Stable Approximate Canonicalization: Stability Obstructions, Selector Ambiguity, and Data-Dependent Trade-offs

Anonymous authors
Paper under double-blind review

Abstract

Canonicalization obtains invariant predictors by selecting one representative from each symmetry orbit, but a continuous exact selection is often topologically impossible. We show that approximation does not remove this obstruction without cost. On any compact obstructed domain, deterministic canonicalizers whose invariance and orbit errors vanish must have unbounded Lipschitz constants. For free finite-group actions on compact connected smooth manifolds, the optimal expected orbit error of an exactly invariant L-Lipschitz relaxation is $\Theta(1/L)$ under every density bounded above and below. An antipodally symmetric population fitted to TUM VI orientations supports this rate: over the reported fit range $L \geq 17$, an explicit quaternion relaxation has a fitted log–log slope -1.017, with L times the error approaching a nonzero plateau. We then show that continuous relaxations can localize their error near construction-specific selector-ambiguity strata, making data mass near those strata the relevant practical quantity. Experiments on LiDAR tracks, molecules, histopathology, and particle clouds show that occupancy, representation instability, and downstream harm are distinct. These results turn a qualitative obstruction into a quantitative, data-dependent account of when canonicalization is useful and what it costs.

1 Introduction

Canonicalization is a simple recipe for symmetry: map every input to a chosen member of its orbit and apply an ordinary predictor. It underlies pose normalization, principal-axis alignment, sign conventions, sorting, and many learned invariant pipelines. Its appeal hides a basic problem. For common actions—including antipodal identification and frame selection—no globally continuous exact representative exists. An implementation must tear, cease to be orbit-valued, or abandon exact invariance.

The obstruction is usually stated as a binary topological fact. Learning systems, however, use approximate maps on nonuniform data, so the useful questions are quantitative: can the defect be made small at a fixed stability budget, how quickly does it decay as that budget grows, and where in the input space is the cost paid? We answer these questions in a common framework.

First, compactness upgrades nonexistence to a stability theorem: if both invariance and orbit fidelity approach zero, the Lipschitz constants must diverge. Second, for a broad smooth finite-group setting, we identify the population rate exactly up to constants, $\Theta(1/L)$. Third, for concrete selectors with ambiguity sets, we construct relaxations whose error is confined to a thin neighborhood of the selector-ambiguity strata; expected error is therefore controlled by the population mass in that neighborhood.

Our experiments test these claims rather than rank architectures. Real TUM VI orientations provide a smooth, full-dimensional antipodally symmetric population on S^3 and a direct test of the $1/L$ law. Waymo tracks expose how eigenvalue and sign selector ambiguities produce temporal tearing or collapse. SPICE/MD17, PatchCamelyon, and JetClass show that high occupancy can be benign while rare events can be severe. The message is precise: topology

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

determines the unavoidable trade-off, geometry determines its rate, and the data and task determine its practical importance.

2 Related work

Recent work formalizes canonicalization and frames as routes to invariant or equivariant models (Dym et al., 2024; Kaba et al., 2023; Ma et al., 2024); probabilistic frames replace a single representative with a distribution (Kim et al., 2023). Regularity results establish when continuous canonical forms can or cannot exist (Tahmasebi and Jegelka, 2025). We complement these structural results by relating approximation error to a stability budget and by isolating the data mass on which the compromise is necessary.

Group convolution and averaging avoid deterministic selection but have different computational or information trade-offs (Cohen and Welling, 2016; Puny et al., 2022). Smooth sorting and spectral perturbation theory similarly replace discontinuous choices with stable surrogates (Prillo and Eisenschlos, 2020; Blondel et al., 2020; Yu et al., 2015). Our results apply at the representation interface: they distinguish invariance, membership in the original orbit, and Lipschitz stability, independently of the downstream architecture.

Our principal new claim relative to these structural results is the optimization law in Theorem 2: a matching population lower and upper rate as a function of a Lipschitz budget. We do not claim novelty for the qualitative obstruction itself.

3 From obstruction to a quantitative rate

Let a compact group G act continuously on a metric space V, and let $K \subset V$ be compact and G-invariant. A canonicalizer $C : K \to K$ is *invariant*, $C(gx) = C(x)$, and *orbit-valued*, $C(x) \in Gx$. Writing $\pi : K \to K/G$ for the quotient map makes the obstruction transparent.

Proposition 1 (Canonicalization is a section)**.** *A continuous canonicalizer on K exists if and only if π admits a continuous section.*

Thus, whenever the quotient bundle has no section, exact continuous canonicalization is impossible. To quantify approximation, for continuous $f : K \to K$ define

$$\mathrm{Inv}_K(f) = \sup_{x \in K, g \in G} d(f(gx), f(x)), \qquad \mathrm{Orb}_K(f) = \sup_{x \in K} d(f(x), Gx).$$

The first term measures invariance; the second measures whether the output remains a representative of the input orbit.

Theorem 1 (Stable approximation)**.** *If continuous maps $f_n : K \to K$ are equicontinuous and $\mathrm{Inv}_K(f_n) \to 0$, $\mathrm{Orb}_K(f_n) \to 0$, then a subsequence converges uniformly to a continuous exact canonicalizer.*

The proof is short: Arzelà–Ascoli supplies a uniformly convergent subsequence, and both defects pass to its limit. Hence approximation cannot evade a genuine obstruction while retaining uniform stability.

Corollary 1 (Lipschitz blow-up)**.** *If K admits no continuous canonicalizer and both defects of f_n vanish, then $\sup_n \mathrm{Lip}(f_n) = \infty$.*

For a fixed budget define

$$\eta_K(L) = \inf_{\mathrm{Lip}(f) \le L} \max\{\mathrm{Inv}_K(f), \mathrm{Orb}_K(f)\}.$$

The same compactness argument gives a finite-budget form.

Corollary 2 (Positive error floor)**.** *On every compact obstructed domain, $\eta_K(L) > 0$ for every finite L.*

The next result changes the approximation problem in an essential way. The preceding defects are uniform and the maps are $K \to K$; the population result imposes exact invariance,

permits ambient-valued maps, and measures *expected* orbit error. In the free finite setting, this distinction is unavoidable. For the nontrivial free action of a finite group on connected K used below, $K \to K/G$ is a covering of degree $|G| > 1$ and has no continuous section: the image of a section would be a nonempty clopen sheet in connected K. Let $\Delta = \min_{x \in K, , g \neq e} |gx - x| > 0$. If a continuous map had $2\,\mathrm{Orb}_K(f) + \mathrm{Inv}_K(f) < \Delta$, every $f(x)$ would have a unique nearest point in Gx. That nearest-point choice is continuous, and the triangle inequality shows that the choices at x and gx differ by less than Δ; hence they coincide. This would give a continuous exact canonicalizer. Therefore

$$2\,\mathrm{Orb}_K(f) + \mathrm{Inv}_K(f) \geq \Delta, \qquad \max\{\mathrm{Orb}_K(f), \mathrm{Inv}_K(f)\} \geq \Delta/3,$$

independently of the Lipschitz budget. Thus the $1/L$ law below does not make both uniform defects vanish: increasing L allows a fixed-amplitude error to be concentrated on a population layer of shrinking mass.

A sharp population rate emerges for finite symmetries. Let a nontrivial finite group G act orthogonally on a finite-dimensional Euclidean space, and let K be a compact connected C^2 G-invariant manifold on which the action is free. Write d_K for the intrinsic Riemannian distance and Lip_K for the corresponding input Lipschitz constant. Compact embeddedness gives a constant $c_K < \infty$ with $|x - y| \leq d_K(x, y) \leq c_K |x - y|$; replacing d_K by ambient distance therefore changes constants but not the $1/L$ exponent. Let P have density p with respect to normalized volume, with $0 < m \leq p \leq M < \infty$. For exactly invariant maps

$f : K \to E$, set

$$\mathcal{E}_P(f) = \mathbb{E}_{X \sim P} d(f(X), GX), \qquad \eta_P^{\mathrm{inv}}(L) = \inf_{\substack{f(gx) = f(x) \\ \mathrm{Lip}_K(f) \leq L}} \mathcal{E}_P(f).$$

Theorem 2 (Finite-group population rate)**.** *Under the assumptions above, there are constants $c, C, L_0 > 0$ such that*

$$\frac{c}{L+1} \leq \eta_P^{\mathrm{inv}}(L) \quad (L \geq 0), \qquad \eta_P^{\mathrm{inv}}(L) \leq \frac{C}{L} \quad (L \geq L_0).$$

Consequently $\eta_P^{\mathrm{inv}}(L) = \Theta(1/L)$.

For the lower bound, topology forces competing local branches separated by a codimension-one bad set; Lipschitz continuity spreads the mismatch over a collar of width proportional to $1/L$, whose probability mass is of this order. For the upper bound, cut the quotient along a codimension-one skeleton, choose local lifts away from it, and interpolate across a collar of width $t \asymp 1/L$. Density bounds convert volume to probability. The appendix gives the complete geometric construction and proof.

What the rate does and does not claim. The theorem is an optimization statement, not merely an analysis of one construction. Its lower bound applies to every exactly invariant map with the stated Lipschitz budget, including maps that do not resemble a branch-cut selector. Its upper bound shows that a smoothed cut attains the same order, but neither constant is universal: orbit separation, quotient geometry, and the density bounds all enter. The assumption $p \geq m > 0$ is essential. If a population avoids every unavoidable transition region, its error can decay faster or vanish on its support; concentration near such a region instead increases the constant. The theorem concerns ambient orbit distance, not a universal conversion from representation error to prediction error, and it does not give the rate for continuous groups or changing stabilizers. Those cases motivate the localization result below.

The order cannot generally be improved. For the antipodal sharp model analyzed in the appendix, the optimized relaxation obeys

$$\lim_{L \to \infty} L\, \eta^{\mathrm{inv}}(L) = \frac{1}{\pi}. \tag{1}$$

The nonzero scaled-error plateau motivates the real-data diagnostic below; the real-data experiment does not test the sharp constant $1/\pi$.

Theorem 3 (Sharp antipodal model)**.** *For the uniform model defined in Appendix A.3, the optimal error has the $1/L$ rate and asymptotic constant $1/\pi$.*

4 Localization near selector-ambiguity sets

We distinguish action singularity from selector ambiguity. The *action-singular* (nonfree) locus consists of inputs with nontrivial stabilizer. Separately, a chosen representative rule can lose uniqueness or conditioning on a construction-specific set Σ—for example at repeated PCA eigenspaces, vanishing sign-fixing moments, sorting ties, or learned-logit ties—even when the action is free. Here Σ denotes this selector-ambiguity set; we assume it is closed and G-invariant and put $R = K \setminus \Sigma$. Suppose the action is orthogonal, the origin is fixed, and an exact canonicalizer C_R is continuous on R. Let $\chi : [0, \infty) \to [0, 1]$ be continuous, with $\chi(u) = 0$ for $u \le 1$ and $\chi(u) = 1$ for $u \ge 2$, and define

$$f_t(x) = \chi\left(\frac{d(x, \Sigma)}{t}\right) C_R(x) \quad (x \in R), \qquad f_t(x) = 0 \quad (x \in \Sigma).$$

Proposition 2 (Localized relaxation)**.** *f_t is continuous and exactly invariant. It equals C_R outside the $2t$-neighborhood $N_{2t}(\Sigma)$ and is zero inside $N_t(\Sigma)$.*

The relaxation sacrifices orbit membership only where the selector becomes ambiguous. Proposition 2 proves continuity and localization, not a general Lipschitz law: continuity of C_R alone does not imply $\mathrm{Lip}(f_t) = O(1/t)$. Such a bound requires quantitative regularity, as in the specific finite-group upper construction. If $\|x\| \le B$ on K, this gives a distribution-dependent bound.

Theorem 4 (Selector-ambiguity mass control)**.** *For any random input X supported on K,*

$$\mathbb{E}\, d(f_t(X), GX) \le \mathbb{E}[\|X\| \mathbf{1}\{X \in N_{2t}(\Sigma)\}] \le B\, P(X \in N_{2t}(\Sigma))\,.$$

Corollary 3 (Score-calibrated mass control)**.** *Assume that the selector-ambiguity set is a finite union $\Sigma = \bigcup_{j=1}^{J} \Sigma_j$. For each j, let $s_j : K \to [0, \infty)$ vanish on Σ_j and be C_j-Lipschitz with respect to the metric used in $N_t(\Sigma)$. Then the localized relaxation in Proposition 2 satisfies*

$$\mathbb{E}\, d(f_t(X), GX) \le B \sum_{j=1}^{J} P\{s_j(X) < 2C_j t\}\,.$$

The corollary is one-sided: it turns calibrated score occupancy into an upper bound on the metric-neighborhood mass, but it does not identify a score threshold with a metric radius without the constants C_j.

This bound suggests two empirical quantities that should not be conflated. *Occupancy* is the mass near a selector-ambiguity stratum. *Severity* is the representation or task change caused there. A third quantity, pose-space sensitivity, need not agree with canonical-object or prediction sensitivity: an unstable frame can act almost trivially on a symmetric object, while a small representation change can cross a sharp decision boundary.

For a normalized degeneracy score s_j associated with a stratum Σ_j, we therefore report both the occupancy and conditional severity profiles

$$p_j(t) = P\{s_j(X) < t\}, \qquad r_j(t) = \mathbb{E}[\ell(X) \mid s_j(X) < t],$$

where ℓ is a mechanism-appropriate failure quantity. These profiles are not generally additive because neighborhoods of different strata can overlap. We call a stratum empirically active when it has appreciable occupancy, elevated conditional severity, or both. This separates an unavoidable but dormant ambiguity from a population-level failure mechanism.

Corollary 3 gives a precise condition under which these score-occupancy curves control the metric mass in Theorem 4. In the experiments below we do not estimate the constants C_j or a reverse score–distance comparison, so normalized scores are treated as diagnostic proxies rather than numerical measurements of $d(x, \Sigma)$.

5 Experiments

We evaluate the rate and localization claims on five domains. Thresholds and normalizations are fixed before downstream evaluation; protocol details, audit tables, and uncertainty

summaries are in the appendix. We use official TUM VI orientations (Schubert et al., 2018), Waymo Open Dataset tracks (Sun et al., 2020), SPICE and MD17 molecules (Eastman et al., 2023; Chmiela et al., 2017), PatchCamelyon (Veeling et al., 2018), and JetClass (Qu et al., 2022). Appendix C asks a separate controlled question: does enforcing or learning invariance remove representative instability, or only change how the trade-off is paid? It compares measured invariance, orbit error, sensitivity, and collapse for analytic and learned constructions. This is a mechanism audit, not a validation of Theorem 4: its bins use diagnostic scores rather than $d(x, \Sigma)$ and are not guaranteed to be jointly far from every selector-ambiguity channel.

Population	Action and scale	Geometry / stability evidence	Task-level evidence
TUM VI	$q \sim -q$; 92,536 poses, 7,786 centers	Smooth KDE slope -1.017; $LE = 0.454$ at $L = 129$	Raw trajectory retained only as a non-smooth diagnostic
Waymo vehicles	$SO(2)$; 347,775 observations	Sign occupancy 2.680%; $> 90°$ tearing 10.55%	Hard PCA ΔRecall@1 $= -0.0152 \pm 0.0053$
Waymo pedestrians	$SO(2)$; 72,183 observations	Eigengap occupancy 5.2× vehicles; tearing 28.36%	Large pose tails can be hidden by object symmetry
SPICE / MD17	$SO(3)$; 136,886 conformers / trajectory control	SPICE $P(s_{\mathrm{PCA}} < .02) = 1.329\%$; MD17 none below .2	Hard-PCA force S_F p99 545.5 versus raw 46.78
PatchCamelyon	exact C_4; 40k train, 12k test	Continuous-orientation sensitivity p95 11.77	Prediction propagation is not ordered by representation sensitivity
JetClass	permutations; 40,000 jets	$P(\Delta < 10^{-3}) =$ 78.3%; hard/soft p95 29.38/0.97	Soft sort: Δaccuracy $+0.0106 \pm 0.0055$

Table 1: **Scale and principal evidence across the real-data studies.** Metrics are domain-specific because occupancy, representation severity, and downstream consequence answer different questions.

5.1 Evaluation and uncertainty

Let $Z(x)$ be the representation passed downstream and $h(x) = \phi(Z(x))$ the prediction. For perturbations δ with $\|\delta\| \le r$, we estimate

$$\widehat{S}_r^{\mathrm{obj}}(x) = \max_{\delta} \frac{D_{\mathrm{obj}}(Z(x+\delta), Z(x))}{\max\{\|\delta\|, \epsilon_{\mathrm{num}}\}}, \qquad \widehat{S}_r^{\mathrm{pred}}(x) = \max_{\delta} \frac{D_{\mathrm{pred}}(h(x+\delta), h(x))}{\max\{\|\delta\|, \epsilon_{\mathrm{num}}\}}.$$

When a deterministic pose $a(x)$ exists, we separately replace D_{obj} by the group geodesic distance between $a(x+\delta)$ and $a(x)$. For a differentiable molecular energy model, $F_\theta(X) = -\nabla_X E_\theta(X)$, we also report the force/gradient-output diagnostic

$$\widehat{S}_r^{F}(X) = \max_{\delta \in \mathcal{A}_r(X)} \frac{\|F_\theta(X+\delta) - F_\theta(X)\|_F}{\max\{\|\delta\|_F, \epsilon_{\mathrm{num}}\}}.$$

This is not $\widehat{S}^{\mathrm{obj}}$: for raw coordinates $Z(X) = X$, the Euclidean representation quotient is one even though $\widehat{S}^F$ can be large. The finite multi-start search makes these empirical lower bounds on worst-case sensitivity rather than certificates. Temporal data additionally use natural pose increments. Relaxed frames are checked through their singular values and determinant, since a matrix can remain continuous precisely by ceasing to encode a rigid pose.

Uncertainty respects the sampling unit. TUM VI KDE curves use 10^6 independent Monte Carlo draws per bandwidth; the dependent raw trajectories use a sequence-stratified moving-block bootstrap. Waymo intervals resample complete tracks. Learned comparisons use seeds $0, 1, 2$ and are paired within seed. Unless explicitly labeled as a standard error or confidence interval, $\pm$ for a learned-model result means mean $\pm$ sample standard deviation across these

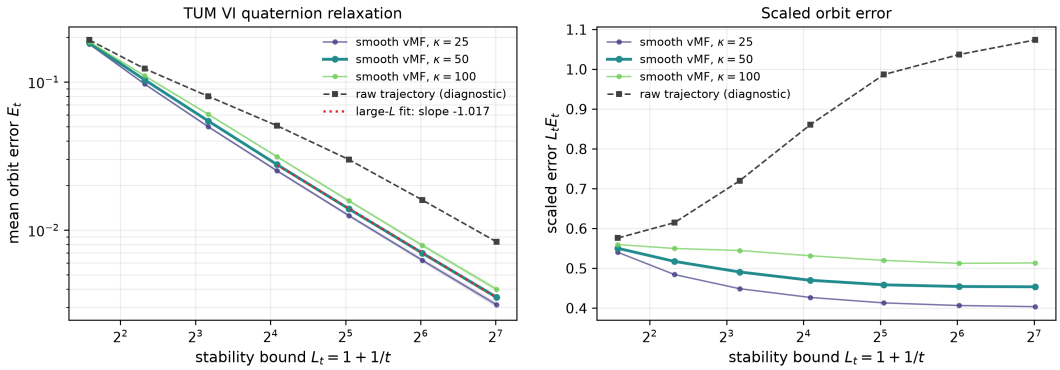

Figure 1: **TUM VI quaternion calibration.** Left: mean orbit error versus stability scale. The fit over $L \geq 17$ has slope -1.0171 (regression SE 0.0055; $R^2 = 0.99994$). Right: $L\mathbb{E}[e_t]$ approaches a nonzero plateau, consistent with $\Theta(1/L)$. Shading/error bars are Monte Carlo uncertainty.

seeds (ddof = 1); with three seeds this is descriptive, not a significance test. Exact finite actions are enumerated where feasible; arbitrary rotations of a pixel grid are not treated as exact $SO(2)$ actions, which is why PatchCamelyon uses quarter turns as its primary test. Full thresholds and protocols appear in Appendix B.

5.2 Real-data calibration of the $1/L$ law

We pool the 92,536 motion-capture quaternions from TUM VI room1–room6, renormalize them, and temporally subsample to 7,786 centers. The fitted population is the antipodally symmetric von Mises–Fisher mixture

$$p_\kappa(q) \propto \frac{1}{N} \sum_i \{e^{\kappa q^\top q_i} + e^{-\kappa q^\top q_i}\},$$

which is positive on S^3 and satisfies $p(q) = p(-q)$. From one million Monte Carlo draws, we evaluate the exactly invariant map

$$f_t(q) = \mathrm{clip}(q_w/t, -1, 1)q, \quad L_t = 1 + 1/t, \quad e_t(q) = \min\{\|f_t(q) - q\|, \|f_t(q) + q\|\}.$$

Numerically, $f_t(-q) = f_t(q)$ to machine precision.

Figure 1 is consistent with the predicted -1 exponent for this explicit family. Nearby bandwidths $\kappa \in \{25, 100\}$ give slopes -1.0270 and -1.0174; the scaled error at $L = 129$ changes from 0.404 to 0.514, showing that the constant is population-dependent. Direct evaluation on the raw trajectory gives scaled error 1.074. We report it only as a diagnostic: a finite set of trajectory samples is not the full-dimensional smooth population assumed by Theorem 2.

5.3 Which strata are active?

For PCA alignment, we normalize eigenvalue gaps by total variance and sign statistics by object scale. The active stratum depends strongly on the domain (Figure 2). At threshold 0.02, only 0.0193% of Waymo vehicles are near an eigengap but 2.680% are near a sign boundary, a 139× difference. Pedestrian eigengap occupancy is 5.2× the vehicle value. In SPICE, 1.329% of polyatomic conformers have normalized PCA gap below 0.02 and 8.16% below 0.1, while sampled MD17 conformers have none below 0.2. JetClass sits close to sorting ties: its median minimum normalized gap is 2.86×10^{-4} and 78.3% fall below 10^{-3}.

5.4 Waymo: tearing, collapse, and task geometry

On 347,775 vehicle and 72,183 pedestrian observations, hard PCA frames exhibit both rare catastrophic changes and frequent class-dependent ambiguity. Across 315,900 vehicle

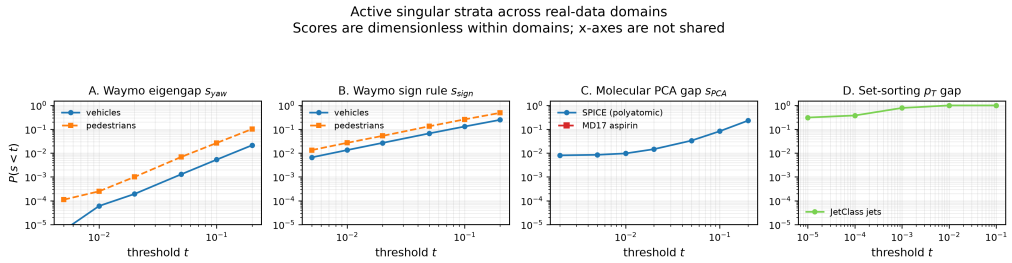

Figure 2: **Active selector-ambiguity strata are data-dependent.** Occupancy curves use domain-normalized gaps or sign margins. The same canonicalization rule can operate far from ambiguity in one population and predominantly near it in another.

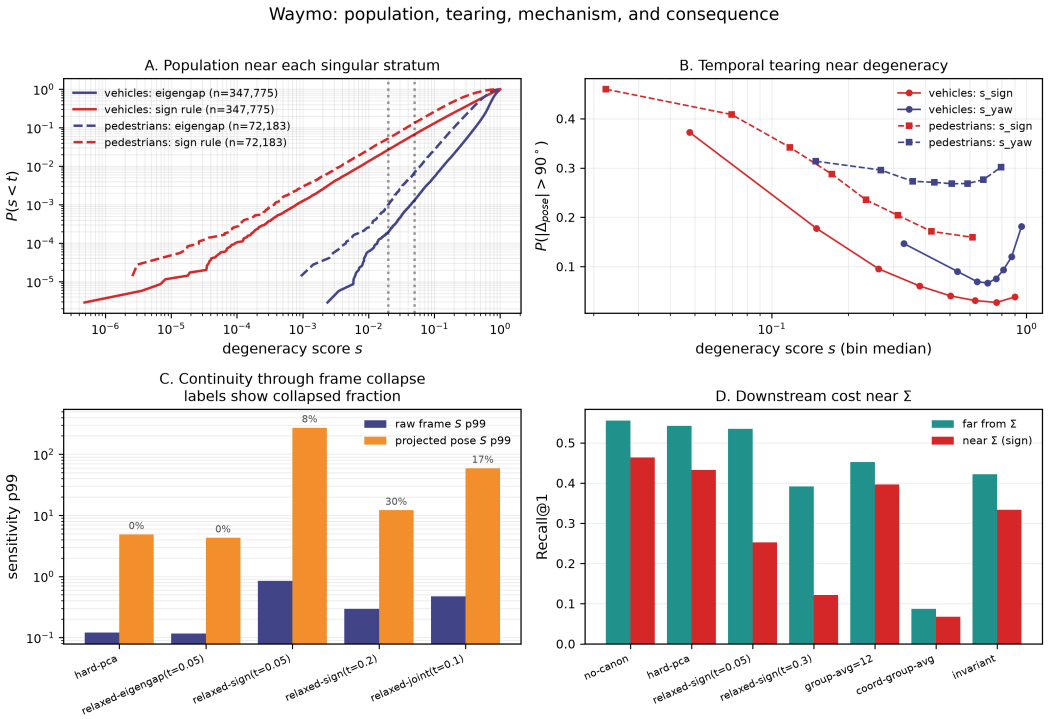

Figure 3: **Waymo diagnosis.** (A) selector-ambiguity occupancy; (B) temporal frame tearing; (C) collapse under a continuous sign relaxation; and (D) retrieval changes. Stability improves by representing ambiguity away from the original orbit, and downstream effects depend on the task.

and 64,156 pedestrian transitions, frame changes over 90° occur in 10.55% and 28.36%, respectively. For vehicles, the rate near versus far from the sign stratum is 43.05% versus 4.95% (risk ratio 8.70); for pedestrians it is 45.75% versus 19.30% (ratio 2.37).

The localized relaxation makes the alternative failure mode explicit: sign-coordinate collapse rises from 8.40% at $t = 0.05$ to 30.08% at $t = 0.2$. Pose sensitivity is not object sensitivity. Vehicle pose-space tail ratios reach 870.2, 338.2, and 880.8 for yaw, sign, and joint strata, whereas the corresponding pedestrian ratios are 1.0, 38.1, and 165.1; near-symmetric objects can hide large pose changes after transformation. Finally, retrieval performance is not monotone in representation smoothness: hard PCA changes accuracy by -0.0152 ± 0.0053 relative to no canonicalization, while a strong sign relaxation ($t = 0.3$) changes it by -0.1998 ± 0.0040. The relaxation removes a discontinuity by discarding frame magnitude, which this task uses.

5.5 Cross-domain consequences

Figure 4 separates representation-level and task-level diagnostics. In SPICE force prediction, hard PCA increases the force-sensitivity p_{99}, $\widehat{S}^F$, from 46.78 to 545.5, while force-MAE changes vary in sign across seeds $(+0.607, -0.190, +0.251)$. These are force/gradient sensitivities, not representation sensitivities. On PatchCamelyon, an exact C_4 orbit representation has p95 sensitivity 1.00, whereas a continuous orientation representation reaches 11.77; neither number alone determines accuracy. On JetClass, hard sorting has p95 sensitivity 29.38, while soft sorting at temperature 0.2 reduces it to 0.97, preserves permutation consistency at 1.0, and changes accuracy by $+0.0106 \pm 0.0055$.

Soft sorting uses a mixing matrix, so its output need not be a literal permutation of the input blocks or remain on the permutation orbit. The result supports only the task-level conclusion that the tested relaxation retains enough information for classification while reducing measured representation sensitivity.

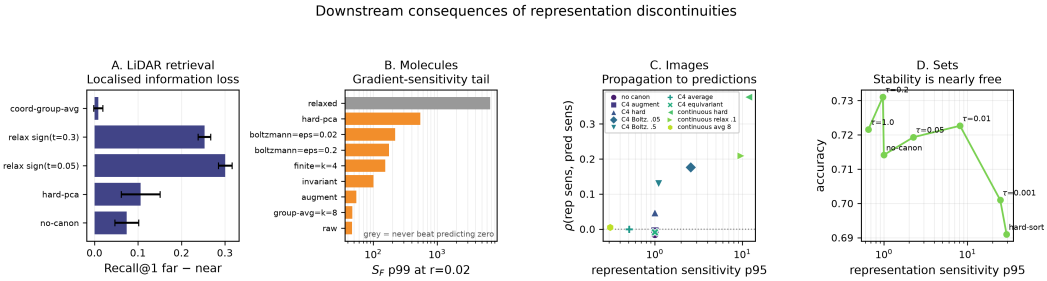

Figure 4: **Cross-domain downstream effects.** Canonicalization can improve or degrade task metrics independently of representation-level diagnostics. The molecular panel reports force/gradient sensitivity rather than S^{obj} . Panels use domain-appropriate metrics and report paired seed uncertainty where available; full numerical tables appear in the appendix.

6 Implications for design and evaluation

Canonicalization is not universally beneficial or harmful: each mechanism pays for branch ambiguity differently. Hard selectors preserve orbit membership but can tear; continuous relaxations trade tearing for off-orbit collapse; averaging or multi-frame methods trade identity or compute. Evaluation should therefore report a vector of outcomes—invariance, orbit fidelity, stability, cost, and task performance—rather than a single accuracy number. In particular, use dimensionless degeneracy scores, inspect upper tails and near/far risk, and separate pose/frame, canonical-object, and prediction sensitivity. Matched controls should keep the encoder and training budget fixed. A compact design summary is given in Appendix B.4.

The experiments illustrate why these diagnostics must remain distinct. TUM VI calibrates an attainable $1/L$ mechanism for one invariant family; it does not solve the optimization defining $\eta_P^{\text{inv}}(L)$, while the bandwidth sweep shows that the leading constant is population-dependent. Waymo shows that the most occupied ambiguity need not dominate failure, and JetClass shows that a relaxation can reduce measured representation sensitivity without literal orbit preservation. For families with an explicit $O(1/t)$ or analogous conditioning bound, t is an operating point that trades conditioning against the width of the off-orbit collar; Proposition 2 alone does not supply this law.

7 Limitations and open directions

The matching $\Theta(1/L)$ law currently requires a free finite action on a compact connected smooth manifold; it does not determine the exponent for continuous groups, changing stabilizers, noncompact domains, or general stratified spaces. The localized construction

covers a useful part of the selector-ambiguity picture, including free actions with nontrivial cuts, but gives an upper bound rather than a matching rate. Density bounds exclude lower-dimensional or transition-avoiding populations, and the theorem's constants are geometric rather than dataset-specific. Empirically, finite perturbation searches are lower bounds on local worst-case sensitivity, the trained models are diagnostic rather than leaderboard systems, and three seeds do not support significance claims. Finite/probabilistic frames and equivariant models also change the output space or computational cost and therefore lie outside the single-representative obstruction studied here.

8 Conclusion

Canonicalization on an obstructed domain cannot be simultaneously exact, invariant, and uniformly stable under arbitrarily accurate deterministic approximation. For smooth free finite-group populations this sharpens to an optimal expected orbit error of order $1/L$. TUM VI calibrates the exponent and its population-dependent constant, while construction-specific selector ambiguities localize where relaxations pay their cost. The practical conclusion is an auditable trade-off, not universal failure: hard selectors tear, relaxations may leave the orbit or collapse, and averaging or multi-frame methods trade information or computational cost. Reporting occupancy, representation severity, and task effect together makes that trade-off explicit.

AI Use Statement

Generative AI tools were used to assist with language editing, improving readability, manuscript organization, and the drafting or refactoring of selected software code. They were not used to develop the theoretical framework, formulate mathematical claims or hypotheses, provide substantive steps in proofs, design the research methodology or experiments, generate or clean datasets, or interpret experimental results.

Any AI-assisted code was manually reviewed, integrated, and tested by the authors against the intended algorithms and experimental protocols. All AI-assisted text and code were critically reviewed by the authors. The scientific claims, proofs, numerical results, citations, and implementation details were independently checked by the authors, who take full responsibility for the final content of the paper.

References

Mathieu Blondel, Olivier Teboul, Quentin Berthet, and Josip Djolonga. Fast differentiable sorting and ranking. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 950–959. PMLR, 2020. URL https://proceedings.mlr.press/v119/blondel20a.html.

Stefan Chmiela, Alexandre Tkatchenko, Huziel E. Sauceda, Igor Poltavsky, Kristof T. Schütt, and Klaus-Robert Müller. Machine learning of accurate energy-conserving molecular force fields. *Science Advances*, 3(5):e1603015, 2017. doi: 10.1126/sciadv.1603015.

Taco Cohen and Max Welling. Group equivariant convolutional networks. In *Proceedings of the 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pages 2990–2999. PMLR, 2016. URL https://proceedings.mlr.press/v48/cohenc16.html.

Nadav Dym, Hannah Lawrence, and Jonathan W. Siegel. Equivariant frames and the impossibility of continuous canonicalization. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pages 12228–12267. PMLR, 2024. URL https://proceedings.mlr.press/v235/dym24a.html.

Peter Eastman, Pavan Kumar Behara, David L. Dotson, Raimondas Galvelis, John E. Herr, Josh T. Horton, Yuezhi Mao, John D. Chodera, Benjamin P. Pritchard, Yuanqing Wang, Gianni De Fabritiis, and Thomas E. Markland. SPICE, a dataset of drug-like molecules and peptides for training machine learning potentials. *Scientific Data*, 10(1):11, 2023. doi: 10.1038/s41597-022-01882-6.

Sékou-Oumar Kaba, Arnab Kumar Mondal, Yan Zhang, Yoshua Bengio, and Siamak Ravanbakhsh. Equivariance with learned canonicalization functions. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 15546–15566. PMLR, 2023. URL https://proceedings.mlr.press/v202/kaba23a.html.

Jinwoo Kim, Tien Dat Nguyen, Ayhan Suleymanzade, Hyeokjun An, and Seunghoon Hong. Learning probabilistic symmetrization for architecture agnostic equivariance. In *Advances in Neural Information Processing Systems*, volume 36, 2023. URL https://papers.nips.cc/paper_files/paper/2023/hash/3b5c7c9c5c7bd77eb73d0baec7a07165-Abstract-Conference.html.

George Ma, Yifei Wang, Derek Lim, Stefanie Jegelka, and Yisen Wang. A canonicalization perspective on invariant and equivariant learning. In *Advances in Neural Information Processing Systems*, volume 37, 2024. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/702b67152ec4435795f681865b67999c-Abstract-Conference.html.

Sebastian Prillo and Julian Eisenschlos. SoftSort: A continuous relaxation for the argsort operator. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 7793–7802. PMLR, 2020. URL https://proceedings.mlr.press/v119/prillo20a.html.

Omri Puny, Matan Atzmon, Edward J. Smith, Ishan Misra, Aditya Grover, Heli Ben-Hamu, and Yaron Lipman. Frame averaging for invariant and equivariant network design. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=zIUyj55nXR.

Huilin Qu, Congqiao Li, and Sitian Qian. Particle transformer for jet tagging. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 18281–18292. PMLR, 2022.

David Schubert, Tobias Goll, Nikolaus Demmel, Vladyslav Usenko, Joerg Stueckler, and Daniel Cremers. The TUM VI benchmark for evaluating visual-inertial odometry. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1680–1687, 2018. doi: 10.1109/IROS.2018.8593419.

Pei Sun, Henrik Kretzschmar, Xerxes Dotiwalla, Aurélien Chouard, Vijaysai Patnaik, Paul Tsui, James Guo, Yin Zhou, Yuning Chai, Benjamin Caine, Vijay Vasudevan, Wei Han, Jiquan Ngiam, Hang Zhao, Aleksei Timofeev, Scott Ettinger, Maxim Krivokon, Amy Gao, Aditya Joshi, Yu Zhang, Jonathon Shlens, Zhifeng Chen, and Dragomir Anguelov. Scalability in perception for autonomous driving: Waymo Open Dataset. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2446–2454, 2020.

Behrooz Tahmasebi and Stefanie Jegelka. Regularity in canonicalized models: A theoretical perspective. In *Proceedings of the 28th International Conference on Artificial Intelligence and Statistics*, volume 258 of *Proceedings of Machine Learning Research*, pages 4789–4797. PMLR, 2025. URL https://proceedings.mlr.press/v258/tahmasebi25a.html.

Bastiaan S. Veeling, Jasper Linmans, Jim Winkens, Taco Cohen, and Max Welling. Rotation equivariant CNNs for digital pathology. *arXiv preprint arXiv:1806.03962*, 2018.

Yi Yu, Tengyao Wang, and Richard J. Samworth. A useful variant of the davis–kahan theorem for statisticians. *Biometrika*, 102(2):315–323, 2015. doi: 10.1093/biomet/asv008. URL https://doi.org/10.1093/biomet/asv008.

A Core proofs

We retain only the arguments needed to verify the main theorems; auxiliary examples and alternative frame constructions are omitted.

A.1 Sections, compactness, and fixed stability budgets

Proof of Proposition 1. If $C : K \to K$ is invariant, $s([x]) = C(x)$ is well-defined on K/G and $s \circ \pi = C$. The universal property of the quotient topology makes s continuous, and orbit membership gives $\pi \circ s = \mathrm{id}$. In the other direction, a continuous section s defines $C = s \circ \pi$; this map is continuous, constant on orbits, and satisfies $\pi(C(x)) = \pi(x)$. □

Proof of Theorem 1. The maps take values in compact K and are equicontinuous, so Arzelà–Ascoli gives a subsequence f_{n_j} converging uniformly to a continuous $C : K \to K$. For $x \in K$ and $g \in G$,

$$d(C(gx), C(x)) \le 2\|f_{n_j} - C\|_\infty + \mathrm{Inv}_K(f_{n_j}) \longrightarrow 0,$$

hence C is invariant. Because distance to the compact set Gx is 1-Lipschitz,

$$d(C(x), Gx) \le \|f_{n_j} - C\|_\infty + \mathrm{Orb}_K(f_{n_j}) \longrightarrow 0.$$

Thus $C(x) \in Gx$, so C is an exact continuous canonicalizer. □

Proofs of Corollaries 1 and 2. A common Lipschitz bound makes a family equicontinuous, so vanishing defects would contradict Theorem 1 on an obstructed domain. For the fixed-budget claim, if $\eta_K(L) = 0$, choose f_n with $\mathrm{Lip}(f_n) \le L$ and both defects below $1/n$ and apply the same argument. □

A.2 The finite-group population rate

We use Riemannian distance d_K and normalized volume ν on K. Because K is a compact connected C^2 submanifold of Euclidean space, its intrinsic and ambient metrics are globally bi-Lipschitz equivalent: there is a constant $c_K < \infty$ such that $\|x-y\| \le d_K(x,y) \le c_K\|x-y\|$ for all $x, y \in K$. Hence, for Euclidean-valued maps, $\mathrm{Lip}_K(f) \le \mathrm{Lip}_E(f) \le c_K \mathrm{Lip}_K(f)$. The choice of d_K in Theorem 2 therefore changes the constants in the stability budget but not the $1/L$ exponent; empirical Euclidean/Frobenius sensitivities should not be identified numerically with Lip_K without this constant.

Uniform versus population approximation. The free finite-action setting also gives an L-independent uniform error floor, which is distinct from the expected-error rate proved below. Let $\varepsilon_{\mathrm{orb}} = \mathrm{Orb}_K(f)$ and $\varepsilon_{\mathrm{inv}} = \mathrm{Inv}_K(f)$ for a continuous map $f : K \to E$, with all output distances Euclidean, and let $\Delta = \min_{x \in K,\, g \ne e} \|gx - x\| > 0$. If $2\varepsilon_{\mathrm{orb}} + \varepsilon_{\mathrm{inv}} < \Delta$, then $\varepsilon_{\mathrm{orb}} < \Delta/2$ and the nearest orbit point to $f(x)$ is unique. Calling it $C_f(x)$ gives a continuous on-orbit map. For any $h \in G$,

$$\|C_f(hx) - C_f(x)\| \le 2\varepsilon_{\mathrm{orb}} + \varepsilon_{\mathrm{inv}} < \Delta,$$

so the two orbit points must coincide. Thus C_f would be a continuous exact canonicalizer, a contradiction. Consequently $2\,\mathrm{Orb}_K(f) + \mathrm{Inv}_K(f) \ge \Delta$ and $\max\{\mathrm{Orb}_K(f), \mathrm{Inv}_K(f)\} \ge \Delta/3$. The $\Theta(1/L)$ law below therefore does not describe simultaneous uniform decay of both defects. It describes how, under exact invariance, a fixed-amplitude orbit error can be confined to a population layer whose mass shrinks like $1/L$.

Lemma 1 (Uniform boundary-tube growth)**.** *For every $a \in (0, 1/2]$ there are $c_0, t_0 > 0$ such that every open $A \subset K$ with $a \le \nu(A) \le 1 - a$ satisfies*

$$\nu(N_t(\partial A)) \ge c_0 t \qquad (0 < t \le t_0).$$

Proof. The isoperimetric profile of a compact connected manifold has a positive minimum I_* on $[a, 1 - a/2]$. Put $A_s = \{x : d_K(x, A) < s\}$. If $\nu(N_t(\partial A)) \ge a/2$, the result follows after reducing c_0. Otherwise $\nu(A_s) \in [a, 1 - a/2]$ for $0 < s \le t$. The coarea formula gives $\frac{d}{ds}\nu(A_s) = \mathrm{Per}(A_s) \ge I_*$ almost everywhere. Integrating and using $A_t \setminus A \subset N_t(\partial A)$ yields $\nu(N_t(\partial A)) \ge I_* t$. □

Proof of Theorem 2. Let $q = |G| \geq 2$. Freeness, finiteness, and compactness give a positive minimum orbit separation

$$\Delta = \min_{x \in K,\, g \neq e} \|gx - x\| > 0, \qquad \rho = \Delta/2.$$

Fix an exactly invariant f with $\mathrm{Lip}_K(f) \leq L$ and set $r(x) = d(f(x), Gx)$. Orthogonality implies

$$|r(x) - r(y)| \leq (L+1) d_K(x, y). \tag{A.1}$$

Let $B = \{r \geq \rho\}$ and $\Omega = K \setminus B$. At each point of Ω there is a unique closest member of Gx to $f(x)$. For $g \in G$, let A_g be the open region where that member is gx. The A_g partition Ω, $\partial A_g \subset B$, and exact invariance gives $hA_g = A_{gh^{-1}}$. Hence

$$\nu(A_g) = \frac{1 - \nu(B)}{q}. \tag{A.2}$$

If $\nu(B) \geq 1/2$, then $p \geq m$ directly gives $\mathcal{E}_P(f) \geq m\rho/2$. Otherwise every A_g has volume between $1/(2q)$ and $1/2$. Lemma 1, applied to A_e, and $\partial A_e \subset B$ imply

$$\nu(N_t(B)) \geq c_0 t \qquad (0 < t \leq t_0).$$

Choose $t_L = \min\{t_0, \rho/[2(L+1)]\}$. By (A.1), $r \geq \rho/2$ on $N_{t_L}(B)$, and therefore

$$\mathcal{E}_P(f) \geq \frac{m\rho}{2} \nu(N_{t_L}(B)) \geq \frac{c}{L+1}.$$

This proves the lower bound after taking the infimum over f.

For the upper bound, $Q = K/G$ is a compact smooth manifold and $K \to Q$ a finite covering. Triangulate Q and let S be its codimension-one skeleton. Its lift $\Sigma \subset K$ is invariant and piecewise smooth, so for small s,

$$\nu(N_s(\Sigma)) \leq C_\Sigma s. \tag{A.3}$$

Choose one lift over each open top-dimensional simplex. This gives an invariant canonicalizer C_R on $R = K \setminus \Sigma$; on each component it is a fixed orthogonal group action. For a 1-Lipschitz cutoff χ that is zero on $[0,1]$ and one on $[2, \infty)$, define

$$f_t(x) = \chi(d_K(x, \Sigma)/t) C_R(x) \quad (x \in R), \qquad f_t(x) = 0 \quad (x \in \Sigma).$$

It is exactly invariant. If $R_K = \sup_{x \in K} \|x\|$, comparisons within a component give Lipschitz constant at most $1 + R_K/t$; comparisons across components cross the zero collar and give at most R_K/t. Thus $\mathrm{Lip}_K(f_t) \leq A/t$ for a constant A depending only on K.

For completeness, the cross-component estimate can be made explicit. Put $r_x = d_K(x, \Sigma)$ and $\alpha_x = \chi(r_x/t)$. Since distance to Σ is 1-Lipschitz and χ is 1-Lipschitz, points x, y in the same component satisfy

$$\|f_t(x) - f_t(y)\| \leq \|x - y\| + R_K |\alpha_x - \alpha_y| \leq (1 + R_K/t) d_K(x, y).$$

If x and y lie in distinct components of $K \setminus \Sigma$, every path joining them meets Σ, hence $d_K(x, y) \geq r_x + r_y$. Moreover $0 \leq \alpha_x \leq r_x/t$ and similarly for y, so

$$\|f_t(x) - f_t(y)\| \leq R_K(\alpha_x + \alpha_y) \leq \frac{R_K}{t}(r_x + r_y) \leq \frac{R_K}{t} d_K(x, y).$$

This supplies the mixed-component step in the stated A/t bound.

Outside $N_{2t}(\Sigma)$ the map is orbit-valued, while everywhere

$$d(f_t(x), Gx) \leq R_K \mathbf{1}_{N_{2t}(\Sigma)}(x).$$

Using $p \leq M$ and (A.3),

$$\mathcal{E}_P(f_t) \leq 2 R_K M C_\Sigma t.$$

Taking $t = A/L$ for sufficiently large L yields $\eta_P^{\mathrm{inv}}(L) \leq C/L$. □

A.3 Sharp antipodal calibration

Let $u_\theta = (\cos\theta, \sin\theta)$ and

$$X_\theta = \frac{1}{\sqrt{2}}[u_\theta, -u_\theta], \qquad K_{\mathrm{dbl}} = \{X_\theta : 0 \le \theta < 2\pi\}.$$

Column exchange sends X_θ to $X_{\theta+\pi} = -X_\theta$, and $\|X_\theta - X_\varphi\|_F = 2|\sin((\theta-\varphi)/2)|$. For the uniform population, let $\bar\eta(L)$ be the optimal expected orbit error among maps f satisfying $f(-X) = f(X)$ and $\mathrm{Lip}(f) \le L$.

Proof of Theorem 3. For any admissible f, put $r(\theta) = d_F(f(X_\theta), \{\pm X_\theta\})$. Some $r(\theta_0) \ge 1$: otherwise the closest orbit point is unique and its sign is continuous, hence constant on the circle, contradicting invariance under $\theta \mapsto \theta + \pi$. Moreover, r is $(L+1)$-Lipschitz in angular distance and is π-periodic. The two triangular profiles around θ_0 and $\theta_0 + \pi$ therefore give

$$\bar\eta(L) \ge \frac{1}{\pi(L+1)}. \tag{A.4}$$

For the upper bound define

$$c_L(s) = \mathrm{clip}(Ls, -1, 1), \qquad f_L(X_\theta) = c_L(\cos\theta) X_\theta.$$

Oddness of c_L makes f_L exactly invariant. On each unsaturated arc the double-angle identity gives Lipschitz ratio at most L; on saturated arcs the ratio is one; The remaining mixed-arc case needs a separate argument. For $L \ge 1$, choose lifts with shortest angular separation $\delta = |\theta - \varphi| \le \pi$, and put $h = \delta/2$, $m = (\theta+\varphi)/2$, $a = c_L(\cos\theta)$, and $b = c_L(\cos\varphi)$. Then

$$\|aX_\theta - bX_\varphi\|_F^2 = (a-b)^2 + 4ab\sin^2 h, \qquad \|X_\theta - X_\varphi\|_F = 2\sin h.$$

If both points are unsaturated, substituting $a = L\cos(m+h)$ and $b = L\cos(m-h)$ gives a squared quotient $L^2\cos^2 h \le L^2$. If both are saturated with the same sign, the quotient is 1. If they are saturated with opposite signs, write $\theta_* = \arccos(1/L)$. Their separation obeys $h \ge \pi/2 - \theta_*$, so the quotient is $\cot h \le \tan\theta_* = \sqrt{L^2-1} \le L$. Finally, suppose one point is saturated and the other is not. By symmetry take $a = 1$ and fix the unsaturated value b. The squared quotient is

$$R^2(c) = \frac{1 + b^2 - 2bc}{2 - 2c}, \qquad c = \cos\delta,$$

with $dR^2/dc = 2(1-b)^2/(2-2c)^2 \ge 0$. Thus the worst case is the smallest angular separation, attained when the saturated point is moved to the boundary θ_*; at that boundary the preceding unsaturated calculation applies. Hence every pair has quotient at most L. Hence $\mathrm{Lip}(f_L) \le L$. Its orbit error is $(1 - L|\cos\theta|)_+$, so direct integration yields

$$\bar\eta(L) \le \frac{2}{\pi}\left(\arcsin\frac{1}{L} - L + \sqrt{L^2-1}\right). \tag{A.5}$$

The right side is $1/(\pi L) + O(L^{-3})$, while (A.4) is $1/(\pi L) + O(L^{-2})$. The squeeze theorem gives $\lim_{L\to\infty} L\bar\eta(L) = 1/\pi$. □

A.4 Localized relaxation and selector-ambiguity mass control

Here Σ is a construction-specific selector-ambiguity (or cut) set, not necessarily the nonfree locus of the group action. In particular, the finite-group theorem above assumes a free action everywhere on K, yet its upper-bound construction still requires an invariant cut Σ. For Proposition 2, Σ is taken to be closed and G-invariant and the cutoff $\chi : [0,\infty) \to [0,1]$ is continuous.

Proof of Proposition 2. Invariance of Σ and orthogonality give $d(gx, \Sigma) = d(x, \Sigma)$. Thus the cutoff and C_R are invariant, so f_t is invariant. It is continuous on R. If $x_0 \in \Sigma$, then $d(x, x_0) < t$ implies $d(x, \Sigma) < t$ and hence $f_t(x) = 0$; therefore it is continuous at Σ as well. The defining properties of χ give $f_t = C_R$ outside $N_{2t}(\Sigma)$ and $f_t = 0$ on $N_t(\Sigma)$. □

Proposition 2 is qualitative: continuity of C_R alone does not imply a general Lipschitz bound for f_t, nor a general rule that increasing t lowers a Lipschitz constant. The $O(1/t)$ estimate in Theorem 2 uses the additional fact that its piecewise canonicalizer is a fixed orthogonal group action on each component. Any quantitative "larger t means smaller Lipschitz scale" statement is therefore restricted to constructions with corresponding regularity bounds.

Proof of Theorem 4. For $x \in R$ write $\alpha = \chi(d(x,\Sigma)/t)$. Orthogonality and $C_R(x) \in Gx$ give $\|C_R(x)\| = \|x\|$, and the reverse triangle inequality shows the exact identity

$$d(\alpha C_R(x), Gx) = (1-\alpha)\|x\|.$$

On Σ the error is $\|x\|$. Since $1-\alpha$ vanishes outside $N_{2t}(\Sigma)$,

$$d(f_t(x), Gx) \le \|x\| \mathbf{1}_{N_{2t}(\Sigma)}(x).$$

Taking expectations and using $\|X\| \le B$ proves the claim. □

Proof of Corollary 3. If $x \in N_{2t}(\Sigma_j)$, choose $y \in \Sigma_j$ with $d(x,y) < 2t$. Since $s_j(y) = 0$ and s_j is C_j-Lipschitz, $s_j(x) < 2C_j t$. Therefore

$$N_{2t}(\Sigma) = \bigcup_{j=1}^{J} N_{2t}(\Sigma_j) \subseteq \bigcup_{j=1}^{J} \{x : s_j(x) < 2C_j t\}.$$

Theorem 4 followed by the union bound proves the result. □

B Reproducibility and numerical details

This section records data provenance, uncertainty units, and the numerical values needed to audit the claims in the main text.

Unless a quantity is explicitly labeled as a standard error or a confidence interval, every $\pm$ attached to a learned-model summary denotes the mean plus/minus the sample standard deviation across seeds $0, 1, 2$ (ddof $= 1$). These three-seed spreads are descriptive uncertainty summaries, not standard errors, confidence intervals, or significance tests.

This appendix separates two reproducibility goals. Sections B.1–B.3 record data provenance, sampling/evaluation units, thresholds, splits where used, uncertainty aggregation, and the numerical contrasts cited in the main text. Appendix C separately records optimizer, update-count, width, seed, and checkpoint-reuse details for the controlled learned constructions. The real-data sections are intended as auditable mechanism studies rather than architecture benchmarks.

B.1 TUM VI quaternion population

We use only the official EuRoC/DSO 512×512 room1–room6 archives and retain `mav0/mocap0/data.csv`. Quaternion ordering is read from each CSV header; samples are renormalized. Sequence sizes are 16,541, 15,477, 15,200, 13,075, 16,999, and 15,244, totaling 92,536 poses. Subsampling at 10 Hz yields 7,786 KDE centers. For each $\kappa \in \{25, 50, 100\}$ we draw 10^6 independent samples and evaluate $t \in \{1/2, 1/4, 1/8, 1/16, 1/32, 1/64, 1/128\}$, corresponding to $L_t \in \{3, 5, 9, 17, 33, 65, 129\}$. Antipodal density symmetry and $f_t(-q) = f_t(q)$ hold to reported numerical precision.

t	L_t	Mean error	SE	95% CI	$L_t E_t$
.5000000	3	.1834745	.0003099	[.1828671,.1840820]	.5504236
.2500000	5	.1034895	.0002454	[.1030086,.1039705]	.5174476
.1250000	9	.0545414	.0001834	[.0541819,.0549010]	.4908726
.0625000	17	.0276530	.0001331	[.0273921,.0279140]	.4701012
.0312500	33	.0138990	.0000954	[.0137121,.0140860]	.4586684
.0156250	65	.0069912	.0000680	[.0068579,.0071245]	.4544299
.0078125	129	.0035167	.0000484	[.0034218,.0036115]	.4536493

Table 2: Primary smooth-population results ($\kappa = 50$). Intervals are Monte Carlo intervals conditional on the fitted KDE.

The large-L fits over $L \geq 17$ have slopes -1.027046, -1.017131, and -1.017413 for $\kappa = 25, 50, 100$, respectively. Their scaled errors at $L = 129$ are .403818, .453649, and .513565. The primary slope has SE .005509 and $R^2 = .999941$. The raw empirical control uses a sequence-stratified circular moving-block bootstrap with 2,000 replicates and block length 120. At $L = 129$ its mean error is .008323, 95% CI [.005920, .011252], and scaled error 1.0736.

B.2 Waymo protocols and audit statistics

Vehicle geometry uses 347,775 observations from 4,067 tracks; pedestrians use 72,183 observations from 768 tracks. Temporal analyses contain 315,900 and 64,156 transitions. Crops are expressed in annotated box-local coordinates to remove physical heading change. The yaw score is the horizontal covariance eigengap divided by total horizontal variance; the sign score is the magnitude of the equivariant odd moment divided by object scale. "Near" means score below .05 and "far" above .2. Confidence intervals resample tracks.

t	Veh. eigengap	Ped. eigengap	Veh. sign	Ped. sign
.005	.000575%	.01108%	.6605%	1.334%
.010	.006038%	.02494%	1.338%	2.693%
.020	.01927%	.09975%	2.680%	5.277%
.050	.1291%	.6899%	6.706%	13.35%
.100	.5311%	2.659%	13.04%	26.04%
.200	2.134%	10.24%	24.82%	48.68%

Table 3: Waymo mass satisfying normalized degeneracy score $s < t$.

Pose changes over 30°/90°/150° occur for 13.53/10.55/8.53% of vehicle transitions and 36.62/28.36/21.65% of pedestrian transitions. Vehicle $> 90°$ rates near/far from the sign stratum are 43.05/4.95% (risk ratio 8.70); pedestrian rates are 45.75/19.30% (ratio 2.37).

Class	Channel	Object p99 near/far	Pose p99 near/far
Vehicle	Eigengap	111.0×	870.2×
Vehicle	Sign	170.8×	338.2×
Vehicle	Joint	185.7×	880.8×
Pedestrian	Eigengap	1.0×	1.0×
Pedestrian	Sign	28.7×	38.1×
Pedestrian	Joint	41.5×	165.1×

Table 4: Paired near/far sensitivity ratios at perturbation radius .2.

The retrieval split has 210,269 training and 79,049 test observations, split by log; no track or scene crosses the split. Query-positive pairs are at least one second apart. Across three seeds, hard PCA changes Recall@1 relative to no canonicalization by $-.0152 \pm .0053$ and sign relaxation at $t = .3$ by $-.1998 \pm .0040$. At $t = .3$, reference-run Recall@1 is .1216 near and .3915 far from the sign stratum.

B.3 Cross-domain audit statistics

SPICE geometry uses 136,886 conformers from 17,999 molecules; the diagnostic models are deliberately small and are not accuracy benchmarks. PatchCamelyon uses 40,000 training and 12,000 test patches under exact quarter turns. JetClass uses 40,000 jets and exact input permutations. Learned contrasts are paired across seeds $0, 1, 2$.

For the molecular diagnostic, $E_\theta(X)$ is the learned scalar energy and the reported force is obtained by differentiation, $F_\theta(X) = -\nabla_X E_\theta(X)$. Its sensitivity is a gradient-output diagnostic,

$$\widehat{S}_r^F(X) = \max_{\delta \in \mathcal{A}_r(X)} \frac{\|F_\theta(X+\delta) - F_\theta(X)\|_F}{\max\{\|\delta\|_F, \epsilon_{\text{num}}\}},$$

using the molecular perturbation search for that experiment. This quantity is separate from canonical-object or representation sensitivity $\widehat{S}^{\text{obj}}$. In particular, for the raw representation $Z(X) = X$, the Euclidean representation quotient is one, whereas the force map can have a large gradient-level sensitivity.

Domain	Method / comparison	Diagnostic	Task statistic
SPICE	Raw coordinates	force S_F p99 46.78	reference force MAE 6.282
SPICE	Hard PCA	force S_F p99 545.5	paired MAE changes $+.607, -.190, +.251$
PatchCamelyon	Hard exact C_4	repr. p95 1.00	repr.–prediction $\rho = .047$
PatchCamelyon	Continuous orientation	repr. p95 11.77	repr.–prediction $\rho = .377$
JetClass	Hard sorting	repr. p95 29.38	baseline comparison
JetClass	Soft sorting, $\tau = .2$	repr. p95 .97	Δaccuracy $+.0106 \pm .0055$

Table 5: Numerical values supporting the cross-domain claims.

Soft-sort p95 sensitivity is $24.66, 8.15, 2.26, .97, .65$ for $\tau = .001, .01, .05, .2, 1.0$, with permutation consistency 1.0 throughout. Permutation consistency here should not be read as literal preservation of the input particle set. Soft sorting replaces a discrete permutation by a soft mixing matrix, so the relaxed output can mix particle blocks and need not lie on the original permutation orbit. The supported JetClass conclusion is that, over the tested temperatures, the relaxation retains enough task-relevant information for classification while reducing the measured representation sensitivity. SPICE has $P(s_{\text{PCA}} < .02) = 1.329\%$ and $P(s_{\text{PCA}} < .1) = 8.16\%$; the sampled MD17 aspirin control has no conformer below .2. JetClass has median normalized minimum gap 2.86×10^{-4} and $P(\Delta < 10^{-3}) = 78.3\%$.

B.4 Design and evaluation summary

Mechanism	Primary benefit	Typical cost	Minimum diagnostic
Hard selector	Orbit-valued compact representation	Tearing near branch boundaries	Temporal/perturbation tails
Continuous relaxation	Continuous invariant map	Off-orbit collapse in a transition collar	Orbit error and collapse
Orbit/frame average	Invariance without one branch	Compute and possible identity loss	Group-size scaling/task ablation
Probabilistic/multi-frame	Represents branch ambiguity	Sampling variance and compute	Calibration/sample-count sweep

Table 6: Design summary. Each mechanism resolves branch ambiguity by paying a different cost; diagnostics should target the quantity it sacrifices.

C Synthetic diagnostics of analytic and learned canonicalizers

These experiments complement the real-data diagnostics in Appendix B. They compare invariance, orbit fidelity, and canonical-object response on synthetic populations approaching selected ambiguities of analytic and learned canonicalizers. Fixed-radius searches provide the primary sensitivity evidence for the analytic $SO(3)$ comparison; a separate mixed-radius diagnostic covers the full set of methods.

C.1 Populations, constructions, and measurement protocols

Populations and realized degeneracy. The base pool contains 8192 clouds, with 1024 at each generator parameter

$$a \in \{1.0,\ 0.316,\ 0.1,\ 0.0316,\ 0.01,\ 0.00316,\ 0.001,\ 0\}.$$

Top-up sampling adds clouds until the near and far buckets reach their required counts, giving augmented pools of 12 595 clouds for $SO(3)$ and 38 888 for S_6. Each evaluation seed $(0, 1, 2)$ draws 2048 clouds uniformly without replacement from the corresponding augmented pool; all pool entries have equal weight. The $SO(3)$ evaluation inputs have 256 points in $\mathbb{R}^3$; the S_6 action permutes six coordinate blocks.

Training uses newly generated batches at every optimizer update, not samples from the evaluation pool. Evaluation uses a separately generated, frozen Stage A cache that is never accessed during training. No evaluation clouds are used for optimization: this is separation by generation, not a partition of one finite dataset.

The parameter $a = 0$ does not impose an exact sample-covariance multiplicity or sorting-key tie. The minimum realized normalized eigengap is 6.67×10^{-4} and the minimum normalized sorting-key gap is 3.26×10^{-5}. Thus the evaluated inputs contain neither exact covariance multiplicities nor exact sorting-key ties; this experiment does not test the selectors' behavior on those exact strata.

The eigengap is normalized by total variance, the odd-moment sign statistic by object scale with consistent point-count normalization, and the minimum adjacent sorting-key gap by object scale. As in Appendix B.2, *near* means a score below 0.05 and *far* a score above 0.20. All evaluated $SO(3)$ clouds are near in the sign channel. Accordingly, their near/far comparisons use *only the eigengap*; S_6 comparisons use the sorting-key gap. The two populations have separate no-canonicalization references and are not put on a common numerical scale merely by using the same norm.

Defects and aggregation. Object distances and perturbation norms are Frobenius norms on the full $N \times 3$ array, equivalently Euclidean norms after flattening. For a canonical-object map C,

$$\mathrm{Inv}(x) = \sup_{g \in G} \|C(gx) - C(x)\|_F, \qquad \mathrm{Orb}(x) = \inf_{g \in G} \|C(x) - gx\|_F.$$

Invariance is estimated over 16 Haar draws, augmented for finite groups by an evenly subsampled grid of at most 48 elements. Orbit error uses Kabsch alignment for $SO(3)$ and exact minimization over S_6; the latter does not mean that invariance was exhaustively evaluated over all 720 permutations. Within each seed, $\widehat{\text{Inv}}$ and $\widehat{\text{Orb}}$ are cloud-wise means (`nanmean`), whereas sensitivity is a cloud-wise p_{95}. The defect audit found zero NaNs in 6144 cloud evaluations per method, so no observations were removed by `nanmean`. This count concerns defects, not deliberately undefined frame diagnostics or zero-over-zero ratios.

Collapse is the fraction with $\sigma_{\min}(F(x)) < 0.1$, for the returned 3×3 frame or 6×6 soft permutation matrix. Pose regression returns a matrix and has measured collapse zero; hard sorting returns no frame. Near/far orbit ratios compare conditional means, and near/far sensitivity ratios compare conditional p_{99} values. Each ratio is formed inside its seed before averaging; zero-over-zero orbit ratios remain undefined. Per-seed statistics are averaged over the three seeds. Printed $\pm$ values are sample standard deviations (ddof $= 1$), not confidence intervals. Means are retained even for high-dispersion rows, with seedwise min–max ranges added in brackets. A mean without $\pm$ has no displayed standard deviation; it is not a zero-variance claim.

Two finite-search sensitivity diagnostics. We distinguish the candidate-search protocol from the sensitivity definition in Section 5.1. For a generated perturbation δ and requested radius ρ, the implemented discrepancy quotient is

$$q_C(x; \delta, \rho) = \frac{\|C(x+\delta) - C(x)\|_F}{\max\{\rho, \epsilon_{\text{num}}\}}, \qquad \|\delta\|_F \le \rho.$$

The two reported diagnostics are

$$\widehat{S}^{\text{obj}}_{\text{mix},r}(x) = \max_{(\delta,\rho)\in\mathcal{A}^{\text{mix}}_r(x)} q_C(x; \delta, \rho), \qquad \widehat{S}^{\text{obj}}_{\text{fix},r}(x) = \max_{\delta\in\mathcal{D}_r(x)} q_C(x; \delta, r).$$

The mixed-radius search samples requested radii log-uniformly up to $r = 0.2$; the fixed-radius search spends its full search budget at one requested radius. Each uses eight restarts of 40 steps, with clouds processed in chunks of 512, and $\epsilon_{\text{num}} = 10^{-4}$. Perturbations are clipped as $\delta \leftarrow \min\{1, \rho/\|\delta\|_F\}\delta$. Analytic methods use double precision and learned methods the single precision of their weights. The finite candidate sets are not assumed to be nested.

In a paired normalization check, all reported quantiles of $\|\delta\|_F/\rho$ were 1.0000, and 100% of ratios exceeded 0.99. Recomputing the summary on the same perturbations with the perturbation-norm denominator gave the same result to four decimal places. The formulas above retain the requested-radius normalization used by the implementation.

A separate reconstruction check measured the difference between the two represented inputs, including rounding after perturbation addition:

$$\delta_{\text{rec}} = \text{fp64}(\text{fp32}(x+\delta)) - \text{fp64}(\text{fp32}(x)),$$
$$e_{\text{rec}} = \frac{\|\delta_{\text{rec}} - \delta\|_F}{\|\delta\|_F}.$$

Here fp32 rounds to single precision and fp64 promotes the result for double-precision subtraction. The reported relative error is 0.0099 (0.99%) at $\|\delta\|_F = 10^{-4}$ and 0.001 at 10^{-3}. This calibrates the chosen numerical floor; it is distinct from the denominator comparison and does not establish equivalence of the two finite searches.

Analytic constructions. Hard PCA selects an oriented covariance frame. The localized family uses a score-based attenuation with width t, motivated by the construction in Proposition 2. Hard sorting orders the blocks by scalar keys; exact ties are resolved by the sorting routine and are outside the evaluated strict-order inputs. Soft sorting uses

$$(P_\tau)_{ij} = \text{softmax}_j\left(-|s_{(i)} - s_j|/\tau\right).$$

Away from ties it approaches hard sorting as $\tau \to 0$. Learned-key sorting retains this relaxation while learning the keys.

Learned frame constructions. Unconstrained pose and matrix regression control for learning without imposing the frame transformation law. For row-stacked centered points X_c, we use

$$g \cdot X = Xg^\top, \qquad C(X) = X_c F(X)^\top, \qquad F(g \cdot X) = F(X)g^\top.$$

The last equality makes C invariant. The invariant-gain vector heads preserve vector channels and compute scalar gains from their norms, with rigid and relaxed versions of the returned frame.

For eigenframe selection, let $U(X)$ be a covariance eigenbasis with $\det U = +1$, and let M_j be the 24 proper signed permutation matrices. The candidate row frames are $F_j(X) = M_j U(X)^\top$. On the simple-spectrum locus they transform up to reindexing:

$$F_j(g \cdot X) = F_{\pi_g(j)}(X)g^\top.$$

The scorer receives covariance eigenvalues and axis moments computed in each candidate's coordinates, with one shared network across candidates:

$$\phi_j(X) = \big(\lambda(X), \mathrm{mom}(X_c F_j(X)^\top)\big), \qquad \ell_j(X) = h_\theta(\phi_j(X)).$$

Since $(Xg^\top)_c F_j(g \cdot X)^\top = X_c F_{\pi_g(j)}(X)^\top$, both the features and the logits undergo the same reindexing as the frames. An implementation check, after enforcing $\det U = +1$ and matching candidates, found candidate-frame discrepancies within 2×10^{-10} and feature discrepancies within 5×10^{-11}. Thus

$$F_\tau(X) = \sum_j \frac{\exp(\ell_j(X)/\tau)}{\sum_k \exp(\ell_k(X)/\tau)} F_j(X)$$

has the required transformation law on the simple-spectrum locus. Hard selection takes the first maximizing candidate; its invariance guarantee requires a unique maximizer. Soft weighting smooths selection among the candidates, not the covariance eigenbasis at a repeated eigenvalue, and is not asserted to give a globally continuous frame.

Training and evaluation controls. Learned frame objectives combine PCA fitting and invariance losses, $\mathcal{L}_{\mathrm{pca}} + \mathcal{L}_{\mathrm{inv}}$. Relaxed rows use Σ-aware weighting and the prescribed target $\chi(s/t)C_{\mathrm{PCA}}(x)$. These rows therefore test the costs of *implementing an imposed relaxation*, not spontaneous discovery of its localization pattern. Learned runs use Adam at 10^{-3} and seeds $0, 1, 2$, without a learning-rate schedule, early stopping, or per-method tuning.

Each vector-feature and eigenframe run uses 60 epochs of 100 fresh batches, with one optimizer update per batch: 6000 updates at batch size 32 on 192-point clouds. The degeneracy parameter and batch are resampled at each update, giving 192 000 generated training clouds per run without replaying stored examples. Evaluation uses 256 points for $SO(3)$. Hidden widths are 32 for the invariant-gain vector heads and 64 for the shared eigenframe scorer. These are representation diagnostics, not prediction-accuracy benchmarks.

Both learned temperature sweeps hold the trained network fixed within each seed. Eigenframe temperatures $1, 10, 30$ share one checkpoint, while learned-key sorting at $\tau = 0.01, 0.1$ shares one key-network checkpoint. Matching recorded checkpoint hashes confirm the latter reuse. In each family only the evaluation-time temperature changes; no retraining is performed for the sweep.

C.2 Fixed-radius sensitivity and search coverage

Table 7 is the primary radius-resolved sensitivity comparison for the analytic $SO(3)$ methods. Each requested radius uses three seeds and 2048 clouds per seed, with relative across-seed standard deviations at most 0.06. The same requested radius and restart budget are used for each method within a column.

Search coverage. On the same 512 clouds in a paired hard-PCA check, the fixed-0.05 search found a larger quotient than the mixed-radius search up to 0.2 on 91.6% of clouds; their

Method	$\widehat{S}^{\mathrm{obj}}_{\mathrm{fix},r}$, p_{95} .001	.01	.05	.2	$p_{95}\, r$ .05	.2
No canonicalization, $SO(3)$	1.000	1.000	1.000	1.000	0.050	0.200
Hard PCA	10.54	12.03	259.8	69.1	12.99	13.82
Localized, $t = 0.05$	2.190	2.189	2.193	2.203	0.110	0.441
Localized, $t = 10^{-4}$	99.8	117.5	144.7	64.3	7.24	12.87

Table 7: **Primary fixed requested-radius diagnostic for analytic $SO(3)$ methods.** Column headings give r. The last two columns report r times the sensitivity quantile, an empirical output-discrepancy scale rather than a certified jump size. Quantiles need not be monotone in r. These are not the mixed-radius estimates in Table 8.

p_{95} values were 263 and 122, respectively. The total restart and step budgets were equal, but only 12.9% of mixed-radius draws fell in the audited neighborhood of 0.05, allocating about one of eight restarts there rather than eight. A larger admissible radius budget therefore did not give better finite-search coverage. In particular, the mixed-radius column is not an empirical upper envelope of the fixed-radius sweep. The fixed-radius results lead the analytic sensitivity analysis; mixed-radius results below remain explicitly protocol-specific and are not substituted for a converged local supremum.

The wide relaxation remains nearly flat (2.189–2.203) over the tested 200-fold radius range. Hard PCA instead reaches 259.8 at $r = 0.05$ and 69.1 at $r = 0.2$, giving discrepancy products 12.99 and 13.82. These are consistent with comparable branch-crossing displacements divided by different radii. The narrow relaxation also varies strongly with radius. Thus the fixed-radius sweep distinguishes the wide attenuation from the hard and narrow constructions without equating a large finite increment with a mathematical discontinuity.

At $r = 10^{-2}$, ordinary least squares on paired log–log observations from 512 clouds gives $\widehat{S}^{\mathrm{obj}}_{\mathrm{fix},r} \propto s_{\mathrm{eig}}^{-\widehat{\alpha}}$, with $\widehat{\alpha} = 1.05$ and a 95% percentile bootstrap interval $[1.01, 1.12]$. The bootstrap uses 1000 full-size paired case resamples with replacement and degree-one refits (seed 0); endpoints are the 2.5th and 97.5th percentiles. This fit concerns inverse-eigengap conditioning, not the population $1/L$ law of Theorem 2.

C.3 Joint diagnostics across constructions

Table 8 reports defects, score-conditioned orbit ratios, collapse, and the supplementary mixed-radius sensitivity diagnostic. The $SO(3)$ and S_6 no-canonicalization invariance defects are 0.574 and 1.14 ± 0.046, respectively; numerical comparisons are made within each action and population. The sensitivity column is labeled by its actual search protocol, not interpreted as the strongest search available.

Rigid selectors have zero orbit error but can produce large discrepancy quotients. Unconstrained pose regression also has invariance defect 0.494: a rigid output alone does not make canonicalization invariant. The relaxed unconstrained matrix combines a smaller observed quotient with orbit error 0.277 and collapse 0.999.

Dependence on the learned construction. The relaxed invariant-gain vector head has mixed-radius sensitivity mean 528 and seed range $[163, 902]$, compared with 34.8 ± 2.2 for eigenframe selection at $\tau = 1$. The three nonnegative eigenframe values sum to about 104.4, below the vector-head minimum, so their *reported diagnostic* ordering holds across the three seeds. Eigenframe selection also has lower mean orbit error and collapse. This is a contrast between the tested constructions under this finite search, not a search-independent stability ranking or evidence that all between-seed variation is caused by architecture.

For orbit-valued selectors, orbit error alone cannot expose branch response. The mixed-radius near/far p_{99} sensitivity ratios are 2.24 for hard PCA, 1639 for hard sorting, 2.30 for the rigid invariant-gain vector head, 4.25 for hard eigenframe selection, and 0.54 for pose regression. These are associations with the chosen input score under the stated search.

Method	$\widehat{\text{Inv}}$	$\widehat{\text{Orb}}$	Orbit near/far	$\widehat{S}^{\text{obj}}_{\text{mix},0.2}$ p_{95}	Collapse
Reference					
No canonicalization, $SO(3)$	0.574	~ 0	–	1.000	–
No canonicalization, S_6	1.14 ± 0.046	0	–	1.000	–
Analytic constructions					
Hard PCA, $SO(3)$	1.12×10^{-15}	0	–	126 ± 12	0
Localized, $t = 10^{-4}$	1.48×10^{-14}	0.109	2.48	102 ± 2.7	0.232
Localized, $t = 0.05$	6.34×10^{-16}	0.295	0.96	1.70 ± 0.01	1.00
Hard sorting, S_6	0	0	–	485 ± 20	–
Soft sorting, $\tau = 0.01$	2.2×10^{-17}	0.0246	324	9.44 ± 0.13	0.401
Soft sorting, $\tau = 0.1$	8.4×10^{-17}	0.0433	1.82	1.30 ± 0.01	0.476
Learned: unconstrained $SO(3)$ heads					
Pose regression (rigid)	0.494	0	–	147 ± 40	0
Matrix regression (relaxed)	0.156	0.277	0.93	4.73 ± 2.2	0.999
Learned: equivariant constructions on the regular locus					
Vector head, invariant gains (rigid)	2.94×10^{-5}	0	–	1227 ± 113	0
Vector head, invariant gains (relaxed)	$(1.62 \pm 1.7) \times 10^{-5}$	0.216	1.37	528 $[163, 902]$	1.00
Eigenvector selection (hard)	5.63×10^{-7}	0	–	125 ± 6.7	0
Eigenvector selection, $\tau = 1$	6.38×10^{-7}	0.148	7.83	34.8 ± 2.2	0.477
Eigenvector selection, $\tau = 10$	3.29×10^{-7}	0.197	2.50	12.3 ± 0.5	0.544
Eigenvector selection, $\tau = 30$	1.99×10^{-7}	0.253	1.37	4.96 ± 0.52	0.738
Learned soft sorting, $\tau = 0.01$	3.7×10^{-8}	0.0466	4.97	1.84 ± 0.16	0.507
Learned soft sorting, $\tau = 0.1$	9.2×10^{-8}	0.124	2.89	2.04 ± 1.55 $[1.13, 3.82]$	0.901

Table 8: **Joint diagnostics with supplementary mixed-radius sensitivity.** Inv and Orb are cloud-wise means; sensitivity is a cloud-wise p_{95}, followed by aggregation across three seeds. Printed $\pm$ values are sample standard deviations; brackets add seedwise ranges without replacing means. The vector-head sensitivity SD is not displayed. Orbit near/far uses the eigengap for $SO(3)$ and the key gap for S_6, not distance to the full selector-ambiguity set. Collapse is the fraction with $\sigma_{\min}(F) < 0.1$; dashes mark undefined ratios or unavailable frame diagnostics. The mixed-radius search has the coverage limitation established in Section C.2.

C.4 Relaxation settings and score-conditioned error

Table 9 records changes in the supplementary mixed-radius diagnostic together with orbit-error contrast and collapse. All four families show a lower near/far orbit ratio and more collapse at the wider setting. Three show a lower mean mixed-radius sensitivity; learned-key sorting shows an increase in its mean and a seed-dependent tail response.

Family	Setting change	Mean p_{95} $\widehat{S}^{\mathrm{obj}}_{\mathrm{mix},0.2}$	Orbit near/far	Collapse
Analytic localized, $SO(3)$	$t : 10^{-4} \to 0.05$	$102 \to 1.70$	$2.48 \to 0.96$	$0.232 \to 1.00$
Analytic soft sorting, S_6	$\tau : 0.01 \to 0.1$	$9.44 \to 1.30$	$324 \to 1.82$	$0.401 \to 0.476$
Learned-key soft sorting, S_6	$\tau : 0.01 \to 0.1$	$1.84 \to 2.04$	$4.97 \to 2.89$	$0.507 \to 0.901$
Learned eigenvector selection, $SO(3)$	$\tau : 1 \to 30$	$34.8 \to 4.96$	$7.83 \to 1.37$	$0.477 \to 0.738$

Table 9: **Relaxation settings, mixed-radius response, and conditional error.** All sensitivity endpoints are means of per-seed p_{95} values from Table 8; uncertainty and supplementary ranges are reported there. Both learned sweeps change only the evaluation-time temperature, reusing one checkpoint per seed within each family. Near/far ratios compare conditional mean orbit errors in score-defined bins.

For the analytic localized family, increasing t raises mean orbit error from 0.109 to 0.295 and collapse from 0.232 to 1.00. Its lower sensitivity is also visible in the primary fixed-radius sweep, rather than relying only on the mixed-radius endpoints. Analytic soft sorting similarly raises orbit error from 0.0246 to 0.0433 as its mixed-radius sensitivity falls. For the same eigenframe checkpoint, raising temperature from 1 to 30 lowers that diagnostic from 34.8 to 4.96, while orbit error rises from 0.148 to 0.253 and collapse from 0.477 to 0.738.

For learned-key sorting, the keys are held fixed within each seed while the evaluation-time temperature increases from 0.01 to 0.1. At $\tau = 0.1$, the three p_{95} values are 1.128, 1.155, and 3.822: their mean is 2.035, above 1.84 at $\tau = 0.01$ by about 11%, and their sample SD is 1.548. The narrower setting has seedwise p_{95} in $[1.66, 1.95]$, so two seeds improve and one worsens under this temperature change. Seedwise medians at $\tau = 0.1$ remain similar (0.853–0.856), but one seed puts 10.1% of clouds above sensitivity 2 and the other two put none. The mean increase and the lack of a seed-consistent improvement are both part of the result.

What the conditional ratios do not identify. Every $SO(3)$ cloud is near the sign boundary by the chosen threshold; even an eigengap-*far* cloud is not far from all analyzed selector-ambiguity channels. Consequently, ratios such as 0.96 or 1.37 describe eigengap-conditioned means within a sign-near population. They do not show loss of localization relative to the full selector-ambiguity set, nor uniform error across the population. Moreover, no comparison bound between the normalized scores and $d(x, \Sigma)$ is established. The metric-neighborhood localization in Proposition 2 is therefore not directly tested by these bins. The input scores also do not measure all learned ambiguities, such as logit ties, vanishing vector features, or learned-key margins. This experiment diagnoses score-conditioned error, not localization with respect to the complete selector-ambiguity geometry of every construction.

C.5 Conclusions and scope

The primary fixed-radius comparison shows that the wide analytic relaxation has a quieter canonical-object response than the hard and narrow selectors at the tested radii. The joint diagnostics show that such attenuation can coexist with substantial orbit error and loss of rigidity, while rigid learned outputs need not be invariant. The learned relaxed rows realize an imposed training target, while both learned temperature sweeps vary only the evaluation-time temperature of a fixed checkpoint within each seed. These are distinct experimental contrasts, not interchangeable evidence of an independently learned law.

All sensitivity values are finite-search diagnostics, not certified suprema or Lipschitz constants; the mixed-radius results do not establish search-independent rankings. The $SO(3)$

experiments do not test the free finite-group population rate of Theorem 2, and the score-conditioned comparisons do not test metric localization relative to the complete selector-ambiguity set. They also do not estimate the score Lipschitz constants C_j required to turn the occupancy curves into the numerical bound of Corollary 3. Three seeds do not separate training, population, and search contributions to variation. The qualitative relation to JetClass in Appendix B.3 concerns the costs of smoothing, not matched numerical scales or temperatures.