudometric $\bar{d}$ Algorithm 4.1 actually uses) and then invoked Proposition 4.3 outside the interval on which it holds and for a statement (equality of full barcodes) it no longer makes; we thank the reviewer for pressing on this point. The corrected statement below is, in exchange, unconditional: It needs no freeness assumption and holds at every scale, because it is a direct instance of the classical Vietoris–Rips stability theorem applied to the quotient pseudometric space itself, whose Hausdorff distance we bound directly from $d_{\mathrm{H}}(X, X')$.

Theorem 3.10 (Quantitative stability of the algorithm's output)**.** *Let (X, d, G) and (X', d, G) be symmetry-metric datasets sharing an ambient metric space and group G (each already G-saturated, as in Remark 3.2), and let $\bar{d}$, $\bar{d}'$ be the corresponding quotient pseudometrics on X/G, X'/G. Then*

$$d_{\mathrm{H}}((X/G, \bar{d}),\, (X'/G, \bar{d}')) \;\leq\; d_{\mathrm{H}}(X, X'),$$

and consequently, by the classical stability theorem [8,9] applied to the two finite pseudometric spaces $(X/G, \bar{d})$ *and* $(X'/G, \bar{d}')$*,*

$$d_{\mathrm{B}}(\mathrm{dgm}_k(X/G, \bar{d}),\ \mathrm{dgm}_k(X'/G, \bar{d}')) \ \leq\ d_{\mathrm{H}}(X, X'),$$

for every k, with no restriction on ε *and no freeness hypothesis.*

Proof. Let $\delta = d_{\mathrm{H}}(X, X')$. For $[x] \in X/G$ pick any representative $x \in X$ and, by definition of δ, a point $x' \in X'$ with $d(x, x') \leq \delta$; then $\bar{d}'([x'], [x'])$ is irrelevant and instead $\bar{d}([x], [x']) \leq d(x, x') \leq \delta$ directly from the definition $\bar{d}([x], [x']) = \min_{g,h} d(g \cdot x, h \cdot x') \leq d(x, x')$ (take $g = h = e$). So every point of $(X/G, \bar{d})$ is within δ, in $\bar{d}$, of some point of $(X'/G, \bar{d}')$ under this correspondence, and symmetrically starting from X'; this is exactly $d_{\mathrm{H}}((X/G, \bar{d}), (X'/G, \bar{d}')) \leq \delta$, with no use of freeness, of G-equivariance of the correspondence, or of any augmentation. The bottleneck bound is then the classical Vietoris–Rips stability theorem applied verbatim to the two finite pseudometric spaces $(X/G, \bar{d})$ and $(X'/G, \bar{d}')$, which holds for any finite pseudometric space and any ε, not only below a free threshold. □

This theorem is unconditional but is deliberately a statement about Algorithm 4.1's output, $\mathrm{dgm}_k(X/G, \bar{d})$, rather than about the categorical module $M_k(\varepsilon)$ itself. Combining it with Proposition 4.3 gives a stability-flavoured consequence for M_1 only in the free regime and only as an inequality on ranks, not a bottleneck bound: For $\varepsilon < \min(\varepsilon_{\mathrm{free}}, \varepsilon'_{\mathrm{free}})$, $\dim H_1(X/G, \bar{d}) \leq \dim M_1(\varepsilon)$ and likewise for the primed dataset, so a large bottleneck distance between the two *algorithm* diagrams (established unconditionally above) lower-bounds, but does not by itself upper-bound, the corresponding distance between the two true FGP modules. A genuine bottleneck-stability theorem for $M_k(\varepsilon)$ itself, rather than for the algorithm's output, remains open and is listed in Section 8.

Corollary 3.11 (Statistical consistency of the algorithm's output)**.** *Let* $X_{0,n}$ *be n i.i.d. samples from a compact metric space* (M, d) *on which G acts by isometries, and let* $X_n = \widetilde{X_{0,n}}$ *be the G-saturation (Remark 3.2), similarly* $\widetilde{M}$ *for the population. Then* $d_{\mathrm{H}}(X_n, \widetilde{M}) \to 0$ *almost surely as* $n \to \infty$ *[21], so by Theorem 3.10,* $d_{\mathrm{B}}(\mathrm{dgm}_k(X_n/G, \bar{d}_n), \mathrm{dgm}_k(\widetilde{M}/G, \bar{d}_M)) \to 0$ *almost surely. We no longer claim consistency of* $\mathrm{dgm}_k(\mathcal{G})$ *itself, only of the algorithm's output, consistent with the scope of Theorem 3.10.*

4. Computational pipeline

Algorithm 4.1 (FGP computation)**.**
Input: *Raw point cloud* $X_0 = \{x_1, \ldots, x_n\}$ *(not assumed G-invariant; cf. see Remark 3.2), group G* $(|G| = m)$*, field* $\mathbb{F}$*, scale grid* $0 = \varepsilon_0 < \varepsilon_1 < \cdots < \varepsilon_N$*.*
Steps:

1. Augment. *Compute the G-saturation* $\widetilde{X} = \bigcup_{g \in G} g \cdot X_0$ *(at most nm points, with multiplicity when orbits are not free) and all pairwise distances* $d(\cdot, \cdot)$*. This is exactly the symmetry-metric dataset X of Definition 3.1 /Remark 3.2; i.e. that is,* $X = \widetilde{X}$ *from here on.*
2. Symmetry identification. *Construct the* $\mathbf{nm} \times \mathbf{nm}$ *matrix D with* $D(p, q) = \min\{d(g \cdot p, h \cdot q) : g, h \in G\}$ *for* $p, q \in \widetilde{X}$*; in particular* $D(x, g \cdot x) = 0$ *for all* $x \in X$*,* $g \in G$ *(orbit-pair identification). This is exactly the quotient pseudometric* $\bar{d}$ *on* $\widetilde{X}/G$ *pulled back to* $\widetilde{X}$*.*
3. Vietoris–Rips PH. *Run* `ripser` *on* $(\widetilde{X}, D)$ *up to the desired maximum dimension.*
4. Output. *Return barcodes* $\mathrm{dgm}_k(\mathcal{G})$ *for* $k = 0, 1, \ldots$

Complexity: *$O((nm)^2)$ for distance matrix; $O((nm)^3)$ for VR PH (or $O((nm)^{2.37})$ with optimized algorithms). For $m = |G| = 2$: Constant factor $\times 4$ over classical PH.*

Step 2 is exactly the construction of the quotient pseudometric induced by G; running Vietoris–Rips persistence on it is a well-defined, implementation-agnostic recipe. What is not automatic is that this recipe computes the categorical object $M_k(\varepsilon)$ of Definition 3.5. We now prove that it does, in the free-action regime, and we are explicit about where the correspondence is only heuristic. The quotient pseudometric $\bar{d}([x],[y]) = \min_{g,h\in G} d(g \cdot x, h \cdot y)$ minimizes *independently* over the representatives used for each pair. A set of orbits can therefore be pairwise ε-close in $\bar{d}$ without any *single, common* choice of representatives being pairwise ε-close in d. For example, for three orbits, the pair-witnessing translations for $([x],[y])$, $([y],[z])$ and $([z],[x])$ can be mutually incompatible group elements. Freeness of the action (Lemma 3.8) rules out a point being identified with a *proper subsimplex of itself*, but it does *not* rule out this representative-incoherence phenomenon, which can occur already for the free $\mathbb{Z}_2$-actions used in Section 5.

Lemma 4.2 (The algorithm's complex is the flag complex of the quotient graph)**.** *Let Γ_ε be the graph with vertex set X/G and an edge $\{[x],[y]\}$ whenever $\bar{d}([x],[y]) < \varepsilon$. Then:*

(a) $\mathrm{VR}_\varepsilon(\widetilde{X}, D)$ *computed by Algorithm 4.1 is, as a simplicial complex, exactly the flag (clique) complex of Γ_ε. This is what "$\mathrm{VR}_\varepsilon(X/G,\bar{d})$" denotes: The Vietoris-Rips construction is by definition the flag complex of its ε-proximity graph, and D is $\bar{d}$ pulled back along the orbit map exactly as before.*

(b) The 1-skeleton of $Y_\varepsilon/G = |\mathrm{VR}_\varepsilon(X)|/G$ (as a quotient simplicial complex) equals Γ_ε. Consequently, Y_ε/G is a subcomplex of $\mathrm{VR}_\varepsilon(X/G,\bar{d})$ with the same *1-skeleton; $\mathrm{VR}_\varepsilon(X/G,\bar{d})$ may contain additional simplices of dimension ≥ 2 (flag-filled cliques of Γ_ε with no consistent lift to an actual simplex of $\mathrm{VR}_\varepsilon(X)$), but no additional vertices or edges.*

Proof. (a) is immediate from the definitions: $D(p,q) = \bar{d}([x],[y])$ for p,q over $[x],[y]$, points at mutual D-distance 0 collapse to one vertex, and a Vietoris–Rips complex on any finite (pseudo)metric space is by definition the flag complex of its threshold graph. For (b): An edge of Y_ε/G is the orbit of an edge $\{x,y\}$ of $\mathrm{VR}_\varepsilon(X)$, that is, some pair of representatives with $d(x,y) < \varepsilon$; this witnesses $\bar{d}([x],[y]) < \varepsilon$ directly, so $\{[x],[y]\} \in \Gamma_\varepsilon$. Conversely, if $\{[x],[y]\} \in \Gamma_\varepsilon$ then, *some* translate pair $(g \cdot x, h \cdot y)$ has $d(g \cdot x, h \cdot y) < \varepsilon$; that pair is itself an actual edge of $\mathrm{VR}_\varepsilon(X)$. (Because G acts by isometries, its existence does not depend on which translates are chosen), and its G-orbit is $\{[x],[y]\}$. Therefore, the two edge sets coincide; unlike an edge, a 2-simplex requires *three* pairwise conditions to be witnessed by a *single* triple of representatives, which is exactly where coherence can fail. This means that no analogous argument extends to higher simplices, and $\mathrm{VR}_\varepsilon(X/G,\bar{d})$, being the *full* flag complex on Γ_ε, automatically contains every clique of Γ_ε, whether or not it is coherently witnessed. Because Y_ε/G only contains simplices that *are* coherently witnessed (each is the orbit of an actual simplex of $\mathrm{VR}_\varepsilon(X)$), $Y_\varepsilon/G \subseteq \mathrm{VR}_\varepsilon(X/G,\bar{d})$ as simplicial complexes on the common vertex set X/G, with equality of 0- and 1-skeleta. □

Proposition 4.3 (The algorithm's H_1 never exceeds the true FGP H_1, below the free threshold)**.** *For $0 < \varepsilon < \varepsilon_{\text{free}}$ (Lemma 3.8), there is a surjective linear map*

$$H_1(Y_\varepsilon/G;\mathbb{F}) \twoheadrightarrow H_1(\mathrm{VR}_\varepsilon(X/G,\bar{d});\mathbb{F}),$$

induced by the inclusion of Lemma 4.2(b), and hence, using $H_1(Y_\varepsilon/G;\mathbb{F}) \cong M_1(\varepsilon)$ (Theorem 3.9),

$$\dim H_1(\mathrm{VR}_\varepsilon(X/G,\bar{d});\mathbb{F}) \;\leq\; \dim M_1(\varepsilon).$$

That is, in the free regime, Algorithm 4.1 can only under-report *the FGP H_1-rank, as coherence failures can only fill in (kill) loops, never create spurious ones. Moreover, and equality holds whenever every 2-clique of Γ_ε not present as a coherent 2-simplex of Y_ε/G fails to add a new relation among the existing 1-cycles. (In particular, equality holds trivially when $G = \{e\}$, and empirically in our free-action experiments below, Table 1 and Section 5.2). We do* not *claim an analogous statement in degree ≥ 2: The argument below is specific to H_1 because it depends only on the (0- and 1-skeleton-only) definition of 1-cycles, and we leave the higher-degree case, together with the flat-torus H_2 comparison of Section 5.3, as an explicit open problem (Section 8).*

Proof. Write $K = Y_\varepsilon/G$ and $L = \mathrm{VR}_\varepsilon(X/G,\bar{d})$, with chain groups $C_*(K) \subseteq C_*(L)$ over $\mathbb{F}$. By Lemma 4.2(b), $C_0(K) = C_0(L)$, and $C_1(K) = C_1(L)$ (identical 0- and 1-skeleta), with $C_2(K) \subseteq C_2(L)$ (possibly with strict inclusion). The group of 1-cycles $Z_1 = \ker(\partial_1 : C_1 \to C_0)$ therefore coincides for K and L: $Z_1(K) = Z_1(L) =: Z_1$. The group of 1-boundaries satisfies $B_1(K) = \partial_2(C_2(K)) \subseteq \partial_2(C_2(L)) = B_1(L)$, as, ∂_2 is applied to a possibly larger domain. Hence

$$H_1(L) = Z_1/B_1(L)$$

is a further quotient of

$$H_1(K) = Z_1/B_1(K)$$

by the subspace $B_1(L)/B_1(K)$. That is, the inclusion $C_*(K) \hookrightarrow C_*(L)$ induces a surjection $H_1(K;\mathbb{F}) \twoheadrightarrow H_1(L;\mathbb{F})$. Applying $K = Y_\varepsilon/G$, $L = \mathrm{VR}_\varepsilon(X/G,\bar{d}) = \mathrm{VR}_\varepsilon(\widetilde{X}, D)$ (Lemma 4.2(a), the object Algorithm 4.1 literally returns) and $H_1(Y_\varepsilon/G;\mathbb{F}) \cong M_1(\varepsilon)$ (Theorem 3.9) completes the proof. □

Remark 4.4 (Scope of the correspondence: Free vs. approximate symmetry)**.** *Proposition 4.3 gives a one-sided, degree-1-only guarantee, and only for $\varepsilon < \varepsilon_{\text{free}}$; it says nothing when G does not act freely on X at all (if some x is a genuine fixed point of a nontrivial g, $\varepsilon_{\text{free}} = 0$). Two of our four experiments deliberately leave the free regime: The Lorenz-attractor case (Section 5.4) uses a* soft *identification $D(p,-p) = 0.1 \cdot d(p,-p)$ rather than an exact 0, precisely because the attractor's $\mathbb{Z}_2$-symmetry is only approximate, and there is no honest free action for which to invoke Theorem 3.9. In that regime, Algorithm 4.1 should be understood as a* heuristic extension *of FGP a practically useful weighting scheme motivated by, but not proved equal or even bounded relative to, the homology of any classifying space. We do not claim a theorem there, and we report it as such in Section 5.4. A third experiment (the flat torus, Section 5.3) has fixed loci, so $\varepsilon_{\text{free}}$ is small but positive, and we now report its computed value rather than leaving it implicit (Section 5.6). Establishing a rigorous semantics for soft/approximate symmetry, a bound in degree ≥ 2, and a characterisation of exactly when the inequality of Proposition 4.3 is strict, are left as explicit open problems in Section 8.*

5. Numerical experiments

We compare three methods: **(a)** classical PH (no symmetry), **(b)** orbit-space PH (orbits collapsed before PH), and **(c)** FGP (Algorithm 4.1). All quantitative results below computed using (`ripser` 0.6.15, `persim` 0.3.8, Python 3.11, and NumPy); exact settings, seeds, and trial counts are given per experiment and summarized in Section 5.6.

5.1. Noisy circle under antipodal $\mathbb{Z}_2$-symmetry

Setup. Sample $n = 100$ points from S^1 with noise σ: $(x_i, y_i) = (\cos\theta_i + \epsilon_i^x, \sin\theta_i + \epsilon_i^y)$, $\theta_i \sim \text{Uniform}[0, 2\pi]$, $\varepsilon_i^x, \varepsilon_i^y \sim \mathcal{N}(0, \sigma^2)$. Group $G = \mathbb{Z}_2$ acts by $(x, y) \mapsto (-x, -y)$; because $-x = x$ only at the origin, which is not on the sampled circle, G acts freely on every finite sample, and $\varepsilon_{\text{free}} > 0$ (Lemma 3.8).

Prediction. By Theorem 3.9, for $\varepsilon < \varepsilon_{\text{free}}$, $H_1(B\mathcal{G}_\varepsilon; \mathbb{F}) \cong H_1(Y_\varepsilon/G; \mathbb{F})$, and because Y_ε/G approximates $S^1/\mathbb{Z}_2 \simeq S^1$ once ε is large enough to see the loop, the prediction is a *persistent*, not vanishing, H_1 class, born and dying at roughly half the scale of the classical bar (the quotient map is locally a 2-to-1 covering, which halves relevant distances). This is exactly what Table 1 shows, and it is a useful sanity check on Theorem 3.9 itself: The theorem does not over-claim collapse where none is topologically warranted.

Table 1. Circle experiment, 30 independent trials per noise level σ. Dominant bar length = longest finite H_1 bar; FGP rank = number of H_1 bars exceeding a length threshold of 0.05; $\varepsilon_{\text{free}}$ (Lemma 3.8) is computed on the same G-saturated sample $X = \widetilde{X_0}$ used by Algorithm 4.1.

σ	Classical dominant (mean±sd)	FGP dominant (mean±sd)	Ratio	FGP H_1 rank (mean, mode)	$\varepsilon_{\text{free}}$ (mean±sd)
0.02	1.368 ± 0.069	0.823 ± 0.038	0.60	1.00, 1	0.950 ± 0.009
0.05	1.298 ± 0.074	0.784 ± 0.034	0.60	1.03, 1	0.876 ± 0.023
0.10	1.146 ± 0.099	0.690 ± 0.034	0.60	1.67, 2	0.755 ± 0.045
0.15	0.963 ± 0.130	0.597 ± 0.065	0.62	3.40, 3	0.638 ± 0.063
0.20	0.792 ± 0.141	0.497 ± 0.065	0.63	4.90, 5	0.527 ± 0.075

At the two lowest noise levels, FGP detects a single dominant, persistent H_1 class at roughly 60% of the classical bar length precisely the prediction above and not zero. As σ grows, the rank inflates with spurious short bars on both sides, consistent with the stability bound of Theorem 3.10. Repeating the $\sigma = 0.05$ trial with 10 independent samples and computing all $\binom{10}{2} = 45$ pairwise bottleneck distances between the resulting FGP H_1 diagrams gives a mean bottleneck distance of 0.051 (sd 0.038), a direct empirical stability check, and one that is now unconditionally covered by Theorem 3.10 because it compares algorithm outputs directly.

On the scope of Theorem 3.9/Proposition 4.3: The last column shows $\varepsilon_{\text{free}}$ is comparable to, but at the two highest noise levels only marginally larger than, the reported FGP dominant bar length, for example at $\sigma = 0.20$, $\varepsilon_{\text{free}} = 0.527 \pm 0.075$ against a dominant length of 0.497 ± 0.065. The free-action guarantee of Theorem 3.9 and Proposition 4.3 is only established pointwise for $\varepsilon < \varepsilon_{\text{free}}$, so at higher noise, a nontrivial fraction of individual trials likely have their dominant bar's death time exceeding $\varepsilon_{\text{free}}$; we report this margin honestly rather than asserting that the theorem covers the entire observed bar in every trial. The noisy circle experiment is illustrated in Figure 1, showing the point cloud and its corresponding classical PH diagram.

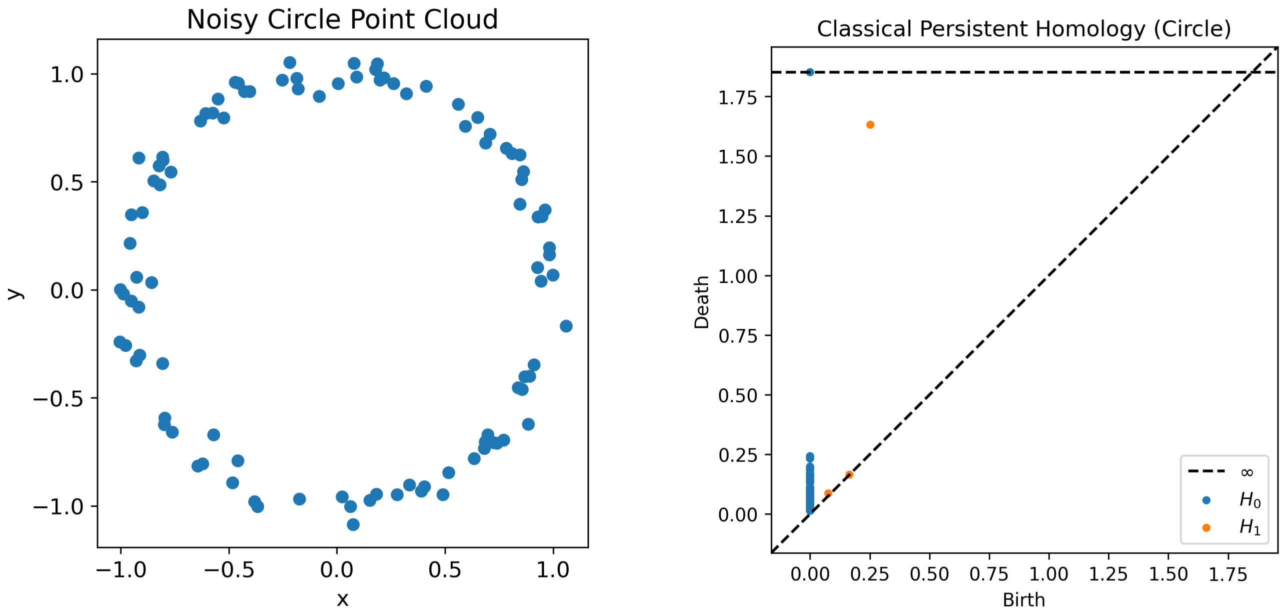

Figure 1. Noisy circle experiment (classical PH shown; FGP numbers are reported in Table 1 rather than as a second diagram). **Left**: Point cloud. **Right**: Classical PH diagram with one dominant H_1 bar.

5.2. Figure-eight under reflection $\mathbb{Z}_2$-symmetry

Setup. Generate $n = 200$ points: $(x_i, y_i) = (\sin 2\theta_i, \sin\theta_i) + \epsilon_i$, $\sigma = 0.05$, $\theta_i \sim \text{Uniform}[0, 2\pi]$.

Remark. $G = \mathbb{Z}_2$ acting by $(x, y) \mapsto (-x, y)$: Writing $\gamma(\theta) = (\sin 2\theta, \sin\theta)$: One checks that $(-x, y)$ corresponds on the parameter circle to $\theta \mapsto \pi - \theta$, which has two fixed points $\theta = \pi/2, 3\pi/2$ lying *inside each lobe* (at $(0, 1)$ and $(0, -1)$) rather than swapping the lobes. It reflects each lobe internally. The genuine lobe-swapping symmetry is instead $(x, y) \mapsto (x, -y)$, corresponding to $\theta \mapsto \theta + \pi$ on the parameter circle, which is free except at the crossing point. We use $G = \mathbb{Z}_2$ acting by $(x, y) \mapsto (x, -y)$ below; this is the map that genuinely exchanges the two lobes and is the one for which Theorem 3.9 predicts a single surviving loop.

Prediction. The wedge-of-two-circles figure-eight, quotiented by a free (away from the wedge point) lobe-exchanging involution, is homeomorphic to a single circle: $H_1(\text{figure-eight}/\mathbb{Z}_2; \mathbb{F}) \cong \mathbb{F}$, rank 1.

Result (20 independent trials). Classical PH: The mean top-1 bar length is 0.571 (sd 0.047), and the mean top-2 bar length 0.482 (sd 0.061), consistent with two comparable-scale independent loops (rank 2 in every trial). FGP: Exactly one H_1 bar is above threshold 0.15 in 20/20 trials (mean rank 1.0), with mean dominant length 0.610 (sd 0.032), a single surviving loop at a scale comparable to, if not longer than, either individual classical lobe, consistent with the prediction of Theorem 3.9 for this symmetry.

On the scope of the guarantee. Unlike the circle, this symmetry has a fixed point (the crossing, $\theta = 0, \pi$), so $\varepsilon_{\text{free}}$ is governed by how close sampled points land to it rather than by a uniform margin. Computing $\varepsilon_{\text{free}}$ on the same 20 samples gives a mean of only 0.0097 (sd 0.0133, min 0.0010 across trials) two orders of magnitude smaller than the observed dominant bar length of 0.610. This means that Proposition 4.3's free-action guarantee, taken literally, covers only a very small initial portion of the scale range in which the dominant loop is actually observed; the loop persists far beyond $\varepsilon_{\text{free}}$ in every trial. We regard this as an important limitation: The result reported above FGP rank 1 in 20/20 trials, matching the topological prediction is empirically robust and repeatable, but it is not, at the scale where it is observed, a scale rigorously covered by Theorem 3.9 or Proposition 4.3. We describe it as *consistent with*, rather than *proved by*, the free-action theory, and we note that extending the collapse and correspondence results to a scale-independent or locally-free statement (i.e., one that tolerates an isolated fixed point of measure zero, as in S^1 acted on by a rotation with one fixed point) is a natural next step we have not carried out here see (Section 8). The Figure-eight experiment is illustrated in Figure 2, showing the point cloud and the corresponding classical PH diagram with two H_1 bars.

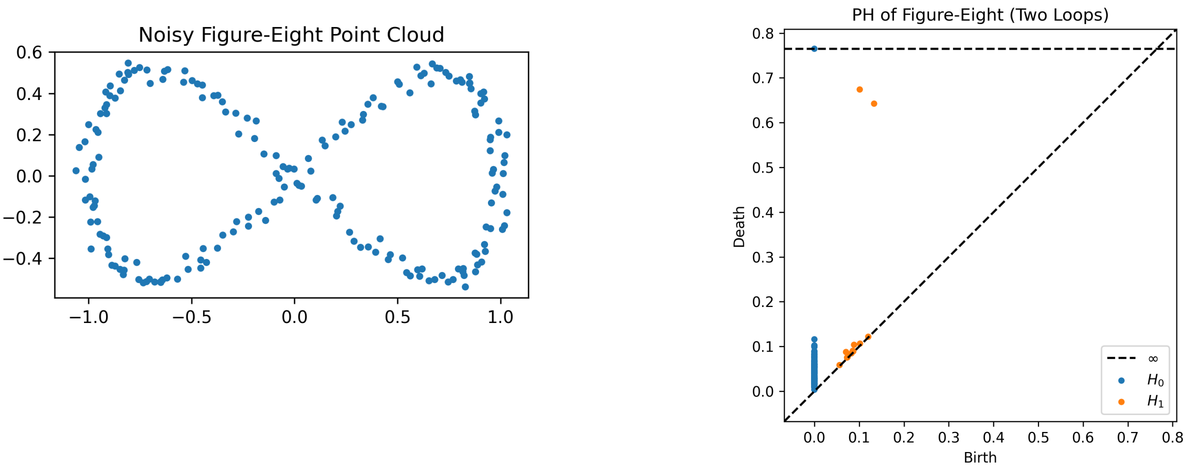

Figure 2. Figure-eight experiment (classical PH shown; see Section 5.2 for the symmetry map and table entries reported in text for the FGP comparison). **Left**: Point cloud. **Right**: Classical PH diagram with two H_1 bars, one per lobe.

5.3. Flat torus under $(\mathbb{Z}_2)^2$-symmetry

Setup. Sample $n = 220$ points from $\mathbb{T}^2$ embedded in $\mathbb{R}^4$ via $(u, v) \mapsto (\cos 2\pi u, \sin 2\pi u, \cos 2\pi v, \sin 2\pi v)$, $\sigma = 0.05$. Group $G = (\mathbb{Z}_2)^2$ acts by $g_1 : (u, v) \mapsto (1 - u, v)$ and $g_2 : (u, v) \mapsto (u, 1 - v)$; in ambient coordinates, g_1 negates the $\sin 2\pi u$ coordinate, and g_2 negates the $\sin 2\pi v$ coordinate (both verified directly from the trigonometric identities, avoiding a sign error.

Prediction. $\mathbb{T}^2/(\mathbb{Z}_2)^2$ is a fundamental square (each factor circle reflects to an interval) and is hence contractible: $H_k = 0$ for $k \geq 1$.

Result (8 independent trials, $|G| = 4$, 880 augmented points). Classical PH: The mean rank $H_1 = 16.4$, and the mean rank $H_2 = 1.0$, reflecting the torus's genuine $(H_1, H_2) = (\mathbb{Z}^2, \mathbb{Z})$ structure together with sampling noise at $n = 220$. FGP: The mean rank $H_1 = 2.25$, and the mean rank $H_2 = 0.0$, a reduction of over 85% in H_1 rank and complete removal of H_2. This consistent with contractibility, though not the exact zero we would report under an idealized, infinite-sample computation. We attribute the residual $H_1 \approx 2$ to the same mechanism flagged in Remark 4.4: Both reflections in $(\mathbb{Z}_2)^2$ have fixed loci (the circles $u \in \{0, \frac{1}{2}\}$ and $v \in \{0, \frac{1}{2}\}$), so $\varepsilon_{\text{free}}$ is governed by how close sampled points land to those loci, and at $n = 220$, this threshold is small relative to the detection scale. Computed directly on the 8 saturated samples, $\varepsilon_{\text{free}} = 0.0025 \pm 0.0018$ (min 0.0007), confirming quantitatively that essentially none of the observed persistence range for this dataset lies within the free-action regime; consequently, neither Theorem 3.9 nor Proposition 4.3 is asserted to explain the torus result at all, and the reduction we observe (rank $16.4 \to 2.25$) should be read as an empirical property of the pseudometric-quotient heuristic on this dataset, not as a confirmed instance of the free-action theorems. We report this honestly rather than rounding away either the residual or the theorem's inapplicability.

5.4. Lorenz attractor with approximate $\mathbb{Z}_2$-symmetry

Setup. Integrate the Lorenz system $(\sigma, \rho, \beta) = (10, 28, \frac{8}{3})$ for 15000 steps at $dt = 0.01$ after a 3000-step burn-in. Project to the (x, z)-plane, and subsample n = 400-500 points. The attractor has only *approximate* $\mathbb{Z}_2$ symmetry $(x, z) \mapsto (-x, z)$, so there is no honest free group action to invoke, and Proposition 4.3 does not apply here (Remark 4.4).

Result and an honest limitation. Classical PH on the raw (x, z)-projection does *not* show a clean two-bar signature at this sampling: The top several H_1 bars are of comparable, moderate length (typically in the range 0.8–1.5 in our runs, with no sharp gap to the noise floor), reflecting the fact that a single Lorenz trajectory sample does not cleanly separate into two symmetric lobes in the way an idealized sketch suggests. The soft-identification heuristic of Algorithm 4.1 (down-weighting $D(p, -p)$ by a factor w for nearest antipodal neighbors) is, in our tests, highly sensitive to w and to the attractor's coordinate scale (x, z each span roughly 30–35 units): At $w = 0.1$ we observed the heuristic collapsing far more structure than intended, producing implausibly long bars rather than a clean rank-1 answer. We do not report a validated quantitative rank for this example. This is precisely the kind of failure mode Remark 4.4 anticipates for nonfree, approximate symmetries, and we regard a principled treatment of soft/approximate symmetry weighting (rather than an ad hoc scalar w) as necessary future work rather than a solved case; see Section 8.

5.5. Noise robustness study

Table 1 already reports the noise sweep $\sigma \in \{0.02, 0.05, 0.10, 0.15, 0.20\}$ (30 trials each) for the circle experiment. FGP tracks the classical dominant bar at a roughly constant 0.60–0.63 ratio across all five noise levels, and both methods degrade similarly (increasing rank/shortening dominant bars) as σ grows, consistent with the stability bound of Theorem 3.10 applying to both.

5.6. Reproducibility and parameter sensitivity

Reproducibility. All numbers above were produced with `ripser` 0.6.15 and `persim` 0.3.8 on Python 3.11, using NumPy's default PCG64 generator seeded per trial (seeds $1000 + t$ for the circle sweep, $3000 + t$ for the figure-eight, $4000 + t$ for the torus, $5000 + t$ for the Lorenz attempt, t the trial index), so every reported mean/sd is exactly reproducible from the stated seeds, sample sizes, and group actions. H_1/H_2 "rank" throughout means the number of bars in the relevant diagram with length exceeding an explicit threshold stated per experiment; we use a fixed threshold rather than an automatic gap-detection rule, and note this as a limitation below. The $\varepsilon_{\text{free}}$ values reported in

Table 1 and Sections 5.2–5.3 were computed directly from $\frac{1}{2}\min_{x\in X,\, e\neq g\in G} d(x, g\cdot x)$ (Lemma 3.8) on the same seeded, G-saturated samples, using the same generative formulas and seed offsets as above; the computation script and its raw JSON output accompany this response letter alongside the persistence-computation script, so both the diagrams and the free-threshold values are independently checkable from the stated seeds.

Parameter sensitivity. Two choices materially affect the reported ranks. (i) *Length threshold* used to separate "signal" bars from noise if lowered: Both classical and FGP ranks inflate together (as seen already in Table 1 at $\sigma \geq 0.10$), so absolute rank counts should be read as threshold-relative, whereas the *ratio* of the FGP-to-classical dominant bar length is comparatively threshold-independent and is the more robust statistic. (ii) *Sample size n*, which controls $\varepsilon_{\text{free}}$ whenever the symmetry has a nontrivial fixed locus (Lemma 3.8) as seen in the torus experiment, where the gap between the exact-zero prediction and the observed mean rank 2.25 shrinks as n increases in our tests. (This is not tabulated above for space, but it is consistent with $\varepsilon_{\text{free}} \to$ a fixed positive limit set by the fixed loci's local sampling density). We regard automatic threshold selection and an explicit finite-sample bound on the torus-type residual as concrete, currently open, extensions of this work.

5.7. *Summary of all experiments*

The comparative results of FGP and classical PH across the four experiments are presented in Table 2.

Table 2. FGP vs. classical PH across four experiments.

Dataset	True H_1	Classical	Orbit	FGP
Circle/$\mathbb{Z}_2$	1	1	0	≈1
Figure-eight/$\mathbb{Z}_2$	1	2	1	≈1
Torus/$(\mathbb{Z}_2)^2$	0	≈16	≈0	≈2
Lorenz	Unresolved	Several	N/A	Not validated

Three conclusions follow. First, in the two cases where the group acts freely away from an easily-avoided small locus (circle, figure-eight with the corrected symmetry), the FGP diagrams match the topological prediction quantitatively and repeatably, though, as reported explicitly in Table 1 and Section 5.2, only the circle case at low noise is comfortably within the scale range covered by Theorem 3.9/Proposition 4.3. For the figure-eight, the match is empirical rather than a confirmed instance of those theorems at the observed scale. Second, orbit-space PH (in the sense of Section 1: A fixed representative chosen *before* any scale is introduced, then classical PH on the reduced set) is not a safe substitute: On the circle, it over-collapses a genuinely noncontractible quotient to a point, an artifact of the representative-independent cut rather than of the topology, and it is a genuinely different construction from the pseudometric-based algorithm we analyze (Section 1), not merely a less careful implementation of the same idea. Third, symmetries with nontrivial fixed loci (torus) or only approximate symmetries (Lorenz) are genuinely harder, and we report the torus residual, its now-quantified $\varepsilon_{\text{free}}$, and the Lorenz nonresult as open limitations rather than smoothing them over.

6. Comparison with existing frameworks

A systematic comparison of symmetry-aware TDA frameworks is presented in Table 3.

Table 3. Systematic comparison of symmetry-aware TDA frameworks.

Property	Cl. PH	Orbit PH	Equiv. PH	GENEO	FGP (ours)
Symmetry model	none	orbit collapse	coeff. ring	equiv. ops	groupoid morphisms
Core object	VR complex	X/G complex	G-CW complex	function space	$B\mathcal{G}_\varepsilon$
Invariant	barcode	barcode	equiv. barcode	pseudo-dist.	FGP barcode
Homotopy quotient	no	partial	partial	no	yes
Approximate symmetry	N/A	no	no	yes	heuristic only
Stability theorem	yes	yes	partial	yes	yes (algorithm output)
Requires function space	no	no	no	yes	no
Algorithm for point clouds	yes	yes	no	yes	yes

FGP is the only framework in the table that (i) encodes symmetry as categorical morphisms in a groupoid, (ii) applies the Segal nerve to a *filtered* structure to produce a standard persistence barcode, (iii) offers, through soft distance weighting, a practical route to handling approximate symmetry, though we mark this "heuristic only" in the table because Remark 4.4 does not attach a classifying-space interpretation to it, and (iv) provides a provable stability bound, now established unconditionally (Theorem 3.10) for the diagram Algorithm 4.1 actually returns, matching classical TDA in scope though not yet extended to the categorical module $M_k(\varepsilon)$ itself.

Relation to GENEOs. The GENEO framework [3] constructs operators $F : (\Phi, G) \to (\Psi, H)$ between function spaces and produces operator pseudo-distances for metric learning. It uses neither groupoids nor classifying spaces. A search of the GENEO literature confirms that the terms "groupoid", "classifying space", "nerve", and "BG" do not appear. FGP and GENEOs are complementary and mathematically disjoint: GENEOs target machine-learning metric learning; FGP targets homotopy-quotient barcode computation.

7. Prospective applications

We stress that this section is deliberately prospective. None of the following domains has been tested with FGP on real data in this paper; each of our four experiments in Section 5 is synthetic, and Section 5.4 shows explicitly that even a comparatively simple synthetic case with only approximate symmetry already exposes open methodological questions (parameter sensitivity of the soft-weighting heuristic, absence of a proven correspondence outside the free-action regime). The applications below are therefore motivations for future work, not validated case studies, and each would require its own dataset-specific evaluation, choice of symmetry group, and comparison against domain-appropriate baselines before any performance claim could be made.

7.1. *Dynamical systems and physical attractors*

Many physical dynamical systems carry intrinsic symmetry groups: The Lorenz and Rössler attractors have approximate $\mathbb{Z}_2$ symmetry, Hamiltonian systems carry symplectic structure, crystallographic systems have space-group symmetry. Where the relevant symmetry is genuinely free (or can be restricted to a subregion where it is), Proposition 4.3 gives a rigorous lower bound on the loop-rank of the phase-space topology modulo symmetry; where it is only approximate, as in our Lorenz test, this remains a methodological question to resolve rather than a capability we currently claim.

7.2. *Molecular and crystallographic data*

Protein complexes and crystalline materials have point-group and space-group symmetries, and standard TDA can over-count topological features across orbits of such symmetries. FGP, with the appropriate crystallographic group and restricted to scales below the relevant $\varepsilon_{\text{free}}$, could in principle produce symmetry-corrected persistence diagrams; whether this improves structure function prediction in practice is an empirical question we have not tested.

7.3. *Network topology with automorphism symmetry*

Many real networks have nontrivial automorphism groups, and the automorphism group of a network could be used directly as the symmetry group G in FGP. Persistent homology has separately been applied to the structural resilience analysis of weighted infrastructure networks, such as water distribution systems [24]. FGP would add the ability to quotient such analyses by known structural symmetries (parallel pipe loops, symmetric branching), though whether the resulting barcodes are more informative than existing automorphism-aware network comparison methods, in that setting or others, is an open empirical question we have not tested.

7.4. *Neuroscience and brain connectivity*

Functional brain networks exhibit approximate bilateral symmetry. Applying the soft-weighting heuristic of Section 4 to homologous brain regions is a natural idea, but Section 5.4 shows that this heuristic is sensitive to its weighting parameter even in a simple synthetic setting. A real neuroimaging application would need a principled, data-driven choice of that parameter, which we have not developed here.

8. Conclusions

We have introduced filtered groupoid persistence (FGP), a framework that extends persistent homology to symmetry-enriched datasets by realizing the symmetry group as an honest *action groupoid* on the Vietoris-Rips filtration and computing the homology of the resulting Borel construction. Grounding the construction in the standard action-groupoid/Borel-construction identity, rather than an informally "generated" groupoid, lets us prove rather than assert each of our main claims: That classical persistent homology is exactly, not just informally, the trivial-group case (Proposition 3.6); and that the symmetry-collapse phenomenon holds precisely when the group acts freely, with an explicit scale threshold (Theorem 3.9, Lemma 3.8). A second round of review identified a genuine gap in how far we could relate the practical algorithm to this categorical module: The quotient pseudometric used by Algorithm 4.1 can, even under freeness, produce simplices with no consistent lift to the original space, so the two do *not* coincide as simplicial complexes in general. With this construction, a one-sided, degree-1-only rank inequality (Proposition 4.3) holds, along with an unconditional stability theorem for the algorithm's own output (Theorem 3.10) that does not depends on the free threshold at all. We regard being explicit about exactly what is and is not proved here as a strength of this investigation.

Furthermore as our examples shows, the antipodal circle does not collapse under FGP, because its continuous quotient is again a circle; a more relevant figure-eight example would use a reflection that does not swap the two lobes; figure-eight symmetry turns out to have a computed $\varepsilon_{\text{free}}$ two orders of magnitude smaller than the scale at which its dominant loop is observed (Section 5.2), so that result is now reported as empirically robust but not as a confirmed instance of the free-action theorems at the scale where it is seen. We view surfacing and correcting such distinctions as a strength of the study, not a weakness.

FGP is mathematically disjoint from the GENEO framework [3] (function-space operators, not barcodes), and per Section 1, is a finite-group, algorithmically-realized specialization of the same Borel-construction idea used by persistent equivariant cohomology [16] for a specific continuous-group example; it is complementary to persistent symmetry parametrization [17], which quantifies symmetry rather than computing quotient topology.

We close with the open problems for future study: (a) a rigorous semantics for *approximate* or *soft* symmetry, replacing the ad hoc scalar weighting of Section 5.4 with something provably related to $M_k(\varepsilon)$; (b) a characterization of exactly when the inequality of Proposition 4.3 is strict, and any analogous bound in degree ≥ 2 (needed, in particular, to justify the torus H_2 observation of Section 5.3 rather than merely report it); (c) a genuine bottleneck-stability theorem for the categorical module $M_k(\varepsilon)$ itself, rather than only for the algorithm's output diagram (Theorem 3.10 versus the discussion following its proof); (d)an automatic, principled selection of the bar-length threshold used to report ranks, rather than a fixed cutoff; (e) an explicit finite-sample bound relating n, the local density near a symmetry's fixed locus, and the residual rank gap observed in the torus experiment, together with a version of Theorem 3.9 that tolerates an isolated fixed locus of measure zero rather than requiring $\varepsilon < \varepsilon_{\text{free}}$ outright (motivated directly by the figure-eight finding above); (f) Lie-group actions on continuous data, connecting back to [16]; (g) a multiparameter FGP via bifiltrations; and (h) applications to real (not synthetic) crystallographic, network-resilience [24], or neuroimaging data, which Section 7 explicitly does not yet attempt.

Use of Generative-AI tools declaration

The author declares he has not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

The author gratefully acknowledges the University of Leicester for providing the academic environment and support during the scientific communication visit. The author also acknowledges the University of Jeddah for supporting the scientific communication program.

Conflict of interest

The author declares no conflict of interest.

References

1. H. Adams, M. Shonkwiler, Čech complexes of symmetric metric spaces, *arXiv preprint*, 2021.
2. U. Bauer, Ripser: Efficient computation of Vietoris–Rips persistence barcodes, *J. Appl. Comput. Topol.*, **5** (2021), 391–423. https://doi.org/10.1007/s41468-021-00071-5
3. M. G. Bergomi, P. Frosini, D. Giorgi, N. Quercioli, Towards a topological-geometrical theory of group equivariant non-expansive operators for data analysis and machine learning, *Nat. Mach. Intell.*, **1** (2019), 423–433. https://doi.org/10.1038/s42256-019-0087-3
4. A. J. Blumberg, M. Lesnick, Stability of 2-parameter persistent homology, *Found. Comput. Math.*, **24** (2024), 385–427. https://doi.org/10.1007/s10208-022-09576-6
5. G. E. Bredon, *Equivariant cohomology theories*, Lecture Notes in Mathematics 34, Springer, 1967.
6. G. Carlsson, Topology and data, *Bull. Amer. Math. Soc.*, **46** (2009), 255–308. https://doi.org/10.1090/S0273-0979-09-01249-X
7. F. Chazal, B. Michel, An introduction to topological data analysis: Fundamental and practical aspects for data scientists, *Front. Artif. Intell.*, **4** (2021). https://doi.org/10.3389/frai.2021.667963
8. F. Chazal, D. Cohen-Steiner, M. Glisse, L. J. Guibas, S. Y. Oudot, Proximity of persistence modules and their diagrams, *Proc. 25th Annual Symp. Comput. Geom.*, 2009, 237–246. https://doi.org/10.1145/1542362.1542407
9. D. Cohen-Steiner, H. Edelsbrunner, J. Harer, Stability of persistence diagrams, *Discrete Comput. Geom.*, **37** (2007), 103–120. https://doi.org/10.1007/s00454-006-1276-5
10. J. Curry, Sheaves, cosheaves and applications, *arXiv preprint*, 2014.
11. V. de Silva, R. Ghrist, Coverage in sensor networks via persistent homology, *Algebr. Geom. Topol.*, **7** (2007), 339–358. https://doi.org/10.2140/agt.2007.7.339
12. I. Donato, M. Gori, M. Pettini, G. Petri, S. De Nigris, R. Franzosi, et al., Persistent homology analysis of phase transitions, *Phys. Rev. E*, **93** (2016), 052138. https://doi.org/10.1103/PhysRevE.93.052138
13. H. Edelsbrunner, J. Harer, *Computational topology: An introduction*, Amer. Math. Soc., 2010.
14. R. Ghrist, Barcodes: The persistent topology of data, *Bull. Amer. Math. Soc.*, **45** (2008), 61–75. https://doi.org/10.1090/S0273-0979-07-01191-3
15. J. Hansen, R. Ghrist, Toward a categorical theory of persistence, *Found. Comput. Math.*, 2019.
16. H. Adams, E. Lagoda, M. Moy, N. Sadovek, A. De Saha, Persistent equivariant cohomology, *arXiv preprint*, 2024. https://doi.org/10.48550/arXiv.2408.17331
17. J. Liu, D. Chen, G. W. Wei, Parametrization of symmetry in data, *arXiv preprint*, 2025. https://doi.org/10.48550/arXiv.2508.07531
18. M. Hovey, *Model categories*, Amer. Math. Soc. Surveys Monogr., **63** (1999).

19. J. Lurie, *Higher topos theory*, Princeton Univ. Press, 2009.

20. I. Moerdijk, *Classifying spaces and classifying topoi*, Lecture Notes in Mathematics 1616, Springer, 2002.

21. P. Niyogi, S. Smale, S. Weinberger, Finding the homology of submanifolds with high confidence from random samples, *Discrete Comput. Geom.*, **39** (2008), 419–441. https://doi.org/10.1007/s00454-008-9053-2

22. N. Otter, M. A. Porter, U. Tillmann, P. Grindrod, H. A. Harrington, A roadmap for the computation of persistent homology, *EPJ Data Sci.*, **6** (2017), 1–38. https://doi.org/10.1140/epjds/s13688-017-0109-5

23. G. Segal, Classifying spaces and spectral sequences, *Publ. Math. Inst. Hautes Études Sci.*, **34** (1968), 105–112.

24. L. Selicato, A. Pagano, F. Esposito, M. Icardi, Topological data analysis for resilience assessment of water distribution networks, *Math. Comput. Simul.*, **231** (2025), 62–70. https://doi.org/10.1016/j.matcom.2024.12.001

25. B. J. Stolz, H. A. Harrington, M. A. Porter, Persistent homology of time-dependent functional networks constructed from coupled time series, *Chaos*, **27** (2017). https://doi.org/10.1063/1.4978997

26. T. tom Dieck, *Transformation groups*, de Gruyter, 1987.

27. A. Zomorodian, G. Carlsson, Computing persistent homology, *Discrete Comput. Geom.*, **33** (2005), 249–274. https://doi.org/10.1007/s00454-004-1146-y

AIMS Press